Determinism and the Method of Difference[*]

Urs HOFMANN and Michael BAUMGARTNER

Received: 01.11.2010
Final Version: 31.01.2011

BIBLID [0495-4548 (2011) 26: 71; pp. 155-176]

ABSTRACT: The first part of this paper reveals a conflict between the core principles of deterministic causation and the standard method of difference, which is widely seen (and used) as a correct method of causally analyzing deterministic structures. We show that applying the method of difference to deterministic structures can give rise to causal inferences that contradict the principles of deterministic causation. The second part then locates the source of this conflict in an inference rule implemented in the method of difference according to which factors that can make a difference to investigated effects relative to one particular test setup are to be identified as causes, provided the causal background of the corresponding setup is homogeneous. The paper ends by modifying the method of difference in a way that renders it compatible with the principles of deterministic causation.

Keywords: method of difference; deterministic causation; interventionism; causal discovery; causal reasoning.

1. *Introduction*

In contrast to the popularity of methods of causal discovery that implement Bayesian networks and analyze probabilistic data, the problem of causally interpreting deterministic dependencies among factors or variables has received comparably little attention in recent years.[1] On the face of it, this reduced interest in the causal analysis of deterministic data is remarkable for at least two reasons. First, analyzing deterministic data cannot be considered a special case of analyzing probabilistic data by means of Bayes-nets methods, because deterministic dependencies give rise to violations of the so-called *faithfulness* assumption which, in one way or another, is presupposed by all Bayes-nets methods (Spirtes et al. 2000; Pearl 2000; Ramsey et al. 2006; Glymour 2007). In consequence, the latter are not applicable to deterministic dependencies. Second, probabilities in empirical data on macroscopic causal processes are commonly seen to be due to mere epistemic limitations. Ontically, myriads of macroscopic processes are taken to be of deterministic nature, which, accordingly, constitute a very widespread type of phenomenon.

---

[*] We thank Craig Callender, Sebastian Leugger, Alexandre Marcellesi, Fabio Molo, Wolfgang Spohn, Christian Wüthrich and two anonymous referees of this journal for very helpful comments on earlier drafts. Moreover, we have profited a lot from discussions at the philosophy of science research colloquia at the University of California at San Diego and the University of Konstanz. Finally, Michael Baumgartner is indebted to the Deutsche Forschungsgemeinschaft (DFG) for generous support of this work (project *CausaProba*).

[1] Among the few studies that explicitly focus on the discovery of deterministic dependencies are (Luo 2006), (Glymour 2007), or (Baumgartner 2009).

Nonetheless, explanations for the little attention deterministic methodologies have received as of late are not difficult to come by. For one, deterministic dependencies, notwithstanding their (ontic) prevalence, rarely (phenomenally) manifest themselves in data. Ordinary causal structures are of such high complexity and so sensitive to confounding influences that data are seldom homogeneous enough to actually exhibit deterministic dependencies. Only data that are collected against highly controlled causal backgrounds, as for instance given in specific laboratory contexts, de facto feature deterministic relations. Furthermore, in homogeneous laboratory contexts causal reasoning is normally considered to be much less problematic than in contexts with uncontrolled causal backgrounds. Laboratory contexts permit systematic manipulations of investigated factors which renders it possible to uncover causal structures along the lines of the well-established method of difference (MoD). Even though, since the times of Mill (1843), MoD has repeatedly been adapted to modern theories of causation and to the constraints of modern scientific practice (cf. Ragin 1987, 2000, 2008; May 1999; Woodward 2003; Baumgartner 2009), the basic idea behind the method has remained unaltered over the past 160 years. Roughly, MoD determines a factor A to be causally relevant to a factor E, if a manipulation of A in a first test situation $\mathcal{S}_1$ is followed by a variation of E, while in a second test situation $\mathcal{S}_2$ that lacks a manipulation of A and that is causally homogeneous with $\mathcal{S}_1$, i.e. that accords with $\mathcal{S}_1$ in regard to causes of E not located on a path from A to E, no variation of E occurs. MoD is generally considered to be a *correct* method to uncover deterministic causal structures; that is, if it is applied to deterministic structures and yields that a factor A is causally relevant to a factor E, this causal dependency indeed exists. That is, within homogeneous laboratory contexts that allow for systematic manipulations of deterministic structures a simple rule that induces reliable causal inferences is usually presumed to be available.

This paper shows that reliably uncovering deterministic structures, even under perfectly controlled circumstances, is not as straightforward as it may seem at first sight. We shall argue that in case of deterministic dependencies that are investigated against homogeneous causal backgrounds the correctness of the method of difference is far from obvious. More specifically, the paper exhibits that causal inferences drawn on the basis of MoD may conflict with fundamental principles that are commonly taken to characterize deterministic causation, as the principle of determinism ("Same cause, same effect"), the principle of causality ("No effect without at least one of its causes"), or the principle of non-redundancy ("Causal structures do not contain redundant elements"). We are going to present a simple deterministic process such that, when this process is investigated under ideally homogeneous conditions, MoD yields that a particular factor A is (part of) a deterministic cause of another factor E, where, in fact, such a dependency violates at least one of the core principles of deterministic causation. Hence, the claim that MoD is a correct method to uncover deterministic causal structures and the claim that the principles of determinism, causality, and non-redundancy all hold for such structures are incompatible.

The second part of the paper then locates the source of this conflict in an inference rule that has been implemented, more or less explicitly, in all available formulations of MoD—most prominently in modern interventionist methodologies: if there exists *at least one* manipulation of an investigated cause factor A with respect to one particular

test setup such that this manipulation is followed by a change in the effect E, A is causally relevant to E, provided that the causal background of the corresponding setup is homogeneous. We take the incompatibility of available variants of MoD and the principles of deterministic causation to count against the correctness of this inference rule. The paper ends by replacing this rule by an alternative one that renders the method of difference compatible with the principles of deterministic causation.

Section 2 presents the relevant principles of deterministic causation and establishes their intuitive plausibility. In section 3, we exhibit the details of the method of difference as it has been conceived in modern studies on causal reasoning. Section 4 then introduces the conflict between the principles of deterministic causation and inferences induced by MoD. Finally, section 5 suggests a modification of MoD that resolves the conflict.

2. *The Principles of Deterministic Causation*

For the purposes of this paper, we do not have to presuppose a full-blown theory of deterministic causation, rather it suffices to introduce three principles a causal structure has to satisfy in order to be of deterministic nature. That is, we can confine our discussion to three necessary conditions of deterministic causation.

To present the details of those principles, some conceptual preliminaries are required. We are going to focus on causation on type level in this paper. Moreover, for simplicity we shall only consider causal models that exclusively involve binary variables, which we call *factors* for short. A causal analysis must be relativized to a set of examined factors, which shall be referred to as the *factor frame* of the analysis. Factors are taken to be similarity sets of event tokens. They are sets of type identical token events. Whenever a member of such a similarity set occurs, the corresponding factor is said to be *instantiated*. Factors are symbolized by italicized capital letters A, B, C, etc. Factors that are related in terms of type-level causation are not related in terms of some metaphysically stronger form of dependence, as logical dependence, supervenience, emergence, mereological containment or the like, i.e. they are *non-causally independent*. As absences are often causally interpreted as well, we take factors to be negatable. The negation of a factor A is written thus: $\overline{A}$. $\overline{A}$ simply represents the absence of an instance of A. Alternatively, factors can be seen as binary variables that take the value 1 whenever an event of the corresponding type occurs and the value 0 whenever no such event occurs.

Ordinarily, deterministic causes are highly complex and one effect type may be brought about by several alternative causes. Deterministic causes are parts of whole causing *compounds*. A compound only becomes *causally effective*, i.e. actually brings about its effect, if all of its constituents are co-instantiated, i.e. instantiated spatiotemporally close-by or coincidently. Coincidently instantiated factors are instantiated in the same *situation*.[2] Often, not all factors contained in a deterministic cause are known or of interest in a corresponding context of causal discovery. Compounds shall be symbolized

[2] Which spatiotemporal interval determines what counts as *one situation* is notoriously vague and depends on the specificity of the causal structure under investigation. We are not going to address this question here, but are simply going to assume that the structures discussed in this paper are sufficiently well known that the coincidence relation is properly interpretable. For more details cf. Baumgartner (2008).

by simple concatenations of factors, with placeholders $\mathcal{X}$, $\mathcal{X}_1$, $\mathcal{X}_2$ etc. standing for open sequences of unknown or unmentioned factors, for example $ABC\mathcal{X}_1$. If A is part of a compound which is a deterministic cause of E, A is said to be *causally relevant* to E. The set σ of relevance relations holding among the factors contained in a given factor frame μ (relative to pertinent data) constitutes a *causal structure* over μ. By a *deterministic causal structure* we mean a causal structure that only comprises deterministic dependencies.[3] Finally, a factor A is said to be *exogenous* relative to a structure μ iff no other factors Z_i in μ are causally relevant to A.

Based on this conceptual background, we can now state the relevant principles of deterministic causation. If a compound $\mathcal{X}$ is said to be a deterministic (type-level) cause of a factor E, what is claimed, among other things, is that coincident instantiations of the components of $\mathcal{X}$ determine E to be instantiated. That is, whenever the factors in $\mathcal{X}$ are instantiated in a situation $\mathcal{S}_i$ there also is an instance of E in $\mathcal{S}_i$. Generalizing this conditional for whole causal structures yields our first principle:

Determinism (D): If a causal structure σ is deterministic, any two situations $\mathcal{S}_i$ and $\mathcal{S}_j$ that accord with respect to instantiations of exogenous factors in σ accord with respect to instantiations of *all* factors in σ.

Second, causal structures satisfy the principle of causality, according to which effects do not occur spontaneously, i.e. without at least one of their alternative causes being instantiated as well. Adapting this principle to deterministic structures yields:[4]

Causality (C): If a factor E is an effect within a deterministic causal structure σ, E is not instantiated in a situation $\mathcal{S}_i$ without at least one of its alternative (complex) causes in σ being instantiated in $\mathcal{S}_i$ as well.

And third, deterministic structures do not feature redundancies. To illustrate this principle of non-redundancy, assume that striking a match, factor S, in combination with the presence of oxygen, O, and the dryness of the match, D, determines the match to catch fire, F. It then also holds that the compound $SODQ$ determines F to be instantiated, where Q stands for any arbitrary factor like singing a song or baptizing an elephant. However, the deterministic cause of F is not $SODQ$, but SOD. $SODQ$ has a proper part, *viz.* SOD, which alone determines F. Q is redundant. Or suppose we define a factor A such that A is instantiated if and only if the match is struck or no oxygen is present: $A \leftrightarrow S \vee \overline{O}$. It then follows that AOD also determines the match to catch fire, for whenever A occurs in combination with OD, A must be instantiated by a struck match and not by the absence of oxygen, because oxygen cannot be both present and absent in the same situation. Nonetheless, one disjunct in the definiens of A plays no causal role whatsoever for the lighting of matches. For mere logical reasons, instances

[3] Hence, we limit our discussion in this paper to *completely* deterministic structures, that is, to causal structures that do not contain both deterministic and indeterministic dependencies. For a treatment of so-called *semi-deterministic* structures cf. e.g. (Luo 2006).

[4] Often, the principles of determinism and causality are combined to one principle of deterministic causation in the literature. We furnish them with different labels here for the mere purpose of facilitated reference later on.

of $\overline{O}$ cannot be co-instantiated with instances of OD, hence, they are redundant for the bringing about of F. If SOD is a deterministic cause of F, it follows, among other things, that any instances of S, of O, and of D can occur in the same situation and that, if that is the case, F is instantiated as well. In other words, deterministic causes only comprise factors all of whose instances are *compossible*. This constraint is violated by AOD. Finally, assume that whenever F is instantiated, either SOD is instantiated or the dry match is exposed to some chemical C while oxygen is present, i.e. $F \rightarrow SOD \vee COD$. Now let factor G be defined such that G is instantiated if and only if the match is both struck and exposed to the flammable chemical: $G \leftrightarrow S \wedge C$. It follows that whenever F is instantiated, so is the disjunction $SOD \vee COD \vee GOD$. Yet, analogously to the redundant Q or the redundant instances of A, GOD makes no difference to F over and above $SOD \vee COD$. GOD is not an additional alternative cause of F, i.e. it is redundant.

For mere logical reasons it is excluded that Q, a proper subset of A, and GOD are ever indispensable for the bringing about of F. Yet, each element of a deterministic causal structure σ possibly makes a difference to the effects of σ. Deterministic structures do not contain elements that are dispensable for mere logical reasons.

Non-Redundancy (NR): If a causal structure σ comprising the set ε of effects is of deterministic nature, σ only involves factors Z_i that are indispensable for the bringing about of the members of ε in *at least one* possible situation $\mathcal{S}_m$, such that *any* instance of Z_i is causally effective in $\mathcal{S}_m$.

Combining (NR) with (D) and (C), respectively, has implications that allow for a convenient formal aggregation of the three principles. According to (D), a compound $\mathcal{X}$ which is a deterministic cause of a factor E is a *sufficient* condition of E, i.e. $\mathcal{X} \rightarrow E$. As to (NR), such a sufficient condition must not contain redundancies. That is, first, it must not be the case that a proper part α of $\mathcal{X}$ is itself sufficient for E, where a proper part of the conjunction $\mathcal{X}$ amounts to a reduction of $\mathcal{X}$ by at least one conjunct. Hence, $\alpha \rightarrow E$ must be false for all proper parts α of $\mathcal{X}$. If $\mathcal{X}$ satisfies that constraint, $\mathcal{X}$ is a *minimally sufficient* condition of E.[5] Second, no component of $\mathcal{X}$ must have a subset of instances that, for logical reasons, cannot be co-instantiated with the other factors in $\mathcal{X}$. All instances of the components of $\mathcal{X}$ must be compossible. Deterministic causes hence are minimally sufficient conditions of their effects, such that all of the instances of their component factors are compossible.

Furthermore, according to (C), the disjunction of all alternative deterministic causes $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n$ of an effect E constitutes a necessary condition of E, i.e. $E \rightarrow \mathcal{X}_1 \vee \mathcal{X}_2 \vee \cdots \vee \mathcal{X}_n$. Subject to (NR), such a necessary condition must not contain redundancies. More specifically, it must not be the case that a proper part β of $\mathcal{X}_1 \vee \mathcal{X}_2 \vee \cdots \vee \mathcal{X}_n$, i.e. $\mathcal{X}_1 \vee \mathcal{X}_2 \vee \cdots \vee \mathcal{X}_n$ reduced by at least one disjunct, is itself necessary for E. That is, $E \rightarrow \beta$ must be false for all proper parts β of $\mathcal{X}_1 \vee \mathcal{X}_2 \vee \cdots \vee \mathcal{X}_n$. If $\mathcal{X}_1 \vee \mathcal{X}_2 \vee \cdots \vee \mathcal{X}_n$ satisfies that constraint, it is a *minimally necessary* condition of E. In sum, deterministic causal structures can be represented by a double-conditional of type (1), where (i) each compound $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n$ is composed of factors all of whose

[5] Cf. (Broad 1930; Mackie 1974; Graßhoff and May 2001; Baumgartner 2008).

instances are compossible, (ii) each compound $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n$ is a minimally sufficient condition of E, and (iii) $\mathcal{X}_1 \vee \mathcal{X}_2 \vee \cdots \vee \mathcal{X}_n$ is minimally necessary for E.

$$\mathcal{X}_1 \vee \mathcal{X}_2 \vee \cdots \vee \mathcal{X}_n \Leftrightarrow E\,. \tag{1}$$

For brevity, we refer to such double-conditionals that satisfy (NR) as *minimal theories* of E (cf. Graßhoff and May 2001; Baumgartner 2008).

Deterministic structures can be represented on various levels of specification, i.e. by various minimal theories. To illustrate, reconsider the structure regulating the lighting of matches. It can be modeled by this (rather coarse-grained) minimal theory:

$$SOD \vee COD \Leftrightarrow F\,. \tag{2}$$

There also exist more fine-grained descriptions, as can e.g. be attained by specifying factors involved in (2). For instance, there exist two types of matches: matches whose head is made of red phosphorus, and others whose head consists of phosphorus sesquisulfide. That is, the set of events represented by the factor "striking a match" (S) can be decomposed into the subset of events of type "striking a red phosphorus match" (S_1) and the subset of events of type "striking a phosphorus sesquisulfide match" (S_2). Decomposing S in this vein yields a more fine-grained minimal theory:

$$S_1OD \vee S_2OD \vee COD \Leftrightarrow F\,. \tag{3}$$

While there may be numerous minimal theories that adequately represent a deterministic structure σ, subject to (D), (C), and (NR) it holds that there exists *at least one* minimal theory for every σ. Or differently, the principles of deterministic causation entail:

Existence of a Minimal Theory (MT): If a factor A is part of a deterministic cause of E, there exists at least one minimal theory Φ such that A is part of at least one disjunct in the antecedent of Φ, i.e. $A\mathcal{X}_1 \vee \mathcal{X}_2 \vee \cdots \vee \mathcal{X}_n \Leftrightarrow E$. [*from (D), (C), (NR)*]

In order to show that the principles of deterministic causation can conflict with inferences drawn on the basis of the method of difference, it must—for obvious reasons—be guaranteed that there in fact exist deterministic causal structures in nature. That is, we moreover need the following widely accepted assumption:[6]

Existence of Deterministic Structures (ED): On macro levels, i.e. on levels above the quantum domain, there exist causal structures that are ultimately of deterministic nature.

[6] As is well known, there is an ongoing debate in the literature as to whether, if we choose to adopt an indeterministic interpretation of quantum mechanics, we are forced to settle for universal macro indeterminism as well. For a survey of the debate from a neurobiological perspective cf. Weber (2005). In the literature on causal reasoning, the existence of deterministic structures on macro levels is uncontroversial (cf. Glymour 2007).

For each of these deterministic causal structures there exists at least one minimal theory. If that is not the case for a particular structure, it is not a deterministic causal structure. Given the existence of deterministic structures, this innocuous presupposition is all we need for the sequel of the argument developed in this paper.

3. The Method of Difference

The standard method to uncover deterministic structures in controlled experimental contexts dates back to Mill (1843, vol. I, 455): the method of difference (MoD). The kernel of MoD has remained unaltered over the past 160 years: by comparison of test situations that agree in relevant respects except for instantiations of investigated causes and effects, MoD experimentally reveals causal dependencies. In this section, this basic methodological approach is made more explicit and precise.

It is virtually a truism of causal reasoning that correlations among instantiations of two factors A and E—even perfect correlations—are not sufficient for a causal dependency between A and E. Correlations of A and E can also result from uncontrolled variations of common causes of A and E. More generally, systematic covariations of A and E may either be due to a causal dependency between A and E or to the uncontrolled behavior of so-called *confounders*, where a confounder is a cause of an investigated effect that can change the value of that effect independently of the factors in an analyzed frame. Tailored to the analysis of deterministic structures, the notion of a confounder can be more precisely defined as follows: A factor O is a confounder of an effect Z_n relative to an analyzed factor frame $\{Z_1, \ldots, Z_n\}$ iff O is located on a causal path to Z_n on which none of the factors $Z_1, \ldots, Z_{n-1}$ are located.

In order to infer causal dependencies from covariations, test situations must be compared that are uniform with respect to instantiations of confounders. Test situations that satisfy this constraint are termed *causally homogeneous*:

Causal Homogeneity (CH): Two test situations $\mathcal{S}_i$ and $\mathcal{S}_j$ that are compared in order to investigate the causal structure behind the behavior of an effect Z_n relative to the frame $\{Z_1, \ldots, Z_n\}$ are causally homogeneous iff $\mathcal{S}_i$ and $\mathcal{S}_j$ agree with respect to instantiations of confounders of Z_n relative to $\{Z_1, \ldots, Z_n\}$.

Given two causally homogeneous test situations $\mathcal{S}_1$ and $\mathcal{S}_2$, the method of difference requires that in $\mathcal{S}_1$ the value of at least one of the factors $Z_1, \ldots, Z_{n-1}$ is changed by intervention, while in $\mathcal{S}_2$ no such interventions are performed (or vice versa). Intervening on a factor Z_1 amounts to surgically inducing Z_1 to change its value—most of all, interventions on Z_1 are not connected to the analyzed effect on a causal path that does not go through Z_1 (cf. Woodward 2003, 98). If the interventions in $\mathcal{S}_1$ then turn out to be accompanied by a change in the value of Z_n while no such change occurs in $\mathcal{S}_2$, it follows that the manipulated factors are causally relevant to Z_n (or $\overline{Z_n}$, respectively). This can be seen by the following reasoning. According to the principle of causality, the change in the value of Z_n in $\mathcal{S}_1$ does not occur spontaneously, that is, it must have a cause. Provided that Z_n indeed is an effect of a deterministic structure, the fact that the value of Z_n remains unaltered in $\mathcal{S}_2$ implies that no cause of Z_n—most of all,

no uncontrolled confounder of Z_n—is instantiated in $\mathcal{S}_2$. From this, in combination with the causal homogeneity of $\mathcal{S}_1$ and $\mathcal{S}_2$, it follows that no uncontrolled variation of a confounder of Z_n accounts for Z_n changing its value in $\mathcal{S}_1$. The only remaining difference that can possibly account for the change of the value of Z_n in $\mathcal{S}_1$ then are the intervention-induced changes in $\mathcal{S}_1$. Therefore, the manipulated factors are parts of complex causes of Z_n, i.e. they are causally relevant for Z_n. This, in general terms, is the method of difference.

To make things more concrete, let us illustrate causal reasoning based on MoD by means of the simplest possible test design: suppose we want to investigate whether a single factor A is causally relevant to an effect E. The investigated factor frame for our exemplary case, hence, shall be $\{A, E\}$. First, we need two test situations $\mathcal{S}_1$ and $\mathcal{S}_2$ that are causally homogeneous for E with respect to $\{A, E\}$. Since homogeneity amounts to uniformity of confounders across $\mathcal{S}_1$ and $\mathcal{S}_2$ and since confounders are (per definition) not controlled for in our test design, the satisfaction of (CH) by $\mathcal{S}_1$ and $\mathcal{S}_2$ can only be ascertained in idealized contexts. In real-life experimental circumstances homogeneity can merely be rendered more or less plausible, for instance, by means of randomization or isolation of experimental setups in laboratory environments. As we are only going to be concerned with idealized discovery contexts in this paper, we shall simply assume the availability of test situations $\mathcal{S}_1$ and $\mathcal{S}_2$ that are causally homogeneous for E with respect to $\{A, E\}$. Next, MoD calls for an intervention on A that induces an instantiation of A in one of the two test situations, i.e. that sets A's value to 1, while A is left uninstantiated in the other situation. For simplicity, we stipulate that, throughout this paper, investigated cause factors are always instantiated in $\mathcal{S}_1$ and absent in $\mathcal{S}_2$. In sum, this is the design of the simplest application of MoD. To facilitate later reference, we refer to this test design as a *difference test*, or a *d-test* for short, and to the homogeneous configuration of background factors as a *d-test setup*.[7]

Relative to a pair of situations $\langle \mathcal{S}_1, \mathcal{S}_2 \rangle$ such that A is instantiated in $\mathcal{S}_1$ and absent in $\mathcal{S}_2$, an investigated effect factor E can be instantiated and absent in four possible configurations. If E is instantiated in $\mathcal{S}_1$ and absent in $\mathcal{S}_2$, the d-test is said to generate a *1-0-outcome*, where "1" symbolizes an instantiation and "0" a non-instantiation of E. Likewise, a d-test can yield a *0-1-outcome*, a *1-1-outcome*, or a *0-0-outcome*. Each of these four descriptions of d-test outcomes is complete, i.e. it provides all relevant information about a corresponding d-test outcome.

Only two of these four possible outcomes are causally interpretable. A 1-0-outcome induces an inference to the causal relevance of A for E: as E occurs in $\mathcal{S}_1$, at least one cause of E must be instantiated in $\mathcal{S}_1$; necessarily, A is part of that (these) cause(s), for otherwise E would occur in the homogeneous situation $\mathcal{S}_2$ as well. Based on an analogous reasoning, an 0-1-outcome entails the causal relevance of $\overline{A}$ for E. By contrast, 0-0- and 1-1-outcomes are *not* causally interpretable, because they are compatible with incompatible causal models. A 0-0-outcome could result from the simple causal irrelevance of A.

[7] Note that as this test design involves a comparison of two test situations only, it presupposes that investigated causal structures are encoded in binary terms. Often, more complex test designs for MoD also require binary encodings, for this allows for a straightforward implementation of Boolean optimization techniques (cf. Baumgartner 2009).

However, such an outcome could also be due to the fact that, even though A indeed is part of a complex cause $\mathcal{X}_i$ of E, not all of the other factors in $\mathcal{X}_i$ are instantiated along with A in $\mathcal{S}_1$, such that $\mathcal{X}_i$ does not become causally effective in $\mathcal{S}_1$. Finally, a 1-1-outcome is realized whenever at least one confounder of E with respect to $\{A, E\}$ is instantiated in both test situations, irrespective of whether or not A is part of a cause of E. In short, outcomes of types 0-0 and 1-1 are compatible both with A being causally relevant to E and with A not being causally relevant to E.

One consequence of the method of difference deserves separate mention at this point: according to MoD, one single intervention is sufficient to establish A or $\overline{A}$ as cause of E, provided that a corresponding d-test produces an 1-0- or 0-1-outcome. Woodward (2003, 59) has recently restated (or modally generalized) this consequence—which is already contained in Mill's original formulations of MoD—in the following often cited passage taken from the definitional core of Woodward's acclaimed interventionist theory of causation:[8]

> A necessary and sufficient condition for X to be a (type-level) direct cause of Y with respect to a variable set $\mathbf{V}$ is that there be a possible intervention on X that will change Y or the probability distribution of Y when one holds fixed at some value all other variables Z_i in $\mathbf{V}$.

That is, relative to an appropriate test setup it holds that if the effect variable changes its value after at least one intervention on the investigated cause variable, the latter is entailed to be a cause of the former according to MoD.

Note that the design of d-tests constitutes the simplest possible application of the method of difference. MoD allows for uncovering causal structures of arbitrary complexity. Analyzing more extensive factor frames, however, requires more intricate test designs. Since we are going to focus on deterministic dependencies between pairs of factors in the following, we can sidestep more complex test designs here.[9] All that matters for our purposes is that according to MoD a single d-test result of type 1-0 implies the causal relevance of A for E, given the causal homogeneity of corresponding test situations and given that E is an effect of a deterministic causal structure in the first place. Hence, in case of simple d-tests the method of difference infers causal relevance relationships based on the following inference rule:

Difference-making (DM): A factor A is causally relevant to a factor E if there exists at least one d-test setup δ such that intervening on A with respect to E in one test situation of type δ generates an 1-0-outcome.

Finally, a caveat must be introduced that will turn out to be relevant in the remainder of this paper: factors analyzed by means of MoD must be *suitable for causal modeling*. Procedures of causal reasoning can only be expected to produce adequate outputs if the variables that are fed into the procedures satisfy certain conditions. For instance, defining

[8] We can confine ourselves to Woodward's account of direct causation here, because his analysis of indirect (or contributing) causation introduces no additional elements that would be relevant for our current purposes. The *variable set* $\mathbf{V}$ Woodward mentions in this passage is what we have been calling the *factor frame* in this paper.

[9] For more details on uncovering complex causal structures on the basis of MoD cf. Baumgartner (2009).

a factor B as $\overline{C} \vee E$, i.e. $B \leftrightarrow \overline{C} \vee E$, yields dependencies between B and E that we do not want to interpret causally, even though there exist d-test setups relative to which intervening on B with respect to E generates a 1-0-outcome. B is not a suitable cause factor for E, because B represents a disjunctive and gerrymandered property some of whose instances have no resemblances whatsoever. Moreover, B and E are logically dependent (which causes and effects are not). Unfortunately, suitability conditions are commonly not explicitly discussed in the relevant literature—often, discussions of such conditions are explicitly sidestepped (cf. Spirtes et al. 2000, 21, 91-92). There exist only a few gradual constraints on the suitability of variables: causally modeled variables must not be non-causally dependent (cf. sect. 2), they must neither represent gerrymandered nor gruelike properties some of whose instances have no resemblances (cf. Lewis 1999; Fodor 1997; Glynn forthcoming), rather they must stand for (imperfectly) natural properties all of whose instances mutually resemble each other (cf. Lewis 1999), or they should be salient in a corresponding research context (cf. Halpern and Hitchcock 2010). Plainly, all of these conditions are vague and only yield suitability by degree. The only precise thing that can be said about the suitability of variables is that, on pain of circularity and of trivializing corresponding procedures of causal reasoning, suitability must not be characterized in causal terms. That is, A must not only count as a suitable causal variable with respect to E if A in fact causes E (or vice versa). In sum, every application of MoD must be preceded by identifying a suitable factor frame consisting of factors that are reasonably natural and salient relative to the relevant context of causal discovery.

4. *The Conflict*

In this section we show that causal inferences drawn on the basis of the method of difference can conflict with the principles of deterministic causation. To this end, we introduce a very simple electric circuit, i.e. an instance of an electrodynamic causal structure, that we assume to be one of those deterministic structures claimed to exist by (ED). Nothing hinges on the electrodynamic nature of our example. All that matters for our purposes is its particular causal structuring. Thus, should the reader, for whatever reason, take electrodynamic processes to be irreducibly indeterministic, she may simply substitute any other deterministic process of the same structuring. Furthermore, we presume to analyze that electric circuit under idealized conditions in which we have complete control over all relevant factors. Based on these assumptions, it will turn out that d-tests conducted on our sample structure induce inferences to deterministic dependencies that do not satisfy all principles of deterministic causation. Hence, the presumption that MoD is a correct method of causal discovery and that our electric circuit is of deterministic nature, on the one hand, and (D), (C), and (NR), on the other, imply a contradiction.

Our sample circuit is depicted in figure 1. The burning of a light bulb "⊗" is regulated by two electric subcircuits, one on the left-hand side and one on the right-hand side. Both subcircuits are powered by a battery "|ı" (b1 and b2, respectively), which shall be assumed to be fully charged by default in the following. The light is on iff either the left

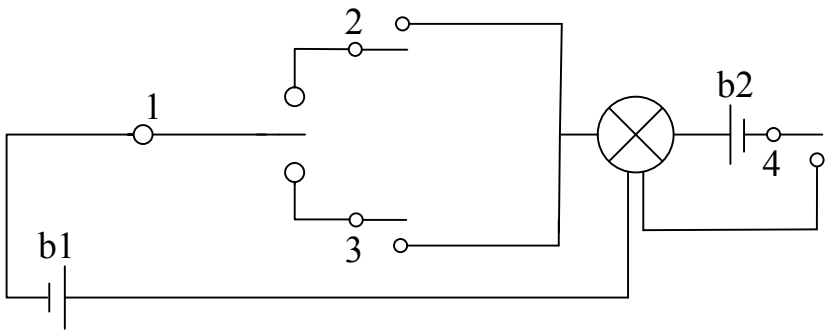

Figure 1: An electric circuit.

or the right subcircuit is closed.[10] The left circuit is closed iff switch 2 is closed while switch 1 is closed upwards or switch 3 is closed while switch 1 is closed downwards. The right circuit is closed iff switch 4 is closed. The structure is assumed to be complete, that is, there are no other ways to turn the lamp on. Furthermore, it is presumed that all switches can only be either open or closed. There is no such thing as a half-closed switch.

As indicated in section 2, causal structures can be analyzed on different levels of specification, i.e. relative to different factor frames. When it comes to causally analyzing our electric circuit, differences in the specificity or grain of the analysis turn out to be of particular importance, as shall be shown in what follows. To begin with, suppose we choose to analyze the circuit relative to the following factor frame, which we assume to be suitable for our purposes:

$(\mathcal{F}_1)$

A_1 : switch 1 closed upwards	F : battery b1 charged
A_2 : switch 1 closed downwards	K : battery b2 charged
B : switch 2 closed	H : switch 4 closed
C : switch 3 closed	E : light on

Can factors A_1 and A_2 be said to be causally relevant to E relative to $\mathcal{F}_1$? We first answer this question based on the method of difference. To this end, we need a proper d-test setup, i.e. homogeneous test situations for E with respect to $\mathcal{F}_1$. Within the idealized laboratory context presumed for our example such a setup is not hard to come by. For instance, take two situations in which switches 2 and 3 are closed, switch 4 is open, and both batteries are fully charged. If we now intervene to instantiate A_1 or A_2, respectively, in one of the two situations, while, in the other, A_1 and A_2 are not instantiated, we get a causally interpretable d-test outcome of type 1-0: E is instantiated in the situation

[10] We purposefully choose a causal structure whose effect can be brought about on two independent causal paths, because in deterministic structures of this complexity causes and effects can be experimentally distinguished without recourse to external asymmetries as the direction of time (cf. Baumgartner 2008).

in which A_1 or A_2 are manipulated and not instantiated in the other test situation. According to (DM), this outcome induces an inference to the causal relevance of A_1 and A_2 for E.

Provided that electric circuits are deterministic structures, the method of difference hence yields that A_1 and A_2 are each part of a deterministic cause of E. Subject to the principles of deterministic causation, it thus follows that A_1 and A_2 are contained in a minimal theory of E. In light of the dependencies among the elements of the circuit specified above, this minimal theory is easily stated:

$$A_1BF \vee A_2CF \vee KH \Leftrightarrow E\,. \tag{4}$$

All of the three disjuncts on the left-hand side of (4) are minimally sufficient for E, the instances of their components are compossible, and the disjunction as a whole is minimally necessary. The lamp is turned on if and only if at least one of the compounds A_1BF or A_2CF or KH is instantiated. That is, relative to the level of analysis adopted in $\mathcal{F}_1$, the electric circuit of figure 1 can be modeled as a deterministic causal structure which can be straightforwardly uncovered by the method of difference—at least within our ideal discovery context.

Matters are different if we choose to analyze the circuit relative to the more coarse-grained frame

$(\mathcal{F}_2)$	A : switch 1 closed	C : switch 3 closed	K : battery b2 charged
	B : switch 2 closed	F : battery b1 charged	H : switch 4 closed
		E : light on	

$\mathcal{F}_2$ differs from $\mathcal{F}_1$ only insofar as the behavior of switch 1 is represented by one factor A in $\mathcal{F}_2$, whereas in $\mathcal{F}_1$ A is decomposed into two factors A_1 and A_2. A does not represent a gruelike or gerrymandered property but a reasonably natural one, which is salient relative to our context of causal discovery. Even though the instances of A are divisible into different subsets (switch 1 closed upwards/downwards/before midnight/after midnight/forcefully/slowly. . .) the elements of all of those subsets resemble each other in relevant respects: they are all closures of switch 1. Moreover, we have already shown that decomposing A into A_1 and A_2 yields two parts of deterministic causes of E. That is, given that $\mathcal{F}_1$ is a suitable factor frame for our purposes, so is $\mathcal{F}_2$. Let us thus investigate whether the method of difference also yields that A is causally relevant to E. To answer this, homogeneous test situations for E with respect to $\mathcal{F}_2$ are required. The same setup of the circuit which satisfies (CH) for E with respect to $\mathcal{F}_1$ also satisfies (CH) for E with respect to $\mathcal{F}_2$: switches 2 and 3 are closed, switch 4 is open, and both batteries are fully charged. Instantiating A in one situation of this type while suppressing A in another such situation yields a d-test outcome that is causally interpretable: E co-varies with the manipulation of A. Hence, there exists an intervention that wiggles the investigated cause variable such that the investigated effect variable wiggles along. As to (DM), this 1-0-outcome induces an inference to the causal relevance of A to E.

Given our sample circuit is deterministically structured, this finding entails that not only A_1 and A_2 are parts of a deterministic cause of E but also A. This, in turn, implies

that there exists a minimal theory of E containing A. Let us try to state that theory. As a first attempt one might simply substitute A_1 and A_2 in (4) by A:

$$ABF \vee ACF \vee KH \Leftrightarrow E\,. \tag{5}$$

It can easily be seen, however, that (5) is not a minimal theory, because neither ABF nor ACF are sufficient for E. For instance, in a constellation in which switches 3 and 4 are open, switch 2 is closed, the batteries are charged, and switch 1 is closed downwards, ABF is instantiated, yet the lamp does not burn, i.e. E is not instantiated. Analogously, closing switch 3, opening switches 2 and 4, and closing switch 1 upwards yields a constellation in which ACF is instantiated along with $\overline{E}$. ABF and ACF, hence, are not deterministic causes of E. (5) is not an adequate reproduction of the causal structure regulating the behavior of E, for it violates the principle of determinism (D).

Closing switch 1 and only requiring one of the switches 2 and 3 to be closed as well, does not determine the lamp to burn. This suggests that A might be part of a deterministic cause of E which comprises both B and C. Thus, another candidate minimal theory of E containing A would be (6):

$$ABCF \vee KH \Leftrightarrow E\,. \tag{6}$$

The compound $ABCF$ indeed is sufficient and even minimally sufficient for E, because—as we have seen above—neither ABF nor ACF are sufficient for E and without a fully charged battery b1 (F) the lamp obviously cannot burn. Nonetheless, (6) is not a minimal theory of E either, for there are scenarios in which neither $ABCF$ nor KH are instantiated even though the light is on. Hence, $ABCF \vee KH$ is not necessary for E, which shows that (6) violates the principle of causality (C). To illustrate, suppose switch 4 is open, switches 1 and 2 are closed, switch 3 is open, and battery b1 is charged. If switch 1 happens to be closed upwards in this setup, the lamp burns while neither $ABCF$ nor KH are instantiated, because C and H are not instantiated—call this scenario $\mathcal{S}$. As the lamp does not burn spontaneously in $\mathcal{S}$, there must be a cause of this instance of E in its spatiotemporal neighborhood. Relative to the idealized design of our example, we can presuppose complete knowledge about the causal structure behind the circuit and can thus easily account for the instance of E in $\mathcal{S}$. In $\mathcal{S}$ the light is on because switches 1 and 2 are closed (upwards). Additionally closing switch 3 is not necessary. Nonetheless, as we have seen above, $ABCF$ is minimally sufficient for E, because A can be instantiated by closing switch 1 either upwards or downwards. In the first case, an instance of B is required to turn the light on, in the second case there must be an instance of C.

Contrary to $\mathcal{F}_1$, the analytic inventory provided by the frame $\mathcal{F}_2$ is not fine-grained enough to adequately model the cause that is responsible for E in $\mathcal{S}$. $\mathcal{F}_2$ does not allow for complementing (6) by missing alternative causes of E in accordance with the principles of deterministic causation. On the level of specification given by $\mathcal{F}_2$ there does not exist a minimal theory which represents the deterministic causal structure behind the circuit of figure 1. More specifically, there does not exist a minimal theory of E with

respect to the frame $\mathcal{F}_2$ comprising A. Nonetheless, as we have shown above, MoD identifies A as being part of a deterministic cause of E.

The principles of deterministic causation do not require that deterministic structures can be modeled relative to one *particular* factor frame. The fact that A is determined to be causally relevant to E on the basis of MoD merely implies that there exists at least one frame—containing A, of course—which allows for stating a minimal theory of E containing A (cf. MT above). The pertaining minimal theory, however, must not be stated on the basis of the conceptual inventory provided by $\mathcal{F}_2$. Rather, the latter can be expanded by introducing other factors that enable a more fine-grained description of our exemplary circuit. The question thus arises as to how to expand $\mathcal{F}_2$ such that A can be shown to be part of a deterministic structure causing E which—unlike (5) and (6)—accords with the principles of deterministic causation.

The reason why we have not yet succeeded in reproducing the structure behind the circuit in figure 1 in a way that complies with (D), (C), and (NR) and that features A as a part of a cause of E is at hand: closing switch 1 can cause the lamp to burn along two different causal paths and $\mathcal{F}_2$ does not allow for specifying the path which is activated by a particular instance of A. The strategy to remedy this deficiency suggests itself. We need to introduce variables that specify whether switch 1 is flipped upwards or downwards. Hence, let us introduce the following two factors:

$$D_1 : \text{switch 1 flipped upwards} \qquad D_2 : \text{switch 1 flipped downwards}$$

Introducing D_1 and D_2 into $\mathcal{F}_2$ yields frame $\mathcal{F}_3$: $\mathcal{F}_3 = \mathcal{F}_2 \cup \{D_1, D_2\}$. $\mathcal{F}_3$ enables us to further specify the complex cause A could be part of by conjunctively adding D_1 and D_2, respectively, to pertaining compounds. Plainly, adding either D_1 or D_2 to the compound $ABCF$ contained in (6) will not yield a minimally sufficient condition of E. Our electric circuit is structured in such a way that, if switch 1 is closed upwards, the position of switch 3 is rendered irrelevant, and analogously if switch 1 is closed downwards, switch 2 is of no relevance any longer. Hence, complementing (6) by D_1 and D_2 would yield a model of the circuit that inevitably features redundancies and, thus, violates (NR). In contrast, introducing D_1 and D_2 into (5), on the face of it, seems to yield just the specification of our model that accords with (D), (C), and (NR):

$$AD_1BF \vee AD_2CF \vee KH \Leftrightarrow E\,. \tag{7}$$

Does (7) indeed amount to a minimal theory of E? Clearly, an instance of E occurs if and only if either AD_1BF or AD_2CF or KH are instantiated. Thus, (7) features both sufficient and necessary conditions of E, i.e. it accords with (D) and (C). Yet, are these conditions free of redundancies, i.e. does (7) also accord with (NR)? That the answer to that question must be in the negative, as both AD_1BF and AD_2CF involve redundancies, can be seen by the following reasoning. In virtue of the structuring of the electric circuit it holds that whenever switch 1 is flipped upwards or downwards, it is closed. That means the set of instances of D_1 and of D_2 are proper subsets of the set of instances of A, i.e. $D_1 \rightarrow A$ and $D_2 \rightarrow A$. In consequence, both AD_1BF and AD_2CF contain proper parts that are sufficient for E, *viz.* D_1BF and D_2CF. Flipping switch 1 upwards (downwards) and closing switch 2 (switch 3) while battery b1

is charged determines the light to be on. Additionally requiring switch 1 to be closed is redundant. That is, introducing D_1 and D_2 into (5) does not result in a minimal theory of E containing A, but renders A redundant and, thus, violates (NR). Contrary to first appearances, (7) is not a minimal theory of E either. In sum, while $\mathcal{F}_2$ is not fine-grained enough to model the circuit in accordance with (C), $\mathcal{F}_3$ is too fine-grained to model it in such a way that A has a non-redundant causal function in accordance with (NR). Thus, a factor frame relative to which A can be said to be causally relevant to E in accordance with all principles of deterministic causation must be somewhat more specific than $\mathcal{F}_2$ and somewhat less specific than $\mathcal{F}_3$.

Such as not to render A redundant, additional factors introduced into $\mathcal{F}_2$ must be less specific than D_1 and D_2. Candidates are not hard to come by:

D_3 : something flipped upwards $\qquad$ D_4 : something flipped downwards

Introducing D_3 and D_4 into $\mathcal{F}_2$ results in the frame $\mathcal{F}_4$: $\mathcal{F}_4 = \mathcal{F}_2 \cup \{D_3, D_4\}$. In contrast to D_1 and D_2, D_3 and D_4 can be instantiated by other things than switch 1. Relative to the design of our exemplary circuit, D_3 can also be instantiated by switch 2 and D_4 by switches 3 or 4. This guarantees that the sets of instances of D_3 and D_4 are not proper subsets of the instances of A, which, in turn, guarantees that A is not rendered redundant by admitting D_3 and D_4. These considerations furnish a further candidate model of the structure regulating the behavior of E:

$$AD_3BF \vee AD_4CF \vee KH \Leftrightarrow E\,. \tag{8}$$

Is (8) a minimal theory of E? Again, that is not the case. Analogously to (4), (8) does not accord with (D), for neither AD_3BF nor AD_4CF are sufficient for E. To see this, consider a scenario in which switch 2 is closed (upwards), switch 1 is closed downwards, switches 3 and 4 are open, and the batteries are fully charged. In such a scenario, the compound AD_3BF is instantiated, yet the lamp does not burn. Furthermore, if switch 3 is closed (downwards), switch 1 is closed upwards, switches 2 and 4 are open, and the batteries are fully charged, AD_4CF is instantiated, yet no instance of E occurs. Hence, neither AD_3BF nor AD_4CF are deterministic causes of E. The electric circuit of figure 1 is structured in such a way that it is of crucial importance that switch 1, and not something else, is flipped upwards when switch 2 is closed and the battery is charged. Similarly, it is switch 1 which must be switched downwards, and not something else, in cases when switch 3 is closed and the battery charged. However, frame $\mathcal{F}_4$—just as $\mathcal{F}_2$—is too coarse-grained to allow for an adequate reproduction of these dependencies.

We still have not found a factor frame containing A relative to which A could indeed be said to be part of a deterministic cause of E. In order to state a minimal theory of E comprising A we need a frame which is somewhat more specific than $\mathcal{F}_4$ and somewhat less specific than $\mathcal{F}_3$. We are looking for additional factors that can only be instantiated by switch 1, yet whose instances are not completely contained in the set of instances of A. As a final attempt, let us investigate whether a disjunctive coarse-graining of D_1 and D_2 might do the job:

D_5 : switch 1 flipped upwards or switch 1 left open $\qquad$ D_6 : switch 1 flipped downwards or switch 1 left open

The frame that results from introducing D_5 and D_6 into $\mathcal{F}_2$ will be referred to as $\mathcal{F}_5$: $\mathcal{F}_5 = \mathcal{F}_2 \cup \{D_5, D_6\}$. As not all instances of D_5 and D_6 are also instances of A, introducing these factors into (5) does not render A redundant:

$$AD_5BF \vee AD_6CF \vee KH \Leftrightarrow E\,. \tag{9}$$

It may justifiably be doubted that D_5 and D_6 meet the suitability standards introduced at the end of section 3, for their instances fall into two subsets whose elements seem to lack resemblance. By contrast, D_5 is equivalent to $\overline{D_2}$ and D_6 to $\overline{D_1}$. Since we took both D_1 and D_2 to be suitable for causal modeling, let us, for the sake of the argument, also accept the suitability of D_5 and D_6. Hence, does (9) not only assign a non-redundant function to A, but moreover satisfy the other constraints imposed on minimal theories? As can easily be seen from the definitions of D_5 and D_6, that again is not the case. Both D_5 and D_6 have proper subsets of instances that, for logical reasons, cannot be co-instantiated with A. Whenever switch 1 is open, both D_5 and D_6 are instantiated, yet these instances of D_5 and D_6 are not compossible with A. For mere logical reasons, all of these instances of D_5 and D_6 cannot ever be causally effective in turning the lamp on, i.e. they are redundant. One disjunct in the definiens of D_5 and D_6, *viz.* leaving switch 1 open, is not only irrelevant for E, but moreover causally relevant for $\overline{E}$. That is, (9) violates (NR). It is not a minimal theory of E.

All of our attempts at specifying the initial frame $\mathcal{F}_2$ in order to find a minimal theory of E containing A have missed the mark. While (7) and (9) introduce redundancies, (8) does not satisfy the principle of determinism. Of course, this does not conclusively prove that there does not exist a minimal theory of E containing A. Negative existentials that are not formal truths cannot normally be conclusively proven. Nonetheless, we presume to have exhausted the realm of possible expansions of $\mathcal{F}_2$. $\mathcal{F}_3$ is too fine-grained, as it renders A redundant. We have tried to coarse-grain $\mathcal{F}_3$ both by means of existential ($\mathcal{F}_4$) and disjunctive ($\mathcal{F}_5$) generalization, none of which has been successful. There does not exist a minimal theory that would feature A as part of a deterministic cause of E. From this it follows that A cannot be said to be part of a deterministic cause of E in accordance with all the principles of deterministic causation, notwithstanding the fact that MoD identifies the closure of switch 1 as part of a deterministic cause of the light being on. The claim that MoD is a correct method of uncovering deterministic structures, the claim that our sample circuit is of deterministic nature, and the claim that deterministic structures are regulated by the principles of determinism, causality, and non-redundancy are not compatible.

5. *Resolving the Conflict*

What are we to conclude from this contradictory finding? A possible conclusion would be that the causal structure regulating the behavior of our circuit is not of deterministic nature after all. In consequence, (D), (C), and (NR) do not apply to the circuit, which, in turn, prevents the conflict between MoD and the principles of deterministic causation from arising in the first place. However, as our findings in section 4 in no way hinge on the specific (electrodynamic) nature of our example, successfully resolving the conflict along

these lines would also require that it be shown that *no other* deterministic structure with the same causal 'switching' pattern can be substituted for our circuit. Every structure with the same form as our circuit would have to be claimed to be of irreducibly indeterministic nature. Drawing such a far-reaching consequence is clearly uncalled for on the mere basis of an a priori philosophical argument as the one presented in the previous section.

Alternatively, it could be held that deterministic structures, contrary to first appearances, do not satisfy all of the principles of deterministic causation put forward in section 2. As a consequence, one would have to postulate that there are deterministic causes that do not determine their effects, or effects of deterministic structures that occur without any of their causes, or causal structures that contain elements that cannot possibly make a difference to the effects contained in pertaining structures. Any of these consequences, in our view, would amount to a straight-out contradiction in terms. Rejecting any of the principles of deterministic causation and still speak of deterministic causal structures is not a viable option.

It might also be argued that the conflict stems from the fact that A, after all, violates the suitability caveat introduced at the end of section 3. Even though "switch 1 closed" does neither represent a gruelike nor a gerrymandered property, A could be rejected as a variable that is suitable for modeling the causes of E, because A is an *ambiguous* manipulation variable in the sense of Spirtes and Scheines (2004), and ambiguous manipulation variables might be declared unsuitable for causal modeling. Spirtes and Scheines introduce their notion of an *ambiguous manipulation* with a hypothetical example (834):

> (. . .) there are two sorts of cholesterol: LDL cholesterol causes heart disease, and HDL cholesterol prevents heart disease. Low-cholesterol diets differ, in the proportions of the two kinds of cholesterol. Consequently, experiments with low-cholesterol regimens can differ considerably in their outcomes. In such a case the variable identified as causal—total cholesterol—is actually a deterministic function of two underlying factors, one of which is actually causal, the other preventative. The manipulations (diets) are actually manipulations on the underlying factors, but in different proportions. When specification of the value of a variable, such as total cholesterol, underdetermines the values of underlying causal variables, such as LDL cholesterol and HDL cholesterol, we will say that manipulation of that variable is ambiguous.

Hence, could the conflict between MoD and the principles of deterministic causation be resolved by stipulating that MoD is only applicable to factor frames that do not feature ambiguous manipulation variables—which requirement is violated by $\mathcal{F}_2$? First, it must be noted that coarse-grained variables (e.g. variables representing supervening properties) commonly underdetermine the values of underlying fine-grained variables (representing corresponding supervenience bases). That is, manipulations of coarse-grained variables generally tend to be ambiguous in the sense of Spirtes and Scheines (2004). Accordingly, restricting the applicability of MoD to ambiguity-free factor frames would restrict that applicability to factor frames without coarse-graining. Such a sweeping restriction is certainly not called for, because as long as the underlying variables of an ambiguous manipulation variable M *do not have opposite* effects, the overall causal influence of M on an investigated effect variable is still determinate—and should hence be uncoverable by a proper procedure of causal inference (cf. Spirtes and Scheines 2004, 844). By contrast, if underlying variables have opposite effects, as in the cholesterol example, the overall

effect of M is inevitably indeterminate. However, restricting the application of MoD to factor frames that do not feature ambiguous manipulation variables of this problematic type does not resolve the conflict of the previous section, for A_1 and A_2—the two fine-grained constituents of A—do not have opposite effects on E. A is not an ambiguous manipulation variable of the problematic type.

Moreover, restricting the applicability of methods of causal discovery to factor frames that do not feature ambiguous manipulation variables with opposite effects would amount to determining the suitability of variables for causal modeling based on *causal conditions*. Yet, whether or not these conditions are satisfied may not be assessable in contexts of causal discovery. In the example of the previous section, we applied MoD to determine whether A_1, A_2, and A are positively or negatively relevant to E. Hence, knowledge of the causal dependencies among A, A_1, A_2, and E cannot be presupposed when assessing the suitability of corresponding factor frames. The suitability of factor frames for causal modeling must be determined in non-causal terms. A method of causal reasoning must not be induced to draw incorrect causal inferences in cases of ambiguous manipulation variables, rather, it must be able to detect whether analyzed factor frames feature ambiguities of the problematic kind or not.

Section 4 has shown that MoD does not meet that benchmark. MoD erroneously identifies A as part of a deterministic cause of E. Thus, the only remaining consequence to draw from the conflict between MoD and the principles of deterministic causation is that available variants of the method of difference can give rise to incorrect causal inferences. Both closing switch 1 upwards and closing it downwards are (positively) causally relevant for the light to be on, but the closure of switch 1 simpliciter is not—even though the latter is nothing but the union of the former. There are two independent causal routes from switch 1 to the lamp. Different variables that are independent of the closure of switch 1 are involved in these routes. Causally relevant factors in deterministic structures, however, are not connected to their effects via multiple routes that are influenced by factors that are not controlled by (i.e. that are not effects of) those causally relevant factors. In contrast, if A were connected to E on routes that do not differ in relevant respects, A could easily be identified as part of a deterministic cause of E. For instance, if it were not possible to interrupt the upper and lower connections between switch 1 and the lamp by virtue of switches 2 and 3, a minimal theory of E containing A could easily be stated: $AF \vee KH \Leftrightarrow E$. In the circuit of figure 1, however, switches 2 and 3 override the causal relevance of the closure of switch 1 simpliciter to the light being on. Due to switches 2 and 3 there does not exist a deterministic cause of E for which A would play a non-redundant role.

That means a single intervention on a suitable cause variable A, even in ideally homogeneous circumstances, that is followed by a change in the value of an investigated effect variable E is not sufficient to establish the causal relevance of A to E. Accordingly, the inference rule (DM) implemented in traditional formulations of the method of difference is not correct. A single 1-0-outcome does not even in perfectly homogenous d-test setups entail causal relevance. This is the proper consequence to draw from the conflict between MoD and the principles of deterministic causation.

This finding, of course, raises the follow-up question as to how the method of difference is to be amended such that all of its inferences are compatible with the principles

of deterministic causation. If one intervention on a proper d-test setup generating a 1-0-outcome is not enough to unfold causal relevancies, what else is required? In order to answer that question, let us reconsider our electric circuit. If both switches 2 and 3 are open or if switch 4 is closed, all interventions on switch 1 yield outcomes of type 1-1 or 0-0, which are not causally interpretable. There are three setups of the circuit that provide homogeneous test situations for E relative to which an intervention on A can generate causally interpretable outcomes:

Setup δ_1: Switch 2 is closed, switches 3 and 4 are open, battery b1 is charged.
Setup δ_2: Switch 3 is closed, switches 2 and 4 are open, battery b1 is charged.
Setup δ_3: Switches 2 and 3 are closed, switch 4 is open, battery b1 is charged.

Setups δ_1 and δ_2 are of particular interest for our purposes. For instance, if A is manipulated by closing switch 1 in a situation of type δ_1, the lamp only burns if A happens to be instantiated by closing switch 1 upwards. If the manner of intervening on A, i.e. of closing switch 1, is varied in another test situation of type δ_1 such that switch 1 is now closed downwards, the lamp does not burn, in spite of no other variable having changed its value. That is, in situations of type δ_1, the closure of switch 1 is sometimes followed by the light being on and sometimes not—and analogously for situations of type δ_2. In other words, upon equal instantiations of potential cause variables, the investigated effect sometimes occurs and sometimes it does not. Clearly, inferring that A is a cause of E based on such test results would induce a violation of the principle of determinism, which, as indicated above, we do not want for processes of the type under consideration. That A is not part of a deterministic cause of E, however, is not revealed if causal inferences are based on singular interventions on A in d-test setups in which switch 1 is closed upwards. Only *systematically varying the manner of manipulating* A in test situations of type δ_1 and δ_2 exhibits that A cannot in fact be interpreted as part of a deterministic cause of E. Varying the manner of manipulating A amounts to varying the causes of A used as interventions on A with respect to E. Merely using one particular cause of A as intervention on A does not induce reliable causal inferences, even in ideally homogeneous laboratory contexts. Reliable causal inferences with respect to deterministic structures are only to be had, if the manner of intervening on investigated cause variables is systematically varied and the outcome of such test iterations *remains stable* across these variations.[11]

Further qualifications are required, though. Consider a situation in which our electric circuit is set in δ_3. All possible variations of intervening on A in such a situation will be accompanied by a change in the value of E. Switch 1 can either be closed upwards or downwards. If switches 2 and 3 are closed, closing switch 1 in either way generates stable 1-0-outcomes, for the lamp burns in both cases. That is, stability of test outcomes across variations of intervening on A must not only be attained relative to one particular d-test setup but relative to all setups that can generate causally interpretable outcomes, i.e. relative to all of δ_1, δ_2, and δ_3. More generally put, the inference rule for d-tests

[11] As Woodward (2003, ch. 6) shows, interventions must not be varied arbitrarily, but only within some range of variability that is suitable for a pertaining process. We tacitly assume here that interventions are only varied within such a range; for instance, switch 1 must not be shattered with a hammer.

implemented in the method of difference (DM) must be amended along the following lines:

Stable difference-making (SDM): A factor A is causally relevant to a factor E if there exists a d-test setup δ such that intervening on A with respect to E in one test situation of type δ generates an 1-0-outcome, and for all d-test setups δ' for which there exists a possible intervention I on A with respect to E generating an 1-0-outcome there does not exist an intervention I' on A with respect to E *not* generating an 1-0-outcome.

It is plain that (SDM) is only conclusively applicable under idealized conditions to the effect that complete control over all relevant factors is on hand. Only then is it possible to assess whether there in fact does not exist an intervention I' on A with respect to E *not* generating an 1-0-outcome. Without ideal isolability of an analyzed process the truth value of such a negative existential cannot be determined in a finite number of steps. In real experimental contexts, (SDM) is only applicable inductively. That is, an experimenter will vary the manner of intervening on a tested factor A to a certain finite degree, which he takes to be representative for the causal structure under investigation. If the tested factor stably makes a difference to the investigated effect across a significant number of variations, the result will be inductively generalized such that (SDM) is applicable and gives rise to a causal inference.

That difference-making should be stable across a significant amount of varied manipulations in order to infer that two factors are causally related is not a new idea. Woodward (2003, ch. 6), for instance, has emphatically stressed the importance of stable or invariant difference-making, especially for deciding among rival causal explanations.[12] The requirement of stable difference-making, however, has commonly been seen as a heuristic means to uncover causal dependencies under non-ideal epistemic conditions where unknown and uncontrolled factors tend to confound test results. Producing stable results across systematically varied interventions within uncontrolled backgrounds significantly raises the probability that pertaining backgrounds are homogeneous, which, in turn, enhances the reliability of corresponding causal inferences. Yet, the standard opinion in the literature, from Mill to Woodward, has been that under homogeneous experimental conditions, i.e. when possible confounders of an investigated deterministic structure are controlled, a single positive d-test result is sufficient for a causal inference.

We take the conflict between MoD-guided causal reasoning and the principles of deterministic causation revealed in section 4 to show that the single-intervention conjecture has been too optimistic. Even under ideal circumstances, single interventions generating a d-test outcome to the effect that a change in a factor A is followed by a change in a factor E can, at best, be seen to entail that A or an element of one of its many decompositions $A_1, A_2, \ldots, A_n$, where $A \leftrightarrow A_1 \vee A_2 \vee \cdots \vee A_n$, is causally relevant to E. Single interventions, however, are under no circumstances sufficient to establish the relevance of A to E. The fact that difference-making must be stable in order for it

[12] Also, Spirtes and Scheines (2004) suggest that in order to reliably estimate the effects of interventions on causal structures, those effects must be stable.

to reliably shed light on causal relationships only partly stems from epistemic or experimental limitations resulting in hampered controllability of causal backgrounds. Varying d-test setups and manipulations of investigated cause variables, first and foremost, serves the purpose of finding the adequate level of analysis, i.e. of determining whether *A* or its decomposition or both are causally relevant. Causal structures cannot adequately be modeled on any arbitrary level of specification. Section 4 has shown that the grain of the analysis is crucial for correct causal inferences, in particular, and successful causal modeling, in general. In order to find the proper level of analysis, systematic variations of test setups and manipulations are essential, independently of how well the investigated structure is known or controlled.

6. Conclusion

The first part of this paper has shown that applying traditional versions of the method of difference to deterministic causal structures—as simple electric circuits—may yield causal inferences that contradict fundamental principles of deterministic causation. The second part has located the source of this conflict in an inference rule that has, more or less explicitly, been implemented in all available formulations of the method of difference: single d-tests generating a 1-0-outcome are sufficient to reveal causal relevancies, provided that pertaining causal backgrounds are homogeneous. We have argued that even complete control over the factors involved in an investigated causal structure does not pave the way for a straightforward inference rule which would uncover deterministic structures based on a handful of successful experimental manipulations. One of the primary tasks that must be fulfilled on the way to an adequate causal model is to find a proper level of analysis. Not any level is suited to model a causal process in terms of a deterministic structure. Stability of test results across systematic variations of experimental manipulations not only increases the probability of homogeneous causal backgrounds in contexts of limited control, but is also required for identifying adequate levels of analysis in contexts of perfect control.

Apart from refining the inference rule connecting difference-making to causal dependencies, this paper has shown that reliably uncovering deterministic causal structures is considerably more intricate than it is often taken to be in the literature—laboratory circumstances notwithstanding. It is time that the causal analysis of deterministic data receives an amount of attention by the interested community that matches the gravity of the problems that come with it.

REFERENCES

Baumgartner, M. 2008. Regularity theories reassessed. *Philosophia* 36: 327-354.

Baumgartner, M. 2009. Uncovering deterministic causal structures: A Boolean approach. *Synthese* 170: 71-96.

Broad, C. D. 1930. The principles of demonstrative induction I-II. *Mind* 39: 302-317, 426-439.

Fodor, J. 1997. Special sciences: Still autonomous after all these years. *Noûs* 31: 149-163.

Glymour, C. 2007. Learning the structure of deterministic systems. In *Causal learning. Psychology, philosophy, and computation*, eds. A. Gopnick and L. Schulz, 231-240. New York: Oxford University Press.

Glynn, L. (forthcoming). A probabilistic analysis of causation. *British Journal for the Philosophy of Science*.

Graßhoff, G. and M. May 2001. Causal regularities. In *Current issues in causation*, eds. W. Spohn, M. Ledwig, and M. Esfeld, 85-114. Paderborn: Mentis.

Halpern, J. Y. and C. Hitchcock 2010. Actual causation and the art of modelling. In *Heuristics, probability, and causality*, 383-406. London: College Publications.

Lewis, D. 1999. New work for a theory of universals. In *Papers in metaphysics and epistemology*, 8-55. Cambridge: Cambridge University Press.

Luo, W. 2006. Learning Bayesian networks in semi-deterministic systems. In *Advances in Artificial Intelligence*, eds. L. Lamontagne and M. Marchand. Volume 4013 of *Lecture Notes in Computer Science*, 230-241. Berlin: Springer.

Mackie, J. L. 1974. *The cement of the universe. A study of causation.* Oxford: Clarendon Press.

May, M. 1999. *Kausales Schliessen. Eine Untersuchung über kausale Erklärungen und Theorienbildung.* Ph. D. thesis, Universität Hamburg, Hamburg.

Mill, J. S. 1843. *A system of logic.* London: John W. Parker.

Pearl, J. 2000. *Causality. Models, reasoning, and inference.* Cambridge: Cambridge University Press.

Ragin, C. C. 1987. *The comparative method.* Berkeley: University of California Press.

Ragin, C. C. 2000. *Fuzzy-set social science.* Chicago: University of Chicago Press.

Ragin, C. C. 2008. *Redesigning social inquiry: Fuzzy sets and beyond.* Chicago: University of Chicago Press.

Ramsey, J., J. Zhang, and P. Spirtes 2006. Adjacency-faithfulness and conservative causal inference. In *Proceedings of the Twenty-Second Annual Conference on Uncertainty in Artificial Intelligence (UAI-06)*, 401-408. Arlington, Virginia: AUAI Press.

Spirtes, P., C. Glymour, and R. Scheines 200). *Causation, prediction, and search* (2 ed.). Cambridge: MIT Press.

Spirtes, P. and R. Scheines 2004. Causal inference of ambiguous manipulations. *Philosophy of Science* 71: 833-845.

Weber, M. 2005. Indeterminism in neurobiology. *Philosophy of Science* 72: 663-674.

Woodward, J. 2003. *Making things happen.* Oxford: Oxford University Press.

Urs HOFMANN studied physics at the ETH Zürich and philosophy of science at the University of Bern and the University of California at San Diego. He was working on issues in the philosophy of physics as well as on causation. Urs sadly passed away before the final publication of this article, which is dedicated to his memory.

Michael BAUMGARTNER is a lecturer at the Department of Philosophy of the University of Konstanz. His publications include an introduction to the philosophy of causation as well as various papers on causation, causal reasoning, regularity theories, interventionism, non-reductive physicalism, logical formalization, and the slingshot argument.

ADDRESS: Dept. of Philosophy, University of Konstanz, Universitätsstrasse 10, 78464 Konstanz, Germany. Email: michael.baumgartner@uni-konstanz.de