Introduction to the Epistemology of Causation

Frederick Eberhardt
Department of Philosophy
Washington University in St Louis
eberhardt@wustl.edu

Abstract

This survey presents some of the main principles involved in discovering causal relations. They belong to a large array of possible assumptions and conditions about causal relations, whose various combinations limit the possibilities of acquiring causal knowledge in different ways. How much and in what detail the causal structure can be discovered from what kinds of data depends on the particular set of assumptions one is able to make. The assumptions considered here provide a starting point to explore further the foundations of causal discovery procedures, and how they can be improved.

1 Introduction

David Hume's skeptical argument on induction concluded that causal knowledge must be based on custom and habit [17]. Hume had argued that causal knowledge is not a priori and that there is no perceptual component beyond constant conjunction, time order and spatio-temporal proximity – which are jointly insufficient – to detect a causal relation. If we took this conclusion to have dealt the devastating final blow to any hope of scientifically justified causal knowledge, then all research efforts into policy analysis (e.g. introduction of public healthcare, climate change or bailout plans for the financial sector etc.) would be futile not just because we do not have enough data, but because we are in principle unable to provide a justification for any influence that our policy might have, no matter how much research we do. Worse, every careful deliberation of our own actions could at best be described as a consideration of the habits of our mind rather than an assessment of the expected outcome of an intervention. The reason to go to a doctor when ill would have to be described not as hope of finding a cure that has been shown to be successful, but as a desire to explore the mental associations the doctor has with our ailment (admittedly, it does sometimes feel like that). Ultimately, the whole idea of knowing a reason for something would stand on a somewhat dubious footing.

We know that in particular circumstances nothing other than causal knowledge will do: To predict the behavior of a system under intervention (e.g. the

behavior of the economy subject to a new policy, or the effect of new treatment on a patient etc.) simple dependencies will not do. For example, finding that arthritis is correlated with diabetes provides no indication of what the incidence rate of arthritis will be once diabetes can be cured. Knowing that there is a common cause of arthritis and diabetes implies that the incidence rate of arthritis will depend on where in the process the cure of diabetes impacts. It therefore comes as no surprise that scholars have tried to find ways out of Hume's dilemma and supply alternative accounts of how we come to have causal knowledge. Here I will not provide a historical review – such can be found in more detail in [24], with Bacon, Kant, Mill and Reichenbach as the main non-contemporary philosophical figures with constructive attempts. Instead, I will outline some of the main principles underlying current procedures for causal discovery, and describe how they fit together to form an epistemology of causation.

2 Bridge Principles

In the philosophical literature the metaphysics of causality arguably occupies the prime position. In light of Hume's argument one may regard this as a result of an attempt to first establish what it *is* that we talk about in making causal claims, before we then figure out how we come to know it. But the metaphysical accounts have provided essentially no guidance for methods of discovery, because it remains unclear how they could be operationalized into discovery procedures that do not depend on the availability of causal knowledge in the first place.[1]

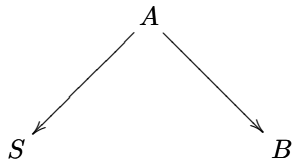
Epistemological headway was made by a completely different strategy that largely ignored metaphysical considerations. The approach, in broad strokes, was this: Let us accept with Hume that there is no immediate perceptual or measurable feature that identifies uniquely a causal relation. (This assumption is by no means obvious, since many psychological experiments show that humans and animals seem very sensitive to purely perceptual features when making causal inferences. [13, 19]) If no direct identification of causal features is possible, then assumptions are required that connect what can be observed to the underlying causal structure that generates the phenomena: bridge principles.

With the Principle of Common Cause, Reichenbach [23] was one of the first to pursue this strategy. The Principle of Common Cause was an attempt to connect the constraints implied by a causal structure with measurable constraints in the resulting probability distribution generated by the variables that are causally connected. If the distribution over two variables exhibits a depen-

[1] In Lewis' account of causation in terms of counterfactuals a measure of distances between possible worlds is needed to assess the counterfactuals. Proposed measures depend on knowledge of laws, which presumably represent causal facts in the first place. In accounts that are based on causation as a transfer of a conserved quantity (Salmon, Dowe) it remains a mystery how one comes to know which quantity is conserved without knowing the causal relations first. A similar concern applies to attempts using the mechanistic approach of Machamer et al.. Mackie's "INUS condition" specifies criteria for a cause, but it supplies no procedure. It is unclear whether "INUS-causes" can be discovered by an analysis of regularities or by an analysis of the differences between individual instances, or both.

dence, then this dependence may result from one variable causing the other. It is also possible that there is a third variable, a common cause, that causes both: For example, the occurrence of storms and particular barometer readings are correlated, although neither is the cause of the other (shifting the barometer needle does not prevent the storm, and the barometer reading generally precedes the storm, and since one generally assumes that there is a time order from cause to effect, the storm cannot be the cause of the barometer reading). A third variable, atmospheric pressure, which causes both, explains the correlation. The Principle of Common Cause takes a dependence to be indicative of either one variable causing the other, or the existence of a common cause. It excludes the possibility of the variables being causally disconnected, but leaves the exact connection still underdetermined.

The motivating idea of the Principle of Common Cause is spelled out in full generality in the Causal Markov Condition, which is widely seen as the most fundamental bridge principle in causal epistemology. It uses the causal Bayes net framework [29, 21], which represents the causal structure as a directed graph over a set of variables. For example, the causal structure between storm (S), the barometer reading (B) and atmospheric pressure (A) would be represented as follows:



Causal effects between variables result in a probability distribution over the graph, e.g. for the above example, $p(A, S, B) = p(A)p(S|A)p(B|A)$. The causal Markov condition matches the absence of particular causal relations (arrows between vertices in the directed graph) with (conditional) independence relations in this distribution, and dependence relations with the presence of causal connections.

> **Causal Markov Condition:** Let G be a causal graph with vertex set V and P be a probability distribution over the vertices in V generated by the causal structure represented by G. G and P satisfy the Causal Markov Condition if and only if for every W in V, W is independent of $V \setminus$ Descendants(W) given Parents(W). (see [29], p. 29). [The genealogical terminology is interpreted in the obvious way with reference to the directed graph.]

For the barometer-storm example the Markov condition implies that given the value of atmospheric pressure, the occurrence of storm and the state of the barometer needle are independent. While the Markov condition constrains the dependencies in the probability distribution over a set of variables of interest, independencies are constrained by a converse principle:

> **Faithfulness Condition:** A causal graph G and a probability distribution P over G are faithful to one another if all and only

> the conditional independence relations true in P are entailed by the Markov condition applied to G. ([29], p. 13).

Together, causal Markov and faithfulness enable the inference from independence and dependence constraints in a probability distribution to features of the underlying causal structure, even though they do not identify the causal structure uniquely in all cases. For example, if we measure the variables storm (S), barometer reading (B) and atmospheric pressure (A), then we might find that storm and barometer reading are dependent ($S \not\perp\!\!\!\perp B$), as are barometer reading and atmospheric pressure ($B \not\perp\!\!\!\perp A$) and storm and atmospheric pressure ($S \not\perp\!\!\!\perp A$), while storm and barometer reading *given* atmospheric pressure are independent ($S \perp\!\!\!\perp B|A$). From these constraints alone (assuming Markov and faithfulness) we can only infer that one of the following three causal structures is true, but we do not know which:

$$\begin{aligned} S \rightarrow \quad & A \quad \rightarrow B \\ S \leftarrow \quad & A \quad \leftarrow B \\ S \leftarrow \quad & A \quad \rightarrow B \end{aligned}$$

These three structures form a Markov equivalence class, since they imply the same independence constraints. If we in addition know that the measurements of A always preceded the measurements of S and B, then we can uniquely identify the third structure.

In some cases Markov and faithfulness are sufficient for unique identification of the causal structure. For example, consider a car where we measure the state of the gas tank (G), the battery (B) and whether the motor (M) starts. We might find the following distributional constraints:

$$G \perp\!\!\!\perp B \qquad G \not\perp\!\!\!\perp M \qquad B \not\perp\!\!\!\perp M \quad \text{and} \quad G \not\perp\!\!\!\perp B|M.$$

These constraints uniquely imply (given Markov and faithfulness) that the underlying causal structure is

$$G \rightarrow M \leftarrow B.$$

The fact that conditional on a common effect, two variables that are otherwise independent, become dependent (such as B and G here) provides a unique signature of the underlying causal structure (referred to as v-structure or unshielded collider) in the dependence and independence constraints of the distribution. This can be used for causal discovery (and the orientation of causal influences!) between variables, even when no time order information is available.

The examples show that causal Markov and faithfulness provide a set of assumptions that enable first steps towards a justification of inferences from observable features to particular causal relations. They do not necessarily uniquely identify the causal structure – so they can certainly not be viewed as defining causal relations – and they do not endorse a particular metaphysical account of causation. Instead, they pick up on particular features of causality that are relevant to knowledge acquisition and that are shared by many metaphysical accounts. Markov and faithfulness can only be viewed as one starting point: They

provide constraints that enable discovery, but in many cases one is interested in a further reduction of the remaining underdetermination which may require additional assumptions. In other cases one might be unsure whether even Markov or faithfulness are satisfied, and so one might wonder whether there are alternative assumptions one can use. This suggests a general epistemological approach for the discovery of causal relations: Given a set of assumptions that function as bridge principles, one can determine the possibilities of discovery and develop appropriate procedures. One can work out how the limits of discovery depend on different assumptions. One can ask what justifications are available for any particular set of assumptions, and investigate what happens if an assumption fails.

The remainder of this paper takes precisely this approach: It considers some of the most common sets of assumptions used for causal discovery, indicates, where possible, justifications for those assumptions and points to procedures that use these assumptions. Often a distinction is made between causal discovery in observational data and causal discovery using experiments. It is certainly true that these two circumstances imply different limits to causal discovery, but the distinction is not as clear cut as it may initially seem. First, there is a continuum of circumstances from purely observational studies over weak experimental to fully randomized controlled trials. Second, the power of inferences in experimental circumstances is not solely a result of the experimental setting, but a result of the combination of the experimental setting with other assumptions. Third, we often are in a situation in which experimentation is impossible or unethical, but we will find that some of the following assumptions that are not specifically experimental enable rather powerful inferences even in these circumstances. Consequently, experimental circumstances will be discussed as their own set of assumptions in a section at the end, suggesting that those assumptions stand on a similar epistemic footing as the other assumptions that are not explicitly constrained to the experimental setting.

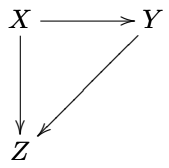
3 Discovery Strategy

What are the desiderata of a search procedure? – Minimally, any discovery procedure should be correct. That is, if its assumptions are satisfied it should not return any false claims. However, since discovery procedures are based on statistical data, and no informative procedure can guarantee not to make errors on finite samples, this minimal correctness must be framed in terms of a limiting condition: In the large sample limit, the procedure should not make errors. For this to be possible, the output of a search procedure based on a particular set of assumptions must be sensitive to the remaining underdetermination that cannot be avoided in the large sample limit. (In the example with the barometer above there were three causal structures that are indistinguishable by dependence and independence constraints, no matter how much data is available.) Given a specification of the underdetermination that will remain in the output (such as given by a Markov equivalence class), a procedure that is guaranteed to make

no errors in the large sample limit is said to be *consistent* with respect to this residual underdetermination. So as not to be trivially consistent, the residual underdetermination should be in some sense (close to) minimal.

Consistency alone is generally considered a necessary requirement, but since we are bound to finite samples, guarantees about the limiting behavior of a procedure provide little assurance. Ideally, one would like a characterization or a bound on the error for the output of any procedure given a particular sample size. For other statistical procedures such re-assurance is often specified in terms of confidence intervals or p-values that provide a probabilistic bound on the size of the error. Causal search procedures that support confidence intervals are said to be uniformly consistent, ones that do not are pointwise consistent. The difference between the two types of consistency is ultimately one of quantifier order: For uniform consistency there is one sample size that bounds the error for all possible distributions over the variables. For pointwise consistency for each distributions there is a sample size that bounds the error. If one does not know which distribution one is dealing with, then the appropriate sample size is similarly unknown.

Whether a procedure is pointwise or uniformly consistent depends on its assumptions. There are search algorithms of both types, and in some cases a pointwise consistent procedure can be made uniformly consistent by a strengthening of some of its assumptions. Without such strengthening, an identification of causal structures using a pointwise consistent procedure can be difficult, because two different causal structures can have distributions that are arbitrarily close to each other, and therefore make a distinction virtually impossible with finite samples. For example, suppose there are three variables X, Y and Z, and the causal structure between them is



but the direct causal influence of X on Z is almost balanced out by the indirect causal influence of X on Z via Y. For a finite sample, X and Z might appear independent, even though there is no violation of faithfulness (in the large sample limit the two variables are not independent). If X and Z appear independent in the finite sample (and $X \not\perp\!\!\!\perp Y$, $Y \not\perp\!\!\!\perp Z$, and $X \not\perp\!\!\!\perp Z|Y$), Markov and faithfulness alone imply that the generating causal structure is a collider: $X \rightarrow Y \leftarrow Z$. Depending on how closely the two causal connections between X and Z cancel each other out in the true causal structure, an arbitrarily large sample may be required to distinguish the two qualitatively very different causal structures even though no assumption was violated.

This example highlights two issues: First, Markov and faithfulness are assumptions that relate causal structure with distributional structure, not with sample associations. Consequently, given only a finite sample, the question of

whether one has reason to believe that the observed associations and independence features in the sample generalize to the distribution, is separate from what inferences one can make about the causal structure given a particular distributional structure. Second, it is an open question how one should treat the output of procedures that are only pointwise consistent: Should one in light of these finite sample problems abandon such procedures altogether, should one search for a justification of further assumptions that ensure uniform consistency (e.g. time order or the possibility of an intervention would help here), or should one try what one has anyway? Reichenbach argued that the minimal guarantee that pointwise consistent procedures converge at some point, even though we do not know when, gives us – at least in the absence of other procedures with stronger guarantees – a rational foundation for inference [22]. If Keynes' slogan that in the long run we all die, can be taken to be indicative of his views of search procedures as well, then he presumably took the opposite view.

Even if we could provide a definitive answer on the question of uniform vs. pointwise consistent procedures, open questions remain on how to treat the output of discovery procedures. Such procedures build on a set of assumptions that mark out the hypothesis space they search. In general we have no guarantee that a particular assumption holds with certainty – none are a priori or necessary. We can at best have a certain belief, more or less justified, that a particular assumptions holds. Consequently, even the confidence interval or the p-value returned by a uniformly consistent search procedure cannot on its own provide the basis for some absolute credence in the output. It is doubtful, whether we can quantify our belief in the search space assumptions sufficiently to appropriately integrate such a belief with the confidence (or lack thereof) in a causal relation determined by the search procedure. (Although again, Reichenbach believed we could provide whole hierarchies of higher order probabilities that were grounded in observable phenomena. [22]) There remains a genuine open question of which epistemic stand one should take towards the output of search procedures.

4 Assumptions for Causal Discovery

Unfortunately, much of the debate in philosophy has focused primarily on the justification and criticism of the Markov and faithfulness conditions. This focus is damaging in two ways: It disregards the possibility of alternatives (especially in the case of faithfulness) and ignores the difficulty of providing alternatives (in the case of Markov). I will cast the net a little further.

Causal Markov

There is a general consensus that the Markov condition describes a central feature of many metaphysical accounts of causation. It generalizes and formalizes the intuition that causes "screen off" their effects from any causal processes that do not descend from the effects. Criticism consequently has taken two forms:

On the one hand it disputes that the real causal metaphysics satisfies such a condition – much of Cartwright's criticism [1, 3, 4] takes this form. Woodward and Hausman [11, 12] have attempted to respond by relating the Markov condition to the idea of modularity (see also [31, 21]) – that one ought to be able to intervene on a causal variable without affecting variables other than the intervened variable and its causal descendents. Alternatively, critics have argued against the inferences the Markov condition licenses. (see e.g. Sober [28] and Cartwright [2]). Most of the purported counterexamples of this latter type involve cases where there is by assumption no causal connection between two variables but an association between the variables is still observed. These examples risk a confusion of the statistical question of generalizability of sample features with the actual claim of the Markov condition that licenses the inference from distributional features to causal structure. We will not pursue the debate here (but see [30, 14, 9, 10] for responses). The real problem is that the Markov condition appears central to just about any scientific investigation, and so far no (weaker) alternative to the Markov condition has been proposed. So while the criticism can be taken as a caution against universal application of the Markov condition, it appears to be a rather tough task to propose any alternative that comes anywhere close to a similar starting point for causal inference.

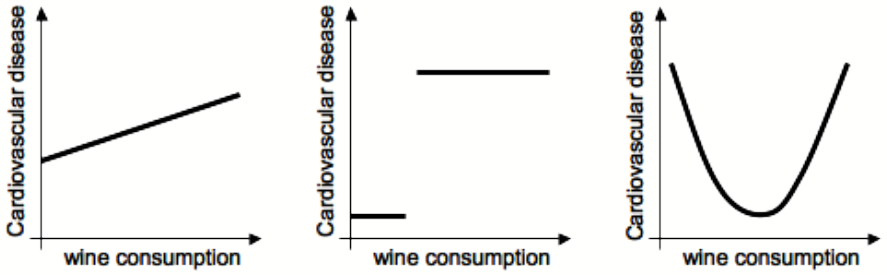
Causal Faithfulness

The situation is different with regard to faithfulness. The justification for faithfulness is pragmatic, to enable particular inferences in search procedures. Counterexamples are more easily constructed and more plausible – as are remedies. For example, if the causal effects along the two paths in the example in Section 3 cancelled exactly, then X and Z would be independent although they are causally connected, thereby violating faithfulness. Independencies that are not due to the Markov condition can arise in the distributional structure for all sorts of reasons, including cancelling pathways, deterministic causal relations or particular discrete causal relations along causal chains. One can provide a mathematical argument that most probability distributions are faithful to the generating causal structure (see [29], p. 41, Theorem 3.2), but there are circumstances where it is reasonable to argue that precisely the unfaithful cases are of scientific interest. The good news is that in the case of faithfulness, the space of alternatives is much better understood. There are many discovery procedures that do not require the faithfulness assumption [25, 20], and there has been a precise investigation into the types of violations of faithfulness that can be detected [33], or where a slight strengthening of other assumptions enables such a detection [32].

Linearity

In general, more than just the causal structure is of interest. One is also interested in the precise quantitative effect one variable has on another. Independence constraints place very minor constraints on the functional form of a

causal effect, and consequently estimation of the quantitative influence would be difficult without further assumptions. One very common assumption about the functional form of the causal relation is linearity. Linearity requires that the value of a variable is determined by a linear sum of the values of its causes plus some error term that is generally, though not necessarily, assumed to have a Gaussian distribution. Consider the relation of wine consumption (x-axis) and the risk of cardiovascular disease (y-axis) in the following three graphs. Only



the first graph satisfies the assumption of a linear relation. Causal relations that involve a threshold below (or above) which there is no, or a different causal effect (middle graph), or causal relations with a non-monotonic effect (third graph) are examples that violate linearity. When investigating the effect of wine consumption on cardiovascular disease it is not obvious even within plausible ranges of wine consumption, whether a linearity assumption is justified. It is perfectly plausible that a positive effect can be observed for small quantities, but that there is a jump in risk as soon as one drinks more than a certain quantity (middle graph). It is also plausible that the relation is quadratic, with a positive effect only for small quantities (right graph).

Apart from that the fact that linearity captures formally something like a dose-reponse relation, the justification for its application arguably has a historical origin: Linear relations are simple, the statistical computations have been tractable for a long time and consequently are well understood (regression analysis). In addition, the assumption enables an enormous reduction in the residual underdetermination of search procedures, especially when there are unmeasured variables [16, 5]. Linearity implies that the causal effect of one particular causal connection (pathway) between two variables can be considered independently of any other connections between the two variables. Consequently, there are many implied constraints on the probability distribution that can be used to identify causal relations.

Gaussianity

Commonly, a linearity assumption is combined with an assumption that the errors of a variable have a Gaussian distribution (bell-curve). Anyone familiar with linear regression will recognize this assumption from the parameter estimation. Part of the motivation to assume Gaussian errors derives from their com-

putational simplicity, since combinations of Gaussian distributions again form joint Gaussian distributions. But more importantly, Gaussian distributions are supported by the central limit theorem: If multiple independent and identically distributed effects (with finite mean and variance) combine, then their sum will approximately follow a Gaussian distribution. Consequently, if the errors in a linear model are taken to represent the many minor individual influences not captured in the model, then for a large sample, the error distribution can reasonably be expected to be Gaussian. Of course, it would be a mistake to assume that error distributions are always or necessarily Gaussian, even when one has large sample sizes. Often the independency assumption of the errors (and their identical distribution) is difficult to guarantee. The assumption would be of minor importance if it made no difference to discovery procedures. Surprisingly, non-Gaussian error distributions can actually make the discovery task easier in the sense that in the large sample limit, more detail can be discovered about the causal structure. Some of the very recent developments of search procedures build on a linearity assumption combined with non-Gaussian errors, and it can be shown, that such procedures actually enable the unique identification of the underlying causal structure [26, 25, 18].

Causal Sufficiency

In many circumstances one cannot assume that all the relevant variables have been measured. There might be unmeasured variables that lead to spurious correlations in the data, that – if not recognized as spurious – can lead to incorrect conclusions about the underlying causal structure. For example, if we find that wine consumption is correlated with a lower incidence rate of cardiovascular disease then this might be due to wine consumption causing the reduction, or it may be the case that people with a higher income can afford to buy wine and can afford proper health care and a balanced lifestyle, resulting in the observed correlation even though wine consumption may have no effect whatsoever on cardiovascular disease. If such variables as socio-economic status are not measured in the study then what inferences can one draw about the true causal structure? The assumption of causal sufficiency, i.e. that there are no unmeasured common causes, is a very strong assumption. In some cases, one can detect that this assumption fails. For example, given four observed variables W, X, Y and Z, if the *only* independencies (represented by $\perp\!\!\!\perp$) between the variables are

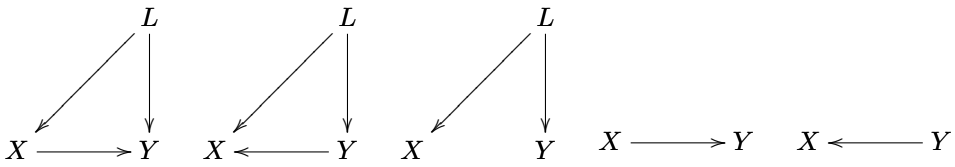
$$\begin{array}{llll} W \perp\!\!\!\perp Y & W \perp\!\!\!\perp Z & X \perp\!\!\!\perp Z & \\ W \perp\!\!\!\perp Y | Z & W \perp\!\!\!\perp Z | X & W \perp\!\!\!\perp Z | Y & X \perp\!\!\!\perp Z | W \end{array}$$

then the only causal structure satisfying these constraints is

$$W \rightarrow X \leftarrow L \rightarrow Y \leftarrow Z$$

where L is an unobserved variable. The presence of the unobserved variable L can be inferred from the independence constraints. But this is not generally

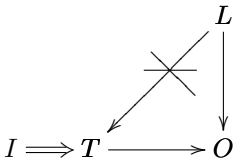
the case. If causal sufficiency cannot be assumed, then the residual underdetermination in the output is generally much larger. For example, just assuming Markov and faithfulness, if a dependence between X and Y is observed, it could have been generated by any of the following structures (again we assume that L is unobserved):



Only the last two are possible candidates if causal sufficiency is assumed. Unless one is dealing with a closed system, there rarely seems to be a justification for the assumption of causal sufficiency, other than that the discovery problem is otherwise often not tractable. There are, however, procedures that are designed to identify whatever can be identified about the causal structure in the case of causal insufficiency, and the situation is not as bleak as might be expected: In general, the residual underdetermination – though more substantial – can be precisely characterized (see FCI-algorithm in [29]), and with the assumption of linearity with non-Gaussian errors, the underdetermination can be reduced substantially (in the large sample limit), and the presence and location of latent variables can (to a certain extent) be determined [15, 16]. Furthermore, there are procedures that can then be used to investigate the structure between latent variables as well [27].

Interventions

The more common way to avoid spurious correlations due to unmeasured variables is to use interventions. Experimental interventions on a system of variables are often hailed as the golden standard of causal discovery. A large literature in statistics is dedicated to the experimental design of randomized controlled trials. The basic idea of randomized controlled trials is that the values of the purported cause variable (the treatment) are set by a distribution that is (causally) independent of the system of variables under consideration. This "setting" of the values of the treatment variable by the intervention (I) implies that the causal influence of any other variable on the treatment variable is broken, and thereby any spurious correlation between treatment (T) and outcome (O) due to any unmeasured common cause (L) can be avoided.



If a correlation between treatment and outcome is measured in the experiment, it must then be due to the treatment causing the outcome. However, such an

inference is only licensed if it is known that the intervention only influences the intervened variable. If the intervention is "fat hand" – influencing more than one variable similar to someone who has insufficiently slim fingers to manipulate an intricate mechanism – then a correlation between variables might be due to the fat-hand intervention rather than a causal influence between variables.

The "surgical" aspect (breaking the causal influences in the treatment variable) of a randomized controlled trial is one way of dealing with unmeasured common causes in an experimental setting. But even if the intervention is weaker than a randomized controlled trial in that it does not make the treatment variable independent of its other causes, there are circumstances in which causal discovery is possible despite the presence of latent variables. In fact, for particular types of "soft" interventions, if one can assume that causal relations are linear, one can uniquely identify the causal structure among an arbitrarily large set of causal variables in a single experiment, and one can detect the presence and location of latent variables. The theory in this case is very similar to that of instrumental variables in economics [7, 5]. These results can also be utilized in circumstances when it is unethical, uneconomical or physically impossible to perform a randomized controlled trial. Even a small external influence on a variable may result in constraints on the system that can be used for causal discovery. Needless to say, there might well be more residual underdetermination, which will require additional assumptions of the type discussed in previous sections to determine the unique causal structure.

How much one can discover about the causal relations, and in what detail, depends on the exact type of intervention, how many variables one can subject to interventions, how many experiments with interventions one can perform and what other assumptions one can make about the causal relations.

The main advantage of interventions therefore is not so much their surgical feature, but that they introduce an external influence into the system under investigation, which helps to disentangle causal relations and, importantly, to orient causal influences, if no time order is available in the data.

There is an additional set of interesting questions that I can only hint at in this overview: Unlike the passive observational setting in which questions of efficient discovery are limited to sample size and the computational aspects of model selection (e.g. the order of statistical tests, such as independence tests, or the computation of scores of particular models), we have to additionally decide in the experimental setting which experiments to perform, i.e. which variables to subject to an intervention, and in what order to perform the experiments. Although these questions do not necessarily imply a difference in the limit of what can be discovered in any sequence of experiments, it may well make a practical difference in terms of what can be discovered in a finite number of experiments or with a finite amount of resources. One example of a result of this kind is that one might naively think that the most efficient way to discover a causal relations among a set of variables is to intervene on one variable while holding all but one other variable fixed at a certain value. In fact it turns out (and has been known at least since the seminal work on experimental design by R.A. Fisher [8]) that much more efficient sequences of experiments are possible

when multiple variables are subject to an intervention simultaneously in each experiment (see also [6]). More generally this raises many questions of how aspects of decision theory can guide causal discovery.[2]

Needless to say, there are many other forms of assumptions that I have not discussed here (e.g. time order, determinism, cyclicity etc.) which have their own implications for causal discovery. In each case, various procedures are already available.

5 Conclusion

This brief survey was intended to suggest a perspective that seems to have been lost in the philosophical discussions focused on particular assumptions: There are many assumptions beyond causal Markov and faithfulness that are used for causal discovery. These assumptions can be combined in different ways that enable a variety of search procedures that (in the large sample limit) are guaranteed to yield different insights into the underlying causal structure. In some cases, combinations of assumptions can be used to make others, such as faithfulness or causal sufficiency, redundant. In many cases the exact implications for discovery given a set of assumptions have been worked out already, but many open questions remain, since general conclusions for all discovery procedures are not available.

While I cannot even begin to list the most important open problems for causal discovery here, this context lends itself to point to one area in desperate need of further investigation: the robustness of procedures. I have indicated that discovery procedures can be based on a variety of different assumptions. But very little is known about how these procedures perform when one of their assumptions fails or is incompletely satisfied. Since there is no obvious hierarchy of assumptions, one cannot always guarantee that one can fall back into a safety net of weaker assumptions. At present we simply do not know how well procedures perform when some of their assumptions fail, or what the nature of their errors then is. Obviously, an epistemology of causation would benefit enormously from a better understanding of this matter.

References

[1] N. Cartwright. *How the Laws of Physics Lie*. Oxford University Press, 1983.

[2] N. Cartwright. What is wrong with Bayes nets? *The Monist*, 84:242–264, 2001.

[2]Note that causal decision theory considers a decision theory that is guided by causal constraints. The suggestion here is that there are many open questions concerning causal discovery that is guided by decision theory.

[3] N. Cartwright. Against modularity, the causal Markov condition and any link between the two. *British Journal of Philosophy of Science*, 53:411–453, 2002.

[4] N. Cartwright. From metaphysics to method: Comments on manipulability and the causal Markov condition. *British Journal of Philosophy of Science*, 57:197–218, 2006.

[5] F. Eberhardt. *Causation and Intervention*. PhD thesis, Department of Philosophy, Carnegie Mellon University, 2007.

[6] F. Eberhardt, C. Glymour, and R. Scheines. On the number of experiments sufficient and in the worst case necessary to identify all causal relations among n variables. In F. Bacchus and T. Jaakkola, editors, *Proceedings of the 21st Conference on Uncertainty and Artificial Intelligence(UAI-05)*, pages 178–184. AUAI Press, Corvallis, Oregon, 2005.

[7] F. Eberhardt and R. Scheines. Interventions and causal inference. *Philosophy of Science*, 74(5):981–995, 2007.

[8] R. Fisher. *The design of experiments*. Hafner, 1935.

[9] C. Glymour. Markov properties and quantum experiments. In W. Demopoulos and I. Pitowsky, editors, *Physical Theory and its Interpretation*, volume 72, pages 117–126. Springer, 2006. Series title: The Western Ontario Series in Philosophy of Science.

[10] C. Glymour. Review of Jon Williamson, *Bayesian Nets and Causality*, Oxford University Press, 2005. forthcoming in British Journal of Philosophy of Science, 2008.

[11] D. Hausman and J. Woodward. Independence, invariance, and the causal markov condition. *British Journal for the Philosophy of Science*, 50:521–583, 1999.

[12] D. Hausman and J. Woodward. Modularity and the causal Markov condition: A restatement. *British Journal for the Philosophy of Science*, 55:147–161, 2004.

[13] F. Heider and M. Simmel. An experimental study of apparent behaviour. *American Journal of Psychology*, 13(2):243–259, 1944.

[14] K. Hoover. Nonstationary time series, cointegration, and the Principle of Common Cause. *British Journal of Philosophy of Science*, 54:527–551, 2003.

[15] P. O. Hoyer, S. Shimizu, and A. J. Kerminen. Estimation of linear, non-gaussian causal models in the presence of confounding latent variables. In *Proceedings of 3rd European Workshop on Probabilistic Graphical Models (PGM-06)*, pages 155–162, 2006.

[16] P. O. Hoyer, S. Shimizu, A. J. Kerminen, and M. Palviainen. Estimation of causal effects using linear non-gaussian causal models with hidden variables. *Intl. Journal of Approximate Reasoning*, 49:362–378, 2008.

[17] D. Hume. *An enquiry concerning human understanding*. Clarendon, 1748.

[18] G. Lacerda, P. Spirtes, J. Ramsey, and P. O. Hoyer. Discovering cyclic causal models by independent components analysis. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI-08)*, pages 366–374, 2008.

[19] A. Michotte. *La perception de la causalité*. Publications Universitaires de Louvain, 1954.

[20] E. Nyberg and K. Korb. Informative interventions. In F. Russo and J. Williamson, editors, *Causality and Probability in the Sciences*. College Publications, London, 2006.

[21] J. Pearl. *Causality*. Oxford University Press, 2000.

[22] H. Reichenbach. *The Theory of Probability. An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability*. University of California Press, 1949.

[23] H. Reichenbach. *The Direction of Time*. University of California Press, 1956.

[24] J. Reiss. Causation: An opinionated introduction. draft available: http://jreiss.org/Causality%20Manuscript.pdf.

[25] S. Shimizu, P. O. Hoyer, A. Hyvärinen, and A. J. Kerminen. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7:2003–2030, 2006.

[26] S. Shimizu, A. Hyvärinen, Y. Kano, and P. O. Hoyer. Discovery of non-gaussian linear causal models using ICA. In *Proceedings of the 21st Conference in Uncertainty in Artificial Intelligence (UAI-05)*, pages 525–533. AUAI Press, 2005.

[27] R. Silva, R. Scheines, C. Glymour, and P. Spirtes. Learning the structure of linear latent variable models. *Journal of Machine Learning Research*, 7:191–246, 2006.

[28] E. Sober. Venetian sea levels, British bread prices, and the Principle of Common Cause. *British Journal of Philosophy of Science*, 52:331–346, 2001.

[29] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction and Search*. MIT Press, 2 edition, 2000.

[30] D. Steel. Indeterminism and the causal Markov condition. *British Journal of Philosophy of Science*, 56:3–26, 2005.

[31] J. Woodward. *Making Things Happen*. Oxford University Press, 2003.

[32] J. Zhang. Consistency in causal inference under a variety of assumptions. Master's thesis, Carnegie Mellon University, 2002.

[33] J. Zhang and P. Spirtes. Detection of unfaithfulness and robust causal inference. *Minds and Machines*, 18(2):239–271, 2008.