Causal explanation:
recursive decompositions and mechanisms *

Michel MOUCHART[a] and Federica RUSSO[b]

[a] Statistics, [b] Philosophy,

University of Louvain, Belgium.

July 24, 2009

Abstract

This paper deals with causal explanation in quantitative-oriented social sciences. In the framework of statistical modelling, we first develop a formal structural modelling approach which is meant to shape causal explanation. Recursive decomposition and exogeneity are given a major role for explaining social phenomena. Then, based on the main features of structural models, the recursive decomposition is interpreted as a mechanism and exogenous variables as causal factors. Arguments from statistical methodology are first offered and then submitted to critical evaluation.

1 The quest for causal explanations in the social sciences

Emile Durkheim (1960) had the ambitious goal to explain suicide as a social phenomenon. Accordingly, in his masterpiece *Le suicide*, he looked for the social causes of suicide. Durkheim's interest in the determinants of suicide was motivated by the observation of a great variability in the suicide rate. This variability appeared to be quite irrelevant across time within the same population, but was instead considerable across different societies. By examining how the suicide ratio varied across societies, Durkheim aimed to detect the social factors this variation depended on and thus to explain why, for instance, societies with a more integrated family structure had lower suicide rates. More recently, the demographer John Caldwell proposed a model to explain child survival in developing

*This paper was originally prepared for presentation at the conference 'CAPITS 2008', University of Kent 10–12 September 2008. We thank the conference participants, Guillaume Wunsch and two referees for extremely helpful and stimulating comments. F. Russo acknowledges financial support from the FRS-FNRS (Belgium). M. Mouchart acknowledges financial support from the IAP Research network grant nr P6/03 of the Belgian Government (Belgian Science Policy).

countries. Notably, Caldwell (1979) investigated maternal education as a major causal factor. He observed that maternal education alone could account for more variance than all other relevant socio-economic factors altogether (*e.g.* mother's place of residence, husband's occupation and education, type of marriage, *etc.*), and therefore *this* factor deserved special attention. López-Ríos et al. (1992), to give another example, were interested in explaining a significant lower mortality rate in Spain in the Eighties after socio-economic policies in the Seventies were carried out. In particular, they were interested in assessing the causal effect of factors such as economic and social development on the one hand, and of the use of sanitary infrastructure on the other hand.

What these examples from various areas in the social sciences have in common is that they all seek to provide an explanation of a phenomenon of interest—more specifically, they seek to provide a *causal* explanation. A causal explanation is provided, intuitively, once the factors or causes that bring about the phenomenon are identified. This, however, is still too loose a characterisation. In this paper we investigate how causal explanations are built in the social sciences.

We proceed as follows. Section 2 presents the structure of the explanation that a statistician provides for a phenomenon of interest. Three aspects are highlighted. First, explanation is incomplete, or partial, because it is based on a stochastic representation of the world where the stochastic component stands for what is *not* explained. Second, an explanation is given by decomposing a complex causal mechanism into a sequence of 'simpler' explanatory mechanisms. Third, explanation is causal, that is we identify cause-effect relations through the condition of exogeneity. Section 3 then addresses the question of the interpretation of the recursive decomposition and of why it carries explanatory power. We argue that a recursive decomposition is to be interpreted as a causal mechanism and that what allows the causal interpretation is exogeneity. The goal is not to provide (yet another) definition of the concept of mechanism, but rather to clarify what it is meant when social scientists interpret a component, or factor, of a recursive decomposition as a 'mechanism'.

2 The structure of the statistician's explanation

2.1 Explanation in a stochastic environment

To explain a social phenomenon, the statistician is usually provided with a data set containing observations coming, for instance, from a survey or a census. The statistician's explanation of a phenomenon related to a data set is based on a statistical model, which is basically a set of probability distributions on an observation space, namely:

$$\mathcal{M} = \{\mathbf{S}, P^{\theta} \quad \theta \in \Theta\} \tag{1}$$

where $\mathbf{S}$ is the sample, or observation, space and for each $\theta \in \Theta$, P^{θ} is a probability distribution on $\mathbf{S}$. Thus, θ characterizes a particular distribution P^{θ} and is called a parameter. The statistician analyses data *as if* it had been generated by one of these distributions; at this stage, the distributions in (1) have only a representational role without structural nor causal implications—a topic to be considered later in this section. Thus a statistical model can be thought of as a set of plausible hypotheses to uncover the so-called data generating process.

Earlier papers (Mouchart et al., 2009; Russo et al., 2008), noticed that a statistical model can be seen as a stochastic representation of the phenomenon of interest. This is due to partial and incomplete knowledge of a phenomenon that leads the statistician to model the phenomenon stochastically. This stochastic representation is the cornerstone of the statistician's explanation: the interpretation of parameter θ provides the explanation of the phenomenon whereas the stochastic component of the model stands for the unexplained part of the phenomenon. Here, 'hazard' (stochastic, random) is here used as opposed to 'explainable'. We use it as an epistemic concept independent of the metaphysical issue of whether the real world is deterministic or indeterministic.

Let us illustrate with a very simple example. Suppose we weight an object n times with an imperfect scales. We accordingly observe $X = (X_1, X_2, ...X_n)$. A simple statistical model might be $X_i \sim ind.N(\mu, \sigma^2)$, *i.e.* each X_i is assumed to be identically independently distributed (iid) as a normal distribution with mean μ and variance σ^2. The statistician's explanation then runs as follows: because of the imperfection of the scales, the measurements X_i are interpreted as a realisation of a probability distribution and by interpreting the parameter $\theta = (\mu, \sigma^2) \in I\!R \times I\!R_+$ the statistician explains that each measurement X_i is relative to a same 'true' weight μ and that the error distribution is characterised by a variance σ^2. The statistician is able to account for the fact that the measurements X_i tend to cluster around a same value μ but not for the fact that the measurements X_i are at a distance from that value. In this sense the statistician's explanation is partial.

2.2 The multivariate aspect of social phenomena

In social contexts, interest is usually given to the multivariate aspect of a phenomenon of interest. The statistician who analyses a social phenomenon considers a vector of variables selected on the basis of background knowledge and of the available data; this is far from being an easy task but a thorough discussion of the criteria to select the variables of interest falls beyond the scope of this paper. This multivariate aspect makes the task of the statistician's explanation far more complex. In the sequel, we develop three steps that should be distinguished: (i) marginal-conditional decomposition, (ii) structural modelling, and (iii) recursive decomposition.

2.2.1 Marginal-conditional decomposition

Explaining a complex multivariate phenomenon is typically operated by decomposing it into a sequence of simpler 'pieces'. The natural decomposition of a multivariate probability distribution is obtained by a marginal-conditional decomposition, *i.e.* in the bivariate case $X = (Y, Z)$:

$$p_X = p_Z \, p_{Y|Z} \tag{2}$$

Here the bivariate process generating $X = (Y, Z)$ is decomposed into two univariate processes: a marginal process generating Z and a conditional process generating Y given Z ('generating' in the sense suggested in section 2.1). To illustrate, consider the following example. Suppose that the statistician observes data on the price, Y, and the quantity, Z, of fish transacted upon the fishermen's return from the sea. A simple model might decompose this bivariate process, generating $X = (Y, Z)$, into a marginal process generating Z and deemed to represent the good or bad fortune of the fishing activity, and a conditional process, generating $(Y \mid Z)$, representing the operation of the auctioning process. Let us now assume, for the sake of simplicity, that each distribution is normal. Thus the left-hand side of (2) would be written:

$$\begin{pmatrix} Y \\ Z \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mu_Y \\ \mu_Z \end{pmatrix}, \begin{pmatrix} \sigma_Y^2 & \sigma_{YZ} \\ \sigma_{YZ} & \sigma_Z^2 \end{pmatrix}\right] \tag{3}$$

whereas the right hand side would be written

$$Z \sim \mathcal{N}(\mu_Z, \sigma_Z^2) \qquad (Y \mid Z) \sim \mathcal{N}(\alpha + \beta Z, \sigma_{Y|Z}^2) \tag{4}$$

Although models (3) and (4) are mathematically equivalent, model (4) is explanatory because the parameters (μ_Z, σ_Z^2) and $(\alpha, \beta, \sigma_{Y|Z}^2)$ characterise the marginal process representing the fishing activity and the auctioning process, respectively, at variance from the parameters of (3) that only characterise the random variability of the data. The statistician's explanation essentially lies in the interpretation of the parameters that characterise the distributions, *e.g.* if $\beta < 0$, the conditional process generating $(Y|Z)$ explains that, when Z increases, Y tends to decrease on conditional average. Similarly, the converse decomposition $p_X = p_Y \, p_{Z|Y}$ would not provide a convincing explanation from an economic point of view.

2.2.2 Structural modelling

Alone, the marginal-conditional decomposition is not enough for explanation because, in the example of the bivariate case, it is mathematically arbitrary to choose $p_X = p_Z \, p_{Y|Z}$ or $p_X = p_Y \, p_{Z|Y}$. Thus, an explanation also requires the statistical model to uncover the *structure* of the data generating process. As long as such a structure is latent, *i.e.* not directly observable, the statistician will make systematic use of two ingredients: background knowledge and invariance.

Background knowledge Broadly speaking, background knowledge, or field-knowledge, stands for the whole body of knowledge we have of a given field, and may incorporate different aspects such as: knowledge of the socio-demo-political context, knowledge of same/similar phenomena relative to the same population at different times or to other populations, results of analyses performed with different methods, knowledge of related fields such as biomedicine in epidemiology, *etc.* Sometimes such knowledge may be gathered into ‘well established’ theories. In the fish market example given above, background knowledge supports the decomposition $p_X = p_Z \, p_{Y|Z}$ rather than $p_X = p_Y \, p_{Z|Y}$ because the first one allows us to associate a contextually interpretable univariate subprocess to each component of the relevant decomposition: this was precisely the rationale behind (4).

Invariance We mentioned earlier that an observation of the data set is interpreted as *a* realisation of one of the distributions constituting the statistical model. However, a model that specifies a different process for each observation of the data set would be rather useless for explanation as it would not be *structural*. Thus a major aim of structural modelling is to distinguish incidental from structural components of a data generating process. This means that a structural model should display an adequate level of stability or invariance under a suitable class of modifications of the environment and/or of interventions. Invariance is a condition of stability of the marginal-conditional structure of the model and of the characteristics (parameters) of the distribution; parameter stability is indeed an important object of statistical testing. Moreover, the stability of the marginal-conditional decomposition may also be tested by evaluating the parametric stability of alternative decompositions. The specification of the invariance property of a structural model is also a basic ingredient of the definition of the population of interest or the population of reference. In the fish market example, it would be important to test whether the auctioning process has the same characteristics in different seasons and/or different harbours. It should be stressed that invariance and stability properties are requirements *complementary* to congruence with background knowledge, that does not necessarily imply stability or invariance. For this reason invariance and stability are the object of (statistical) testing. In particular, intervention may raise difficulties in the stability of the system, a problem already pointed out by Lucas (1976).

2.2.3 Recursive decomposition

Let us now consider the general case where a vector of variables X is decomposed into g components, namely $X = (X_1, X_2, \cdots X_g)$ (with g typically much larger than 2), and suppose that the components of X have been ordered in such a way that in the complete marginal-conditional decomposition

$$\begin{aligned} p_X(x \mid \omega) \quad = \quad & p_{X_g|X_1,X_2,\cdots X_{g-1}}(x_g \mid x_1, x_2, \cdots x_{g-1}, \theta_{g|1,\cdots g-1}) \\ & \cdot p_{X_{g-1}|X_1,X_2,\cdots X_{g-2}}(x_{g-1} \mid x_1, x_2, \cdots x_{g-2}, \theta_{g-1|1,\cdots g-2}) \cdots \\ & \cdot p_{X_j|X_1,X_2,\cdots X_{j-1}}(x_j \mid x_1, x_2, \cdots x_{j-1}, \theta_{j|1,\cdots j-1}) \cdots p_{X_1}(x_1 \mid \theta_1) \end{aligned} \tag{5}$$

each component of the right hand side may be considered, in a first step, as a structural component with mutually independent parameters, *i.e.* (in a sampling theory framework):

$$\omega = (\theta_{g|1,\cdots g-1}, \theta_{g-1|1,\cdots g-2} \cdots, \theta_1) \in \Theta_{g|1,\cdots g-1} \times \Theta_{g-1|1,\cdots g-2} \cdots \times \Theta_1 \tag{6}$$

Under property (6) the conditioning variables $X_1, \ldots, X_{j-1}$ of each factor of (5), $p_{X_j|X_1,X_2,\cdots X_{j-1}}$, are called *exogenous for* the parameter of the corresponding conditional distribution, $\theta_{j|1,\cdots j-1}$. That is to say, the inference on the parameter $\theta_{j|1,\cdots j-1}$ of the conditional distribution should not depend on the specification of the data generating process of the conditioning variables $X_1, X_2, \cdots X_{j-1}$. Therefore, exogeneity is a condition that allows the separation of inferences, notably of the inferences on $\theta_{j|1,\cdots,j-1}$ characterising the conditional distribution and on $\theta_{1,\cdots,j-1}$ characterising the marginal distribution of the conditioning variables. More explicitly, in a likelihood approach, the separation of inference means that any inference concerning (any function of) $\theta_{1,\cdots,j-1}$ would be based on a likelihood function derived from the marginal distributions $p_{X_1,X_2,\cdots X_{j-1}}$, independently of the specification of the conditional distribution $p_{X_j|X_1,X_2,\cdots X_{j-1}}$, whereas any inference concerning (any function of) $\theta_{j|1,\cdots j-1}$ would be based on a likelihood function derived from the conditional distributions $p_{X_j|X_1,X_2,\cdots X_{j-1}}$ independently of the specification of the marginal distribution $p_{X_1,X_2,\cdots X_{j-1}}$. It follows that such a separation of inference takes advantage from more a parsimonius modelling and eventually enjoys of more robust properties. Later, in section 3.3, we stress other important features of exogeneity.

Remark The concept of exogeneity has a long history in econometrics. The works of the Cowles Commission in the late Forties and the early Fifties have been path-breaking and are still influential nowadays; in particular, Koopmans (1950) puts emphasis on exogeneity in dynamic models. Barndorff-Nielsen (1978) is significant in the development of conditions for separation of inference. Florens and Mouchart (1980, 1985) and Florens et al. (1980) bridge Koopmans (1950) and Barndorff-Nielsen (1978) works and provide a coherent account of exogeneity integrating the separation of inference in dynamic and non-dynamic models. Engle et al. (1983) present a list of different concepts of the econometric literature and display their connections with exogeneity through the introduction of supplementary conditions. Florens and Mouchart (1985) not only provide a basic concept of exogeneity, but also make the concept explicit in different levels of model specification, namely, global, initial, and sequential, before combining those concepts of exogeneity with non-causality. This analysis is further developed in Florens et al. (1993). ■

Equations (5) and (6) characterize a *completely recursive system*. A recursive decomposition is not complete when, in equation (5), some components are random vectors rather than random variables. This typically happens when we cannot order some of the variables due to a lack of knowledge about their causal or temporal priority. In such a case, in the factorization (5) there is (at least) one factor being a distribution of a *vector* of variables, say X_j, conditional on the antecedent ones $(X_1, \cdots, X_{j-1})$. In other words, the conditional process generating $(X_j \mid X_1, \cdots, X_{j-1})$ is *not* decomposed into a sequence of univariate conditional processes, hence the recursive decomposition is not complete. This situation is met under the heading of 'simultaneity' in the econometric literature.

The case of simultaneity is an interesting and quite disputed issue. A classroom example is provided by a simple two-equation market model: supply and demand. When it is known, and specified, that one side is price-setter and the other side is price-taker, a complete recursive decomposition may be operated and the econometric model completely explains the mechanism generating the observed price and quantity and the market equilibrium process is completely understood. However, when a standard competitive approach is adopted, the market-clearing process becomes a black box generating the equilibrium price and quantity, and the econometric explanation has a different nature. Indeed, the supply and demand equation represents now notional concepts, rather than statistical entities such as marginal and conditional distributions. These concepts are of the same nature of the counterfactuals used in a large portion of the literature on causation. In such a case the identification of the parameters requires identifying restrictions that are not empirically testable, but only supported by contextual knowledge and/or economic theory. In other words, the explanation provided by the econometric model is of a speculative nature and the recursive decomposition among the endogenous variable is not operating.

Let us emphasise that a recursive decomposition is essentially an ordering of the variables in such a way that each factor of the right-hand side of (5) is structurally valid. Once the number of components p increases, background knowledge, possibly substantiated by statistical tests, typically provides a simplification of the factors in the form of conditional independence properties. More specifically, it is often the case that the distribution of $(X_j \mid X_1, \cdots, X_{j-1})$ is known not to depend on some of the conditioning variables. Thus there is a subset of variables $\mathcal{I}_j \subset \{X_1, \cdots, X_{j-1}\}$ actually relevant for the conditional process generating $X_j \mid X_1, \cdots, X_{j-1}$ as defined by the property

$$X_j \perp\!\!\!\perp X_1, \cdots, X_{j-1} \mid \mathcal{I}_j \tag{7}$$

implying that the factor $p_{X_j|X_1,X_2,\cdots X_{j-1}}$ in (5) is actually simplified into $p_{X_j|\mathcal{I}_j}$ and $\mathcal{I}_j$ may be called the *relevant information of the j-th factor*. Once $\mathcal{I}_j$ has been specified for each factor, (5) is condensed into

$$p_{X_1,X_2,\cdots X_g} = \prod_{1\leq j\leq g} p_{X_j|\mathcal{I}_j} \tag{8}$$

This form is accordingly called a *condensed recursive decomposition*. As argued in Mouchart et al. (2009), *causes* may then be viewed as exogenous variables in the condensed recursive decomposition, or, alternatively, as the relevant information of a structurally valid conditional distribution.

Readers familiar with the literature on graph models may recognize that a directed acyclic graph (DAG) is a graphic representation of a condensed recursive decomposition and that the causal structure is depicted by the set of ancestors. Also, in the literature on graph models, the concepts of completeness and of recursivity are not identical to those developed in the statistical tradition and discussed in this paper. This is due to the fact that in the statistical tradition these concepts relate to multivariate distributions; not all structures of probabilistic independences can be represented by graph models. For binary variables, a simple example is the case of a trivariate distribution with pairwise independence but not complete mutual independence. In simple cases, however, the two families of concepts coincide.

Summarizing: the statistician's explanation is *partial*, because based on a stochastic representation of a phenomenon of interest, and *structural*, because based on a recursive decomposition that seeks to decompose a vector of variables into structurally valid components.

2.3 Difficulties

As a matter of fact, several problems hinders explanation from being a simple task. Let us focus on three of them: (i) partial observability, (ii) time delay, dynamic structure, feedback effects, and (iii) causal chain.

2.3.1 Partial observability

Many models in the social sciences involve latent, *i.e.* non-observable, variables. Some of these variables could possibly be observed but are non-observable for practical, legal, or ethical reasons whereas other variables are genuinely non-observable because they correspond to theoretical concepts partially observed through indicators, or proxy variables, and are used in framing theoretical models. The statistical model, because bearing on observable variables only, is accordingly constructed in two steps: a first one in the form of a structural model involving both observable and latent variables and a second step, the implied statistical model (also called operational model), obtained by integrating out the non-observable variables.

Statistical models of that type are known in statistical methods as 'mixture models', because a marginal distribution may be viewed as a mixture of conditional distributions, and are characterised by severe difficulties. First, the distributions have a complex analytical structure, making the inference process often cumbersome. Second, the integration of non-observable variables require the introduction of specific supplementary assumptions often impossible to be controlled or to be statistically tested, and justified only by field knowledge. Third, the marginalized distributions no longer represent a data generating process supported by arguments that it is structural, and their parameters may no longer be given a simple structural interpretation. Moreover, integrating the unobservable explanatory variables typically jeopardises the exogeneity of the remaining variables; this problem has been explicitly worked out for a simple trivariate case in Mouchart et al. (2009). The econometric literature on heterogeneity, *i.e.* on unobservable explanatory variables, often suggests to recover loss of exogeneity by introducing further *ad hoc* assumptions, such as the independence between the heterogenous factors and the observable explanatory variables, even though such assumptions may be contextually doubtful and empirically not testable (see also Mouchart et al. (2009) for a deeper discussion of those supplementary assumptions).

2.3.2 Time delay, dynamic structure, feedback effects

In many cases, a reasonable specification of the structure of the data generating process requires the introduction of time delays in order to take into account dynamic features of the phenomenon of interest. This makes the observations a sequence of data not independent. In particular, the effects of a cause require some delay before being operational and feedback effects often take place through adjustments of individual behaviours. These facts generate further difficulties.

Firstly, the specification of dynamic models is substantially more demanding than the specification of models with mutually independent observations.

Secondly, the time frequency of data is often not high enough for identifying the shortest term features. In other words, the available data operate a time-aggregation and therefore should be viewed as a partial observability of the dynamic structure of the data generating process. In line with the above-mentioned difficulties due to partial observability, econometricians, already in the early Fifties (Wold and Jureen (1953) and Bentzel and Hansen (1955)), have argued that that time-aggregation is a main cause of simultaneity because otherwise an econometric model should come into a completely recursive form.

Thirdly, even without time-aggregation, the presence of feedback effects requires a substantially more complex analysis of exogeneity and causality. Moreover, different levels of model specification should be distinguished, namely (i) a global one, modelling at once all available data, (ii) an initial one, modelling data conditionally on initial conditions, and (iii) a sequential one, modelling each data sequentially conditionally on their relative history. Florens and Mouchart (1985) provides an integrated approach to this topic.

2.3.3 Causal chain

The recursive decomposition, be it complete or not, along with the causal interpretation of exogenous factors, makes manifest that variables in $\mathcal{I}_j$ are (direct) causes of X_j and, similarly, other variables are causing X_{j-1}, accordingly producing a chain of causes within the data $X = (X_1, \ldots, X_g)$. Two issues should be made explicit.

First, the 'natural' state of a social phenomenon is not 'one cause–one effect' but rather 'multiple causes–multiple effects'. This leads to identify not only direct but also *indirect* causes. Thus the crucial aspect of the framework presented here is to provide an *ordered structure* in a 'systemic' approach. In other words, it is not enough to say that 'everything depends on everything': a structure ought to be elaborated in order to explain and shed light within an otherwise black box. But there is no free lunch. The cost to be paid is to learn how to manage a complex causal structure. In the social sciences this issue is crucial, in particular, for policy purposes.

Second, the causal chain constructed within a given data set $X = (X_1, \ldots, X_g)$ is essentially truncated. This means that once the data set has been ordered in such a way that X_g is explained by $X_1, \ldots, X_{g-1}, \ldots, X_j$ explained by $X_1, \ldots, X_{j-1}$, *etc.* , the statistician is still left with explaining X_1. One might argue that either there is no plausible explanation for X_1 or, in the causal chain, X_1 is far enough from the variables $X_j, X_{j+1}, \ldots$ to be explained, so that the indirect effect of the explanatory variables of X_1 could be neglected. However, it should be stressed that, although necessary from an operational viewpoint, the explanation provided by means of the statistical model may fail to be robust with respect to the truncation. The social scientist should be particularly aware of that difficulty, and only field knowledge can be of help at this stage. Nevertheless, the fact that appealing to field knowledge helps in those cases does not introduce a vicious circle nor does it make knowledge to be gained through structural modelling radically different from field knowledge. Rather, this reflects the idea that structural modelling (i) does establish knowledge to be used in other studies, but (ii) does not establish immutable and eternal 'causal truths'. Structural modelling is a dynamic process in which field knowledge and new knowledge constantly interplay. How precisely they interplay is, however, the object of another paper.

3 Explanatory mechanisms

3.1 Interpreting recursive decompositions

So far we have argued that the statistician's attempt to explain a given phenomenon of interest involves two aspects. First, a genuinely partial explanation by incorporating in the statistical model a stochastic component deemed to represent what is not explained. Second, a recursive decomposition over an ordered vector of variables deemed to disentangle a multivariate, *i.e.* complex, phenomenon into a sequence of univariate (conditional) processes.

It is worth emphasising that a recursive decomposition is made of two ingredients. First, a sequential marginal-conditional decomposition of a multivariate distribution. This is a standard operation in the calculus of probability that can be operated over an arbitrary order of a vector of variables. Second, a specific order is selected by requiring structural validity of each component of the decomposition. Such a validity requires a close congruence with background knowledge, along with a condition of stability/invariance, which implicitly defines the population of reference. These two elements allow us to introduce a *specific* concept of mechanism that fits structural models. Because this concept is not meant to be a general one and consequently may not attract unanimous consensus, comparison with other alternative approaches may be useful.

Thus, the guiding questions of this section are the following: how to interpret the recursive decomposition? Why is it explanatory? In a nutshell, recursive decompositions are to be interpreted in terms of *mechanisms*, and we discuss below why mechanisms thus conceived *explain*.

3.2 Recursive decompositions and mechanisms

In interpreting the recursive decomposition in terms of 'mechanisms' we distinguish between 'global mechanisms' and 'sub-mechanisms'. The *whole* recursive decomposition characterises a *global* mechanism, whereas *each* conditional distribution within the recursive decomposition characterises an (autonomous) *sub*-mechanism within the global one. In this context, 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. In other words, the explanatory power of a mechanism is operationalised, in structural models, through the recursive decomposition. For instance, in the fish market example, the market, call it the global mechanism, generates a bivariate distribution of the price (Y) and quantity (Z); the econometric explanation consists in distinguishing a supply process represented by the marginal distribution of Z, a demand process represented by the conditional distribution of ($Y|Z$), and a market equilibrium process based on a quantity-taking behaviour of the demand side.

Thus a recursive decomposition carries explanatory power insofar as it disentangles a global mechanism into sub-mechanisms in the above sense. But what are the specific features of a mechanism in this context of structural models?

Stochastic mechanisms This concept of mechanism is a stochastic one: a mechanism is *not* deterministic but it rather singles out a stable/invariant and contextually meaningful aspects of the phenomenon of interest (see below). The fact that in social contexts causal explanation is essentially mechanistic does not imply that it also is *mechanicistic*, in the sense that it essentially requires physical deterministic mechanisms in order to explain social phenomena.

Stable mechanisms Because a mechanism is meant to identify a stable/invariant, and therefore repeatable, aspect of the phenomenon being modelled, identifying a mechanism means to separate incidental from structural features of the data generating process. By so-doing, the statistician is also able to distinguish spurious from causal correlations.

Mixed mechanisms In social contexts, mechanisms are not necessarily 'physical', that is made of physical processes or physical entities interacting in one way or another. This is for several reasons, three of which are:

1 In statistical models used in the social sciences mechanisms try to depict the working forces,

i.e. the motivation or rationale for evolving, characterised by variables that possibly lack 'physical' (or even manifest/observable) counterparts. Many social, demographic, or economic variables are conceptual constructs introduced to shape a 'theory', the development of which leads to building measurement devices by means of a number of relevant indicators that are distinct from the definition of the concept. For instance, 'socio-economic status' might be measured from income and level of education, but these indicators are not meant to provide a unique definition of the concept.

2 Many social scientists are interested in mechanisms where very different types of variables interplay. For instance, health economics or some branches of epidemiology are interested in how economic variables influence health variables and vice-versa. In this case, although some variables might have a 'physical' counterpart (*e.g.* baby's weight at birth), not all of them will (*e.g.* socio-economic status). Consequently, we need a characterisation of mechanism broad enough as to include both 'physical' and 'non-physical' components. That is to say, we have to model 'mixed' mechanisms.

3 Also, health variables do not influence economic variables (or vice-versa) *as such*, but through indirect paths involving intermediate causal variables. Those indirect paths need (or may need) to be specified in order to *explain* the phenomenon of interest. In Caldwell's model mentioned in section 1, maternal education does not influence child mortality *as such*, and in fact a major improvement of Caldwell's framework was provided by Mosley and Chen (1984), who developed an analytical framework explaining the indirect paths through which a social variable such as maternal education can have a causal impact on a health variables such as child survival.

For details on mixed mechanisms and, more generally, on modelling mechanisms in causal modelling, see Russo (2008, ch. 6).

Alternative views

The very notion of mechanism is currently matter of a vivid debate (Little, 1991, 1998; Machamer et al., 2000; Woodward, 2002, 2003; Bunge, 2004; Psillos, 2004; Bechtel and Abrahamsen, 2005; Reiss, 2007; Craver, 2007). Alternative accounts may feed debate about the *concept* or the *role* of mechanisms in the social sciences. For expository purposes, we only selected two, notably Little (1991, 1998) and Craver (2007).

Little (1991) defends the idea that causal analysis in the social sciences is legitimate but that it depends upon identifying social mechanisms. Little goes as far as saying that such social mechanisms work through the actions of individuals—a position also known as methodological individualism. To discuss the plausibility of a microfoundation approach (see for instance Little (1991, 1998)) would lead us too far away from the main track. Yet, Little's characterisation of mechanisms will help us in clarifying our claim that the recursive decomposition represents a social mechanism. According to Little (1991, p. 15) a causal mechanism is a series of events governed by lawlike regularities that lead from the explanans to the explanandum. Mechanisms, within a microfoundational perspective, are grounded in meaningful and intentional behaviour of individuals. The sort of things having causal properties are, for instance, the actions of individuals and groups. One might disagree with Little about the soundness of a microfoundation approach, or about the use of *lawlike* regularities. Notably, Hoover (2001) stresses the causal import of the structural approach in econometrics arguing for a reality of macroeconomic structures that does not boil down to the reality of microeconomic relations and holds the view that mechanisms and causal structures may substitute for laws and do not necessarily need to be supported by appeals to laws.

Yet, notwithstanding divergences on those issues, we surely agree with Little on his account of *statistical analysis* as a form of causal reasoning in social research (Little, 1991, ch.8). Statistical explanation, in his view, has to be accompanied by a causal story indicating the mechanisms. The identification of the mechanism involves (lawlike) statistical regularities, but is not the end of the explanation; it is only the first step in establishing causal relations. So statistical tools serve to uncovering the patterns present in the empirical phenomenon *i.e.* the data set. It is in this sense that Little's and our views are close to each other.

Craver (2007) explores the notion and the explanatory power of mechanisms in the domain of neurosciences. At the beginning of his book, Craver (2007, p. 5) discusses the example of how a neuron releases neurotransmitters and concludes:

> This is a mechanism in the sense that it is a set of entities and activities organized such that they exhibit the phenomenon to be explained.

To our understanding, this "skeletal description", as Craver calls it, is broad enough as to account for mechanisms in various domains. Should you take the entities to be neurons and the activities neurotransmitters release, the above skeletal description will well fit neuro-mechanisms. Should you take entities to be socio-demo-economic variables, and activities to be their influence on other socio-demo-economic variables, the above skeletal description will fit equally well social mechanisms. The degree of 'physical' reality one wishes to give to entities and activities may lead to different accounts—notably, to a different ontological commitment to the existence—of mechanisms. In the social sciences, we do not need to endorse the view that elements and relations should always have *physical* counterparts—see above the discussion of mixed mechanisms.

3.3 Mechanisms, explanation, causality

3.3.1 Mechanisms and explanation

More on partial explanation In section 3.1 we have recalled the partial nature of statistical explanation—the stochastic component delineates the frontier between what we explain and what we do not explain. But there is another sense in which explanation is partial.

In case the statistician can operate a *complete* recursive decomposition, the explanation is complete in the sense that each sub-mechanisms is identified and thus the global mechanism fully disentangled. However, in case the statistician is unable—for a whole variety of different reasons, *e.g.* missing data or insufficient background knowledge—to operate a complete recursive decomposition, the explanation itself is partial. In such a case the conditional distribution bears simultaneously on several variables that are therefore not explained individually.

More on explanatory power However, whether complete or incomplete, the recursive decomposition—the mechanism—provides an explanation of the phenomenon of interest. In other words, mechanisms, we claim, carry explanatory power. The question is to understand *why* it is so. The answer to this question resides in considering the whole modelling procedure as explanatory on the one hand, and in understanding the explanatory import of exogeneity on the other hand. Simply put, when a conditional distribution is a component of the recursive decomposition, the conditioning variables are exogenous and can be interpreted as causal factors. In the next section, we discuss this idea more thoroughly. A related issue concerns the possibility to simulate. The recursive decomposition explains because the distribution generating the data can be simulated in a contextually meaningful way. In this sense to explain also means 'being able to reproduce'.

3.3.2 Causal factors and exogeneity

Why interpreting *exogenous* variables as *causal* factors? This is for three different but related reasons.

First, the *whole modelling procedure* is explanatory. The goal of structural modelling is to characterise clearly identified and interpretable mechanisms. We mentioned in section 2 that the choice of the marginal-conditional decomposition may be arbitrary; this is the reason why we need background knowledge and invariance: to make a selection among the various possible decompositions. In other words, the marginal-conditional decomposition *alone* does not provide a (causal) explanation of a given phenomenon, but the whole modelling procedure does. Indeed, building a structural model is made of a progressive procedure, three steps of which may be identified. In a first stage we select the variables of interest, that is the elements of the mechanism, out of background knowledge. In a second stage we build the statistical model, in particular, we operate a recursive decomposition over the initial joint probability distribution. Finally, in a third stage we confirm or disconfirm the hypothesised mechanism by confrontation with empirical evidence and background knowledge. Briefly put, tests on the recursive decomposition concern, on the one hand, its invariance or stability and, on the other hand, whether it is congruent with background knowledge. Adequacy of the model is also tested by measuring goodness of fit and the amount of variability the model can account for. This is a very concise presentation of the hypothetico-deductive methodology of causal models. For a thorough discussion of hypothetico-deductivism in causal modelling see Little (1998, p.' 9), Cartwright (2007, ch. 2), and Russo (2008, ch. 3).

Such a stepwise methodology provides a *causal* explanation because it aims to provide an understanding of a given phenomenon by showing the causal sub-mechanisms that underlie it. Of course, the question arises as to what guarantees the *causal* interpretation of the relations or the mechanisms established in those models. This is the relation between exogeneity and causality, that is why we interpret exogenous factors as *causal* factors.

Second, before discussing in more detail the relation between exogeneity and causality, another issue is worth pointing out. Borrowing Cartwright's adage, *no causes in, no causes out.* According to the hypothetico-deductive methodology used in structural models, the first stage, that is the hypothesis formulation stage, exactly concerns a *causal* hypothesis. Therefore what we will (dis)confirm exactly is a *causal* structure. Causal relations are not inferred from mere correlations, taken out from a 'magical' statistical hat. Structural models, unlike simple associational or descriptive models, have a rich and sophisticated apparatus of assumptions that underwrite their causal interpretations (see for instance Russo (2008, ch. 3) who divides them into three categories: statistical, extra-statistical, and causal). Also, specific tests—notably, invariance and exogeneity tests—allow us to causally interpret the relations showed in the recursive decomposition and therefore the exogenous variables as causes.

Third, here comes the thorny issue: the relation between exogeneity and causality. Identifying causal factors with exogenous variables is based on the following considerations:

- Because causality is a latent concept, causal inference can only be of the type 'to the best of our knowledge'; causal relations pertain to the *interpretation* of a model (*i.e.* a *representation* of the data generating process) and are, therefore, relative to a model rather than a sole characteristic of the available data. Differently put, structural modelling is not a hunt for the 'true' model nor a device that enables us to discover the 'true' causal relations. Structural modelling is a progressive path toward making intelligible the observed phenomena while adjusting the window of observation to pre-specified targets.

- Exogeneity is a condition of separation of inference. As mentioned earlier, the (partial) explanation of the statistician is cast in the framework of a statistical model, in terms of parameters that characterise the distributions of interest (see end of section 2.2.1). Thus the exogeneity condition (6) does not only allow us to separate the inferences on $\theta_{j|1,\cdots,j-1}$ and on $\theta_{1,\cdots,j-1}$, but it also allows us to distinguish the process generating the causes, characterised by $\theta_{1,\cdots,j-1}$, and the process generating the effect, characterised by $\theta_{j|1,\cdots,j-1}$. Separating causes from effects mirrors the asymmetry of causation. This last point makes clearer and more precise the older expression 'exogenous means generated outside the model'.

Needless to say, this is not to suggest that structural models provide immutable and eternal causal explanations of social phenomena. As mentioned above, explanation is intrinsically relative and partial, that is relative to the specific conceptual framework and dependent on available empirical and theoretical information. This means that nothing prevents future explanations to discard previous ones. Furthermore, nothing prevents different social scientists to provide different explanations to the same phenomenon: a causal explanation crucially depends on background knowledge which also includes the social scientist's personal or political beliefs. Finally, such causal explanations involve an implicit stopping rule in order to avoid an otherwise ad infinitum chain of 'explaining the explanatory'.

3.4 Evaluation / flexibility of explanation

Before closing this section on explanatory mechanisms, two features of causal explanation are worth mentioning. The first concerns the evaluation of explanations and the second their flexibility.

We can evaluate explanations by considering three interrelated aspects: (i) statistical, (ii) epistemic, and (iii) ontological adequacy. We can give (i) a *statistical* evaluation by measuring, with the coefficient of determination, how much variability is accounted for, and by measuring the goodness of fit. We can also give (ii) an *epistemic* evaluation, by asking whether results are coherent with background knowledge. (iii) An *ontological* evaluation is also possible: if ontological homogeneity between the variables acting in the mechanism is lacking (for instance if the mixed mechanism includes both economic and health variables), it may be desirable to identify and justify indirect paths from the causes to the effect. Causal explanations will then be good or bad depending on how well they meet statistical, epistemic, and ontological requirements.

Such explanations also exhibit a high flexibility. The first aspect of flexibility is concerned with 'mixed mechanisms': as discussed earlier, we do not need to stick to a *physical* concept of mechanism. The second aspect relates to the available information we base the explanation upon. The case study on bargaining powers and market segmentation presented below shows that even if available data and background knowledge do not allow to fully explain the phenomenon, *some* explanation is possible. The third aspect is that such explanations allow a *va et vient* between established theories and establishing theories. Established scientific theories are (and ought to be) used to formulate the causal hypothesis and to evaluate the plausibility of results on theoretical grounds. But causal models also participate in establishing new theories by generalising results of single studies. This reflects the idea that science is not monolithic, discovering immutable and eternal truths. If the model fits the data, the components of the recursive decomposition are structurally stable and congruent with background knowledge, then we can say, to the best our knowledge, that we hit upon a mechanism that explains a given social phenomenon. But what if one of these conditions fails? A negative result may trigger further research by improving the modelling strategies or by collecting new data, thus leading to new discoveries that may question background knowledge.

We now conclude by briefly presenting two examples. In the first one, researchers successfully provided a causal explanation by disentangling the mechanism in a recursive decomposition on five variables. In the second one, researchers did not succeed to fully explain the phenomenon by providing the marginal recursive decompositions due to a lack of data and background knowledge.

Health systems and mortality in Spain López-Ríos et al. (1992) were interested in regional mortality in Spain. Spain met deep socio-economic changes in the mid-Seventies, and consequently policy in that period tried to intervene on improving the social and economic situation. This led to a lower mortality rate at the time of the study. This background supported the choice of distinguishing the supply and demand of medical care, 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-Ríos et al. (1992), instead, hypothesised that regional mortality be influenced by the health system which was in turn influenced by the social and economic development. The vector of variables (economic development, social development, sanitary infrastructure, use of the medical care system, age structure, mortality) was decomposed in basically two sub-mechanisms. In the first, 'economic development' was the exogenous variable influencing mortality through 'social development' and 'sanitary infrastructure'; in the second, 'age structure' was the exogenous variable influencing mortality through 'use of the medical care system'.

Bargaining powers and market segmentation in freight transport in Belgium Mouchart and Vandresse (2007) analysed multivariate data obtained by face-to-face interviews with companies using the services of freight transport in Belgium. This data provided information, for each contract of a sample, on several characteristics of the contract, such as the price, the distance, the speed of delivery, *etc.* From the interviews it was clear that each contract was the result of a bargain between the service user and the service provider. For instance, a requirement of quick delivery could be priced at a higher tariff than a slow delivery, depending on the availability of the provider. However, there wasn't available data on every step of the bargaining process: only the final result was known. Moreover, no economic theory, no game-theoretic strategy, nor any field-knowledge was available for substantiating any possible recursive decomposition. Consequently, a recursive decomposition of the data generating process was not possible and Mouchart and Vandresse (2007) could only provide an analysis of the joint distribution of all the available data without incorporating any exogeneity assumption. This analysis nevertheless provided some explanation of the global functioning of the freight transport market in terms of imperfection of the competition and of the bargaining power of the actors, but did not provide an explanation about the data generating process of each variable separately.

4 Concluding remarks

Quite uncontroversially, explanation belongs to the tasks of the social sciences. What is more controversial, however, are the features and characteristics of explanations in social contexts. This paper tackled this issue by analysing structural modelling.

We highlighted three features. (i) Explanation is incomplete, or partial, in the sense that it is based on a stochastic representation of the world where the stochastic component stands for what is not explained. (ii) An explanation is given by decomposing a complex causal mechanism into a sequence of 'simpler' explanatory mechanisms. In a nutshell, explaining a complex social phenomenon involves two ingredients. First, we operate a recursive decomposition on a multivariate distribution that represents the phenomenon of interest; the whole recursive decomposition is interpreted as the 'global mechanism'. Second, we consider each component of the recursive decomposition as an 'autonomous' sub-mechanism within the global mechanism insofar as it is composed by a univariate conditional distribution. 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. In other words, the explanatory power of a mechanism is operationalised, in structural models, through the recursive decomposition. (iii) Explanation is causal, that is we identify cause-effect relations through the condition of exogeneity. We defined exogeneity as a condition of separability of inferences on the parameters of the marginal-conditional distribution, which allows to identify the variables that play a causal role in the mechanism.

We emphasised that providing a complete explanation of a complex phenomenon is not always (and often not) possible and that incomplete recursive decompositions have to be accommodated in the toolkit of the social scientist.

Rather than proposing new definitions of key concepts used in structural modelling, we offered a reassessment of the literature connecting the practice of (structural) statistical modelling to the concepts of explanation and mechanism. Notably, we aimed to clarify how social scientists explain social phenomena by building structural models and what it means to interpret a recursive marginal-conditional decomposition as a mechanism.

References

Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*. John Wiley & Sons.

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.

Bentzel, R. and Hansen, B. (1954-1955). On recursiveness and interdependency in economic models. *Rev. Economic Studies*, 22:153–158.

Bunge, M. (2004). How does it work? The search for explanatory mechanisms. *Philosophy of the Social Sciences*, 34:182–210.

Caldwell, J. (1979). Education as a factor in mortality decline: an examination of nigerian data. *Population Studies*, 33(3):395–413.

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

Craver, C. F. (2007). *Explaining the brain. Mechanisms and the mosaic of neuroscience*. Oxford University Press, Oxford.

Durkheim, E. (1897/1960). *Le suicide*. Presses Universitaires de France.

Engle, R., Hendry, D., and Richard, J.-F. (1983). Exogeneity. *Econometrica*, 51(2):277–304.

Florens, J.-P. and Mouchart, M. (1980). Initial and sequential reduction of bayesian experiments. *CORE Discussion Paper 8015*.

Florens, J.-P. and Mouchart, M. (1985). Conditioning in dynamic models. *Journal of Time Series Analysis*, 53(1):15–35.

Florens, J.-P., Mouchart, M., and Rolin, J.-M. (1980). Réductions dans les expériences bayésiennes séquentielles, paper presented at the colloque processus aléatoires et problèmes de prévision, held in bruxelles 24-25 april 1980. *Cahiers du Centre d'Etudes de Recherche Opérationnelle*, 23(3-4):353–362.

Florens, J.-P., Mouchart, M., and Rolin, J.-M. (1993). Noncausality and marginalization of markov processes. *Econometric Theory*, 9:241–262.

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

Koopmans, T. (1950). When is an equation system complete for statistical purposes? In Koopmans, T., editor, *Statistical Inference in Dynamic Economic Models*, volume Cowles Commission Monograph 10 of *Statistical Inference in Dynamic Economic Models*. John Wiley & Sons.

Little, D. (1991). *Varieties of social explanation: an introduction to the philosophy of social science.* Westview Press, Boulder.

Little, D. (1998). *Microfoundations, method and causation. On the philosophy of the social sciences.* Transaction Publishers, New Brunswick, N.J.

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

Lucas, R. (1976). Econometric policy evaluation. In Brunner, K. and Meltzer, A., editors, *The Phillips curve and labor markets*, volume 1 of *Carnegie-Rochester Conference Series on Public Policy*, pages 161–168. North-Holland, Amsterdam.

Machamer, P., Darden, L., and Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science*, 67:1–25.

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

Mouchart, M., Russo, F., and Wunsch, G. (2009). Structural modelling, exogeneity and causality. In H. Engelhardt, H-P Kohler, A. P., editor, *Causal Analysis in Population Studies: Concepts, Methods, Applications*, chapter 4, pages 59–82. Springer, Dordrecht.

Mouchart, M. and Vandresse, M. (2007). Bargaining powers and market segmentation in freight transport. *Journal of Applied Econometrics*, 22:1295–1313.

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

Reiss, J. (2007). Do we need mechanisms in the social sciences? *Philosophy of the social sciences*, 37(2):163–184.

Russo, F. (2008). *Causality and causal modelling in the social sciences. Measuring variations.* Methodos Series. Springer, New York.

Russo, F., Wunsch, G., and Mouchart, M. (2008). Potential outcomes, counterfactuals, and structural modelling: Causal approaches in the social sciences. In *European Population Conference*, Barcelona.

Wold, H. and Jureen, L. (1953). *Demand Analysis.* Wiley, New York.

Woodward, J. (2002). What is a mechanism? a counterfactual account. *Philosophy of Science*, 69:S366–S377.

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