COWLES COMMISSION DISCUSSION PAPER: STATISTICS NO. 353

The Causal Principle and the Identification Problem

by Herbert A. Simon

December 19, 1950

December 19, 1950

The Causal Principle and the Identification Problem

by Herbert A. Simon

The term "cause," presumably exorcised from empirical science by Hume, has never entirely disappeared from scientific discussions, even those ostensibly conducted within a positivist or operationalist framework. In some contexts, the concept of causality has simply been invoked as a synonym for determinism - e.g., in the determinism-indeterminism controversies of quantum mechanics. With this issue we will not be concerned, but will consider only systems that can be regarded as completely deterministic. Instead we are interested in the concept of causality as an asymmetrical relationship among certain variables in a deterministic system - a relationship that permits us such statements as "A causes B, while B is the effect produced by A;" and that rules out such statements as "A causes B and B causes A."

In view of the generally unsavory epistemological status of asymmetrical causal relationships, it is remarkable to what extent they underlie research in the biological and social sciences. For example, the classical work of Henderson, Cannon, and others on homeostasis is replete with references to asymmetrical relations among the variables. On thirst, Cannon states:

"Thirst is a sensation referred to the inner surface of the mouth and throat, especially to the root of the tongue and the back part of the palate. ... when water is lacking in the body the salivary glands are unfavorably affected... are therefore unable to secrete, the mouth and pharynx become dry and thus the sensation of thirst arises."

Walter B. Cannon. The Wisdom of the Body (W. W. Norton, New York, rev. ed. 1939) pp. 62-66.

The causal chain clearly implied by this statement is:

deficiency of water in body tissues $\longrightarrow$ reduction in salivation $\longrightarrow$ dryness of tongue and palate $\longrightarrow$ stimulation of nervous system (sensation of thirst). To this Cannon adds elsewhere:

$\longrightarrow$ activity of drinking $\longrightarrow$ restoration of water content of tissues.

Hume or no Hume, it is difficult to think or write of these functional relations as symmetrical, or asymmetrical but running in the opposite direction. For example, if there is normal salivation, but the saliva is prevented from reaching the tongue and palate, thirst is produced, but this neither reduces salivation nor produces a deficiency of water in the body tissues.

Similarly in economics, we speak of relations like:

Poor growing weather $\longrightarrow$ small wheat crops $\longrightarrow$ increase in price of wheat.

and we reject the notion that by changing the price of wheat we can affect the weather. The weather is an "exogenous" variable, the price of wheat an "endogenous" variable.

Recent research on the problem of the identification of economic relations has now provided us with an analytic scheme that permits the introduction of an asymmetrical causal relationship among variables, and an analysis of the epistemological assumptions underlying such a scheme. Indeed, the fundamental literature on the identification problem contains numerous references to such a causal principle (Koopmans ed. pp. 56, 60, 62, and particularly 393-409)

1. <u>Self-contained Structures</u>.

In the following discussion we will seek mathematical simplicity by limiting ourselves to systems of linear equations without stochastic disturbances. The basis for our procedure (and the procedure underlying the approach to the identification problem) may be outlined as follows:

1.1 In the classical theory of systems of linear equations, we are interested in properties of a system that are invariant under certain groups of transformations of the coefficients of its matrix. In particular, we may be interested in the solutions of a system (the sets of value of the variables satisfying the system). These are invariant under elementary row transformations of the matrix.

Elementary row transformations are those which: (1) interchange rows of the matrix (i.e. reorder the equations), (2) add to a given row multiples of another row or rows, (3) multiply a row by a non-zero scalar. These all amount to pre-multiplication of the coefficient matrix by a non-singular matrix. This group of transformation thus generated we will call the R-transformations.

(Albert, p. 24, 43)

Any two coefficient matrices that are obtainable from one another by R-transformations we will call <u>R-equivalent</u>. Concentration of interest on those properties (e.g., solutions) that are invariant under the group of R-transformations has led to the replacement of the notion of causality by the notion of mutual dependence. For given a (consistent and independent) set of k linear equations in n ($n \geq k$) variables, then in general, each variable belonging to any subset (see Bocher, p. 46) of k variables can be expressed as a function of the remaining $(n - k)$. The property of "dependency" of a variable or set of variables is not invariant under the group of R-transformations. Since the asymmetrical causal relationship appears to be connected with the notion of dependent variable, we cannot hope to find such an invariant relationship unless we restrict ourselves to a more limited group of transformations than the R-transformations.

<u>Definition</u>. <u>1.1</u>: We say that two coefficient matrices are <u>structurally equivalent</u> (<u>S-equivalent</u>) if the second can be obtained from the first by

premultiplication by a non-singular diagonal matrix (i.e. a product of row transformations of the third type only). The group of transformations thus admitted we will call the group of S-transformations.

1.2 Our procedure now will be to introduce an asymmetrical relationship, having the properties usually attributed to the causal relationship, which is invariant under the group of S-transformations.

Definition 1.2: A linear structure is the augmented coefficient matrix (that is, the $m \times (n + 1)$ matrix the elements of whose last column are the constant coefficients) of a system of linear equations, that possesses the following special properties:

1.2.1 that in any subset of k equations from the linear structure, at least k different variables appear with non-zero coefficients in one or more of the equations of the subset. (At least k columns of the matrix of the subset are not identically zero.)

1.2.2 that in any subset of k equations in which $m \geq k$ variables appear with non-zero coefficients, the values of any $(m - k)$ variables can be determined arbitrarily and the equations can then be solved for unique values of the remaining k variables (Bocher, p. 46)

These conditions are sufficient (although not all necessary) to guarantee the independence and consistency of the equations and of certain subsets with which ~~they~~ we will deal. They can be translated into conditions on the ranks of certain determinants contained in the augmented matrix of the structure.

Definition 1.3: A linear structure is self-contained if it has as many equations as variables. (Marschak, p. 7)

From 1.2.2, we see that a self-contained linear structure possesses a unique solution.

Definition 1.4: A linear model is the class of all linear structures that can be obtained from a given structure by substitution of new non-zero scalars for the non-zero scalars of the original structure without altering the ranks of any sub-determinants (i.e., without violating 1.2.2.)

2. Complete Subsets

2.1 Consider any subset, A, of the equations of a linear structure (a subset of the rows of the augmented coefficient matrix) and the corresponding subset, α , of the variables that appear with a non-zero coefficient in at least one of the equations of A. Let N_A equal the number of equations in A, and n_α the number of variables in α. By 1.2.1, $n_\alpha \geq N_A$

Definition 2.1: If, in $\{A, \alpha\}$, $n_\alpha = N_A$, we will say that $\{A, \alpha\}$ is complete if $n_\alpha > N_A$, we will say that $\{A, \alpha\}$ is incomplete (cf. Marschak, p. 8)

Theorem 2.1 Let $\{A, \alpha\}$ be complete and $\{B, \beta\}$ be complete. Then their intersection $\{C, \gamma\} = \{A \cdot B, \alpha \cdot \beta\}$ is complete, i.e., $n_\gamma = N_C$.

Proof: Designate by A' the complement of C in A, i.e., $(A - C)$, and by B' the set $(B - A)$. Designate by α' and β' the variables appearing in A' and B' respectively. Let $n_{\beta' - \gamma}$ be the number of variables that appear in β' but not in γ , i.e. the cardinal number of the set $(\beta' - \beta' \cdot \gamma)$.

Since $\{A, \alpha\}$ is complete and the variables of γ (n_γ in number) are included in α , we can solve A for unique values of the γ.

We have:

(3.1) $$n_\gamma + n_{\beta' \cdot \gamma} = n_\beta$$

(3.2) and $$N_C + N_{B'} = N_B$$

But since B is complete, $N_B = n_\beta$. Hence

(3.3) $$(n_\gamma - N_C) + (n_{\beta' - \gamma} - N_{B'}) = 0$$

and if we assume, contra the theorem, that $n_{\gamma'} > N_{c'}$, then

(3.4) $$N_{\beta'} > n_{\beta-\gamma}$$

which contradicts 1.2.2.

Hence $n_{\gamma} = N_{c}$ and the theorem is proved.

From this follows immediately:

Theorem 2.2. The minimal complete subsets $\{A_i, \alpha_i\}$ of a linear structure (i.e. the complete subsets that do not include complete proper subsets) are disjunct.

2.2 We can now decompose the linear structure into two parts: a part $\{A, \alpha\}$ which is the sum of the minimal complete subsets; $A = A_1 + A_2 + \dots + A_k$ and $\alpha = \alpha_1 + \alpha_2 + \dots \alpha_k$; and a remainder $\{B, \beta\}$. But, by 1.2.2., and the fact that the values of the α are uniquely determined, it follows that $N_A = n_\alpha$, i.e. that the number of variables appearing in α is equal to the number of equations in A. Further, if β is not null, we must have $n_\beta > N_B$, else $\{B, \beta\}$ were closed, contrary to definition. Hence at least one of the variables of α must belong to β.

It is convenient to distinguish three cases:

I. $\{A, \alpha\}$ consists of a single complete set - the entire structure; i.e., the structure contains ~~one~~ no complete proper subsets. In this case, we will say the structure is <u>completely integrated</u>.

II. $\{A, \alpha\}$ consists of one or more proper subsets of the structure and $\{B, \beta\}$ is not null. In this case, we will say the structure is <u>causally ordered</u>.

III. $\{A, \alpha\}$ consists of one or more proper subsets of the structure and $\{B, \beta\}$ is null. In this case, we will say the structure is <u>unintegrated</u>.
In each case, we will call the minimal complete subsets the complete subsets of zero order.

In case II, we can solve $\{A, \alpha\}$ for unique values of the αs, substitute these in the equations B, and obtain the <u>first derived structure</u>, a self-contained structure of N_β equations in $n_{\beta-\beta\cdot\alpha} = N_\beta$ unknowns. We can now find the minimal complete subsets of the first derived structure (complete subsets of first order) and proceed as before, obtaining Case I, II, or III. If Case II holds, we repeat the process with the second derived structure, and so forth. Since the number of equations in the original structure was finite, we must finally reach a derived structure that falls under Case I or Case III.

2.3 It is clear that each minimal complete subset of the first derived structure must contain at least one variable in α, else the subset were a complete subset of the original structure. Similarly, each minimal complete subset of the k^{th} derived structure must contain at least one variable that appears in a minimal complete subset of the (k-1)st derived structure and that does not appear in a minimal complete subset of any derived structure preceding (k-1) in the sequence.

<u>Definition 2.2</u> The minimal complete subset B is said to be directly causally dependent on the minimal complete subset A ($A \rightarrow B$) if (1) there is a variable x_k that appears in both d B, (2) if B belongs to a derived structure of order higher than A, and (3) if x_k does not belong to a minimal complete subset of the original structure or of a derived structure of order lower than A.

2.4 Since the variables appearing in two minimal subsets of the same order (and not in minimal subsets of lower order) are disjunct, it is clear that our relation $\rightarrow$ partitions into minimal subsets the equations and variables of our structure and partially orders these minimal subsets. (Birkhoff, p.7)

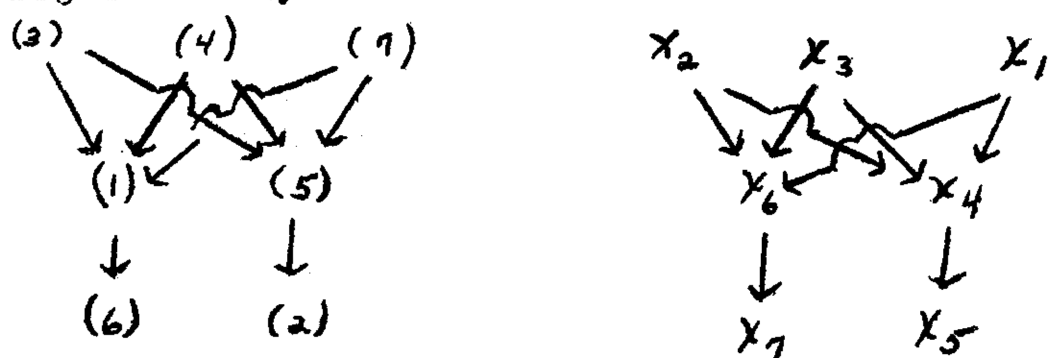
Moreover, since the variables appearing in any equation are the same under the group of S-transformation, our relationship of causality is invariant under this group.

3. Analysis of an Example

3.1 As an example, exhibiting the foregoing definitions consider the following linear structure:

$$
\begin{aligned}
&(1)\quad \alpha_{11} x_1 + \alpha_{12} x_2 + \alpha_{13} x_3 + \alpha_{16} x_6 = \alpha_{10}\\
&(2)\quad \alpha_{24} x_4 + \alpha_{25} x_5 = \alpha_{20}\\
&(3)\quad \alpha_{32} x_2 = \alpha_{30}\\
&(4)\quad \alpha_{43} x_3 = \alpha_{40}\\
&(5)\quad \alpha_{51} x_1 + \alpha_{52} x_2 + \alpha_{53} x_3 + \alpha_{54} x_4 = \alpha_{50}\\
&(6)\quad \alpha_{66} x_6 + \alpha_{67} x_7 = \alpha_{60}\\
&(7)\quad \alpha_{71} x_1 = \alpha_{70}
\end{aligned}
$$

It can be shown that there are three complete subsets of zeroth order: equation (3) and variable x_2; equation (4) and variable x_3; and equation (7) and variable x_1. There are two complete subsets of first order: Equation (1) and x_6; and equation (5) and x_4. Finally there are two complete subsets of second order: equation (2) and x_5; and equation (6) and x_7. In this case, each complete subset consists of one equation in one variable, and we can represent the causal partitioning alternatively thus:



Reordering our equations to correspond with the order of the corresponding variables, the partitioning can also be represented as follows:

eq.	x_1	x_2	x_3	x_4	x_6	x_5	x_7
(7)	x	o	o	o	o	o	o
(3)	o	x	o	o	o	o	o
(4)	o	o	x	o	o	o	o
(5)	x	x	x	x	o	o	o
(1)	x	x	x	o	x	o	o
(2)	o	o	o	x	o	x	o
(6)	o	o	o	o	x	o	x

In this figure, non-zero coefficients in the matrix are designated by x, zero coefficients by 0. The coefficients of the constant term are not displayed.

3.2 We see from this representation that ordering the equations and variables according to their causal relations places the matrix in a canonical form that in some sense is as nearly triangular as the structural equations permit. This suggests that calculation of the causal relations in a structure may have some value in indicating the optimum arrangement of equations and variables in sequencing the computation of their solutions. It would be easy to construct an electrical computing device which, even for very large structures, would rapidly locate the complete subsets from this matrix representation.

The blocks of zeros above and to the right of the main diagonal in the canonical form of the matrix show clearly also that our concept of causality is essentially identical with the concept of unilateral coupling.[1]

---

1. As a matter of fact, the writer originally approached his problem from the standpoint of unilateral coupling, and only at a very much later stage recognized its relationship to the identification problem. (Goodwin, p.)

3.3 The blocks of zeros in the lower left hand corner are really accidental properties of the particular partitioning we are studying - that variables of zero order appear only in equations of zero and first order, and not in equations of second order. This suggests an approach to a classification of structures. For example we might propose:

<u>Definition 3.1</u> A causal structure will be called modular if variables of the k^{th} order appear only in equations of the k^{th} and $(k + 1)$st orders, for all k. (Birkhoff, pp. 9-11, 34)

The causal relation we have defined is a non-transitive relation. $A \rightarrow B$ and $B \rightarrow C$ does not imply $A \longrightarrow C$. We may wish to introduce a transitive relationship of "directly or indirectly caused".

<u>Definition 3.2</u> $A \supset C$ if there exist B_1, B_2 ..., B_k such that $A \rightarrow B_1 \longrightarrow B_2 \longrightarrow \dots \longrightarrow B_k \longrightarrow C$.

If we now add to our minimal complete sets the null set and the set of all the variables we obtain a lattice on the binary asymmetrical relation $\supset$.

4. <u>Causality and Identification</u>

The point has been emphasized that the causal structure of a self-contained system is invariant under the group of S-transformations. It is not, in general, invariant under the group of R-transformations. For example, by adding to one equation a multiple of another equation we can generally alter the set of variables appearing in the first equation.

Hence, causality will have operational meaning if the restriction of admissible transformations to diagonal transformations has operational meaning. But this is exactly the restriction that is necessary to assure identifiability of a linear structure.

Stated differently, if enough <u>a priori</u> restrictions are placed upon the matrix of a self-contained system of equations to assure identifiability of the structure, these restrictions permit us to specify a determinate set of causal relationships among the equations and variables of the structure.

But what is the source of these a priori distinctions? They appear to arise in somewhat the following fashion:

We suppose a group of persons whom we shall call "experimenters." If we like, we may consider "nature" to be a member of the group. The experimenters, severally or separately, are able to choose without restriction the elements of a $n x (n+1)$ coefficient matrix. We may say that they <u>control directly</u> the values of these coefficients. Once the matrix is specified, the values of the n variables, in the n linear equations defined by the matrix are uniquely determined. Hence, the experimenters <u>control indirectly</u> the values of these variables.

The basic problems of inference are the following:

(1) Given certain incomplete information about the behavior of the experimenters (e.g. information as to the zero elements of the matrix), and certain observations of the values of some or all of the n variables, to estimate some or all of the remaining coefficients of the matrix.

(2) Given, as above, information about the behavior of the experimenters, to describe the causal structure of the variables.

The literature on identification is concerned with the first problem; the present paper with the second.

In both cases, in addition to a language describing the model, we require a meta-language describing the relationship between the "experimenters" and the model. The terms "direct control" and "indirect control" are in this metalanguage. Thus, in our metalanguage, we have an asymmetrical relationship: behavior of experimenters $>$ matrix coefficients $>$ values of variables, that must be introduced in order to establish the asymmetrical causal relationship.

In one sense, then, we have not solved our problem so much as moved it from the language of the original model to the metalanguage. But the same may be said of the theorems of identification. The application of all of these depends upon the obtainability of a priori constraints on the coefficient matrix.

5. Causality in Systems Not Self-Contained

5.1 We now proceed to show that it is essential that we assume a self-contained structure in order to introduce the notion of causality.

Consider the structure used as example in the last section. Suppose we omit equations (3) and (7), and replace them with:

(8) $$\alpha_{85} x_5 = \alpha_{80}$$

(9) $$\alpha_{99} x_9 = \alpha_{90}$$

We then obtain the following causal structure:

$$\begin{tikzcd}
x_7 \arrow[d] & x_5 \arrow[d] & x_3 \arrow[ddl] \\
x_6 \arrow[dr] & x_4 \arrow[d] & \\
& (x_1, x_2) &
\end{tikzcd}$$

where (x_1, x_2) represents the complete subset of second order comprising the variables x_1 and x_2. We see that we have not only reversed the direction of causality between x_5 and x_7 on the one hand, and x_1 and x_2 on the other, but we have also changed the relation of x_3 to the remainder of the system. Hence we cannot speak of an "internal" causal structure among the equations of a structure that is segment~~ed~~al (i.e. not self-contained) apart from the particular self-contained structure in which it is imbedded. In our new case the canonical form of the matrix is:

	x_3	x_5	x_7	x_4	x_6	x_1	x_2
(4)	x	o	o	o	o	o	o
(8)	o	x	o	o	o	o	o
(9)	o	o	x	o	o	o	o
(2)	o	x	o	x	o	o	o
(6)	o	o	x	o	x	o	o
(1)	x	o	o	o	x	x	x
(5)	x	o	o	x	o	x	x

Of the five equations common to both structures, only equation (4) has retained the same order. Moreover, the complete subsets of equations are associated with different subsets of variables than they were before.

5.2 In general, we can complete a system that is segmental by adding an appropriate number of additional equations, and in general we can do this in a number of different ways. Each of the resulting self-contained structures will likely have different causal relations among its equations and variables.

In a self-contained system all variables are endogenous. However, if we consider the system $\{B, \beta\}$, the variables $\alpha \cdot \beta$ will be exogeneous with respect to this system. Hence, we can speak of the variables of the k th order as <u>exogenous variables of the (K+1) st order</u>. To specify the exogeneous variables of a segmental structure is equivalent to completing the structure by adding equations in which these exogeneous variables alone occur.

6. <u>Further Analysis of the Identification Problem</u>

We return now to the problem of identification of a structure. We have stated loosely that the operational assumptions necessary to assure identification are the same as those necessary to define the causal structure of the system. This statement needs to be made more precise.

6.1 We consider a linear structure in the variables $x_1, \ldots, x_n$, which may be regarded as an n-vector, x. We designate by $\mathfrak{X}$ the set of points in the x-space that satisfy the equations of the structure.

Definition 6.1: A structure A is completely identified if $\mathfrak{X}_A$ is not identical with $\mathfrak{X}_A^*$, where $\mathfrak{X}_A^*$ is the set of points satisfying A^*, any other structure belonging to the same model (see Definition 1.4).

Definition 6.2: A coefficient α_{ij} is identified - in a structure A if $\mathfrak{X}_A$ is not identical with any $\mathfrak{X}_A^*$, where $\mathfrak{X}_A^*$ belongs to some A^* having $\alpha_{ij}^* \neq \alpha_{ij}$, in the same model.

Definition 6.3: An equation in a structure is identified if all its coefficients are identified.

Now identification of a coefficient is a necessary condition for the estimation of that coefficient from observations satisfying the equations of the structure. For these observations belong to $\mathfrak{X}_A$. But if $\mathfrak{X}_A \equiv \mathfrak{X}_A^*$ with $\alpha_{ij} \neq \alpha_{ij}^*$ then both the values α_{ij} and α_{ij}^* are admissible estimates for that coefficient - no set of observations can distinguish between the two structures.

On the other hand, identification of a coefficient is not a sufficient condition for the estimation of that coefficient from observations satisfying the equations of the structure. This may be seen from the following considerations.

If an identified structure A has $k \leq n$ equations, $\mathfrak{X}_A$ will be a hyperplane of (n-k) dimensions, as will also $\mathfrak{X}_A^*$. Hence if the observations have dimensionality (n-k) (they cannot have higher dimensionality and satisfy the structural equations), they cannot all belong both to $\mathfrak{X}_A$ and $\mathfrak{X}_A^*$. But if the dimensionality of the observations is less than (n-k) they may belong to the intersection of $\mathfrak{X}_A$ and $\mathfrak{X}_A^*$.

Definition 6.4: If a set of observations in n variables has (n-k) dimensions, we will say that it is full relative to a structure of k equations.

Then all the coefficients of a structure can be estimated from a set of observations if and only if the structure is completely identified and the set of observations is full relative to the structure. (This theory is due to Frisch.)

If the latter condition is not fulfilled, we may surmise that the observations are constrained by relationships in addition to the structural equations, and we may construct a new structure that comprises the old together with equations expressing these additional relationships. We can then proceed to estimate the parameters provided that the new structure is identified.

6.2 Now it has been shown (Koopmans) that a linear structure is completely identified if and only if the a priori restrictions on the model (the zeros of the coefficient matrix) are such as to permit only S-transformations upon the matrix. If the identified structure is self-contained, there will then be a unique causal structure associated with it.

On the other hand, the causal structure may be unique even if the structure is not completely identified. Since the causal structure depends only on which subset of variables appears in which subset of equations, it will be invariant over the group of R-transformations upon the coefficients of the equations of such a subset. Suppose the augmented matrix, $[A]$, of a structure be partitioned in terms of its complete subsets:

$$(6.1) \quad [A] = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_k \end{bmatrix}$$

Consider a square matrix $[B]$, such that:

$$(6.2)\quad [B] = \begin{bmatrix} B_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & B_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & B_3 & 0 & \cdots & 0 \\ 0 & \cdot & \cdot & \cdot & \cdots & 0 \\ 0 & 0 & 0 & & & B_k \end{bmatrix}$$

where $B_1, \ldots, B_k$ are non-singular square sub-matrices with numbers of rows equal, respectively, to those of $A_1, \ldots, A_k$. Then $C = BA$ will have the same causal structure as A.

6.3 Koopmans has shown that the necessary and sufficient condition for the identifiability of an equation in a linear structure is that (1) the number of variables absent from the equations be at least equal to the number of equations less one, and (2) that a certain determinant of the matrix of the structure be non-singular. The second condition is implicit in our definition of structure (particularly 1.2.2), which is more restrictive than Koopmans'. Hence, we need consider only the first condition.

Koopmans' result leads immediately to the following theorem:

<u>Theorem 6.1</u> A necessary and sufficient condition for the identifiability of an equation in a self-contained system is that it contain only one variable.

Hence, in a self-contained system, the only equations that are identified are those constituting minimal complete subsets (of zero order) in that system.

<u>Definition 6.5</u> An equation of k^{th} order in a self-contained structure, A is <u>identified to order m</u> ($m \leq k$) if it is identified in the (sectional) system consisting of the equations of A that belong to order m and higher orders.

With this definition, we obtain again from Koopmans' theorem, the following:

<u>Theorem 6.2</u> If and only if an equation of k^{th} order in A is identified in the system consisting of the equations of A of order k alone, it is

identified to order k. For the variables in A of order higher than k are just equal in number to the equations of order higher than k, and these variables are absent from the equations of k th order.

Hence in studying whether an equation of order k is identified to that order, we need to consider only the system of equations of k th order. A stronger theorem of the same kind is:

Theorem 6.3 If and only if an equation of k th order in A is identified in the system of equations belonging to its complete subset (of order k) it is identified to order k. The reasoning follows the same line as in Theorem 6.2.

6.4 The rationale of the usual observational procedure for estimating the coefficients of a structure would appear to be this. We suppose a self-contained structure, A, and a complete subset, A_k, of order k. We suppose A_k to be identified to order k, and we wish to estimate its coefficients.

All the variables of A, of order less than k, that appear in A_k are exogeneous variables relative to A_k. We now suppose that these variables can be arbitrarily varied (by relaxing the structural equations of order less than k) to produce a full set of observations relative to A_k. This full set of observations, together with the condition that the equations of A_k be identifiable, permits us to estimate the coefficients.

It is to be noted that we have here again implicitly introduced the notion of an experimenter who, by his direct control over the parameters of the equations in A of order less than k, can bring about independent variations in the variables that are exogeneous to A_k. If this procedure is operationally meaningful, the experimenter, confronted with a self-contained structure, A, can partition the structure into its complete subsets, and isolating each of these from the whole, proceed to estimate its parameters. This seems to correspond exactly to the procedure of a physiologist who (in the example used earlier) prevents an animal's saliva from reaching the palate and hence explores the thirst mechanism.

7. Causality in Non-Linear Systems

Thus far we have considered only the case of linear, non-stochastic systems. In this paper, the stochastic case will not be considered, but a few comments will be made on the non-linear case.

We consider a system of functional relations of the form:

$$\phi_i(x_1, \ldots, x_n) = 0 \qquad (i = 1, \ldots, n) \tag{7.1}$$

We suppose that no parameters appear in these equations but only universal constants (e.g., the gravitational constant). We assume further that the system has, at most, a denumerably infinite set of solutions.

Now we can again decompose the system into complete subsets of equations of various orders, such that each subset contains as many variables that do not appear in subsets of lower order as it contains equations. If appropriate conditions are imposed on our system, this decomposition will again be unique.

In our linear system, we assumed that an experimenter could directly control the parameters appearing in the equations. In the present case, we assume that an experimenter can relax any equation, or set of equations in the system. Then we have the same general relationship between the problem of defining causal structure and the problem of identification as in the linear case.

8.1 Conclusion

In this paper we have defined a concept of causality that corresponds to the intuitive use of that term in scientific discussion. Causality is an asymmetrical relation among certain variables, or subsets of variables, in a self-contained system of equations. There is no necessary connection between the asymmetry of this relation and asymmetry in time, although an analysis of the causal structure of dynamical systems in econometrics and physics will show that lagged relations (and relations involving derivatives) can generally be interpreted as causal relations. (The writer has carried out this analysis for certain simple mechanical and thermodynamical systems.)

The concept of causality has been shown to be intimately connected with the concept of identifiability, although the conditions under which a system possesses a unique causal structure are somewhat weaker than the conditions under which it is completely identified.

A study of the operational meaning of the causal relation (or of a structural relation) appears to require a metalanguage that permits discussion of the relation between the system of equations and an experimenter who has direct control over the parameters of the system. As the brief discussion of the non-linear case implies, the distinction between parameters and variables can be disregarded if the former are regarded as exogeneous variables (determined by a larger system) with respect to the latter. In this case, the experimenter must be regarded as being able to relax or alter particular equations in this larger system.