Caveats For Causal Reasoning With Equilibrium Models

Denver Dash and Marek Druzdzel**

Intelligent Systems Program, University of Pittsburgh, Pittsburgh, PA 15260, USA,
{ddash,marek}@sis.pitt.edu

Abstract. In this paper[1] we examine the ability to perform causal reasoning with recursive equilibrium models. We identify a critical 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 $\mathcal{F}$ of recursive equilibrium models that violate the Manipulation Postulate. We relate this class to the existing phenomenon of reversibility and show that all models in $\mathcal{F}$ display reversible behavior, thereby providing an explanation for reversibility and suggesting that it is a special case of a more general and perhaps widespread problem. We also show that all models in $\mathcal{F}$ 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 drastically incorrect inferences may be obtained when applying the Manipulation Postulate to equilibrium models, a result which has implications for current work on causal modeling, especially causal discovery from data.

1 Introduction

Manipulation in causal models originated in the early econometrics literature [9, 12] in the context of structural equation models, and has recently been studied in artificial intelligence, building a sound theory from some basic axioms and assumptions regarding the nature of causality [10, 7]; work which has resulted in the development of the *Manipulation Theorem* [10] and in sound and complete axiomatizations for causality [3], including the development of a new language for causal reasoning [4].

Critical to these formalisms is the assumption that when some variable in the model is manipulated, the net result from a structural standpoint will be *the removal of 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. As intuitive as this idea is, there are a few simple physical examples that have been suggested [10, 1] which seem to violate the Manipulation Postulate; in particular, systems have been identified which appear to be reversible. Neither a formal analysis of why reversibility occurs nor an indication of how widespread the problem is has been presented in the

** Currently with ReasonEdge Technologies, Pte, Ltd, 438 Alexandra Road, #03-01A Alexandra Point, Singapore 119958, Republic of Singapore, mjdruzdzel@reasonedge.com.

[1] An extended version of this paper is being submitted to the Journal of Artificial Intelligence Research.

S. Benferhat and P. Besnard (Eds.): ECSQARU 2001, LNAI 2143, pp. 192–203, 2001.
© Springer-Verlag Berlin Heidelberg 2001

causality literature. For these reasons the problem of reversibility has been widely ignored by researchers in causality.[2]

In this paper, we identify a class $\mathcal{F}$ of recursive equilibrium models that are guaranteed to violate the Manipulation Postulate; and in more complicated ways than merely reversing arcs under manipulation. Rather than relying on examples to demonstrate the existence of this class, this work is unique in that it provides a mathematical proof that $\mathcal{F} \neq \emptyset$ based on the existence of dynamic (time-dependent) models that possess recursive equilibrium counterparts. We show that the set of models which belong to $\mathcal{F}$ is surprisingly large, encompassing a wide array of the most common physical systems. We also show that every model in $\mathcal{F}$ displays reversibility, thereby providing a mathematical basis for this phenomenon 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.

Our proofs rely on the results of Iwasaki and Simon [5] who apparently were 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: Fisher [2] discusses the relationship between a time-varying model and its time-averaged counterpart; Kuipers [6] discusses temporal abstraction in dynamic qualitative models with widely varying time-scales; and Richardson [8] discusses the relationship between independencies in dynamic models and in *non*-recursive equilibrium causal models. Due to space limitations, some proofs are only sketched below; however, full proofs are available in an online appendix at: *http://www.sis.pitt.edu/~ddash/papers/caveats/appendix.ps*.

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 $\mathrm{Pa}(v)_G$ and $\mathrm{Ch}(v)_G$ to denote the parents and children, respectively, in G, for some $v \in V$. We will use $\mathrm{Anc}(v)_G$ and $\mathrm{Des}(v)_G$ to denote the ancestors and descendants of a variable v in graph G. If e is an equation then we use $Params(e)$ to denote the set of variables contained in e. If E is a set of equations, we use $Params(E)$ to represent $\bigcup_{e \in E} Params(e)$.

2 Causal Models

We are considering causal models, in the form of *structural equation models*, whereby 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 unique equation $e \in E$:

Definition 1 (total causal mapping). *A total causal mapping over E is a bijection* $\phi : V \rightarrow E$, *where E is a set of n equations with* $V \equiv \mathrm{Params}(E)$. *Obviously* ϕ *can be written equivalently as a list of associations:* $\{\langle v_1, e_1 \rangle, \langle v_2, e_2 \rangle, ..., \langle v_n, e_n \rangle\}$.

The notion of a set of equations being "self-contained" is defined precisely in [9] and [5]. Roughly the term means that the set of equations are logically independent (no equation can be derived by other equations in the set) and all parameters are identifiable. We will use the terms "structural equation model" and "causal model" interchangeably:

[2] Galles and Pearl [3] and subsequently Halpern [4] prove a theorem which they label "reversibility"; however this concept of reversibility has nothing to do with our concept. In particular, their theorem assumes that the Manipulation Postulate holds.

e_1: $f_1(w, z, v, x) = 0$	(mapped to w)
e_2: $x = x_0$	(mapped to x)
e_3: $f_2(x, y, z) = 0$	(mapped to z)
e_4: $f_3(y, z) = 0$	(mapped to y)
e_5: $f_4(v) = 0$	(mapped to v)

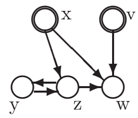

Fig. 1. An example causal model.

Definition 2 (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 over parameters V, and $\phi : V \to E$ is a total causal mapping.*

A structural equation model can be used to represent a joint probability distribution over the variables by including in each equation dependence on an independent random variable that represents the external, non-modeled factors that may introduce noise into the system. It is sufficient for our purposes to consider only models such that any equation $e \in E$ can be freely inverted for any variable $v \in \mathit{Params}(e)$ so that v can be written as a function of the remaining parameters of e, e.g., $v = f(\mathrm{Pa}(v))$. An example of such a model is shown in Figure 1. Such a causal model defines a directed graph G by directing an edge, $p \to v$, for each $p \in \mathit{Params}(e) \setminus \{v\}$.

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

Definition 3 (recursive causal model). *A causal model $M = \langle V, E, \phi \rangle$ with a causal graph G is recursive if and only 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 any causal mapping $\phi' : V \to E$ must be identical to ϕ: i.e., $\phi(v) = \phi'(v)$ for all $v \in V$.*

Proof. (sketch) This can be proven by induction by ordering the variables according to the topological sort of the graph, and showing for any mapping that if all the parents of a variable x are assigned according to ϕ then x must be also. The base case corresponds to an exogenous variable x_0 which must be assigned to $\phi(x_0)$ since that equation must have x_0 as its only parameter. □

Causal inference may require the structure of the causal graph to be altered prior to performing probabilistic inference; in particular, it is made possible by a critical postulate which we call the *Manipulation Postulate.* All formalisms for causal reasoning take the Manipulation Postulate as a fundamental starting point:

Postulate 1 (Manipulation Postulate) *If $G = \langle V, E \rangle$ is a causal graph and $V' \subset V$ is a subset of variables being manipulated, then the causal graph, G', for the manipulated system is such that $G' = \langle V, E' \rangle$, where $E' \subseteq E$ and E' differs from E by at most the set of arcs into V'.*

In plain words, manipulating a variable can cause some of its incoming arcs to be removed from the causal graph, but can effect no other change in the graph. We say that a manipulation on v is *perfect* if all incoming arcs are removed from v in the manipulated graph. For the duration of this paper we will assume that all manipulations are perfect. This postulate is related to the well-known "do" operator of Pearl [7] in that a perfect manipulation on a system specified by a causal graph G will be correctly modelled by applying the do operator to G if and only if the Manipulation Postulate holds.

Manipulation inferences require only graphs (for qualitative inference), and maybe probability distributions (for quantitative predictions). This fact makes common tools used in causal modeling, for example causal discovery, useful from a causal inference perspective. It allows us 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. 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 manipulation inferences.

3 Violating the Manipulation Postulate

Druzdzel [1], and Spirtes *et al.* [10] have pointed out that, contrary to the Manipulation Postulate, some systems appear to exhibit *reversibility* when manipulated. The standard example of a reversible system is the transmission of a bicycle. In normal operation, the rotation rate of the pedals is fixed and the wheels rotate in response. the following causal graph describes this system: *Pedal Rotation Rate* → *Wheel Rotation Rate;* however, if the bike is propped up on a bike rack and the wheel is directly rotated at some rate, then the pedals will rotate in response. The causal ordering of the system under these circumstances yields: *Wheel Rotation Rate* → *Pedal Rotation Rate*. The mere citing of physical examples, however, is not a completely satisfying demonstration that a correctly modeled system can violate the Manipulation Postulate. For example, perhaps there are hidden variables at play in our examples that, once included into the model, will produce a model that does not violate the Manipulation Postulate. Here we provide examples of systems which appear to violate the Manipulation Postulate in ways other than merely flipping arcs between manipulated children and their parents, suggesting that the problem of reversibility is a more general problem than originally supposed. All examples in this section possess recursive graphs, thus according to Lemma 1 their causal mappings are unique.

Reversibility is especially troubling from the point of view of automated causal discovery. It appears that manipulation inferences are possible only on models for which we have a strong understanding of the domain in the form of equations. Unfortunately, after learning a causal model from data, the only knowledge we have typically consists of an automatically discovered graph along with an automatically discovered probability distribution.

The Ideal Gas System Figure 2 displays one of the simplest physical systems. 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.

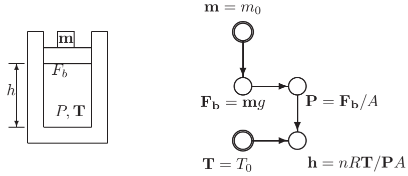

Fig. 2. Causal model of the ideal gas in equilibrium.

The equations presented in Figure 2 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. Figure 2 shows the causal graph given by constructing a causal mapping for this system. In words: *"In equilibrium, the force applied to the bottom of the piston must exactly balance 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."*

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 3. Applying the Manipulation Postulate to the model in Figure 2 yields the graph with the arcs $P \to h$ and $T \to h$ removed, as depicted in Figure 3 (b).

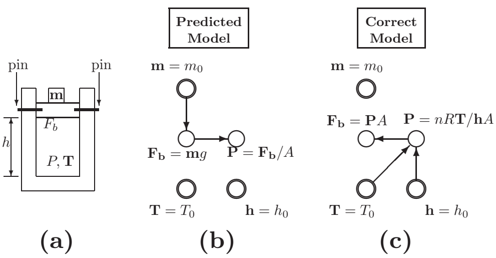

Fig. 3. The ideal gas model violates the Manipulation Postulate when h is manipulated.

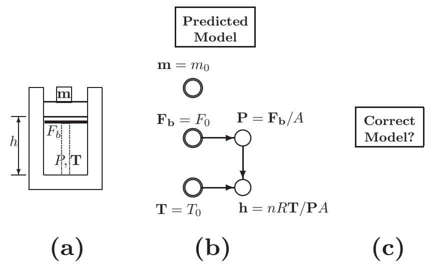

Fig. 4. No equilibrium model exists after manipulating F_b.

What is the true causal graph for this system? Fortunately since this is a simple system which we understand well, we are able to write down the governing equations, given in Figure 3 (c). 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 becomes independent of m. It is clear that the true causal model differs from that predicted by the Manipulation Postulate. Furthermore, although some arcs have been reversed in the graph, one has been deleted ($m \to F_b$) and another has changed in an apparently arbitrary fashion ($T \to h$ changed to $T \to P$). This causal graph is exactly the one that would be learned from data using the *manipulated system* to generate the data, as can be verified by calculating the independencies between variables using the equations of Figure 3 (c) with independent error terms.

There are other, even more dramatic problems with manipulating variables in this model. Refer back to the original ideal gas model of Figure 2. Imagine that for some reason we want to minimize h; it would not be unreasonable, given the graph in Figure 2, to set the value of 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 Figure 2.

Consider what happens when F_b is manipulated in this way. In the real system, the force on the bottom of the piston can be set independently of the mass 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 4 (a). Something very unexpected happens under this manipulation. Rather than getting the model of Figure 4 (b), expected by the Manipulation Postulate, 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 a causal graph alone provides no indication of. If this example seems exaggerated it is only because we have some concrete understanding about the equations underlying this system. However, imagine applying manipulations to automatically learned models of complex socio-economic or medical systems, where our basic knowledge is at most fragmentary. Performing manipulations on such models could have unpredictable effects, to say the least.

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.

The issue of modeling a causal system on varying time scales and relating models on those time scales was addressed by Iwasaki and Simon [5]. 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. After we dropped a new mass on the piston and changed T, we waited some length of time for the piston to come to rest, then measured all of our variables. In this experiment, on the contrary, we will begin measuring our variables some time Δt after we have dropped the mass on the piston. If we repeated this experiment several times we would find that the independencies and the equations governing this dynamic behavior will in general be entirely different from those in equilibrium.

Structural equation models were used in [5] to handle time-dependent systems by modeling the system at fixed, discrete time intervals. This is accomplished by creating new variables for each time slice, and adding differential equations that may relate variables across time slices. From a modeling perspective, time-dependent models and graphs are thus no different in principle from equilibrium structural equation models. 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.

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: $P = kT/h$, 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 2. In addition to these physical laws, the system is constrained by the definition of acceleration and velocity of the piston (expressed in discrete form):

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

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. 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.

This model 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 values of t.

Since 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, it can be represented by a convenient shorthand graph for an infinite sequence of time steps. In this shorthand graph temporal subscripts can be dropped and we use special dashed links, labelled *integration links* [5], to denote that a causal relationship is really occurring through a time slice. The shorthand dynamic causal graph for the ideal gas system is shown later in Figure 5 (a). Since these shorthand graphs are based on differential equations, they always make the assumption that if x and $\dot{x}$ are present in the model then $\dot{x} \rightarrow x$ across time slices.

4.1 Deriving Equilibrium Models from Dynamic Models

The dynamic graph in Figure 5 (a) represents the causal graph for the system modelled over an infinitesimal time scale; whereas, the graph from Figure 2 is modelled 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 5 (a) adds some confusion to the concept of recursivity, since it possesses cycles itself although it really is meant to represent an acyclic graph that is unrolled in time. Thus to clear up confusion we generalize the concept of recursivity for a shorthand graph:

Definition 4 (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 5 (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 $\dot{v} \in \mathrm{Pa}(v)_G$.*

The operation of *equilibration* was presented in Iwasaki and Simon [5] whereby the derivatives of a dynamic variable x are eliminated from a model by assuming that x has achieved equilibrium:

Definition 6 ($\mathbf{V_{del}(x)}$, $\mathbf{E_{del}(x)}$). *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:*

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

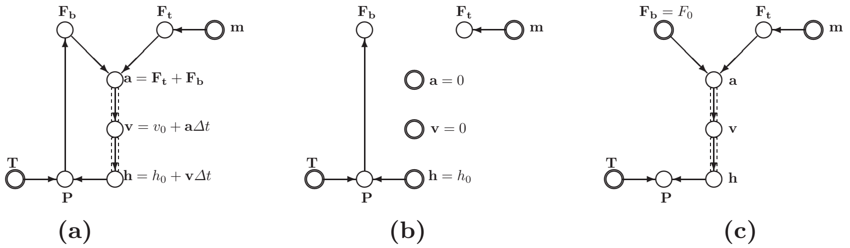

Fig. 5. (a) The dynamic ideal gas causal graph, (b) Manipulating h, (c) Manipulating F_b.

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

Definition 7 (equilibration). *Let $M = \langle V, E, \phi \rangle$ be a causal model and let $x \in V$ be a dynamic variable with $x^{(n)} \in V$ the highest order derivative of x in V. The model $M_{\tilde{x}} = \langle V_{\tilde{x}}, E_{\tilde{x}}, \phi_{\tilde{x}} \rangle$ due to the equilibration of x is obtained by the following procedure:*

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

Equilibration is equivalent to assuming that a dynamic variable x has achieved equilibrium. This implies that all of x's derivatives will be zero. Equilibration can cause the remaining set of equations to be non-self-contained. We call equilibration *well-defined* if this does not happen.

Definition 8 (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 and only if x is not a dynamic variable in M.*

Definition 9 (equilibrated model). *A causal model $M_{\tilde{x}} = \langle V_{\tilde{x}}, E_{\tilde{x}}, \phi_{\tilde{x}} \rangle$ is an equilibrated model with respect to x if and only if $M_{\tilde{x}}$ is derived from a dynamic model $M = \langle V, E, \phi \rangle$ by performing a well-defined equilibration on $x \in V$, and x is a dynamic variable in M.*

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 5 (a) 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 5 (a), we obtain the graph depicted in Figure 5 (b). 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 5 (b) to the manipulated equilibrium ideal gas system of Figure 3 (c), we can see that aside from the extra variables that were added to the dynamic model for clarity (F_t, a and v), Figure 5 (b) 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.*

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, $x = x_0$, then that fixed-point will be a stable fixed-point if and only if the following stability relation [11] holds:

$$\frac{\partial \dot{x}}{\partial x}\Big|_{x_0} < 0, \text{ (Stability condition)}$$

where $\dot{x}$ is the time-derivative of x.

According to this stability condition, the variable $\dot{x}$ must somehow be a function of x for stability to occur. What does this imply about dynamic causal models? 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 10 (The Structural Stability Principle). *Let G be a causal graph with dynamic variable v, and let* $\mathrm{Fb}(v)$ *denote the set* $\mathrm{Fb}(v) = \mathrm{Anc}(v)_G \cap \mathrm{Des}(v)_G$, *then v will possess a stable fixed-point only if* $\mathrm{Fb}(v) \neq \emptyset$.

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 5 (a), the model shown in Figure 5 (c) is obtained. We can see immediately from the causal graph that this manipulation will break the only feedback loop for x in this system, and thus according to the Structural Stability criterion, there does not exist a stable equilibrium point for this model. Our second major observation is therefore that *the dynamic model, together with the Manipulation Postulate and the Structural Stability criterion correctly predict that some manipulations will cause an instability.*

5 Theorems

In this section we formalize the observations suggested by the examples in Section 3. For the remainder of this section, let $M = \langle V, E, \phi \rangle$ be an arbitrary dynamic causal model, let $x \in V$ be a dynamic variable in M and let $M_{\tilde{x}} = \langle V_{\tilde{x}}, E_{\tilde{x}}, \phi_{\tilde{x}} \rangle$ be the causal model obtained by performing a well-defined equilibration operation on x. Let G and $G_{\tilde{x}}$ be the causal graphs for M and $M_{\tilde{x}}$, respectively and $G_x^{(0)}$ be the graph corresponding to G with all of x's integration links removed. We define $\mathrm{Fb}(x)$ to be the set of feedback variables: $\mathrm{Fb}(x) = \{\mathrm{Anc}(x)_G \cap \mathrm{Des}(x)_G\}$, and let $V_{del}(x)$ and $E_{del}(x)$ be defined as in Definition 6.

Definition 11 (RFRE Model, $\mathcal{F}$). *$M_{\mathcal{F}}$ is a recursive feedback-resolved equilibrated (RFRE) model with respect to x if and only if the following conditions hold:*

1. **Equilibration:** *$M_{\mathcal{F}}$ is derived from a dynamic model M_d by equilibrating x in M_d,*
2. **Recursivity:** *$M_{\mathcal{F}}$ and M_d are both recursive, and*
3. **Feedback-resolution:** $\{\mathrm{Fb}(x) \setminus V_{del}(x)\} \cap \mathrm{Ch}(x)_{G_d} \neq \emptyset$.

We denote the class of all RFRE models as $\mathcal{F}$, and use $\mathcal{F}(x)$ to denote the set of RFRE models with respect to 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* $v_i \in \mathrm{Anc}(v_j)_{G_x^{(0)}}$, *and*
2. *the pairs corresponding to* $\mathrm{Fb}(x)$ *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 $v \in \mathrm{Fb}(x)$. Define $\mathrm{Anc}(\mathrm{Fb}(x))_{G_x^{(0)}} \equiv \bigcup_{v \in \mathrm{Fb}(x)} \mathrm{Anc}(v)_{G_x^{(0)}}$ and $\mathrm{Des}(\mathrm{Fb}(x))_{G_x^{(0)}} \equiv \bigcup_{v \in \mathrm{Fb}(x)} \mathrm{Des}(v)_{G_x^{(0)}}$ to be the set of ancestors and descendants, respectively of $\mathrm{Fb}(x)$. By transitivity of the ancestor and descendant relationships, if there exists a $v \in \mathrm{Anc}(\mathrm{Fb}(x)) \cap \mathrm{Des}(\mathrm{Fb}(x))$ then $v \in \mathrm{Fb}(x)$. Thus an ordering can be defined such that $O(v_{anc}) < O(v_{fb}) < O(v_{des})$ for arbitrary variables $v_{anc} \in \mathrm{Anc}(\mathrm{Fb}(x)) \setminus \mathrm{Fb}(x)$, $v_{des} \in \mathrm{Des}(\mathrm{Fb}(x)) \setminus \mathrm{Fb}(x)$, and $v_{fb} \in \mathrm{Fb}(x)$. □

Lemma 3. *Let $\overline{F}$ denote the set $V_{\tilde{x}} \setminus \{\mathrm{Fb}(x) \cup \{x\}\}$. If M and $M_{\tilde{x}}$ are recursive then $\phi_{\tilde{x}}(v) = \phi(v)$ for all $v \in \overline{F}$.*

Proof. (sketch) Using Lemma 2 and a recursive proof similar to that of Lemma 1, it can be proven that it will always be possible to define a mapping ϕ' such that each $v \in \overline{F}$ gets mapped to $\phi(v')$ for some $v' \in \overline{F}$. It then follows by Lemma 1 that since $\phi_{\tilde{x}}$ is recursive, $\phi' = \phi_{\tilde{x}}$. □

The next lemma says, informally, that all ancestors of x in $\mathrm{Fb}(x)$ that are not dynamic variables in $G_x^{(0)}$ must pass through $x^{(n)}$:

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

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: $x^{(n)} \to x^{(n-1)} \to \cdots \to x$, all $x^{(i)}$ such that $i \neq n$ must have a single parent which is connected by an integration link. Therefore, all $v \in \mathrm{Anc}(x)_G \setminus V_{del}(x)$ must be ancestors of $x^{(n)}$, i.e., $\mathrm{Fb}(x) \setminus V_{del}(x) \subseteq \mathrm{Anc}(x^{(n)})_{G_x^{(0)}}$. □

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

Proof. (sketch) First note that the result follows for all $x^{(j)}$ such that $j < n$, because by construction $\mathrm{Pa}(x^{(j)}) = \{x^{(j+1)}\}$ in M. Thus we only need to prove that $x^{(n)} \notin \mathrm{Ch}(x)$. $M_{\tilde{x}}$ is recursive by assumption; therefore, by Lemma 1 there only exists one causal mapping, $\phi_{\tilde{x}}$. However, if $x^{(n)} \in \mathrm{Ch}(x)$ then it can be shown by Lemma 3 that there exists a mapping ϕ' such that $\phi'(x) = \phi(x^{(n)})$, and all other variables in $V_{\tilde{x}}$ retain the associations specified by ϕ. By Lemma 4 it follows in such case that $\mathrm{Anc}(x) \cap \mathrm{Des}(x)$ is non-empty, which contradicts the recursivity of $\phi_{\tilde{x}}$. □

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

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 $\phi_{\tilde{x}}$. By construction $x \neq v_i$, and by Lemma 3, $v_i \in \mathrm{Fb}(x)$. Since $x \in \mathit{Params}(e_i)$ and since $\langle v_i, e_i \rangle \in \phi$ it must be the case that $v_i \in \mathrm{Ch}(x)_G$. 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 $v_i \in \mathrm{Fb}(x) \setminus V_{del}(x)$, and since $v_i \in \mathit{Params}(e_i)$ it must be the case that $v_i \in \mathrm{Pa}(x)_{G_{\tilde{x}}}$. □

Lemma 7. *If $M_{\tilde{x}} \in \mathcal{F}(x)$ 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 $x \to v$ for all $v \in \mathrm{Ch}(x)_G \cap V_{\hat{x}}$.*

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 $G_{\hat{x}}$. □

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

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

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 $M_{\tilde{x}} \in \mathcal{F}(x)$, by Lemma 6 there exists a $v \in \mathrm{Ch}(x)_G$ such that $v \to x$ in $G_{\tilde{x}}$; however, according to Lemma 7, the edge $x \to v$ must exist in $G_{\hat{x}}$. Thus, manipulating x in $G_{\tilde{x}}$ by applying the Manipulation Postulate leads to an incorrect graph $G_{\tilde{x}}|_{\hat{x}}$, because it will not contain an edge between v and x. □

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

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

6 Discussion

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 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 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. What then is the minimal information needed to insure that a model will support manipulation? 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.

Acknowledgments We would like to thank Greg Cooper, Clark Glymour, Richard Scheines, Herbert Simon, and Peter Spirtes for helpful and intense discussions, and Javier Díez and Nanny Wermuth for their comments and encouragement. The views expressed in this paper are not necessarily shared by any of those acknowledged above. Our work was supported by the National Aeronautics and Space Administration under the Graduate Students Research Program (GSRP), grant number S99-GSRP-085, the Air Force Office of Scientific Research under grant number F49620–00–1–0122, and by the National Science Foundation under Faculty Early Career Development (CAREER) Program, grant IRI–9624629.

References

1. 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.
2. Franklin M. Fisher. A correspondence principle for simultaneous equation models. *Econometrica*, 38(1):73–92, January 1970.
3. D. Galles and J. Pearl. Axioms of causal relevance. *Artificial Intelligence*, 97(1–2):9–43, 1997.
4. Joseph Y. Halpern. Axiomatizing causal reasoning. *Journal of Artificial Intelligence Research*, 12:317–337, 2000.
5. Yumi Iwasaki and Herbert A. Simon. Causality and model abstraction. *Artificial Intelligence*, 67(1):143–194, May 1994.
6. 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.
7. 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.
8. Thomas Richardson. *Models of Feedback: Interpretation and Discovery.* PhD dissertation, Carnegie Mellon University, Department of Philosophy, 1996.
9. Herbert A. Simon. Causal ordering and identifiability. In William C. Hood and Tjalling C. Koopmans, editors, *Studies in Econometric Method. Cowles Commission for Research in Economics. Monograph No. 14*, chapter III, pages 49–74. John Wiley & Sons, Inc., New York, NY, 1953.
10. Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search.* Springer Verlag, New York, 1993.
11. Steven H. Strogatz. *Nonlinear Dynamics and Chaos with Applications to Physics, Biology, Chemistry, and Engineering.* Addison-Wesley, Publishers, Reading, MA, 1991.
12. Robert H. Strotz and H.O.A. Wold. Recursive vs. nonrecursive systems: An attempt at synthesis; part I of a triptych on causal chain systems. *Econometrica*, 28(2):417–427, April 1960.