Beware of the DAG!

A. Philip Dawid APD@STATSLAB.CAM.AC.UK
Statistical Laboratory
University of Cambridge
Wilberforce Road
Cambridge CB3 0WB, UK

Editor: Isabelle Guyon, Dominik Janzing and Bernhard Schölkopf

Abstract

Directed acyclic graph (DAG) models are popular tools for describing causal relationships and for guiding attempts to learn them from data. In particular, they appear to supply a means of extracting causal conclusions from probabilistic conditional independence properties inferred from purely observational data. I take a critical look at this enterprise, and suggest that it is in need of more, and more explicit, methodological and philosophical justification than it typically receives. In particular, I argue for the value of a clean separation between formal causal language and intuitive causal assumptions.

Keywords: Directed acyclic graph, conditional independence, probabilistic causality, statistical causality, causal DAG, augmented DAG, intervention DAG, causal discovery, causal Markov condition, reification fallacy, instrumental variable

1. Introduction

This article is a philosophical excursion into the territory of `causal inference`[1] and directed acyclic graphs (DAGs). Like Cartwright (2007, Chapter II), it takes a sceptical attitude to the widespread view that we can learn about causal processes by constructing DAG models of observational data.

My fundamental concern is the relationship between, on the one hand, properties or concepts relating to an external reality, such as probabilistic independence or `causality`, which we wish to elucidate or manipulate; and, on the other hand, formal representations of such properties by means of mathematical or logical structures, such as graphs. It is important to avoid confusing the picture with the reality (the "reification fallacy"), and to be very clear about the nature and limitations of the relationship between the two.

2. Seeing and doing

Following Spirtes et al. (2000) and Pearl (2000), I accept the fundamental importance of distinguishing between the activities of *seeing* and *doing*.

Seeing involves passive observation of a system in its natural state. *Doing*, on the other hand, relates to its behaviour in a disturbed state, typically brought about by some external intervention. A case can be made (Dawid, 2000, 2002b) for regarding the otherwise philo-

1. Throughout, we use `teletype font` to highlight terms relating to informal `causal` concepts.

sophically problematic concept of `causation` as simply describing how the system responds to external intervention—a stripped-down "agency" or "manipulability" interpretation of causality (Hausman, 1998; Woodward, 2003). `Causal inference` then refers to the problem of drawing conclusions, from available data, about such responses.

Now it ought not to need saying, but is nonetheless worthy of continual repetition and emphasis, that there is absolutely no logical reason for there to be any connexion whatsoever between observations made under the different regimes of seeing and doing: a system may very well behave entirely differently when it is kicked than when it is left alone. "To find out what happens to a system when you interfere with it you have to interfere with it (not just passively observe it)"(Box, 1966). So any understanding one might achieve of the system's undisturbed behaviour is at best indirectly relevant to its disturbed behaviour, and thus to `causal inference`.

Often, however, for reasons of cost, practicality, ethics, *etc.*, we can not "interfere," but are confined to passive observation of the undisturbed system. We might then attempt to proceed by assuming connexions between the different regimes, that would allow us to transfer knowledge gained from *seeing* to inferences about the effects of *doing*. But it is important to be entirely explicit about such assumptions; to attempt, so far as is possible, to justify them; and to be fully aware of the sensitivity of any conclusions drawn to their validity.

In recent years there has grown up a body of methodology that purports to extract `causal` (doing) conclusions from purely observational (seeing) data in fairly automatic fashion. This largely revolves around directed acyclic graph (DAG) models, which have interpretations in both seeing and doing contexts. However, these interpretations, while related, are distinct, and it remains important to justify (when possible) the replacement of one interpretation by another. There is thus nothing the least bit automatic about such `causal discovery` algorithms. *Caveat emptor!*

3. Conditional independence

We start by concentrating on the behaviour, under a single stable regime, of a collection of variables of interest. This behaviour will be modelled by means of a fixed joint probability distribution P.[2] If we can obtain and record repeated observations under the same regime, we might hope to estimate P. Here we largely ignore problems of inference, and restrict attention to purely probabilistic properties.

One of the most important of such properties is that of *conditional independence*, CI (Dawid, 1979, 1980). We write $X \perp\!\!\!\perp Y \mid Z$ to denote that variables X and Y are probabilistically independent in their joint distribution given $Z = z$, for any observable value z of Z.

Universal qualitative properties of probabilistic CI include:

2. There are of course many interpretations of probability (Galavotti, 2005). For present purposes a naïve frequentist view, which can also be given a subjective Bayesian intepretation in terms of exchangeability (de Finetti, 1975), will suffice. Williamson (2005) argues for an "objective Bayesian" interpretation as most appropriate for `causal inference`. The formal mathematical framework is the same in all cases.

$$
\begin{array}{lll}
X \perp\!\!\!\perp Y \mid X & & \\
X \perp\!\!\!\perp Y \mid Z & \Rightarrow & Y \perp\!\!\!\perp X \mid Z \\
X \perp\!\!\!\perp Y \mid Z, \quad W \leq Y & \Rightarrow & X \perp\!\!\!\perp W \mid Z \\
X \perp\!\!\!\perp Y \mid Z, \quad W \leq Y & \Rightarrow & X \perp\!\!\!\perp Y \mid (W, Z) \\
\left.\begin{array}{l} X \perp\!\!\!\perp Y \mid Z \\ \text{and} \\ X \perp\!\!\!\perp W \mid (Y, Z) \end{array}\right\} & \Rightarrow & X \perp\!\!\!\perp (Y, W) \mid Z
\end{array}
\tag{1}
$$

(where $W \leq Y$ denotes that W is a function of Y). While these do not exhaust all such relationships (Studený, 1992), they are adequate for many statistical purposes.

4. Graphical representation

It can be helpful to use mathematical constructions of various kinds to represent and manipulate CI (Dawid, 2001). This works when we can make formal analogies between properties of probabilistic CI and certain (non-probabilistic) properties of the representations we use. The representations themselves can look very different from probability distributions, and we need to be very clear as to how we are to interpret properties of such a representation as "saying something about" properties of CI. As with any use of representations to assist understanding and construct arguments, the *semantics* (or *meaning*) of a representation—describing exactly just how it is to be taken as relating to the external "reality" it is intended to represent—is at least as important as its *syntax*—describing its internal grammar.

One of the most popular and useful of such representations is the directed acyclic graph (DAG): a full description and analysis of the formal semantics of the relationship between DAGs and the collections of CI properties they represent, together with the associated notation and terminology, can be found in Cowell et al. (2007). Although this theory will be familiar to many readers, I repeat here the specific features I wish to emphasise—more to clarify what is *not* being said than what is.

A DAG has nodes representing variables, and arrows joining them. Given a joint distribution over an ordered set of random variables $(V_1, \ldots, V_N)$, we can construct an associated DAG with the (V_i) as vertices by inserting arrows to V_{i+1} $(i = 0, \ldots, N-1)$ from the smallest subset, S_i say, of all earlier variables, $V^i = (V_1, \ldots, V_i)$, such that

$$V_{i+1} \perp\!\!\!\perp V^i \mid S_i. \tag{2}$$

As a simple example, Figure 1 shows the unique DAG over four variables (Z, U, X, Y) that represents the following pair of CI properties:

$$U \perp\!\!\!\perp Z \tag{3}$$

$$Y \perp\!\!\!\perp Z \mid (X,\ U). \tag{4}$$

It can be shown that a CI property $S \perp\!\!\!\perp T \mid U$ (where S, T, U are collections of variables) can be logically deduced from the collection of input properties (2) $(i = 0, \ldots, N-1)$ by

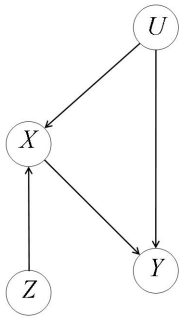

Figure 1: Simple DAG

repeated applications of the algebraic rules (1) if and only if the following purely geometric property of the associated DAG $\mathcal{D}$ is satisfied. First we delete all nodes that are not "ancestors" of some node in $S \cup T \cup U$, and all their incoming arrows; then we add undirected edges between any two nodes that are "parents" of a common "child" variable, if they are not already joined by an arrow; next we delete all arrowheads, so obtaining an undirected graph, the relevant *moralized ancestral graph*. Finally, in this graph we look for paths joining S and T that do not intersect U. If there are none such, we say that S and T are "d-separated"[3] by U in $\mathcal{D}$. In this case, the CI property $S \perp\!\!\!\perp T \mid U$ is indeed a logical consequence of the input assumptions (2) and the transformation rules of (1) (and so valid if our input assumptions are); if d-separation fails, then it is not a logical consequence, and will fail for at least one joint distribution that satisfies (2). We can thus use the DAG as a "theorem-proving machine" to streamline logical manipulations of probabilistic CI properties.

We say that a DAG $\mathcal{D}$ with node-set $\mathcal{V}$, a set of variables, *represents* a collection $\mathcal{C}$ of CI properties over $\mathcal{V}$ if the relation "$S \perp\!\!\!\perp T \mid U$" is in $\mathcal{C}$ if and only if S and T are d-separated by U in $\mathcal{D}$. This relationship between a $\mathcal{D}$ and a collection of CI properties will constitute our *semantic interpretation* of a DAG.

For given $\mathcal{V}$, the collections of CI properties $\mathcal{C}$ that are represented by some DAG are very special, comprising exactly those that can be derived, using (1), from an input collection of the recursive form (2). Thus with $\mathcal{V} = \{X, Y, Z\}$, the pair of properties $X \perp\!\!\!\perp Y$, $X \perp\!\!\!\perp Y \mid Z$ has no DAG representation, and this is indeed the typical state of affairs. Conversely, when a DAG representation is available, it need not be unique. Thus the three DAGs of Figure 2 all represent the CI property $A \perp\!\!\!\perp B \mid C$, and are all equally valid for this purpose. This flexibility is extended when we allow the set $\mathcal{V}$ of variables considered to vary: both DAGs of Figure 3 represent the single CI property $A \perp\!\!\!\perp B$.

3. We have in fact described the "moralization" version of this graph-separation property (Lauritzen et al., 1990); this is logically equivalent to the "d-separation" property as described by Pearl (1986); Verma and Pearl (1990).

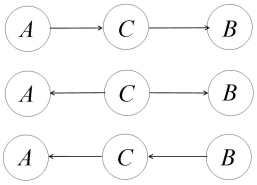

Figure 2: Three DAGs representing $A \perp\!\!\!\perp B \mid C$

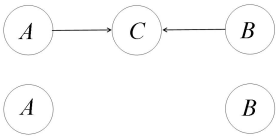

Figure 3: Two DAGs representing $A \perp\!\!\!\perp B$

5. Causal interpretation

According to the theory presented above, the single purpose of a DAG representation is to mirror, *via* the moralization/d-separation semantics described in § 4, the probabilistic relationship of conditional independence—a relationship that, it is worth emphasising, is entirely symmetrical, as is captured by the second line of (1). However, it is in the very nature, and indeed name, of a directed acyclic graph that it contains *directed* arrows between variables, so that its graphical structure necessarily embodies a non-symmetrical relationship between nodes. From the point of view I have so far taken, this is a pure artefact: thus Figure 1, although composed of directed arrows, is nothing but an alternative way of representing the symmetrical CI relationships (3) and (4). The rôle of an arrow in a DAG model is thus much like that of a construction line in an architect's drawing: although it plays an important rôle in the formal syntax of the model, it has no counterpart in the world, and does not contribute directly to the semantic interpretation of the model.

5.1 Reification

Nevertheless, it is very hard to resist the urge to interpret the arrow from node X to node Y in the DAG as representing something meaningful in the real-world system that the DAG is modelling—for example, as embodying some conception of the (non-symmetrical) relation of `cause and effect`: that X is, in some sense, a `direct`[4] `cause` of Y. Likewise, we might

4. This particular interpretation depends, not only on the directionality in the DAG representation, but on what variables we have chosen to include in it. For example, there might be an additional, as yet unmentioned, variable V, jointly distributed with (Z, U, X, Y) in a way that could be represented by inserting V between X and Y in Figure 1. The new DAG would still embody the CI properties (3) and (4) (as well as some further CI properties involving V), so that, as far as these variables are concerned, both DAGs represent their CI properties equally well. But there is an arrow from X to Y—and, thus, a putative interpretation of X as a "`direct cause`" of Y—in only one of them.

be tempted to read off from the graph such intuitive properties as "X `lies on the causal pathway` between Z and Y". But I again emphasise that no such inferences are justified from the formal semantics relating DAG representations to conditional independence. Such an interpretation of an incidental formal attribute of a mathematical representation of the world as corresponding to something real in the external (physical or mental) world[5] may be termed "reification". While reification can often be indicative and fruitful, it is important to be very clear as to when we are reaching beyond the formal semantics by which the representation has been supposed to encode real-world properties, and in that case to consider very carefully whether, when and how this might be justifiable.

There may be several alternative ("Markov equivalent") DAG representations of the same collection of CI statements for a given collection of variables. These must then have the same *skeleton* (undirected version) and *immoralities* (configurations of the form $A \to C \leftarrow B$ where there is no arrow between A and B) (Frydenberg, 1990). Some, but not all, of the directions on the arrows will be the same in all these Markov equivalent DAGs: the reificationalist temptation then is to regard these fixed directions, at least, as meaningful in some `causal` sense.

5.2 Causal DAGs

As just one among many examples of interpretations of DAG models that go beyond mere conditional independence properties, I quote from Hernán and Robins (2006) (their Figure 2 is redrawn here as our Figure 4), in discussion of a certain problem relating to the use of "instrumental variables" (see § 8 below):

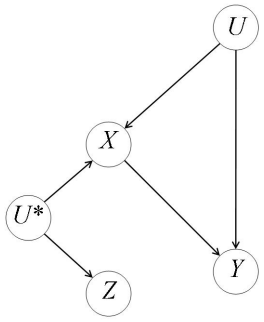

Figure 4: Figure 2 of Hernán and Robins (2006)

> "A `causal` DAG is a DAG in which:
>
> (i). the lack of an arrow from V_j to V_m can be interpreted as the absence of a `direct causal effect` of V_j on V_m `(relative to the other variables on the graph)`
>
> (ii). all `common causes`, even if unmeasured, of any pair of variables on the graph are themselves on the graph. In Figure 2. . . the inclusion of the measured

5. Bourdieu (1977, *p.* 29) speaks of "sliding from the model of reality to the reality of the model"

> variables (Z, X, Y) implies that the `causal` DAG must also include their unmeasured `common causes` (U, U^*)."

Here again, our use of `teletype font` highlights non-mathematical `causal concepts`, that are here supposed to correspond to certain purely graphical aspects of the structure of Figure 4. Whereas those graphical aspects can be stated clearly enough, the `causal concepts` that they are supposed to embody are, to say the least, in need of further clarification—they are certainly distinct from the pure CI properties that I have so far taken to be the purpose of a DAG representation. Until we have such clarification, it will be impossible to decide whether or not a given DAG is indeed a `causal` DAG for the problem it is intended to represent. It is important to realize that such a question can never be addressed solely in formal terms, by reference only to the DAG—that would be a circular argument. What is needed is an externally derived understanding of these `causal concepts`. Only if and when we have such an external `causal` understanding of our problem can we decide whether or not a given DAG is indeed a `causal` DAG representation for it. In particular, it can not be appropriate to define concepts such as `direct causal effect` or `common cause` purely in terms of a DAG model—unless that model has already been justified as "`causal`" by other, necessarily non-formal, considerations.

5.3 Probabilistic causality

One approach to causal interpretation, going back at least to Reichenbach (1956) and developed by Suppes (1970) among others, supposes the existence of connexions between causal properties and conditional independence properties. Essentially, this proceeds by equating probabilistic conditional independence of two variables X and Y given a "suitable" third variable Z with the `causal independence` of X and Y, in the sense that neither of these variables `causally affects` the other. It is hard to make this precise however. In one direction, it apparently implies that, if X and Y are completely independent probabilistically, then neither can `affect` the other. However, this degree of implication is not usually claimed, since such independence could be an accidental result of numerical cancellation between two or more non-null `causal` probabilistic relationships involving these and other variables.[6] Likewise, if we find that X and Y are not independent given any of the "suitable" variables Z currently under consideration, we can not immediately deduce `causal dependence` between X and Y, since we can not rule out the possibility that we have simply not examined enough Zs. And there again, how are we to delimit which variables Z are "suitable"? Certainly if Z occurs later in time (or in some other considered relevant `causal order`) than either X or Y, it would not generally be considered appropriate for us to take a CI property $X \perp\!\!\!\perp Y \mid Z$ as having causal implications for the relationship between X and Y—rather, this CI property would be ascribed to accidental cancellation. But there might

6. Such a state of affairs is sometimes dismissed as being due to a "non-faithful" DAG representation of the problem, where we have a DAG that is regarded as embodying (*via* d-separation) those CI properties that can indeed be regarded as `causal`, while there are still others, not graphically represented, that are "accidental" consequences of the conditional probability specification over the DAG. However, this begs the question of which of the CI relationships in the problem are to be regarded as `causal` in the first place.

well also be other, context-dependent, and less easily explicated, "suitability" constraints on Z.

In the converse direction, it has been claimed (the "weak causal Markov assumption", Scheines and Spirtes, 2008) that, if X and Y are `causally disconnected`, in the sense that neither X nor Y `causally affects` the other and they have no other `common cause` Z, then they must be probabilistically independent. This again is fully meaningful only in the presence of a pre-existing understanding of the `causal` concepts referred to—in which case it should be empirically testable (and what would one do if it were found to fail, as there is no obvious reason to exclude?). Contrapositively, if X and Y are found to be probabilistically dependent, the weak causal Markov assumption would imply, either that one of X and Y does `causally affect` the other, or that they have some other `common cause` Z. This implication could be regarded as supplying an interpretation for the specific, somewhat complex, `causal assertion` in the previous sentence, and thus as a constraint on any attempt at `causal interpretation` of the terms in it.

A fuller causal interpretation for CI and DAGs is *via* the "Causal Markov Condition", CMC (Spirtes et al., 2000) (or "Axiom", Scheines and Spirtes, 2008). This relies on an external concept of one variable being a `cause` of another (which is then an `effect` of the first), and of its special cases of `immediate` (or `direct`) `cause` and `effect`. CMC now asserts that a variable V is probabilistically independent of its `non-effects`, conditional on its `immediate causes`. Accordingly, a `causal` DAG representation is one where the graph parents of V represent its `immediate causes`, while its non-descendants in the graph represent `non-effects` of V. This conception appears close to that in the quotation from Hernán and Robins (2006) in § 5.2. CMC is closely related to the "Principle of the Common Cause" (Reichenbach, 1956).

A simple but nonetheless powerful criticism of CMC is "Why?". The assumed connexions between probability and causality seem arbitrary, and there are numerous telling counterexamples (Williamson, 2005; Cartwright, 2007). And its application is problematic.

CMC relates three different things: causality, probabilistic CI, and DAGs. Correspondingly, it can be construed in at least three distinct ways:

(i). Given an externally defined understanding of the terms `cause` and `immediate cause`, CMC supplies a way of constructing a "`causal` DAG" representation of a problem—which in turn implies various (testable—and so potentially falsifiable!) CI relationships between variables.

(ii). Given the conditional independence properties enjoyed by a collection of variables, and external understandings of the `causal` terms (supposed mutually compatible in accordance with (i) above), CMC acts as a constraint on a probabilistic DAG representation if it is to be interpretable as a `causal` DAG.

(iii). Finally, if we are given a probabilistic DAG that is asserted to be a `causal` DAG, CMC acts as a constraint on any purported interpretations of the `causal` terms employed.

Thus, even if accepted, CMC in no way provides a simple interpretation of `causality`. Rather, it constrains the relationships between three terms. We can not get any more out of it without putting more in: "No causes in, no causes out" (Cartwright, 1994, Chapter 2).

5.4 Causal discovery

The currently very active enterprise of `causal discovery` (Spirtes et al., 2000; Glymour and Cooper, 1999; Neapolitan, 2003) is based on assuming connexions (especially, CMC) between conditional independence and `causality` along the lines of those described in § 5.3 above. There are many variations, but all share the same basic philosophy. Essentially, one analyses observational data in an attempt to identify conditional independencies (possibly involving unobserved variables) in the distribution from which they arise. Some of these might be discarded as "accidental" (perhaps because they are inconsistent with an *a priori* `causal order`); those that remain might be represented by a DAG.[7] Finally this DAG is interpreted as, in some sense, a `causal` DAG. When there are several Markov equivalent DAG representations of the discovered CI relationships, the arrows common to all are interpreted `causally`.

5.5 Other interpretations

Recently other ideas have been suggested for determining directed relationships between variables solely on the basis of their joint probability distribution. For example, Janzing and Schölkopf (2008) aim to distinguish between the two decompositions $p(x, y) = p(x)\, p(y \mid x)$ and $p(x, y) = p(y)\, p(x \mid y)$ in terms of their computational complexity: if this favours, say, the former, then one might regard this as indicating a directional effect $X \to Y$. Similarly, if one can reasonably describe $p(y \mid x)$, but not $p(x \mid y)$, in terms of an implicit additive error structure (Zhang and Hyvärinen, 2008), one might again interpret that as implying a directional effect $X \to Y$. However, whether such directionalities have any connexion with `causality` is a non-obvious metaphysical issue that deserves deeper attention.

6. A formal language for causality

The above approaches to adding `causal` meaning to CI and DAG representations operate with `causal concepts` that remain vague and undefined, and can be criticised for their somewhat arbitrary assumptions about how `causal` and probabilistic properties should be related. An alternative and more helpful approach (Didelez and Sheehan, 2007b) is to specify directly a formal framework, with clear semantics relating mathematical properties of a putative representation to `causal` properties of the external system it is intended to represent. No rigid assumptions about how `causality` relates to probability need be made. Rather, the aim is to present a general language in terms of which we can clearly express and manipulate whatever tentative assumptions we may be making in a specific context. The parallel task for the less ambitious task of modelling probabilistic conditional independence has already been successfully undertaken, as described in Sections 3 and 4. I now describe one way of extending that analysis into the `causal` domain.

I must start by clarifying my own understanding of the concept of `causation`. In the spirit of § 2, I regard this as simply relating to the behaviour of a system when subjected to an external intervention. Having decided on this as the fundamental interpretation, I find I can fruitfully largely dispense with intuitive terms such as `cause`, `effect`, *etc.*, which are

7. or some other graphical CI representation, such as partial ancestral graphs (Richardson and Spirtes, 2002; Zhang, 2008)

so often the cause of much misunderstanding. Any causal concept I might wish to consider should be unambiguously definable in the formal language (and so in effect redundant).

For application of this method to a particular external system, we must have firmly in mind one or more real-world situations ("regimes"), at least one of which involves external intervention,[8] that we wish to understand—and perhaps relate to each other.

We shall in fact focus on one particular kind of relationship between regimes, where we postulate the identity, across two or more regimes, of the *conditional distribution* for one set of variables given another set of variables. When this holds we can regard that conditional distribution as a "modular component," transferable across regimes. Modularity (though more typically conceived in terms of transferable *deterministic* relationships between variables) has often been taken as an essential or defining property of causality, though this view has been challenged by Cartwright (2007, Chapter II-3). While I make no metaphysical commitment to modularity as essential to the understanding of causality, nor to the expression of modularity in terms of invariant conditional distributions, I do nevertheless believe that this particular approach covers a very great deal of ground, and that it is the key to causal inference, including most aspects of "statistical" causality. A similar approach, regarding causality as residing in the "structural stability" of random variation, is taken by Russo (2008).

Suppose then that there is a collection of variables that together describe relevant aspects of the behaviour of a system under each regime of interest. Under any one of these regimes, these variables will have a joint distribution, with conditional independence properties that might be modelled algebraically as in § 3 or, where appropriate, graphically as in § 4. I now extend such mathematical representations to incorporate assumed relationships, as described above in terms of invariant conditional distributions, between the different regimes.[9] To this end, I introduce additional variables to represent the various regimes.

For example, in a system involving variables denoted by Z, U, X, Y, we could introduce an *intervention variable*, F_X, with values corresponding to the different regimes that arise on intervening to set the value of X in various ways. If X is binary, then F_X might have values $\emptyset$, 0 and 1. The interpretation is that, when $F_X = \emptyset$ (the *idle* regime), the other variables arise from the undisturbed system; when $F_X = 0$ [resp., 1] the other variables arise from the system disturbed by an external intervention that forces X to take the value 0 [resp., 1]. Note that F_X is not a *random* variable, but rather has the status of a parameter, determining the regime operating: in particular, any probability or independence statements must be explicitly or implicitly conditioned on the value of F_X.

8. We must have a clear understanding of the nature of any intervention considered, as a real-world phenomenon, prior to any modelling exercise.

9. This invariance or (stochastic) "regularity" concept, in addition to being fundamental to my interpretation of causality, has other useful applications, arguably outside "causal inference", which can be modelled and analysed in essentially the same way. Thus we might consider the differing probabilistic behaviours of some collection of random variables in various different hospitals. We could then introduce a non-random regime indicator (but now without an interventional interpretation) to index which hospital we are looking at: this would allow us to express an assumption that a certain conditional distribution is the same in all hospitals. Or, we could express the property that a certain imperfect diagnostic test has the same error probabilities, no matter who it is used on. Such "reusable invariant modules" can be conveniently implemented in "object-oriented" software such as Hugin 6 <http://www.hugin.com/> (Dawid et al., 2007), and have been found useful in generic schemes for handling and interpreting evidence (Hepler et al., 2007).

In general, the joint distributions of (Z, U, X, Y) (and their independence properties) under these three different regimes (*i.e.*, given $F_X = \emptyset$, 0 or 1) will be unrelated to each other. But should we wish to specify or describe connexions between them, we can usefully do so using, once again, the language of conditional independence—now naturally extended (Dawid, 1979, 2002b) to allow some of the variables entering to be non-random intervention variables (or more general decision or parameter variables). This *extended conditional independence* (ECI) theory turns out to supply the appropriate formal language (syntax and semantics) for describing and manipulating `causality` in our modular decision-theoretic understanding of the term. Although this programme can be effected in entirely algebraic fashion, it is most easily and fruitfully described in graphical terms.

6.1 Augmented DAGs

Just as for regular CI it is often (though by no means always) possible and helpful to represent a collection of ECI properties by means of a DAG[10]—now explicitly including nodes to represent non-random regime or decision variables (generally drawn as square) in addition to nodes representing random variables (generally drawn as round). Indeed, this can be done with essentially[11] the identical constructions and interpretations as for regular DAGs. Such a DAG is termed an *influence diagram* (ID) (Dawid, 2002b). Many of the IDs considered in a causal context have a specific form, as "*augmented DAGs*"[12] (Pearl, 1993). Figure 5 shows an augmented DAG, a variation on the simple, purely probabilistic, DAG of Figure 1, that also incorporates (but in a particular way) an *intervention node* F_X.

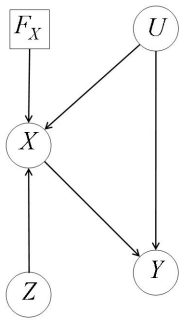

Figure 5: Augmented DAG

What would it mean to say that a particular system is represented by this augmented DAG? If we apply the "d-separation semantics" described in § 4 to Figure 5, ignoring the distinction between random and decision variables, we can read off the following conditional independence properties:

10. Other kinds of graphical CI representations can be similarly extended to include intervention variables (Dawid, 2002a; Zhang, 2008; Eichler and Didelez, 2009).
11. The only new constraint is that, in constructing a DAG using (2), all non-random variables be considered before all random variables.
12. Murphy (2008) points out the usefulness of more general IDs in this context.

$$(U, Z) \quad \perp\!\!\!\perp \quad F_X \tag{5}$$
$$U \quad \perp\!\!\!\perp \quad Z \mid F_X \tag{6}$$
$$Y \quad \perp\!\!\!\perp \quad F_X \mid (X, U) \tag{7}$$
$$Y \quad \perp\!\!\!\perp \quad Z \mid (X, U; F_X). \tag{8}$$

Since F_X is not a random variable, a little care is need in interpreting these statements, but this can nevertheless be done in straightforward fashion. Thus property (5) is to be interpreted as saying that the joint distribution of (U, Z) is independent of the regime F_X: *i.e.*, it is the same in all three regimes. In particular, it is unaffected by whether, and if so how, we intervene to set the value of X. The identity of this joint distribution across the two interventional regimes $F_X = 0$ and $F_X = 1$ could be interpreted as expressing a `causal` property: manipulating X has no (probabilistic) `effect` on the pair of variables (U, Z). Furthermore, since this common joint distribution is also supposed the same in the idle regime, $F_X = \emptyset$, we could in principle use observational data to estimate it—thus opening up the possibility of `causal inference`.

Property (6) asserts that, in their (common) joint distribution in any regime, U and Z are independent: this however is a purely probabilistic, not a `causal`, property.

Property (7) says that the conditional distribution of Y given (X, U) is the same in both interventional regimes, as well as in the observational regime, and can thus be considered as a modular component, fully transferable between the three regimes — again, I regard this as expressing a `causal` property.

Finally, property (8) asserts that this common conditional distribution is unaffected by further conditioning on Z (not in itself a `causal` property).

Just as for regular CI, it is possible for a collection of ECI properties to have more than one representation as an augmented DAG. This is the case for Figure 6, where the direction of the arrow between U and V is not determined.

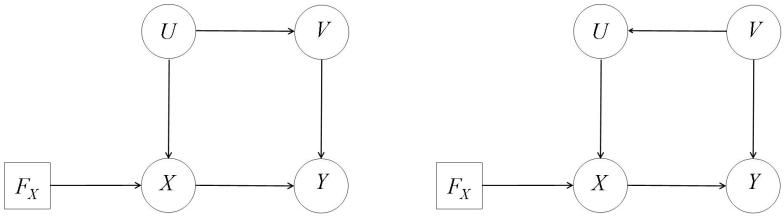

Figure 6: Two Markov-equivalent augmented DAGs

We see that, taking a decision-theoretic understanding of `causality`, the essential ingredients required for certain "`causal`" assertions and inference—namely that certain marginal or conditional distributions be unaffected by whether or how certain interventions are made—are readily expressible using the familiar language of conditional independence (specifically, they arise when the second argument of an ECI relation is a decision variable).

They are just as readily manipulated by means of the rules embodied in (1). And in those special cases that it is possible to express all the `causal` and probabilistic assumptions made in a form that can be represented by an augmented DAG,[13] we can use the d-separation semantics of § 4 as a "theorem-proving machine" to discover their logical implications.

7. Intervention DAGs

One approach to reinterpreting a *probabilistic* DAG as `causal` is first to elaborate it with further intervention variables, and then to interpret the resulting augmented DAG according to the extended d-separation semantics described above. This is the essence of the influential approach of Pearl (2000). In its simplest and most widespread form, this explicitly or implicitly associates with *every* random variable V in the probabilistic DAG an intervention node F_V, with state-space $\mathcal{V} \cup \{\emptyset\}$, where $\mathcal{V}$ is the state-space for V; and an arrow pointing from F_V to V. In the resulting augmented DAG, the conditional distribution of V given its parents (which now include F_V) is taken, when $F_V = \emptyset$, to be the same as it is in the original probabilistic DAG, given only V's random parents; whereas when $F_V = v \in \mathcal{V}$ (interpreted as "setting" V to v by an external intervention) this conditional distribution assigns probability 1 to $V = v$.

Such an augmented DAG represents and relates all regimes that intervene on any number of variables simultaneously, or on none. In this representation, the conditional distribution of any non-manipulated variable given its random parents is taken to be a modular component, invariant across all such regimes. In particular, given the DAG, any interventional joint distribution is easily computed in terms of the observational joint distribution.

In Pearl's earlier work (Pearl, 1993, 1995) he moved backwards and forwards between explicit and implicit representation of the intervention variables in the DAG. More recently he, and most of those following him, have been using only the implicit version, an *intervention DAG*, in which the intervention variables F_V are not explicitly included in the diagram, but the DAG is nevertheless to be interpreted as if they were—thus giving it a specific interpretation as a `causal` DAG. I regard this graphical demotion of the intervention indicators as a retrograde move, since the graphical representation, although now imbued with new `causal` semantics, is then visually indistinguishable from that used to describe purely probabilistic CI. Consequently, great care is needed to be clear just what a given DAG is intended to represent, and to avoid slipping unthinkingly from one interpre-

13. Such cases are indeed quite special. Thus in an augmented DAG incorporating an intervention node F_X pointing at X, we will have (a) $X \perp\!\!\!\perp \{F_X, \mathrm{nd}(X)\} \mid \mathrm{pa}(X)$, where $\mathrm{nd}(X)$ and $\mathrm{pa}(X)$ denote, respectively, the non-descendants and the parents of X in the corresponding unaugmented graph. In particular, this implies *both* the `causal` property (b) $X \perp\!\!\!\perp F_X \mid \mathrm{pa}(X)$, expressing the invariance of the conditional distribution for X given its graph parents, *and* the probabilistic CI property (c) $X \perp\!\!\!\perp \mathrm{nd}(X) \mid \{\mathrm{pa}(X), F_X\}$, asserting that, in the observational regime (the only regime for which this property is non-trivial), X is independent of its non-descendants given its parents. While this appears very close to being a statement of CMC, it is not, because: (i) it does not make any logical connexion between the distinct properties (b) and (c), it merely claims that, in the problem modelled, they both happen to hold; and (ii) this simultaneous assertion of both (b) and (c) is in any case a consequence of using an augmented DAG representation, which can embody only very special sets of ECI relationships, and may simply not be appropriate for the case at hand.

tation to another. Explicit representation of intervention nodes helps to guard against such confusion, as well as simplifying interpretation and manipulation.[14]

Pearl's approach has the great advantage of being explicit as to the meaning to be attached to `causality`: indeed, it uses exactly the same modular decision-theoretic interpretation as I do. Further, if (but this is a big if!) we have agreed that a given problem can be modelled, uniquely, by means of a certain intervention DAG, we can use it to assign meaning to certain other `causal` concepts, such as those entering into the "Causal Markov assumption": for example, an `immediate cause` of a variable V might now be simply *defined* as a parent of V in the associated intervention DAG.

But what is typically omitted in this enterprise is explicit consideration of the relationship between a DAG model and the real world. In particular, because every DAG model can now be given both a probabilistic and a `causal` interpretation, it is easy to conclude that, once we have derived a DAG model to describe observational conditional independencies, it must necessarily also be interpretable according to the more sophisticated `causal` semantics of intervention DAGs. While this is evidently untrue (in particular, distinct DAG models representing identical observational CI properties will always have different implications when interpreted `causally`), such "reification" of a graphical DAG structure is all too common.

When this facile interpretation of DAG models is combined with the CI-seeking techniques of "causal discovery", as described in § 5.4, it lends further support to the widespread but nevertheless totally unfounded belief that we are now in possession of a soundly-based technology for drawing justifiable `causal conclusions` from purely observational data, without further assumptions.[15] The force of Cartwright's maxim "No causes in, no causes out" can not be so easily evaded.

My own approach is different. It makes no explicit or implicit assumptions determining how different regimes *should* be related, and in general allows that there may be no relationship whatsoever between them. But if and when we do wish to express specific relationships between regimes, for example because we feel we have external reasons for assuming or hypothetically considering them, we have at our disposal a formal language (quite generally, the algebraic language of ECI, or, more restrictedly, representations in terms of augmented DAG models or influence diagrams) by means of which we can represent and manipulate such relationships. The initial task of deciding just how the different regimes might indeed be related is not (indeed, could not be) addressed within the formal language, but has to be considered from an entirely external, context-dependent, standpoint, taking into account what is known, and what it is reasonable to assume or postulate, about the effects of real interventions in the real-world system under consideration.

8. Instrumental variables

To clarify the similarities and differences between augmented DAG representations and other `causal` DAG representations, we revisit the example of § 5.2. Hernán and Robins (2006)

14. For example, Pearl (1995) derives his "do-calculus" rules using an explicit augmented DAG representation, but then re-expresses them in terms of the unaugmented graph—when they become considerably more complex. It is not clear what is gained to compensate for this loss of transparency.
15. See Geneletti (2005) for an elucidation of the hidden assumptions that are being made in this enterprise.

present the causal DAG of Figure 4 as a counterexample to the supposition (Martens et al., 2006) that the following conditions are necessary for a variable Z to qualify as an "instrumental variable" for estimating the causal effect of an "exposure" X on a "response" Y, in the presence of an additional, unmeasured, variable U (a confounder), that affects both X and Y, when we can not directly manipulate X:

(i). Z has a causal effect on X

(ii). Z affects the outcome Y only through X (*i.e.*, no direct effect of Z on Y)

(iii). Z does not share common causes with the outcome Y (*i.e.*, no confounding for the effect of Z on Y).

The causal DAG presented by Hernán and Robins (2006) as embodying these assumptions is essentially the same as our Figure 1. This is contrasted with the causal DAG of Figure 4, which is not regarded as embodying condition (i), since Z has no direct causal effect on X, but is merely associated with it through sharing a common cause U^*.

Note that the descriptions of both problems employ intuitive causal terms, and that these are associated with the presence and directionality of the arrows in the causal DAG representations.

My decision-theoretic versions of these stories are presented in Figure 5 and Figure 7. In each case I have added an intervention node F_X associated with X, describing three regimes of interest: the idle regime $F_X = \emptyset$ corresponding to pure observation, and the two interventional regimes $F_X = 0$ and 1, corresponding to an intervention in which X is externally manipulated to take values 0 and 1, respectively. Data can be gathered only under the idle regime, which is thus all that can be directly estimated; but our interest is nevertheless in estimating (if possible), and, especially, comparing, the distributions of the response Y under the interventional regimes, $F_X = 0$ and $F_X = 1$.

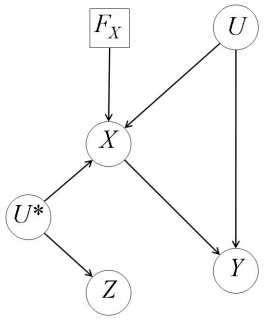

Figure 7: Augmented DAG corresponding to Figure 4

Now in the story represented by Figure 4 or Figure 7, the variable U^*, while apparently required for a full specification of the structure of the problem, plays no rôle in the analysis of Z as an instrumental variable. So we can restrict attention to the joint distribution, under the various regimes, of the variables (U, X, Y, Z), and their independence properties. We then find that the augmented DAG of Figure 7 embodies the identical conditional

independence properties (5)–(8) as the alternative augmented DAG of Figure 5, in which U^* does not figure at all.[16] Consequently, for our purposes this is just as good an augmented DAG representation of the problem as Figure 7. This equivalence should be contrasted with the apparent attitude of Hernán and Robins (2006), that if Figure 4 is a "true" `causal` DAG representation of the problem, then Figure 1 is not.

The important interpretive difference between `causal` DAGs and augmented DAGs is that, in the former, causal meaning is understood as carried by the arrows, whereas, in the latter, it is entirely carried by extended conditional independence properties, involving intervention variables, which are represented only indirectly in the DAG, *via* d-separation. In particular, in Figure 5 the arrow from Z to X is *not* to be construed as representing a relationship of `cause and effect` between Z and X (see Didelez and Sheehan (2007a) for more on this in the context of Mendelian randomization).

The ECI properties (5)–(6) are "core conditions" for a variable Z to be an *instrument* for the effect of X on Y.[17] Once so characterised, these properties can be manipulated algebraically using the rules of (1) (together with properties such as $F_X = 0 \Rightarrow X = 0$), without reference to any graphical representation: the "theorem-proving" properties of DAG representations, while immensely useful, are logically inessential. And if we do want to use graphical representations to help us, there is no point in arguing whether it is Figure 5 or Figure 7 that is "correct"—since each of them embodies (5)–(6) equally well.

9. Justifying assumptions

An important characteristic of my approach to causality using extended CI and augmented DAGs is that it is descriptive, not prescriptive. It makes no assumptions as to how causality ought to behave or be represented; rather, it supplies a language by which we are able clearly to express and manipulate any such assumptions we might wish to make in any given context. In this respect it differs from other theories of "`probabilistic causality`"[18] in much the same way as Kolmogorov's purely formal theory of probability differs from other theories such as the "classical theory" based on the assumption that intuitively "equally possible" outcomes should be assigned equal probabilities, or von Mises's theory of collectives, which sought to represent assumed empirical properties of probability, such as the existence and stability of limiting relative frequencies, directly within the formal theory. This strict separation of the formal general-purpose language from any special assumptions that might be made in specific contexts allows for much greater clarity and flexibility. It also protects

16. This essential identity between Figure 7 and Figure 5 would be destroyed if we were also to allow intervention in Z, as for example if we interpreted the DAG as an intervention DAG in the sense of § 7. However I am here considering the case where Z can be observed, but not manipulated.
17. There is one more "core condition", expressible in terms of CI but not graphically representable: $X \not\perp\!\!\!\perp Z \mid F_X = \emptyset$. In addition to these core conditions, precise identification of a causal effect by means of an instrumental variable requires further modelling assumptions, such as linear regressions (Didelez and Sheehan, 2007a).
18. By this term I do not mean to include general theories of "`statistical causality`," such as that of Rubin (1978), which likewise make no prescriptive assumptions. See Dawid (2000, 2002b) for comparisons and contrasts between my own approach and other approaches to `statistical causality`. The general points I have made could have been developed from the viewpoint of those other theories, though these mostly do not focus, as I do, on modularity at the level of conditional distributions, which supplies a natural point of contact with the intuitive concepts of "`probabilistic causality`".

against the ever-present danger of unthinking reification of incidental formal properties of our representations. In particular, it does not in itself support `causal` interpretation of a probabilistic DAG. If we wish to represent this, we have very explicitly to introduce (using ECI) whatever additional assertions we are making about effects of interventions.

This formal approach does, of necessity, leave entirely untouched such questions as "Where do we get our `causal` assumptions from?" and "How can they be justified?" It is at this point, entirely removed from representational issues, that we might find a place for informal arguments based on intuitive understandings of `cause and effect`.

In principle, the meaning of ECI assumptions such as (5)–(8) is straightforward; and they could indeed all be tested empirically if we had access to data collected on (U, Z, X, Y) under the various regimes. In practice, however, we will usually not have such data (and it may not even be clear which unobserved external variable or variables are represented by the symbol U). Then the appropriateness of the assumptions made requires and deserves further, necessarily context-dependent, argument.

For example, physical randomization of a treatment T in the "idle" regime, when present, is generally agreed to provide a convincing reason for believing that the observational distribution of a response Y, given $T = t$, is the same as its distribution would be under an intervention to set T to t (formally: $Y \perp\!\!\!\perp F_T \mid T$), thus justifying `causal` interpretation of these conditional distributions.

Although this property of randomization is usually taken as intuitively obvious, I am not aware of any argument for it based on deeper principles. One such argument could be based on the assumed existence of some *sufficient covariate* U, such that (a) $U \perp\!\!\!\perp F_T$ and (b) $Y \perp\!\!\!\perp F_T \mid (T, U)$ (Dawid, 2002b). Here, (a) says that the distribution of U is unaffected by which regime is operating—typically believable if U is a "pre-treatment" variable; while (b) says that, conditional on U and *which* treatment T is applied, the response Y of the system is unaffected by *how* (*i.e.*, in which regime) it is applied. While it may not be easy to identify a specific pre-treatment variable U with this property, one might be willing to accept that some such variable does exist. Randomization, and the pretreatment status of U, now gives good cause to accept $T \perp\!\!\!\perp U \mid F_T = \emptyset$, whence (since T is in any case non-random in any interventional regime) (c) $T \perp\!\!\!\perp U \mid F_T$. Using the rules of (1), it is straightforward to deduce, from the three CI properties (a), (c), (b), the desired conclusion $Y \perp\!\!\!\perp F_T \mid T$. Alternatively, these CI properties can be represented by the augmented DAG of Figure 8, from which we can readily read off $Y \perp\!\!\!\perp F_T \mid T$. Similar arguments can be made to justify suitably expressed causal interpretations of data generated under more complex randomization schemes. But the appropriateness of any such argument needs to be carefully considered, not just taken for granted.[19]

When physical randomization is not possible, it will be necessary to attempt to justify causal CI assumptions on other grounds. For example, in the instrumental variable problem

19. For example, even in a randomized double-blind clinical trial—the "gold standard" of evidence-based medicine—one could argue that the very artificiality of the trial negates assumption (b) above: we would not expect the same response process to operate for future treated patients as for those in the trial. To make progress we might make weaker assumptions, such as transferability from the clinical trial into general practice of the "specific causal effect": $\mathrm{E}\{Y \mid T = 1, U = u, F_T = \emptyset\} - \mathrm{E}\{Y \mid T = 0, U = u, F_T = \emptyset\} = \mathrm{E}\{Y \mid F_T = 1, U = u\} - \mathrm{E}\{Y \mid F_T = 0, U = u\}$. While not expressible in terms of ECI, such an assumption still relates to the invariance of probabilistic properties across different regimes. Again, it should be made explicit, and justified (ideally empirically).

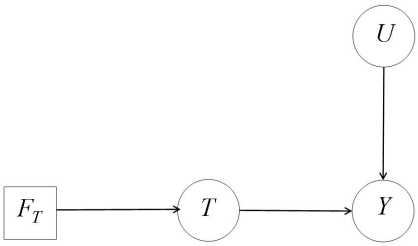

Figure 8: Augmented DAG for randomization

of § 8, we need to argue for the appropriateness of the assumptions (5)–(8). (Once again, it is enough that there exist *some* variable U, which we need not however specify in detail, for which the conditions can be assumed to hold.)

Property (5) essentially requires that both U and Z be pre-treatment variables, and then (7) implies that U must be a sufficient covariate.

Properties (6) and (8) are more problematic.

Property (6) could be plausible if Z is itself determined by randomization: a scenario in which this occurs is that of "incomplete compliance" (Dawid, 2003), where patients are randomized to treatment, with randomization indicator Z, but the treatment X actually taken might not be the same as that assigned. Alternatively, in "Mendelian randomization" (Didelez and Sheehan, 2007a), Z might be a gene that naturally affects X: property (6) might then be justified, for suitable U, on the basis of the random assortment of genes under Mendelian genetics. As described by Didelez and Sheehan (2002): "If we think of U as some behavioural pattern or life style, this independence condition can be justified as long as we are reasonably certain that any possible genetic factors influencing the behavioural pattern are unrelated to this particular gene".

Finally, (8) requires that the distribution of Y given (X, U) (which has been assumed the same in all regimes) is unaffected, in any regime, by further conditioning on Z—intuitively expressed as "no `direct effect` of Z on Y". This might be plausible in the imperfect compliance context, where we could believe that behaviour of the response Y could depend on the treatment X actually taken and further pre-existing individual characteristics U, but not further on the treatment Z that the individual was supposed to take. In the context of Mendelian randomization, we require that "there is no association between the genotype and the disease status given the intermediate phenotype and the life style" (Didelez and Sheehan, 2002). (However, core conditions (6) and (8) can be violated in the presence of various complications, such as linkage disequilibrium, pleiotropy, genetic heterogeneity or population stratification (Didelez and Sheehan, 2007a).)

When attempting to justify the core conditions in a specific context, it is plausible that thinking about the problem in terms of further unobserved variables, such as U^* in Figure 7, can play a valuable rôle in the process. However, once these conditions have been settled on as the assumptions we wish to introduce, there is no need to make irrelevant distinctions between alternative, equally valid representations of them, such as Figure 5 and Figure 7.

In our ECI framework, attention is clearly drawn to any assumptions we may choose to make, since these have to be clearly expressed as explicit ingredients added to our model. In other approaches the assumptions are often hidden, and it is easy to be misled into believing that they are not in need of justification. For example, the weak causal Markov assumption (§ 5.3) rules out certain ECI representations purely on the basis of ordinary CI properties in the observational regime; but there is no logical reason why this should be so, and its validity should be carefully considered in every intended application.

10. Conclusion

We have contrasted two approaches to the interpretation of graphical models of probabilistic causal processes. Each of these purports to relate properties of the mathematical model and properties of the process.

The most common approach, which may loosely be termed "probabilistic causality" (see § 5.3) works with intuitive understandings of causal terms, which are often taken as undefined and self-evident primitives (although it would be an interesting exercise to conduct a survey to investigate the extent to which different practitioners do in fact share a common understanding of such terms). Its most important feature is that it assumes firm links (*via e.g.* the "Causal Markov Condition") between such causal concepts and certain probabilistic conditional independence properties (typically represented in a DAG) of observational regimes—though these are generally insufficient fully to determine causal relationships on the basis of purely observational data (though they may act as constraints on such relationships).

In contrast the approach I am promoting, based on the algebraic theory of extended conditional independence and its graphical representations (see § 6), is based on a clearly defined internal mathematical structure (syntax), and clearly described rules of interpretation (semantics)—the latter informed by a clearly-articulated understanding of causality in terms of the probabilistic consequences of interventions. It does not assume any special relationship between causality and conditional independence. It merely supplies a formal language by means of which we can express and explore interesting causal conjectures—phrased as the identity ("modularity") of certain conditional distributions across a variety of different regimes, which will typically encompass both intervention and pure observation. This surgical separation of the formal language from *ad hoc* causal assumptions enforces clear and unambiguous articulation of those assumptions, allows us to develop the logical implications of our assumptions, and clarifies exactly what needs to be justified in any particular context. That justification is itself, however, a separate task, that can not rely on formal representations of any kind but must relate to the real-world context of the problem.

What rôle for "causal discovery"?

The enterprise of "causal discovery", based on the first of the above interpretations of DAGs, purports to extract causal conclusions from observational conditional independencies. In fact, however, it can not do so without additional causal inputs (for example, to distinguish between conditional independencies taken to have causal content, and those regarded as "accidental"). Such assumptions are rarely articulated explicitly, and even more rarely justified. Instead, it is common to interpret a constructed conditional independence DAG as

if it were also a "causal DAG". This might be "justified" by dismissing certain observational conditional independencies as non-causal because they are not explicitly expressed by the constructed DAG ("non-faithfulness" of the DAG representation)—but this argument is invalid. We need to start from a pre-existing understanding of our causal terms, and then seek a mathematical representation of them; to *define* our primitive concepts in terms of a tentative representation would constitute both a circular argument and a reification fallacy. Perhaps the best that can be said for such methods is that they can suggest interesting causal conjectures for further investigation. Ideally we would then gather data from appropriate truly interventional studies, to investigate, and if necessary revise, the validity of conjectures made purely on the basis of observational data about the effects of interventions. Williamson (2005) argues for such a "hybrid hypothetico-deductive/inductive" approach.

Alternatively, if we can start from data collected under a variety of regimes, including interventional studies, we can directly apply variations of causal discovery techniques to uncover genuinely causal properties. Thus, if we had data on variables (U, Z, X, Y) under all three regimes $F_X = \emptyset$, $F_X = 0$, $F_X = 1$, we could empirically test the ECI properties (5) and (7), by (for example) simple χ^2-tests (which are equally valid for testing homogeneity of conditional distributions as they are for testing conditional independence); alternatively, Bayesian techniques could be used (Cooper and Yoo, 1999). Only with such experimental data could we hope to obtain genuine empirical evidence in favour of a causal DAG representation such as Figure 5.

Acknowledgments

Thanks to Nancy Cartwright, Vanessa Didelez, Sara Geneletti, Paul Rosenbaum, Federica Russo and Jon Williamson for helpful comments.

References

Pierre Bourdieu. *Outline of a Theory of Practice*. Cambridge University Press, 1977.

George E. P. Box. Use and abuse of regression. *Technometrics*, 8:625–629, 1966.

Nancy Cartwright. *Nature's Capacities and Their Measurement*. Clarendon Press, Oxford, 1994.

Nancy Cartwright. *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge University Press, 2007.

Gregory F. Cooper and Changwon Yoo. Causal discovery from a mixture of data. In *Proceedings of the 15th Annual Conference on Uncertainty in Artificial Intelligence (UAI-99)*, pages 116–12, San Francisco, CA, 1999. Morgan Kaufmann.

Robert G. Cowell, A. Philip Dawid, Steffen L. Lauritzen, and David J. Spiegelhalter. *Probabilistic Networks and Expert Systems: Exact Computational Methods for Bayesian Networks*. Springer, New York, 2007.

A. Philip Dawid. Conditional independence in statistical theory (with Discussion). *Journal of the Royal Statistical Society, Series B*, 41:1–31, 1979.

A. Philip Dawid. Conditional independence for statistical operations. *Annals of Statistics*, 8:598–617, 1980.

A. Philip Dawid. Causal inference without counterfactuals (with Discussion). *Journal of the American Statistical Association*, 95:407–448, 2000.

A. Philip Dawid. Separoids: A mathematical framework for conditional independence and irrelevance. *Annals of Mathematics and Artificial Intelligence*, 32:335–372, 2001.

A. Philip Dawid. In discussion of Lauritzen and Richardson (2002). *Journal of the Royal Statistical Society, Series B*, 64:348–351, 2002a.

A. Philip Dawid. Influence diagrams for causal modelling and inference. *International Statistical Review*, 70:161–189, 2002b. Corrigenda, *ibid*., 437.

A. Philip Dawid. Causal inference using influence diagrams: The problem of partial compliance (with Discussion). In P. J. Green, N. L. Hjort, and S. Richardson, editors, *Highly Structured Stochastic Systems*, pages 45–81. Oxford University Press, 2003.

A. Philip Dawid, Julia Mortera, and Paola Vicard. Object-oriented Bayesian networks for complex forensic DNA profiling problems. *Forensic Science International*, 169:195–205, 2007.

Bruno de Finetti. *Theory of Probability (Volumes 1 and 2)*. John Wiley and Sons, New York, 1975. (Italian original Einaudi, 1970).

Vanessa Didelez and Nuala A. Sheehan. Mendelian randomisation and instrumental variables: What can and what can't be done. Technical Report 05-02, University of Leicester Department of Health Sciences, 2002.

Vanessa Didelez and Nuala A. Sheehan. Mendelian randomisation as an instrumental variable approach to causal inference. *Statistical Methods in Medical Research*, 16:309–330, 2007a.

Vanessa Didelez and Nuala A. Sheehan. Mendelian randomisation: Why epidemiology needs a formal language for causality. In F. Russo and J. Williamson, editors, *Causality and Probability in the Sciences*, volume 5 of *Texts In Philosophy Series*, pages 263–292. College Publications, London, 2007b.

Michael Eichler and Vanessa Didelez. On Granger-causality and the effect of interventions in time series. Research Report 09:01, Statistics Group, University of Bristol, 2009.

Morten Frydenberg. The chain graph Markov property. *Scandinavian Journal of Statistics*, 17:333–353, 1990.

Maria Carla Galavotti. *Philosophical Introduction to Probability*. CSLI Publications, Stanford, 2005.

Sara G. Geneletti. *Aspects of Causal Inference in a Non-Counterfactual Framework*. PhD thesis, Department of Statistical Science, University College London, 2005.

Clark Glymour and Gregory F. Cooper, editors. *Computation, Causation and Discovery*. AAAI Press, Menlo Park, CA, 1999.

Daniel Hausman. *Causal Asymmetries*. Cambridge University Press, Cambridge, 1998.

Amanda B. Hepler, A. Philip Dawid, and Valentina Leucari. Object-oriented graphical representations of complex patterns of evidence. *Law, Probability & Risk*, 6:275–293, 2007.

Miguel A. Hernán and James M. Robins. Instruments for causal inference: An epidemiologist's dream? *Epidemiology*, 17:360–372, 2006.

Dominik Janzing and Bernhard Schölkopf. Distinguishing between cause and effect via the algorithmic Markov condition. Paper presented at NIPS 2008 Workshop "Causality: Objectives and Assessment", Whistler, Canada, 2008.

Steffen L. Lauritzen, A. Philip Dawid, Birgitte N. Larsen, and Hanns-Georg Leimer. Independence properties of directed Markov fields. *Networks*, 20:491–505, 1990.

Steffen L. Lauritzen and Thomas S. Richardson. Chain graph models and their causal interpretations (with Discussion). *Journal of the Royal Statistical Society, Series B*, 64: 321–361, 2002.

Edwin P. Martens, Wiebe R. Pestman, Anthonius de Boer, Svetlana V. Belitser, and Olaf H. Klungel. Instrumental variables: Applications and limitations. *Epidemiology*, 17:260–267, 2006.

Kevin Murphy. Modeling uncertain interventions. Paper presented at NIPS 2008 Workshop "Causality: Objectives and Assessment", Whistler, Canada, 2008.

Richard E. Neapolitan. *Learning Bayesian Networks*. Prentice Hall, Upper Saddle River, New Jersey, 2003.

Judea Pearl. A constraint–propagation approach to probabilistic reasoning. In L. N. Kanal and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 357–370, Amsterdam, 1986. North-Holland.

Judea Pearl. Comment: Graphical models, causality and intervention. *Statistical Science*, 8:266–269, 1993.

Judea Pearl. Causal diagrams for empirical research (with Discussion). *Biometrika*, 82: 669–710, 1995.

Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2000.

Hans Reichenbach. *The Direction of Time*. University of Los Angeles Press, Berkeley, 1956.

Thomas S. Richardson and Peter Spirtes. Ancestral graph Markov models. *Annals of Statistics*, 30:962–1030, 2002.

Donald B. Rubin. Bayesian inference for causal effects: the role of randomization. *Annals of Statistics*, 6:34–68, 1978.

Federica Russo. *Causality and Causal Modelling in the Social Sciences: Measuring Variations*, volume 5 of *Methodos Series*. Springer, 2008.

Richard Scheines and Peter Spirtes. Causal structure search: Philosophical foundations and future problems. Paper presented at NIPS 2008 Workshop "Causality: Objectives and Assessment", Whistler, Canada, 2008.

Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction and Search*. Springer-Verlag, New York, Second edition, 2000.

Milan Studený. Conditional independence relations have no finite complete characterization. In *Transactions of the Eleventh Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, volume B, pages 377–396, Prague, 1992. Academia.

Patrick Suppes. *A Probabilistic Theory of Causality*. North Holland, Amsterdam, 1970.

Thomas Verma and Judea Pearl. Causal networks: Semantics and expressiveness. In R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 4*, pages 69–76, Amsterdam, 1990. North-Holland.

Jon Williamson. *Bayesian Nets and Causality: Philosophical and Computational Foundations*. Oxford University Press, Oxford, 2005.

James Woodward. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, Oxford, 2003.

Jiji Zhang. Causal reasoning with ancestral graphs. *Journal of Machine Learning Research*, 9:1437–1474, 2008.

Kun Zhang and Aapo Hyvärinen. Distinguishing causes from effects using nonlinear acyclic causal models. Paper presented at NIPS 2008 Workshop "Causality: Objectives and Assessment", Whistler, Canada, 2008.