Determinism and the Method of Difference

Urs Hofmann, Michael Baumgartner*

Abstract

The first part of this paper reveals a conflict between the core principles of deterministic causation, on the one hand, and the standard method of difference which is widely seen (and used) as a correct method of causally analyzing deterministic data. 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 this inference rule in a way that renders the method of difference compatible with the principles of deterministic causation.

1 Introduction

In contrast to the popularity of probabilistic—especially Bayesian—methods of causal discovery, 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, because the former violate the so-called faithfulness assumption which figures prominently in probabilistic discovery procedures (Spirtes et al. 2000, 53-57, or 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.

Nonetheless, explanations for the sporadic publicity 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 analyzed data. Ordinary causal structures are of such high

*We thank ...

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

complexity and sensitive to such a great many confounding influences that data are seldom homogeneous enough to actually exhibit deterministic dependencies. Only empirical data that are collected against highly controlled causal backgrounds, as for instance given in specific laboratory contexts, de facto feature deterministic relations among investigated factors. Furthermore, in homogeneous laboratory contexts causal reasoning is commonly 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 pertaining 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. In sum, within homogeneous laboratory contexts that allow for systematic manipulations of deterministic structures a simple rule that induces reliable causal inferences is commonly presumed to be available.

This paper intends to show 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"). That is, we are going to present a simple deterministic process such that, when this process is investigated under ideally homogeneous conditions, MOD yields that a factor A is (part of) a deterministic cause of 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 entail a contradiction.

The second part of the paper then locates the source of this conflict in an inference rule that has, more or less explicitly, been implemented in all available formulations of MOD: 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 modifying this rule in a way that renders the method of difference compatible with the principles of deterministic causation.

Yet, independently of where exactly the source of the conflict between traditional formulations of MoD and the principles of deterministic causation is located in the end, this paper aims to show that reliably uncovering deterministic causal structures is considerably more intricate than it is usually taken to be in the literature—ideal 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.

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

As is well known, the notions of determinism and causation have given rise to some of the most intense controversies in the philosophy of science of the past century. Many different and often incompatible analyses of both determinism and causation have been developed in the literature.[2] However, most of these complications are of no relevance for the purposes of this paper. We are neither going to be concerned with the metaphysical question whether the world or particular causal processes are deterministic, nor are we going to inquire about criteria identifying deterministic causal processes or deciding on the deterministic structuring of the world. In consequence, we do neither have to presuppose a full-blown analysis of causation nor an account of what it means for the world, a law of nature, or a scientific theory to be deterministic. Rather, the argument presented here only requires answers to the question as to what features are intuitively ascribed to a causal dependency, *given* that it is judged to be of deterministic nature. In a nutshell, the essential conditions of deterministic causation we need for our purposes are the following conditionals: If $\mathcal{D}$ is a deterministic causal dependency, then

(i) $\mathcal{D}$ satisfies the principle of determinism;
(ii) $\mathcal{D}$ satisfies the principle of causality;

[2] While the literature on causation is too extensive and complex to be informatively cited in a note, there are some serviceable monographs presenting the main theories of determinism: Berofsky (1971), Earman (1986), Sobel (1998, ch. 3).

(iii) $\mathcal{D}$ does not feature redundancies.

To see that all of these necessary conditions for deterministic causation are intuitively or pre-theoretically plausible (to say the least), 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 examples that exclusively involve binary variables, which we call *event types*, or *factors* for short. Analyses of causal processes 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, of events that share at least one feature. Whenever a member of a similarity set that corresponds to an event type occurs, the latter is said to be *instantiated*. Factors are symbolized by italicized capital letters A, B, C, etc., with variables Z, Z_1, Z_2 etc. running over the domain of factors. As absences are often causally interpreted as well, we take factors to be negatable.[3] The negation of a factor A is written thus: $\overline{A}$. $\overline{A}$ simply represents the absence of an instance of A. In short, factors are binary variables that take the value 1 whenever an event of the corresponding type occurs and the value 0 whenever no such event occurs.

Deterministic causes are highly complex and one effect type may be brought about by several alternative causes. Factors do not causally determine their effects in isolation. Rather, they are parts of whole causing *compounds*. A compound only becomes causally effective if all of its constituents are co-instantiated, i.e. instantiated spatiotemporally close-by or *coincidently*. What spatiotemporal interval counts as coincident is notoriously vague, for it depends on the specificity of the causal process under investigation. We are not going to address this question here, but are simply going to assume that the processes discussed in this paper are sufficiently well known that the coincidence relation is properly interpretable.[4] Coincidently instantiated factors are instantiated in the same *situation*. Often, not all factors contained in a deterministic cause are known or of interest to a pertaining causal investigation. Compounds shall be symbolized by simple concatenations of corresponding factors, with variables X, X_1, X_2 etc. standing for open (but finite) sequences of unknown or unmentioned factors, thus for example $ABCX_1$. 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. Finally, by a *deterministic causal structure* we mean a causal structure that only comprises deterministic dependencies.[5]

[3] The controversial question as to what instantiates absences shall be sidestepped in the present context.

[4] For more details on the problem of suitably interpreting the coincidence relation for a given causal process cf. Baumgartner (2008).

[5] 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).

If a compound X is said to be a deterministic (type-level) cause of a factor Z, what is claimed, among other things, is that coincident instantiations of the components of X determine Z to be instantiated. That is, whenever the factors in X are instantiated coincidently there also is an instance of Z. Generalizing this conditional for whole causal structures yields our first principle of deterministic causation:

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, i.e. factors that have no parents in σ, accord with respect to instantiations of *all* factors in σ.

Second, causal structures are taken to 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. Applying this principle to deterministic structures yields:[6]

Causality (C): If a factor Z is an effect within a deterministic causal structure σ, Z is not instantiated without at least one of its alternative (complex) causes in σ being instantiated 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 the 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, we would not want to say that the deterministic cause of F is $SODQ$. Rather, F is only caused by SOD. $SODQ$ has a proper part, *viz.* SOD, which alone determines F. In consequence, Q is redundant, irrelevant to the bringing about of F. Moreover, 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, for, trivially, 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 of $\overline{O}$ cannot be co-instantiated with instances of OD, hence, they are redundant for the bringing about of F. Still worse redundancies can result from analogous factor definitions. Let factor B be defined such that $B \leftrightarrow S \vee Q$, where Q stands for "baptizing an elephant". While all instances of Q are compatible with the instances of OD, BOD does not determine the match to burn, for when an elephant is baptized in the presence of oxygen and of a dry match, BOD

[6] It shall not be claimed that the principles of determinism and causality are logically independent. Often, they are combined to one principle of deterministic causation in the literature. We furnish them with different labels here for the purpose of facilitated reference later on.

is instantiated, yet the match is in no way entailed to light. Nonetheless, BOD can easily be turned into a sufficient condition of F by introduction of a new factor $\overline{Q}$ representing the absence of an elephant's baptism. $B\overline{Q}OD$ determines the match to light, for $\overline{Q}$ ensures that B can only be instantiated by struck matches in combination with $\overline{Q}OD$. Clearly though, baptisms of elephants or their absence do not causally contribute in any way to the lighting of matches. By claiming that, say, SOD is a deterministic cause of F, what is claimed, among other things, is 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 and $B\overline{Q}OD$. 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. $F \rightarrow SOD \vee COD$. Moreover, suppose factor G is 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 one of the disjuncts in $SOD \vee COD \vee GOD$. Yet, analogously to the redundant Q or the redundant instances of A or B, we would not want to count GOD among the alternative causes of F. All instances of F can be accounted for by merely drawing on the disjunction $SOD \vee COD$. GOD being coincidently instantiated entails that both SOD and COD are instantiated. In such cases we would not say that F is caused by a third alternative cause apart from SOD and COD, rather, we would say that the lighting of the match is overdetermined. GOD is not an additional cause of F, i.e. it is redundant.

For mere logical reasons it is excluded that Q, a proper subset of A and B, and GOD are ever indispensable for the bringing about of F. Yet, for each factor and its instances as well as for each compound involved in a deterministic causal structure σ it holds that it 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 contains factors Z_i and compounds X_j 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 and X_j 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 X which is a deterministic cause of a factor Z is a *sufficient* condition of Z, i.e. $X \rightarrow Z$. 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 X is itself sufficient for Z. Hence, $\alpha \rightarrow Z$ must be false for all $\alpha \subset X$. If X satisfies that constraint, X is a *minimally sufficient* condition of Z.[7] Second, no component of X must have a subset of instances that, for logical reasons, cannot be co-instantiated with

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

the other factors in X. All instances of the components of 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 $X_1, X_2, \ldots, X_n$ of an effect Z constitutes a necessary condition of Z, i.e. $Z \rightarrow X_1 \vee X_2 \vee \ldots \vee X_n$. 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 $X_1 \vee X_2 \vee \ldots \vee X_n$, i.e. $X_1 \vee X_2 \vee \ldots \vee X_n$ reduced by at least one disjunct, is itself necessary for Z. That is, $Z \rightarrow \beta$ must be false for all $\beta \subset X_1 \vee X_2 \vee \ldots \vee X_n$. If $X_1 \vee X_2 \vee \ldots \vee X_n$ satisfies that constraint, it is a *minimally necessary* condition of Z.[8] In sum, deterministic causal structures can be represented by a double-conditional of type (1), where (i) each compound $X_1, X_2, \ldots, X_n$ is composed of factors all of whose instances are compossible, (ii) each compound $X_1, X_2, \ldots, X_n$ is a minimally sufficient condition of Z, and (iii) $X_1 \vee X_2 \vee \ldots \vee X_n$ is minimally necessary for Z.

$$X_1 \vee X_2 \vee \ldots \vee X_n \Leftrightarrow Z \tag{1}$$

For brevity, we refer to such double-conditionals that satisfy (NR) as *minimal theories* of Z.[9]

Causal structures can be represented on various levels of specification. To illustrate, reconsider the structure regulating the lighting of matches. It can be described by the (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:

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

The coarse-grained compound SOD and its decomposition $S_1OD \vee S_2OD$ are biconditionally dependent: $SOD \leftrightarrow S_1OD \vee S_2OD$. In light of this logical interdependence, (NR) precludes SOD and $S_1OD \vee S_2OD$ from being part of the same minimal theory of F. For relative to a minimal theory which contains

[8] Cf. Graßhoff and May (2001), Baumgartner (2008).

[9] In order to properly express the relational constraints imposed on instances of causes and effects, as e.g. spatiotemporal proximity, first-order formalisms would be required. Since these complications are of no relevance to the argument of this paper, we use propositional expressions as (1) as convenient abbreviations of the complete logical form of minimal theories. For more details on the first-order form of minimal theories cf. Baumgartner (2008).

SOD, $S_1OD \vee S_2OD$ is redundant, whereas relative to a minimal theory containing $S_1OD \vee S_2OD$, SOD is redundant. In this sense, (NR) guarantees that levels of specification are not mixed. (NR) assigns different minimal theories of the same effect to different levels of specification.

While there may be numerous minimal theories that adequately represent a deterministic structure σ, in light of (D), (C), and (NR) it holds that there exists *at least one* minimal theory for every σ. Or differently, the principles of deterministic causation require that in order for A to be part of a deterministic cause of Z there must exist at least one minimal theory of Z which A is part of. (D), (C), and (NR) hence entail:

Existence of a Minimal Theory (MT): If a factor A is part of a deterministic cause of Z, 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. $AX_1 \vee X_2 \ldots \vee X_n \Leftrightarrow Z$. [*from (D), (C), (NR)*]

Such as to exhibit 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 uncontroversial assumption:

Existence of Deterministic Structures (ED): On macro levels, i.e. on levels above the quantum domain, most—if not all—causal structures are ultimately of deterministic nature. In particular, processes in the domain of classical mechanics, electrodynamics etc. are of deterministic nature.

All in all, independently of the question as to what are both sufficient and necessary conditions for a process to be of causal or deterministic nature and independently of whether all causal processes in the world, i.e. the world as a whole, are deterministic, it is beyond doubt that countless causal processes on macro level in fact are deterministically structured. These processes satisfy the principles of determinism, causality, and non-redundancy. That means for each of these processes there exists at least one minimal theory. If that is not the case for a particular process, it is not a deterministic causal process. Given the uncontroversial 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, 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 cause and effect variables, 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, for example, also result from uncontrolled variations of common causes of A and E in the background of causally analyzed test situations. 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*. In order to clarify what a confounder amounts to, the notion of a causal path is required: A sequence of factors $\langle Z_1, \ldots, Z_k \rangle$, $k \geq 2$, constitutes a *causal path* from Z_1 to Z_k iff for each Z_i and Z_{i+1}, $1 \leq i < k$, in the sequence: Z_i is directly causally relevant to Z_{i+1}. A compound X_j is said to be part of a causal path, if at least one component of X_j is contained in the sequence constituting that path. Now the notion of a confounder needed for the analysis of deterministic structures can be specified: A compound X_j is a confounder of an effect Z_n relative to an analyzed factor frame $\{Z_1, \ldots, Z_n\}$ iff X_j is part of a causal path leading to Z_n not containing any of the factors $Z_1, \ldots, Z_{n-1}$. Less technically put, a confounder is a cause of an investigated effect by means of which the latter can be manipulated independently of the factors in the frame.

In order to infer causal dependencies from covariations, test situations must be compared that are uniform with respect to instantiations of confounders of investigated effects. 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 to Z_{n-1} is changed by intervention, while in $\mathcal{S}_2$ no such interventions are performed. 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 effect under investigation 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 followed 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

A	$\overline{A}$
E	$\overline{E}$

(a)

A	$\overline{A}$
$\overline{E}$	E

(b)

A	$\overline{A}$
E	E

(c)

A	$\overline{A}$
$\overline{E}$	$\overline{E}$

(d)

Tab. 1: Possible outcomes of a d-test investigating the causal relevance of a factor A to an effect E.

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 for an effect E, i.e. part of deterministic cause of E. The investigated factor frame for our exemplary case hence shall be $\{A, E\}$. In order to determine whether A is causally relevant to E based on MOD, we first 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 experimental contexts. In real-life experimental circumstances homogeneity can merely be rendered more or less plausible, for instance, by means of randomization procedures or isolation of experimental setups in laboratory environments. In the end, all inferences drawn on the basis of MOD must be relativized to the assumption that resorted to test situations are homogeneous. As we are only going to be concerned with idealized laboratory contexts in this paper, we do not have to further discuss how to render the homogeneity of test situations maximally plausible. 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 $\{A, E\}$. MOD then 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. This is the design of the simplest application of MOD. For easy reference later on, we refer to this test design as a *difference test*, or a *d-test* for short, and to the homogeneous configuration of background factors not contained in the frame as a *d-test setup*.

A d-test can generate four possible outcomes, which we list in *coincidence tables* as given in table 1. The two columns in coincidence tables represent two homogeneous test situations, i.e. one particular d-test setup. The first row specifies the value of the manipulated potential cause factor A, where 'A' symbolizes an instantiation of A, i.e. $A = 1$, and '$\overline{A}$' stands for the absence of A, i.e. $A = 0$. The second row then indicates whether the investigated effect is instantiated in a pertaining test situation or not, where 'E' stands for $E = 1$ and '$\overline{E}$' for $E = 0$. For example, table (1a) represents a test result such that E occurs when A is instantiated and does not occur when A is not instantiated—and analogously for (1b), (1c), and (1d). For brevity, we shall refer to an outcome of type (1a) as a *1-0-outcome*, to one of type (1b) as *0-1-outcome*, to one of type (1c) as a *1-1-outcome*, and to an outcome of type (1d) as a *0-0-outcome*.

Only two of the four possible d-test outcomes are causally interpretable. Table (1a) induces an inference to the causal relevance of A to E. In this table's first column the effect occurs, thus, at least one cause of E must be instantiated in the corresponding test situation. Necessarily, A is part of one of these causes, for otherwise E would occur in the homogeneous test situation represented in column 2 as well. Based on an analogous reasoning, table (1b) entails the causal relevance of $\overline{A}$ for E. In contrast, tables (1c) and (1d) do not induce any causal inferences. The causal background of the test situations represented by table (1c) apparently features causes of E that do neither contain A nor $\overline{A}$. That is, in the test situations compared in (1c) confounders of E with respect to $\{A, E\}$ are uniformly instantiated which cause E to be present in both situations. In consequence, nothing with respect to a possible causal relevance of A or $\overline{A}$ can be inferred from (1c). Similarly, test situations compared in (1d) do not feature a causally interpretable difference. Nonetheless, it might be thought that (1d) authorizes an inference to the *causal irrelevance* of A or $\overline{A}$ for E. That, however, is not the case. The method of difference does not require all factors possibly constituting a complex cause of E in combination with A or $\overline{A}$ to be instantiated in d-test setup. Hence, a d-test result as the one depicted in table (1d) may occur even if A in fact is part of a complex cause X_1 of E, namely because other components of X_1 are not instantiated in the setup of (1d). In sum, outcomes of d-tests are causally interpretable if and only if there is a difference between compared test situations. Only 1-0- and 0-1-outcomes are causally interpretable.

One methodological principle implemented in the method of difference deserves separate mention at this point: According to MoD, one single intervention is sufficient to establish 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 methodological principle, 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:[10]

> 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 $\mathbf{V}$.

[10] We can confine ourselves to Woodward's account of direct causation here, because his analysis of indirect (or contributing) causation introduces no elements that would override the principle according to which the existence of a singular intervention of type (1a) or (1b) is sufficient for causation. The *variable set* $\mathbf{V}$ Woodward mentions in this passage is what we have been calling the *factor frame* in this paper.

That is, relative to an appropriate d-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. As this methodological principle will turn out to be of crucial importance for the sequel of this paper we furnish it with a label. It shall be referred to as the *single-intervention principle*.

Finally, 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.[11] 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 Z_1 for Z_n, given the causal homogeneity of corresponding test situations and given that Z_n 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 Z_1 is causally relevant to a factor Z_n if there exists at least one d-test setup δ such that intervening on Z_1 with respect to Z_n in one test situation of type δ generates an 1-0-outcome.

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 presume to analyze under perfectly idealized conditions in which we have complete control over all relevant factors. It turns out that d-tests conducted on this 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 electrodynamic processes on macro level are of deterministic nature (ED), on the one hand, and (D), (C), and (NR), on the other, imply a contradiction.

Consider the circuit depicted in figure 1. The burning of the light bulb '$\otimes$' 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 or the right subcircuit is closed.[12] The left circuit is closed

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

[12] 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).

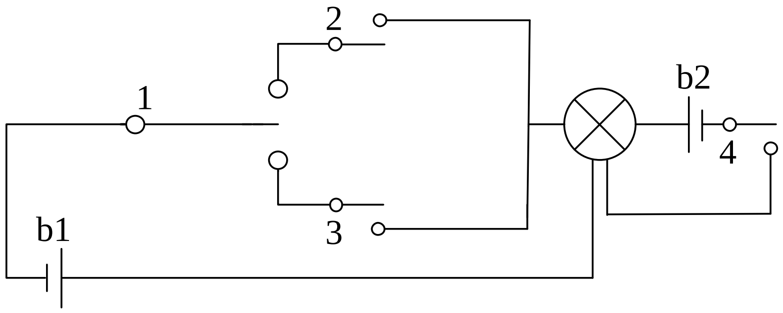

Fig. 1: An electric circuit.

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. When it comes to causally analyzing our exemplary 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:

$(\mathcal{F}_1)$	A_1:	closing switch 1 upwards	F :	battery b1 charged
	A_2:	closing switch 1 downwards	K :	battery b2 charged
	B :	closing switch 2	H :	closing switch 4
	C :	closing switch 3	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 A_1 and A_2 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 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 to E.

Provided that electric circuits are deterministic structures as stipulated by (ED), 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 in the antecedent 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 laboratory context.

Matters are different if we choose to analyze the circuit relative to the more coarse-grained frame $\mathcal{F}_2$:

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

$\mathcal{F}_2$ differs from $\mathcal{F}_1$ only insofar as the behavior of switch 1 is represented by one binary variable A in $\mathcal{F}_2$, whereas in $\mathcal{F}_1$ A is decomposed into two variables A_1 and A_2. Let us investigate whether the method of difference also yields that A is causally relevant to E. To answer this, homogeneous test situations for A with respect to $\mathcal{F}_2$ are required. The same setup of the circuit which satisfies (CH) for A_1 and A_2 with respect to $\mathcal{F}_1$ also satisfies (CH) for A 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, the single-intervention principle is satisfied; 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 that electric circuits are 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 the deterministic causes of E. (5) is not a 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 exemplary circuit, we, of course, 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 electric 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 cause of E relative to $\mathcal{F}_2$. This finding seems paradoxical. A is experimentally identifiable as causally relevant to E relative to the level of specification given in $\mathcal{F}_2$, yet the corresponding causal structure cannot be modeled on that same level of specification.

The finding, however, is paradoxical only in the weak sense of the term, i.e. paradoxical in the sense of ‘puzzling’. It is not contradictory. The principles of deterministic causation do not require that deterministic causal structures are reproducible on a *particular* level of specification. The fact that A is determined to be causally relevant to E on the basis of MOD relative to $\mathcal{F}_2$ merely implies that there exists at least one level of description which allows for stating a minimal theory of E containing A (cf. (MT)). 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 complemented with additional variables that enable a more fine-grained description of our exemplary circuit. The question thus arises as to how to complement $\mathcal{F}_2$ 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 on two causal paths that differ in relevant respects 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: flipping switch 1 upwards D_2: flipping switch 1 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 representation 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 coincidently. 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, but rather 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 reproduce to causal structure behind the circuit in accordance with (C), $\mathcal{F}_3$ is too fine-grained to reproduce that structure in such a way that A has a non-redundant causal function in accordance with (NR). Thus, a level of specification relative to which A can be said to be causally relevant 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, the additional variables introduced into $\mathcal{F}_2$ must be less specific than D_1 and D_2. More explicitly, the additional variables must be defined in such a way that (i) they allow for determining whether the upper or the lower path from A to E is activated, and that (ii) the set of their instances is not a subset of the instances of A. Candidates possibly satisfying (i) and (ii) are not hard to come by:

D_3: flipping something upwards D_4: flipping something downwards

Introducing D_3 and D_4 into $\mathcal{F}_2$ results in 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 an 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. The additional factors D_3 and D_4 meet condition (ii), but not condition (i).

We still have not found the adequate level of specification 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 containing 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: flipping switch 1 upwards or leaving switch 1 open

D_6: flipping switch 1 downwards or leaving switch 1 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:

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

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. Hence, all 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 turning the light on, but moreover causally relevant for the lamp *not* burning, i.e. 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. Plainly, negative existentials that are not formal truths cannot be conclusively proven in principle. Nonetheless, we presume to have exhausted the realm of possible adaptations 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 in combination with (ED) entails that closing switch 1 is 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 electrodynamic processes on macro level are 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 straightforward conclusion would be that the causal structure regulating the behavior of our electric circuit is not of deterministic nature after all, i.e. that (ED) is false. As shown in section 3, MoD only generates correct results if the effect under investigation indeed is an effect of a deterministic structure. Hence, if the circuit in figure 1 is not deterministic, we have not properly applied MoD. That would explain why A was incorrectly ascribed causal relevance for E in the previous section. Plainly, this line of reasoning would entail that there are irreducibly indeterministic causal dependencies way above the quantum domain. Recently, Glynn (2009) has presented an argument in favor of the existence of objectively indeterministic (chancy) processes on macro levels. Glynn's reasons for this claim, however, have nothing to do with the argument advanced in the paper at hand, and he represents a narrow minority position. Tendencies in the literature point in the opposite direction, as non-standard deterministic interpretations of quantum mechanics are continuously gaining popularity. Most of all, we have shown that relative to a proper frame as $\mathcal{F}_1$ there exists a minimal theory of our exemplary electric circuit, that is, the latter can in fact be modeled in terms of a wholly deterministic structure. In consequence, we are not ready to settle for the indeterministic nature of the dependency between closing switch 1 and the light being on on the mere basis of an a priori philosophical argument as the one presented in the previous section. Drawing such a far-reaching consequence, in our view, is not called for without independent (scientific) evidence.

Alternatively, it could be held that deterministic causal 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 dependencies is not a viable option.

The only remaining consequence to draw from the findings of the previous section, hence, is that available variants of the method of difference can give rise to incorrect causal inferences. There does not exist a level of specification (or a frame) relative to which A is part of a deterministic cause of E. Both closing switch 1 upwards and closing it downwards are causally relevant for the light to be on, but closing switch 1 simpliciter is not—even though the latter is nothing but the union of the former. There are two independent causal paths from switch 1 to the lamp. Different variables that are independent of the closing of switch 1 are involved in these paths. Causally relevant factors in deterministic structures, however, are not connected to their effects through multiple paths that are influenced by factors that are not controlled by (i.e. that are not effects of) those relevant factors. In contrast, if A were connected to E on paths 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 closing 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 potential cause variable A, even in ideal homogeneous laboratory 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 single-intervention principle implemented in traditional formulations of the method of difference is false. A single 1-0-outcome does not even in perfect d-test setups entail causal relevance. The inference rule (DM) is not correct. 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 the application of MoD that we erroneously took to induce the inference to the causal relevance of closing switch 1 to the light being on in the previous section. 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 that are not causally interpretable. Overall, there are three setups of our electric circuit that provide homogeneous test situations for A with respect to E relative to which a proper 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, closing 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 identical instantiations of potential cause variables, the investigated effect sometimes occurs and sometimes it does not. Clearly, identifying A as 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 electrodynamic processes of the type under consideration. That A cannot be 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.

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

Stable difference-making (SDM): A factor Z_1 is causally relevant to a factor Z_n if there exists a d-test setup δ such that intervening on Z_1 with respect to Z_n 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 Z_1 with respect to Z_n generating an 1-0-outcome there does not exist an intervention I' on Z_1 with respect to Z_n *not* generating an 1-0-outcome.

It is plain that (SDM) is only conclusively applicable under idealized experimental 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 Z_1 with respect to Z_n *not* generating an 1-0-outcome. Without ideal isolability of an analyzed process the truth value of such a negative existential, of course, 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 Z_1 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) gives rise to a causal inference.

That difference-making should be stable across a significant amount of varied manipulations in order for an investigated relationship between two factors to be of causal nature 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. 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 the previous section 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 $A \leftrightarrow A_1 \vee A_2 \vee \ldots \vee A_n$, 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 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. The previous section 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 a methodological principle that has, more or less explicitly, been implemented in all available formulations of the method of difference: the single-intervention principle according to which single d-tests generating a 1-0-outcome reliably 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 uncovering deterministic causal structures is considerably more intricate than it is often presumed to be. Nonetheless, problems of causal discovery in deterministic contexts have received far less attention in the pertaining literature than their probabilistic counterparts. One upshot of this paper is that this unbalanced focus should be reconsidered.

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.

Berofsky, B. (1971). *Determinism*. Princeton: Princeton University Press.

Broad, C. D. (1930). The principles of demonstrative induction i-ii. *Mind 39*, 302–317, 426–439.

Dowe, P. (2002). What is determinism? In H. Atmanspacher and R. C. Bishop (Eds.), *Between Chance and Choice: Interdisciplinary Perspectives on Determinism*, pp. 309–319. Thorverton: Imprint Academic.

Earman, J. (1986). *A Primer on Determinism*. Dordrecht: D. Reidel.

Glymour, C. (2007). Learning the structure of deterministic systems. In A. Gopnick and L. Schulz (Eds.), *Causal Learning. Psychology, Philosophy, and Computation*, pp. 231–240. New York: Oxford University Press.

Glynn, L. (2009). Deterministic chance. *British Journal for the Philosophy of Science*.

Graßhoff, G. and M. May (2001). Causal regularities. In W. Spohn, M. Ledwig, and M. Esfeld (Eds.), *Current Issues in Causation*, pp. 85–114. Paderborn: Mentis.

Hesslow, G. (1981). Causality and determinism. *Philosophy of Science*, 591–605.

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

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.

Sobel, J. H. (1998). *Puzzles for the Will – Fatalism, Newcomb and Samarra, Determinism and Omnsicience*. Toronto: University of Toronto Press.

Spirtes, P., C. Glymour, and R. Scheines (2000). *Causation, Prediction, and Search* (2 ed.). Cambridge: MIT Press.

Woodward, J. (2003). *Making Things Happen*. Oxford: Oxford University Press.