Correlational data, causal hypotheses, and validity

Federica Russo
Philosophy, University of Kent
f.russo@kent.ac.uk
25 October 2010

Abstract A shared problem across the sciences is to make sense of correlational data coming from observations and/or from experiments. Arguably, this means establishing when correlations are causal and when they are not. This is an old problem in philosophy. This paper, narrowing down the scope to quantitative causal analysis in social science, reformulates the problem in terms of the validity of statistical models. Two strategies to make sense of correlational data are presented: first, a 'structural strategy', the goal of which is to model and test causal structures that explain correlational data; second, a 'manipulationist or interventionist strategy', that hinges upon the notion of invariance under intervention. It is argued that while the former can offer a solution the latter cannot.
Keywords Causal hypotheses; Causal modelling; Causation; Correlation; Manipulationism; Intervention; Mechanism; Recursive decomposition; Structural modelling; Validity.

1. Introduction

The sciences try to make sense of correlational data coming from experiments and/or from observations. 'Making sense' arguably means deciding whether correlations found in a data set are causal or not, which is, needless to say, an old and hotly disputed issue. The goal of the paper is to reformulate the question 'when are correlation causal?', which has been given all sort of answers without reaching any definite agreement, in terms of the question 'when is a statistical model built to make sense of correlations valid?', which, on the contrary, has not received the right kind of attention yet.

In order to do that, I narrow down the scope to quantitative causal analysis in social science. Quantitative causal analysis is a tradition of scientific inquiry that traces back to the pioneering works of Quetelet (1869) and Durkheim (1895 and 1897) in demography and sociology respectively. Blalock (1964) and Duncan (1975) made significant improvements, and since then quantitative analysis has shown noteworthy progress. Nowadays, there is a variety of statistical models available in demography, sociology, or econometrics—see for instance the recent handbook edited by Kaplan (2004). There are currently two main traditions in quantitative research: one in economics / econometrics—see for instance the works of Hoover (2001) and Heckman (2008)—and the other focusing more in graphical methods and appealing to artificial intelligence techniques—see for instance the works of Pearl (2000) and of people at Canergie Mellon (e.g., Spirtes et al. (1993)). A survey and discussion of this 'methodological pluralism' in social research is offered in Russo (2006) and Russo (2009a, ch.3). Although methods may vary a great deal in, say, estimation methods, ultimately they all aim to answer the

same question—that is to make (causal) sense of correlational data—and, I argue, this is a question about the *validity of the model*.

It is commonly agreed that causal inference in well-designed experiments is more reliable than in purely observational contexts. In fact, in laboratory experiments, ideal conditions are more often met because uncontrolled variations in the environment are much better known. On the contrary, if the social scientist cannot directly manipulate the data, how can she infer what is the causal story behind the observed correlations? What is peculiar to the social sciences is that they most often deal with observational rather than experimental data.[1] The usual situation the social scientist is confronted with is having a data set coming from surveys, interviews, or censuses, and no possibility to directly intervene on the data. Granted, to some extent social scientists do interact with subjects, for instance when they give them questionnaires; however, these situations do not count as experiments as such. Also, I am not claiming that only the social sciences are observational—astronomy also faces the same kind of problems I am concerned with in this paper.

Let me emphasise that the paper does not investigate what is the extra-content to add to correlations to make them causal—which is a question about the *metaphysics* of causality. The metaphysics of causality is interested in establishing what causality in fact is, and what are the concepts in terms of which causality has to be cashed out. Instead, the question here addressed is how do we know or how do we test that correlations are (plausibly) causal—which are questions about the *methodology/epistemology* of causality. Thus the methodology and epistemology of causality are concerned with developing and implementing successful methods for the discovery and confirmation of causal relations and with the notions guiding causal reasoning in these methods.

The paper tackles the question of evaluating the validity of the model by comparing and contrasting two possible strategies: the 'structural strategy' and the 'interventionist/manipulationist strategy'. The paper is organised as follows. *First*, it considers the 'structural strategy'. The structural strategy is presented as the broad methodological framework embracing various statistical methods. It is argued that in quantitative social science, to decide or to establish whether correlations are causal is to establish whether a model is valid or not. The intermediate steps are given by the tasks of modelling and testing causal structures—i.e. mechanisms—that supposedly explain the correlations found in the data. *Second*, it considers the 'interventionist/manipulationist strategy'. In this approach the key notion is 'invariance under intervention', cashed out as follows. If the relations are causal, intervening on the cause-variable, would produce changes in the effect-variable, and the relation between the cause- and the effect-variables would remain stable. I point to two types of problems. On the one hand, if the project is given a purely metaphysical reading, then it lacks the complementary

---

[1] True, there are areas in social science where experimental methods are becoming increasingly popular. Yet, randomised experiments may raise more troubles than the ones they are actually meant to solve. The reason is that special care is needed in setting up experiments in social contexts because the intervention that is supposed to test a putative causal relation may change the structure altogether. This problem in social science is known at least since the so-called Lucas critique (Lucas 1976). For a discussion of structure-altering interventions in social contexts, the interested reader may also look at Steel (2008a, ch.8).

methodological story of how to evaluate the validity of the causal model, which is instead supplied by the structural strategy. On the other hand, if the project is mainly about the methodological implications about testing—i.e., it says how things would or must change in response to certain interventions—then the interventionist / manipulationist framework is insufficient in order to evaluate the validity of the model in observational contexts.

The main reason to reformulate the old question 'when are correlations causal?' into 'when is a statistical model valid?' is that the former conveys the misleading idea that *a* condition, or a set of necessary and sufficient conditions, can ensure the inference from correlation to causation. More specifically, as the recent literature polarises around the 'interventionist strategy', the notion of invariance under intervention seems to assume the role of 'condicio sine qua non'.

There are two take-home messages to get from the comparison of the two strategies. *First*, to decide whether correlations are causal is to evaluate the validity of the *whole* model. The evaluation of the validity of the model goes well beyond results of statistical tests—background knowledge and an explicit mechanistic interpretation loom large in this. *Second*, and as a consequence from the first, whether correlations are causal cannot be decided just on the basis of results of statistical tests, let alone just on the basis of *one* condition, notably the condition of invariance under intervention, which is so much acclaimed by manipulationist modellers.

2. The structural strategy

The 'structural strategy' is the general methodological framework embracing various statistical methods, and it provides the general principles of the test 'set-up' for causal hypotheses. The expression test 'set-up' is preferable because, as I shall explain later, causal hypotheses cannot be tested by *a* condition (or a finite set of necessary and sufficient conditions). The structural strategy is an account of the model-building and the model-testing process through which causal hypotheses are formulated and put forward for empirical testing and the whole model's validity is evaluated. I present the structural strategy in four steps. *First*, I explain that the goal of the structural strategy is to look for structures, i.e. for mechanisms. *Second*, I give the general features of how these mechanisms are modelled. *Third*, I discuss the kind of tests and considerations needed in order to (dis)confirm causal hypotheses and thus, *fourth*, to decide whether the model is valid.

2.1 Looking for structures

Structural modelling means looking for structures, i.e. for mechanisms. Before qualifying this claim in more detail, let me warn the reader of a possible confusion. The philosophical and scientific literature on causal modelling often uses 'structural modelling' and 'structural equation modelling' as synonyms. However, these are not coextensive terms and their linguistic proximity is the source of misunderstandings. Structural *equation* modelling is a particular type of statistical model used in quantitative social science, especially in econometrics. Structural modelling, instead, does not denote a particular (statistical) model (e.g., structural equation models, covariance models,

multilevel models, etc.) but refers to a general methodological account of model-building and model-testing. Structural modelling, thus, is a general methodological framework for causal analysis. This distinction between a particular statistical model and a general methodological framework is often not clear enough.

For instance, Woodward (2003, ch.7) devotes a long chapter to 'structural models'. However, what he discusses are, in fact, structural *equation* models, focussing on regression techniques and on the invariance condition. His account does not do justice to the variety of statistical models used in quantitative causal analysis, it does not present a *general* methodological framework for causal analysis, and it puts emphasis on the *counterfactual* definition of invariance. However, not all causal models are based on counterfactuals, and invariance is not the only feature making them 'causal'.

Pearl (2000) also deals with structural models sometimes meaning structural modelling and sometimes more specifically structural equation modelling. More recently, Pearl (2011) offers a 'structural theory of causation', that is a 'general theory of causation'. I take it to mean a 'general methodological framework' for causal analysis. In the paper, he develops a formalism that is general enough to subsume, as special cases, particular models such as structural equation models, potential outcome models, and graphical models. The core idea behind the formalism is that information about how the probability distributions over a set of variables would differ if external conditions were to change is given by causal assumptions; thanks to those assumptions, we can identify relationships that remain invariant when external conditions change. Pearl's approach very well complements my arguments, by offering the technical details that I will simply evoke later in section 2.2 and 2.3. Nevertheless, what is left unexplained in Pearl's approach is what makes his structural theory *structural*. This I attempt to do in section 2.1 and 2.2 in advancing an explicit *mechanistic interpretation* of structural modelling. Moreover, all along his work, Pearl distinguishes three types of queries: (i) about the effects of potential interventions (ii) about counterfactuals, and (iii) about direct and indirect effects. According to the methodology of structural modelling detailed throughout section 2, some structure needs to be identified *in order to* answer any of the queries before. Such structure, or mechanism, is identified by the recursive decomposition, as explained in section 2.3.

Nancy Cartwright has also extensively written on quantitative causal analysis, although she tackles questions different from mines. In *Hunting Causes and Using Them*, Cartwright discusses econometric techniques and theoretical models used for the purposes of causal analysis in economics. The scope of her discussion is thus much more restricted than mine, as the general methodological framework hereby presented also embraces disciplines in social science outside economics. Yet, the conclusion of one of her arguments indirectly supports mine. Cartwright criticises the idea that controlled experiments are the 'ideal' test for causal hypotheses. In particular, she is against the claim that sometimes we do not need to run the experiment, as Nature does it for us. Notice that, by and large, manipulationist modellers to be discussed in section 3 would subscribe to this claim too. But Cartwright is concerned that we are very rarely in the conditions to use this argument. It is a matter of fact—argues Cartwright—that causal hypotheses are difficult to test. Hence, we need reasons that are *outside* the regime over

which the test is conducted in order to draw causal conclusions. As the same worries apply to testing hypotheses and drawing inferences in lab setting, Cartwright's argument can be seen as a jolt to the view that experimentalism is the panacea for all problems in causal inference, which is indeed the critical target of section 3. The positive argument in favour of the idea of 'testing validity' developed in section 2.4 is more similar in spirit to her discussions on external validity. In addressing question of 'use' of causal claims, Cartwright (2011) challenges the widespread view that studying invariance properties of the probability distribution describing the population under analysis is no guarantee of the external validity of causal claims. Simply put, Cartwright is worried that even when we can single out very stable and invariant properties of a population, this does not ensure that the claim will be valid 'outside' the data being analysed. External validity requires much more than simply statistics and probability, notably it requires establishing what she calls 'tendency claims'. Section 2.4 offers an analogue argument for internal validity.

Finally, Hoover (2011) takes the goal of quantitative causal analysis to represent causal relationships by invariant parametrizations of a system of equations. This, however, does not imply endorsing altogether a counterfactual approach, and in fact Hoover criticises Woodward for impoverishing the analysis and bestowing too much importance to counterfactual manipulability. As it will be clear from the arguments given in section 3, I take side with Hoover on that. However, when Hoover presents the structural account as based on Herbert Simon's causal ordering (see, e.g., Simon (1953 and 1954) and Fennell (2011)) and on the condition of exogeneity, the question that is left unanswered is, again, what makes a structural account *structural*.

In the following, I focus on some selected aspects of structural modelling. In particular, I am interested in dwelling on the following interrelated ideas: (i) structural models model mechanisms, (ii) to model mechanisms means to formulate suitable causal hypotheses to put forward for empirical testing, (iii) to decide about the results of tests is to decide whether the model is valid. [2]

Two remarks are in order. *First*, the philosophical and methodological literature has devoted some room to the relation between structural (equation) modelling and mechanisms, but not in the way the connection between the two is developed here. For instance, classical economists such as Adam Smith, David Ricardo, Thomas Malthus and John Stuart Mill or the Chicago School of Economics developed the economic theory so that the theory *dictates*, rather than model, what the mechanism is. More recently, in Kevin Hoover's approach, the causal structure represented by a set of structural equations is a "network of counterfactual relations that maps out the underlying mechanisms through which one thing is used to control or manipulate another" (Hoover 2001, p.24)—a view that falls under the 'interventionist/manipulationist strategy' to be discussed later in section 3. Also, in the tradition of statistics and econometrics, it is usually said that the goal is to model the data generating process. The specification of the data generating process, in turn, depends on the statistical model, and, consequently, the assumptions of the model play a decisive role. Freedman (2004), for instance,

---

[2] The reader interested in more technical and formal aspects of structural modelling, especially related to the condition of exogeneity and to recursive decompositions, may look at Mouchart et al. (2009) and Mouchart and Russo (2010).

distinguishes between statistical and causal assumptions and requires interventions to grant causal inferences. The crucial assumptions, in his account, are the causal ones, which eventually consist in *assuming* that structural equations give the causal mechanism that generate the observed data. However, none of the available approaches explains how we get a mechanism out of a statistical model, which is exactly what the 'structural strategy' hereby presented instead does.

Second, the philosophical and methodological literature interested in problems of testing causal relations (rather than *defining* causality) has narrowly focused on just *one* condition, i.e. invariance under intervention. It is this narrow focus that I criticise, and, as an alternative, I offer the 'structural strategy', which instead aims to evaluate the *validity of the whole model.*

It is worth emphasising that structural modelling aims, from the very start, to go beyond mere descriptive knowledge of a phenomenon. It aims, in fact, to unveil—in the sense of modelling out—the mechanism that brought it about. It is in this sense that structural modelling means looking for structures, in particular, the causal structures or mechanisms underlying the correlational data we want to make sense of. Thus the mechanism is not simply assumed or imposed by theory. The mechanism is exactly what the structural modeller tries to build out of background knowledge and what she tests against empirical data—the modelling and testing aspects of mechanisms will be discussed later in section 2.2 and 2.3.

The qualification of this last claim, notice, does not depend on the specific definition of mechanism (that is on the metaphysical account of mechanism) one might offer or endorse, but rather on the more general—and arguably more widely accepted—position that mechanisms carry explanatory power. Thus, to say that structural modelling looks for structures is in line with the characterisation of Machamer, Darden and Craver (2000), in that the structure being modelled assemblies 'things' (entities, in MDC vocabulary; variables, in the jargon of statistical models) that interact with and influence each other in a specified way (the activities, in MDC vocabulary; the statistical relevance relations, in the jargon of statistical models). It is also in line with the characterisation of Bechtel and Abrahamsen (2005), who put a lot more emphasis on the structure, organisation, and decomposition of the mechanism. However, this 'mechanistic interpretation' of structural models is at variance with the characterisation of Woodward (2003), who conceives of mechanisms as chains of invariance relations.

Structural modelling makes no ontological commitment as to the (degree of) physical existence of mechanisms. In other words, the choice of a particular ontological account of mechanisms is perpendicular to the epistemological issue of the explanatory power of mechanisms. The understanding of mechanisms in structural modelling is rather *epistemic*—it is more concerned with how we can (causally) make sense of correlational data. Structural modelling achieves this goal by offering a story about, or a description of, a mechanism. Such epistemic understanding is akin to 'mechanism schemata', as discussed in Machamer (2004), Machamer, Darden and Craver (2000), and Darden and Craver (2002). Simply put, mechanism schemata are abstract and idealised descriptions of a type of mechanism; in other words, mechanism schemata are concerned with the description of the behaviour of the mechanism. Yet, in looking for mechanisms that

explain correlational data, structural modellers do track something real—that is what *actually* happens, rather than what would happen. This brings me to the next point.

More importantly, this understanding of mechanisms in structural modelling is in the tradition of "making sense of causation in terms of 'what actually happens' without appealing to counterfactuals" (Woodward 2004, p.51), a program that Jim Woodward ascribes to Jim Bogen (2004) and Peter Machamer (2004) and, more generally, to what he calls the "anti-counterfactualist Pittsburgh tradition" (Woodward 2004, p.43). This last remark is important to understand the project I undertake in the remainder of the paper. The question being asked is not what is the right definition of 'causality' or of 'causal relation' or of 'mechanism'. Instead, recall, what I am interested in is the following issue: in social science contexts, how can we best make sense of correlational data coming from observations? In other words, how do we establish whether there is in fact some causal structure in the data?

Consequently, the question immediately arises as to *how* one can find out about such causal structures or mechanisms. The works of e.g. William Bechtel, Carl Craver and Lindley Darden (Bechtel (2008), Craver (2007), Darden (2006)) discuss methods to find out about mechanisms in other contexts, notably in neuroscience and biology; instead, the structural strategy discussed here, in Russo (2009a) and in Mouchart and Russo (2011) is a methodology to find out about mechanisms in social science.

2.2 Modelling structures

Modelling structures, or mechanisms, is performed in three stages. In a first stage, causal hypotheses are formulated and a statistical model is built. In a second stage, causal hypotheses are tested in the statistical model. And, finally, results of tests are evaluated to conclude to the validity or invalidity of the model.

This broad characterisation, however, hinges upon 'causal hypotheses' rather than mechanisms. Here is how the two are linked. Causal hypotheses formulated in order to make sense of correlational data are not simply of the form '*X* causes *Y*'. The structural modeller formulates a whole set of hypotheses and makes a number of different assumptions that, altogether, have to be interpreted as hypothesising the mechanism explaining the correlations in the data. The assumptions made in a structural model are themselves of different types. Some have merely *statistical* import (e.g., the assumption about normality of the distributions of the variables figuring in the equations), whilst others have more profound *causal* meaning (e.g., the non-correlation between the putative cause-variables and the errors figuring in the equations).[3] Altogether, the assumptions and the causal hypotheses are 'formally translated' in what the structural modeller calls the 'recursive decomposition'.[4] I direct the reader interested in the technical aspects of the recursive decomposition to Mouchart and Russo (2011), and give here the core idea distilled from complicated formulae.

---

[3] A thorough discussion of the assumptions of causal models is given in Russo (2009, ch.4).

[4] Structural modellers try, as much as possible, to build recursive or acyclic models. In simple terms, this means that, given a graphical representation of the mechanism by means of a set of vertices and of directed edges connecting one vertex to another, we do not run into a loop going through the various possible paths in the graph. As a matter of fact, this assumptions of acyclicity is often violated and in fact much research in methodology is devoted to develop models that cope with this aspect.

The structural modeller is usually provided with a data set containing observations of a certain number of variables and with the joint probability distribution over these variables—the goal is to dig out the causal structure underlying it. The initial joint probability distribution is thus factorised into a sequence of products of marginal and conditional components, where the variables to condition upon play the explicit role of the causes. The whole recursive decomposition may be interpreted as characterising a global mechanism, whereas each conditional distribution within the recursive decomposition may be interpreted as characterising a (autonomous) sub-mechanism within the global one. Thus, decomposing a global mechanism into a sequence of (autonomous) sub-mechanisms is tantamount to disentangling the action of each component in a sequence of the sub-mechanisms operating in a global mechanism.

Two remarks are in order. *First*, I say that the recursive decomposition *may be* interpreted as a mechanism. Formalisms are usually subject to more than one interpretation. For instance, structural equations and the associated decomposition may also be interpreted in terms of intervention: a simple structural equation $Y = \beta X + \varepsilon$ is typically read by manipulationist modellers as 'were we to intervene on X, Y would change accordingly'. But this isn't, by all means, the only or the most basic interpretation. The structural strategy hereby presented, in fact, does not take the 'interventionist' reading of the equations and of the recursive decomposition as the primary one. *Second*, many specific modelling tools such as structural equation models or Bayesian Nets decompose initial joint probability distributions into sequences of marginal and conditional components, exactly in the way I just described. The question is to what extent modellers use those *tools* (i) accompanied by a mechanistic interpretation of the decomposition and (ii) having in mind the *validity of the whole* model rather than one particular test condition (invariance for structural equation models and the Markov Condition for Bayesian Nets) to make causal sense of correlational data. The whole point of adopting the structural strategy is that it does endorse an explicit mechanistic interpretation of the decomposition and within this framework correlational data are given a causal interpretation if the *whole model* is valid in the sense explained in section 2.4.

Here is an example from social science research. López, Mompart, Wunsch (1992) conducted a study on regional mortality in Spain. Spain met deep socio-economic changes in the mid-Seventies: policy in that period tried to intervene to improve the social and economic situation, including sanitary infrastructures. The goal of the analysis was to explain an observed lower mortality rate at the time of the study. Background knowledge about the Spanish situation around the Seventies and about the types of policies implemented supported the choice of distinguishing the supply and demand of medical care (that is the infrastructure and its use are distinguished), unlike the majority of similar studies. In fact, previous studies in demography and medical geography examined the incidence of the health system on regional mortality coming to the conclusion that regional differences in mortality could not possibly be explained by regional differences in the health system. López, Mompart, Wunsch (1992), instead, hypothesised that regional mortality was influenced by the health system which was in turn influenced by the social and economic development.

They decided to model a structure made of the following five variables: economic development (X_1), social development (X_2), sanitary infrastructure (X_3), use of sanitary

infrastructure (X_4), age structure (X_5), and mortality (Y). The initial joint probability distribution over the whole vector of variables was decomposed into two sub-mechanisms, where economic development and age structure are exogenous.[5] In plain English, this means that economic development influences social development and sanitary infrastructure, thus creating two distinct causal paths to regional mortality. Analogously, age structure influences the use of sanitary infrastructure, which in turn influence regional mortality. In the first, 'economic development' is the exogenous variable influencing mortality through 'social development' and 'sanitary infrastructure'; in the second, 'age structure' is the exogenous variable influencing mortality through 'use of the medical care system'. The corresponding directed acyclic graph would look like figure 1.

Two remarks are now in order. *First*, it may be objected that it is arbitrary to choose a particular recursive decomposition over another. In other words, why would the acyclic graph in figure 1 be a faithful representation of the mechanisms underlying the correlational data we want to make sense of? Figure 1 represents the *hypothesised* mechanisms. The mechanism then underwent testing and, based on the results and interpretations of testing, it was validated in a sense to be better specified later in section 2.4. This is to say that there is no a priori way to know whether the arrows in figure 1 really are causal rather than just correlational. The proof of the pudding, as we say, is in the eating. *Second*, and related to the remark above, it may asked why some variables are not linked at all; that is, why does figure 1 exclude that X_2 causally act on X_4 (or vice-versa)? Again, there is no principled reason why the two are not (causally) linked in either direction. Background knowledge and preliminary analyses of data suggested, in *this* case, that there was no action of X_2 on X_4. But this is by no means a universal truth. The situation may be different in another population. A different analysis of the same data may even question such modelling assumption. The question is, again, to what extent the mechanism thus modelled is validated in the sense to be further discussed in section 2.4.

Fig. 1. Regional mortality in Spain

---

[5] Exogeneity is a thorny issue for structural modellers. Intuitively, when we say that a variable X is exogenous for a variable Y, this means that some conditions about the parameters of X and Y are satisfied such that it is legitimate to interpret X as the cause and Y as the effect. It is in this sense that it is commonly said that exogenous variables are generated outside the model, i.e. they are not caused by other variables in the model. For a discussion, see for instance Mouchart and Russo (2011).
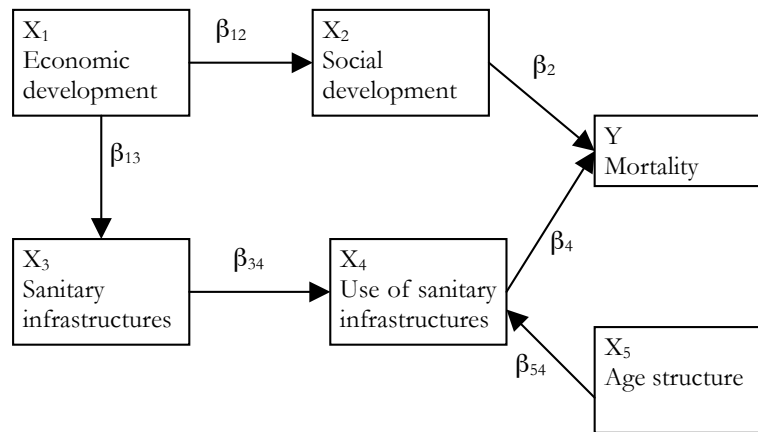
It is for the purpose of testing that the whole set of hypotheses and assumptions is broken down in smaller pieces, so to speak. For instance, tests may specifically concern whether a parameter is significant, or aim to estimate the causal impact of variable X_1 on variable Y, or aim to test whether the relation between variable X_1 and Y is invariant, etc.

Contents and scope of tests are discussed later. The key question, at this stage, is *how* causal hypotheses are formulated. Structural modelling makes prominent use of background knowledge.[6] Background knowledge is notoriously a hard to define and elusive notion. Here is an attempt to specify what qualifies as background knowledge. It certainly includes available scientific theories but it may also include general knowledge of the socio-political context, knowledge of demographic characteristics of the population under investigation, or 'institutional' knowledge (i.e., knowledge of the functioning and procedures of an institution such as the Central Bank). Background knowledge, however, is more than knowledge of 'the social'. In fact, in areas such as epidemiology, we need to include biological variables too, and background knowledge will also include biomedical knowledge. It is worth noting that 'knowledge of the social' and 'knowledge of the biological' are not completely or necessarily separated and independent. For instance, in many cases background knowledge includes knowledge of mixed mechanisms. Simply put, in mixed mechanisms, the social is used to explain the biological, and vice-versa. These kinds of 'mixed' mechanisms are also one of the sounding boards for the evaluation of the validity of the model, as we shall see later in section 2.4.

Evidence is also important for background knowledge. Notably, evidence of the same putative mechanism operating in different populations may justify further research, or evidence about different mechanisms operating in other populations may justify a different modelling strategy. In short, the use of different/similar data and/or models also belongs to background knowledge.

Background knowledge thus participates in the formulation of causal hypotheses because it helps with the choice of variables and with the specification of the relations between variables. In other words, background knowledge helps in modelling the

---

[6] Philosophers of science have been long debating on the status of models and on their relation to theory and to background knowledge. It is widely agreed that models are not tested by empirical data alone and that background knowledge and general theories contribute to model-building and model-testing also in scientific domains other than the one discussed here. The interested reader may have a look at Frigg and Hartmann (2009) and at the references therein, in particular Morgan and Morrison (1999).

mechanism. The later stage of 'testing structures' is meant to (dis)confirm this hypothesised mechanism.

Since so much importance is bestowed to background knowledge, the objection of circularity is around the corner. If causal hypotheses are formulated mainly on the basis of background knowledge, what do we actually learn from the structural model that we did not already know? Put differently, how does structural modelling take us beyond background knowledge? There are two answers to this question.

One is that (*pace* classical economists) background knowledge *suggests* a causal mechanism that possibly makes sense of correlational data, but tests are required in order to accept what background knowledge suggests—this is to decide whether the model is valid. However, background knowledge does not *dictate* what hypotheses to make, and in fact the hypothesis formulation stage also includes preliminary analysis of data—a stage that has been also called 'data snooping' (Freedman 2005). Notice, however, that this liberty of making wise and responsible use of background knowledge is at variance with much of scientific practice in economics, where the economic theory gives altogether the hypotheses and the model, and all is left is to perform statistical tests to check to what extent data conforms to the theory. This trend, however, is the object of critiques (see for instance Hoover (1988), Moneta (2007), Chao (2009)).

The other answer is that structural modelling is a dynamic process in which we witness a *va et vient* between established knowledge and knowledge yet to be established. Thus, background knowledge is essential in order to formulate causal hypotheses, but, in turn, results of tests may question and even discard background knowledge, thus triggering further research. That is to say, structural modelling does not establish immutable and eternal (causal) truths.

2.3 Testing structures

The next stage is 'testing the structure'. In quantitative social science making use of statistical models, important tests concern goodness of fit and significance of parameters. Leaving technicalities aside, tests for goodness of fit indicate how well the equations in the statistical model provide an accurate description of the data. It may happen, in fact, that the hypothesised causal mechanism be theoretically plausible, but the equations poorly describe data due to, for instance, problems of confounding. This typically happens, to put it roughly, when a variable Z is a common cause of both the putative cause X and of the effect Y, thus 'confounding' the relation between X and Y.[7]

Nevertheless, what I am most concerned with is the so-called condition of invariance or stability. Standard quantitative social science methodology requires that, in order to interpret correlations causally, the relations between variables described by the equations in the model be stable in a sense to be specified. In experimental contexts, this would turn out to be a relatively easy task. If we want to know whether X causes Y, we wiggle X holding fixed everything else around X and then see what happens to Y. However, in purely observational contexts, this isn't possible. We are provided with a data set on which we cannot directly intervene manipulating and changing the values of the

---

[7] See for instance Bollen (1989) and Wunsch (2007).

variables. And even when 'social' interventions are performed, i.e. a 'policy experiment' is carried out, such interventions may raise difficulties in the stability of the system, a problem pointed out by Lucas as early as 1976.

The way out, then, consists in seeing whether the relation between the cause- and effect-variable(s) holds stable in different partitions of the data set or in different panel data used in the study. An invariance test, then, is not about whether the relation between X and Y *would* remain stable *were we* to wiggle X. Instead, the test is whether the relation *actually* remains stable when we consider different portions of the data set, or different data sets with observations from different populations or different time periods. This is how, in the spirit of the 'anti-counterfactualist Pittsburgh tradition', the structural modeller makes sense of what *actually* happens.

It is worth emphasising that invariance is *a test* to check whether correlations satisfy certain conditions that warrant the causal interpretation. According to the structural strategy, invariance is *not* an identity conditions for causal relations, nor does it exhaust the question of making causal sense of correlational data. I shall get back to this point in next section, but let me anticipate what I think are the issues at stake. I am not denying that invariance play a role in modelling. Indeed it does, but (i) it is not implied by the structural strategy that causality is metaphysically cashed out in terms of invariance, and (ii) invariance is but *one* among the tests.

2.4 Validity

The final stage is to decide whether the model is valid. To state it in non-formal terms, this means to decide whether the story about the mechanism supposedly making sense of correlational data provides a plausible enough explanation about what is really going on in the world.

In the social science literature, the locus classicus is the work of Cook and Campbell (1979). Cook and Campbell (1979) borrow from Campbell and Stanley (1963) the terms 'internal' and 'external' validity, which refer to the best available approximation of the truth of causal statements. They present quasi-experimental designs to be applied in a variety of research settings. In their view, the design is meant to probe causal hypotheses and to this end it has to be evaluated with respect to four types of validity: statistical, internal, construct, and external validity. Simply put, given a particular study, this will be statistically valid if statements about covariation can be made with reasonable confidence, internally valid if a causal relation is confirmed within the specific population at hand, constructively valid if alternative constructs for cause- and effect-variables deliver consistent results, and it will be externally valid if the results can be generalised to other populations. Each type of validity raises issues that are equally important, but I shall concentrate here on internal validity.[8]

Cook and Campbell offer a thorough discussion of the threats to internal validity. They mention, in particular, knowledge of time ordering, sources of random error, and

---

[8] A general discussion of validity in causal analysis in social science is given in Russo (2009, ch.3). The reader interested in external validity—an emerging topic in the philosophical debate—may also have a look at Guala (2005), Steel (2008a), at their two contributions at the PSA 2008 (Guala 2008 and Steel 2008b) and at Jiménez and Miller (2010).

the appropriate use of statistics and of statistical tests. There is, however, an aspect overlooked in their account, namely congruence with background knowledge and explanatory power. I want to argue, in fact, that to decide whether a model is valid or not, goes beyond 'statistical' considerations. Given that the goal, recall, is to make sense of correlational data, we have to make sure that the story being told takes into account where the data comes from (background knowledge) and that the story actually is a plausible (mechanistic) explanation of the correlations found in the data. Let me clarify this idea further.

Of course, there is no ready-made recipe. However, the structural strategy here presented, recall, aims to make sense of correlations by hypothesising and testing a mechanism that thus explains a given phenomenon. This means that, besides the 'statistical considerations' emphasised by Cook and Campbell, validity is a matter of assessing the goodness of explanation.

Such evaluation is certainly concerned with statistics, to some extent. In statistics textbooks, in fact, explanatory power is (statistically) measured by a coefficient (called coefficient of determination) quantifying the amount of variability of the 'effect-variable' that is taken into account by including the chosen 'cause-variables' in the model. However, this coefficient should be seen as a measure of goodness of fit rather than of explanatory power. Whether the model convincingly makes (causal) sense of correlations ought to be assessed on more *theoretical* grounds. Those theoretical grounds concern (i) 'congruence with background knowledge' on the one hand, and (ii) 'ontological homogeneity' in the mechanism on the other hand.

First, as mentioned above, the hypothesis formulation stage quite importantly depends on background knowledge. But I also said that background knowledge suggests, rather than dictates, the hypothesised mechanism. We started with background knowledge and we have to go back to it. Thus, the structural modeller has to answer a number of questions. For instance, does the model give a causal story congruent with what we know or it doesn't? If not, is it more plausible to reject the model or to question background knowledge? Background knowledge plays the role of an additional sounding board for the model, but we have to be prepared to question background knowledge too.

Second, if ontological homogeneity between the variables acting in the mechanism is lacking, it may be desirable to identify and justify indirect paths from the causes to the effect. Let me explain this idea. In many domains of social science, for instance demography or social epidemiology, to make sense of correlational data modellers may need to include variables of very different nature. For instance, Mosley and Chen (1984) explain child survival in developing countries by means of both social and biological factors. This means modelling a mechanism that is 'mixed', so to speak. Now, to say that a social factor such as maternal education influences a health variable such as child survival has theoretical plausibility only if we can specify *how* this is possible, namely by identifying indirect paths from the cause(s) to the effect. The identification of indirect paths is of course dependent on available background knowledge, data, and modelling tools. That is to say, the more we can specify the mixed mechanism, the better for explanatory purposes. The interested reader may look at Russo (2009a, ch.6), Russo (2009b), and Mouchart and Russo (2011) for further details.

The case study on regional mortality in Spain discussed earlier well illustrates the two theoretical aspects to be considered in evaluating the validity of the model. In fact, systematic confrontation with background knowledge, both at the hypothesis-formulation and at the testing stage, bear out the plausibility of the mechanism. Also, although there are more blatantly cases of mixed mechanisms, the one modelled by López, Mompart, Wunsch (1992) involves variables that are heterogeneous enough (e.g., age structure and social or economic development) to illustrate my point. Without those considerations, positive results of statistical tests on the decomposition (including invariance tests) are insufficient to establish whether the hypothesised mechanism causally makes sense of the correlational data the researchers had to analyse.

In sum, the social scientist adopting the structural strategy will make sense of correlational data in the following way. Out of background knowledge, including field knowledge and previous studies, and of preliminary analyses of data she will first hypothesise a mechanism to be (dis)confirmed by means of empirical testing. The main test concerns the actual (rather than counterfactual) stability or invariance of the relation between the cause- and the effect-variable(s). However, to decide whether the mechanism provides a good enough causal story for the correlations found in the data set is to decide whether the *whole model* is valid or not. The evaluation of validity, in turn, goes beyond mere statistical considerations about, say, sources of random error, but also has to invoke *theoretical* considerations about congruence with background knowledge and about the ontological homogeneity of the mechanism.

3. The interventionist/manipulationist strategy

Let me now present and discuss an alternative strategy to make sense of correlational data. This is the interventionist or manipulationist account. Mainly defended by Jim Woodward (1997, 2000, 2002, 2003, 2004), the approach is also advocated, among others, by Hausman (1986, 1997), Hausman and Woodward (2004), Woodward and Hitchcock (2003), or Pearl (2000), and, in the social science literature, by Heckman (2008) or Morgan and Winship (2007). In a nutshell, this is an account of causality and of causal explanation that hinges upon the notion of invariance under intervention, which is, in turn, *counterfactually* defined. I begin by spelling out the notion of invariance under intervention as it is the hub of the interventionist/manipulationist strategy. I then discuss the role of experiments in manipulationism to assess whether and to what extent this strategy offers a solution to the problem of making (causal) sense of correlational data in social contexts where data is most often observational. Finally, I point out that even a staunch supporter of manipulationism such as Jim Woodward needs to develop a notion of invariance under intervention for observational studies that is not manipulationist in character.

3.1 Invariance under intervention

'Invariance under intervention' is invoked by interventionist modellers because it provides a definition of causality, or it bestows empirical generalisations explanatory power, or some combination of the two.

In the Woodward-Hitchcock approach, in particular, conceptual and explanatory considerations intertwine. Here is an example of the conceptual considerations: invariance is what makes a generalisation *causal* rather than accidental.

> The notion of invariance is closely related to the notion of an intervention. A generalization *G* (relating, say, changes in the values of *X* to changes in the values of *Y*) is invariant if *G* would continue to hold under some intervention that changes the value of X in such a way that, according to *G*, the value of *Y* would change—"continue to hold" in the sense that *G* correctly describes how the value of *Y* would change under this intervention. […] The notion of *invariance under intervention is intended to do the work (the work of distinguishing between causal and merely accidental generalizations) that is done by the notion of a law of nature* in other philosophical accounts. (Woodward 2003, p.15-16, emphasis added)

Invariance under intervention, however, requires some further clarification, notably about the fact that, under an intervention, the generalisation "*would* continue to hold". That is to say, invariance is counterfactually characterised:

> We are now in a position to characterize the notion of invariance: a relationship *R* between variables *X* and *Y* is invariant if it would continue to be true (or approximately true) in at least some hypothetical situations or possible worlds in which the value of *X* is changed as the result of an intervention. That is, *there must be some non-actual value x of X such that the following counterfactual is true: 'if X were equal to x, then the values of X and Y would stand (approximately) in the relation R'.* (Woodward and Hitchcock 2003, p.15, emphasis added)

> As I have emphasized, the notion of invariance is a modal or counterfactual notion: it has to do with whether a relationship would remain stable if, perhaps contrary to actual fact, certain changes or interventions occur. (Woodward 2003, p. 279)

Now, all this becomes relevant for explanation. In the Woodward-Hitchcock approach, in fact, there is a shift from from nomothetic laws (typical of the deductive-nomological model) to empirical generalisations that have the ability to support a particular kind of counterfactuals in virtue of their being invariant. And this is what gives empirical generalisations explanatory power:

> Explanation has to do with the exhibition of counterfactual dependence describing how the system whose behaviour we wish to explain would change under various conditions. As we will see, whether a generalization can figure in such a pattern of dependence and hence can be used to explain has to do with whether it is invariant, rather than with whether it is lawful. (Woodward and Hitchcock 2003, p.2)

Let us step back now. We started with a problem that is widespread across the sciences: to make sense of correlational data coming from observations or experiments. I then suggested focusing on quantitative social science, where data most often comes from observations and therefore cannot be manipulated. How does the interventionist/manipulationist strategy fare with this challenge?

3.2 Manipulationism and experiments

The interventionist/manipulationist modeller would answer by saying that what is required is that correlations found in the data set be sufficiently stable under intervention, in the sense explained above. The question then arises whether, since what is required is invariance under *intervention*, the manipulationist approach presupposes an experimentalist stance. The question is legitimate since, if it does, manipulationist

strategies are in a hard position for making sense of correlational data in observational contexts.

So we first have to assess whether and to what extent manipulationist modellers appeal to experiments. At times, there seems to be a strong reliance on experiments:

> […] the kind of counterfactuals that are relevant to understanding causation are connected to *experiments*—either actual or hypothetical. […] counterfactuals are understood as claims about what would happen if a certain sort of experiment were to be performed. (Woodward 2002, p.4, emphasis in the original)

What is the real weight of experiments in the manipulationist approach, then? In order to respond this question, we have to consider the status of the project itself. I shall consider below two possibilities: (i) the manipulationist project is essentially conceptual, i.e., it aims to provide a metaphysical account of causation and explanation, and (ii) the manipulationist project is not entirely conceptual—it rather aims to provide an account of how manipulation and experiments enlighten explanatory tasks. It is worth discussing both variants of manipulationism for several reasons.

One reason is that it is controversial whether manipulationist modellers, and in particular Woodward, support either version, in spite of the claims they make in various places. A number of authors read the project as a contribution to the metaphysics of causality (Campaner 2003 and 2006, Galavotti 2001, Strevens 2007 and 2008). In particular, Strevens argues that the metaphysical reading is the only cogent way to make sense of the manipulationist programme, and yet Woodward (2008) vehemently rebuts casting doubts on the sole and main metaphysical import of the manipulationist approach and reinforces the methodological import of the programme. Given the disagreement on the status of the project, it is worth examining both options.

Another reason is that choosing either version has consequences for my argument. On the one hand, structural modellers may well live with conceptual manipulationism and supply the methodology of model-building and model-testing that conceptual manipulationists lack, but this would imply that metaphysics and methodology are quite unrelated. Arguably, this is not a desirable situation. On the other hand, the methodological story about model-building and model-testing offered by explanatory manipulationists proves wanting in the contexts analysed in this paper, and this is where structural modellers take over.

Finally, because the interventionist approach is advocated by some methodologists in social science too (see, e.g., Morgan and Winship 2007), it is important to decide whether they are proposing too narrow a strategy, unable to make sense of correlational data in observational contexts.

Conceptual manipulationism

Let us consider the first possibility. The project is conceptual, i.e. it is about the *nature* of causation. This means that the account aims to say what is the extra content of causation with respect to correlation and to give *identity conditions* for causal relations.

If this reading of the manipulationist approach is right, then presupposing experimentalism is not a problem directly relevant for the task of making sense of

correlations in observational contexts, because it aims to say what causality is, not how causal relations are to be established (which is the problem of this paper). Causality, in this approach, lies in manipula*bility*. So, in an experimental context, something is a cause if, *were we to intervene* on it, the effect would change accordingly. But the setting needs not be necessarily experimental. Even in an observational context one might support a metaphysical view according to which something is a cause if, under manipulation, it would yield changes in the effect(s).

One line of counter-argument that I will just mention in passing without further elaboration is that causation does not lies in counterfactuality. Here is as Jim Bogen (2005, p. 415) puts it:

> How can the chemicals in birth control pills cause non-pregnancy in fertile females, when they can't prevent non-pregnancy for males? Counterfactualists say it's because an invariant counterfactual obtains between non-pregnancy and interventions that introduce those chemicals in the female, but not in the male bloodstream (Woodward 2003, pp. 154ff., 172-173, 198-220). Counterfactual reasoning can be *epistemically* important to the discovery of causal structures. But Counterfactualism is not an epistemological idea. It is an ontological idea, or a piece of conceptual analysis to the effect that there is no causality without counterfactual regularities.

Another line of counter-argument that instead I want to pursue is the following. Arguably, a manipulationist metaphysics needs to be accompanied by a story about *testing* too. The reason for this would be that once we have an account of what causality *is* (identity conditions), it helps a great deal to provide an account of *how* to ascertain whether a relation is causal (test conditions). However, conceptual manipulationism lacks a sensible methodological story, which is instead supplied by the structural strategy.

The interventionist modeller might say that this is the stage where experimental design enters, in order to operationalise the variables and to perform experiments to test that relations between variables are invariant under intervention. Nonetheless, the problem with the interventionist strategy arises exactly at this point. Arguably, test conditions are different in experimental and observational contexts. In particular, as in observational contexts we cannot manipulate the variables to check out what changes would occur, we need an alternative test for invariance. Additionally, establishing causal relations requires more than testing *a* condition (or a finite set of conditions), but it calls for the evaluation of the validity of *whole model*, which includes confrontation with background knowledge, as discussed earlier in section 2.4.

Explanatory manipulationism

The second possibility is that the manipulationist project be not entirely conceptual. The focus would rather be *explanation*, namely *how* a phenomenon would change after certain interventions.[9] Here, there is very tight a knot between how- and why-questions. Invariance under intervention is first needed to distinguish accidental from causal generalisations. Then, the extra bonus we gain with invariance is that it also confers empirical generalisations explanatory power. In fact, invariant (under intervention) generalisations can then be used to ask *counterfactual* questions about the conditions under

---

[9] 'Explanatory manipulationism' is a coinage, as far as I know, of Strevens.

which the explanandum would have been different: these are Woodward's *withbd-questions*, i.e., what-if-things-had-been-different-questions.

Withbd-questions serve the role, in Woodward's account, to ensure that correlations are causal. The stock example is that of Mr Jones taking birth control pills and never getting pregnant. According to Woodward, we exclude that this relation is causal because to the question 'What would happen if Mr Jones hadn't taken birth control pills?' we still answer that he would have not got pregnant. That is to say, manipulating the cause-variable (taking birth control) does not lead to any change in the effect-variable (getting pregnant).[10]

Thus, there is a sense in which why-questions, in the manipulationist account, receive an answer through the 'how': why does X cause Y? Because the empirical generalisations between X and Y is invariant under intervention, and this allows us to ask counterfactual questions about what would happen to Y, were X be different. The extent to which this account would be favourably viewed by contemporary 'mechanistas' (see especially Bechtel's work)—who put so much emphasis on explaining a phenomenon by decomposing it—is beyond the scope of the present work.

If the 'explanatory' reading of manipulationism is correct, then experimentalism does play a decisive role for the problem raised in this paper. This central role of experiments is recognised by manipulationist theorists themselves (Woodward 2002, 2006), and by a number of commentators (Galavotti 2001, Psillos 2004, Strevens 2007 and 2008). But, if experiments are the keystone of explanation, then, I contend, explanatory manipulationism cannot be successfully applied to observational contexts.

Granted, manipulationists modellers do offer a story that suits experimental contexts. In particular, there is a sense in which the use of counterfactuals is pretty straightforward:

> My views about the use of counterfactuals in connection with understanding causal claims and problems of causal inference are grounded in pragmatic considerations, not apriori metaphysics. For it to be legitimate to use counterfactuals for these purposes, I think that it is enough that (a) they be useful in solving problems, clarifying concepts, and facilitating inference, that (b) *we be able to explain how the kinds of counterfactual claims we are using can be tested or how empirical evidence can be brought to bear on them*, and (c) we have some system for representing counterfactual claims that allows us to reason with them and draw inferences in a way that is precise, truth-preserving etc. I think that these conditions are met in many scientific contexts in which counterfactuals are used. (Woodward 2002, p. 4, emphasis added)

Consider a stock example of manipulationist modellers. Ohm's law states that the voltage E of a current is equal to the product of its intensity I times the resistance R of the wire. Consider now the two following counterfactuals:

> (1) If the resistance were set to $R=r$ at time t, and the voltage were set to $E=e$ at t, then the current I would be $i=e/r$ at t;
> (2) If the resistance were set to $R=r$ at time t, and the voltage were set to $E=e$ at time t, then the current I would be $i* \neq$ (*is not equal to*) e/r at t.

---

[10] The question arises whether we really need the whole manipulationist machinery in cases like this. It may be argued, in fact, that in such cases background knowledge is sufficient to exclude spurious causal relations. For a discussion of the explanatory import of manipulationism, see also Russo (2009b).

Non-trivial truth-values can be given to those counterfactuals "as long as we can describe how to test them" (Woodward 2004, p. 45). In experimental setups, it will suffice to make further measurements (if *t* is in the future) or to appeal to previous results (if *t* is in the past) in order to test the counterfactuals above. It might be argued that once we make an experiment *to test the counterfactual,* we appeal to *factual* evidence. So it seems that the *counter*factual element dissolves anyway. But I will not enter this quarrel here. The interested reader might refer to the thrust and counter-thrust between Bogen (2004) and Woodward (2004).

The manipulationist strategy apparently suggests that once you test for invariance the job is completed. But this is a wrong picture of scientific practice. First, as discussed in section 2.3, there are *many* tests to perform, even though invariance is arguably the most important one. Second, as discussed in section 2.4, to decide whether there is plausible causal story that makes sense of correlational data we have to evaluate the *validity of the whole model.* This is exactly what the structural strategy discussed above offered.

Manipulationist modellers may still play the card of ideal experiments. What is required is not to *actually* perform experiments, but to evaluate counterfactuals that evaluate *ideal* experiments describing hypothetical interventions. This move, however, begs the question and reintroduces the problem mentioned in 'conceptual manipulationism': even if we accept it as a conceptual analysis of causation, a story about evaluating the validity of a model to make (causal) sense of correlational data is missing.

At best the interventionist/manipulationist strategy can offer a test condition for experimental contexts. But as I tried to show earlier, making sense of correlations involves more than simply satisfying *a* condition of invariance. Instead, it involves deciding about the validity of the whole model. So even ideal experiments are insufficient to this end. What is more, the manipulationist strategy misses a story about testing for the particular data I am concerned with: those on which we cannot intervene. In other words, the manipulationist/interventionist strategy fails to offer a separate account for observational contexts. As a consequence, explanatory manipulationism is not an option for observational contexts, whilst the structural strategy discussed earlier is.

It is worth emphasising that, of course, one can support the metaphysics she likes, but the chosen metaphysics is useless if there are no implications for testing. A structural modeller could indeed hold a manipulationist metaphysics, but then either she is left with no indication whatsoever as to how to test causal relations, or she has to hold the questionable view that metaphysics and methodology are completely unrelated. As mentioned in section 2.1, the structural strategy is not committed to any particular metaphysics, while the trouble with the manipulationist strategy is twofold: (i) conceptual manipulationism is a particular metaphysics that is unrelated to large part of scientific practice, notably to observational social science, and (ii) explanatory manipulationism provides a story about testing that is of no use to scientists that cannot intervene on data.

3.3 Observational data and 'weak invariance'

A way to go round the stumbling block of observational data is to say that in such cases *we don't* perform any experiment, but *Nature* does the experiment for us. Fair enough, but we are left with the same problem as before: analysing a data set, the hard job is to figure

out whether or not Nature performed an experiment from which to draw the conclusion that correlations are invariant enough to be deemed causal. How do we do that? I presented before the structural strategy. In a nutshell, we model and test structures, i.e. mechanisms, that explain correlational data; the plausibility of such a story is a matter of evaluating the validity of the *whole model*, and confrontation with background knowledge at the various stages looms large in this. Let us now consider the solution of one of the most strenuous manipulationists.

Beside invariant empirical generalisations, Woodward also discusses the concept of 'possible-cause' generalisation (chapter 5.8 of his 2003), for which 'weak invariance' has to hold (chapter 6.15 and 7.8). Possible-cause generalisations, following his presentation, are the kind of generalisations that are established by means of structural equation models in social science, and weak invariance is not to test whether the generalisation *would* remain stable were we to intervene, but whether the generalisation *is* stable across subpopulations or different partitions of the data set.

In his long and detailed chapter on invariance, Woodward discusses Cornfield et al. (1959) paper on the relations between smoking and lung cancer. Woodward (2003, p.312) correctly notices that this paper was written in 1959, when detailed knowledge about the biochemical mechanism through which smoking produces cancer was still lacking. Consequently, scientists largely relied on epidemiological evidence—that is on *observational* data—and only to a lesser extent on experimental studies of laboratory animals. Woodward then points out that the scientists do not aim to formulate 'exceptionless generalisations', i.e. laws; instead, they establish a causal link between smoking and lung cancer because the relation turns out to be *invariant*. What kind of invariance is Woodward referring to? Let us read the full passage:

> For example, the authors note that some association appears between smoking and lung cancer in every well-designed study on sufficiently large and representative populations with which they are familiar. There is evidence of a higher frequency of lung cancer among smokers than among nonsmokers, when potentially confounding variables are controlled for, among both men and women, among people of different genetic backgrounds, across different diets, different environments, and different socioeconomic conditions […]. The precise level and quantitative details of the association do vary, for example, the incidence of lung cancer among smokers is higher in lower socioeconomic groups, but the fact that there is some association or other *is stable or robust across a wide variety or different groups and background circumstances*. […] Thus, although Cornfield et al. do not exhibit a precise deterministic or probabilistic generalization that is invariant across different circumstances [meaning: across interventions] the cumulative impact of their evidence is to show that the relationship between smoking and lung cancer is relatively invariant in the weak sense described above [i.e., chapter 5.8]. (Woodward 2003, p.312, emphasis and brackets added)

The kind of invariance invoked here *is not* the kind of *counterfactual* invariance under intervention presented in section 3.1. Rather, as discussed in the structural strategy (section 2.3), invariance is tested across partitions of the data set. This prompts a number of remarks.

A charitable interpretation of Woodward's notion of 'weak invariance' is that, after all, we don't disagree that much. Perhaps all this emphasis on manipulation and intervention is 'just' misplaced. Unfortunately, misplacing manipulation and intervention has the

serious side effect of overshadowing that making sense of correlational data is a matter of evaluating the validity of the whole model, *not* of evaluating a single condition. Testing invariance, be it 'counterfactual' or 'weak', is but a part of the more ample story about validity.

There is however a stronger point to be made. Weak invariance may solve the problem of testing invariance in contexts where we cannot directly intervene on the data. That is to say, it is a workable alternative to the kind of *counterfactual* invariance advocated by manipulationist modellers. Yet, the problem of confusing testing *a condition* and evaluating the validity of the *whole model* arises again, because in observational contexts, where the goal is to establish possible-cause generalisations (in Woodward's terminology) we don't *just* test various conditions of weak invariance one by one. Thus, the manipulationist/interventionist strategy, both in its strong form adapted to experimental contexts and in its 'weak' form adapted to observational contexts, can be subsumed under the more general framework of the structural strategy. The structural strategy offers a broad methodological framework, namely it is an umbrella for various particular modelling tools. Manipulationist strategies in the 'strong' and 'weak' versions are exactly specific tools in this sense.

Let me emphasise that this is more than a philosophy of science point. This is a point about methodology too. Even armed with 'weak' invariance besides 'counterfactual' invariance, the manipulationist modeller wouldn't be able to reproduce the explanatory machinery that structural modellers such as López, Mompart, Wunsch (1992) put in motion to explain the low mortality rate in Spain in the early Nineties. The reason is that manipulationism discriminates between correlations and causal relations on the basis of just *one* condition, i.e. invariance. More importantly, manipulationism is not accompanied by a mechanistic interpretation of the recursive decomposition—this is what does the explanatory job. To conceive of mechanisms simply as chains of invariance relations does not shed any light on *why* the decomposition represents a mechanism and on *why* such formal tool can be explanatory. This is instead provided by the structural strategy, as argued in section 2.1 and 2.2, and in Mouchart and Russo (2011). Thus, manipulationist modellers lack a 'global view' of modelling, where the notion of validity of the model is vital. This includes confrontation with background knowledge at all stages and does not reduce testing causal hypothesis to checking *a* single condition.

4. Conclusion

Making sense of correlational data coming from observations or experiments is a shared problem across the sciences. What this 'making sense' arguably means is whether correlations are causal or not. If we could always perform experiments, this would be a relatively easy task: wiggle the putative cause, hold fix the rest, and see what happens to the putative effect. Unfortunately, we can't. There are in fact contexts—I have in mind many social science domains—in which no manipulation on the data can be performed. The scientist is provided with a data set and her task is to tell a story—a causal story—that accounts for the correlations found in the data.

This paper suggested that a way to tell causal stories in order to make sense of correlations: the 'structural strategy'. I presented the structural strategy as a general

methodological framework rather than as a particular model, such as structural equation models, covariance models, or hierarchical models. The main idea behind this structural approach is to look for structures, that is for mechanisms, that are able to explain why we observe the correlations we do. Once a mechanism is hypothesised, it is put forward for empirical testing and, finally, the validity of the model will be evaluated on the basis of results of empirical testing and of other theoretical considerations. Thus, the structural strategy is not offering an answer to the old question of what is the extra content that correlations must have in order to be causal. The structural strategy locates in the domain of epistemology and methodology of causality and reformulates the old problem 'when are correlations causal' into 'how do establish the validity of a causal model'.

I then turned the attention to another possible strategy: the manipulationist or interventionist strategy. I argued that this account, based on the notion of invariance under intervention, cannot solve the problem of making sense of correlational data when manipulations are not possible. On the one hand, if the manipulationist account is given a purely metaphysical reading—that is it aims to tell what causality is—then a story about testing is missing and consequently there is no possible answer to the problem of the paper. On the other hand, if the manipulationist account is given a more epistemological and methodological reading—that is it gives an account of testing how things would/must change in response to certain interventions, then the story told is not suitable for purely observational contexts, exactly because too much importance is bestowed on experiments.

References

Bechtel, W. (2008). *Mental mechanisms: philosophical perspectives in cognitive neuroscience*. Oxford: Routledge.

Bechtel, W. and Abrahamsen, A. (2005). Explanation: a mechanist alternative, *Studies in the History and Philosophy of the Biological and Biomedical Sciences*, 36, 421-441.

Blalock, H. M. (1964), *Causal Inference in Nonexperimental Research*. Chapel Hill: University of North Carolina Press.

Bogen, J. (2004). Analysing Causality: The Opposite of Counterfactual is Factual. *International Studies in the Philosophy of Science*, 18(1):3 − 26.

Bogen, J. (2005). Regularities and causality; generalizations and causal explanations. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36, 397-420.

Bollen, K.A. (1989). *Structural Equations with Latent Variables*. New York: John Wiley & Sons.

Campaner, R. (2003). Sulle teorie manipolative della causalità. *Rivista di Filosofia*, 94(1), 89-106.

Campaner, R. (2006). Mechanisms and Counterfactuals: A Different Glimpse of the (Secret?) Connexion. *Philosophica*, 77 15-44.

Campbell, D. T. and Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Chicago: Rand McNall.

Cartwright, N. (2007). *Hunting causes and using them*. Cambridge: Cambridge University Press.

Cartwright, N. (2011). Predicting 'It Will Work for Us': (Way) Beyond Statistics. In P. McKay Illari, F. Russo, J. Williamson (eds), *Causality in the Sciences*. Oxford University Press.

Chao, H-K (2009). *Representation and structure in economics. The methodology of econometric models of the consumption function*. New York: Routledge.

Cook, T. and Campbell, D. (1979). *Quasi-Experimentation. Design and Analysis Issues for Field Settings*. Chicago: Rand MacNally.

Craver, C. (2007). *Explaining the brain*, New York: Oxford University Press.

Darden, L. (2006). *Reasoning in biological discoveries*. New York: Cambridge University Press.

Darden, L. and Craver, C. (2002). Strategies in the interfield discovery of the mechanism of protein synthesis. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 33(1), 1-28.

Duncan, O. D. (1975). *Introduction to structural equation models*. New York: Academic Press.

Durkheim, E. (1895 - 1912). *Les règles de la méthode sociologique*. Paris: Libraire Félix Arcan. 6th edition.

Durkheim, E. (1897 - 1960). *Le suicide*. Paris: Presses Universitaires de France.

Fennell, D.J. (2011). The structural theory of causation. In P. McKay Illari, F. Russo, J. Williamson (eds), *Causality in the Sciences*. Oxford University Press.

Freedman, D. A. (2004). On specifying graphical models for causation, and the identification problem. *Evaluation Review*, 26, 267–93.

Freedman, D. A. (2005). *Statistical models. Theory and practice*. Cambridge: Cambridge University Press.

Frigg, R. and Hartmann, S. (2009). Models in Science, *The Stanford Encyclopedia of Philosophy* (Summer 2009 Edition), E. N. Zalta (ed.), URL= <http://plato.stanford.edu/archives/sum2009/entries/models-science/>

Galavotti, M. C. (2001). Causality, mechanisms and manipulation. Available at PhilSci Archive, University of Pittsburgh, http://philsci-archive.pitt.edu/archive/00000132/, accessed 9 July 2009.

Guala, F. (2005). *The methodology of experimental economics*. New York: Cambridge University Press.

Guala, F. (2008). Extrapolation without process tracing. Paper presented at the Philosophy of Science Biannual Meeting, Pittsburgh.

Hausman, D. (1986). Causation and Experimentation. *American Philosophical Quarterly*, 23, 143-154.

Hausman, D. (1997). Causation, Agency, and Independence. *Philosophy of Science*, 64(4), Supplement, S15-S25.

Hausman, D. and Woodward, J. (2004) Manipulation and the Causal Markov Condition. *Philosophy of Science*, 71(5), 846-856.

Heckman, J. (2008). Econometric causality. IZA DP No. 3425. http://www.nber.org/papers/w13934 accessed 11 April 2009.

Hoover, K. (1988). *The new classical macroeconomics: a skeptical inquiry*. Oxford: Basil Blackwell.

Hoover, K. (2001). *Causality in macroeconomics*. Cambridge: Cambridge University Press.

Hoover, K. (2011). Counterfactuals and Causal Structure. In P. McKay Illari, F. Russo, J. Williamson (eds), *Causality in the Sciences*. Oxford University Press.

Jiménez-Buedo, M. and Miller, L. M. (2010). Why a trade off? The relationship between external and internal validity in experiments. *Theoria*, 25(69), 301-321.

Kaplan, D. (eds) (2004), *The SAGE handbook of quantitative methodology for the social sciences*. Thousands Oaks, Cal., London: SAGE.

López-Rios, O., Mompart, A., Wunsch, G. (1992). Système de soins et mortalité régionale: une analyse causale. *European Journal of Population*, 8(4), 363-379.

Lucas, R.E. (1976). Econometric policy evaluation. In K.Brunner and A.H.Meltzer (eds.), *The Phillips curve and labor markets, Carnegie-Rochester Conference Series on Public Policy*, vol.1, Spring. Amsterdam: North-Holland, 161-168.

Machamer, P. (2004). Activities and Causation: The Metaphysics and Epistemology of Mechanisms. *International Studies in the Philosophy of Science*, 18(1), 27-39.

Machamer, P., Darden, L. and Craver, C. F. (2000). Thinking about Mechanisms. *Philosophy of Science*, 67, 1-25.

Moneta, A. (2007). Mediating between causes and probabilities: the use of graphical models in econometrics (pp. 109-130). In F. Russo and J. Williamson (eds), *Causality and Probability in the Sciences*. London: College Publications.

Morgan, M. and Morrison, M. (1999). *Models as Mediators. Perspectives on Natural and Social Science*. Cambridge: Cambridge University Press.

Morgan, S.L. and Winship, C. (2007). *Counterfactuals and Causal Inference*. New York: Cambridge University Press.

Mosley, W. H., Chen, L. C. (1984). An analytical framework for the study of child survival in developing countries. *Population and Development Review*, 10 (Supplement), 25–45.

Mouchart, M. and 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.

Mouchart, M., Russo, F., Wunsch, G. (2009). Structural modelling, exogeneity and causality. In H. Engelhardt, H-P Kohler, A. Prskwetz (Ed.), *Causal analysis in population studies: concepts, methods, applications* (59-82). Dordrecht: Springer.

Pearl, J. (2000). *Causality: models, reasoning, and inference*. Cambridge: Cambridge University Press.

Pearl, J. (2011). The structural theory of causation. In P. McKay Illari, F. Russo, J. Williamson (eds), *Causality in the Sciences*. Oxford University Press.

Popper, K. (1959). *The logic of scientific discovery*. London: Hutchinson.

Psillos, S. (2004). A glimpse on the secret connexion: harmonizing mechanisms with counterfactuals. *Perspectives on Science*, 12(3), 288-319.

Quetelet, A. (1869). *Physique sociale. Ou Essai sur le developpement des facultés de l'homme*. Bruxelles: Muquardt.

Russo, F. (2006). The rationale of variations in methodological and evidential pluralism. *Philosophica*, 77, 97-124.

Russo, F. (2009a). *Causality and causal modelling in the social sciences. Measuring variations*. Methodos Series. New York: Springer.

Russo, F. (2009b). Explaining Causal Modelling. Or, What a Causal Model Ought to Explain. In Proocedings of the SILFS Conference, Milan 8-10 October 2007. London: College Publications.

Simon, H. (1953). Causal ordering and identifiability. In W. C. Hood, T. C. Koopmans (Ed.), *Studies in econometric method* (pp. 49–74). New York: Wiley.

Simon, H. (1954). Spurious correlation: a causal interpretation. *Journal of the American Statistical Association*, 49(267), 467–479.

Spirtes, P., Glymour, C., Scheines, R. (1993). Causation, prediction, and search. New York: Springer.

Steel, D. (2008a). *Across the boundaries. Extrapolation in biology and social science*. Oxford University Press.

Steel, D. (2008b). A new approach to argument by analogy: extrapolation and chain graphs. Paper presented at the Philosophy of Science Biannual Meeting, Pittsburgh.

Strevens, M. (2007). Essay review of Woodward, Making Things Happen. *Philosophy and Phenomenological Research*, 74:233–249. 2007.

Strevens, M. (2008). Comments on Woodward, Making Things Happen. *Philosophy and Phenomenological Research*, 77:171–192.

Woodward, J. (1997). Explanation, invariance, and intervention. *Philosophy of Science*, 64, S26-S41.

Woodward, J. (2000). Explanation and invariance in the special sciences. *The British Journal for Philosophy of Science*, 51, 1197-254.

Woodward, J. (2002). Counterfactuals and causal explanation. PSA 2002: contributed papers, http://philsci-archive.pitt.edu/archive/00000839/, accessed 29 April 2008.

Woodward, J. (2003). *Making things happen: a theory of causal explanation*. New York: Oxford University Press.

Woodward, J. (2004). Counterfactuals and Causal Explanation, *International Studies in the Philosophy of Science*, 18(1), 41-72.

Woodward, J. (2006). Author's response. *Metascience*, 15, 53-66.

Woodward, J. (2008). Reply to Strevens. *Philosophy and Phenomenological Research*, 77, 193-212.

Woodward, J. and Hitchcock, C. (2003). Explanatory generalizations, Part I: A counterfactual account. *Noûs*, 37(1), 1-24.

Wunsch, G. (2007). Confounding and Control. *Demographic Research*, 16: 15–35.