The Interventionist Account of Causation and Non-causal Association Laws

Max Kistler
Université Paris 1
mkistler@univ-paris1.fr
http://max.kistler.free.fr

Abstract

The key idea of the interventionist account of causation is that a variable A causes a variable B if and only if B would change if A were manipulated in an appropriate way. I argue that Woodward's (2003) version of interventionism does not provide a sufficient condition for causation, insofar as it is not adequate for manipulations grounded on association laws. Such laws, which express relations of mutual dependence between variables, ground manipulative relationships which are not causal. I suggest that the interventionist analysis is sufficient for nomological dependence rather than for causation.

1. The interventionist analysis of causation

According to the interventionist account, (type level) causation is a relation between variables[1]. Its fundamental hypothesis is that a variable A causes a variable B if and only if there are circumstances in which it is possible to manipulate B by intervening on A. According to Woodward (2003; 2008), this idea underlies scientific research for causes across all sciences. He gives the following example from social science. One can observe, in the contemporary US, a statistical correlation between children's attendance of private schools (P) and their scholastic achievements (A). A randomized experiment would be a straightforward way by which a social scientist could try to find out whether this correlation stems from the fact that attendance of private schools causes better scholastic achievement or whether both variables are effects of some common cause, such as the parents' higher socio-economic status (S). Such an experiment requires randomly attributing children from a group of fixed S to two sub-groups: one sub-group is sent to a public school, the other to a private school. This is equivalent to attributing one value of P to the individuals in the experimental group and another value to those in the control group. Making the assignment to the two subgroups random is intended to make it independent of other factors that could influence A independently of P. It is also supposed to make sure that the decision to assign any child to the experimental group (which attends a private school) or to the control group (which attends a public school) does not influence her scholastic achievements, through paths that do not run through the type of school she attends. After a suitable lapse of time, A is measured in the two subgroups. Any correlation that is found between A and P can be taken to reflect the existence of a causal influence of P on A. The possibility that A and P are the effects of some common cause such as S has been excluded by randomizing the attribution of a value to variable P for each individual. This is supposed to ensure that P is statistically uncorrelated with S, and indeed with any other variables that might be common causes of P and A.

The interventionist account of causation is even more plausible as an analysis of causation in experimental sciences such as physics or chemistry, where it is often practically easier to control the values of variables than in the social sciences. It is more difficult to

[1] The interventionist account has been generalized in order to be applicable to « actual causation », where the relata are values of variables rather than variables. Cf. Woodward (2003, p. 74-86). Woodward takes actual causation between values of variables to be equivalent to what has often been called « token causation ». Token causation relates particular events localized at a unique place and time, whereas type causation relates variables that can have many instantiations at many different times and places.

arbitrarily set the value of socio-economic status for a given individual, than to set the value of variables such as the intensity of electric current in a copper wire in a physics laboratory. Let S, A and P represent physical variables characterizing observable and manipulable properties of copper wires. Let P be the electric current flowing through the wire, A the heat release from the wire, and S the room temperature. The fundamental idea of the interventionist account is that P causes A if and only if the following is true: if the room temperature S, as well as all other variables that might influence A are held fixed (except of course P and A themselves), then, if one intervened on P by changing its value, without thereby intervening on A through paths that run through other potential causes of A but do not run through P, then the value of A would change.

Relations of manipulation between variables can be represented within the framework of graphs. A graph is a pair whose constituents are 1) a set V of variables and 2) a set of edges connecting these variables pairwise. Edges represent relations of possible manipulation. These possible manipulations impose a direction on the edges: If X and Y are two variables connected by an edge, the edge is directed toward Y if and only if there is a possible intervention on X, such that, if the intervention changed the value of X while the values of all other variables in the set V were held fixed at some value, the value of Y would undergo a change.

This framework allows various causal concepts to be defined. For the purposes of this paper, it will be sufficient to concentrate on the notion of a *direct cause*. "A necessary and sufficient condition for X to be a direct cause of Y with respect to some variable set V is that there be a possible intervention on X that will change Y (or the probability distribution of Y) when all other variables in V besides X and Y are held fixed at some value by interventions." (Woodward 2003, p. 55).

This definition makes crucial use of the notion of an intervention. Typically – but not necessarily – the value of an intervention variable is determined by a human experimenter, and typically - but not necessarily - intervention variables are exogeneous, i.e. variables whose values are not determined by the values of other variables within V. Here is Woodward's (2003) definition of an intervention variable.

"I is an intervention variable for X with respect to Y if and only if I meets the following conditions:

I1. I causes X.

I2. I acts as a switch for all the other variables that cause X. That is, certain values of I are such that when I attains those values, X ceases to depend on the values of other variables that cause X and instead depends only on the value taken by I.

I3. Any directed path from I to Y goes through X. That is, I does not directly cause Y and is not a cause of any causes of Y that are distinct from X except, of course, for those causes of Y, if any, that are built into the I-X-Y connection itself; that is, except for (a) any causes of Y that are effects of X (i.e., variables that are causally between X and Y[2]) and (b) any causes of Y that are between I and X and have no effect on Y independently of X.

I4. I is (statistically) independent of any variable Z that causes Y and that is on a directed path that does not go through X." (Woodward 2003, p. 98)

In Woodward's example mentioned above, the intervention variable represents the experimenter's decision to send a given child to private school. This is an intervention according to the definition because 1) it causes P, in the sense that the intervention determines whether the child attends a private school. 2) It is part of the idea of a randomized experiment that only the experimenter's decision to attribute a given child to the experimental or control group determines whether she attends a private school or not. 3) The decision to put a child in

[2] This clause does not apply to the case of direct causation.

the experimental group that attends a private school does not *directly* influence the child's scholastic achievements, i.e. if it influences them, it does so only by way of her attending a private school. 4) The very idea of a randomized experiment consists in making the determination of the value of P independent of all other variables, and in particular of variables that might influence A.

In the physical example mentioned above, raising the voltage across the copper wire satisfies these conditions on an intervention on the electric current P: 1) A change of the voltage I causes a change in the electric current P. 2) The electric current P is determined only by the voltage I (given the electrical resistance of the wire). 3) The voltage does not directly cause the wire to release heat but only through the flow of electrical current it causes. 4) The voltage is statistically independent of other causes of the wire's releasing heat, such as the room temperature, or various kinds of radiation.

2. Association Laws

Woodward's conditions are not sufficient for X being a direct cause of Y. My argument for this claim involves functional association laws, which are symmetric in the sense that they express mutual functional dependence between two variables X and Y, given other variables. Such laws create the following conceptual problem for Woodward's analysis of direct causation. Intervening on X (while holding other variables fixed) changes Y, so that X should be a direct cause of Y; but intervening on Y changes X, so that Y should be a direct cause of X. This mutual dependence holds for every particular system to which the law applies, at every instant. This can be made explicit by taking the relevant variables to be specific for a system s and a time t[3]. Let P be the generic variable representing electric current and A the generic variable representing heat release. The formalism of graphs is mostly used at this generic level: statistical tools are used to determine whether P is correlated with A. But causal processes take place at determinate places and times. This can be made explicit by using specific variables: P(s,t) is the variable representing the electric current in wire s at time t. In general, let X(s,t) be the specific variable representing the value X takes in system s at t.

In terms of specific variables, the problem is this. If X(s,t) and Y(s,t) are related by an association law, the interventionist analysis yields the result that X(s, t) is a direct cause of Y(s, t), and that Y(s, t) is a direct cause of X(s, t). This is incompatible with the asymmetry of causation. I conclude that the relation characterized by Woodward's conditions cannot be causation. Rather, it is a relation that is not asymmetric and easily confounded with causation. I suggest it is nomological dependence.

It might seem that using specific variables implies changing the subject from type level causation, which is Woodward's topic, to token level causation[4]. But Woodward certainly intends his account to apply to specific variables as well as to generic variables. Take the example he offers for a direct cause: "With respect to a set that includes variables like A's desire for revenge, A's pulling the trigger of the gun, A's hitting B on the head with a rock, A's poisoning B's drink, and B's death, A's desire for revenge may be a direct cause of his pulling the trigger, which may in turn be a direct cause of B's death." (Woodward 2003, p. 55) Clearly, things like A's pulling the trigger are particular, dated events. The variables

[3] Spohn (2001a; 2001b ; 2006) uses specific variables in this sense, whereas Woodward (2003) following Spirtes, Glymour and Scheines (2000) and Pearl (2000), uses generic variables.

[4] The causal relation between specific variables bears close resemblance to but should not be confused with (what Woodward and others call) "actual causation". Actual causation, as defined by Woodward (2003, p. 77) is a relation between specific *values* of variables, whereas specific variables are still variables.

representing them are specific variables: they represent properties of particular systems at particular times, such as *A's* pulling the trigger *at t*[5].

The law I will take as an example is a system law[6] valid for all devices containing rotating electrically charged masses. Magnetic stirrers obeying this law are used in chemistry for mixing substances. The law says that the angular momentum due to the rotation of the mass is proportional to the magnetic moment due to the rotation of the electric charge. It can be easily derived from two more general laws of nature.

(1) The angular momentum **L** of a mass m rotating with speed **v** in a circle with diameter **r** is

$$\vec{L} = \vec{r} \times m\vec{v}$$

where **r** is the particle's position in a coordinate system centered at the centre of rotation, and × denotes the cross product. Variables in boldface are vectors.

(2) The magnetic moment **μ** of an electric charge e rotating with speed **v** in a circle with radius **r** is

$$\vec{\mu} = \frac{1}{2} e\vec{r} \times \vec{v}.$$

A little calculation putting (1) and (2) together yields

$$\text{(MS)}\ \vec{L} = \vec{\mu}\, \frac{2m}{e}$$

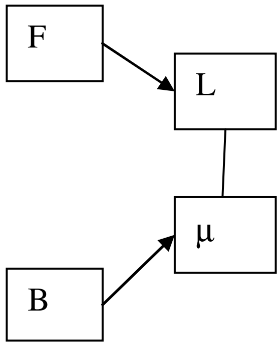
This "magnetic stirrer law" (MS), according to which **L** and **μ** are functions of each other, is an association law. The corresponding specific variables, characterizing a determinate magnetic stirrer s at a determinate instant t, are **L**(s, t), **μ**(s, t). Each particular magnetic stirrer obeys at each moment t a specific law that is an instance of the general magnetic stirrer law.

$$\vec{L}(s,t) = \vec{\mu}(s,t)\, \frac{2m}{e}$$

The law can be tested either by manipulating **L** by exerting a mechanical force **F** on the rotating object while holding fixed its mass m and charge e, and observing the change in **μ**, or by manipulating **μ**. The latter can be done by increasing the strength of the magnetic field **B**, which accelerates the rotation of the charge. The law (MS) is then tested by observing whether and how much this influences **L**. Fig. 1 shows a graph representing variables **L** and **μ** connected by (MS), as well as two intervention variables: **F** represents the mechanical force exerted on the rotating object; **B** represents the magnetic field. The controversial question is whether the edge between **L** and **μ** can be given a causal interpretation.

[5] I thank an anonymous referee for helping me clarify this point.

[6] The domain of such a law is not universal, as is the case with general laws of nature, but consists of all systems of a given type. Cf. Schurz (2002). Cummins (2000) calls them "in situ" laws. Cartwright (1999) calls systems obeying such laws "nomological machines".



F = mechanical force on rotating object **L** = magnetic moment
B = magnetic field **μ** = angular momentum

Fig. 1. Graph representing the variables relevant for testing the magnetic stirrer law (MS)

Let us take the mass m and charge e of the rotating object to be fixed. An intervention on **L**(s,t) by exerting mechanical force **F** will change **μ**(s,t) and an intervention on **μ**(s,t) by manipulating the strength of the magnetic field **B** will change **L**(s,t). Therefore, the interventionist model leads to the conclusion that **L**(s,t) and **μ**(s,t) are mutual causes of each other. It is crucial here to distinguish between generic and specific variables. Contrary to what the interventionist analysis implies, two specific variables, representing different properties of the same system at the same time, cannot be mutual causes of each other. Here is a simple argument for this claim. Let us assume the interventionist analysis. Given that manipulating **L**(s,t) changes **μ**(s,t) and manipulating **μ**(s,t) changes **L**(s,t), **L**(s,t) is cause of **μ**(s,t) and **μ**(s, t) is cause of **L**(s,t). Now suppose causation is transitive. Then we get the absurd consequence that **L**(s,t) and **μ**(s,t) are causes of themselves, because, e.g., **L**(s,t) causes **μ**(s,t) which causes **L**(s,t), so that **L**(s,t) causes **L**(s,t). This argument makes crucial use of the transitivity of causation. However, it is controversial whether causation is transitive. I will argue later that the burden of proof is on my opponent who holds that one cannot use transitivity in this case. I will show that this claim is not well grounded insofar as all cases which have been claimed to show that causation is not transitive are very different from the present case. But I think that although the use of transitivity makes my argument particularly strong (because it leads to the absurd result that some specific variables are causes of themselves), there is a weaker form of the argument that does not use transitivity. The asymmetry of cause and effect belongs to the conceptual core of causation: insofar as causes and effects are taken to be localized in space and time, it is part of the content of the notion of causation that if x causes y, then y cannot at the same time cause x. So even without transitivity, my argument shows that application of the interventionist account to the magnetic stirrer system yields an absurd result: that **L**(s,t) is cause of **μ**(s,t) and **μ**(s, t) is cause of **L**(s,t).

Graphs are mostly used to represent relations between generic variables. At the level of generic variables, feedback cycles are common, in which the causal influence of variable X on variable Y coexists with a reverse influence of variable Y on variable X[7]. However, such feedback cycles are very different from relations of mutual determination grounded on association laws. The difference appears clearly as soon as time is explicitly represented. Take the economic feedback circle in which the increase of demand D of a good increases its price P, which increase in turn lowers the demand D. The influences of D on P and of P on D are causal but not simultaneous.

At the level of generic variables, which are independent of the times of their instances, price and demand form a circle (fig. 2). As soon as we switch from generic variables to time-specific variables, the circle is replaced by a zig-zag line (fig. 3). Fig. 3 shows a graph with

[7] Pearl (2000, p. 12ff.).

specific variables and a temporal dimension. It shows that each effect is delayed with respect to its cause. D(s, t_1) influences P(s, t_2), which influences D(s, t_3), which influences P(s, t_4) etc, where, for all i, t_i is earlier than t_{i+1}.

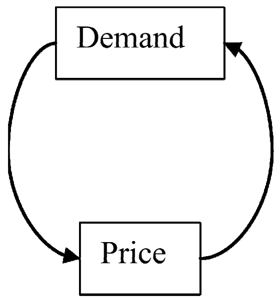


Fig. 2: Cycle involving generic variables

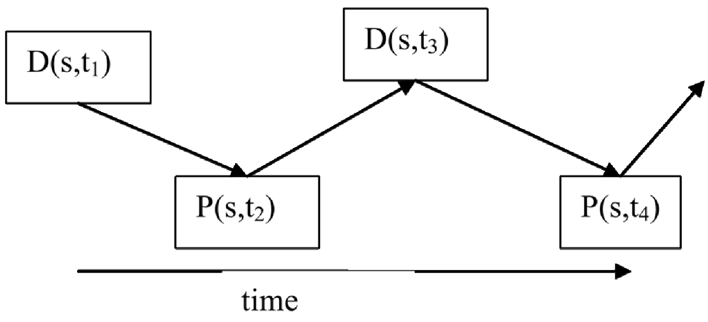


Fig. 3: Delayed influence between different time-specific variables

If we make the same move from generic to time-specific variables for variables related by a law of simultaneous association, the result is different.

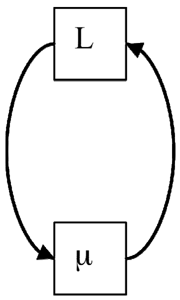


Fig. 4: Cycle involving generic variables

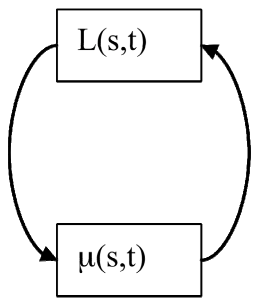


Fig. 5: Mutual influence between different time-specific variables

Fig. 5 shows that the relation between the generic variables **L** and **μ** is no feedback cycle. However, this is not apparent at the level of generic variables (fig. 4), where their relation cannot be distinguished from a feedback cycle. Two variables form a feedback cycle if two conditions are satisfied. 1. The generic variables form a circle (as in fig. 2 and fig. 4). 2. Specific variables taken for the same system at the same time do *not* form a circle. According to these criteria, price and demand really form a feedback cycle. In their case, specific variables for the same system at the same time are not related by a circle, and thus, the problem of self-causation does not arise. However, **L** and **μ** do not form a feedback cycle, because the relevant specific variables also form a circle.

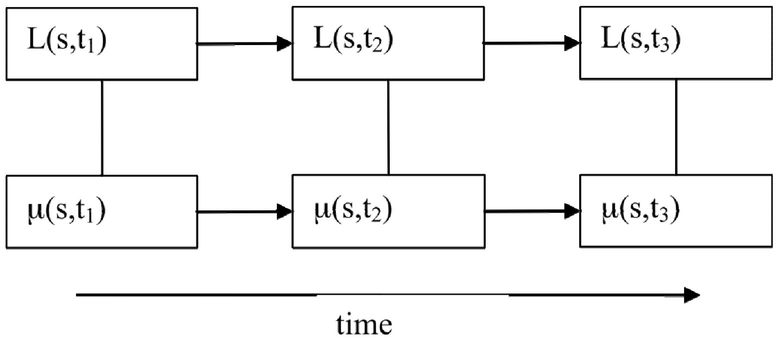


Fig. 6. Time evolution of specific variables linked by a simultaneous association law

The difference between a simultaneous association law and a feedback cycle also appears clearly if we represent the evolution of the specific variables related by such laws in a graph with a temporal dimension (fig. 6). If, for every moment t_i, two specific variables **L**(s,t_i) and **μ** (s,t_i) are related by a law of simultaneous association, at each moment t_i, the variables corresponding to that instant t_i stand in a relation of mutual dependence. These dependence relations appear as vertical lines in fig. 6. If each of these dependence relations were causal, we would, given the transitivity of causation, get the absurd result that, at each moment, each of these specific variables was a cause of itself[8].

Could we not solve the problem raised by the mutual manipulability of variables linked by simultaneous association laws, simply by adding the requirement of a temporal delay to the condition of manipulability? **L**(s,t) does not cause **μ**(s,t) and vice versa, although each can be used to manipulate the other, simply because these specific variables do not satisfy the requirement of temporal delay. In a way, this will be the suggestion I will make at the end. But it seems preferable that this requirement not be simply added ad hoc, but rather follows from more general considerations. Indeed, simply requiring that the effect follows the cause after some finite delay leaves it unclear why simultaneous causation does not exist, and why delayed manipulability reveals causation whereas simultaneous manipulability does not.

Spohn (2001a; 2001b; 2006) makes a move that looks similar but is in fact more radical. He requires that every specific variable represented in a model concerns a different time; in other words, an acceptable model must not contain more than one specific variable concerning a given instant. Contrary to the postulate we considered in the last paragraph, Spohn does not stipulate that relations between simultaneous variables are never causal. Rather, he excludes the possibility to represent such simultaneous specific variables in the first place. This makes it trivially true that there are no causally related specific variables characterizing the same system at the same time. It is true simply because no model can contain such specific variables relative to the same system at the same time at all.

[8] Maybe this distinction between genuine feedback cycles and mutually dependent variables lies behind Pearl's stipulation that "directed graphs may include directed cycles (e.g., X→Y, Y→X), representing mutual causation or feedback processes, but not self-loops (e.g., X→X)" (Pearl 2000, p. 12). At the level of generic variables, this stipulation seems completely unmotivated. Indeed, as Pearl explicitly proves (2000, p. 237), in the absence of other influences, transitivity holds: if X→Y and Y→Z (but no additional influence X→Z along any pathway independent from the pathway running through Y), then X→Z. Transitivity seems to imply directly that every "directed cycle" X→Y, Y→X necessarily entails the existence a "self-loop" X→X. One coherent interpretation of Pearl's remark would be to take "self-loops" to refer to what I have called relations of mutual dependence (which are circular both at the generic and at the specific levels), whereas directed cycles correspond to what I have called feedback cycles (which are circular only at the level of generic variables). However, Pearl's framework cannot give any justification for his exclusion of self-loops, which therefore seems ad hoc.

This requirement guarantees indeed that manipulability is sufficient for causation. It solves our problem. However, from the point of view of the representation of scientific methodology, the requirement of temporal precedence appears to be ad hoc and too strong.

1. First, it seems ad hoc to disallow the representation of relations between variables linked by an association law, insofar as the experimental investigation of their relation, leading to the discovery of that law, seems to follow exactly the same strategy as the discovery of causal relations: observation of statistical correlations and independencies on the one hand, and experimental intervention on the other hand.
2. The second reason is even more important. The existence of a temporal delay can be the object of experimental enquiry. It may be a matter of empirical research whether a change in one of these variables leads to a simultaneous or a delayed change in the other variable. A model that excludes one of these possibilities by stipulation cannot make sense of such an investigation.
3. Another reason for which it is inappropriate to exclude specific variables bearing on the same instant is that this makes it impossible to analyse situations in which two simultaneous factors A_1 and A_2 influence a given variable B. An example might be a collision between two cars (B) caused by the fact that one car changes lanes (A_1) and that simultaneously the other car changes lanes (A_2)[9].

3. Objections and Replies

Several objections may be raised against my counterexample against the interventionist analysis of causation. One is that **F** and **B** are not intervention variables in Woodward's sense in the first place. One might argue that both interventions through **F** and through **B** always *directly* change both variables **L** and **μ**, whereas an intervention must make only one direct change (in the cause variable), all other changes being indirect and mediated by the first change. The objection is that changes of **F** and **B** are not interventions because they are not "focused" enough: They always make *two* variables change, not just one. If this were the case, it would be appropriate to add two arrows in fig.1, one from **B** to **L**, and one from **F** to **μ**. This would show that the situation is no counterexample to Woodward's analysis, because B and F do not satisfy the conditions for intervention variables in this structure.

My reply takes its inspiration from a recent paper of Woodward's (2011). He argues that the interventionist conditions for direct causation and other causal notions can be applied within models containing variables that cannot be independently manipulated, in particular because they are related by definitions or by a supervenience relation. Consider a model containing variables that are related by definition[10]. It contains variables **HDC** (high density cholesterol), **LDC** (low density cholesterol) and **TC** (total cholesterol), which are related by the equation **HDC+LDC=TC**. Consider an inquiry aimed at finding out whether **HDC** causes heart disease **H**. Any intervention on **HDC**, while **LDC** is held fixed, is necessarily accompanied by a change in **TC**. However, this doesn't seem to be a good reason for denying that interventions on **HDC** are possible. It is good scientific strategy to intervene on **HDC** while holding fixed **LDC**, whereas the requirement to hold both **LDC** *and* **TC** fixed would make enquiries on the causal link from **HDC** to **H** impossible, because (by the definition of **TC**) **HDC** cannot vary if **LDC** and **TC** are held fixed. Thus, explains Woodward, on an appropriate construal of what an intervention is, intervening on **HDC** with respect to **H** requires holding fixed all variables that cause **H** and do not lie on the causal path from **HDC**

[9] I owe this reason and the example to an anonymous referee.
[10] The example that follows is due to Spirtes and Scheines (2004).

to **H**, *except* variables that are related to **HDC** (or **H**) "as a result of supervenience relations or relations of definitional dependence" (2011, p. 27). He calls interventions satisfying this condition "IV*-interventions" (Woodward 2011, p. 27). The original definition (IV) in (Woodward 2003) requires that an intervention **I** on **X** with respect to **Y** must not be "a cause of any other causes of **Y** that are distinct from **X** except, of course, for those causes of **Y** that are built into the **I-X-Y** connection itself" (Woodward 2003, p. 98). Thus, for **I** to be an intervention variable on **X** with respect to **Y** in the sense of the original (IV), it must be possible to hold fixed all variables **Z** that cause **Y** through some route that does not run through **X**. (IV*) weakens the requirement on which variables must be held fixed.

Woodward does not mention non-causal laws of association. However, it seems reasonable to relax the conditions on interventions even further, and also exclude variables that are linked to **X** or **Y** by non-causal association laws from the set of variables that must be held fixed during interventions on **X** with respect to **Y**. The reason is the same as in the cases considered by Woodward (2011): if it were necessary, in order to intervene on **X** with respect to **Y**, to hold fixed all variables related to **X** and **Y** by non-causal association laws, no intervention would ever be possible. I therefore suggest to relax the condition on an intervention even further, introducing the notion of an (IV**)-intervention. The definition of an (IV**)-intervention **I** on **X** with respect to **Y**, is obtained by modifying clause (I3), in the original definition of (IV) in (Woodward 2003, p. 98), in the following way. According to the modified clause (I3**), an intervention **I** on **X** with respect to **Y** must not be a cause of any other causes of **Y** that are distinct from **X** except for those causes of **Y** that are built into the **I-X-Y** connection itself, except for variables that are related to **X** and **Y** as a result of supervenience relations or relations of definitional dependence, *and except for variables related to X or Y by non-causal association laws*. The original clause (I4) from (Woodward 2003), which says that "**I** is (statistically) independent of any variable **Z** that causes **Y** and that is on a directed path that does not go through **X**", must me modified in a similar way, "any variable **Z**" being replaced by "any variable **Z** other than those in the supervenience bases of **X** and **Y** and other than those related to **X** or **Y** by relations of definitional dependence or by non-causal association laws".

With this concept of intervention, **F** counts as an intervention on **L** (if it otherwise satisfies the conditions on an intervention given in (IV**)) although all interventions **F** that change **L** are (by virtue of a law of association between **L** and **μ**) accompanied by changes in **μ**.

A second objection might be that each apparent case of non-causal dependence between variables is really a case of *identity*. According to this objection, if one can manipulate **L** by manipulating **μ** and manipulate **μ** by manipulating **L** although there is no causal relation between **L** and **μ**, this means that **L** and **μ** are not different variables. Rather, they are two names of the same variable. If this were correct, the argument would not after all establish that there are manipulative relations between variables which are nevertheless not causal. It is not surprising that the relation of a variable to itself is not causal! However, **L** and **μ** (whether general or specific) are not identical. First, as their definitions (1) and (2) above show, they don't have the same numerical value. Second, they do not even have the same dimension: **L** is measured in kg m^2/s^2, whereas **μ** is measured in C m^2/s^2. It makes no sense to suppose that something that is measured in units of weight (multiplied by the square of distance and divided by the square of time) is identical to something that is measured in units of charge (multiplied by the square of distance and divided by the square of time).

Here is a third, more fundamental, way of challenging the alleged counterexample to the interventionist analysis of causation. It may be said that it is illegitimate to consider either **F** as an intervention variable for **L** or **B** as an intervention variable for **μ**, because both **L** and μ are functions of another variable, which is not represented in fig. 1: the speed **v** of the rotating object. The real causal structure is represented by fig. 7.

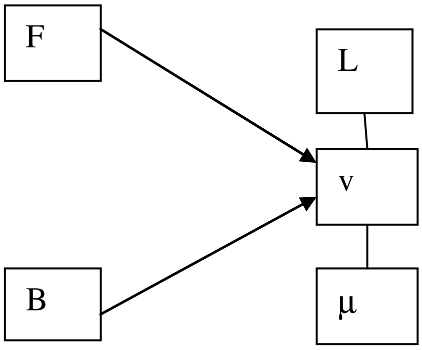


Fig. 7. Alternative graph representing the variables relevant for testing the magnetic stirrer law

Fig. 7 suggests that both **F** and **B** are intervention variables with respect to **v**, rather than with respect to either **L** or **μ**. If it is impossible to intervene either on **L** or on **μ**, testing the magnetic stirrer law cannot yield a counterexample to the interventionist analysis. The point of this objection is not that **F** is not an intervention variable on **L** (nor **B** an intervention variable on **μ**) because manipulating **F** cannot change **L** without changing **v** (and because manipulating **B** cannot change **μ** without changing **v**). Indeed, for the same reasons (given in my reply to the first objection) for which it is not appropriate to require that **μ** should be held fixed during an intervention on **L**, it would be inappropriate to require that **v** must be held fixed during an intervention on **L**. Rather, the objection is this. Given that **F** is always followed by changes both of **L** and of **v**, it must be determined by scientific reasons whether **F** is an intervention on **v** or on **L**. Now, from a physical point of view, it seems clear that **F** directly changes **v**, whereas **L** changes concomitantly with **v** because there is a law of nature that determines **L** as a function of **v** (and other variables). An analogous reasoning applies to **B**: **B** also directly intervenes on **v** rather than **μ**; **μ** varies together with **v** because there is a law according to which **μ** is determined by **v** (and other variables). Thus, the correct representation of the situation is given by Fig. 7: there are arrows from **F** to **v** and from **B** to **v**, but no arrow from **F** to **L** and no arrow from **B** to **μ**. If this is correct, the interventionist framework does not after all lead to the conclusion that there is a symmetric and mutual relation of manipulability between **L** and **μ**.

Two points in reply. First, the objection is beside the point: My thesis is that the interventionist conditions are compatible with the graph of fig. 1 although this has the intuitively incorrect result that **L** and **μ** are causes of each other. The present objection against my argument says that the situation represented in fig. 7 – which includes **v** as an additional variable – is no counterexample against the interventionist analysis. My point is that the interventionist analysis cannot rule out the false result of fig. 1; it is not relevant to point out against this thesis that there is a different representation of the situation (fig. 7), which does not raise the same problem for interventionism. Interventionism is supposed to analyze the reasons scientists have for taking certain relations between the variables used to represent a given system to be causal and others not. Such reasons must not presuppose – as the present objection does – that one already knows about the causal relations before choosing the variables.

Second, the problem I have raised still arises for the variables represented in the graph of fig. 2, insofar as the interventionist analysis does not rule out the relations between **v** and **L** and between **v** and **μ** from being causal, because it does not have any means to make the distinction between causal dependence and non-causal nomic dependence. The fact that **L** and **μ** are dependent on **v** by non-causal association laws is incompatible with the hypothesis that the relations of dependence between them are causal. My point is that the dependence between variables that is expressed by association laws is too tight to be causal. However, the conditions the interventionist analysis imposes on the relations between variables make no room for the distinction between nomic but non-causal dependence and causal dependence.

This is why the edges between **v** and **L** and between **v** and **μ** in fig. 7 are not directed. **L** depends, functionally or nomologically, on **v**. The fact that there is a structural equation **L= f(v)** might suggest that there should be an edge between **v** and **L**, directed towards **L**. But this would only be a good reason to take the edge to be directed if there were no structural equation suggesting the opposite direction. In this case, there is also a structural equation **v= f(L)** that shows that **L** is a function of **v**. The edge is not directed because there is functional and nomological dependence in both directions.[11]

The objection says that only some models containing variables related by relations of non-causal association laws lead to the absurd result that some pairs of variables are mutual causes of each other, and suggests that this result can be avoided by adding further variables to the set, in this case **v**. This seems to fit well with Woodward's reply to Strevens in their recent exchange on the question of whether the interventionist account of causation presented in (Woodward 2003) has the unintended consequence of making causal relations relative to sets of variables. Woodward (2008a) denies Strevens' (2007; 2009) claim that his (Woodward 2003) account makes causal relations relative to variable sets[12]. He points out that although his definitions of the concepts of direct cause and contributing cause make them relative to a variable set, there is also a nonrelative concept of contributing cause, according to which "X is a contributing cause *simpliciter* (in the sense that it isn't relativized to any particular variable set) as long as it is true that there exists a variable set **V**, such that **X** is correctly represented as a contributing cause of **Y** with respect to **V**" (Woodward 2008a, p. 209). According to the definition (M) in (Woodward 2003, p. 59), the notion of a contribution cause is more general than that of a direct cause: **X** can be a contributing cause of **Y** even if all paths linking **X** to **Y** run through intermediate variables $\mathbf{Z_1},\ldots,\mathbf{Z_n}$. Now, according to the definition of a nonrelativized contributing cause suggested in Woodward (2008a), **L** and **μ** are contributing causes of each other, because *there exists* a variable set relative to which **L** and **μ** are contributing causes of each other. As we have seen, {**F, B, L, μ**} is such a set; thus, **L** and **μ** are contributing causes of each other.

Here is still another way of defending interventionism. It might be argued that the variables appearing in fig. 1 and 7 are too closely related to play the role of candidates for causal relations. All of the variables in the set **S=**{**F, B, v, L, μ**} represent the state of the stirrer system. The state of the system has causes, such as someone's switching on the electric current creating the field **B**, and effects such as the mixing of the liquid in which the stirrer is immersed. However, it is wrong to apply the interventionist analysis to the variables *within* the set **S**. One way to make this idea precise uses the condition (IF) of independent fixability introduced in Woodward (2011).

"(IF): a set of variables V satisfies independent fixability of values if and only if for each value it is possible for a variable to take individually, it is possible (that is, possible in terms of their assumed definitional, logical, mathematical, or mereological relations or "metaphysically possible") to set the variable to that value via an intervention, concurrently

[11] I thank an anonymous referee for suggesting this way of arguing for the non-directedness of these edges.

[12] According to Woodward, what is relative to sets of variables, are causal *judgments*, insofar as such judgments depend on "what we regard as serious possibilities" (Woodward 2008a, p. 205).

with each of the other variables in V also being set to any of its individually possible values by independent interventions" (Woodward 2011, p. 12).

The advocate of interventionism may say that variables that do not satisfy (IF) are not even candidates for being causally related; thus the problem raised by the mutual relations between **L** and **μ** does not arise because the existence of a non-causal association law relating them makes it impossible to fix **L** and **μ** independently of each other. In reply, it is important to note that Woodward's definition of (IF) doesn't mention non-causal association laws. Strictly speaking, according to Woodward's formulation of (IF) quoted above, **L** and **μ** are *not* excluded from independent fixability because their correlation, and the impossibility to modify them independently of each other is not due to "definitional, logical, mathematical, or mereological relations" ; more generally, it is not a case of "metaphysical impossibility" (to modify variables independently of each other) in the sense in which these other relations give rise to such metaphysical impossibility.

But couldn't we simply modify Woodward's condition (IF) and include "non-causal nomological dependence" alongside with definitional, logical, mathematical and mereological relations in the set of relations that violate (IF)? Call this the enlarged condition of independent fixability (IF*). I think there are reasons for avoiding this move; maybe these reasons explain why Woodward has not mentioned association laws in his (2011) and in his formulation of (IF). The problem is that, from the interventionist perspective, it is not clear what distinguishes such non-causal association laws between variables from causal laws (or type-level causal relations) between variables. The main claim of the present paper is precisely that interventionism cannot make that distinction because non-causal association laws also ground manipulation relations. The problem for the defense of interventionism we are presently considering is that non-causal nomological association laws resemble causal laws more than the relations in the set mentioned in (IF), i.e. definitional, logical, mathematical, mereological relations and, more generally, relations of "metaphysical necessity". Unlike the latter, and like causal laws, nomological relations are empirical rather than logical, definitional or mathematical, and unlike the latter and like causal laws, they are generally taken to be metaphysically contingent. I would not put too much weight on the latter point though. It doesn't seem relevant that it is in some sense metaphysically possible (supposing that laws of nature are contingent) that **L** varies independently of **μ**. In the context of exploring causal and nomological relations, the relevant possibilities are always nomological possibilities; and it is *not nomologically* possible that **L** varies independently of **μ**. The difficulty for the present defense of interventionism is rather that the nature and strength of non-causal association laws seems to be of the same sort as the nature and strength of causal laws. They only appear to be different insofar as one considers complex situations which are incompletely described by a given set of variables. Causal relations between such variables can always be "interrupted" or "cut" thanks to influences grounded on factors that are not explicitly represented in the model. To see this, it may be useful to consider a system that is so simple that it is possible to represent it completely. In such a complete model, it is just as impossible to "independently fix" causally related variables as it is impossible to independently fix variables related by non-causal association laws. Consider a system composed just by a proton and an electron that is perfectly isolated from its surroundings and assume we have a complete model of this 2-particle system. In such a model, the causally related variables (the states of the particles) are not "independently fixable", because there is nothing within the system that would allow breaking these causal relations, and the system is by hypothesis completely isolated from outside influences. Thus, in this case, the enlarged (IF*)-condition would not be satisfied by causally related variables although there seems to be no reason why this system should not contain any causal relations.

Here is a straightforward way of defending the hypothesis that manipulability is sufficient for causation, even if laws of simultaneous association are taken into account. One might argue that there can be no "causal loop" from **L**(s,t) to **μ**(s,t) and back to **L**(s,t) because it is impossible to intervene both on **L**(s,t) and on **μ**(s,t) at the same time t in the same system s. Indeed, given that these variables stand in a relation of mutual functional dependence, one cannot independently fix both **L** and **μ**. However, in the framework of the interventionist account, in order to justify the claim that both **L**(s,t) causes **μ**(s,t) and **μ**(s,t) causes **L**(s,t), it is only required that for all s and t, it is both possible to intervene on **L**(s,t) and also possible to intervene on **μ**(s,t). The defense of interventionism under consideration fails by committing a fallacy relative to the scope of the possibility of manipulation.

To justify that **μ**(s,t) causes **L**(s,t) and **L**(s,t) causes **μ**(s,t), it is necessary and sufficient that: for all s and all t, it is possible to manipulate **L**(s,t) by intervening on **μ**(s,t), and it is possible to manipulate **μ**(s,t) by intervening on **L**(s,t). However, it is *not* necessary that: for all s and all t, it is possible to manipulate both **L**(s,t) by intervening on **μ**(s,t) and **μ**(s,t) by intervening on **L**(s,t). Only the latter condition cannot be satisfied because one cannot perform both manipulations on the same system at the same moment. The point can be made in terms of counterfactuals and possible worlds[13]. According to interventionism, **L**(s,t) causes **μ**(s,t) iff it is true that, if one changed **L**(s,t), **μ**(s,t) would change. This is true iff **μ**(s,t) changes in the closest possible worlds in which **L**(s,t) changes. In the same way, **μ**(s,t) causes **L**(s,t) iff it is true that, if one changed **μ**(s,t), **L**(s,t) would change; and this is true iff **L**(s,t) changes in the closest possible worlds in which **μ**(s,t) changes. Both claims can be true at the same time; they are made true by different possible worlds. In other words, for both claims (that **L**(s,t) causes **μ**(s,t) and that **μ**(s,t) causes **L**(s,t)) to be true it is not necessary to consider (im)possible worlds in which one changes both **L**(s,t) and **μ**(s,t).

A more radical move to defend interventionism in the same spirit would be to claim that the direction of causation is not objectively determined, but depends on perspective[14]. According to whether one *considers* an intervention on **L** or on **μ**, **μ** is cause of **L** or **L** is cause of **μ**. Then it seems to be enough not to take both perspectives at the same time to avoid the result that there is a causal loop at the level of specific variables. I take it to be unsatisfactory to consider the direction of causation as not objectively determinate. But this move cannot really rescue interventionism anyway. Let us assume that the direction of the causal relation between **L** and **μ** is determined by the direction a given cognitive agent considers. True, it is impossible to intervene on both **L** and **μ** in the same system at the same instant. However, nothing stands in the way of *considering* both interventions at the same time. First, there may be two agents each of whom considers one of the two directions. Second, a single agent may consider both directions at the same time. She can, e.g., draw two diagrams next to each other, one of which represents an intervention on **μ** and the other an intervention on **L**, and then consider both diagrams at the same time. Both possibilities show that it is possible to endorse both perspectives (the perspective according to which **L** causes **μ**, and the perspective according to which **μ** causes **L**) at the same time. Therefore, if the direction of a causal relation is determined by the perspective of a cognitive agent, causal relations in both directions can coexist at the same time. The refutation by reductio goes through as before: Given transitivity, each of the relata is its own cause.

Maybe the direction of the causal dependence between **L** and **μ** depends on the experimental setup, rather than on the cognitive state of an observer. In an experimental setup in which one intervenes on **L** (via **F**), **L** is the cause of **μ**, whereas in an experimental setup in

---

[13] Thanks to an anonymous referee for suggesting me this way of putting the point.
[14] I criticize this move, suggested by Fair (1979), in Kistler (2006).

which one intervenes on **μ** (via **B**), **μ** is the cause of **L**. This suggestion invites the same reply as the one we have just considered: Causal relations exist before and independently of their experimental discovery. The physical system of the magnetic stirrer may be experimentally explored in both ways. Thus, before one makes one of the two interventions (on **μ** via **B** and on **L** via **F**), it is both true that **L** causes **μ** because there is an experimental setup which allows manipulating **L** via **F** *and* that **μ** causes **L** because there is an experimental setup which allows manipulating **μ** via **B**.

Let me consider one last objection. My argument depends on the controversial thesis that causation is transitive. However, the chain **L – μ – L** may be a case where causation is not transitive. In order for this objection not to be question-begging, one would need an independent reason for thinking that transitivity does not apply to the chain **L – μ – L**. However, we have on the contrary some reason for thinking that it is no exception to transitivity: Most counterexamples to the transitivity of causation that can be found in the literature belong to one of two categories, but association laws belong to neither of them. In counterexamples of the first category, transitivity seems to be violated to the extent that it is left unspecified whether the items that are causally related are events or facts. The second category concerns cases where the cause or the effect (or both) is a *negative* fact; double prevention is a particular case[15].

The following case described by Ehring (1987) belongs to the first category. Smith puts potassium salts in the fireplace, making the fire in the fireplace purple. The fire then lights a log lying nearby. There is a causal chain from Smith's throwing potassium salts in the fireplace to the log's taking fire, but it seems wrong to say the former event is a cause of the latter[16]. One way to account for this scenario without abandoning the requirement of transitivity is to take the terms of causal relations to be facts (Kistler 2001) or aspects of events (Paul 2004). At the level of facts, there is no causal chain relating Smith's act to the log's inflammation, because there is no common middle term. The effect of the first causal process, the fire in the fireplace being purple, is not identical with the cause of the second process, the fire being hot. Without a causal chain, the question of transitivity does not even arise. There is an illusory appearance of a causal chain as long as one doesn't distinguish the fire's becoming purple from the fire's being hot.

The causal chain leading from **L**(s,t) to **μ**(s,t) and then again to **L**(s,t) does not belong to this category of counterexamples to transitivity. The appearance of a chain in the potassium salts case depends on the ambiguous specification of the middle term. Here, there is no such ambiguity. The middle term is exactly **μ**(s,t). Even if we consider **μ**(s,t) as a fact or an aspect of an event, we still get the result that **L**(s,t) (and **μ**(s,t) for that matter) causes itself.

The second category of counterexamples to transitivity involves causal relations where the cause or the effect is a negative fact. Let us consider a case of so-called double prevention that Hitchcock (2001) attributes to Ned Hall[17]. A hiker sees a rock falling, which makes her duck to avoid it. The fact that the hiker didn't get hurt by the rock makes her continue her trip. This is a case of double prevention, in the sense that the hiker's ducking prevents the falling of the rock from preventing her from continuing her trip. It seems wrong to say that the rock's falling causes the continuation of the trip, although there seems to be a causal chain linking the former, via the hiker's ducking, to the latter. As with the first category of counterexamples, it is possible to defend the transitivity of causation by denying that there is a causal chain. There are independent reasons for denying that negative facts, such as the fact

[15] See Bennett (1987), Hall (2004a).
[16] Other examples can be found in McDermott (1995), Hall (2000/2004b), Paul (2004).
[17] Hitchcock (2001, p. 276) indicates that it figures in an unpublished version of Hall (2004a).

that the hiker is not hit by the rock, can be causes or effects[18]. Negative facts enter into relations of explanation, which may be indirectly grounded on (and made true by) causal processes, but they are no terms of causal relations. However, the chain from **L**(s,t) to **μ**(s,t) and then back to **L**(s,t) does not belong to this second category either, simply because **L**(s,t) and **μ**(s,t) are no negative facts.

This suggests[19] that the relations between **L**(s,t) and **μ**(s,t) do not belong to any category of relations that give rise to the illusion of a causal chain. But then there really is a chain of determination by which each of the variables **L**(s,t) and **μ**(s,t) indirectly determines itself via the other variable and two instances of (MS). However, it is absurd that a specific variable causes itself. I conclude that the relation of determination expressed by an association law such as (MS) is not causation.

I do not deny that there are other counterexamples to the transitivity of causation that belong to neither of these two categories. Hall (2000/2004b) mentions the following situation in which a switch influences the route on which a causal influence travels between two variables. A railroad track branches in two tracks at point A, but the two tracks merge again into a single track at a second point B. Intuitions may diverge on the question whether the switching is a cause of the train's arrival at some point C beyond the merging point B. We need not take a stand on this question. It suffices to note that the relations between and **L**(s,t) and **μ**(s,t) do not belong to the category of switches, so that there is no non ad-hoc reason to deny that transitivity holds in their case.

Conclusion

I have raised a problem for Woodward's version of the interventionist analysis of causation. Given that association laws guarantee manipulability, I have argued that the existence of a manipulability relation between two specific variables X(s,t) and Y(s,t), i.e. the fact that an intervention on one variable can be used to manipulate the other, is not sufficient for the existence of a causal relation between these variables. The fact that X(s,t) makes a difference to Y(s,t) is not in itself sufficient for X(s,t) being a cause of Y(s,t). If these variables are linked by a law of functional dependence, they cannot be causally related because the opposite hypothesis (i.e. that manipulability is sufficient for causation, together with the assumption of the transitivity of causation), leads to the absurd result that X(s,t) and Y(s,t) are both causes of themselves. Furthermore, I have argued that the existence of a relation of mutual dependence between specific variables shows that the corresponding generic variables do not form a feedback cycle. It remains to be seen whether causation between *generic* variables can be construed in such a way that it may be true that generic variables related by an association law are nevertheless causally related or even mutual causes of each other.

The aim of this paper was mainly critical. However, let me end by suggesting that manipulability may provide a satisfactory analysis of nomic dependence rather than of causation. A relation between variables X(s,t) and Y(s*,t*) that satisfies interventionist criteria can be either a relation of non-causal nomic dependence or a relation of causal dependence. If this is correct, it opens up the possibility of building upon this analysis to construct a new analysis of causation. Let me sketch one possible way in which this might be done. If t < t*, manipulability cannot be mutual. In general, the variable X(s,t) may characterize a system s, whereas variable Y(s*,t*) may characterize a different system s*. If t

[18] I have argued for this claim elsewhere (Kistler 2006).

[19] It does not establish it. Maybe the relations in the chain **L** – **μ** – **L** are not transitive for some other reason. But I can't think of any such reason.

< t*, and if we assume that there is no backwards causation, it is not possible both that intervening on X(s,t) changes Y(s*,t*), and that intervening on Y(s*,t*) changes X(s,t). Thus, if intervening on X(s, t) changes Y(s*, t*), this may be sufficient for the existence of a relation of *causal* nomic dependence between X(s, t) and Y(s*, t*). One way to analyse the relation of causal nomic dependence is in terms of nomic dependence and of a causal process. This is not the place to go into the analysis of the notion of a causal process. Let us suppose processes are grounded on the transmission of an amount of energy or some other conserved quantity[20]. Then X(s,t) can be a cause of Y(s*,t*) if there is a causal process extending from system s at t to system s* at t*. A special case is when s* is identical to s, so that the causal process consists in the evolution of one system, which is characterized by variables X and Y related by a functional association law.

If the existence of a transmission process between s at t and s* at t* is what makes the relation between X(s, t) and Y(s*, t*) causal, it may be tempting to conclude that one need not include manipulability in the sufficient condition for causation, or in other words, that manipulability is redundant for this sufficient condition for causation[21]. However, insofar as what is at issue is the question whether one specific variable X(s,t) causes another specific variable Y(s*,t*), rather than whether there is a causal process relating system s at t to system s* at t*, the existence of a transmission process is not enough. My suggestion is that X(s,t) causes Y(s,t) iff 1) there is a causal process between system s at t and system s* at t* and 2) there is a relation of nomological dependence between the variables X and Y (both general and specific). I suggest that the interventionist analysis fits nomological dependence rather than causation itself.

In our example of the magnetic stirrer, let an intervention change the value of $\mathbf{L}$ at t, to set it to $\mathbf{L_1}$. The value of $\boldsymbol{\mu}$ changes simultaneously, becoming $\boldsymbol{\mu_1}$. Without further intervention, variables $\mathbf{L}$ and $\boldsymbol{\mu}$ still have the values $\mathbf{L_1}$ and $\boldsymbol{\mu_1}$ at a later time t*. The persistence of the rotating object from t to t* is a causal process. Therefore, one can justify the claim that setting $\mathbf{L}$ to $\mathbf{L_1}$ at t caused $\boldsymbol{\mu}$'s having the value $\boldsymbol{\mu_1}$ at t*, and also the claim that setting $\boldsymbol{\mu}$ to $\boldsymbol{\mu_1}$ at t makes a causal difference to $\mathbf{L}$'s having the value $\mathbf{L_1}$ at t*, in two steps.

1) One justifies the existence of a causal process linking the system at t to the system at t*. This may take the form of showing that the system at t is "genidentical"[22] to the system at t*. 2) One justifies the nomic dependence of $\boldsymbol{\mu}$ on $\mathbf{L}$, given that **r**, **m**, and **e** are held fixed by the existence of the simultaneous association law (MS) relating these variables.

What makes true a causal claim relating two specific variables has two components: the existence of a transmission process and the existence of a relation of nomological dependence between the two variables. Interventionism analyses nomological dependence, which is one of the components. In this paper, I have tried to establish this claim by showing that the interventionist condition can be satisfied in cases where this component, i.e. nomological dependence, comes alone, without causation[23].

References

- Carnap, R. (1928), *Der logische Aufbau der Welt*, Hamburg, Felix Meiner, 1998.
- Cartwright, N. (1999). *The Dappled World: A Study of the Boundaries of Science*.

[20] Kistler (1998; 2006).
[21] This conclusion has been suggested by an anonymous referee.
[22] Carnap (1928, § 159), Hawley (1999), Kistler (2001a).
[23] I thank Olivier Darrigol, Isabelle Drouet and Gernot Kleiter, as well as my auditors in Toulouse, Liblice, Salzburg and Konstanz, where I have presented earlier versions of this paper, for many helpful suggestions. The paper has been greatly improved thanks to the sharp and helpful critical remarks of two anonymous referees and of the editors of this special issue.

Cambridge: Cambridge University Press.
- Collins, J., Hall, N. and Paul L. A. (eds.), Causation and Counterfactuals. Cambridge (Massachusetts), MIT Press, 2004.
- Cummins, R. (2000), How Does it Work? Vs. What Are the Laws? Two Conceptions of Psychological Explanation, in F. Keil & R. Wilson (eds.), *Explanation and Cognition*, Cambridge, MA: MIT Press, p. 117-45.
- Ehring, D. (1987), Causal Relata, Synthese, 73, p. 319–328.
- Fair D. (1979), Causation and the Flow of Energy, *Erkenntnis* 14, p. 219-250.
- Hall, N. (2004a), Two Concepts of Causation, in Collins et al. (2004), p. 225-276.
- Hall, N. (2000/2004b), Causation and the Price of Transitivity, in Collins et al. (2004), p. 181-204.
- Hawley, K. (1999), Persistence and Non-supervenient Relations, Mind 108, pp. 53-67.
- Hitchcock, C. (2001), The Intransitivity of Causation Revealed in Equations and Graphs, Journal of Philosophy 98, p. 273-99.
- Kistler, M. (1998), Reducing Causality to Transmission, Erkenntnis 48, p. 1-24.
- Kistler, M. (2001), Causation as transference and responsibility, in Wolfgang Spohn, Marion Ledwig & Michael Esfeld (eds.), Current Issues in Causation, Paderborn, Mentis, p. 115-133.
- Kistler, M. (2001a), Le concept de génidentité chez Carnap et Russell, in: Sandra Laugier (ed.), Carnap et la construction logique du monde, Paris, Vrin, 2001, p. 163-188.
- Kistler, M. (2006), Causation and Laws of Nature, Londres, Routledge.
- McDermott, M. (1995), Redundant Causation, *British Journal for the Philosophy of Science* 46, p. 523-544.
- Paul, L.A. (2004), Aspect Causation, in Collins et al. (2004), p. 205-224.
- Pearl J. (2000), *Causality: Models, Reasoning, and Inference*, Cambridge University Press.
- Schurz, G. (2002), Ceteris paribus Laws: Classification and Deconstruction, *Erkenntnis* 57, p. 351-372.
- Spirtes, P., Glymour, C. and Scheines, R. (2000), *Causation, Prediction and Search*, Second edition, Cambridge (Mass.), MIT Press.
- Spirtes, P. and Scheines, R. (2004), Causal Inference of Ambiguous Manipulations, *Philosophy of Science*, 71, p. 833–845.
- Spohn, W. (2001a), Bayesian Nets Are All There is to Causal Dependence, in M.C. Galavotti et al (eds.), *Stochastic Causality*, Stanford: CSLI, and in W. Spohn, Causation, Coherence, and Concepts: A Collection of Essays, Springer 2009.
- Spohn, W. (2001b), Deterministic Causation, in Wolfgang Spohn, Marion Ledwig & Michael Esfeld (eds.), *Current Issues in Causation*, Paderborn, Mentis, 2001, pp. 21-46.
- Spohn, W. (2006), Causation: An Alternative, *British Journal for the Philosophy of Science*, 57, p. 93-119.
- Strevens, M. (2007), Review of Woodward, *Making Things Happen*, *Philosophy and Phenomenological Research*, 74, pp. 233-249.
- Strevens, M. (2008), Comments on Woodward, *Making Things Happen*, *Philosophy and Phenomenological Research*, 77, pp. 171-192.
- Woodward, J. (2003), *Making Things Happen*, Oxford University Press.
- Woodward, J. (2008), Mental Causation and Neural Mechanism, in J. Hohwy and J. Kallestrup, *Being Reduced*, Oxford, Oxford University Press, p. 218-262.
- Woodward, J. (2008a) Response to Strevens, *Philosophy and Phenomenological Research*, 77, pp. 193-212.

- Woodward, J. (2011), Interventionism and Causal Exclusion, http://philsci-archive.pitt.edu/8651.