# Causal Ordering in a Mixed Structure[1]

**Yumi Iwasaki**[2]

**Department of Computer Science**
**Carnegie Mellon University**
**Pittsburgh, Pennsylvania 15213**

## Abstract

This paper describes a computational approach, based on the theory of causal ordering, for inferring causality from an acausal, formal description of a phenomena. Causal ordering is an asymmetric relation among the variables in a self-contained equilibrium and dynamic structure, which seems to reflect people's intuitive notion of causal dependency relations among variables in a system. This paper extends the theory to cover models consisting of mixture of dynamic and equilibrium equations. When people's intuitive causal understanding of a situation is based on a mixed description, the causal ordering produced by the extension reflects this intuititve understanding better than that of an equilibrium description. The paper also discusses the view of a mixed model as an approximation to a completely dynamic model.

## 1. Introduction

Much of sciences and engineering is concerned with characterizations of processes by equations that describe the relations that hold among parameters of objects and that govern their behavior over time. Formal treatment of the foundations of sciences have avoided notations of causation and spoke only of functional relations among variables. Nevertheless, the notion of causality plays an important role in our understanding of phenomena. Even when a formal description of a situation is given in terms of acausal, mathematical relations, informal explanations of the form, "A causes B" are exceedingly common. People are able to explain the behavior in causal terms while using an acausal formal description of a situation.

This paper describes a computational approach, based on the theory of causal ordering, for inferring causality from an acausal, formal description of a phenomena. Causal ordering, first presented by Simon [Simon 52], is an asymmetric relation among the variables in a self-contained equilibrium or dynamic model, which seems to reflect people's intuitive notion of causal dependency relations among variables. This intuitive causal understanding of a situation is based on a mixed description, the causal ordering produced by the extension reflects this intuitive understanding better than that of an equilibrium description. The paper also discusses the view of a mixed model as an approximation to a completely dynamic model.

## 2. Causal ordering in an equilibrium structure

Causal ordering was initially defined by Simon for an equilibrium structure consisting of equilibrium equations [Simon 52]. First, we define an equilibrium structure:

**Definition 1:** Self-contained equilibrium structure

A self-contained equilibrium structure is a system of $n$ equilibrium equations in $n$ variables that possesses the following special properties:

1. That in any subset of $k$ equations taken from the structure at least $k$ different variables appear with nonzero coefficients in one or more of the equations of the subset.
2. That in any subset of $k$ equations in which $m \geq k$ variables appear with nonzero coefficients, if the values of any $(m - k)$ variables are chosen arbitrarily, then the equations can be solved for unique values of the remaining $k$ variables.

The condition (1) above ensures that no part of the structure is over-determined. The condition (2) ensures that the equations are not dependent because if they are the equations cannot be solved for unique values of the variables.

The idea of causal ordering in a self-contained equilibrium structure can be described roughly as follows. A system of $n$ equations is called *self-contained* if it has exactly $n$ unknowns. Given a self-contained system, $S$, if there is a proper subset, $s$, of $S$ that is also self-contained and that does not contain a proper self-contained subset, $s$ is called a *minimal complete subset*. Let $S_0$ be the union of all such minimal complete subsets of $S$; then $S_0$ is called the set of *minimal complete subsets of zero order*. Since $S_0$ is self-contained, the values of all the variables in $S_0$ can, in general, be obtained by solving the equations in $S_0$. By substituting these values for all the occurrences of these variables in the equations of the set $(S - S_0)$, one obtains a new self-contained structure, which is called the *derived structure of first order*. Let $S_1$ be the set of minimal complete subsets of this derived structure. It is

[1]This research was sponsored by the Defense Advanced Research Projects Agency (DOD), ARPA Order No. 4976 under contract F33615-87-C-1499.

[2]The author's current address is: Knowledge Systems Laboratory, Department of Computer Science, Stanford University, 701 Welch Rd., Palo Alto, CA 94304

called the set of *complete subsets of 1st order*. Repeat the above procedure until the derived structure of the highest order contains no proper subset that is self-contained. For each equation $e_i$ in $S$, let $V_i$ denote the set of variables appearing in $e_i$, and let $W_i$ denote the subset of $V_i$ containing the variables belonging to the complete subsets of the highest order among those in $V_i$. Then, the variables in $W_i$ are said to be *directly causally dependent* on the elements in $(V_i - W_i)$.

In order for the above procedure to produce causal relations in the model that agrees with our intuitive understanding of the causal relations in the real situation, the equations comprising a model come from an understanding of mechanisms. The term *mechanism* is used here in a general sense to refer to distinct conceptual parts in terms of whose functions the working of the whole system is to be explained. Mechanisms are such things as laws describing physical processes or local components that can be described as operating according to such laws. An equation representing such a mechanism is called a *structural equation*, and every equation in the model should be a structural equation standing for a mechanism through which variables influence other variables.

One thing to note about the method of causal ordering is that it does not require knowledge about the precise functional forms of equations. The only information that the method makes use of is what variables appear with a non-zero coefficient in what equations, which in terms of mechanisms translates to what variables are causally linked by each mechanism.

## 3. Example : Bathtub

Though the causal structure produced by the method of causal ordering usually agrees with people's intuitive notions of causal relations [Iwasaki and Simon 86, Iwasaki 87], sometimes cases arise where a causal structure produced does not agree with human intuition. We present one such case to motivate extension of the method to dynamic and mixed structures.

The device used as an example is a bathtub as shown in Figure 3-1 [Kuipers 87a]. There are four variables; the input and output flow rates $Q_{in}$ and $Q_{out}$, the amount of water in the tub, $A$, the valve opening, $K$, and the pressure in the bottom of the tub, $P$.

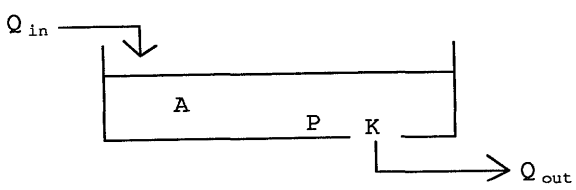


**Figure 3-1:** Bathtub

The situation can be characterized by the following four equations, where $c_1$, $c_2$, and $c_3$ represent positive constants.

$$Q_{out} = KP \tag{1}$$

The output flow rate is proportional to the pressure.

$$A = c_1 P \tag{2}$$

The pressure is proportional to the amount of water

$$Q_{out} = Q_{in} \tag{3}$$

When the system is in equilibrium, the input flow equals the output flow.

$$Q_{in} = c_2 \tag{4}$$

$$K = c_3 \tag{5}$$

The input flow rate and the valve opening are exogenous. (externally controlled)

The causal ordering produced for this bathtub model is shown in Figure 3-2. "$x$ --> $y$" means that variable $y$ is causally dependent on $x$.

$$\begin{tikzcd} Q_{in} \arrow[r] & Q_{out} \arrow[r] & P \arrow[r] & A \\ & K \arrow[u] & & \end{tikzcd}$$

**Figure 3-2:** Equilibrium Causal Ordering of Bathtub

The causal structure shown in Figure 3-2 may seem counter-intuitive. It shows that the output flow rate directly depends on the input flow rate, the pressure depends on the output flow rate, and the amount depends on the pressure. However, intuitively speaking, adding water to the tub increases the amount ($A$), which increases the pressure ($P$), which in turn increases the output flow rate ($Q_{out}$). Figure 3-3 shows this "intuitive" causal ordering.

$$\begin{tikzcd} Q_{in} \arrow[r] & A \arrow[r] & P \arrow[r] & Q_{out} \\ & & & K \arrow[u] \end{tikzcd}$$

**Figure 3-3:** Intuitive Causal Ordering of Bathtub

In what follows I will first show that the causal ordering in Figure 3-2 is in fact the correct ordering for an equilibrium model. In the next section, I will show that the "intuitive" causal ordering can be obtained by the extension of causal ordering to make it applicable to systems of dynamic equation as well as mixture of dynamic and equilibrium equations.

In order to see that the ordering in Figure 3-2 is correct, one must realize that the model is an equilibrium one. In an equilibrium model, quantities represent the final values assumed by variables when equilibrium is attained and not transient values. In the bathtub example, if the input flow is decreased suddenly, it will cause immediate disturbances in the values of other variables. However, the entire system will be in a steady state only when the output flow again becomes equal to the input flow, which is the situation the equilibrium model depicts.

Suppose that the value of $K$ is changed by opening up the valve a little more, then an immediate reaction will be that $Q_{out}$ will increase. However, when equilibrium is restored eventually, assuming that it will, the equilibrium value of $Q_{out}$ must be equal to $Q_{in}$ (otherwise, the system would not be in equilibrium). Thus, changing $K$ only affects the equilibrium values of $P$ and $A$ but not $Q_{out}$. Therefore, equilibrium value of $Q_{out}$ cannot be dependent on $P$ or $A$, a fact correctly reflected in the ordering of Figure 3-2 but not in Figure 3-3.

Nevertheless, it is true that ordering in Figure 3-3 seems to capture *some* intuitive notion of causality in the situation. It is because in this case our "intuitive" causal understanding is of a dynamic situation rather than that of the equilibrium situation represented by the model above. The next section presents causal ordering in dynamic and mixed systems.

# 4. Causal ordering in a dynamic and mixed structures

In this section, we define self-containment and causal ordering for structures consisting of differential equations and mixture of differential and equilibrium equations. We will then show that the "intuitive" causal ordering similar to that in Figure 3-3 emerges as the causal ordering in a mixed model of the bathtub.

## 4.1. Causal ordering in a dynamic structure

Dynamic causal ordering is defined for systems consisting of first order differential equations. Since a differential equation of higher order can be converted into a set of first order equations by introducing new variables to stand for derivatives, the definition of causal ordering presented here applies to a very wide class of dynamic systems.

Following is the definition of a self-contained dynamic structure [Simon and Rescher 66]:

**Definition 2:** A self-contained dynamic structure

A self contained dynamic structure is a set of $n$ first-order differential equations involving $n$ variables such that:

1. In any subset of $k$ equations of the structure the first derivative of at least $k$ different variables appear.
2. In any subset of $k$ equations in which $r$ ($r \geq k$) first derivatives appear, if the values of any $(r - k)$ first derivatives are chosen arbitrarily, then the remaining $k$ are determined uniquely as functions of the $n$ variables.

The above definition of self-containment for a dynamic structure is analogous to that for an equilibrium structure. The condition (1) above ensures that no part of the structure is over-determined while the condition (2) ensures that the structure is not under-constrained.

Given a self-contained dynamic structure, one can perform elementary row operations to the equations to solve them for the $n$ derivatives. This operation produces an equivalent system of equations in *canonical form*. A differential equation is said to be in *canonical form* if and only if there is only one derivative in the equation, and the derivative is the only thing appearing on the left-hand-side of the equations. A self-contained dynamic structure in $n$ variables, $x_1, \ldots, x_n$, in canonical form consists of $n$ equations of the following form, where $f_i$'s ($1 \leq i \leq n$) are functions of their arguments:

$$x_i' = f_i(x_1, x_2, \ldots, x_n)$$

We interpret the equations of structure in this form to be *mechanisms* of the system. Therefore, the $i$th equation, the only one containing $x_i'$, is regarded as the mechanism determining the time path of $x_i$. Furthermore, variable $x_i$, whose derivative appear in the $i$th equation, is said to be directly causally dependent on the variables that appear with a non-zero coefficient in the equation.

## 4.2. Causal ordering in a mixed model

Systems are in practice often described in terms of a combination of equilibrium and dynamic equations. A such mixed structure is a natural extension of dynamic structures.

Before defining self-containment for mixed structures, we must introduce some notations. Let $M$ be a system of $n$ equations in $n$ variables such that some of the equations are equilibrium equations and others are first-order differential equations. Then, let *Dynamic*($M$) be the subset of $M$ consisting of all the differential equations in $M$, and let *Static*($M$) be the set consisting of all the equilibrium equations in $M$ and one constant equation for every variable $v$ whose derivative appears in *Dynamic*($M$). A constant equation of a variable is an equation of the form, $v = c$, where $c$ is a constant.

The intuitive meaning of the set *Static*($M$) may be understood as follows: the equilibrium equations in a mixed set represent mechanisms that restore equilibrium so quickly that they can be considered to hold in 0 units of time within some time-frame (e.g. days if the time-frame is centuries). On the other hand, the dynamic equations represent slower mechanisms that require non-zero amounts of time for the variables on their right hand sides to affect the variable on their left hand sides. Therefore, in a very short period of time -- shorter than is required for the variables on the right hand side of the differential equation of a slow mechanism to appreciably affect the variable on the left hand side -- the variable on the left hand side can be considered unchanging. Thus, the set *Static*($M$) represents a snap-shot picture (i.e., a very short-term equilibrium description) of the dynamic behavior of mixed structure $M$.

Let $M$ be a system of $n$ equations in $n$ variables such that some of the equations are static equations and others are dynamic equations of the type defined in the previous section.

**Definition 3:** The set $M$ of $n$ equations in $n$ variables is a self-contained mixed structure iff:

1. One or more of the $n$ equations are first-order

differential equations and the rest are equilibrium equations.

2. In any subset of size $k$ of *Dynamic(M)*, the first derivative of at least $k$ different variables appear.

3. In any subset of size $k$ of *Dynamic(M)* in which $r$ ($r \geq k$) first derivatives appear, if the values of any ($r - k$) first derivatives are chosen arbitrarily, then the remaining $k$ are determined uniquely as function of the $n$ variables.

4. The first derivatives of exactly $d$ different variables appear in *Dynamic(M)* where $d$ is the size of the set *Dynamic(M)*.

5. *Static(M)* is a self-contained equilibrium structure.

Given a self-contained mixed structure, as defined above, the causal ordering among its variables and derivatives follow the definitions of causal ordering in dynamic and static structures. In other words, the causal ordering in a mixed structure can be determined as follows:

1. The ordering among $n$ variables and $m$ derivative in subset *Dynamic(M)* is given by the definition of causal ordering in a dynamics structure.

2. The ordering among variables (but not their derivatives) in *Static(M)* is given by the definition of causal ordering in an equilibrium structure.

### 4.3. Mixed model of the bathtub

Now, we are ready to look at the bathtub example again. Let $M$ be a mixed structure consisting of equations (1), (2), (4), (5) and the following differential equation in place of (3);

$$A' = Q_{in} - Q_{out} \qquad (3d)$$

$M$ is a self-contained, mixed structure according to the definition given above. *Dynamic(M)* consists of equation (3d) alone, and *Static(M)* consists of equations (1), (2), (4), (5) and the following constant equation:

$$A = c_3 \qquad (3e)$$

The causal ordering in $M$ is shown in Figure 4-1. In the figure, an integration link, which is an edge connecting a derivative of a variable to the variable itself, is marked by $i$, causal links in the dynamic part of the model (*Dynamic(M)*) are indicated by arrows of broken lines. The causal structure indicates existence of a feedback loop. The structure may be explained informally in English as follows:

> The output flow rate depends on the pressure, which depends on the amount of water the in the tub. The rate of change of the amount of water is determined by the input and output flow rates.

$$\begin{tikzcd}
Q_{in} \arrow[d] & & & K \arrow[d] \\
A' \arrow[r, "i"] & A \arrow[r] & P \arrow[r] & Q_{out} \arrow[lll, bend left]
\end{tikzcd}$$

**Figure 4-1:** Mixed Causal Ordering of Bathtub

The reason for mixing in one model equilibrium equations (1) and (2) and differential equation (3d) is because the equilibrium relations represented by the first three equations are restored much more quickly (in fact, almost instantaneously) when disturbed than the relation represented by equation (3). Therefore, in a model of a medium temporal grain-size, it is reasonable to treat (3) as taking time but to treat others as instantaneous.

## 5. A mixed structure as an approximation to a dynamic structure

A mixed structure can be viewed as an approximation to a dynamic structure. When a mechanism in a dynamic structure acts very quickly to restore relative equilibrium, one can regard it to be acting instantaneously. Or, when a mechanism acts so much more slowly than other mechanisms in the system that its effect on the variable it controls is negligible, the variable may be considered constant. In these cases, the description of the system's dynamic behavior may be simplified by replacing the fast-acting mechanism by an equilibrium equation or the slow mechanism by a constant equation. This section discusses generating a mixed structure from a dynamic structure as an approximation to the latter through these two techniques.

### 5.1. Equilibrating Dynamic Equations

We will use the term *equilibrating* to refer to the operation of replacing a dynamic equation by its corresponding equilibrium equation. Since the differential equations are assumed to be in canonical form, equilibration is accomplished by replacing the left hand side by 0.

There are a whole range of mixed structures between the completely dynamic structure and the equilibrium structure depending on the temporal grain size selected for the model. However, substituting an arbitrary subset of a dynamic self-contained structure with the corresponding static equations will not necessarily produce a self-contained mixed structure. Moreover, not every self-contained dynamic structure produces a self-contained equilibrium structure when every equation is replaced by the corresponding equilibrium equation.

Let us call a variable *self-regulating* if its derivative is a function of the variable itself, and *non-self-regulating* otherwise.

**Definition 4:** Self-regulating variables and equations

A differential equation in canonical form is called *self-regulating* if the variable whose derivative is the left hand side of the equation also appears on the right hand side with a non-zero coefficient. Such a variable is also called a self-regulating variable.

It can be proved that equilibrating any number of self-regulating equations in a self-contained dynamic or mixed structure will always produce a self-contained mixed structure (or a self-contained equilibrium structure if no more dynamic equations are left). However, equilibrating a non-self-regulating equation may produce an over-constrained structure. The following theorem states this fact. The proof is given elsewhere [Iwasaki 88].

> **Theorem 5:** Equilibrating any number of self-regulating equations in a self-contained dynamic or mixed structure always produces a self-contained mixed structure (or a self-contained equilibrium structure if all the dynamic equations in the original structure have been equilibrated.)

### 5.2. Exogenizing Dynamic Equations

In contrast to variables that adjust to changes in other variables very quickly to restore relative equilibrium, some variables responds so slowly to changes in other variables that they can be regarded as independent of other variables. The equation corresponding to such a variable can be replaced by an exogenous variable equation, which amounts to deleting from the system under consideration the slow mechanism through which others influence this variable. We will call this operation of replacing a dynamic equation by an exogenous variable equation *exogenizing*. There are two ways to exogenize a variable:

Case 1: If a variable $x_i$ is changing but the rate of change depends mostly on $x_i$ itself and very little on other variables, they can be deleted from the expression on the right hand side of the differential equation to make it a function of $x_i$ alone.

Case 2: If a variable is not only unaffected by other variables but is hardly changing, the dynamic equation can be replaced by a constant equation of the variable.

Conceptually, exogenizing is the opposite of equilibrating, because exogenizing a variable assumes it is unaffected by other variables while equilibrating a variable assumes it responds to changes in other variables extremely quickly to restore equilibrium. Exogenizing a variable amounts to deleting a mechanism from the system by placing the mechanism determining the value of the variable outside the scope of the system under consideration, and it is reasonable to do so only when the feedback to the variable from the variables inside the mechanism is negligible. Exogenizing a variable in a self-contained structure always produces a self-contained structure. The proof, given elsewhere, follows directly from the definition of self-containment of a mixed structure [Iwasaki 88].

> **Theorem 6:** Exogenizing an equation in a self-contained dynamic or mixed structure always produces a self-contained structure.

### 5.3. Bathtub example revisited

Consider a totally dynamic model, $D$, of the bathtub example consisting of equation (3d) and the following equations. The causal ordering in this dynamic structure is shown in Figure 5-1.

$$Q'_{out} = c_4(Q_{out} - KP) \quad (1d)$$

$$P' = c_5(rP - A) \quad (2d)$$

$$Q'_{in} = c_6 \quad (4d)$$

$$K' = c_7 \quad (5d)$$

$$\begin{tikzcd}
Q_{in} \arrow[d] & Q'_{in} \arrow[l, "i"'] & & K' \arrow[r, "i"] & K \arrow[d] & \\
A' \arrow[r, "i"] & A \arrow[r] & P' \arrow[r, "i"] & P \arrow[r] \arrow[l, bend right] & Q'_{out} \arrow[r, "i"] & Q_{out} \arrow[l, bend right] \arrow[lllll, bend left]
\end{tikzcd}$$

**Figure 5-1:** Causal Ordering in $D$

If it assumed that the mechanism represented by equation (1d) acts very quickly to restore equilibrium, one can replace the equation by the corresponding equilibrium equation (1). It can be easily verified that the resulting mixed structure is self-contained. Likewise, replacing equation (2d) by the corresponding equilibrium equation also results in a self-contained mixed structure. If both equation (1d) and (2d) are equilibrated, the result is also a self-contained mixed structure. The mixed structure $M$ in Section 4.3 is produced by assuming the mechanisms of (1d) and (2d) to act very quickly and also assuming at the same time that $Q_{in}$ and $K$ are hardly changing.

However, if it was assumed that the mechanism represented by equation (3d) acted very quickly but that the mechanisms of (1d) (2d) were slow, the resulting mixed structure, $M'$, consisting of equations (1d), (2d), (4d), (5d) and (3) would not be self-contained because *Static*($M'$) is not self-contained.

This fact can be intuitively explained by examining the causal structure in Figure 5-1. Since the only causal path from $Q_{in}$ to $Q_{out}$ in the causal graph is ($Q_{in}$, $A'$, $A$, $P'$, $P$, $Q_{out}'$, $Q_{out}$), the equilibrium between $Q_{in}$ and $Q_{out}$ cannot be restored before $A$ and $P$ are restored to equilibrium. Therefore, it produces a contradiction to equilibrate equation (3d) without equilibrating (1d) and (2d) at the same time.

## 6. Discussion

We have extended the method of causal ordering to dynamic and mixed structures. Making assumptions about relative speeds of adjustment in mechanisms in a dynamic structure amounts to classifying the variables into three categories;

1. Variables whose rates of change are influenced only very little by other variables;
2. Variables that adjust so quickly that they are always close to relative equilibrium with other variables;
3. All other variables.

This idea is closely related to that of aggregation of nearly decomposable dynamic system by Simon and Ando. Nearly decomposable systems are those consisting of subsystems of variables such that the variables within a subsystem interact strongly while the interactions among subsystems are much weaker. They showed that when a given dynamic system is nearly decomposable, and if one is only interested in the long-term dynamics of the system, then one can aggregate the subsystems, assuming them to be always in steady-state relative equilibrium, and consider only the movements of the aggregated systems [Simon and Ando 61]. Their work provides theoretical justification for generation of a mixed structures as an abstraction of a completely dynamic structures using the techniques discussed in Section 5.

The idea of abstraction by time-scale is used by Kuipers [Kuipers 87b] in order to control the exponential growth of the number of possible courses of behavior in qualitative simulation. The techniques discussed here can be used to generate models of different time-scales.

The approaches described in this paper have been fully implemented as part of a computer program named CAOS for reasoning about system behavior in the domain of a coal power plant. The program consists of a collection of modules for generation of equation models, causal analysis of models, dynamic stability analysis, and qualitative prediction of the effects of external disturbance. The method of causal ordering in a mixed system has also been used in a program called YAKA to perform diagnosis of faults in oil refinery plant [Lambert et al. 88].

## Acknowledgment

The author would like to thank Professors Herb Simon and John McDermott for their guidance in conducting this research, and Professor Ben Kuipers for helpful discussions of the bathtub example.

## References

[Iwasaki 87] Iwasaki, Y. *Generating Behavior Equations from Explicit Representation of Mechanisms.* Technical Report CMU-CS-87-131, Department of Computer Science, Carnegie Mellon University, 1987

[Iwasaki 88] Iwasaki, Y. *Model-Based Reasoning of Device Behavior with Causal Ordering.* PhD thesis, Department of Computer Science, Carnegie Mellon University, 1988

[Iwasaki and Simon 86] Iwasaki, Y., and Simon, H. A. Causality in Device Behavior. *Artificial Intelligence29* 1986

[Kuipers 87a] Kuipers, B. The Bathtub Counterexample. Unpublished manuscript.

[Kuipers 87b] Kuipers, B. Abstraction by Time-Scale in Qualitative Simulation. In *Proceedings, 6th National Conference on Artificial Intelligence,* 1987

[Lambert et al., 88] Lambert, H., Eshelman, L., and Iwasaki, Y. Using Qualitative Physics to Guide the Acquisition of Diagnostic Knowledge. In *Proceedings of the Third International Conference on Applications of Artificial Intelligence in Engineering,* 1988

[Simon 52] Simon, H. A. On the definition of the causal relation. *Journal of Philosophy*49:517-28, 1952

[Simon and Ando 61] Simon, H. A. and Ando, A. Aggregation of Variables in Dynamic Systems. *Econometrica*29, 1961

[Simon 66] Simon, H. A. and Rescher, N. Causes and Counterfactual. *Philosophy of Science*33:323-40, 1966