

Caveats For Causal Reasoning With Equilibrium Models.

Denver Dash
Decision Systems Laboratory
Intelligent Systems Program
University of Pittsburgh
ddash@sis.pitt.edu

Marek J. Druzdzal
Decision Systems Laboratory
Intelligent Systems Program
School of Information Sciences
University of Pittsburgh
marek@sis.pitt.edu

October 12, 2002

Abstract

In this paper we examine the ability to perform causal reasoning with equilibrium models. We explicate a postulate, which we term the *Manipulation Postulate*, that is required in order to perform causal inference, and we prove that there exists a general class of recursive equilibrium models that are guaranteed to violate the Manipulation Postulate. In addition, we show that all models in this class possess a set of variables V' whose manipulation will cause an instability such that no equilibrium model will exist for the system. We define the *Structural Stability Principle* which provides a graphical criterion for stability in causal models. Our theorems suggest that caution should be exercised when applying causal reasoning to equilibrium models or to models learned from databases wherein features were not measured simultaneously.

1 Introduction

An explicit representation of causality is important in artificial intelligence primarily because it provides foundations for predicting the effects of an agent's actions. A causal model releases an agent from the need to store a combinatorially large set of pairs $\{action \Rightarrow effect\}$, so that the result of external manipulation on the model variables can be predicted directly from the model. Causal reasoning plus the ability to learn causal models from data would enable an intelligent agent to build and test hypotheses about its environment and could help automate the process of scientific discovery from data. All of these are topics that sit on the forefront of artificial intelligence research.

In addition to artificial intelligence, causal modelling is of fundamental importance in econometrics, psychology, sociology, and biology, where causal theories, often in the form of linear structural equation models (SEMs), are an

important aspect of a research paradigm that aids researchers in designing randomized studies and interpreting the results of those studies. Indeed, causal reasoning originated in early genetics and developed over several decades in the econometrics literature [Wright, 1921, Haavelmo, 1943, Koopmans, 1953, Simon, 1953, Strotz and Wold, 1960] in the context of structural equation models.

Critical to these formalisms is the assumption that when some variable in a causal model is manipulated, the net result from a structural standpoint will be *the removal of causal arcs coming into that variable*. In this paper we label this fundamental assumption the *Manipulation Postulate*. The Manipulation Postulate, which will be formally defined in Section 2, is based on our conception of what a “causal model” is together with our conception of what it means to “manipulate” a variable. In this paper, we identify a class \mathcal{F} of causal models that are guaranteed to violate the Manipulation Postulate. We prove that whether or not a causal model obeys the Manipulation Postulate depends on at least two factors: First is the time-scale over which the system is being modelled, second is the resolution, or abstraction level of the model. Our results apply to the simplest, presumably most well-understood types of causal models: recursive (i.e., acyclic) linear models with independent noise terms in the absence of latent-variables (however, our theorems do not rely on the linearity assumption).

To our knowledge, there does not exist a discussion in either the econometrics/SEM literature or in the artificial intelligence/causal theory literature of the extent to which the Manipulation Postulate will be obeyed. The closest example is that of *reversibility* [Druzdzel, 1992, Spirtes *et al.*, 1993, Druzdzel and van Leijen, 2001]¹. However, examples of reversibility have heretofore required the “releasing” of components of the system in order for the reversibility to be apparent; whereas the phenomenon discussed in this paper will occur even while maintaining the state of all exogenous variables (reversibility will be discussed in more detail in Section 2). Despite the complete absence of consideration of this issue in the past literature, we show that the size of the class \mathcal{F} is surprisingly large, encompassing a wide array of the most common physical systems. We also show that every model in \mathcal{F} displays reversibility-type behavior, thereby possibly providing a mathematical basis for reversibility and a set of sufficient conditions for it to occur, while at the same time indicating that it is a more general and perhaps widespread problem than previously suspected.

We consider dynamic models based on continuous differential equations, and make use of a technique developed by Iwasaki and Simon [1994] to derive equilibrium causal models from their dynamic counterparts. Iwasaki and Simon were apparently the first to discuss the relationship between dynamic causal models and recursive equilibrium causal models. However, there has been other work relating dynamic models to non-dynamic models in general: Richardson

¹Galles and Pearl [1997] and subsequently Halpern [2000] prove a theorem which they label “reversibility”; however this concept of reversibility, essentially a re-statement of their condition of uniqueness of solutions, has nothing to do with the reversal of causal arcs under manipulation.

[1996] discusses the relationship between independencies in dynamic models and in non-recursive equilibrium causal models; Fisher [1970] discusses the relationship between a time-varying model and its time-averaged counterpart; and Kuipers [1987] discusses temporal abstraction in dynamic qualitative models with widely varying time-scales.

In Section 2 we review the formalisms for modelling causality and manipulation, and we define the Manipulation Postulate. In Section 3 we use a simple physical example to illustrate how changing the resolution of a model can cause it to violate the Manipulation Postulate. In Section 4 we show that the causal model of our example *obeys* the Manipulation Postulate and explains the behavior observed in our example when it is modelled over shorter time scales. In Section 5 we define a general class of recursive equilibrium models and prove that this class is guaranteed to violate the Manipulation Postulate in at least two different ways. Finally, in Section 6 we discuss the implications of our theorems.

We will use the following notation throughout the remainder of the paper: If $G = \langle V, A \rangle$ is a directed graph with vertex set V and arc set A , we will use $\text{Pa}(v)_G$ and $\text{Ch}(v)_G$ to denote the parents and children, respectively, in G , for some $v \in V$. We will use $\text{Anc}(v)_G$ and $\text{Des}(v)_G$ to denote the ancestors and descendants of a variable v in a directed graph G . If e is an equation then we use $\text{Params}(e)$ to denote the set of variables constrained by e . If E is a set of equations, we use $\text{Params}(E)$ to represent $\bigcup_{e \in E} \text{Params}(e)$.

2 Causal Modelling and Reasoning

There exist several distinct notions of a *causal model*, all of which involve in some way a directed graph wherein if an arc $x \rightarrow y$ exists between two variables x and y , then x is deemed to be a cause of y . These conceptions may differ as to, given a certain specification of a system, which arcs should be present in the directed graph.

In this section we define the predominant concept of causality, and we define a set of assumptions which are intended to make other conceptions agree as to the set of arcs that should be present in the model. Finally, we define how to use a causal model to predict the effect of intervening on components of the system being modelled.

2.1 Structural Equation Models

The predominant conception of causality, known as either a *structural equation model* [Simon, 1953, Strotz and Wold, 1960, Spirtes *et al.*, 1993] or a *causal theory* [Pearl, 2000] is based on a set of equations which are deemed to determine the values of variables in the model.

In this conception, a system is summarized by a set of feature variables V , relations are specified by a set of equations E which determine unique solutions for all $v \in V$, and each variable $v \in V$ is associated with a single equation

$e \in E$. Such a specification of a causal system defines a directed graph over the variables by defining the parent set of v to be the remaining parameters of e which are also in V . An example of such a model is shown in Figure 1. It is

$e_1: f_1(w, z, v, x, \gamma_1) = 0$	(mapped to w)	
$e_2: f_2(x, \gamma_2) = 0$	(mapped to x)	
$e_3: f_3(x, y, z, \gamma_3) = 0$	(mapped to z)	
$e_4: f_4(y, z, \gamma_4) = 0$	(mapped to y)	
$e_5: f_5(v, \gamma_5) = 0$	(mapped to v)	

Figure 1: An example causal model.

sufficient for our purposes to consider only models such that any equation $e \in E$ can be freely inverted for any variable $v \in Params(e)$, e.g., e_3 in Figure 1 could be rewritten as $y = f_3^{-1}(x, z, \gamma_3)$.

A set of equations that make up a SEM must be *self-contained*. This notion is defined precisely in [Simon, 1953] and [Iwasaki and Simon, 1994]. We take the definition to be any set of equations which specifies a unique solution set for a corresponding set of variables V :

Definition 1 (self-contained set of equations) *A set of equations E is self-contained with respect to a set of variables V if $|V| = |E|$ and if E entails a unique solution for every $v \in V$.*

In Figure 1 we intentionally wrote each equation as an *implicit* function of the parameters. This was done to emphasize that in general there may be multiple ways to map equations to variables, each of which could result in a different causal structure. For example, in Figure 1 if e_3 was mapped to y and e_4 was mapped to z , then x would be a parent of y in the causal graph instead of z . To specify a structural equation model one must therefore first specify a *total causal mapping*:

Definition 2 (total causal mapping) *If E is a self-contained set of n equations with respect to a set of variables V , then a total causal mapping is an onto mapping $\phi : V \rightarrow E$ such that $v \in Params(\phi(v))$ for all $v \in V$. A total causal mapping ϕ can be written equivalently as a list of associations: $\phi = \{\langle v_1, e_1 \rangle, \langle v_2, e_2 \rangle, \dots, \langle v_n, e_n \rangle\}$.*

A set of equations E plus a set of variables V and a total causal mapping formally define a *structural equation model*:

Definition 3 (structural equation model) *A structural equation model M is a triple $M = \langle V, E, \phi \rangle$, where E is a self-contained set of equations with respect to V , and $\phi : V \rightarrow E$ is a total causal mapping.*

We will use the terms “structural equation model” and “causal model” interchangeably. A graphical representation of a causal model is a *causal graph*:

Definition 4 (causal graph) A causal graph G is a pair $\langle V, A \rangle$ where V is a set of vertices and A is a set of directed arcs $x_i \rightarrow x_j$ for vertices $x_i, x_j \in V$.

A structural equation model $M = \langle V, E, \phi \rangle$ defines a causal graph $G = \langle V', A \rangle$ by defining $V' = V$ and by defining the parent set of each variable $v \in V$ to be the remaining set of parameters of $\phi(v)$: $A = \{p \rightarrow v \mid v \in V \wedge p \in \text{Params}(\phi(v)) \setminus v\}$

It will be sufficient for the purposes of this paper to consider recursive models only:

Definition 5 (recursive causal model) A causal model $M = \langle V, E, \phi \rangle$ with a causal graph G is recursive if G is acyclic.

The following lemma shows that if M is a recursive model, then there exists exactly one mapping from equations to variables:

Lemma 1 If $M = \langle V, E, \phi \rangle$ is a recursive structural equation model then ϕ is unique, i.e., for any other structural equation model $M = \langle V, E, \phi' \rangle$, it must be the case that $\phi(v) = \phi'(v)$ for all $v \in V$.

Proof: A more general version of this proof is given in [Dash and Druzdzel, 2000]. We prove this result by induction. Let G be the causal graph corresponding to M . Define an ordering O of associations in ϕ such that $O(\langle v_i, e_i \rangle) < O(\langle v_l, e_l \rangle)$ if $v_i \in \text{Anc}(v_l)_G$. We will label the variables and equations according to the order in which their pair appears in O , thus e_i is the equation that appears in the i th pair in O . Assume that ϕ' is some other causal mapping. Let $\langle v_j, e_k \rangle \in \phi'$ be an arbitrary pair and assume that $\phi'(v_m) = \phi(v_m)$ for all $m < k$. We have ordered our equations such that $O(v) \leq O(e_k)$ for all $v \in \text{Params}(e_k)$. Thus it must be the case that $j \leq k$; however by the induction hypothesis it must be the case that $j \geq k$. Thus $j = k$. Finally, e_0 is an equation for an exogenous variable so $\text{Params}(e_0) = \{v_0\}$; thus e_0 cannot be assigned to some variable other than v_0 . Therefore, $\phi = \phi'$. \square

A structural equation model can be used to represent a joint probability distribution over the variables by allowing each equation e_i to include an arbitrarily distributed random variable γ_i that represents the external, non-modelled factors that may introduce noise into the system. We assume that γ_i is independent from γ_j for all $i \neq j$, in effect tacitly assuming that there are no latent (hidden) variables confounding interactions between the variables in the model. Under this assumption it has been shown [Pearl and Verma, 1991] that if the model is recursive, then the causal graph will obey the Markov condition with respect to the probability distribution parameterized by the model, and can thus be represented by a Bayesian network. Conversely, Druzdzel and Simon [1993] showed that any Bayesian network can be represented as a structural equation model.

We further assume that all systems being modelled will obey the faithfulness assumption [Pearl, 1988, Spirtes *et al.*, 1993]. Under these assumptions the causal graph learned using a constraint-based causal discovery algorithm such as the PC algorithm [Spirtes *et al.*, 1993] will be guaranteed to produce (in the infinite sample limit) the same causal graph (up to an equivalence class) given by the causal model.

2.2 Causal Reasoning

The primary reason that causal models are considered to be valuable in practice is that they allow us to reason about the effect of performing manipulations on a system. For example we can manipulate the rotation of a spinning wheel by grabbing the wheel and forcing it to rotate at a given rate regardless of on what setting the wheel’s motor may be set, or we can perform a randomized study to manipulate the intake level of a particular drug D in a sample of patients. Causal inference, or using the model to predict the effect of such a manipulation, is made possible by a critical postulate which we call the *Manipulation Postulate*²:

Postulate 1 (Manipulation Postulate) *If $G = \langle V, A \rangle$ is a causal graph and $V' \subset V$ is a set of variables being manipulated, then the causal graph, $G' = \langle V, A' \rangle$, for the manipulated system is such that $\text{Pa}(V')_{G'} = \emptyset$.*

In plain words, manipulating a variable will cause its incoming arcs to be removed from the causal graph, but can effect no other change in the graph. In terms of a structural equation model $\langle V, E, \phi \rangle$, manipulating $v \in V$ and applying the Manipulation Postulate amounts to striking $\phi(v)$ from the set E and replacing it with an equation of the form $v = v_0$.

The Manipulation Postulate was probably first used in [Strotz and Wold, 1960] in describing manipulation as replacing equations in structural equation models, and was later given a graphical interpretation by [Spirtes *et al.*, 1993] in the development of the *Manipulation Theorem*, and by Pearl as the *Do* operator [Pearl, 1995] in development of his *Calculus of Interventions*.

Manipulation inferences require only graphs (for qualitative inference), and maybe probability distributions (for quantitative predictions). It is this fact which makes the Manipulation Postulate so important, because without it a causal graph and a probability distribution would not be sufficient to allow causal reasoning. This fact allows us, for example, to learn a causal graph from data and feel confident that such a graph can be used to predict the effects of manipulation, without detailed knowledge of equations underlying the graph.

A subtle point arises when applying the Manipulation Postulate in practice: it is obviously not valid for all possible manipulations. That is, this postulate assumes that the manipulations obey certain modularity conditions, i.e., if you set the value of a variable v to some value v_0 , you do not alter the conditional distributions of any other variables in the model given their parents. Translating this condition into real world actions such as “gripping the wheel with the brake”, etc., is not always a straightforward matter. Furthermore, the Manipulation Postulate does not specify how the structure of a causal model can be affected by “releasing” variables. For example, if I “cease gripping the wheel with the brake”, the Manipulation Postulate does not address how the change in structure will be reflected by this operation. This type of action is addressed by the causal ordering algorithm of [Simon, 1953], and is further explored by [Bogers, 1997,

²[Spirtes *et al.*, 1993] distinguish between a *perfect* and *non-perfect* manipulation. Our definition corresponds to their concept of a perfect manipulation.

Druzdzel and van Leijen, 2001] and [Lu and Druzdzel, 2001], but we do not consider these types of actions in this paper.

3 Violations of the Manipulation Postulate

Rarely in the literature has the validity of the Manipulation Postulate been questioned. Spirtes, *et al.*, [1993] cite some shortcomings of blindly applying the Manipulation Theorem, and there is also a discussion of how widely applicable the theorem is. They also mention the problem of reversibility which we discuss in more detail here. However, for the most part manipulation in modern causal research is defined according to the Manipulation Postulate, and no caveats are considered. In this section we provide examples showing how changing the *resolution* of a model can cause the model to violate the Manipulation Postulate.

Druzdzel [Druzdzel, 1992], and Spirtes *et al.* [Spirtes *et al.*, 1993] have pointed out that some systems appear to exhibit *reversibility* when manipulated. The standard example they use of a reversible system is the transmission of a bicycle. In normal operation, the rotation rate of the pedals is fixed: $P = P_0$ and the wheels rotate in response: $W = \alpha P$. the following causal graph describes this system:

$$\textit{Pedal Rotation Rate} \rightarrow \textit{Wheel Rotation Rate};$$

however, if the bike is propped up on a bike rack and the wheel is directly rotated at some rate: $W = W_0$, then the pedals will rotate in response: $P = W/\alpha$. The causal ordering of the system under these circumstances yields:

$$\textit{Wheel Rotation Rate} \rightarrow \textit{Pedal Rotation Rate}.$$

The bicycle transmission is not an example of a violation of the Manipulation Postulate, however, because this scenario requires the *releasing* of the bicycle pedals before reversibility will be apparent. If, on the contrary, the pedals are not released and the rotation rate of the wheel is set independently, then the bicycle chain will obviously be broken, and the pedals will rotate independently from the wheel, exactly as the Manipulation Postulate predicts. While this example is not a *per se* violation of the Manipulation Postulate, it is an interesting phenomenon because we will show in Section 5 that all models in the class \mathcal{F} display reversibility as a true violation of the Manipulation Postulate.

3.1 The Ideal Gas System: Description

Figure 2 displays one of the most well-understood systems in physics. This system is comprised of an ideal gas trapped in a chamber with a movable piston, on top of which sits a mass, m . The temperature, T , of the gas is controlled externally by a temperature reservoir placed in contact with the chamber. Therefore, m and T can be controlled directly and so will be exogenous variables in our model of this system. When the values of either m or T are altered, the

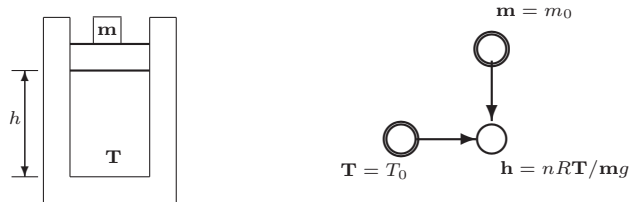


Figure 2: Causal model of the ideal gas in equilibrium.

height of the piston will change: If m is increased then the height will decrease; whereas if T is increased then h will increase. The precise expression of h in terms of T and m is a combination the ideal gas law together with the equilibrium assumption, as given in Figure 2 (g , n , R , m_0 , and T_0 are constants.).

By the “resolution of the model”, we mean the level of abstraction of the model. For example, we could increase the resolution of the ideal gas model in order to explain in more detail how the equation for h in Figure 2 comes about. In Figure 3 we have added two intermediate variables: the total force on the bottom of the piston, F_b , and the pressure of the gas, P (A is a constant). In

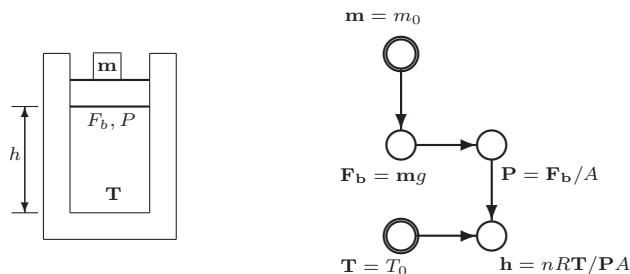


Figure 3: The ideal gas model with increased resolution.

words the causal ordering can be described as follows: “*In equilibrium, the force applied to the bottom of the piston must equal the weight of the mass on top of the piston. Given the force on the bottom of the piston, the pressure of the gas must be determined, which together with the temperature determines the height of the piston through the ideal gas law.*”

The equations presented in Figures 2 and 3 assume that the system is in equilibrium. That is, in a hypothetical experiment where m and T are set to some arbitrary values, there is an implicit time delay in measuring the remaining variables sufficient to allow all time-variation in their values to stabilize.

3.1.1 Manipulating the Height of the Piston

Consider what happens when the height of the piston is set to a constant value: $h = h_0$. Physically this can be achieved by inserting pins into the walls of the chamber at the desired height, as shown in Figure 4a. Applying the Manipulation Postulate to the models in Figures 2 and 3 yields the graphs depicted in Figure 4b and Figure 4d, respectively.

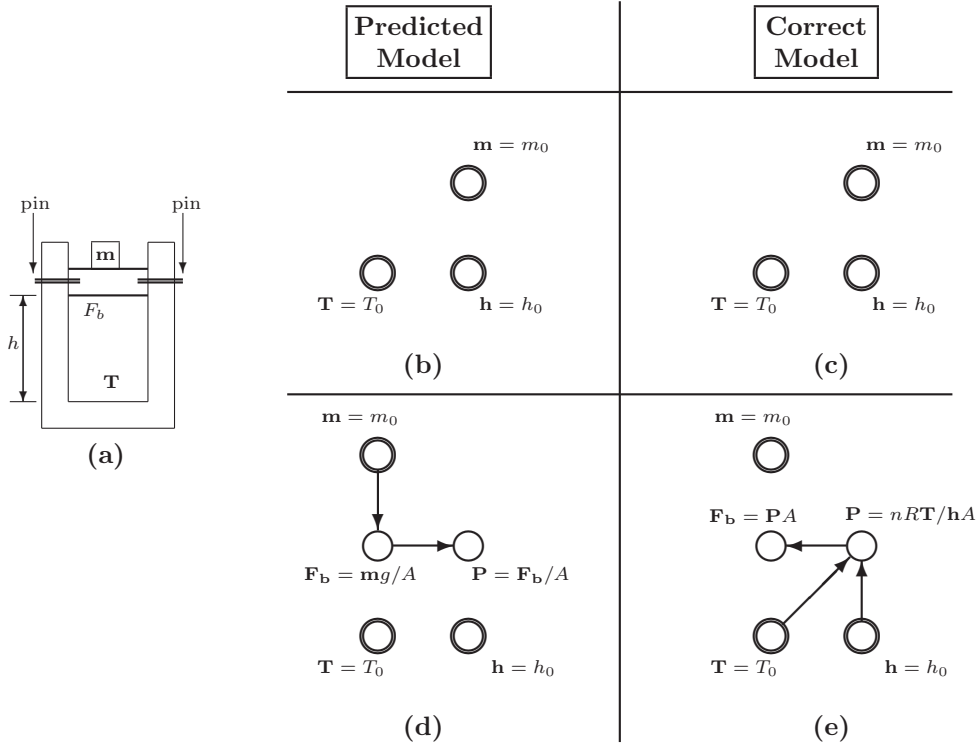


Figure 4: Whether or not the ideal gas model violates the Manipulation Postulate when h is manipulated depends on the resolution of the model.

Now let us consider what the *true* causal graph for these models should look like. For the non-resolved model of Figure 4b, all variables in the model are being set independently; in this case the correct model is given by the Manipulation Postulate. For the resolved model of Figure 4d, things are less straightforward; however, since this is a simple system which we understand well, we are able to write down the governing equations for the manipulated system, given in Figure 4e. Constructing the causal mapping (unique by Lemma 1) for these equations yields the graph shown. In words: *Since h and T are both fixed, P is determined by the ideal gas law, $P = kT/h$. Since the gas is the only source of force on the bottom of the piston, F_b is determined by P : $F_b = PA$. Thus, P is no longer determined by F_b , and F_b is independent of m .* It is clear that the true causal model shown in Figure 4e differs from that predicted by the

Manipulation Postulate shown in Figure 4d.

The fact that changing the resolution of a model can cause it to violate the Manipulation Postulate is a disturbing conclusion. Causal modelers are accustomed to being able to switch back and forth between different levels of abstraction for ease of model construction and explanation. Considered from the standpoint of causal discovery these results are also disheartening. Using data from the equation system of Figure 3 with independent error terms, the causal graph shown there would be learned by a constraint-based discovery algorithm such as the PC algorithm. On the other hand, using data from the equations governing the manipulated system would yield the causal graph in Figure 4e. Both of these facts can readily be verified by calculating the independencies given by the respective equation systems with independent error terms. This fact was also verified empirically using simulation by the authors. The end result is clear: a causal graph learned based on the equilibrium ideal gas system and altered according to the Manipulation Postulate will yield the incorrect causal graph of Figure 4d.

3.1.2 Manipulating the Force on the Bottom of the Piston

There are other, even more dramatic problems with manipulating variables in the expanded-resolution model. Referring back to Figure 3, imagine that for some reason we want to minimize the value of h . It would not be unreasonable, given the graph and the equations in Figure 3, to set h by applying a manipulation to F_b , since F_b is a causal ancestor of h . In particular, in order to make h as small as possible, we would want to make F_b as large as possible according to the equations in Figure 3.

Consider what happens when F_b is manipulated in this way: Again, the Manipulation Postulate predicts that a manipulation will cause the arc from m to F_b to be removed from the model, but otherwise the model will be unchanged. This model is depicted in Figure 5b. In the real system, the force on the bottom of the piston can be set independently of the pressure of the gas by raising a movable stage up through the chamber and directly applying the desired force to the piston with the stage, as shown in Figure 5a. Something very unexpected happens under this manipulation. Unless by coincidence the force applied exactly balances the force due to the mass, the piston will continually be accelerated out of the cylinder, and h , which we intended to minimize, instead grows without bound. Not only does this manipulation violate the postulate, but even worse, we have discovered a *dynamic instability* in the system, i.e., there *is no* equilibrium model; a fact which an equilibrium causal graph alone provides no indication of.

The most disturbing fact about this example is that the instability caused by our manipulation created exactly the opposite effect we were attempting to achieve. Imagine for instance that, instead of the height of a piston, h represented the cancer level in a population of patients. If this example seems exaggerated it is only because we have some concrete understanding about the

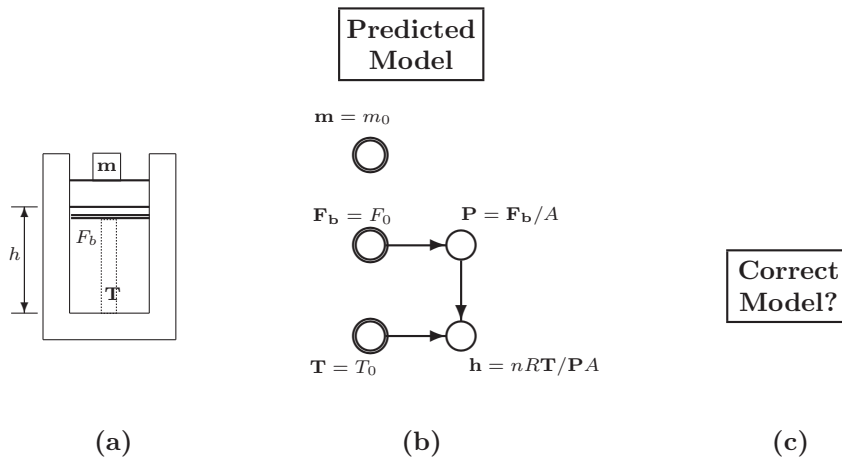


Figure 5: A more severe violation of the Manipulation postulate: no equilibrium model exists after manipulating F_b .

equations underlying the ideal gas. However, imagine applying manipulations to automatically learned models of complex socio-economic or medical systems, where our basic knowledge is typically much less.

4 Dynamic Causal Models

Manipulating the force in the ideal gas model led to an instability. This effect gives us a clue as to what is happening, namely, underlying the equilibrium ideal gas model is a dynamic system. When certain manipulations are made, this dynamic system may not possess an equilibrium point; the result is the hidden instability discovered in the ideal gas system. To understand the phenomenon, we must first discuss how to model this system on a finer time scale and we must show how to relate the fine-time-scale models to the equilibrium models.

The issue of modelling a causal system on varying time scales and relating models on those time scales was addressed by Iwasaki and Simon [Iwasaki and Simon, 1994] in their development of the *Dynamic Causal Ordering* algorithm. The key points that we take from their work are the following:

1. It is possible to model dynamic systems on many different time scales,
2. The causal graphs will not necessarily be the same for different time scales, and
3. The causal models based on shorter time scales can be used to derive models on longer time scales by applying the *equilibration* operator.

Consider again the experiment we performed in Section 3.1. After we dropped a new mass on the piston and/or changed the temperature, we waited some length of time for the piston to come to rest, then measured all of our variables. On the contrary, here we consider the independence and functional relations

between the variables that occur the instant we drop a new mass on the piston or change the temperature of the gas. There is no reason to expect that the independencies found among the variables in this experiment would be the same as in the equilibrium experiment, and in fact, the independencies and the equations governing this dynamic behavior will in general be entirely different. The causal graphs that describe these systems will in general be entirely different as well.

The Dynamic Causal Ordering algorithm considered continuous-time variables that could be represented by differential equations. It used structural equation models to build dynamic causal graphs of these systems, and approximated the continuous time by modelling the system at fixed, discrete time intervals. Graphically, this was accomplished by creating new variables (vertices) for each time slice, and adding integration links to relate variables across time slices. In terms of structural equation models and causal graphs, time-dependent systems are thus no different in principle from equilibrium systems. Finding a causal mapping over these sets of equations would again define a directed acyclic graph (in the recursive case), where some arcs might go across time slices. The Dynamic Causal Ordering algorithm demonstrated that by modelling equilibrium causal systems on shorter time scales it is possible to generate totally different causal models.

We will illustrate the features of this technique by presenting the dynamic causal model of the ideal gas system. There are four physical laws: (1) Weight of a mass: $F_t = mg$, (2) Newton's second law: $\Sigma_i F_i = ma$, (3) the Ideal gas law: $P = nRT/h$, and (4) the Pressure-force relationship: $P = F_b/A$, where a is the acceleration of the piston and all other variables are as defined in Figure 3. In addition to these physical laws, the system is constrained by the definition of acceleration and velocity of the piston (expressed in discrete form):

$$a_{(t)} \approx \frac{v_{(t+1)} - v_{(t)}}{\Delta t} \quad \text{and} \quad v_{(t)} \approx \frac{h_{(t+1)} - h_{(t)}}{\Delta t},$$

where we have used the notation that $x_{(t)}$ refers to the value of variable x at time slice t , and Δt is the (constant) time between slices. These can be rewritten as the recurrence relations:

$$v_{(t)} = v_{(t-1)} + a_{(t-1)}\Delta t$$

$$h_{(t)} = h_{(t-1)} + v_{(t-1)}\Delta t$$

In order to specify a particular solution to these difference equations, initial conditions must be given for h : $h_{(0)} = h_0$ and for v : $v_{(0)} = v_0$, where h_0 and v_0 are constants. Finally, since m and T are exogenous, we have $m_{(t)} = m_0$ and $T_{(t)} = T_0$, for all t .

The recursive graph for this set of equations is shown in Figure 6a. This graph relates all the variables in our model at $t = 0$ with each other and with v and h at $t = 1$. Since $h_{(1)}$ and $v_{(1)}$ are now determined at $t = 1$, we can recursively iterate this procedure to generate causal graphs for arbitrary number of time steps.

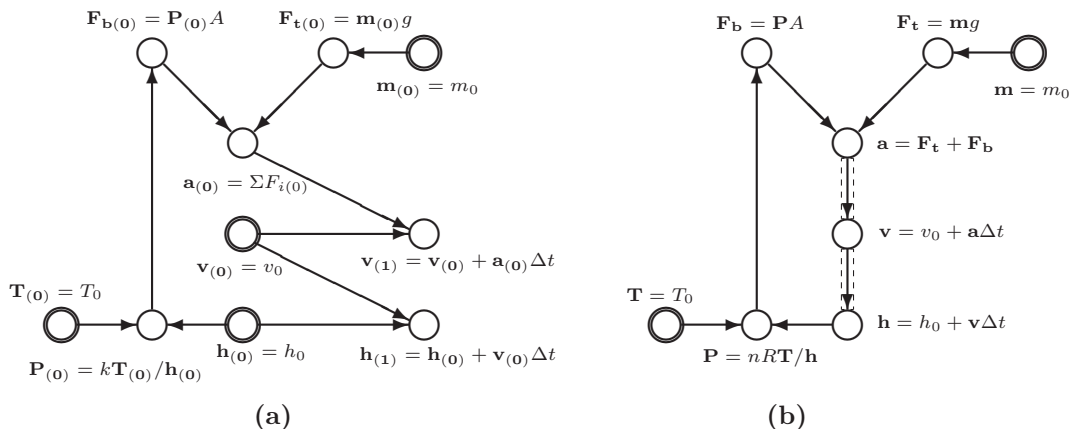


Figure 6: (a) A partially “unrolled” dynamic causal model, and (b) the same graph in shorthand form.

Since this graph is based on continuous differential equations, all causation across time slices will only occur to a variable from its derivative. Thus, this graph is Markovian through time i.e., the variables in the future are d-separated from variables in the past by variables in the present, and it can be represented by the shorthand graph for an infinite sequence of time steps shown in Figure 6b. In this shorthand graph temporal subscripts have been dropped and special dashed links, labelled *integration links* [Iwasaki and Simon, 1994], have been created to denote that a causal relationship is really occurring through a time slice. We will use this shorthand throughout the remainder of the paper. Together with a quantification of the system, a recursive dynamic graph defines a dynamic Bayesian network, and thus is a graphical representation of the temporal probability distribution for this system.

4.1 Deriving Equilibrium Models from Dynamic Models

The dynamic graph in Figure 6 represents the causal graph for the system modeled over an infinitesimal time scale; whereas, the graph from Figure 3 is modeled over a time scale that is long enough for the system to come to equilibrium. Here we formally define dynamic models and we review how to use the equilibration operator to derive an equilibrium model from the dynamic model. We will use the notation that $\dot{v} \equiv dv/dt$ and that $v^{(0)} \equiv v$ and $v^{(i+1)} \equiv dv^{(i)}/dt$.

The shorthand dynamic graph presented in Figure 6b adds some confusion to the concept of recursivity, since it possesses cycles itself although it really is meant to represent the acyclic graph in Figure 6a. Thus to clear up confusion we generalize the concept of recursivity for a shorthand graph:

Definition 6 (recursive causal model) A dynamic causal model $M = \langle V, E, \phi \rangle$ with a causal graph G is recursive if and only if the causal graph $G^{(0)}$, obtained by removing all integration links from G , is acyclic.

Definition 7 (dynamic variable) Given a causal model $M = \langle V, E, \phi \rangle$ with graph G , a variable $v \in V$ is a dynamic variable if and only if $v \in \text{Pa}(v)_G$.

The operation of *equilibration* was presented in Iwasaki and Simonv[Iwasaki and Simon, 1994] whereby the derivatives of a dynamic variable x are eliminated from a model by assuming that x has achieved equilibrium:

Definition 8 ($V_{\text{del}}(\mathbf{x})$, $E_{\text{del}}(\mathbf{x})$) Let $M = \langle V, E, \phi \rangle$ be a causal model with $x \in V$ and with $x^{(n)} \in V$ the highest order derivative of x in the model, then:

$$V_{\text{del}}(x) = \{x^{(i)} \mid 0 < i \leq n, i \neq 0\} \text{ and}$$

$$E_{\text{del}}(x) = \{\phi(x^{(i)}) \mid 0 \leq i < n\}$$

Note that $x \notin V_{\text{del}}(x)$ and $\phi(x^{(n)}) \notin E_{\text{del}}(x)$.

Definition 9 (equilibration) Let $M = \langle V, E, \phi \rangle$ be a causal model and let $x \in V$ be a variable with $x^{(n)} \in V$ the highest order derivative of x in V . The model $M_{\bar{x}} = \langle V_{\bar{x}}, E_{\bar{x}}, \phi_{\bar{x}} \rangle$ due to the equilibration of x is obtained by the following procedure:

1. Let $V_{\bar{x}} = V \setminus V_{\text{del}}(x)$,
2. Let $E_{\bar{x}} = E \setminus E_{\text{del}}(x)$,
3. For each $e \in E_{\bar{x}}$ set $v = 0$ for all $v \in V_{\text{del}}(x)$.
4. Construct a new total causal mapping $\phi_{\bar{x}} : V_{\bar{x}} \rightarrow E_{\bar{x}}$.

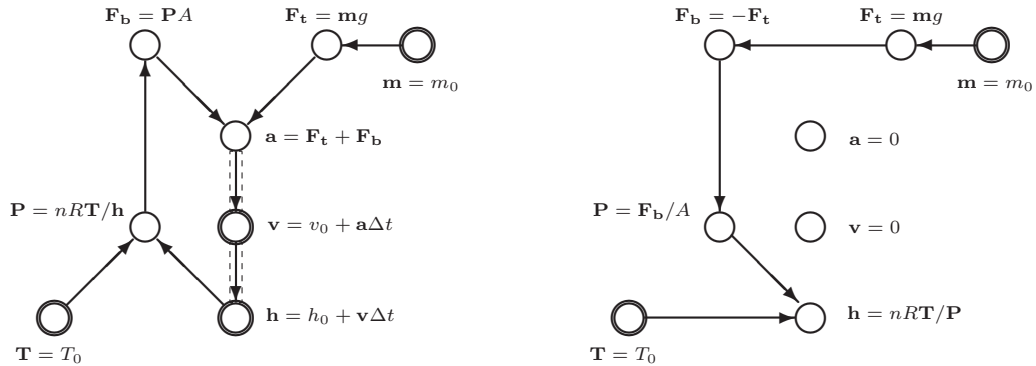


Figure 7: Applying the Equilibration operator.

Equilibration is equivalent to assuming that a dynamic variable x has achieved equilibrium. This implies that all of x 's derivatives will be zero. Applying this operator to the dynamic ideal gas model is shown in Figure 7. It is apparent that the mapping from equations to variables may be very different after equilibration. In this case, the dynamic ideal gas model of Figure 6 reduces to the equilibrium ideal gas model of Figure 3 after the equilibration operator is applied.

Equilibration can cause the remaining set of equations to be non-self-contained. For example, setting variables to zero could cause two equations that were initially independent to become dependent, or it could cause the system to be over-constrained if some variable drops out of an equation. We call equilibration *well-defined* if these do not happen:

Definition 10 (well-defined equilibration) *If $M = \langle V, E, \phi \rangle$ is a causal model and $V_{\bar{x}}$ and $E_{\bar{x}}$ are the respective variables and equations that result when variable $x \in V$ is equilibrated, then we say that equilibrating x is well-defined if and only if $E_{\bar{x}}$ is self-contained with respect to $V_{\bar{x}}$.*

Using the Dynamic Causal Ordering algorithm and the assumption that arcs through time can only occur to a variable from its derivative, we can now define precisely the concepts of *equilibrium model* and *dynamic model*:

Definition 11 (equilibrium model) *A causal model $M = \langle V, E, \phi \rangle$ is an equilibrium model with respect to x for some $x \in V$ if x is not a dynamic variable in M .*

A *dynamic* model is any model which is not an equilibrium model.

4.2 Manipulating Dynamic Models

We now examine the phenomena observed in the ideal gas system from the viewpoint of dynamics. Let us again fix the height of the piston, using the model of Figure 6 to describe the ideal gas system. To fix the piston, we must set h to some constant value for all time, $h_{(t)} = h_0$. We also must stop the piston from moving so we must set $v_{(t)} = 0$ and $a_{(t)} = 0$. Thus, in the dynamic graph with integration links, we can think of this one action of setting the height of the piston as three separate actions. If we assume that the Manipulation Postulate holds on the dynamic model in Figure 8a, we obtain the graph depicted in Figure 8b. Since h is being held constant, this graph is already an equilibrium graph with respect to h (i.e., no equilibration operation is required). By comparing Figure 8b to the manipulated equilibrium ideal gas system of Figure 4c, we can see that aside from the extra variables that were added to the dynamic model for clarity (F_t , a and v), Figure 8b is identical to the expected manipulated model. Therefore, *the Manipulation Postulate holds for this model, and it produces precisely the graph that we originally expected to get but were unable to get from the equilibrium model.*

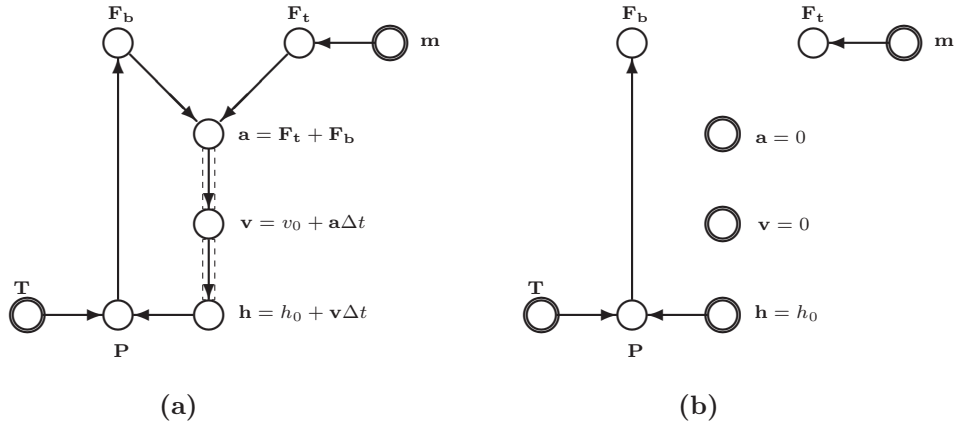


Figure 8: Manipulating h using the dynamic model obeys the Manipulation Postulate.

4.3 Detecting Instabilities

Dynamic models can also be used to predict when a manipulation will cause an instability. In order to demonstrate this, we first need to review a key result about stability in dynamic systems. If, within a dynamic model, a dynamic variable x possesses a fixed-point solution at, say, $x = x_0$, that fixed-point will be a stable fixed-point if and only if the following stability relation holds:³

$$\left. \frac{\partial \dot{x}}{\partial x} \right|_{x_0} < 0, \quad (\text{Stability condition})$$

where \dot{x} is the time-derivative of x . To see this result intuitively, imagine a ball resting in a valley: if the ball is displaced slightly in the positive direction, a velocity (and acceleration) will be created which tends to push it back towards the fixed-point; whereas the displacement of a ball on the crest of a hill will cause a velocity which tends to push the ball away from the fixed-point. That is, in order for a fixed-point to be stable, when we displace the ball in the positive x -direction, a negative velocity must develop in the ball to push it back toward the fixed-point.

According to this stability condition, in the differential equation for x , the variable \dot{x} must somehow be a function of x for stability to occur. In terms of the structure of dynamic causal models, this implies that in order for stability to occur, there must exist some regulation process by which $\dot{x}(t)$ can get information about $x(t')$ for some $t' \leq t$. In our dynamic model, for example, this regulation takes place through the feedback loop: $h_{(t)} \rightarrow P \rightarrow F_b \rightarrow a_{(t)} \rightarrow v_{(t+1)}$. The stability condition thus suggests a structural condition for stability in a causal graph:

Definition 12 (The Structural Stability Principle) *Let G be a causal graph with dynamic variable v , and let $\text{Fb}(v)$ denote the set $\text{Fb}(v) = \text{Anc}(v)_G \cap \text{Des}(v)_G$, then v will possess a stable fixed-point only if $\text{Fb}(v) \neq \emptyset$.*

³See [Strogatz, 1991] for example.

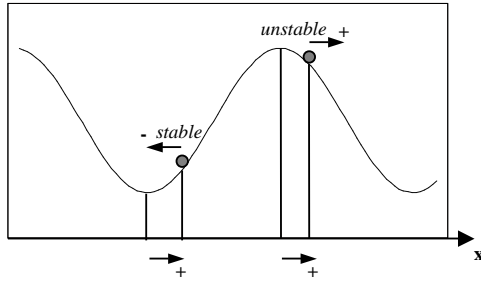


Figure 9: At a stable fixed-point, displacing the ball in the + direction causes a negative velocity to push it back. At an unstable fixed-point, the arising velocity is positive, pushing the ball away.

Consider the implications of manipulating F_b in the dynamic model of the ideal gas system. If we again assume that the Manipulation postulate holds for the dynamic model, when F_b is manipulated in Figure 10a, the model shown in Figure 10b is obtained. We can see immediately from the causal graph that

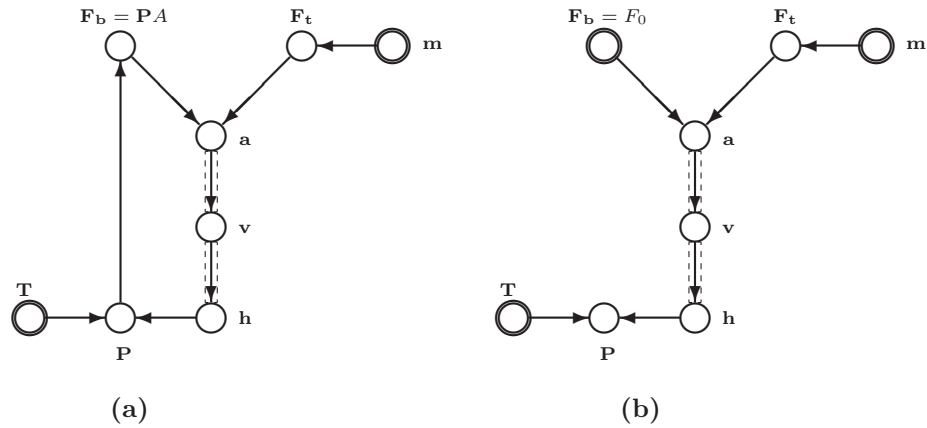


Figure 10: The Structural Stability Principle: manipulating F_b breaks the only feedback loop, causing an instability.

this manipulation will break the only feedback loop for x in this system, and thus according to the Structural Stability principle, there does not exist an equilibrium point for this model. Our second major observation is therefore that *the dynamic model, together with the Manipulation Postulate and the Structural Stability condition correctly predict that some manipulations will cause an instability.*

5 Theorems

In this section we formalize the observations suggested by the examples in Sections 3 and 4. After all, despite the neatness with which the dynamic models appear to explain the phenomena, we have nonetheless only presented physical examples to illustrate our points. Thus we are vulnerable to the criticism that our physical systems might be modelled incorrectly, or that hidden variables may be present that we are not aware of. In this section we prove that the peculiarities which we have presented are a mathematical property of certain abstract dynamical systems, rather than an artifact due to incorrectly modelling the systems in our example.

For the remainder of this section, let $M = \langle V, E, \phi \rangle$ be an arbitrary causal model, let $x \in V$ be a variable being modelled and let $M_{\bar{x}} = \langle V_{\bar{x}}, E_{\bar{x}}, \phi_{\bar{x}} \rangle$ be an equilibrium model with respect to $x \in V_{\bar{x}}$. Let G and $G_{\bar{x}}$ be the causal graphs for M and $M_{\bar{x}}$, respectively and $G_x^{(0)}$ be the graph corresponding to G with all integration links for x and its derivatives removed. We define $\text{Fb}(x)$ to be the set of feedback variables: $\text{Fb}(x) = \{\text{Anc}(x)_G \cap \text{Des}(x)_G\}$, and let $V_{del}(x)$ and $E_{del}(x)$ be defined as in Definition 8.

Definition 13 (RFRE Model, $\mathcal{F}(M, x)$) $M_{\bar{x}}$ is a recursive feedback-resolved equilibrated (RFRE) model with respect to M and x if and only if the following conditions hold:

1. **Equilibration:** $M_{\bar{x}}$ is derived from M by performing a well-defined equilibration of x in M ,
2. **Recursivity:** $M_{\bar{x}}$ and M are both recursive, and
3. **Feedback-resolution:** $\{\text{Fb}(x) \cap V_{\bar{x}}\} \neq \emptyset$.

We denote the class of all RFRE models as \mathcal{F} , and use $\mathcal{F}(M, x)$ to denote the set of RFRE models with respect to M and x .

Lemma 2 *If M is recursive, then there exists an ordering relation O on the associations of ϕ such that:*

1. $O(\langle v_i, e_i \rangle) < O(\langle v_j, e_j \rangle)$ if $v_i \in \text{Anc}(v_j)_{G_x^{(0)}}$, and
2. the pairs corresponding to $\text{Fb}(x)$ form a contiguous sequence in O .

Proof: In $G_x^{(0)}$, all $x^{(i)}$ such that $i \neq n$ are exogenous by construction (they are specified by the initial conditions in the model). Thus they can be ordered before all other $v \in \text{Fb}(x)$. Define $\text{Anc}(\text{Fb}(x))$ and $\text{Des}(\text{Fb}(x))$ to be the set of ancestors and descendants, respectively of $\text{Fb}(x)$: $\text{Anc}(\text{Fb}(x))_{G_x^{(0)}} \equiv \bigcup_{v \in \text{Fb}(x)} \text{Anc}(v)_{G_x^{(0)}}$ and $\text{Des}(\text{Fb}(x))_{G_x^{(0)}} \equiv \bigcup_{v \in \text{Fb}(x)} \text{Des}(v)_{G_x^{(0)}}$. By transitivity of the ancestor and descendant relationships, if there exists a $v \in \text{Anc}(\text{Fb}(x)) \cap \text{Des}(\text{Fb}(x))$ then $v \in \text{Fb}(x)$. Thus an ordering can be defined such that $O(v_{anc}) <$

$O(v_{fb}) < O(v_{des})$ for arbitrary variables $v_{anc} \in \text{Anc}(\text{Fb}(x)) \setminus \text{Fb}(x)$, $v_{des} \in \text{Des}(\text{Fb}(x)) \setminus \text{Fb}(x)$, and $v_{fb} \in \text{Fb}(x)$. \square

Lemma 3 *Let \bar{F} denote the set $V_{\bar{x}} \setminus \{\text{Fb}(x) \cup \{x\}\}$. If $M_{\bar{x}}$ is the model obtained by a well-defined equilibration of x in M and $M_{\bar{x}}$ are recursive then $\phi_{\bar{x}}(v) = \phi(v)$ for all $v \in \bar{F}$.*

Proof: Define an ordering O for ϕ and label the pairs $\langle x_i, e_i \rangle$ in ϕ as in the proof of Lemma 1. By Lemma 2, O can be defined such that all associations for variables in $\text{Fb}(x)$ form a contiguous sequence in O with $O(x) < O(v)$ for all $v \in \text{Fb}(x)$. Define $o_x = O(x)$ and $o_{fb} = O(v_i)$ where $O(v_i) \geq O(v_j)$ for all $v_i, v_j \in \text{Fb}(x)$.

Partition the variables and equations into three sets: $V = \{V_{pre}, V_{fb}, V_{post}\}$ and $E = \{E_{pre}, E_{fb}, E_{post}\}$, where

$$\begin{aligned} V_{fb} &\equiv \{v \mid v \in \text{Fb}(x) \cup \{x\}\}, \\ V_{pre} &\equiv \{v \mid v \in V \text{ and } O(v) < o_x\}, \\ V_{post} &\equiv \{v \mid v \in V \text{ and } O(v) > o_{fb}\}, \end{aligned}$$

and E_{fb} , E_{pre} , E_{post} are the sets of equations associated with V_{fb} , V_{pre} , and V_{post} , respectively, in ϕ .

First we show that all $e \in E_{pre}$ get assigned to some $v \in V_{pre}$ and all $e \in E_{post}$ get assigned to some $v \in V_{post}$, then the result follows by the fact that within the post and pre sets ϕ already provides a causal mapping because none of the equations or variables in E_{pre} have changed and no dependency on the variables in V_{post} have changed in the equations in E_{post} by the assumption of a well-defined equilibration. Since $\phi_{\bar{x}}$ is unique it must possess the same mapping between these sets.

By an argument identical to that of Lemma 1 it can be proven that the first pair $\langle v_j, e_i \rangle \in \phi_{\bar{x}}$ such that $j \neq i$ will occur when $j = o_x$; thus all $v \in V_{pre}$ get mapped only to equations in E_{pre} . Finally, no $e_{fb} \in E_{fb}$ can be assigned to some $v \in V_{post}$ because by construction no equation of order i can be assigned to a variable of order $j > i$ since $O(v_e) < O(e)$ for all $v_e \in \text{Params}(e)$. Therefore, no $e_{post} \in E_{post}$ can be assigned to a $v \in \text{Fb}(x)$ because there would be no equation to replace it in E_{fb} . \square

The next lemma shows that all ancestors of x that are not dynamic variables in $G_x^{(0)}$ must pass through $x^{(n)}$:

Lemma 4 *The following relation holds: $\text{Anc}(x) \setminus V_{del}(x) = \text{Anc}(x^{(n)})_{G_x^{(0)}}$.*

Proof: First note that if v is a dynamic variable, then in $G_x^{(0)}$, by construction v must be given by initial conditions and so must be exogenous. Therefore, in the chain of derivatives:

$$x^{(n)} \rightarrow x^{(n-1)} \rightarrow \dots \rightarrow x$$

all $x^{(i)}$ such that $i \neq n$ must have a single parent which is connected by an integration link. Therefore, all $v \in \text{Anc}(x)_G \setminus V_{del}(x)$ must be ancestors of $x^{(n)}$. \square

Lemma 5 *If $M_{\bar{x}} \in \mathcal{F}(M, x)$ then there does not exist an $x^{(i)}$ such that $x^{(i)} \in \text{Ch}(x)_G$.*

Proof: First note that the result follows for all $x^{(j)}$ such that $j < n$, because by construction $\text{Pa}(x^{(j)}) = \{x^{(j+1)}\}$ in M . Thus we only need to prove that $x^{(n)} \notin \text{Ch}(x)$. We prove this result by contradiction. Define $\{V_{pre}, V_{fb}, V_{post}\}$ and $\{E_{pre}, E_{fb}, E_{post}\}$ as in the proof of Lemma 3. $M_{\bar{x}}$ is recursive by assumption; therefore there only exists one causal mapping, $\phi_{\bar{x}}$, by Lemma 1. We show that if $x \in \text{Pa}(x^{(n)})_G$ then $G_{\bar{x}}$ will not be acyclic, which violates the assumption that $M_{\bar{x}} \in \mathcal{F}(M, x)$.

According to Lemma 3, $\phi_{\bar{x}}(v) = \phi(v)$ for all $v \in V_{pre} \cup V_{post}$. We therefore only need to construct a mapping over V_{fb} . Assume $x \in \text{Pa}(x^{(n)})_G$ and let $\langle x^{(n)}, e^{(n)} \rangle, \langle x, e_x \rangle \in \phi$. Then when $x^{(n)}$ is removed from the set of variables during equilibration, $e^{(n)}$ will be of the form $f(x, V'_{fb}, P) = 0$, where $V'_{fb} \subset V_{fb}$ and P is a set of variables that are not in $\text{Fb}(x)$. During equilibration e_x will be removed from the model, thus x can be assigned to $e^{(n)}$. Furthermore, all remaining variables in $\text{Fb}(x)$ can be associated with their original equations in ϕ . Let $\phi' : V_{\bar{x}} \rightarrow E_{\bar{x}}$ be the causal mapping defined such that $\phi'(v) = \phi(v)$ for all $v \neq x$ and $\phi'(x) = e^{(n)}$. This forms a valid causal mapping and therefore by uniqueness $\phi' = \phi_{\bar{x}}$. By Lemma 4, since $\text{Fb}(x) \setminus V_{del}(x) \subset \text{Anc}(x^{(n)})_{G_x^{(0)}}$ it must be the case that $\text{Fb}(x) \setminus V_{del}(x) \subset \text{Anc}(x)_{G_{\bar{x}}}$, because the equation that was assigned to $x^{(n)}$ is now assigned to x and all the remaining associations are unchanged. However, since $M_{\bar{x}} \in \mathcal{F}(M, x)$ it must be the case that $\{\text{Fb}(x) \setminus V_{del}(x)\} \cap \text{Ch}(x) \neq \emptyset$, therefore, there exists an $f \in \text{Fb}(x) \setminus V_{del}(x)$ such that $x \in \text{Pa}(f)$. Thus $G_{\bar{x}}$ is cyclic, which is a contradiction. \square

Lemma 6 *If $M_{\bar{x}} \in \mathcal{F}(M, x)$, then there exists a $v \in V_{\bar{x}}$ such that $v \in \text{Pa}(x)_{G_{\bar{x}}}$ and such that $v \in \text{Ch}(x)_G$.*

Proof: Define an ordering O for ϕ and label the pairs $\langle v_l, e_l \rangle$ in ϕ according to O as in the proof of Lemmas 1 and 3. Let $\langle x, e_i \rangle$ be the association for x in $\phi_{\bar{x}}$. By construction $x \neq v_i$, and by Lemma 3, $v_i \in \text{Fb}(x)$. Since $x \in \text{Params}(e_i)$ and since $\langle v_i, e_i \rangle \in \phi$ it must be the case that $v_i \in \text{Ch}(x)_G$. Since $x^{(l)}$ is exogenous in $G_x^{(0)}$ for all $l \neq n$ and since, by Lemma 5, $v_i \neq v^{(n)}$, it follows that $v_i \notin V_{del}(x)$. Therefore $v_i \in \text{Fb}(x) \setminus V_{del}(x)$, and since $v_i \in \text{Params}(e_i)$ it must be the case that $v_i \in \text{Pa}(x)_{G_{\bar{x}}}$. \square

Lemma 7 *If $M_{\bar{x}} \in \mathcal{F}(M, x)$ and $M_{\hat{x}} = \langle V_{\hat{x}}, E_{\hat{x}}, \phi_{\hat{x}} \rangle$, with causal graph $G_{\hat{x}}$, is the causal model resulting when x is manipulated in M , then in $G_{\hat{x}}$ there will exist an edge $x \rightarrow v$ for all $v \in \text{Ch}(x)_G \cap V_{\hat{x}}$.*

Proof: Since M obeys the Manipulation Postulate, the only arcs that will be removed from M when x is manipulated will be the arcs coming into x and into x 's derivatives $x^{(i)}$. Since by Lemma 5, x is not a parent of any $x^{(i)}$ the children of x must be preserved in $G_{\hat{x}}$. \square

Finally, Theorem 1 presents conditions which are sufficient for $M_{\bar{x}}$ to violate the Manipulation Postulate.

Theorem 1 (reversibility) *If $M_{\hat{x}} \in \mathcal{F}(M, x)$ and the Manipulation Postulate holds for M , then the Manipulation Postulate does not hold for $M_{\hat{x}}$.*

Proof: Manipulating x in M produces an equilibrium model with respect to x , $M_{\hat{x}}$, which must be the correct model that is obtained when x is manipulated, by definition of the Manipulation Postulate. Let $G_{\hat{x}}$ be the causal graph corresponding to $M_{\hat{x}}$. Since $M_{\hat{x}} \in \mathcal{F}(M, x)$, by Lemma 6 there exists a $v \in \text{Ch}(x)_G$ such that $v \rightarrow x$ in $G_{\hat{x}}$; however, according to Lemma 7, the edge $x \rightarrow v$ must exist in $G_{\hat{x}}$. Thus, manipulating x in $G_{\hat{x}}$ by applying the Manipulation Postulate leads to an incorrect graph $G_{\hat{x}, \hat{x}}$, because it will not contain an edge between v and x . \square

The theorem is labeled “reversibility” because its proof relies on the guaranteed reversal of an arc; nonetheless, it is clear by the examples given in Section 3.1 that there is more complex behavior being exhibited in these systems than mere reversibility.

The last theorem proves that hidden dynamic instabilities are a mathematical feature of some equilibrium causal models:

Theorem 2 (instability) *If $M_{\hat{x}} \in \mathcal{F}(M, x)$, the Manipulation postulate holds for M and the Structural Stability condition holds then there exists a set of variables $V' \subset V_{\hat{x}}$ such that if V' is manipulated in M , the variable x will become unstable.*

Proof: Define $V' \equiv \text{Fb}(x) \setminus V_{\text{del}}(x)$. It must be the case that $V' \neq \emptyset$ by definition of $\mathcal{F}(M, x)$. According to the Manipulation Postulate, manipulating V' in G will create a new graph $G_{V'}$ with $\text{Fb}(x)_{G_{V'}} = \emptyset$. Therefore, according to the Structural Stability principle, x will be unstable in $G_{V'}$. \square

6 Discussion

To summarize, we have asserted that the Manipulation Postulate is critical to all the existing formalisms aimed at performing causal inference. Furthermore, we have shown that there exist simple physical systems that can be modelled as recursive equilibrium causal systems but cannot be correctly used for causal inference because they do not obey this postulate. We have related these problems to the existing problem of “reversibility” and have shown that this characterization of the problem is overly simplistic, because these systems can violate the Manipulation Postulate in much more complex ways than merely switching arcs of manipulated variables. One particularly severe violation is the phenomenon of “hidden instabilities” that can occur when an appropriate manipulation is made on some equilibrium models. We have related these problems to some research on constructing dynamic causal models which asserts that the same system can possess many different causal graphs depending on the time scale over which it is modelled. Furthermore, we have shown that the problem systems modelled dynamically, over an infinitesimal time scale, do in fact obey

the Manipulation Postulate and allow us to predict when a manipulation will trigger an instability.

6.1 The Size of the RFRE Class

One important issue we have not addressed is that of the generality and thus the significance of Theorems 1 and 2. The feedback resolution condition of Definition 13 stipulates that feedback variables that are children of $x \in G$ be present in the equilibrium model. Whether or not this condition is satisfied has little to do with the system itself and more to do with the modeler. The condition that the models be recursive is also very weak, since most models used in practice are recursive. For example, any dynamic model such that set of variables $\text{Fb}(x) \cap V_{\bar{x}}$ forms a linear chain from x to $x^{(n)}$:

$$x \rightarrow x_{fb}^1 \rightarrow x_{fb}^2 \rightarrow \dots \rightarrow x^{(n)}$$

will produce a recursive model when x is equilibrated.

To give a feel for the scope of these conditions, Table 1 provides a sample of physical systems which can possess RFRE models. This table is a virtual laun-

System	\mathbf{x}	$\{\text{Fb}(\mathbf{x}) \setminus \mathbf{V}_{\text{del}}(\mathbf{x})\}$
Ideal gas	height of piston	pressure
Body in viscous medium	velocity	damping force
Simple harm. oscillator	position of mass	spring force
RC circuit	charge on capacitor	current
Op-amp w/neg feedback	charge on neg term	output voltage

Table 1: Examples of physical systems that can be modeled as RFRE models.

dry list of some of the simplest physical systems known. The simple harmonic oscillator system is especially general since many other systems are mathematically isomorphic to it.

Exacerbating the ubiquity of this class is the fact that membership in $\mathcal{F}(M, x)$ is only sufficient, not necessary for violation of the Manipulation Postulate. There exist feedback-resolved non-recursive models, for example, that also violate the Manipulation Postulate. Consider the dynamic model shown in Figure 11. When this model is equilibrated there are two possible equilibrium causal structures, both of which include an arc from $f_1 \rightarrow x$ even though in the dynamic model x was a parent of f_1 . Thus, any equilibrated model for this system that includes f_1 will display reversibility when x is manipulated.

6.2 Future Directions

We have tried to emphasize the severity of our conclusions on the practice of causal discovery from equilibrium data. Because the examples we have presented are based on simple systems about which most readers are likely to have a good general understanding, the consequences of violating the Manipulation

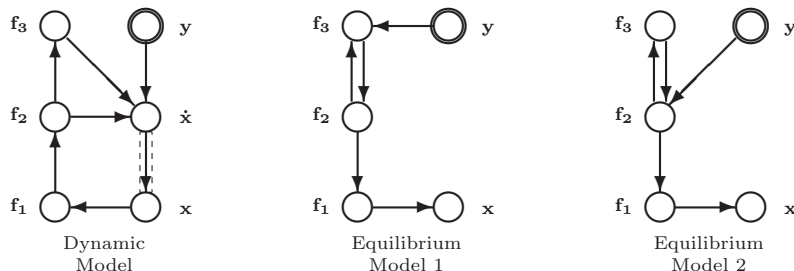


Figure 11: Some non-recursive feedback-resolved equilibrated models will also exhibit reversibility.

Postulate may not be fully appreciated. However, in domains where causal discovery procedures are used to elicit causal graphs from data, typically little or no background knowledge is present. After discovery, therefore, all knowledge that the modeler possesses is in the form of a causal graph and maybe a sample probability distribution. The theorems presented in this paper shed significant doubt on the usefulness of a graph so obtained for performing causal reasoning, because we would have no knowledge of the dynamics underlying this system. One obvious remedy is to use time-series data to learn dynamic causal graphs instead of equilibrium models when causal inferences are required.

This paper raises many new open research questions. Obviously this work raises the importance of being able to learn dynamic causal graphs from time-series data. However, an alternative to focusing efforts on constructing dynamic models would be instead to attempt to more precisely characterize the relationship between a dynamic model and its equilibrium counterpart. In this way it may be possible to extract general rules for when an equilibrium model will be well-behaved under manipulation. Among the questions to be answered are: What are necessary/sufficient conditions for a non-dynamic model to support causal inference? Under what circumstances will a manipulation cause an instability? Constructing a dynamic model instead of a non-dynamic model requires the modeler to know more about the system being modeled; for example one might need detailed time-series data. However, maybe it is possible to insure that a model will support manipulation with less than full time-series data, for example, temporal ordering information may help decide when an equilibrium model is feedback-resolved. What then is the minimal information needed to insure that a model will support manipulation? And most generally, are there general relationships between dynamic models and equilibrium models that can allow us to answer these questions for arbitrary models? We believe these are hard questions but whose answers would be of significance to future work in causal reasoning.

References

- [Bogers, 1997] Jeroen J.J. Bogers. Supporting the change in structure in a decision support system based on structural equations. Master's thesis, Department of Technical Mathematics and Informatics, Delft University of Technology, Delft, The Netherlands, August 1997.
- [Dash and Druzdzel, 2000] Denver Dash and Marek J. Druzdzel. A note on the correctness of the Causal Ordering Algorithm. *Technical Report. Intelligent Systems Program, University of Pittsburgh*, 2000.
- [Druzdzel and Simon, 1993] Marek J. Druzdzel and Herbert A. Simon. Causality in Bayesian belief networks. In *Proceedings of the Ninth Annual Conference on Uncertainty in Artificial Intelligence (UAI-93)*, pages 3–11, San Francisco, CA, 1993. Morgan Kaufmann Publishers, Inc.
- [Druzdzel and van Leijen, 2001] Marek J. Druzdzel and Hans van Leijen. Causal reversibility in Bayesian networks. *Journal of Experimental and Theoretical Artificial Intelligence*, 13(1):45–62, Jan 2001.
- [Druzdzel, 1992] Marek J. Druzdzel. *Probabilistic Reasoning in Decision Support Systems: From Computation to Common Sense*. PhD thesis, Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, December 1992.
- [Fisher, 1970] Franklin M. Fisher. A correspondence principle for simultaneous equation models. *Econometrica*, 38(1):73–92, January 1970.
- [Galles and Pearl, 1997] D. Galles and J. Pearl. Axioms of causal relevance. *Artificial Intelligence*, 97(1-2):9–43, 1997.
- [Haavelmo, 1943] Trygve Haavelmo. The statistical implications of a system of simultaneous equations. *Econometrica*, 11(1):1–12, January 1943.
- [Halpern, 2000] Joseph Y. Halpern. Axiomatizing causal reasoning. *Journal of Artificial Intelligence Research*, 12:317–337, 2000.
- [Hood and Koopmans, 1953] William C. Hood and Tjalling C. Koopmans, editors. *Studies in Econometric Method. Cowles Commission for Research in Economics. Monograph No. 14*. John Wiley & Sons, Inc., New York, NY, 1953.
- [Iwasaki and Simon, 1994] Yumi Iwasaki and Herbert A. Simon. Causality and model abstraction. *Artificial Intelligence*, 67(1):143–194, May 1994.
- [Koopmans, 1953] Tjalling C. Koopmans. Identification problems in economic model construction. In Hood and Koopmans [1953], chapter II, pages 27–48.

- [Kuipers, 1987] Benjamin Kuipers. Abstraction by time-scale in qualitative simulation. In *Proceedings of the National Conference on Artificial Intelligence, AAAI-87*, pages 621–625, Seattle, WA, July 1987. American Association for Artificial Intelligence, Morgan Kaufmann Publishers, Inc., San Mateo, CA.
- [Lu and Druzdzal, 2001] Tsai-Ching Lu and Marek J. Druzdzal. Supporting changes in structure in causal model construction. In Salem Benferhat and Philippe Besnard, editors, *Proceeding of the Sixth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (EC-SQARU 2001)*, volume LNAI 2143 of *Lecture Notes in Artificial Intelligence*, pages 204–215, Toulouse, France, 2001. Springer-Verlag Berlin Heidelberg.
- [Pearl and Verma, 1991] Judea Pearl and Thomas S. Verma. A theory of inferred causation. In J.A. Allen, R. Fikes, and E. Sandewall, editors, *KR-91, Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, pages 441–452, Cambridge, MA, 1991. Morgan Kaufmann Publishers, Inc., San Mateo, CA.
- [Pearl, 1988] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, Inc., San Mateo, CA, 1988.
- [Pearl, 1995] Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–710, 1995.
- [Pearl, 2000] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, UK, 2000.
- [Richardson, 1996] Thomas Richardson. *Models of Feedback: Interpretation and Discovery*. PhD dissertation, Carnegie Mellon University, Department of Philosophy, 1996.
- [Simon, 1953] Herbert A. Simon. Causal ordering and identifiability. In Hood and Koopmans [1953], chapter III, pages 49–74.
- [Spirtes *et al.*, 1993] Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. Springer Verlag, New York, 1993.
- [Strogatz, 1991] Steven H. Strogatz. *Nonlinear Dynamics and Chaos with Applications to Physics, Biology, Chemistry, and Engineering*. Addison-Wesley, Publishers, Reading, MA, 1991.
- [Strotz and Wold, 1960] Robert H. Strotz and H.O.A. Wold. Recursive vs. non-recursive systems: An attempt at synthesis; Part I of a triptych on causal chain systems. *Econometrica*, 28(2):417–427, April 1960.
- [Wright, 1921] S. Wright. Correlation and causation. *Journal of Agricultural Research*, 20:557–585, 1921.