REVUE
DE LA SOCIÉTÉ
DE PHILOSOPHIE
DES SCIENCES

Vol 1 N°1 2014

Louis Vervoort

The Manipulability Account of Causation Applied to Typical Physical Systems

SOCIÉTÉ DE PHILOSOPHIE DES SCIENCES (SPS)
École normale supérieure
45, rue d'Ulm
75005 Paris
www.sps-philoscience.org

REVUE
DE LA SOCIÉTÉ
DE PHILOSOPHIE
DES SCIENCES

Louis Vervoort

The Manipulability Account of Causation Applied to Typical Physical Systems

Sommaire

In the following we will apply the manipulability theory of causation of Woodward (2003) to physical systems, and show that, in the context of physical systems, the theory can be simplified. Elaborating on an argument by Cartwright, we will argue that the notions of 'modularity' and 'intervention' of the cited work should be adapted for typical physical systems, in order to take into account the coupling of system equations. We will show that this allows the reduction of all cause types discussed in Woodward (2003) to only one, namely that of 'total cause'. We therefore claim that the manipulability account can be drastically simplified when applied to coupled physical systems

Key words: cause, causation, manipulability, modularity, physical system, coupling

1 – Introduction

In the present article we will examine an influential theory of causation, namely the 'manipulability' account advanced by James Woodward. We will focus our attention on Woodward's well-known textbook (Woodward 2003). Woodward has authored or co-authored other texts on the topic, such as Hausman and Woodward 2004 (this was a reaction to Cartwright 2002); but these articles were written (at least in part) to support the full-blown theory of Woodward 2003. As is well known, the manipulability account was developed to apply to a wide spectrum of systems, from sociology, economy, physics etc. to everyday life. In the present article we will focus on systems studied in physics, and argue that the theory should be adapted in that case, in order to be fully compatible with the particular mathematics of typical physical systems. The singularity of the systems studied in physics is that they are describable by complete sets of equations involving *all* system variables - which is rarely possible in other contexts. More precisely, we will argue that the notions of 'modularity' and 'intervention', as they are defined in Woodward 2003, are incompatible with a feature of system equations termed 'coupling'. Coupling, *a property of most if not all realistic physical systems*, will be explained in Section 2. Its incompatibility with modularity and intervention will be explained in Section 3. In particular, we will be led to conclude that typical system equations from physics are not modular, but coupled.

Allowing for coupling has an unexpected effect for the manipulability theory of Woodward 2003: it allows us to simplify it greatly. Indeed, we will show (Section 4) that, for typical physical systems, of all the causal notions that are defined in Woodward 2003, only that of 'total cause' remains. 'Total cause' would therefore be, for such systems, nothing other than cause *simpliciter*.

Nancy Cartwright has extensively criticized the manipulability account of causation, and in particular modularity, using a series of arguments (see her 2006 and references therein). In one of the articles, namely in Cartwright 2004, she has identified the point we will investigate here in detail. She discusses a model of a carburetor, and comments (Cartwright 2004, 810): "By design the different causal laws are harnessed together and cannot be changed simply. So modularity fails." As will be detailed in the following, this 'harnessing of laws' corresponds to what will be termed here (following a physics and mathematics tradition) 'coupling of system equations'. Our aim is to study in detail *all* the consequences of this central concept, not only for modularity, but also for the definitions of cause proposed in Woodward 2003. Such a task is made possible thanks to the formal detail and logical coherence of the cited book by Woodward, which are just two of the merits

REVUE DE LA SOCIÉTÉ DE PHILOSOPHIE DES SCIENCES

of this text.

We said above that we will restrict our analysis to physical systems. But we would be more precise in stating that we will focus on systems described by coupled system equations. Physics doubtlessly provides the paradigmatic case, but there exists countless examples of coupled systems in other fields that can be mathematized (see Giordano and Weir 1988, Ch. 7 for a wide variety of examples from other fields). Let us also emphasize that the fact that the theory of Woodward 2003 should be adapted for coupled systems, does not necessarily imply that it must be modified for systems from econometrics, biology, everyday life etc.: in these cases system equations are most of the time decoupled.

2 – Physical Systems. Coupling of System Equations

In the following we will use the term 'physical systems' to denote systems of 'things' (entities and their interactions) *that are studied in physics*. Note that therefore not every 'physical system' in a broader sense will be in our focus. We suspect that Woodward's theory remains valid as it stands for such 'physical systems' *in a broader sense*, e.g. for systems studied in biology, economics, etc.. Let us start by making a few introductory observations linked to the notions of 'variables', 'constants' and 'equations'. Many of these remarks will be known to a majority of readers, especially those with a hands-on knowledge of physics, but they will allow us to define with clarity the concepts we need.

Physical systems are described in the practice of physics by sets of mathematical equations that quantify the relations that exist between the different parameters or *variables* of the system. In other words, for our present purposes 'systems' can roughly be equated with 'sets of (system) equations'; obviously, such equations are typical for the physical subfield or domain in question (e.g. relativistic mechanics, thermodynamics, quantum mechanics, etc.).

We will focus our attention in the following on physical systems that are realistic and therefore somewhat complex, namely systems that are described by several variables (not just one) *that intervene in general in several equations* (not just one) *describing the system*. This property, to be defined more precisely in the following, is called 'coupling' of the system equations. *Let us note from the start that a majority, if not all, of the systems studied in physics are 'coupled'*. Indeed, as a rule of thumb, if a system is characterized by N (dependent) variables, it must be described by N equations in order that the system has a unique solution to have a unique solution. These equations are coupled, else mathematical solution of the system would be much easier than it is: every equation could then be handled 'on its own' to be solved for one variable (see e.g. Boyce and DiPrima 1997, Ch. 7). That is hardly ever the case in physics.[1]

To illustrate the notion of coupling, let us briefly examine a pertinent example, namely Maxwell's equations of classical electromagnetism – pertinent, since describing a potentially infinite class of systems, belonging to such diverse subfields as electrostatics, magnetostatics, and electrodynamics. Maxwell's equations (Jackson 1999, Ch. 1) are a set of four 'coupled partial differential equations', expressing a relationship between the electric field ($\underline{E}$), the electric displacement field ($\underline{D}$), the magnetic field ($\underline{B}$), the magnetic field intensity ($\underline{H}$), and their so-called sources, namely the electric charge (ρ) and current density ($\underline{J}$):

$$\underline{\nabla} \times \underline{H} = (4\pi/c)\,\underline{J} + (1/c)\,\partial\underline{D}/\partial t$$
$$\underline{\nabla} \times E = -(1/c)\,\partial\underline{B}/\partial t$$
$$\underline{\nabla} \cdot \underline{B} = 0$$
$$\underline{\nabla} \cdot \underline{D} = 4\pi\rho$$

(The operator '$\underline{\nabla}$' effectuates derivation in the space coordinates, but that is a detail of no importance in the following.) In the above equations, 'c' is the speed of light. We explicitly display the equations, because they allow us to illustrate an essential point for the rest of the article. One notices that the set contains variables ($\underline{D}$ and $\underline{B}$) that belong to different equations; more precisely, *all equations contain variables that also belong to at least one other equation* – i.e., the equations are *coupled* (this is the more precise definition). The coupling makes mathematical solution difficult: none of the equations can be mathematically manipulated 'on its own' to derive a solution for a variable; they have to be treated in parallel.[2]

Another example refers to figure 1, which represents an electric circuit, involving currents I, I_1, I_2, I_3, an electric potential source V, and three constant resistors Ri (constants of the system). This system is fully described by following four relations: $I_1 = V/R_1$, $I_2 = V/R_2$, $I_3 = V/R_3$ and $I = I_1 + I_2 + I_3$. Here, too, the representative equations are coupled: the system variables appear in general in more than one equation (this is

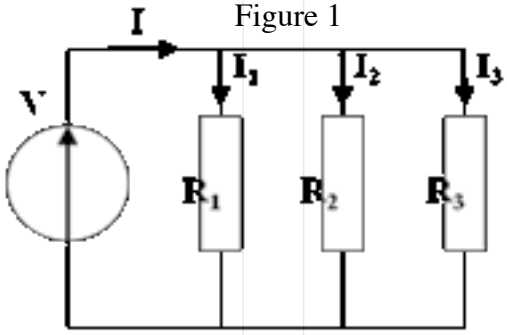
1 - The only exception to the ubiquitous coupling of equations we are aware of are (artificial) systems of non-interacting subsystems each described by only one dependent variable (say x), for instance a collection of N linear springs (evolving according to Hooke's law). One could describe this system by N equations $\{m_i.d^2x_i/dt^2 = -k_ix_i,\ i = 1,...,N\}$. (The k_i and m_i are constants of the system, not variables, and the time t is the independent variable.) This set of equations is decoupled: each dependent variable $x_1, x_2, \ldots x_N$ only intervenes in one equation. However, such a 'system' is highly untypical and in practice not studied in physics: all the relevant information is already contained in the first equation involving x_1. Solving the other equations is trivially redundant: they are the same.

2 - Let us note that the coupling is even more intricate than the Maxwell equations suggest, since in order to solve the set one needs additional equations, the so-called 'constitutive relations', depending on the properties of the environment, the media and materials involved in the particular system studied. One example of such an equation is $\underline{D} = \varepsilon.\underline{E}$, with ε the electric permittivity, a constant of the system. These auxiliary relations introduce additional coupling between the equations. But that only reinforces our argument.

REVUE
DE LA SOCIÉTÉ
DE PHILOSOPHIE
DES SCIENCES

the case for V, I_1, I_2, and I_3, but not for I); more precisely, all equations contain at least one variable that also appears in at least one other equation.

Figure 1



Let us repeat that only the simplest systems, those that involve just one dependent variable and that therefore can be described by one equation, evidently do not exhibit coupling. But such systems are a small minority of all physical systems studied; and indeed they are not treated by Woodward 2003 (the equations of the text always exhibit more than one dependent variable). As already stated, the modification of Woodward's theory we undertake in the following concerns only coupled systems - but all of them, whether from physics or other fields. Indeed, many real world systems that can be mathematized via equations are coupled. Physical systems are the most telling case, but see Giordano and Weir 1988, Ch. 7 for a series of systems from economy, biology, population dynamics, etc., all described by coupled equations.

For the purpose of the article it is important to remember that the role of the 'variables' and 'constants' in the system equations is clearly distinct – as will be obvious to many readers. For any given physical system, the constants are fixed, i.e. they keep the same numerical value independently of the values of the variables. In above examples, 'c' and the 'Ri' are such system constants: their value is assumed to be given. 'Variables' of course play a very different role: within one system with fixed constants, the variables may assume different values; the system equations numerically describe how a variation in one variable is linked ('*coupled*', one could say) to the variations of the other variables. For our purposes we cannot allow any doubt or imprecision to remain; therefore it is useful to return for a moment to the examples we gave above. If in electrodynamical systems, described by the Maxwell equations wecited, the source (or independent) variables $\underline{J}$ and ρ change (for instance due to human intervention, or due to changes in the natural environment), then the other (dependent) variables will also change (always for fixed constants). Similarly, if in Figure 1 the potential source doubles its output voltage V, then all currents will double their value (while the R_i remain constant), as immediately follows from the mathematics of the equations.

Let us now study in some detail the relevant parts of James Woodward's theory of manipulability.

3 – Coupling versus Modularity

Woodward 2003 heavily relies on the notions of 'modularity' and 'intervention' in constructing his theory of causation and explanation.[3] Our aim is to show that, in the case of physical systems, the definition of 'intervention' should be modified, and that 'modularity' does not hold for coupled systems. In Woodward 2003 modularity is defined as follows (quote 1):

> More generally, a system of equations will be modular if it is possible to disrupt or replace (the relationships represented by) any one of the equations in the system by means of an intervention on (the magnitude corresponding to) the dependent variable in that equation, without disrupting any of the other equations. (Woodward 2003, 48)

A similar definition is offered in Woodward 2003, 329. Observe that, a priori, the expression 'to disrupt an equation', intervening twice in above definition, could have different meanings. It could be interpreted as I1) 'to change the form of an equation', or as I2) 'to determine the numerical values of the variables of an equation' - or there may be other meanings. Under interpretation I1) modularity seems not to be a property of equations describing physical systems, because interventions, in the sense defined by Woodward below (and in the sense in which the word is used in physics, for that matter) do *not* change the form of an equation; rather, they set a variable in the equations to a given value (more on this below). More surprisingly, also under interpretation I2) of the term 'disrupt', modularity appears *not* to be a property of sets of equations, as soon as they are coupled, as will be explicitly shown below. But maybe we overlook valuable interpretations; let us therefore see how the concept is used. To that end we need the definition of 'intervention'[4] appearing in quote 1, involving the variables X and Y (quote 2):

> More generally and slightly more precisely, we may think of an intervention on X with respect to Y as an exogenous causal process that changes X in such a way and under conditions such that if any change occurs in Y, it occurs in virtue of Y's relationship to X and not in any other way. (Woodward 2003, 47)

The many examples offered in the text allow us to draw a clear picture of the precise meanings of above concepts. Let us study the most revealing one in detail. In Woodward (2003, 330) the following system of equations is discussed (we keep the original equation indices):

(7.4.3) $Y = aX + U$

3 - These notions have been discussed (and defined differently) by other philosophers, e.g. Spirtes et al. 2000 (see Woodward 2003, 46-48 for a complete list of references).
4 - The more detailed definition on p. 98 leads to exactly the same analysis as is presented here.

REVUE DE LA SOCIÉTÉ DE PHILOSOPHIE DES SCIENCES

(7.4.4) Z = bX + cY + V.

The equations – which are manifestly coupled - are rewritten as follows by algebraic rearrangement:

(7.4.3) Y = aX + U
(7.4.5) Z = dX + W, where d = b+ac and W = cU + V.

The author states (quote 3):

> Despite their observational equivalence, if (7.4.3) – (7.4.4) is modular, then (7.4.3) – (7.4.5) cannot be (and vice versa). To see this, consider an intervention on the variable Y in (7.4.3) that replaces (7.4.3) with the new equation (7.4.3*) Y = y. In effect, what this intervention does is to set the coefficient a in (7.4.3) equal to 0. (Woodward 2003, 330)

(The argument is that, if a = 0, d becomes equal to b, and thus (7.4.5) 'changes' (or is 'disrupted'), so that (7.4.3) – (7.4.5) is not 'modular', according to the definition of quote 1 we displayed above. It would seem here that the author uses interpretation I1) above.) Our point is that the phrases in quote 3 do not correspond to a legitimate interpretation of the mathematical apparatus used to model *coupled* physical systems. First, as highlighted in Section 2, a coefficient such as 'a' is, in such contexts, a constant of the system: its value cannot change, and in particular cannot become 0. What is possible, is that by intervention one does indeed set Y to a certain value, y. But then y = aX + U according to (7.4.3), implying a = (y-U)/X (instead of a = 0). The point to observe is that this intervention on Y will influence X *and* Z, whichever set of equivalent equations one starts from, *due to the fact that the equations are coupled*. Indeed, if the equations were not coupled (if they had no common variables), they could be mathematically handled in a fully separate manner, as if, so to speak, the other equation did not exist. This 'parallel' influence of Y on X and Z may be obvious if one considers (7.4.3) – (7.4.4) since Y intervenes in both equations; but it also follows from (7.4.3) – (7.4.5). If this is not obvious already, one could easily prove it explicitly (in a trivial manner). Whatever value Y takes, whether these values are set by intervention or not, according to (7.4.3) X can be rewritten as (Y – U) / a, so that (7.4.4) becomes, by substitution of X: Z = (c + b/a) Y + V – b U/a (7.4.4#). By inspection of equations (7.4.3) and the simultaneously valid (7.4.4#), one immediately sees that by intervening on Y (by causing it to assume the value y), one *necessarily* changes or influences the value of X *and* of Z. Again, the reason lies in the coupling of the equations. This result holds independently of the representation one chooses for the system. Indeed, equations (7.4.3) – (7.4.5) lead (of course) to the same equation (7.4.4#) (one could say, based on (7.4.3) – (7.4.5), that Y influences Z through its link with X; *X 'couples' Y and Z*). This is a mathematically necessary property of sets of equations that only differ by algebraic rearrangements, such as sets (7.4.3) – (7.4.4) and (7.4.3) – (7.4.5). There is no mathematical or physical information that can differ in both representations. We arrive here at a conclusion that differs from Woodward's, since he claims (quote 4):

> By contrast, according to (7.4.3) – (7.4.5), Y does not cause Z and hence there are no interventions on Y that will change Z. (Recall that an intervention on Y should be uncorrelated with other causes of Y such as X.) (Woodward 2003, 331)

The phrase between brackets points out what is the central issue. The author stipulates that under 'interventions' one variable (Y) can be set in such a manner that another variable (Z) is determined in a fully unrelated, '*decoupled*' manner. That may be an appropriate characterization for systems from economy, sociology etc., but such interventions do not exist in systems as soon as coupling between the equations occurs. We have explicitly proven this fact via our derivation of equation (7.4.4#) from (7.4.3) – (7.4.5): Y does cause Z, and this fact holds irrespectively of the equation set one starts from. For the system discussed, *all variables are linked to each other via coupled system equations* (coupling between equations induces coupling between variables). Clearly, this property can be generalized to all coupled systems (for those who prefer an explicit argument, see footnote 7).

Thus it appears that in coupled systems 'modularity' does not apply, and that the unwarranted use of the concept can be traced back to an 'unphysical' use of 'intervention' – more precisely a usage that does not apply to coupled systems, and in particular not the example just discussed. Note that this particular interpretation of the mathematical representation of intervention is not yet obvious in the definition of the concept as reproduced in quote 2. There is however a shift in the meaning of 'intervention' throughout the text, as was already noticed by Cartwright (2002, 414-15).[5] Careful inspection of the text shows that the same use of 'intervention', violating coupling, appears in other passages,[6] and, most importantly, in the definitions of cause, as we will show in the next Section. Therefore, and for reasons that will become clear in the following, we will arrive at the conclusion it is this element of the manipulability account of causation that should be modified to make it generally applicable to physical systems.

An important corollary of above findings is the following. Our analysis points to the idea that systems represented by coupled equations should not be conceived of as being made up of variables that are directly linked to only a subset of the other variables, but to *all* of the other variables. Indeed, from the definition of coupling, it follows that no variable can be solved for by using just one equation: one needs all equations. But that means that a change (e.g. under an interven-

5 - In the detailed definition of intervention (Woodward 2003, 98), the incompatibility with coupling is manifestly built-in.

6 - See e.g. Woodward 2003 p. 52, p. 100.

REVUE DE LA SOCIÉTÉ DE PHILOSOPHIE DES SCIENCES

tion) in one variable 'propagates' itself to all other variables.[7] In other words, all variables are linked to all other variables. Now, this has essential implications for the 'graph' representation that is extensively used in Woodward 2003 to depict and characterize causal networks – i.e. networks representing 'which variable influences which other variable'. Indeed, it would seem that, for physical (in general coupled) systems, *the correct representation is one in which each vertex is connected to each other vertex*. Note that the graph representation of Woodward 2003 is only based on the equations connecting the variables X, Y, Z,..., not on any time-relation between those variables: see how the notion of graph is introduced on p. 38ff., and examples such as Woodward 2003 p. 49, p. 51, p. 328 and others. But as we have argued, due to the coupling of the equations there is no privileged connection or direction in the graph (when one takes the equations alone into account): all variables are symmetrically (and therefore maximally) connected. (As briefly argued below, we believe a causal directionality *can* be recovered by taking the time parameter into account, but time directionality is certainly not implicit in physical system equations, which are well known to be invariant under time reversal. In real systems, the direction of time is imposed by the boundary conditions (of the time coordinate), depending on the real-world contingency of the system under study. For instance, in the current circuit of Fig. 1, in one experiment the potential V might first be turned on – by a button - and therefore *subsequently* induce the currents; in another experiment the time sequence might be reversed, when the currents are first generated, for instance by a moving magnet, generating a potential V (the equations linking the potential and the currents remain of course identical). Therefore a graph can only be given a directionality by placing, 'a posteriori' so to say, time arrows on it, indicating which parameter (node) of each branch takes on its value first. In sum, in the case of a typical physical system the directionality does not automatically follow from the equations – an idea that seems not always well represented in Woodward 2003, if the theory wants to include coupled systems.). We will not elaborate here on this idea, but this conclusion of a maximally interconnected network of system variables seems inescapable in the case of coupled equations. Since we have not presented a detailed proof of this thesis (it is not essential for our argument), let us introduce it as a conjecture, which we will call in the following the 'dense grid conjecture'.

Notice that the graphs of Woodward 2003 are defined in a different manner than in e.g., Spirtes, Glymour, and Scheines 2000, or in Pearl 2009. In these works, graphs are attached to *probabilistic systems and are not directly derived from deterministic and coupled system equations*. Such graphs can be 'directed', again if one knows causes and effects, i.e. what comes first. (But we conjecture that even such graphs for probabilistic systems will be maximally dense as soon as they represent coupled system equations, for the reasons indicated above.)

In conclusion, following Cartwright 2002, 2006, we believe modularity as defined in Woodward 2003 is not a property of (most) physical systems, in any case not of coupled systems. We have argued in detail that the concept of 'intervention' as employed in the definition of modularity is incompatible with the coupling that exists between the equations of the systems studied in physics. This conclusion may be seen as an elaboration of a claim of Cartwright (2004, 810, where the 'harnessing of laws' is discussed). We believe it is therefore safe to stick to the meaning of 'intervention' *as understood in many passages in Woodward* 2003, and in physics, as based on the idea that an intervention sets the value of a variable to a given value. *One cannot ask more of an intervention, if one does not want to violate the property of coupling*. (So in other words, even if a broader definition of 'intervention' may seem logically possible at first sight, it is not mathematically sound when applied to physical system equations.) Also, with this interpretation of intervention, the only case in which equations can be called 'modular' is that of 'decoupled' equations in the usual mathematical sense. Only if two equations are decoupled, if they share no common variables, can an intervention set the value of a variable in one equation without changing the value of a different variable appearing in the other equation. This may be the case in, for instance, econometrics.

4 – Implications for the Manipulationist Concept of Cause

In the present Section we will investigate in detail the consequences of abandoning modularity for the theory of causation that is proposed in Woodward 2003. As is well known, the theory strongly links modularity to the different notions of 'cause' it defines. One reads:

> (...) I will assume that when causal relationships are correctly and fully represented by systems of equations, each equation will correspond to a distinct causal mechanism and that the equation system will be modular. As we will see, the notion of direct causation and the related notion of causal route is closely bound up with these ideas. (Woodward 2003, 49)

Closer inspection of the text reveals that, according to the author, a set of equations that allows us to identify 'direct' causes is a 'modular' set, representing 'distinct causal mecha-

7 - Suppose one changes variable x_1 in equation e_1; in a coupled system e1 will also contain, by definition, at least one other variable, say x_2 (if not, x_1 will intervene in a second equation containing a second variable x_2); therefore a change in x_1 will change x_2. If there are three variables, x_3 necessarily intervenes in an equation that contains, due to coupling, at least one other variable, i.e. x_1 or x_2; so a change in x1 causes a change in x_2 and x_3; and so on if there are more variables. This can easily be seen in the system equations we exhibited as examples in the previous Section.

REVUE DE LA SOCIÉTÉ DE PHILOSOPHIE DES SCIENCES

nisms'. But we have just argued that 'modularity' is not an appropriate description of most or all physical systems. As we will explain in the following, it is then intuitively compelling that *abandoning the concept of modularity leads to abandoning the distinction between the different types of causes that are introduced in Woodward 2003, namely 'direct', 'contributing', and 'total' causes*. If that were to be the case, our investigation would lead, for physical and coupled systems in general, to a substantial simplification of the theory. Indeed, especially in Chapter 2 of the book, each of above cause types is given a separate definition, which is striking in its logical detail (see especially Woodward 2003, 51 – 61). In view of the importance of above conclusion, let us proceed carefully to prove it, by two different (and independent) routes.

The concept of 'direct cause' (DC) is defined as follows (our italics):

> (DC) 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, 55)

The italicized part shows that the definition of direct cause directly relies on the concept of intervention in its 'decoupled' form, i.e. conflicting with the mathematics of coupled equations. Such a definition may be logically consistent 'on its own'; it is however not synthetically applicable to real coupled systems – it does not describe such systems. As we have studied in detail in the preceding Section, a real intervention on a coupled system will influence all variables, so that 'all other variables besides X and Y' *cannot* 'be held fixed at some value'. Therefore 'direct cause' cannot be a legitimate concept for coupled systems. Since 'contributing cause' is introduced as the complement of 'direct cause' (any cause is either a direct or a contributing cause),[8] the former is likely to be not a good concept either, in the context considered. Indeed, the definition of 'contributing cause' in (Woodward 2003, 57)[9] explicitly uses the concept of 'direct cause' (and therefore intervention), which we have shown to be problematic. Therefore, for coupled systems 'contributing causes' do not exist either - as defined.

This argument suffices as a proof, but in view of the importance of the conclusion, it is instructive to give a second, *independent* argument for the idea that the proposed distinction between direct and contributing causes is inadequate for coupled systems.[10] Since this argument is based on the 'dense grid conjecture' of the previous Section, it may be found less compelling by some than our first argument. But let us emphasize that this second proof is fully redundant, in view of the first one. As well as on the concept of direct cause, the definition, then, of contributing cause (Woodward 2003, 57) explicitly relies on the notions of 'causal path' and of 'causal chain' (see e.g. the definition in our footnote 9). Causal paths and chains are parts of graphs. It is essential for our second proof to recall that Woodward 2003 considers graphs as fully derivable from the system equations alone, *not* from any other information related to boundary conditions in the time domain (remember how graphs are introduced on p. 38ff. and used in examples such as Woodward 2003 p. 49, p. 51, p. 328 etc.). The strategy of Woodward 2003 is to derive graphs from systems equations, and causes from graphs (or equations), *without using any other information*. Now, a rough restatement of the definitions of cause could be as follows: either a cause is 'direct' either 'contributing' if it is part of a causal chain in a graph (see the last phrase of the definition). However intuitive this may be, we have argued in Section 3 that the image of a causal chain is not appropriate for coupled systems (at least not if these chains are to be derived from the equations alone, as is Woodward's premise). Indeed, we have argued that the adequate graphs as derived from the system equations alone are maximally dense grids – maximally interconnected networks, connecting all vertices to all vertices. In this sense, before considering time, there are no chains: geometrically all variables play a perfectly symmetric role. If that is true, our point follows immediately; the distinction between types of causes, as *defined*, vanishes trivially. Since we have not elaborated these ideas in detail, we submit our second proof as a conjecture; but it corroborates our first proof, which seems compelling.

Let us stress that the truly problematic notions for coupled systems are 'modularity' and 'intervention', *not* graph. With sufficient care, there seems little doubt that the notion of 'directed graph' can be consistently defined also for coupled systems (as is clear by now, we believe this should be done by simply superimposing time arrows on the dense grid; these time arrows follow from the contingent boundary conditions in the time domain, see the examples given above). Therefore

8 - See the last sentence in the definition (NC) in Woodward 2003, 57.*

9 - Here is the first part of the definition of contributing cause (Woodward 2003, 57, our italics): "(NC) If X is a contributing type-level cause of Y with respect to the variable set V, then there is a directed path from X to Y such that each link in this path is a* direct causal relationship*; that is, there are intermediate variables along this path, Z1...Zn, such that X is a* direct cause *of Z1, which is..." (This is the necessary condition reappearing in the full definition on p. 59.)*

10 - A third argument is the following: we believe that the original intuition of the author (see p. 50) to introduce the notion of 'contributing cause' is questionable in the case of physical systems. On p. 50 an example of a system is discussed, for which the constants (a, b, and c) assume values such that a = -bc, implying that one of the variables (Y) becomes 0, so that X would not have a direct influence on Y. We cannot elaborate our arguments, but we believe one cannot infer causal relations by giving ad hoc values to the constants (see also the discussion following quote 3). In contrast to variables, system constants have fixed numerical values. Even if for a very peculiar system (a real situation is not provided) they were such that a = -bc, then the original equations still hold, and allow us to infer the causes without accounting for the information a = -bc. (Also, observe that in real systems a will never be exactly *equal to –bc.)*

one could presumably recover good definitions of e.g. direct cause, *but without using the notion of intervention*.[11]

Also, a concept as 'direct cause' as defined in works as Suppes 1970, Spirtes et al. 2000, and Pearl 2009, may very well be legitimate: the systems considered in these works are probabilistic and decoupled. Let us note, in passing, that it is highly relevant that Woodward explicitly states (Woodward 2003, 39) that the works of Pearl especially were his direct source of inspiration, and that these works use models inspired by econometrics (p. 39). We hope it is obvious by now that the mathematics of equations from econometrics (as from biology, sociology etc.) differ in at least one essential respect from typical physics equations.

If the notions of direct and contributing cause do not apply, as defined, to coupled systems, the only remaining cause type for such systems in the manipulability theory of Woodward 2003 is that of 'total cause'. It is defined as follows:

> (TC) X is a total cause of Y if and only if there is a possible intervention on X that will change Y or the probability distribution of Y. (Woodward 2003, 51)

Without surprise, the definition of total cause is the only one that accepts the ordinary meaning of 'intervention on X', *as an operation that sets the numerical value of X*. As we have shown, the other definitions of cause types are based on a conception of intervention that is illegitimate for coupled systems. We are thus led to the conclusion that the above definition of 'total cause' is the only remaining valid definition of cause in Woodward 2003 for most physical systems. We believe it therefore captures the essence of the notion of causation in Woodward 2003 for such systems. As we will show elsewhere, such a simplified notion of cause is indeed in perfect agreement with the use the physics community makes of the concept of cause.

5 – Conclusion

We began this paper with a study of Woodward's notions of modularity and intervention (in his 2003), which play an essential role in the manipulability theory of causation and explanation. We corroborated an earlier argument of Cartwright (2004, 810), and showed that these concepts are in conflict with a ubiquitous property of physical systems, namely the coupling between system equations (and therefore between system variables). Next, we showed that modifying the manipulability account of causation in order to take coupling into account has drastic consequences: it substantially simplifies the theory, leaving - for coupled systems - one generally applicable notion of cause, namely that of 'total cause'. For such coupled systems direct and indirect causes should be defined in a different manner than in Woodward 2003, presumably by referring to time sequences.

Throughout the text we emphasized that our conclusions hold for typical systems studied in physics, more precisely for systems characterized by coupled equations. But let us bear in mind that Woodward's rich and interesting theory may remain valid as it stands for other systems, such as 'physical systems' in a broader sense, systems studied in biology, economics, and other fields.

ACKNOWLEDGEMENTS

I would like to acknowledge detailed discussions of the present article with Frédéric Bouchard, Yvon Gauthier, Jean-Pierre Marquis; with the students of the philosophy of science seminar of the Université de Montréal; and with participants at the bi-annual congrès of the Société de Philosophie des Sciences, Montréal 2012. Research was funded by a grant of the Philosophy Department of the Université de Montréal ("Bourse d'Excellence du Département").

11 - The idea of causal chain is indeed intuitive – think of the paradigmatic example of colliding billiard balls. Here it seems one could easily introduce the notion of 'direct' and 'indirect' cause (some balls directly cause a given ball to move, others indirectly). As stated, we believe this could be done if one introduces time *into the picture, but not by means of the definition of 'contributing cause' provided in Woodward 2003, essentially because it relies on the definition of 'direct cause', and thus on a problematic use of the concept of intervention.*

REVUE
DE LA SOCIÉTÉ
DE PHILOSOPHIE
DES SCIENCES

The Manipulability Account of Causation Applied to Typical Physical Systems

REFERENCES

BOYCE, William, DIPRIMA, Richard. 1997. *Elementary Differential Equations*. New York: John Wiley.
CARTWRIGHT, Nancy. 2002. Against Modularity, the Causal Markov Condition and any Link between the Two. *British Journal for the Philosophy of Science* 53, 411 - 453.
CARTWRIGHT, Nancy. 2004. Causation: One Word, Many Things. *Philosophy of Science* 71, 805 - 819.
CARTWRIGHT, Nancy. 2006. From Metaphysics to Method: Comments on Manipulability and the Causal Markov Condition." *British Journal for the Philosophy of Science* 57, 197 - 218.
GIORDANO, Frank, WEIR, Maurice. 1988. *Differential Equations. A Modeling Approach*. Reading: Addison-Wesley Publishing Company.
HAUSMAN, Daniel, WOODWARD, James. 2004. Modularity and the Causal Markov Condition: A Restatement. *British Journal for the Philosophy of Science* 55, 147 – 161.
JACKSON, John. 1999. *Classical Electrodynamics*. New York: John Wiley.
PEARL, Judea. 2009. *Causality. Models, Reasoning, and Inference*, 2nd Ed. New York: Cambridge University Press.
SPIRTES, Peter, Clark GLYMOUR, and Richard SCHEINES. 2000. *Causation, Prediction, and Search*, 2nd Ed. Cambridge, MA: MIT Press.
SUPPES, Patrick. 1970. *A Probabilistic Theory of Causality*. Amsterdam: North-Holland Publishing Company.
WOODWARD, James. 2003. *Making Things Happen. A Theory of Causal Explanation*. Oxford: Oxford University Press.

HISTORIQUE
Article initialement soumis le 6 juillet 2012.

SITE WEB DE LA REVUE
sites.uclouvain.be/latosensu/index.php/latosensu/index

ISSN 2295-8029

CONTACT ET COORDONNÉES :
Louis Vervoort,
Centre Interuniversitaire de Recherche sur la Science et la Technologie,
University of Quebec at Montreal (UQAM),
Bur. N-8310, Pavillon N,
1205, rue Saint-Denis,
H2X 3R9 Montréal, Québec
Canada
vervoort.louis@courrier.uqam.ca
louisvervoort@hotmail.com

SOCIÉTÉ DE PHILOSOPHIE DES SCIENCES (SPS)
École normale supérieure
45, rue d'Ulm
75005 Paris
www.sps-philoscience.org