An Internal Limit of the Structural Analysis of Causation

ABSTRACT

Structural models of systems of causal connections have become a common tool in the analysis of the concept of causation. In the present paper I offer a general argument to show that one of the most powerful definitions of the concept of actual cause, provided within the structural models framework, is not sufficient to grant a full account of our intuitive judgements about actual causation, so that we are still waiting for a comprehensive definition. This is done not simply by focusing on a set of case studies, but by arguing that our intuitions about two different kinds of causal patterns, i.e., overdetermination and counterdetermination, cannot be addressed using that definition.

KEYWORDS: causality, counterfactuals, causal graphs, causal models, structural equations, overdetermination.

Introduction

In the last two decades structural models of causal connections have played a central role in the analysis of the concepts of causation and actual cause, and in Halpern and Pearl (2005) an interesting and highly sophisticated definition of actual cause, based on a structural account, is provided[1]. However, a number of scholars have put into question the adequacy of any definition of this kind, highlighting that further conditions are required in order to account for our intuitive judgements about actual causation[2]. In the present paper, I present a structural reason why this attractive definition cannot account for such judgements and that, as a consequence, we are still waiting for a comprehensive definition of actual causation. The

[1] As far as I know, this is the most complete definition now at our disposal. It constitutes a development of the definition given in Halpern & Pearl (2001), and benefits from the analysis provided by Hitchcock (2001).

[2] See, for instance, Menzies (2004), Woodward (2006), Hall (2007), Hitchcock (2007), and Halpern & Hitchcock (2014).

paper is subdivided into four parts: in the first one, the basic concepts underlying the construction of structural models are introduced; in the second one, two special kinds of causal determination, i.e., over- and counterdetermination, are defined and analysed; in the third one, the difficulties posed by cases of these kinds are discussed; finally, it is shown that it is not possible to give a uniform account of the intuitive judgements about the causal role played by events involved in over- and counterdetermination.

1. The structural analysis of causation

According to the standpoint proposed in Pearl (2000)[3], a causal model is a model of a system of causal connections. The connected events are represented as variables and particular values of a variable X are represented by equations like $X = x$.

1.1. Causal Models

The variables in a causal model are subdivided into *endogenous* and *exogenous* ones: exogenous variables are the ones whose values are taken as given, while the values of endogenous variables are determined by means of a set of structural equations, representing the causal connections modelled in the system.

Definition 1.1: causal model $\langle \mathcal{U},\mathcal{V},\mathcal{R},\mathcal{E} \rangle$.

A causal model $\mathcal{M}$ is a 4-tuple $\langle \mathcal{U},\mathcal{V},\mathcal{R},\mathcal{E} \rangle$ where

1) $\mathcal{U}$ is a set of variables (exogenous variables)
2) $\mathcal{V}$ is a set of variables (endogenous variables)
3) $\mathcal{R}$ is a function assigning a set of values to variables in $\mathcal{U} \cup \mathcal{V}$
4) $\mathcal{E}$ is a function assigning a structural equation to each variable in $\mathcal{V}$

Intuitively, a structural equation $\mathcal{E}(X)$, usually written as $X = f(\boldsymbol{X})$, where $\boldsymbol{X}$ is a sequence of variables in $\mathcal{U} \cup \mathcal{V}$, is a function determining the value of X given the

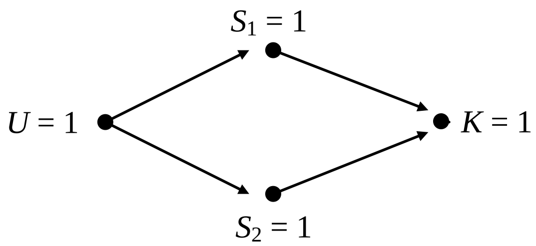
[3] This standpoint, adopted with minor modifications by Halpern & Pearl (2001), Hitchcock (2001), Woodward (2003), Halpern & Pearl (2005), can be now considered the standard view about modelling causation using structural equations.

values of the variables in $\boldsymbol{X}$. It is essential to notice that a structural equation codifies both the kind of dependence on $\boldsymbol{X}$ and the fact that X is the dependent variable, namely that X is the determined variable[4]. In addition, if $X = f(\boldsymbol{X})$ then every variable $X_i \in \boldsymbol{X}$ is an essential argument, in the sense that there is at least one setting of values of the other variables in $\boldsymbol{X}$ and a pair of distinct values x_i and x_i' of X_i, such that X takes distinct values depending on $X_i = x_i$ or x_i' in the given setting. Therefore, if $X = f(\boldsymbol{X})$ and $X_i \in \boldsymbol{X}$, then X_i can be said to be a *potential cause* of X, since there is a context in which the value of X depends only on the value taken by X_i.

Example 1.1: Let's suppose we intend to model the case of a man shot by two killers. Then the causal model would be the 4-tuple $\langle \boldsymbol{\mathcal{U}}, \boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{E}} \rangle$ in which:

- $\boldsymbol{\mathcal{U}}$ includes a variable U necessary to render the connections stated in the structural equations deterministic.
- $\boldsymbol{\mathcal{V}}$ includes the three significant variables: S_1 for the first shooting; S_2 for the second shooting; K for the killing.
- $\boldsymbol{\mathcal{R}}$ is such that $\boldsymbol{\mathcal{R}}(U) = \boldsymbol{\mathcal{R}}(S_1) = \boldsymbol{\mathcal{R}}(S_2) = \boldsymbol{\mathcal{R}}(K) = \{1,0\}$; 1 if the event represented by the variable occurred, 0 otherwise.
- $\boldsymbol{\mathcal{E}}$ is such that $S_1 = U$, $S_2 = U$, $K = \max(S_1,S_2)$, so that the man would die, if either the first or the second killer shoots.

When U takes the value 1, the information conveyed by this model can be partially described using the following graphical representation:



[4] Thus, $X = Y$ and $Y = X$ are different structural equations, given that only the first one states that X is dependent, indeed causally dependent, on Y.

In a graph of this kind nodes represent variables and arrows causal dependencies, while the kind of dependence is not represented.

In what follows only acyclic models are considered. These are models in which $\mathcal{E}$ defines a set of acyclic structural equations, i.e., a set of equations linking variables ordered in $\mathcal{V}$ by a total ordering <, such that if X < Y, then *X* is independent of *Y*. Intuitively, acyclic models are models in which, if a variable *Y* is dependent on a variable *X*, then *X* is independent of *Y* and of any variable directly or indirectly dependent on *Y*. Acyclic models are described using acyclic graphs and are deterministic: once the values of the exogenous variables are specified, the set of equations suffices to determine the value of each endogenous variable.

1.2. Variants of Causal Models

A set of structural equations fixes the background knowledge concerning general causal dependencies between event types. This knowledge is then used to identify actual causal dependencies between event tokens. In order to do so, we need a method to describe counterfactual connections. Such a method is provided by introducing variants of causal models, representing the way in which models change under external interventions[5].

Definition 1.2: *X*-variant of a causal model $\mathcal{M}$: $\mathcal{M}(X = x)$.

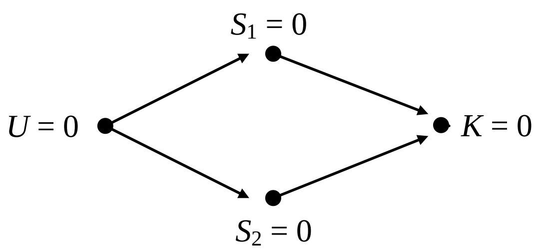
Let $\mathcal{M} = \langle \mathcal{U}, \mathcal{V}, \mathcal{R}, \mathcal{E} \rangle$ be a causal model.
An *X*-variant $\mathcal{M}(X = x)$ of $\mathcal{M}$ a is a 4-tuple $\langle \mathcal{U}', \mathcal{V}', \mathcal{R}', \mathcal{E}' \rangle$ where

1) $\mathcal{U}' = \mathcal{U}$ is the set of exogenous variables
2) $\mathcal{V}' = \mathcal{V}$ is the set of endogenous variables
3) $\mathcal{R}' = \mathcal{R}$ is the function assigning a set of values to variables in $\mathcal{U}' \cup \mathcal{V}'$
4) $\mathcal{E}' = \mathcal{E}$ for variables other than *X*, while $\mathcal{E}'(X)$ is the equation $X = x$

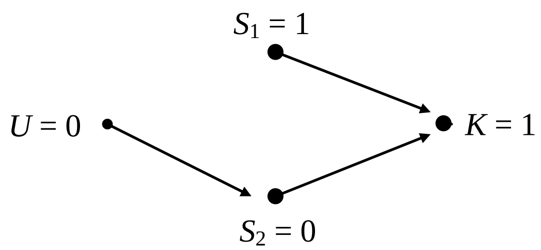
[5] Thus, in a sense, the proposed theory is a counterfactual theory of causation, capturing many aspects of Lewis' first theory. See Lewis (1973a), (1973b), (1979).

The definition of X-variant of $\mathcal{M}$ can be generalized in a straightforward way to sets $\boldsymbol{X}$ of variables. Using variants of a causal model we can construe a counterfactual statement as a statement describing how the world would have been given a change of the value of some variables.

Example 1.2: Let's consider the model introduced before. If the exogenous variable U is set to 0, the graph turns into



We could now want to know what would have happened if the first killer had shot the man. In the present framework, the answer is given by considering the $S_1 = 1$ variant of the model:



In this variant the man would have been killed. Notice that there is no arrow connecting U to S_1, since, as a result of substituting $S_1 = 1$ for the original equation, the dependence of S_1 on U is not working any more.

1.3. Semantics

Following Pearl and Halpern (2005), I will use capital letters both as linguistic and as metalinguistic signs for general variables, and small letters both as linguistic and as metalinguistic signs for values of a variable. A basic formula is one of the form $X = x$, for $X \in \mathcal{V}$ and $x \in \mathcal{R}(X)$. A basic causal formula is a one of the form $[\boldsymbol{X} = \boldsymbol{x}](X = x)$, where $\boldsymbol{X} = \boldsymbol{x}$ states that the variables in the sequence $\boldsymbol{X}$ take the values in the sequence $\boldsymbol{x}$. A formula is valuated at a possible world in a model, where a possible world is specified, in acyclic models, by setting the exogenous variables to some values.

i) $\mathcal{M} \models_{\mathbf{u}} X = x$ stands for
$X = x$ is true at **u** in the model $\mathcal{M}$.

ii) $\mathcal{M} \models_{\mathbf{u}} [\boldsymbol{X} = \boldsymbol{x}](X = x)$ stands for
$X = x$ is true at **u** in the $(\boldsymbol{X} = \boldsymbol{x})$-variant of $\mathcal{M}$.

Being true at **u** means being true when the variables in $\mathcal{U}$ take the values in **u**. Boolean combinations of basic formulas are defined in the usual way.

Definition 1.3: actual cause at the world **u** in a model $\mathcal{M}$.
$X = x$ is an actual cause of φ at **u** in $\mathcal{M}$ iff

AC1: $\mathcal{M} \models_{\mathbf{u}} X = x \wedge \varphi$
AC2: there exists a set $\boldsymbol{X} \subseteq \mathcal{V} \backslash X$ and a setting $\boldsymbol{x}, x'$ of $\boldsymbol{X}, X'$ such that

i) $\mathcal{M} \models_{\mathbf{u}} [\boldsymbol{X} = \boldsymbol{x}, X = x']\neg\varphi$
ii) $\forall\, \boldsymbol{A} \subseteq \boldsymbol{X}, \boldsymbol{B} \subseteq \mathcal{V} \backslash \boldsymbol{X}, \boldsymbol{b}\ (\mathcal{M} \models_{\mathbf{u}} \boldsymbol{A} = \boldsymbol{a} \Rightarrow \mathcal{M} \models_{\mathbf{u}} [\boldsymbol{A} = \boldsymbol{a}, \boldsymbol{B} = \boldsymbol{b}, X = x]\varphi)$ [6]

The idea underlying the definition is simple. AC1 states that both the cause and the effect are actual events. AC2 states that $X = x$ is, in a sense, a necessary and sufficient condition for φ to obtain: necessary, since, according to i), there are circumstances, determined by $\boldsymbol{X} = \boldsymbol{x}$, where if $X = x$ was not actual, φ would not obtain;

[6] The present definition is a slight modification of the original definition 3.1 in Halpern & Pearl (2005). In definition 3 only basic events are allowed as actual causes, but this limitation is only apparent. The original definition includes a further condition stating that the set of events identifying the actual cause has to be minimal; given this condition, it can be proved that such a set can only be a singleton.

sufficient, since, according to ii), changing the actual values of the variables not in $\boldsymbol{X}$ cannot prevent $X = x$ from causing φ.[7]

Remark: In what follows we will consider only models with three endogenous binary variables. In such models, definition 3 can be so restated:

$X_1 = x$ is an actual cause of φ at **u** in $\mathcal{M}$ if and only if

AC1: $X_1 = x_1 \wedge \varphi$

AC2, there exist $x^*_1 \in \mathcal{R}(X_1)$, $x^*_2 \in \mathcal{R}(X_2)$ such that

i) $\mathcal{M} \models_{\mathbf{u}} [X_1 = x^*_1, X_2 = x^*_2]\neg\varphi$

ii) $\mathcal{M} \models_{\mathbf{u}} [X_1 = x_1, X_2 = x^*_2]\varphi \wedge [X_1 = x_1, X_2 = x_2]\varphi$

where $X_2 = x_2$ is the actual value of X_2: $\mathcal{M} \models_{\mathbf{u}} X_2 = x_2$.

2. Basic cases of over- and counterdetermination

In this section the central concepts of over- and counterdetermination are introduced. Cases of causal overdetermination are well-known in literature about causation, since they are used to provide counterexamples to simple counterfactual analyses of causation. Cases of counterdetermination are, in a sense, the negative counterparts of the former ones and can be used to block more sophisticated counterfactual analyses of causation[8].

2.1. Primary and backup overdetermination (pre-emption)

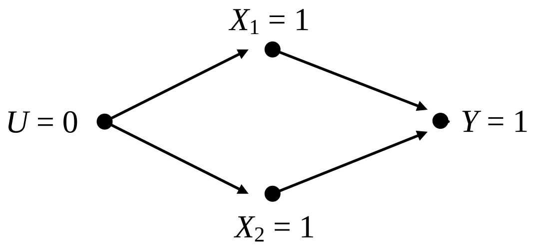
Definition 2.1: (i) an event is overdetermined, by primary overdetermination, when it occurs in a context where it has two potential causes, both of which are actual; (ii) an event is overdetermined, by pre-emption, when it occurs in a context where

[7] The same basic idea constitutes the starting point for the substantially similar definitions proposed in Hitchcock (2001), pp. 286-290, and Woodward (2003), pp. 77-85.

[8] By the way, the classical examples suggesting that causation is not in general transitive are cases of counterdetermination. See, for instance, McDermott (1995) and Hitchcock (2001).

it has two potential causes, the second one being activated if the first one is not actual or does not succeed in becoming actual.

Case 1: the basic model of primary overdetermination.



Equations
$X_1 = U$
$X_2 = U$
$Y = \max(X_1,X_2)$

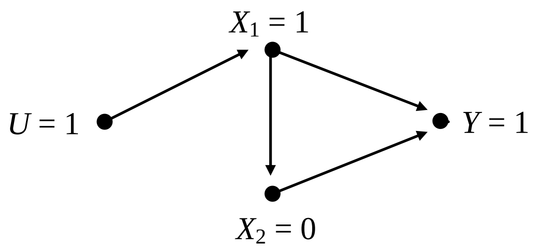
Example 2.1: two killers shoot at a man; each shot is sufficient to cause the man to die.

$X_1 = 1$:= the target is hit by bullet 1
$X_2 = 1$:= the target is hit by bullet 2
$Y = 1$:= the target dies

In accordance with intuitions I take to be standard, definition 1.3 picks out both $X_1 = 1$ and $X_2 = 1$ as actual causes of $Y = 1$. To verify that $X_1 = 1$ is an actual cause of $Y = 1$, put $\boldsymbol{X} = \langle X_2 \rangle$ and $\boldsymbol{x} = \langle 0 \rangle$:

AC2, i) $\mathcal{M} \models_{\mathbf{u}} [X_1 = 0, X_2 = 0](Y = 0)$
AC2, ii) $\mathcal{M} \models_{\mathbf{u}} [X_1 = 1, X_2 = 0](Y = 1) \wedge [X_1 = 1, X_2 = 1](Y = 1)$

Case 2: basic model of backup overdetermination = pre-emption.



Equations:
$X_1 = U$
$X_2 = 1 – X_1$
$Y = \max(X_1,X_2)$

Example 2.2: a killer shoots at a man, his shot being sufficient to cause the man to die; another killer supports the first, being ready to shoot the man, if the first one does not hit him.

$X_1 = 1$:= the target is hit by bullet 1
$X_2 = 0$:= the target is not hit by bullet 2
$Y = 1$:= the target dies

In this case too, definition 1.3 agrees with our intuitions in picking out only $X_1 = 1$ as actual causes of $Y = 1$. Put $\boldsymbol{X} = \langle X_2 \rangle$ and $\boldsymbol{x} = \langle 0 \rangle$:

AC2, i) $\mathcal{M} \models_{\mathbf{u}} [X_1 = 0, X_2 = 0](Y = 0)$
AC2, ii) $\mathcal{M} \models_{\mathbf{u}} [X_1 = 1, X_2 = 0](Y = 1)$

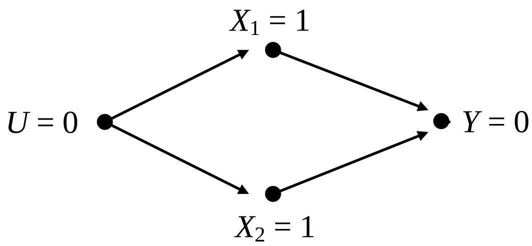
$X_2 = 0$ is not an actual cause, since $\mathcal{M} \models_{\mathbf{u}} [X_1 = 0, X_2 = 0](Y = 0)$.

Focusing on the models just illustrated, we can conclude that the intuitive concept of overdetermination is captured by $Y = \max(X_1,X_2)$ and that the distinction between primary and backup overdetermination is captured by the pair of equations determining the value of X_2.

2.2. Primary and backup counterdetermination (prevention)

Definition 2.2: (i) an event is counterdetermined by primary counterdetermination when it occurs in a context where there are an agent cause and a counteragent cause, both of which are actual causes; (ii) an event is prevented when it occurs in a context where there are an agent cause and a counteragent cause, the second one being activated by the fact that the first one is or becomes actual.

Case 3: the basic model of primary counterdetermination.



Equations

$X_1 = U$

$X_2 = U$

$Y = \max(X_1, 1\text{-}X_2)$

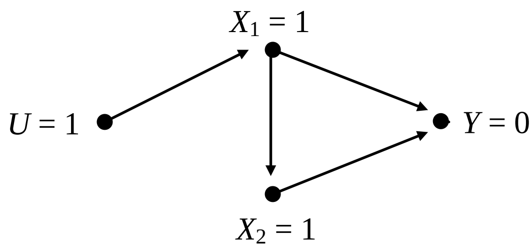
Example 2.3: a killer puts poison into the king's tea; the king's tea is usually checked for poisons; the check detects the presence of the poison; the king is given another cup of tea.

$X_1 = 1$:= the killer poisons the king’s tea

$X_2 = 1$:= the king’s tea is substitute

$Y = 0$:= the king does not die

Case 4: basic model of backup counterdetermination = prevention.



Equations:

$X_1 = U$

$X_2 = X_1$

$Y = \min(X_1, 1\text{-}X_2)$

Example 2.4: a killer puts poison in the king's tea; a bodyguard responds by pouring an antidote in the king's tea; the antidote is not lethal when taken by itself and neutralizes the poison in the king’s tea.

$X_1 = 1$:= the killer poisons the king’s tea

$X_2 = 1$:= the bodyguard pours the antidote

$Y = 0$:= the king does not die

In both cases, the standard causal judgement seems to be that $X_2 = 1$ is the only actual cause of $Y = 0$, and definition 1.3 picks out this very cause. In fact, choose $\boldsymbol{X} = \langle X_1 \rangle$ and $\boldsymbol{x} = \langle 1 \rangle$:

AC2, i) $\mathcal{M} \models_{\mathbf{u}} [X_1 = 1, X_2 = 0](Y = 1)$
AC2, ii) $\mathcal{M} \models_{\mathbf{u}} [X_1 = 1, X_2 = 1](Y = 0)$

$X_1 = 1$ is not an actual cause, since $\mathcal{M} \models_{\mathbf{u}} [X_1 = 1, X_2 = 1](Y = 0)$.

Focusing on these models, we can conclude that the intuitive concept of counterdetermination is captured by $Y = \min(X_1, 1–X_2)$ and that the distinction between primary and backup counterdetermination is captured by the pair of equation determining the value of X_2.

At first sight, definition 1.3 copes well with cases of both over- and counterdetermination. Still, it is possible to present situations in which our intuitive judgements are not in agreement with the outcomes of the definition.

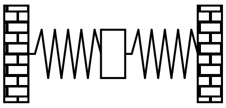
3. Puzzling cases of overdetermination and counterdetermination

Let us begin with a presentation of what I take to be our intuitive causal judgements about cases of over- and counterdetermination, where the initial conditions are possibly changed by varying the value of one or both of the variables representing the potential causes. These judgements can be either just instinctive judgements or the consequences of a critical application of some general principle. The principle I will take into consideration, as the one underlying most of our judgements, is the following

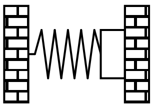
Principle of existential inertia (**PEI**):
i) If no cause is activated, then a variable maintains its actual value.
ii) If a variable maintains its actual value, then a no net cause is activated.

The first part of **PEI** is the ontological counterpart of the physical *Principle of inertia*, while the second one is the ontological counterpart of a physical principle of equilibrium.

In stating the principle of existential inertia it was assumed that a net cause is a potential cause whose effect is not counterbalanced by other potential causes in the circumstances. Thus, if a potential cause is activated, and a dependent variable, say Y, maintains its value, say $Y = 0$, then other potential causes, balancing the effect of the first one, are active or activated, so that no net effect occurs. In cases of balance, causes can act either symmetrically or asymmetrically: they act symmetrically when each of them, in the absence of the other, would cause a change of the value of Y, and act asymmetrically when only one of them, in the absence of the other, would cause that change. As an example, consider the following situations:

a) 

a spring acting on a given object
is counterbalanced by a similar spring.

b) 

a spring acting on a given object
is counterbalanced by a wall.

If the causes act symmetrically, both of them are considered as actual causes of the maintenance of a value: in case a) the object is at rest because two causes act in opposite directions; removing one cause would change the state of rest. If, on the contrary, they act asymmetrically, only the cause that would not change the object's state is considered as an actual cause: in b) the object is at rest because the wall prevents an otherwise possible motion; removing the spring would not change the state of rest.

3.1. Primary overdetermination

Primary overdetermination is given by the presence of two causes acting at once, or almost at once, in the same direction[9]. Having example 2.1 in mind, it seems to be standard to assume that, when the variable X_i takes the value 1, then $X_i = 1$ is an actual cause of $Y = 1$, whereas, if both variables take the value 0, then no change occurs and nothing counts as a cause of $Y = 0$.

PEI leads to the same result. If Y turns to 1, then a cause is activated. The cause can be $X_1 = 1$, $X_2 = 1$, or the conjunction of both. If Y maintains the value 0, then no net cause is activated; but $X_1 = 0$ and $X_2 = 0$ are not counterbalancing causes; therefore neither $X_1 = 0$ nor $X_2 = 0$ is an actual cause of $Y = 0$.

Remark: In the following tables, "I-cause = ☑" indicates that the corresponding actual event is judged as an actual cause according to our intuition, while "D-cause = ☑" indicates that the actual event is judged as an actual cause according to definition 3.

Four possible worlds:

variable	equation	value	I-cause	D-cause

world 1

X_1	X_1 = U	1	☑	☑
X_2	X_2 = U	1	☑	☑
Y	$Y = \max(X_1,X_2)$	1		

world 2

X_1	X_1 = 1–U	0	☐	☐
X_2	X_2 = U	1	☑	☑
Y	$Y = \max(X_1,X_2)$	1		

world 3

[9] Primary overdetermination is also called symmetric overdetermination.

X_1	X_1 = U	1	☑	☑
X_2	X_2 = 1–U	0	☐	☐
Y	$Y = \max(X_1,X_2)$	1		

world 4

X_1	X_1 = 1–U	0	☐	☑
X_2	X_2 = 1–U	0	☐	☑
Y	$Y = \max(X_1,X_2)$	0		

In the last case, the intuitive judgements differ from the ones based on definition 3, since, following the definition, both $X_1 = 0$ and $X_2 = 0$ are actual causes of $Y = 0$. To see why, for example, $X_1 = 0$ is an actual cause of $Y = 0$, choose $\boldsymbol{X} = \langle X_2 \rangle$ and $\boldsymbol{x} = \langle 0 \rangle$. Then:

AC2, i) $\mathcal{M} \models_{\mathbf{u}} [X_1 = 1, X_2 = 0](Y = 1)$
AC2, ii) $\mathcal{M} \models_{\mathbf{u}} [X_1 = 0, X_2 = 0](Y = 0)$

However, it could be said that the situation depicted as world 4 should not be modelled by introducing three variables, since the negative events play no role in that situation. If we accept this suggestion, the problem concerning judgements about causation at world 4 does not arise.

3.2. Backup overdetermination

Backup overdetermination is given by the presence of two causes, acting in the same direction, where the second cause is activated by the first one not being or not becoming actual[10]. Having example 2.2 in mind, it seems to be standard to assume that when a variable X_i takes the value 1, the other being = 0, then $X_i = 1$ is the sole actual cause of $Y = 1$, unless $X_i = 1$ prevents effects that in any case would be prevented by an already existing condition[11].

[10] Notice that early preemption, late preemption, and trumping preemption are all species of backup overdetermination.

[11] A classic example of this exception is given in McDermott (1995): a man catches a cricket ball in a context in which the next things in the ball's direction of motion are a solid brick wall and a win-

PEI leads to the same result. Since Y turns to 1 anyway, a cause is activated, and the cause can be $X_1 = 1$, $X_2 = 1$, or the conjunction of both.

Two possible worlds:

variable	equation	value	I-cause	D-cause

world 1

X_1	$X_1 = U$	1	☑	☑
X_2	$X_2 = 1–X_1$	0	☐	☐
Y	$Y = \max(X_1,X_2)$	1		

world 2

X_1	$X_1 = 1–U$	0	☐	☐
X_2	$X_2 = 1–X_1$	1	☑	☑
Y	$Y = \max(X_1,X_2)$	1		

In this case, judgements based on definition 3 seem to be wholly consistent, or almost wholly consistent, with intuitive judgements.

3.3. Primary counterdetermination

Primary counterdetermination is given by the presence of two causes acting at once in opposite directions, where one of them prevails over the other one[12]. Having example 2.3 in mind, it seems to be standard to assume that, when the variable X_1 takes the value 1, within a context where X_2 takes the value 0, then $X_1 = 1$ is an actual cause of $Y = 1$, while, within a context where X_2 takes the value 1, $X_2 = 1$ is an actual cause of $Y = 0$. Moreover, we are inclined to assume that, when the variable X_1 takes the value 0, there are no actual causes of $Y = 0$, independently of the value taken by X_2: $Y = 0$ because nothing acted for changing this value.

---

dow. In this case only few people would say that the man’s action prevented the ball hitting the window.

[12] See Hiddelston (2005) and Halpern & Hitchcock (2014) for a more extended presentation.

PEI leads to the same result. If Y turns to 1, then a cause is activated and the cause can be $X_1 = 1$ only, because the occurrence of either $X_2 = 1$ or the conjunction $X_1 = 1$ and $X_2 = 1$ is not sufficient to change the value of Y. If Y maintains the value 0, then no net cause is activated. Thus, either no actual cause or at least two counter-balancing causes occur. $X_1 = 1$ and $X_2 = 1$ constitute a pair of counterbalancing causes, and, since their action is asymmetrical, $X_2 = 1$ is the actual cause of $Y = 0$.

Counterdetermination: the four possible worlds.

variable	equation	value	I-cause	D-cause

world 1

X_1	$X_1 = U$	1	☐	☐
X_2	$X_2 = U$	1	☑	☑
Y	$Y = \min(X_1, 1–X_2)$	0		

world 2

X_1	$X_1 = 1–U$	0	☐	☑
X_2	$X_2 = U$	1	☐	☑
Y	$Y = \min(X_1, 1–X_2)$	0		

world 3

X_1	$X_1 = U$	1	☑	☑
X_2	$X_2 = 1–U$	0	☐	☑
Y	$Y = \min(X_1, 1–X_2)$	1		

world 4

X_1	$X_1 = 1–U$	0	☐	☑
X_2	$X_2 = 1–U$	0	☐	☐
Y	$Y = \min(X_1, 1–X_2)$	0		

Counterdetermination represents the main problem for definition 3: it is not difficult to see that our intuitive judgements differ from the ones based on definition 3 in three of the four possible cases.

3.4. Backup counterdetermination viz. prevention

Prevention is given by the presence of two causes, acting in opposite directions, where the second cause is activated by the first one being or becoming actual. Having example 2.4 in mind, it seems to be safe to assume that, when the agent cause is actual, the induced actuality of the counteragent cause is the actual cause of the fact that the world does not change, while, when the agent cause is not actual, the fact that the world does not change has no causes, because nothing is acting to change anything.

PEI leads to the same result. Since Y maintains the value 0 anyway, then either no cause is activated or two counterbalancing causes occur. If $X_1 = 1$, then $X_1 = 1$ and $X_2 = 1$ are counterbalancing causes, and $X_2 = 1$ is the actual cause of $Y = 0$. If $X_1 = 0$, then no cause is activated, and no actual cause occurs.

Two possible worlds:

variable	equation	value	I-cause	D-cause

world 1

X_1	X_1 = U	1	☐	☐
X_2	$X_2 = X_1$	1	☑	☑
Y	$Y = \min(X_1, 1–X_2)$	0		

world 2

X_1	X_1 = 1–U	0	☐	☑
X_2	$X_2 = X_1$	0	☐	☐
Y	$Y = \min(X_1, 1–X_2)$	0		

In this case too, judgements based on definition 3 are not wholly consistent with intuitive judgements. To see that at world 2 $X_1 = 0$ is an actual cause of $Y = 0$ choose $\boldsymbol{X} = \langle X_2 \rangle$ and $\boldsymbol{x} = \langle 0 \rangle$. Then:

AC2, i) $\mathcal{M} \models_{\mathbf{u}} [X_1 = 1, X_2 = 0](Y = 1)$
AC2, ii) $\mathcal{M} \models_{\mathbf{u}} [X_1 = 0, X_2 = 0](Y = 0)$

3.5. A possible response

How the discrepancies noticed in the precedent analysis can be coped with? In Halpern and Pearl (2005) it is observed that we only have to disallow some worlds as serious possibilities. Halpern and Pearl consider the following situation[13]:

> Suzy goes away on vacation, leaving her favourite plant in the hands of Billy, who has promised to water it. Billy fails to do so, and the plant die. But Vladimir Putin also failed to do so. Therefore, if we allow Billy's omission to water the plant as a cause of its death, then we have to allow Putin's omission as a cause as well.

This situation can be treat using two strategies: the first one consists in disallowing models including exotic variables, i.e., variables not corresponding to serious possibilities; the second and more general one, the one adopted in Halpern and Pearl (2005), consists in allowing all the possible models, but disallowing some possible worlds, i.e., some settings of the variables, e.g. the setting where Putin waters the plant. In this way, modelling becomes much more flexible and almost all the cases are treatable with success[14]. As an example, in cases of primary overdetermination, it is possible to disallow world 4 to achieve a perfect correspondence between intuitive and "official" judgements.

[13] See Halpern and Pearl (2005), p 25. This example is taken into account in Hall and Paul (2003), and exhibits a paradigmatic case of multiplication of causes.

[14] See also Hitchcock (2001), pp.290-98 and Woodward (2003), ch. 2, § 2.8 on causation, omission, and serious possibilities. The strategy followed by Halpern and Pearl can be considered as a formalization of Woodward's approach in coping with puzzling examples.

Is a solution of this kind satisfactory? The problem here is given not only by the introduction of a subjective element in the construction of causal models[15], but mainly by the circularity involved in the process by which the construction of such models is accomplished. Why we are willing to exclude some world as a serious possibility? I think that, in answering this question, one has to make use of her causal intuitions about the world. So, why one makes the choice of disallowing the setting in which Putin waters the plant? Because there is no causal connection linking Putin to the situation depicted in the previous example[16].

However, even if one is disposed to embrace the proposed strategy, other cases seem to put into question its efficacy.

4. New puzzling cases

As we have seen, in our simplified models backup over- and counterdetermination are represented using these sets of equations:

overdetermination	counterdetermination
$X_1 = U$	$X_1 = U$
$X_2 = 1–X_1$	$X_2 = X_1$
$Y = \max(X_1,X_2)$	$Y = \min(X_1,1–X_2)$

However, a different kind of backup over- and counterdetermination can be modelled using the following sets of equations:

overdetermination	counterdetermination
$X_1 = U$	$X_1 = U$

[15] See Halpern and Pearl (2005): "As the examples have shown, much depends on choosing the "right" set of variables with which to model a situation, which ones to make exogenous, and which to make endogenous. While the examples have suggested some heuristics for making appropriate choices, we do not have a general theory for how to make these choices. We view this as an important direction for future research.". See also Woodward (2003), p. 88-91, where the subjectivity problem is faced trying to show that at least in part the subjective choice of models and worlds is based on objective facts about how the world works.

[16] The strategy of distinguishing default and deviant worlds, see Hitchcock (2007) and Hall(2007), seem to undergo a similar critique, unless the distinction is made without relying on our causal intuitions.

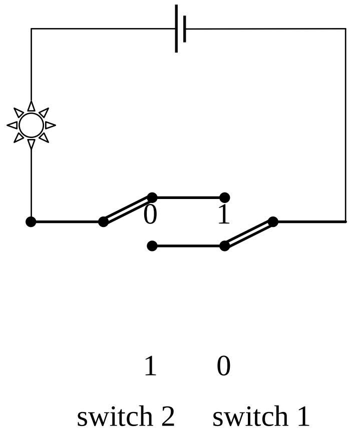
$X_2 = 1–X_1$ $\quad\quad$ $X_2 = X_1$

$Y = \text{if}(X_1{=}X_2;0;1)$ $\quad\quad$ $Y = \text{if}(X_1{=}X_2;0;1)$

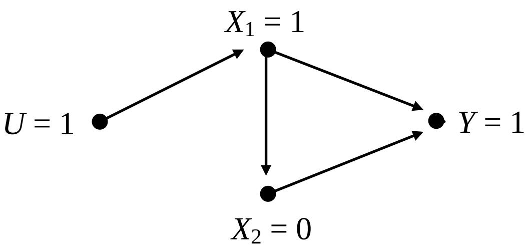
Note: for the sake of brevity I'm going to use the if(C,x,y) function, which returns x, if condition C is true, and y, if the condition is false. Accordingly, Y = if(X_1=X_2;0;1) states that Y = 0, if X_1=X_2, and 1 otherwise.

Which is the difference between the first and the second kind of overdetermination? Let us consider an example.

Example 4.1: circuit consisting of a light bulb and two switches, S_1 and S_2.

If switch 1 is ON, then the light is on. If switch 2 is ON, then the light is ON. But, if both switches are ON, then the light is OFF. In fact, when switch 1 is ON it changes the way in which switch 2 works.	

Now, two men are sent for turning the light on. The task of the first one is to switch S_1 on, while the task of the second one is to switch S_2 on just in case the first man does not achieve his task. Let's assume that the first man has success. The causal model looks as follows:



Equations:

$X_1 = U$

$X_2 = 1–X_1$

$Y = \text{if}(X_1{=}X_2;0;1)$

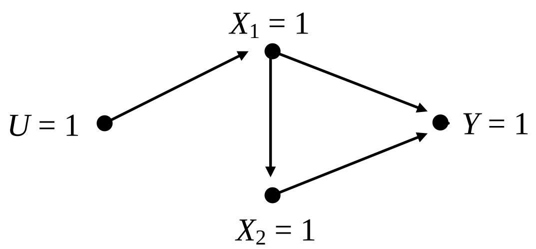
$X_1 = 1$:= S_1 is ON

$X_2 = 0$:= S_2 is OFF

$Y = 1$:= the light is ON

I think it is right to say that $Y = 1$ is overdetermined by backup overdetermination.

Example 4.2: a killer puts poison in the king's tea; a bodyguard responds by pouring an antidote in the king's tea; the antidote neutralizes the poison in the king’s tea, but it is lethal when taken by itself. The model is:



Equations:

$X_1 = U$

$X_2 = 1–X_1$

$Y = \text{if}(X_1=X_2;0;1)$

$X_1 = 1$:= the killer poisons the king’s tea

$X_2 = 1$:= the bodyguard pours the antidote

$Y = 0$:= the king does not die

Example 4.1 is a case of backup overdetermination, since there is a second potential cause that would be activated by the failure of the first one to become actual. Nevertheless, if both causes became actual, there would be no effect. Similarly example 4.2 is a case of backup counterdetermination, since there is a counteragent cause activated by the fact that the first one becomes actual. Nevertheless, if only the counteragent cause became actual, there would be an undesired effect. Notice that, in such cases, the strategy of disallowing some setting of the variables cannot be followed, since there is no exotic setting.

4.1. A second kind of backup overdetermination

What implications do these new kinds of over- and counterdetermination have for the matching of definition 3 with our intuitive judgements?

In such cases we are inclined to say that, when X_1 takes the value 1, then $X_1 = 1$ is the actual cause of $Y = 1$, and, when X_1 takes the value 0, then $X_2 = 1$ is the actual cause of $Y = 1$.

Two possible worlds:

variable	equation	value	I-cause	D-cause

world 1

X_1	$X_1 = U$	1	☑	☑
X_2	$X_2 = 1–X_1$	0	☐	☑
Y	$Y = \text{if}(X_1{=}X_2;0;1)$	1		

world 2

X_1	$X_1 = 1–U$	0	☐	☑
X_2	$X_2 = 1–X_1$	1	☑	☑
Y	$Y = \text{if}(X_1{=}X_2;0;1)$	1		

The output based on definition 3 is rather surprising: both variables have an active causal role. Choose $\boldsymbol{X} = \langle X_1 \rangle$ and $\boldsymbol{x} = \langle 1 \rangle$ to check that $X_2 = 0$ is an actual cause at world 1, and $\boldsymbol{X} = \langle X_2 \rangle$ and $\boldsymbol{x} = \langle 1 \rangle$ to check that $X_1 = 0$ is an actual cause at world 2.

4.2. A second kind of backup counterdetermination

In such cases we are inclined to say that, if X_1 takes the value 1, then $X_2 = 1$ is the actual cause of $Y = 0$, and that, if X_1 takes the value 0, then neither X_1 nor X_2 is an actual cause of $Y = 0$.

Two possible worlds:

variable	equation	value	I-cause	D-cause

world 1

variable	equation	value	I-cause	D-cause
X_1	X_1 = U	1	☐	☑
X_2	$X_2 = X_1$	1	☑	☑
Y	Y = if(X_1=X_2;0;1)	0		

world 2

variable	equation	value	I-cause	D-cause
X_1	X_1 = 1–U	0	☐	☑
X_2	$X_2 = X_1$	0	☐	☑
Y	Y = if(X_1=X_2;0;1)	0		

The output based on definition 3 is: in both cases both variables have an active causal role. To check that $X_1 = 1$ is an actual cause at world 1, choose $\boldsymbol{X} = \langle X_2 \rangle$ and $\boldsymbol{x} = \langle 1 \rangle$. To check that $X_1 = 0$ is an actual cause at world 2, choose $\boldsymbol{X} = \langle X_2 \rangle$ and $\boldsymbol{x} = \langle 0 \rangle$. Finally, to check that $X_2 = 0$ is an actual cause at world 2, choose $\boldsymbol{X} = \langle X_1 \rangle$ and $\boldsymbol{x} = \langle 0 \rangle$.

Remark: Notice that our intuitive judgements concerning causation in the preceding cases are supported by another general principle.

Principle of existential import: it is impossible for the same event in the same model, i.e., given the same causal connections, to be caused both by the occurrence and by the non-occurrence of the same event.

This principle is contradicted by the outputs based on definition 3, since, both in example 4.1 and in example 4.2 a certain event is picked out as a cause both when it occurs and when it does not occur.

4.3. A structural difficulty

Let us now move to what I take to be the last crucial drawback concerning definition 3. It is impossible, within the framework based on this definition, to account for our intuitions concerning the differences between over- and counterdetermination. Actually, basic causal models of over- and counterdetermination share, in a sense, the same structure. Let's denote with X^* the event that is dual with respect to X (e.g. if X stand for being ON, X^* stands for being OFF). We can transform a model of counterdetermination into a model of overdetermination, and vice versa, simply by operating the following transformations: $X_1 \rightarrow X_1^*$; $X_2 \rightarrow X_2$; $Y \rightarrow Y^*$.

Primary overdetermination:

$$X_1 = U \quad \rightarrow \quad X^*_1 = 1-U$$
$$X_2 = U \quad \rightarrow \quad X_2 = U$$
$$Y = \max(X_1,X_2) \rightarrow \quad Y^* = \min(1-X_1,1-X_2) = \min(X^*_1,1-X_2)$$

Backup overdetermination:

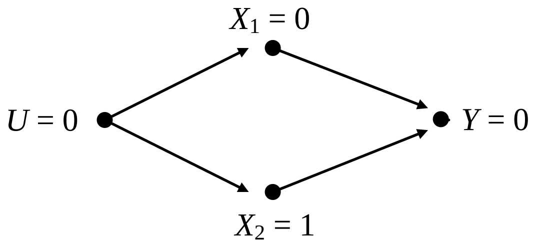
$$X_1 = U \quad \rightarrow \quad X^*_1 = 1-U$$
$$X_2 = 1-X_1 \quad \rightarrow \quad X_2 = X^*_1$$
$$Y = \max(X_1,X_2) \quad Y^* = \min(1-X_1,1-X_2) = \min(X^*_1,1-X_2)$$

Second kind of backup overdetermination:

$$X_1 = U \quad \rightarrow \quad X^*_1 = 1-U$$
$$X_2 = 1-X_1 \quad \rightarrow \quad X_2 = X^*_1$$
$$Y = \mathrm{if}(X_1=X_2;0;1) \rightarrow \quad Y^* = \mathrm{if}(X_1=X_2;1;0) = \mathrm{if}(X^*_1=X_2;0;1)$$

Thus, every possible world in a model for a case of overdetermination can be transformed into a possible world in a model for a case of counterdetermination. However, the intuitive judgements about cases of over- and counterdetermination do not allow for such transformations: indeed, in cases of counterdetermination, it seems impossible for both the agent and the counteragent cause to be actual causes of the same effect, especially when the agent cause is not actual. Therefore, the intuitively acknowledged asymmetry between over- and counterdetermination is not reflected in the treatment based on the application of definition 3.

Example 4.3: consider the following model:



Equations

$X_1 = U$

$X_2 = U$

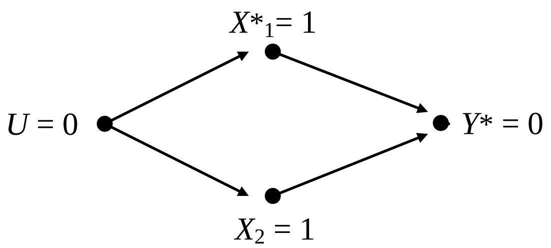
$Y = \min(X_1,1\text{-}X_2)$

$X_1 = 0$:= the king’s tea is not poisoned

$X_2 = 1$:= the king’s tea contains an antidote

$Y = 0$:= the king is not death

The model can be transformed into:



Equations

$X_1 = U$

$X_2 = U$

$Y^* = \max(X^*_1,X_2)$

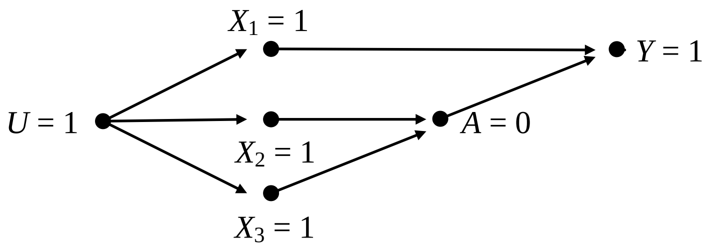
$X^*_1 = 1$:= the king’s tea is poison-free

$X_2 = 1$:= the king’s tea contains an antidote

$Y^* = 1$:= the king lives

According to our intuition on primary overdetermination, both $X^*_1 = 1$ and $X_2 = 1$ should be actual causes of $Y^* = 1$. However, according to our intuition on primary counterdetermination, neither $X_1 = 0$, i.e., $X^*_1 = 1$, nor $X_2 = 1$ should be an actual cause of $Y^* = 1$.

Example 4.4: existent and non-existent threats[17].



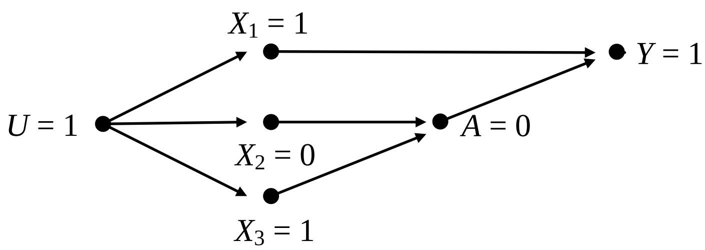
Equations: $X_1 = X_2 = X_3 = U_1$;

$A = \min(X_2, 1–X_3)$;

$Y = \min(X_1, 1–A)$.

Such a model can be used to represent situations of causation by double prevention[18]: if the value of X_3 was not 1, the value of A would be 1, and, consequently, the value of A would be 0. Thus, $X_3 = 1$ prevents A from preventing $Y = 1$, and, in accord with our intuition, definition 3 picks out $X_3 = 1$ as an actual cause of $Y = 1$: choose $\boldsymbol{X} = \langle X_2 \rangle$ and $\boldsymbol{x} = \langle 1 \rangle$.

Let us now consider a different possible world on the same model:



Equations: $X_1 = X_3 = U$; $X_2 = 1–U$;

$A = \min(X_2, 1–X_3)$;

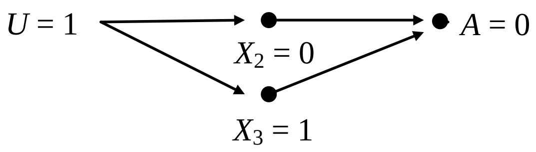
$Y = \min(X_1, 1–A)$.

[17] An analysis of cases of threats, as cases of causation by double prevention, is provided by Hall. See Hall (2007).

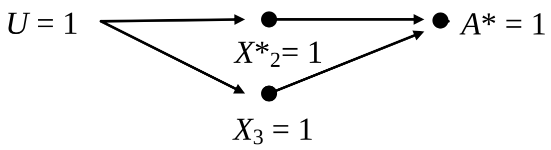
[18] See Hall (2007), p. 120: "The family sleeps peacefully through the night, in part because the watchful police have nabbed the thief before he can enter the house".

In this case, we would hardly say that $X_3 = 1$ is an actual cause of $Y = 1$, since $X_3 = 1$ prevents nothing. Still, this is not the outcome of definition 3: again, choose $\boldsymbol{X} = \langle X_2 \rangle$ and $\boldsymbol{x} = \langle 1 \rangle$.

The above mentioned situation is just another case of failure in distinguishing overdetermination and counterdetermination. The fragment



is an example of counterdetermination. By transforming it, one gets



which is an instance of overdetermination. In the latter case we have no problem in saying that both $X^*_2 = 1$ and $X_3 = 1$ are actual causes of $A^* = 1$, but in the former one we would not count $X_3 = 1$ as an actual cause, given that it seems to cause nothing.

In conclusion, the possibility of transforming models of overdetermination into models of counterdetermination leads our intuition into puzzles. Therefore, since this kind of transformation is allowed if definition 3 is accepted, this definition appears to be not able to meet our basic causal intuitions.

5. Conclusion

The use of causal models for analysing causation has several advantages. It provides a definite representation of causal connections and a general framework in which one can try to determine what actual causation is. Furthermore, it provides a procedure for defining factual and counterfactual dependencies, giving the opportunity of developing a definition of actual causation based on the relation of counterfactual dependency. Finally, the basic principles it involves, i.e., using variables to represent states and events, using structural equations to represent causal connections between events, and using structural interventions to model counterfactual dependencies and to offer a description of the way the world would have been, if some events were not have been actual, fit well both with the philosophical way of addressing causation and with the scientific way of modelling dynamical systems. Still, it seems to be in need of further development in order to cope with all of our intuitive judgements about actual causation. As we have seen, definition 3 is too permissive, while the proposed extended version seems to obscure the progress made in stating that definition, attributing too much power to our subjective choices, indeed our causally oriented subjective choices, concerning the introduction of variables and the selection of allowable settings. Taking into account the results of the foregoing sections, the most promising direction for future research seems to be that of improving the definition proposed by Halpern and Pearl, introducing further necessary conditions, in order i) to limit its permissiveness, and ii) to account for the correspondence between intuitive causal judgements and judgements based on the application of general principles, such as the principle of existential inertia and the principle of existential import. In particular it would seems promising to make the approach dynamic, by taking into account the way in which a causal system evolves from different initial conditions, so to work with structural relations characterizing pair of worlds, modelling initial and final conditions, rather than a single world, modelling the outcome of an intervention.

References

Clark, P. & Hawley, K. (eds.). (2003), *Philosophy of science today*, Oxford: Oxford University Press.

Collins, J., Hall, N., & Paul, L. (eds.) (2004), Counterfactuals and *Causation*, Cambridge: MIT Press.

Hall, N. (2004), Two Conceptions of Causation, in Collins, Hall & Paul (2004), pp. 225-276.

Hall, N. (2007), Structural equations and causation, *Philosophical Studies*, 132, pp.

109-136.

Hall, N. & Paul, L. (2003), Causation and Preemption, in Clark, P. & Hawley, K. (2003), pp. 100-130.

Halpern, J. Y. and C. Hitchcock (2014), Graded causation and defaults. British Journal for the Philosophy of Science, forthcoming.

Halpern, J. & Pearl, J. (2001), Causes and Explanations: A Structural Model Approach–Part I: Causes, in *UAI 2001*, pp.194-201.

Hiddlestone, E. (2005), Causal Powers, British Journal for the Philosophy of Science, 56, pp. 27-59.

Hitchcock, C. (2001), The Intransitivity of Causation Revealed in Equations and Graphs, *Journal of Philosophy*, 98, pp. 273-314.

Hitchcock, C. (2007), Prevention, Preemption, and the Principle of Sufficient Reason. *Philosophical Review*, 116, pp. 495-532.

Lewis, D. (1973a), *Counterfactuals*. Oxford: Basil Blackwell.

Lewis, D. (1973b), Causation, *Journal of Philosophy*, 70, pp. 556-67. Reprinted in Lewis 1986.

Lewis, D. (1979), Counterfactual Dependence and Time's Arrow, *Nous*, 13, pp. 455-76. Reprinted in Lewis 1986.

Lewis, D. (1986), *Philosophical Papers: Volume II*, Oxford: Oxford University Press.

Menzies, P. (2003), Difference-Making in Context, in Collins, Hall, N., & Paul (2003), pp. 139-180.

Menzies, P. (2004), Causal Models, Token Causation, and Processes, *Philosophy of Science*, 71, pp. 820-832.

Pearl, J. (2000), *Causality: Models, Reasoning, and Inference*, Cambridge: Cambridge University Press.

Woodward, J. (2003), *Making Things Happen: A Theory of Causal Explanation*, New York: Oxford University Press.

Woodward, J. (2006), Sensitive and Insensitive Causation, *The Philosophical Review*, 115, pp. 1-50.

Yablo, S. (2002), De facto dependence, *Journal of Philosophy*, 99, pp. 130-148.