November 15. 2010

Inferring Causality through Counterfactuals in Observational Studies
Some epistemological issues

Federica RUSSO[a], Guillaume WUNSCH[b] and Michel MOUCHART[c]
a Philosophy, University of Kent, UK
b Demography, University of Louvain (UCLouvain), Belgium
c Statistics, University of Louvain (UCLouvain), Belgium

Abstract
This paper contributes to the debate on the virtues and vices of counterfactuals as a basis for causal inference. The goal is to put the counterfactual approach in an epistemological perspective. We discuss a number of issues, ranging from its non-observable basis to the parallelisms drawn between the counterfactual approach in statistics and in philosophy. We argue that the question is not to oppose or to endorse the counterfactual approach as a matter of principle, but to decide what modelling framework is best to adopt depending on the research context.

1. Introduction and Background

Arguably, there are two reasons why causal analysis is important in science as well as in everyday life. One is that if we know the causes we are more likely to provide a better explanation and understanding of a given phenomenon. The other is that if we know the causes, we are more likely to take better action or intervention, that is to design, for example, more efficient social or public health policies or to advise on individual treatments.

Controversial as it may sound at first glance, there is a sense in which causal inference is an easy task: it suffices to consider idealised situations. Identify the putative causes and effects, manipulate the causes holding fixed anything else, and see what happens. This is, in essence, the pillar of Baconian science. Without going into the historical details of the revolution Francis Bacon made in scientific method, the reader may like to be reminded that, with Bacon, science becomes a *scientia operativa* (Klein 2008 and 2009): to get to know about the world the scientist does not just passively observe it, but s/he interacts with it. The modern scientist is a "maker" (Ducheyne 2005): s/he performs experiments, that is s/he actively manipulates factors to find out what causes what.

But as science has evolved, methods have become more sophisticated too. A powerful tool introduced by Fisher in the early 1920s is *randomisation*. For the sake of history, the first historically recognised randomised experiment was run by Peirce and Jastrow (1885) in psychometrics, but randomisation had to wait nearly 50 years to receive an adequate conceptualisation and discussion (on this point see for instance Rescher (1978) and Hall (2007)).

Randomisation, in the original thought of Fisher, is a means for eliminating bias in the results due to *uncontrolled* differences in experimental conditions. Whilst we know that in laboratory experiments ideal conditions are more often met because uncontrolled variations in the

environment are much better known, this is certainly not the case in agricultural studies where Fisherian randomisation originated, nor in social and biomedical contexts where phenomena and environmental conditions are highly complex. Randomisation is somehow a heir of Baconian science because it ultimately aims to make causal inference reliable implementing the same ideas holding up the Baconian method: manipulation and control. Randomness, in fact, increases the efficiency of the experiments in the sense that, because unwanted sources of variation are controlled for, the sought level of significance is achieved in fewer trials. Also, by ensuring that unwanted sources of variation are minimised or even eliminated, randomisation ensures that only the cause is manipulated. (See Fisher 1925 and 1935 for the original formulation of randomisation in experimental design, and Rescher 1978, Hacking 1988 and Hall 2007 for historical reconstructions and critical appraisals of the meaning and development of Fisherian randomisation.)

However, as it happens, most studies in the social sciences are constructed on the basis of observational data and not experimental ones. The reason is that randomisation is often unethical or simply not feasible. This makes the reliability of observational studies a real challenge because not only human populations are highly heterogeneous both with respect to know/unknown and non-observable/non-observed factors, but also because, if randomisation cannot be performed, there is less grip on the sources of 'unwanted variations'—as Fisher called them—and on the mechanisms of assignment.

Here is an example that illustrates some difficulties related to heterogeneity; it stems from the work of Tietze, Potter, Henry, and others in the 1970s. In developed societies, women using contraceptives may have a higher fertility than non-contracepting women of fertile ages. This paradoxical result has been explained by the fact that many non-contracepting women are probably sterile or sub-fecund and therefore do not have recourse to contraception, in order to conceive. The measure of the use-effectiveness of methods of fertility control must therefore take the heterogeneity of the fecundity of the population into account, *e.g*. by comparing the fertility of current contraceptors to that of contraceptors who stop using birth control in order to conceive. The groups one compares should therefore be as similar as possible, except for the fact that one group experiences the putative cause and the other does not. The best situation would then be the following: to compare, at the same time, the outcome in the group experiencing the treatment to the outcome in the *same* group not taking the treatment. In this case, the two groups would indeed be perfectly identical, except for the fact that one experiences the cause and the other not.

Needless to say, it is not possible that the *same* individuals take and do not take treatment *at the same time*. But this practical difficulty does not prevent us from *imagining* what would happen if the same individuals did take and did not take the treatment. It is this way of reasoning that led Donald Rubin (1974) to develop a counterfactual framework of causality that we will briefly present in section 2. *Rubin's Causal Model*, as it is now called (Holland, 1986), has become a standard reference in the literature on causality.

The strength of the counterfactual approach seems to lie in the attempt to implement the pillars of Baconian science—that is those principles that most ensure the reliability of causal inference: manipulation and control. On the one hand, if the two groups (actually, the *same* group) only differ as to whether individuals receive the treatment or not, then the action of possible confounders is minimised if not nullified. On the other hand, once we hold fix everything else, the only factor subject to manipulation—albeit *ideal* manipulation—is the putative cause. Because in social science it is not always possible to manipulate or randomise,

the counterfactual framework apparently comes to rescue because it somewhat implements the same ideas of Baconian science—namely manipulation and control—*without* requiring *actual* manipulation. However, the counterfactual approach has its share of problems too, highlighted both by scientists and by philosophers (see later section 3).

This paper adds to the debates on the virtues and vices of counterfactuals, but does not aim to take definite side with the camp of the counterfactualists or the camp of the anti-counterfactualists. The general goal of the paper is to put the counterfactual approach in perspective, from an epistemological point of view. The starting point of the paper is based on two main ideas. Firstly, there is no unanimous consensus on a unique concept of causality: it is the task of the scientist to construct an epistemologically sound concept. When a suggested concept of causality cannot cope with an important and relevant scientific issue, the concept should be amended. Secondly, we believe that the root of the difficulty, when building a concept of causality in social sciences, lies in the intertwined issue that individuals are heterogeneous and that the latter may be due to differences in potentially causal variables. This is so because social processes are typically complex, involving multiple causes- multiple effects mechanisms.

Our position, that we shall develop and articulate in sections 3 and 4, can be summarised as follows. The question is not to oppose the counterfactual approach, randomisation or manipulation *as a matter of principle*. The possibility and, consequently, the decision to use manipulation or counterfactuals or to randomise in a given study depends on *practical* aspects such as the kind of data (for instance experimental or observational) the scientist has access to. On a more epistemological tone, our view is that the concept of causality should not necessarily rely on the concepts of counterfactuality or of manipulability. This, as we shall explain in more detail later, is for several reasons. One reason is that that there may be concepts other than (or in addition to) counterfactuality and manipulation to be used in the explication of the concept of causality. Another reason is that the counterfactual approach should be viewed as one among various possible methods to search for effects of causes and that there is no principled reason why it should necessarily be involved. Actually, causal analysis encompasses many more methods and approaches than just counterfactual models or randomised trials. This is not just a contingency due to the richness of scientific methodology, but it is also due to the fact that one may need different causal methods depending on whether the goal is to explain a phenomenon, to measure effects of known causes, to take action in response to the causal knowledge gathered, to assign causes to observed outcomes, etc.

The paper is organised as follows. In section 2 we recall the main features of the counterfactual approach. In section 3 we discuss six issues concerning the counterfactual approach. The first two issues have already been widely discussed in the literature. One concerns the soundness of the counterfactual approach: some authors have in fact argued that because the *counter*-fact is not observed, this jeopardises its empirical basis. The other concerns the problem of preferring the counterfactual model because it measures effects of causes over alternative models that instead search for causes of effects. The second two issues have to do with the concepts that back up the experimental method: manipulation and randomization. The third and last two issues concern, first, the fact that we are dealing with complex mechanisms and, second, the analogies and parallelisms that have oft been made between the counterfactual model developed in statistics and the counterfactual analysis of causation developed by philosophers. In section 4, we present an application and we assess the extent to which a counterfactual approach enables to answer a given research question. In the final section devoted to discussion and conclusion, after recalling the problems raised by

the counterfactual approach, we go back to the issue of whether alternative frameworks may supply the difficulties encountered by counterfactual models.

2. Counterfactuals and potential outcomes

Consider the classic case of a person who receives a treatment at time t. Simply put, the outcome or response to the treatment is observed at time t + k (k > 0). How does one conclude that the treatment is effective or not? In other words, how do we measure the possible causal effect of the treatment? Donald Rubin's answer to estimating the causal effect of treatments in randomised and nonrandomised studies is based on a counterfactual statement or 'What-if?' question. Philosophers and logicians define counterfactuals as subjunctive conditional statements, the antecedent of which states a contrary-to-fact situation. Consider the aspirin example given also by Rubin (1974). Suppose Mr Jones, suffering from headache, says that "If an hour ago I had taken two aspirins instead of just a glass of water, my headache would now be gone". This conditional statement presupposes that Mr Jones did not take the aspirin and still has headache. Instead, had he taken the aspirins he wouldn't have a headache anymore, and this is why, roughly speaking, we say that aspirin is an effective treatment against headaches.

A number of philosophers have argued, in slightly different ways, that the notion of counterfactuals captures an essential aspect of causation; for a brief overview, see *e.g.* P. Menzies (2009). In philosophy, a full counterfactual account of causation has been developed in the Seventies by David Lewis (Lewis 1973a and 1973b) and is still very influential nowadays (see *e.g.* Woodward 2003, Collins, Hall and Paul, 2004). The intuition that causation has to do with 'what if things had been different' even traces back to Hume, according to some authors. Lewis, in particular, thought that the second part of the well-known definition of cause given by Hume was not just a restatement of the first claim, but a clear encouragement to think of causality in counterfactual terms. As Hume (1748) said, "a cause is [. . .] an object followed by another, and where all the objects similar to the first are followed by objects similar to the second. Or, in other words, *if the first object had not been, the second had never existed"* (italics ours).

Rubin formalised the basic ideas behind counterfactual reasoning as follows. Consider comparing two 'treatments', E and C, in the case of a headache. Let E represent taking two aspirins and C drinking just a glass of water. The potential outcomes Y relating to these two treatments may then be written as two random variables, namely Y (E) and Y (C). The causal effect of treatment E versus treatment C on Y for a particular subject j observed at time t+k is then defined as $Y_j(E) - Y_j(C)$*, i.e.* the differential headache response to taking the aspirins or just drinking a glass of water at time t. If we consider *n* subjects instead of only one subject, we have one causal effect $Y_j(E) - Y_j(C)$ per subject j. The average causal effect for this group of n persons can then be written $\Sigma\,[Y_j(E) - Y_j(C)]/n$, the sum extending from $j = 1$ to n.

Rubin's solution is often called the potential outcome (or response) approach , the two potential outcomes being in this simple case Yj(E) and Yj(C) for each j. Note that the causal effect may differ from one individual to the other; thus a "typical" causal effect (Rubin's term) is obtained as above by taking the average (or any other summary measure) of the individual causal effects. As pointed out by Brand and Xie (2007 p.394), "the potential outcome approach to causal inference extends the conceptual apparatus of randomized

experiments to the analysis of nonexperimental data, with the goal of explicitly estimating causal effects of particular 'treatments' of interest".

In the actual world, one never observes at the same time for the same individual both Y(E) and Y(C). In general, people are indeed assigned either to E or to C but not to both at the same time. Therefore, one can never observe for a *same* individual *j* at the *same* moment of time the causal effect $Y_j(E) - Y_j(C)$. Still following Rubin (1974), suppose there are only two subjects under study, denoted by 1 and 2. The typical causal effect (as defined above in the counterfactual situation) would then be $0.5[Y_1(E) - Y_1(C) + Y_2(E) - Y_2(C)]$. In the actual world, one would observe in a single study either $Y_1(E) - Y_2(C)$ or $Y_2(E) - Y_1(C)$ depending on whether subject 1 or subject 2 is assigned to E, and vice versa subject 2 or subject 1 to C. If treatments are randomly assigned to subjects, we are equally likely to observe one or the other difference. The expected difference in the outcome Y under randomization is then the average $0.5[Y_1(E) - Y_2(C)] + 0.5[Y_2(E) - Y_1(C)]$ which is the same result as that obtained in the counterfactual situation.

Suppose now that subjects 1 and 2 respond identically to the treatments E and C. In that case $Y_1(E) - Y_2(C) = Y_2(E) - Y_1(C)$ and moreover $Y_1(E) - Y_2(C) = Y_1(E) - Y_1(C)$ or $Y_2(E) - Y_1(C) = Y_2(E) - Y_2(C)$. In the situation of perfectly matched subjects with respect to the effects of the treatments, the observed causal effect is therefore equal to the counterfactual causal effect. Results under randomization or perfect matching can easily be extended from two subjects to *n* subjects. Thus the important conclusion: *randomization* and *matching* are two approaches measuring the causal effect in experimental and nonexperimental studies, though randomization cannot often be used in the social sciences and perfect matching is hardly possible in practice (see the thorough review by Morgan and Harding, 2006). In many actual situations in nonexperimental research, the assignment of units to the case and control groups is often prone to selection bias. Thus the assignment procedure is often not "ignorable", in the sense that the likelihood of treatment on the one hand and of the outcome on the other hand are not independent. For example, if the sickest take the new treatment and the healthier the older one, the outcome (*e.g.* recovery) in the treatment group will be due both to the new drug and to the characteristics of the patients at onset. In this case, one must control as best as possible for the assignment factors which have an impact on the outcome. In the above example, one would try to control, *e.g.* by stratification, for the state of health of both groups at the beginning of the trial. Clearly the actual challenge for the researcher is to evaluate how developed the field knowledge is, to ensure that that "all relevant factors" have been controlled for.

It should be noticed that Rubin requires that all subjects be potentially exposable. to the various *k* treatments (E_1, E_2, E_3, ...,E_k) - including possibly no treatment -being compared. In this approach, "causes are only those things that could, in principle, be treatments in experiments" (Holland, 1986). Therefore, an attribute (such as gender or ethnicity) cannot be a cause because potential exposability does not apply to it. In other words, in this framework there is "no causation without manipulation" (Holland *op. cit.*). For example, a study on gender differences in starting salaries cannot be addressed by randomized experiments and therefore gender cannot be a cause of differential salaries among subjects (Rubin, 1986). Gender is an attribute and cannot be considered in the search of effects of causes. According to Rubin, there is no clear causal answer to this issue. We will deal more with this later on.

Let us point out first that a major contribution of Donald Rubin's potential outcome model has been to stress the importance of carefully planning the design stage in observational studies.

In particular, the assignment mechanism by which some units are subjected to the putative cause ("treatment" group) and others are not ("control" group) should be studied in depth prior to any data analysis of the outcomes, and thoroughly explicated if possible: "we should objectively approximate, or attempt to replicate, a randomized experiment when designing an observational study" (Rubin 2007, p.25). For this purpose, Rubin (with others) has developed propensity score methods destined to eliminate bias, at the stage of the initial study design; a propensity score is the probability of being treated given the observed value of a vector of observed covariates, without reference to the outcome data (see *e.g.* Rosenbaum and Rubin, 1983; Rubin, 2001). Propensity score methods can be used to construct treatment and control groups similar as to their distributions of background variables. This approach requires of course that the assignment mechanism is otherwise unconfounded, *i.e*. it assumes that there are no latent confounders influencing the assignment of units between the treatment and control groups (for a discussion of the concept of confounding see *e.g.* Wunsch 2007). This requirement is less demanding in experimental studies where the units are randomly assigned to the treatment and control groups.

3. Counterfactuals: epistemological issues

Though Rubin's potential outcome framework is a significant contribution for analysing cause–effect relations in observational studies, its counterfactual basis nevertheless raises some important epistemological issues, which we now examine. The first two issues are quite often discussed in the literature. One concerns the soundness of the counterfactual approach given that the *counter*-fact is not observed, thus resulting in a lack of sound empirical basis. The other concerns the alternative between a counterfactual model measuring effects of causes and other models concerned instead with the causes of effect. The second two issues concern the concepts that, as we recalled in the introduction, back up the experimental method: manipulation and randomization. The third issue deals with complex mechanisms and the last makes a critical assessment of the too quick and simplistic analogies and parallelisms that have oft been made between the counterfactual model developed in statistics and the counterfactual analysis of causation developed by philosophers.

3.1.Potential outcomes: a "Platonic heaven"?

A major criticism that has been addressed to Rubin's potential outcome (or potential response) model is its *counter*factual basis (Dawid, 2000; Dawid, 2007). Paul W. Holland (1986) has called it 'the fundamental problem of causal inference'. The individual causal effect, as proposed by Rubin, requires taking the difference Yj(E) − Yj(C), though one of the two potential outcomes will never be observed. As Dawid said: "There is no world, actual or conceivable, in which both variables could be observed together. Their simultaneous existence must therefore be confined to some "Platonic heaven" of ideal forms, not fully accessible to real-world observation" (Dawid 2007, p. 510). It is impossible for the same subject j at the same time t to be assigned to both C and E. Rubin himself points out that "E and C are exclusive of each other in the sense that a trial cannot simultaneously be an E trial and a C trial" (Rubin 1974, p.689).

In order to get out of the 'Platonic heaven', the following strategy may be implemented. Either different individuals are assigned to E or to C at the same time, or the same individual is assigned to E and C in different times. In the first case, unknown factors may intervene and bias the causal effect, even when the individuals are matched as best as possible. The second case is known as a cross-over trial: contrary to a parallel-group design, the same subject first

takes treatment A and then after a first period of time crosses-over to taking treatment B during a second period of time. The effects of A and B are then compared on the same individuals. Two major assumptions however limit the scope of this approach (Jones, 2008). A first one is that subjects are in the same state at the beginning of period two as they were at the start of period one, which is a strong assumption indeed. A second limiting factor is a possible carry-over effect: the effect of treatment A might be carried over from the first to the second period, biasing the difference of effects between the two treatments at the end of the trial. Neither approach solves therefore the 'fundamental problem of causal inference'. Thus, the 'true' causal effect remains latent. Actually, we usually have to face the problems of unit heterogeneity and temporal instability in observational studies, though this might not always be the case in experimental ones (see Holland, 1986, section 4).

Although ways out of the 'Platonic heaven' may be found, a conceptual problem about the lack of empirical basis remains at the individual level. Take the aspirin example again: "Had Mr Jones swallowed the aspirin half an hour ago, his headache would have gone now". The fact is that Mr Jones did not swallow the aspirin half an hour ago. This makes it impossible to say what would have happened if he had taken the aspirin, *based on empirical evidence*. Since he did *not* take the aspirin, this hypothesis is completely equivalent to many others: what if Mr Jones went for a walk, or took paracetamol instead, or consulted a holy man or had taken the aspirin later rather than sooner? Here, several putative causes could be equally effective in relieving headache, and consequently there is no a priori reason to claim the counterfactual 'Had Mr Jones taken an aspirin half an hour ago, his headache would have gone now' picks out the right cause whilst 'Had Mr Jones consulted with a holy man, his headache would have gone now' instead doesn't. Moreover, some of these putative causes are statistically not independent; for instance paracetamol would typically be exclusive of aspirin. In this case, the counterfactual itself is clear but more information on *facts*, here on Mr. Jones' actual behaviour, is needed**.**

3.2. Causes of effects

The potential outcome model focuses on the 'effects of cause' problem and can hardly tackle the 'causes of effect' issue, which is central to much of the social sciences (Ni Bhrolchain and Dyson, 2007). Counterfactualists are well aware of this problem, Rubin's causal model having been specifically developed to examine the effects of causes and not the causes of effects. Though debatable, the argument is that causal effects come first in the process of causal inference; therefore one should focus on the measurement of the effects of causes, as in the case of randomised experiments, rather than vice versa on the causes of effects (Holland, 1988). Actually, in many situations one focuses on the causes, such as on the causes of death and on the factors determining mortality and morbidity, rather than on the effects (age at death, in this case). Though favouring a counterfactual approach to causality himself, Heckman (2005, p.2) has nevertheless pointed out that "science is all about constructing models of the causes of effects", and insists on the need of understanding the causes producing the effects, or in other words the determinants of the outcomes. Clearly, both issues, namely "causes of effects" and "effects of causes" are relevant, one or the other or both according to the problem at hand, and, arguably, both issues should be based on a same concept of causality.

The preference for models that measure effects of causes or find out causes of effects brings up more general questions about a unique approach for causal inference. Even if we take for granted that counterfactual models are successful tools to measure effects of causes, it remains an open question of what to do with causes of effects, since this seems an important task in science too. In the final section on discussion and conclusion we will get back to this

issue and suggest that alternative frameworks—in particular a structural framework—are needed to answer questions about causes of effects.

3.3. Manipulation

A major difficulty with the potential outcome framework is that it can hardly take attributes into account (Ni Bhrolchain and Dyson, 2007). Holland (2001) is quite explicit in saying that attributes such as race cannot be manipulated and therefore counterfactuals involving attributes make no sense. For example, the question "What would your life have been had your race been different?" can be viewed as "ridiculous" (Holland op. cit., p. 226). If one accepts the counterfactual/manipulation framework, attributes (such as age, gender, race,…) cannot indeed be causes. Nevertheless, many scientists would consider gender as a cause of initial salary discrimination in many countries, ethnicity as a cause of differential HIV prevalence in Sub-Saharan Africa, ageing as a cause of hearing loss, etc. This is because these attributes are not only associated with their respective effects—they are part of the causal mechanism itself. For example, belonging to different ethnic groups in Africa results in having different reproductive norms, values, and sexual behaviours (such as multi- or single-partnership), and these characteristics are major determinants of exposure to HIV. Any explanatory framework in the social sciences that cannot take attributes into account is therefore necessarily incomplete. The statement "no causation without manipulation" (Holland, 1986) is not adequate in those cases and different test settings have to be developed in order to evaluate effects of non-manipulable causes.

As discussed in the introduction, the manipulative account of causation is based on the idea that one manipulates an independent variable and sees how the value of a response variable depends upon the value of the manipulated variable. If feasible, it has several advantages, as discussed in Sobel (1995). Among others, issues concerning causal priority are easily solved, as manipulation of the putative cause comes first and the possible effects later. However, as manipulation is only *a* means among others for testing causal relations, our point is that it is inappropriate to consider manipulability as an essential condition for causality. In other words, manipulation is only one of the possible ways to test for causal relations, and, more to the point, most often not the one that is actually feasible in observational contexts. Of course, if manipulation is possible, so much the better.

Because many variables cannot be manipulated, *e.g.* attributes and causes that have occurred in the past, the key question around which model building and model testing turn around is: are variations among units in the treatment variable followed by variations in the outcome variable or not? For example, does ageing (a change in the input variable) lead to an increase in physical and mental deficiencies (a change in the outcome variable)? No manipulation and no counterfactuals actually need to be evoked here: one compares individuals of different ages or the same individuals at different ages in order to see if deficiencies are usually more common among the older population than among the younger one. Most probably we will observe that they are. The main problem in a complex situation is however controlling as best as one can for possible confounders, such as period effects in this case. See for example the interesting discussion of gender effects on earnings in Sobel *op.cit.*, pp. 21-22.

Rather than manipulation, the basic idea or rationale underpinning causal analysis is that some form of joint *variation* between variables of interest has to be evaluated. In an experimental context variations come from the manipulation of variables, in the counterfactual approach variations come from thought experiments, in purely observational contexts variations come from the marginal-conditional decomposition of multivariate distributions;; for a more

systematic exposition of model building and model testing based on the notion of variation, see Russo (2009a, 2009b).

To give yet another example taken from Sobel (1995), take the association between a father's occupation and his son's intelligence, measured *e.g.* by his performance at school. A manipulation of the father's occupation will most probably not lead to a change in the child's intelligence, as Sobel rightly states, and the former should not in this case be considered as a cause of the latter. We can nevertheless assume in a longer time-frame that an increase in fathers' occupational level - and more generally socio-economic status - from one generation to another, will be accompanied by an increase in the educational level of their sons, as observed also cross-sectionally among social groups. In this sense, father's SES rightly is a 'cause' of the child's education. What we need here is an understanding of the *social mechanism* (as defined for example in Hedström and Swedberg, 1998) linking father's occupation and child's intelligence, rather than seeing if wiggling one leads to a twinkle in the other.

This example shows again that neither a suitable concept of causality nor the methods of causal inference should be bound to manipulation or counterfactuality. Besides considerations about outcomes of manipulations or counterfactuals, considerations about the underlying mechanism(s) are required in order to decide whether a relation is causal or not. This problem has received recent attention by philosophers. According to some, care is needed in distinguishing between the concept of causality itself and the evidence needed to establish causal relations. This idea, developed by Russo and Williamson (2007 and 2011), and Russo (2009a and 2010) is that, simply put, causal relations have to be established on the basis of multi-fold evidence, in particular evidence about the underlying mechanisms and evidence about difference-making. Concerning the *concept* of causality, it has been suggested that causality has to be understood in *epistemic terms*, that is as the scientist's rational beliefs about causal relations (see Williamson 2005, 2006a, 2006b). On the one hand, the concept of causality is not reduced to the concept of manipulation or of counterfactuals—those are some of the possible *causal methods*—and causality has to do with the rational and scientifically-informed opinions we come to form by means of causal analysis. On the other hand, the concept of causality is not reduced to the concept of mechanism or of difference-making—those are its *evidential components*, that is the types of evidence the scientist needs in order to establish whether a joint variation is rightly deemed to be causal. It is important to emphasise that such an epistemic approach does not lead to a subjective and arbitrary view of causality, because (rational) causal beliefs are formed upon *evidence*, and evidence can be *objectively* evaluated.

It is also worth pointing out that manipulation not only is not necessary for testing causal relations, but also it is not part of the concept of counterfactuality. Indeed, *counter*-factual means "contrary to facts", *i.e.* based on non-realised or non-observed events. However, this does not imply that the non-observed causes be manipulable. In short, manipulability is, when possible, an aid for measuring the possible effect of a putative cause without being a necessary ingredient of counterfactuality nor of causality.

Morgan and Winship (2007 p.280) have supported the argument concerning causal attributes by evoking the construction of counterfactual thought experiments. For example, "the counterfactual model could be used to motivate an attempt to estimate the average gain an employed black male working full time, full year would expect to capture *if all prospective employers believed him to be white*" (italics ours). However, there exists an 'infinity' of

possible thought experiments for each case and no way of testing the validity of their claims with actual data. In the previous example, one could nevertheless estimate the difference in income between Blacks and Whites controlling if possible for all income factors other than race (such as level of education, health status, etc.). No hypothetical counterfactual thought experiment is actually required here. The real problem is both knowing and observing the factors that have to be controlled for, but there is no method of testing if in this way we have made Blacks and Whites exchangeable with respect to the outcome (Kaufman and Cooper, 1999). Only the progress of knowledge can tell us if we have not left out important latent confounders from the analysis.

Some authors such as Paul Holland and James Woodward (for both, see Woodward 2003, chapter 2) contend that the issue in the gender/salary example is actually not to manipulate gender, but in this case, to modify the beliefs concerning gender, or the attitudes and practices of the employer as to hiring females, *i.e.* variables that can be manipulated contrary to gender. Similarly, the Black/White dichotomy is a case of social relations, and these can eventually be changed over time (Muntaner, 1999). Even if we agree with this view, this proposal can nevertheless hardly be extended to all the cases of attributes as causes. Consider the example of sex (male, female) as a major risk factor of breast cancer. No manipulation of beliefs and attitudes towards breast cancer will change the fact that breast cancer is about 100 times less common among men than among women. As hormones seem to play an important role in the aetiology of breast cancer, the biological differences between males and females probably explain the association between sex and breast cancer (American Cancer Society, 2010).

If we see that effects differ among categories and if we can find a suitable stable mechanism or causal narrative explaining why this happens - controlling of course for possible confounders - attributes (such as sex or ethnicity) could indeed be considered as causes of effects even though they cannot be manipulated physically or mentally. In this case, attributes are part of a causal chain (of sub-mechanisms), in which some variables (such as hormones or beliefs) could be manipulated and others not (such as sex or ethnicity).

3.4. Randomization

A randomised experimental study aims to control known and unknown confounders by randomisation: assign randomly individuals to two groups, that then differ only by the fact that one 'receives' the putative cause (the new drug) and the other does not (it usually receives a placebo instead) and after a lapse of time compare them. For simplicity, all units in the same group should receive the same treatment and there should be no interaction among the units themselves. D. Rubin has called these constraints the 'stable-unit-treatment-value-assumption', or SUTVA for short (Rubin, 1990); these conditions can be relaxed in more complex designs. In addition to the major restriction that randomised studies are often ethically or practically unfeasible in the social sciences, experimental results of this kind are also influenced by the placebo/nocibo effects, *i.e.* a favourable or unfavourable effect of the placebo due to subject-expectancy (Amanzio, 2001), and also by (post-treatment) non-compliance with assigned treatment and by missing outcomes, *i.e.* drop-outs (Mealli and Rubin, 2002; Frangakis and Rubin, 2002*).* Moreover, from a strict statistical point of view, placebo/nocibo effects should be considered as "residual", or, better said, "unexplained" effects (see Bouckaert and Mouchart, 2001), if one recognises that a statistical model provides a partial explanation only, the stochastic component corresponding to what is not explained by the model (see Mouchart, Russo and Wunsch 2009, and Mouchart and Russo 2011*).*

We recalled in the introduction that the counterfactual model has its roots in the Fisherian experimental framework, where units are randomly assigned to disjoint sets of treatments (Rubin, 2004). Our point is that although randomization has indeed proved very useful as a method enabling to distinguish causal effects from spurious ones, randomization is by no means *the* essential element of causal modelling. This view is shared by many scientists and philosophers. For instance, Heckman (2008) has stressed that "The claim that causality can only be determined by randomization reifies randomization as the 'gold standard' of causal inference". There are two types of problems with randomization.

On the one hand, the fact that a population can be affected by latent heterogeneity, *i.e.* it is composed of individuals characterised by different values of non observable but potentially causing variables, is a crucial issue. If all individuals were exactly identical, in the sense of being characterised by an identical response distribution, there would be no need to randomize. But because individuals are in fact not identical, randomization may still provide a measure of mean effect although such a measure may be misleading or irrelevant. As a simple example, if in subpopulation A the treatment has a positive effect and in subpopulation B it has an equally negative effect, and if there is no way of distinguishing the two subpopulations with the available data, the mean effect for the whole population may be null without being the effect for any individual.

On the other hand, in the social sciences, randomized experiments are often difficult to conduct for ethical and/or practical reasons. Nevertheless causal patterns have indeed been discovered in all disciplines in the absence of randomized experiments. In those cases randomization is replaced by a careful control of the relevant covariates and by using criteria supportive of causal inference (Ni Brolchain and Dyson 2007; Glasziou et al. 2007).

3.5 Multiple causes-multiple effects.

Counterfactual reasoning is widely used also in everyday contexts; however, an important problem concerns the issue that it is usually unclear what has to be kept fixed in checking what would have happened, had things been different. As Lewis has said: "counterfactuals are infected with vagueness" (Lewis, 1979, p.457). For counterfactualists like Lewis (2004), causation is a relation between events and we need to know precisely what they are. Take the aspirin example again: "Had Mr Jones swallowed the aspirin half an hour ago, his headache would have gone now". The facts are that Mr Jones did not swallow the aspirin half an hour ago and has presently a headache. The counterfactual proposition "Had Mr Jones swallowed the aspirin half an hour ago, his headache would have gone now" asserts that aspirin is a putative cause of relieving headache. As recalled in section 2, Rubin's causal model would compare the effect of Mr Jones not taking aspirin to the effect of Mr Jones taking aspirin. The issue now is that the causes of a headache may be multiple *and* the causes for relieving a headache are also multiple; moreover the effect of the former causes and the effect of the latter causes are possibly not independent. For example, the effect of aspirin might be different according to the fact that the cause of the headache is indigestion or flu.

More generally, even in seemingly simple situations one has to face an issue of multiple causes-multiple effects, involving more than one mechanism at a time. In practice, it is usually not sufficient to compare Jones 1 taking the aspirin to Jones 2 not taking the aspirin. One must control the factors possibly confounding the relationship between aspirin and headache. The two Jones should be matched on all the relevant covariates which could lead to confounding. However if there are many covariates, as is most often the case in social sciences, it will often be impossible to match on the relevant covariates, even using propensity

scores. Concerning the latter more specifically, if samples are small or if assignment bias is important, it can occur that there will be few individuals in the non-treatment group with propensity scores similar to those in the treatment group. Individuals poorly matched are usually dropped from the analysis, leading to further reductions in the sample size. Group overlap (the 'common support' condition) must therefore be substantial for the method to work adequately (Dehejia and Wahba, 2002; Bryson, Dorsett and Purdon, 2002). A major problem is specifying the relevant covariates possibly responsible for confounding. As Rubin himself (Rubin 1974) has pointed out, more well-formulated causal models are needed in the social sciences, because controlling for relevant covariates may not be trivial without a properly developed causal model. This is the reason why Pearl (2009) is in favour of modelling the putative causal relations between treatments, outcomes, observed and unobserved covariates.

Multiple-causation problems can be tackled under different causal frameworks: potential outcomes (*e.g.* Rubin 2004), causal graphs (*e.g.* Pearl 2000), marginal-conditional structural decomposition (*e.g.* in the spirit of the work of the Cowles Commission in the fifties, see in particular Hood and Koopmans 1953). Each approach stresses different specific features. The main issue, as Rubin (2004) has stated, is to propose the " correct conceptual structure", most probably a more difficult issue in observational studies than in experimental ones.

An additional issue is whom should we compare? It has been argued that in some cases the average treatment effect between the treated and the non-treated is not the quantity of interest; one should consider instead the treatment effect for those treated (Heckman, 2005; Winship and Morgan, 1999). These are the cases where, for instance, a policy measure should be beneficial for those who are assigned (or who chose the assignment) to it, and not necessarily for all individuals. For those taking the treatment, the latter can be effective for some individuals and not for others. The heterogeneity of the population treated is the point of interest in this case. For example, why does aspirin work in relieving headaches for some people and not for others?

3.6 The individual or the population?

Many counterfactualists (*e.g.* Holland, 2001), both in the statistical and social science literature, trace the origins of the ideas behind the counterfactual approach in the work of the philosopher David Lewis. Is this filiation valid? We argue here that it is not. Consider again the example of aspirin and headache. On the one hand, the potential outcome model wants to establish whether aspirin is an effective treatment for headache, namely whether aspirin relieves headache. Of course, the fundamental unit *is* the individual. More explicitly, the model concerns a set of single cases, and the individual causal effect is measured using individual data. However, the goal of the potential outcome model is not to know whether Mr Jones would have recovered had he taken an aspirin, but rather whether aspirin is an effective treatment in the target population. On the other hand, Lewis (1973a, 2004) asks what the *truth conditions* of counterfactual statements are. Therefore he asks, given a particular situation, whether the counterfactual claim picks out the right cause. For instance, Mr Jones has been suffering from headache for the last four hours; we now ask whether had he taken the aspirin, his headache would have gone now. This means, in Lewis' approach to ask whether aspirin would be *the* cause of his recovery. True, the analogy is definitively there; Rubin's counterfactual exploits the same idea behind Lewis' counterfactual: had the cause not been, the effect would not have occurred either, but this does not imply that these accounts be the same or that their scope be the same.

This leads us, following also the arguments given in Russo and Williamson (2007 and 2010), to draw a distinction between single-case and generic causal claims. In Lewis' counterfactual reasoning, singular causal relations are established by means of an evaluation of counterfactual statements. In order to know whether taking the aspirin actually relieved Mr Jones' headache, or whether it would have relieved his headache had he took it, we ascertain the truth of the corresponding counterfactual statement. This kind of causal relation is single-case, namely a particular causal relation taking place at a certain time and place. Another story is to evaluate the causal effectiveness of aspirin in relieving headache in a target population, which is exactly the purpose of the potential outcome model. It is true that Rubin's potential outcome model, and more generally counterfactual models, use individual data, but this does not mean that they focus on individual or single-case causal relations per se. The result of a counterfactual model would sound like this: more often than not, taking aspirins relieves headache, therefore, given any individual randomly sampled from the population, had s/he taken the aspirin, his/her headache would most probably have gone. This is not the same as saying that 'had Mr Jones taken the aspirin, his headache would have gone now'. The former counterfactual, although based on individual-level data, is generic, whilst the latter is single-case, that is it concerns a particular causal relation taking place in a given time and place. The reference to Lewis' single-case approach is therefore not appropriate for a generic approach.

4. Application

In this section we exemplify some of the issues we raised earlier in section 3. In particular, by means of a study on the causes of self-rated health, we endeavour to show that (i) as it often happens in social science, randomisation and manipulation cannot be performed; (ii) defining the counterfactual to be evaluated is far from being an obvious task, with important repercussions on model-building; (iii) when real data are analysed and relevant contrasts are chosen, the model loses its *counter*factual character; (iv) the counterfactual approach is in a hard position to answer questions about *how* (causal) variables act upon the effect in a complex network of relations.

This application is based on a recent research on the determinants of self-rated health in the post-communist context of the Baltic countries in the Nineties (Gaumé and Wunsch 2010). The data come from two surveys conducted in 1994 and 1999 in the three Baltic countries. The data are cross-sectional; in spite of criticisms to the possibility of making causal inference in cross-sectional studies, in the present example causal relations could be tested because of substantial background information and because of clear knowledge about temporal priority between most of the variables (on this point see Wunsch *et al* 2010). Notice also that we are here in the case of an observational study, based on *existing* data. Even if the researchers were in the position of deciding how to collect data themselves, randomisation, and a fortiori manipulation, could not be used here for ethical as well as practical reasons.

Suppose a counterfactualist were interested in the effect of 'Education' (putative cause) on 'Self-rated health' (outcome). S/he would ask what would happen to self-rated health for counterfactual values of education. Let us spell this out. In order to assess the causal effect of, *e.g.*, higher education on self-rated health, we have to ask what would happen to the individuals, *had they not* a higher education. This counterfactual is however highly ambiguous. *Not* having a higher education may mean many different things: secondary schooling or less, just secondary schooling, primary schooling or less, just primary schooling, technical education, or no formal schooling at all. Which one is the counterfactual to

evaluate? Each counterfactual would correspond to a different model, leading to a different measure of the effect.

In Gaumé and Wunsch (2010), taking into account the availability of data, three categories of education have been chosen: primary or less, secondary, and higher education. For each of these educational categories, the surveys provide for each individual his/her level of self-rated health according to his/her level of education. Having such data, we can indeed test the effect of different categories of education on self-rated health. But then, the *counter*factual model collapses into a *factual* one, actually into a family of models evaluating the effect of each educational category—for which real data *is* available—on self-rated health.

Suppose for the sake of the argument that we could choose *one* counterfactual counterpart to education. In order to assess the causal effect of education on self-rated health, we should control for possible confounders, taking into account the heterogeneity of the population. From the existing literature, we know for example that social support and physical health could confound the relationship between education and self-rated health as they have an impact on the latter and are also associated to the former (see Wunsch 2007 for a discussion of confounding). Social support and physical health should therefore be added to the model, albeit as control variables.

Moreover, it may be the case that the research question is broader than just estimating the causal effect of education on self-rated health. Suppose it also involves understanding *how* education has an effect (if any) on self-rated health; then, the counterfactual model is in a hard position to give an answer. Background knowledge on the Baltic context tells us for example that education can have an impact on self-rated health via its action on the individual's locus of control and psychological distress and/or on the individual's alcohol consumption. In order to answer the *how*-question, variables such as locus of control or alcohol consumption have to be included in the model not as control variables, but as variables having a possible causal impact per se. The network of relations can thus be presented as in the directed acyclic graph of figure 1. In this graph, the E variables represent residual terms, *i.e.* possible latent factors and other random variations.

But if it is this broader *how*-question we are interested in, the counterfactual model makes it difficult to disentangle the causal effects of each variable in the network via direct and indirect paths. In fact, the counterfactual model does only allow testing each relationship one by one and not as part of a complex network of multiple causes–multiple effects. Actually, the relations in Gaumé and Wunsch's study have been evaluated by way of Bayesian structural equation modelling using a Markov Chain Monte Carlo (MCMC) procedure, thus enabling to test the various relations composing the *network* rather than simple one cause–one effect relations. The reader interested in the results of the study is referred to Gaumé and Wunsch (2010).

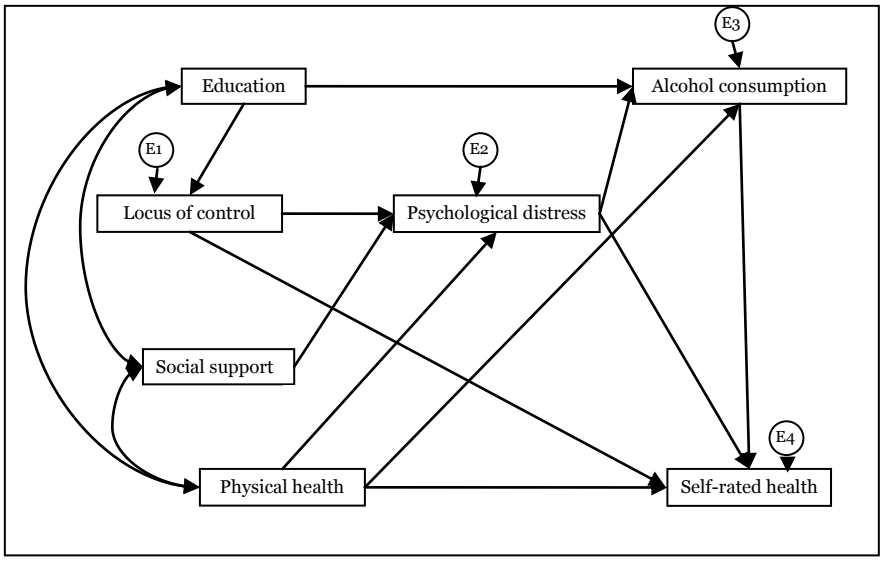

Figure 1 A conceptual model of self-rated health in the Baltic countries 1994-1999

5. Discussion and conclusion

This paper has examined some epistemological issues raised by the counterfactual approach for causal inference in the social sciences, and in particular in observational studies. One strength of counterfactual models developed in statistics by Donald Rubin and others (*e.g.*, Paul Holland) is to regain the power of the experimental (Baconian) method implementing the two pillars of experimental science: manipulation and control through randomisation. Rubin's approach has led to a number of improvements in the quasi-experimental methodology, encouraging researchers to model the mechanisms of assignment more explicitly.

The counterfactual approach raises however several issues which the present paper has discussed.
The first issue concerns the soundness of the counterfactual approach, given that the counterpart of the (putative) cause is not observed, thus undermining a sound empirical basis. Another issue concerns the alternative between a counterfactual model measuring effects of causes and other models concerned instead with the causes of effect. Two other issues concern the concepts that back up the experimental method: manipulation and randomization. Another issue deals with complex mechanisms, and the last issue makes a critical assessment of the parallelism that has been made between the counterfactual model developed notably by D. Rubin in statistics and the counterfactual analysis of causation developed by philosophers, especially D. Lewis. The application discussed in section 4 exemplified some of these issues.

In his seminal book on causality, Judea Pearl (Pearl, 2000) upholds the opinion that there are presently two approaches to causality in science: the potential outcome or counterfactual framework as championed most notably by Donald Rubin, and the structural modelling framework *à la* Wright, Haavelmo, Duncan, Blalock, and others (including Pearl himself). Structural modelling, as the name suggests, aims to model (causal) structures or mechanisms, that is it aims to make explicit how elements of a social system are linked as causes and effects. A structural model or causal mechanism is thus a network of causes and effects proposed as an answer to an explanation-seeking '*Why?*' question, *i.e.* a '*How does it work*?' question, widening the scope of the '*What-if?'* question as in a strict counterfactual framework. Structural modelling avoids many of the issues confronting the counterfactual framework. In particular, it is based on observable outcomes, and manipulation – though

useful - is not mandatory. Consequently, a structural approach can take attributes into account. Finally, structural models can deal with both effects of causes and causes of effects.

Many counterfactualists are however sceptical about the practical usefulness of this type of causal framework, even if they recognize that "understanding and identifying causal mechanisms is, perhaps, the primary driving force of science" (Holland, 2001, p. 224). For Holland, for instance, the danger lies in the fact that almost 'anything' can be considered as a cause "because we are just talking rather than doing", *i.e.* setting up 'treatments' or 'interventions' (*op. cit.* p. 225). Actually, a causal mechanism does not appear from nowhere, like the white rabbit drawn from a conjuror's hat. Nor it necessarily results from adding more and more variables to the predictive set (Sobel, 2000). As we have argued elsewhere (*e.g.* Mouchart, Russo, and Wunsch, 2009; Mouchart and Russo, 2011; Russo, forthcoming), a structural model should be based on the best available knowledge one has of the field; all postulated relations should be accounted for. In particular, it should incorporate those variables deemed to be responsible for possible assignment bias. The postulated mechanism is then represented by a recursive decomposition of the initial multivariate distribution of the data, and the model should display invariance (*i.e.* replication) properties.

The structural modelling framework also has its problems. First of all, to avoid loss of exogeneity, known confounders can be incorporated into the model only on condition that indicators of these confounders are available in the data set. In many situations, especially when one uses secondary data (*i.e.* data collected by others), no information has been obtained for some of the variables in the model. Confounding bias may not be avoidable then, though in some cases omitted variable bias can be controlled for by fixed effects regression or by instrumental variables regression (Stock and Watson, 2003). Unknown latent confounders may however still bias the results. A major drawback is that in many cases one only has a scant knowledge of the underlying mechanism. In this situation, descriptive analysis or exploratory data analysis might be more useful than poor structural modelling. And if one is looking for the effects of causes, the Rubin causal model could be considered, even if we do not adhere to its counterfactual underpinnings.

The challenge does not concern the relevance of structural modelling for causal inference, but rather the procedure to be followed for building a suitable structural model. Among others, the following questions can be raised in this respect:

(i) How can structural models operationalise the integration of field knowledge, in cases of a lack of consensus among experts?
(ii) With respect to graph models, taking into account the criticisms that have been raised (Imbens and Rubin, 1995; see also Pearl's rejoinder, 1995), to what extent should structural models switch the focus from structuring a set of variables, a set of equations, or a graph, to structuring a multivariate distribution?
(iii) Can one take mechanisms as a basis for explanation ?
(iv) Can we then opt for a stochastic view of mechanisms represented by conditional distributions?

Acknowledgments
Comments by two anonymous reviewers are gratefully acknowledged. F. Russo wishes to thank Phyllis McKay Illari, Jon Williamson and Lorenzo Casini for many insights and suggestions on the relations between manipulation, randomisation, and counterfactuals. She also gratefully acknowledges financial support from the British Academy. M. Mouchart

gratefully acknowledges financial support from IAP research network grant nr. P6/03 of the Belgian government (Belgian Science Policy).

References

AMANZIO M. *et al.* (2001). Response variability to analgesics: a role for non-specific activation of endogenous opioids. *Pain*. 90(3):205-15.

AMERICAN CANCER SOCIETY (2010). Breast Cancer Overview, at http://www.cancer.org/, accessed June 29, 2010

BOUCKAERT A., MOUCHART M. (2001). Sure outcomes of random events: a model for clinical trials, *Statistics in Medicine*, (20), 521-543.

BRAND J. E., XIE Y. (2007). Identification and estimation of causal effects with time-varying treatments and time-varying outcomes, *Sociological Methodology*, 37(1), 393-434.

BRYSON A., DORSETT R. PURDON S. (2002). *The use of propensity score matching in the evaluation of active labour market policies*, Working Paper Number 4, Department Work and Pensions, London.

COLLINS J., HALL N., PAUL L.A., eds., (2004). *Causation and counterfactuals*, The MIT Press, Cambridge, Massachusetts.

DAWID A.P. (2000). Causal inference without counterfactuals, *Journal of the American Statistical Association*, 95(450), 407-424.

DAWID A.P. (2007). Counterfactuals, hypotheticals and potential responses: a philosophical examination of statistical causality. In F. Russo and J. Williamson (eds), *Causality and Probability in the Sciences*, College Publications, Texts In Philosophy Series 5, 503-32, London.

DEHEJIA R.H., WAHBA S. (2002). Propensity score-matching methods for nonexperimental causal studies, *The Review of Economics and Statistics*, 84(1), 151-161.

DUCHEYNE S. (2005), Joan Baptiste van Helmont and the Question of Experimental Modernism, *Physis: Rivista Internazionale di Storia della Scienza*, vol.43, pp. 305-332

FISHER R.A. (1925), *Statistical methods for research workers*, Oliver and Boyd, Edinburgh.

FISHER R.A. (1935), *The design of experiments*, Oliver and Boyd, Edinburgh.

FRANGAKIS C.E., RUBIN D.B. (2002). Principal stratification in causal inference, *Biometrics*, 58, 21-29.

GAUMÉ C., WUNSCH G. (2010) Self-rated health in the Baltic countries, 1994-1999. *European Journal of Population*, 26(4), 435-457. DOI10.1007/s10680-010-9217-7.

GLASZIOU P., CHALMERS I., RAWLINS M., MCCULLOCH P. (2007) When are randomised trials unnecessary? Picking signal from noise, *British Medical Journal*, 334, 349-351.

HACKING I. (1988), Telepathy: Origins of randomization in experimental design, *Isis*, Vol. 79, No. 3, A Special Issue on Artefact and Experiment (Sept., 1988), pp. 427- 451.

HALL N.S. (2007), R. A. Fisher and his advocacy of randomization, *Journal of History of Biology*, 40, 295-325.

HECKMAN J. (2005), The Scientific Model of Causality, *Sociological Methodology*, 35(1), 1-97.

HECKMAN J. (2008), Econometric Causality, *International Statistical Review*, 76(1), 1-27.

HEDSTRÖM P., SWEDBERG R. (1998). *Social Mechanisms*, Cambridge University Press, Cambridge.

HOLLAND P.W. (1986) Statistics and causal inference, *Journal of the American Statistical Association*, 81(396), 945-960.

HOLLAND P.W. (1988). Comment: causal mechanism or causal effect: which is best for statistical science?, *Statistical Science*, 3(2), 186-188.

HOLLAND P.W. (2001). The false linking of race and causality: lessons from standardized testing, *Race & Society*, 4, 219-233.

HOOOD W.C., KOOPMANS T.C., eds. (1953) *Studies in Econometric Method*, Wiley, New-York.

HUME D. (1748), *An Enquiry Concerning Human Understanding*, Bobbs-Merrill, Indianapolis, 1955.

IMBENS G.W., RUBIN D.R. (1995). Discussion of 'Causal diagrams for empirical research' by J. Pearl, *Biometrika*, 82(4), 694-695.

JONES B. (2008), The cross-over trial: a subtle knife, *Significance*, 5(3), 135-137.

KAUFMAN J.S., COOPER R.S. (1999). Seeking causal explanations in social epidemiology, *American Journal of Epidemiology*, 150(2), 113-120.

KLEIN J. (2008), Francis Bacon's Scientia Operativa, The Tradition Of The Workshops, And The Secrets Of Nature, in C. Zittel, R. Nanni, G. Engel and N. Karafyllis. Brill (eds), *Philosophies of Technology: Francis Bacon and his Contemporaries*, Brill E-Books. DOI:101163/ej.9789004170506.i-582.1. Accessed 02 February 2010.
KLEIN, J.(2009), Francis Bacon, The Stanford Encyclopaedia of Philosophy (Spring 2009 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2009/entries/francis-bacon/>. Accessed 02 February 2010.
LEWIS D. (1973a), Causation, *Journal of Philosophy*, 70, 556-567. Reprinted with postscripts in D. Lewis (1986), *Philosophical Papers II*, Oxford University Press, Oxford, 159-213.
LEWIS D. (1973b), *Counterfactuals*, Harvard University Press, Cambridge.
LEWIS D. (1979). Counterfactual dependence and time's arrow, *Noûs*, 13(4), 455-476.
LEWIS D. (2004), Causation as influence, in J. Collins, N. Hall, L.A. Paul, eds, (2004), *Causation and counterfactuals*, The MIT Press, Cambridge, Massachusetts.
Mealli F., Rubin D.B. (2002). Assumptions when analyzing randomized experiments with noncompliance and missing outcomes, Health Services & Outcomes Research Methodology, 3, 225-232.
MENZIES, P. (2009), "Counterfactual Theories of Causation", *The Stanford Encyclopedia of Philosophy (Fall 2009 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2009/entries/causation-counterfactual/>.
MORGAN S.L., HARDING D.J. (2006). Matching estimators of causal effects. Prospects and pitfalls in theory and practice, Sociological Methods & Research, 35(1), 3-60.
MORGAN S.L., WINSHIP C. (2007). *Counterfactuals and causal inference*, Cambridge University Press, New York.
MOUCHART, M., RUSSO, F., WUNSCH, G**.** (2009). Structural modelling, exogeneity, and causality. In H. Engelhardt, H-P Kohler, A. Fürnkranz-Prsawetz (eds). *Causal Analysis in Population Studies. Concepts, Methods, Applications*. Dordrecht: Springer, 59-82.
MOUCHART, M., RUSSO, F. (2011). Causal explanation. Mechanisms and recursive decomposition. In P. McKay Illari, F. Russo, J. Williamson (eds), *Causality in the Sciences*. Oxford University Press.
MUNTANER C. (1999). Invited commentary: Social mechanisms, race, and social epidemiology, American Journal of Epidemiology, 150(2), 121-126.
NI BHROLCHAIN M., DYSON T. (2007) On causation in demography: issues and illustrations, *Population and Development Review*, 33(1), 1-36.
PEARL J. (1995). Rejoinder to discussions of 'Causal diagrams for empirical research', *Biometrika*, 82(4), 702-710.
PEARL J. (2000). *Causality. Models, Reasoning, and Inference*, Cambridge University Press, Cambridge.
PEARL J. (2009). Letter to the Editor: Remarks on the method of propensity score, *Statistics in Medicine*, 28, 1415-1416.
PEIRCE C.S., JASTROW J. (1885). "On Small Differences in Sensation". *Memoirs of the National Academy of Sciences* **3**: pp. 73–83. http://psychclassics.yorku.ca/Peirce/small-diffs.htm
RESCHER N. (1978), *Pierce's Philosophy of Science*, University of Notre Dame, Notre Dame - London.
ROSENBAUM P.R., RUBIN D.B. (1983). The central role of the propensity score in observational studies for causal effect, *Biometrika*, 70(1), 41-55.
RUBIN D.B. (1974). Estimating causal effects of treatments in randomized and non randomized studies, *Journal of Educational Psychology*, 66(5), 688-701.
RUBIN D.B. (1986) Which ifs have causal answers, *Journal of the American Statistical Association*, 81(396), 961-962.
RUBIN D.B. (1990). Comment: Neyman (1923) and causal inference in experiments and observational studies, *Statistical Science*, 5(4), 472-480.
RUBIN D.B (2001). Using propensity scores to help design observational studies: application to the tobacco litigation, *Health Services & Outcomes Research Methodology*, 2, 169-188.
RUBIN D.B. (2004). Direct and indirect causal effects via potential outcomes, *Scandinavian Journal of Statistics*, 31, 161-170.

RUBIN D.B. (2007), The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials, *Statistics in Medicine*, 26(1), 20-36.
RUSSO, F. (2009a), "Variational causal claims in epidemiology", in *Perspectives in Biology and Medicine*, 52(4), 540-554.
RUSSO, F. (2009b), *Causality and causal modelling in the social sciences. Measuring variations*, Springer, New York.
RUSSO, F. (2010), Causal webs in epidemiology, *Paradigmi*, Special Issue on the Philosophy of Medicine.
RUSSO, F. (Forthcoming) Correlational data, causal hypotheses, and validity. *Journal of General Philosophy of Science.*
RUSSO, F., WILLIAMSON J. (2007), "Interpreting causality in the health sciences", *International Studies in Philosophy of Science*, 21(2), pp. 157-170, 2007.
RUSSO, F., WILLIAMSON J. (2011), Generic vs. single-case causal knowledge. The case of autopsy, *European Journal for Philosophy of Science*.
SOBEL M.E. (1995). Causal inference in the social and behavioural sciences, chapter 1 in G. Arminger, C. C. Clogg, and M. E. Sobel, eds., *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. NewYork: Plenum, pp. 1-38.
SOBEL M.E. (2000). Causal inference in the social sciences, *Journal of the American Statistical Association*, 95(450), 647-651.
STOCK J.H., WATSON M.W. (2003). *Introduction to Econometrics*, Addison-Wesley, Boston.
WILLIAMSON J. (2005), *Bayesian nets and causality. Philosophical and computational foundations*. Oxford University Press, Oxford.
WILLIAMSON J. (2006a), Causal pluralism versus epistemic causality. *Philosophica*, 77, pp.69-96.
WILLIAMSON, J. (2006b), Dispositional versus epistemic causality. *Minds and Machines*, 16, pp. 259-276.
WINSHIP C., MORGAN S.L. (1999). The estimation of causal effects from observational data, *Annual Review of Sociology*, 25, 659-707.
WOODWARD J. (2003), *Making Things Happen. A Theory of Causal Explanation*, Oxford University Press, New York.
WUNSCH, G. (2007). Confounding and Control. *Demographic Research*, 16: 15–35.
WUNSCH, G., RUSSO, F., MOUCHART, M. (2010). Do we necessarily need longitudinal data to infer causal relations?. *Bulletin ~~the~~ de Méthodologie Sociologique*, 106(1), 1-14.