The final publication is available at Springer via http://dx.doi.org/10.1007/s11135-006-9060-4

Article type: Short communication

Causality as Validity: Some implications for the Social Sciences

Rafael Moreno and Rafael J. Martínez

University of Seville

Address for correspondence:
Department of Experimental Psychology.
Calle Camilo J. Cela s/n
41018 Seville (Spain)

Telephone: +34 954557670
Fax: +34 954551784
e–mail: rmoreno@us.es, rmcervan@us.es

Causality as Validity: Some implications for the Social Sciences

Abstract

We analyse the concept of causality in the social sciences, whose development is insufficient and lesser than the methodology developed for its study. The nature of the causal process as the production of effects remains unclear and the relationships considered to be manifestations of that process cannot be taken for proof of its existence. Given these difficulties, we suggest that, aside from the inherited interpretations, the practice of the concept of causality makes reference to correctly specified relationships not confounded by others; characteristics identical to those which define validity. In that way, causality is equivalent to the validity of a relationship. Beyond merely re-understanding causality, this proposal permits the deduction that the temporal precedence of the cause is a necessary condition only for one type of causality, making it possible to consider other types, not admitted by the traditional notion, in which the cause is consequent or simultaneous to the variable to be explained. Examples and characteristics of these types of causality are presented and considered to be useful for the social sciences.

Key Words: Causality, Validity, Temporal asymmetry, Simultaneous Relations, Non-directional Causality, Causal Models, Social Sciences

1. Introduction

Causality is a central theme in the methodology of the social and natural sciences, which corresponds to the relevant place it holds in our culture (Thompson, 1993). In that sense, experimental methodology holds the study of causal relations as its objective, and the possibilities for the evaluation of causal models for non-experimental data in the social sciences has increased greatly in recent years (e.g. Aickin, 2001; Bentler, 1985; Bollen, 1989; Bollen & Long, 1993; Dalhbäck, 2001; Hayduk, 1987; James et al., 1982; Jöreskog & Sörbom, 1988; Long, 1983; Mc Donald, 1985; Muthén, 1987; Seidel & Eicheler, 1990). This methodological development has been accompanied by a broad consensus about the nature and characteristics of causality: a) It is understood as a process of production or transmission of forces by which the cause brings about the effect; b) It is considered that that process is manifested in specific relationships and that it appears when certain variables, such as common causes and intervening variables, which may influence the components of the relationship under study, are controlled (Wright, 1934, 1960; Simon, 1954; Blalock, 1962; 1964; Duncan, 1966). Relationships which do not occur under these conditions are considered to be spurious or indirect, products of other variables, except in the case of omission of another variable – termed a suppressing variable– with similar effects, but of the opposite sign, on the variables of a causal relationship (Pedhazur, 1982; Bollen, 1989). Finally, c) It is understood that, as manifestations of a process of production, causal relationships are temporally asymmetrical or directional from the cause to the effect.

However, such consensus certainly does not translate into a clear concept of the supposed process of production, which is therefore maintained as one of the frequently discussed themes in social science (Doreian, 2001; Notterman, 2004). As Kenny (1979: 4) affirmed, "It is difficult to convey this notion of causality formally, just as it is to formally define space or time". The principle difficulties of each one of the indicated characteristics are the following, respectively: a) That link is not explained beyond the general declaration that the cause is the generating force of the effect. What exactly the nature of a causal force is and what causal forces in different areas, such as mechanical physics and social processes, have in common, are questions which are not answered, despite being basic for the postulation of the existence of that causal force; a force said to be "almost vitalistic," given its lack of definition (Kenny, 1979: 4); b) The supposition that the causal force has an existence which is made manifest in specific relationships has not been validated, given that the force is not defined independently of the studied relationships. These relationships may not be considered to be the manifestation or proof of the existence of a force which is itself exclusively defined by those relationships. It is simply a circular argument which cannot validate what it intends to, and ; c) Although the temporal asymmetry or direction of the causal relationships is clearly identified in terms of the precedence of a variable, the temporally asymmetric relationship is not a characteristic specific to causation, for example the succession “cockcrow - sunrise”.

Thus, although the causal theme in the social sciences presents an important development of formal procedures, it has serious difficulty to define and validate the supposed existence of a causal force of production. As have indicated Glymour et al. (1989: 58), "mathematical representation is the easiest part; the hard part concerns what inferences are to be drawn from causal claims", which is due precisely to the poor definition of these claims. It appears that there exists a conceptual problem regarding the causal process – a problem compounded by the existing procedures for the study of relationships. To have procedures for the purpose of studying something which is poorly defined is a situation which should be reconsidered in order to recover the sense of the procedures and of the causal notion itself. It is a problem most likely linked to the insufficient attention paid by methodology, as has been indicated by others: "A basic introduction to the concept of causal theories is hard to find, since most of the time the discussion is already put in mathematical form" (Saris & Stronkhorst, 1984: 22). Frequently, methodological work on causality attributes the insufficiencies in definition to lapses in the philosophy of science. That would show that the social sciences have neglected to undertake the necessary conceptual analysis, passing it on to philosophy. Nonetheless, given that that attitude has not been satisfactory, it would appear reasonable to approach the causal question from scientific methodology itself, analysing its specific bases and practices.

2. Causality as validity

As a starting point, we consider that the existing methodological procedures and their continuous application express the concept of causality actually used in scientific practice – a concept which, nonetheless, appears to be masked by the usual and inherited interpretation of causality as a force of production. In our opinion, to adequately define the concept of causality does not insist that one questions the action of the supposed force, but that one questions the facts which are under investigation, and are considered to be reasonable, when researchers claim to be studying causality. Such as with any other concept, it may be true for causality that in identifying the situations in which a community considers the use of the concept to be adequate, we are demonstrating the meaning which that community gives it in practice (Wittgenstein, 1953), coinciding or not with the way that that community describes that practice itself. Therefore, to identify the current concept of causality in the social sciences demands the analysis of the facts used in the methodological bibliography as characteristics of causality, despite the interpretations which have been superimposed over them; analysing what they do, more than what it is said that they do. That should be the necessary and sufficient material to meet the indicated objective.

In the methodological bibliography of the social sciences, principally constituted of structural equation models (SEM), causality is referred to in terms of relationships (e.g. Bollen & Long, 1993; Dalhbäck, 2001; Hayduk, 1987; Jöreskog & Sörbom, 1988; Seidel & Eicheler, 1990). A large part of the relationships of many variables x_j with other variables y_i may be indicated by the additive linear model

$$y_i = \gamma_{i1}x_1 + \gamma_{i2}x_2 + \ldots + \gamma_{iq}x_q + \zeta_i \qquad (1)$$

where each γ_{ij} is the expected change in y_i for a one-unit change in x_1 when the rest of the variables are maintained constant, and where ζ_i represents the random component of y_i or the effects of variables such as errors of measurement or characteristics which are unknown or of little relevance and are therefore not included in the model as independent variables.

The mentioned expression takes into account that the variables of each studied relationship may be related to other variables which they should not confound. Thus, when studying each relationship (y_i, x_j) in equations like (1), it is required that the variables contained in ζ_i, beyond merely having minimal effects and constant variance, are not related among themselves nor among the variables x_j, meaning that $E(\zeta_i)=0$, $V(\zeta_i)=\sigma_{\zeta i}^2$, $COV(\zeta_i\zeta_{i'})=0$ for all $i \neq i'$, and $COV(\zeta_i x_j)=0$, for all i and j. This demands the inclusion of the maximum number of relevant variables, and demands that their control is carried out by operations previous to data collection and/or by statistical means, in experimental or non-experimental research. This is what Steyer et al. (2000) have indicated in terms of causal unbiasedness and unconfoundedness. Additionally, when representative inclusion and control of variables are not sufficient, assumptions or partial restrictions must be made, evaluating the possible relevance of each one of them by successive replications.

Some conclusions may be extracted from what has been expressed. While the traditional notion of causality is understood as a process of production of facts, which is manifested in controlled relationships between variables, in practice regulated by the scientific methodology of SEM, it occurs that: 1) The meaning of causality is exclusively referred to as adequately specified or represented relationships, not confounded by other variables. 2) These two characteristics also define the validity of a relationship, (Kerlinger, 1973; Shadish et al, 2002). In consequence, to identify a relationship as causal is simply, in practice, to identify it as valid. 3) Causality then is formed by the relationships subject to study and by those others which must not confound the first, which means that these and other relationships are equally as necessary and that neither is sufficient to establish causality. Aside from this conjunction, it is true that "correlation does not prove causation" and that "a lack of correlation does not disprove causation" (Bollen, 1989: 52). Finally, 4) the notion sets the causal theme in identifiable terms and rescues it from interpretations with problems of definition or verification; the adequately specified and not confounded relationships should be recognised as necessary and sufficient material to constitute the meaning of causality, instead of being underestimated as simple manifestations of supposed underlying processes never defined independently with respect to those relationships.

To assume the previously enumerated ideas permits a different and perhaps more significant way of thinking about the two characteristics of the traditional notion: Causal force, manifested in relationships. Instead of saying that a relationship reflects causal force studied through specified and not confounded relationships, it may be said that this relationship is termed causal. This last point describes the circumstances in which the concept of causality is applied in scientific practice, different from the supposition of a force based on data considered to be manifestations of that very force. The causal theme therefore becomes a problem for the way it is planted rather than for not having been solved. The conception of causality as the transmission of a hidden force would, by definition, be a superfluous problem which could be abandoned.

3. No necessity of temporal asymmetry in causal relationships

The temporal asymmetry or directionality of relationships, understood as the precedence of one variable, is an inevitable characteristic of the assumption of causality as a productive force. The supposed transmission of that force between variables may only occur if the cause precedes the effect. In consequence, simultaneous relationships are usually ruled out, even considering that in relationships of such appearance - such as some from non-recursive models - the interval between cause and effect is smaller than the observation interval. Within this framework, temporal asymmetry or precedence is not questioned, as it is assumed to be a necessary condition for the relationship and any other possibility would be absurd. Keeping the precedence of the cause constant, the covariation between cause and effect is studied, but not if that covariation depends on the cause preceding the effect. The same occurs in the case of reciprocal causality, in which the cause precedent at time t is the effect variable at a posterior time $t+k$, which assumes bidirectionality or successive temporal asymmetry.

The basis differs from the notion of causality as validity understood in terms of unbiasedness and unconfoundedness. This notion does not include reference to temporal asymmetry of causal relationships, which may indicate that this characteristic is not a necessary condition for all effects. This means that the cause or necessary condition for a valid relationship must not necessarily be precedent in time, and may have several consequences. One may be that if temporal asymmetry or direction is not a necessary condition for causal relationships, then the order of occurrence of determined variables is one thing, and quite another is if that precedence is necessary for that relationship. For that reason, the temporal asymmetry of a determined causal relationship must not be understood as the simple precedence of one variable, but as the conditionality one relationship may have with respect to the temporal precedence of some of its variables. Also, if the temporal precedence of one variable is not a condition for all causal relationships, that means that there may be valid causal relationships without this precedence of the variable considered to be the cause, occurring simultaneous to the effect or following it. In other words, in accordance with the notion of causality as validity used in methodological practice, one may speak of three types of causal or valid relationships in function of the temporal location of the cause or necessary condition: a) Directional causality with preceding cause; b) Directional causality with consequent or posterior cause; and, c) Non-directional causality, with cause simultaneous to the effect. Given the novelty of the latter two types, it is convenient to explore their characteristics and present possible examples.

Abundant examples may be found of directional causal relationships with consequent cause in the psychological literature of operant conditioning, with wide generalization across species. In any of its variations, the phenomenon is described as the increment of those responses of the organism which are followed by determined events or reinforcement. The posterior reinforcer is considered the cause of the increase of the behaviour, and the key criteria to distinguish between behaviour directed by its consequences and that directed by previous stimuli which occur as pavlovian conditioning. Another example of this directional causality with consequent cause is found in the explanations of intentional action or anticipation-guided ends or goals, such as that found in Action Psychology (e.g. Harré et al, 1985).

In reference to non-directional causal relationships, some primary examples may be extracted from the work of Kuhn (1977) on causality in physics. Having objectives quite different from ours, that author referred to explicative relationships in which it was senseless to search for a precedent causal agent. One of those was the relationship between the elliptical orbit of the planet Mars – considered as a phenomenon to be causally explained – and the laws of Newton applied to an isolated system of two bodies which interact with an attraction inversely proportional to the square of the distance, and used as causes or explanations. That may be an appropriate example of non-directional causality in which, assuming the correct specification of the relationship and no confounding with other factors, the two variables are simultaneous. In effect, Newton's model applied to cause or explanation is an interpretation or re-reading of the to-be-explained phenomenon. For that reason, it is senseless to speak of temporal precedence of one or another variable, or to say that the structure described by the model is previous or posterior to the phenomenon. Similar examples may be found among the relationships between determined optical or electromagnetic phenomena and determined configurations of their respective fields. Thus defined the optical deficiencies of some people, such as farsightedness or the phenomenon of the change of light rays when passing through air or water, they are considered to be due to or caused by a determined configuration of the field given parameters such as the curvature of the lens, the focal distance or the angle between the two. Again, it makes no sense to say that the described structure of the model is previous or posterior to the phenomenon which it explains. Examples may also be found in the social sciences in the relationships between phenomena-to-be-explained – such as depression, education or voting behaviour – and models or interpretations of those phenomena, when both covary in valid conditions with adequate specification and the absence of confounding with other terms.

From the examples above it is possible to add some comments on the two types discussed – simultaneous and posterior causes – in addition to causality with temporal precedence. First, the existence of such relationships should not be surprising once the notion of cause as a force or production has been abandoned. There are many factors which do not condition all causal relationships, and we do not find any reason why the temporal precedence of one of its variables should be an exception. In this sense, a factor such as temporal-spatial contiguity between cause and effect, which was taken in the past to be a necessary condition for causation, ceases to be considered as such. Second, the three types imply a notion of causality in terms of a necessary condition, be they precedent, simultaneous or consequent. In other words, every causal relationship establishes an asymmetry between cause and effect, be it temporal, with an antecedent and a consequence, or conceptual as in simultaneous causes in which a model is considered to be an explanation of determined facts or data. Third, given that the three types of causal relationships include the requirements of adequate specification and no confounds, spurious or non-causal relationships are seen as clearly different. Fourth, the simultaneous or non-directional category is also different form the reciprocal or bidirectional category, where the relationship is conditioned by the precedence of different variables at various times, and where the non-directional relationships imply simultaneous variables. Fifth, the three types of causal relationships appear to have different functions. The directional relationships – reciprocal or not – take into account transitions or changes in one variable based on changes in another previous or posterior variable. The non-directional relationships, on the other hand, at least in the above examples, appear to implicate the interpretation of one variable based on other, more structured, variables, relationships between a fact to be understood and a structural model[1]. Finally, the described categories broaden the traditional causal notion of a previous agent, valid in some situations, helping to collect the explicative possibilities used by the natural sciences (Kuhn, 1977) and nonetheless to this day are not adequately recognised or taken advantage of in the social sciences.

4. Conclusions

The analysis of the criteria used in the methodological practice - structural equation models – in the social sciences permits a meaning which reinterprets the usual characteristics of causality. It appears that the material constituted by valid relationships, correctly specified and not confounded with others, is sufficient to realise the meaning given to causality in this scientific practice. From this point of view, it is unnecessary to consider that material as a reflection of a process or underlying force whose existence has never been evaluated independently of those relationships. We understand the notion of causality as a process of production, be it as a firm ontological belief or as a simple conceptual resource, is a problem which greatly distorts the causal theme. For that reason, the question should not be the discovery of relationships in which a causal force is identified, but the discovery of certain configurations or sets of correctly specified and not confounded among one another. To contemplate causality in terms of validity permits us to consider that the cause or necessary condition need not be precedent but also posterior or simultaneous to the variable which it explains, and therefore allows us to approach non-directional causality.

References

Aickin, M. (2001). Graphics for the minimal sufficient cause model. *Quality & Quantity* 35: 49-60.

Bentler, P. (1985). *Theory and Implementation of EQS: A Structural Equations Program*. Los Angeles: BMDP Statistical Software.

Blalock, H. M. (1964). *Causal Inferences in Non Experimental Research*. Chapel Hill: University of North-Carolina.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley.

Bollen, K. A., & Long, J. S. (1993). *Testing Structural Equation Models*. Newbury Park: Sage.

Dalhbäck, O. (2001). Using single-equation models of function-of-functions type in sociological research. *Quality & Quantity* 35: 173-189.

Doreian, P. (2001). Causality in social network analysis. *Sociological Methods and Research* 30: 81-114.

Duncan, O. D. (1966). Path Analysis: Sociological Examples. *American Journal of Sociology* 72: 1-16.

Glymour, C., Scheines, R., Spirtes, P., & Kelly, K. (1987). *Discovering Causal Structure*. New York: Academic Press.

Harré, R., Clarke, D. y De Carlo, N. (1985). Motives and Mechanisms. An introduction to the Psychology of action. London: Methuen (Spanish translation 1989; Barcelona: Paidós).

Hayduk, L. A. (1987). *Structural Equation Modeling with LISREL*. Baltimore: John Hopkins Univ.

James, L., Mulaik, S., & Brett, J. (1982). *Causal Analysis: Assumptions, Models and Data*. Sage: Beverly Hills.

Jöreskog, K. G., & Sörbom, D. (1988). *LISREL 7: A Guide to the Program and Applications*. Chicago: SPSS.

Kenny, D. A.(1979). *Correlation and Causality*. New York. John Wiley.

Kerlinger, F.N. (1973). *Foundations of behavioural research.* New York: Holt, Rinehart & Winston.

Kuhn, T. S. (1977). *The Essential Tension*. Chicago: University Chicago Press.

Long, J. S. (1983). *Covariance Structure Models: An introduction to LISREL.* Beverly Hills: Sage.

Mc Donald, R.P. (1985). *Factor analysis and related methods*. Hillsdale. L.A.E.

Muthén, B. (1987). *LISCOMP: Analysis of linear structural equations with a comprehensive measurement model. Theoretical integration and user's guide*. Mooresville: Scientific Software.

Notterman, J. M. (2004). Persistent conceptual issues in psychology: A selective update. *Theory and Psychology* 14: 239-260

Pedhazur, E, J. (1982). *Multiple Regression in Behavioral Research*. New York: Holt, Rinehart and Winston. (2nd. edition).

Saris, W. E., & Stronkhorst, L. H. (1984). *Causal Modelling in Nonexperimental Research*. Amsterdam. Sociometric Research Foundation.

Seidel, G., & Eicheler, C. (1990). Identification Structure of lineal structural models. *Quality & Quantity* 24: 345-365.

Shadish, W.R., Cook, T. D. & Campbell, D. T. (2002). *Experimental and quasiexperimental designs for generalized causal inference.* Boston: Houghton Mifflin.

Simon, H. A. (1954). Spurious Correlation: A Causal Interpretation. *Journal of the American Statistical Association* 49: 467-479.

Steyer,R., Gabler, S., von Davier, A. y Nachtigall, C. (2000). Causal regression models II: Unconfoundedness and causal unbiasedness. *Methods of Psychological Research* 5: 55-86.

Thompson, G.F. (1993). Causality in economics: Rhetorical ethic or positivist empiric? *Quality & Quantity* 27: 47-71

Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Basil Blackwell.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistical*, 5: 161-215.

Wright, S. (1960). Path coefficients and path regressions: Alternative or complementary concepts? *Biometrics* 16: 189-202.

Note.

1. The proposed types may represent the four aristotelian types of causality. The relationship of the precedent cause appears to correspond to causality of the efficient type, those of consequent cause to final causality and those non-directional causal relationships to the formal and material types, in which a variable is explained by being validly related to a representation or model of its structure or of its component elements, respectively.