

A Comparison of Pearl's Causal Models and Settable Systems

Halbert White¹
Department of Economics
University of California, San Diego

PRELIMINARY: DO NOT CIRCULATE OR CITE
WITHOUT AUTHOR'S PERMISSION

April 1, 2007

¹Halbert White is Chancellor's Associates Distinguished Professor of Economics, Department of Economics 0508, University of California, San Diego 92093-0508 (hwhite@ucsd.edu). The author is grateful for discussion and comments on previous work by Judea Pearl that stimulated this work, and for the comments and suggestions of Karim Chalak, Douglas R. White, and Scott White. Any errors are solely the author's responsibility.

Abstract

This paper examines the relations between the causal models of Pearl and the settable systems framework recently introduced by White and Chalak. We pay particular attention to the suitability of these two approaches for analyzing the behavior of optimizing, interacting agents, the central concern of economics. We show that Pearl's causal model is nested in the settable system framework, and we describe a number of ways in which Pearl's causal models are not well suited to the study of interacting, optimizing agents. In contrast, settable systems have been explicitly designed to facilitate this study, accommodating systems of agents that not only optimize and interact, but that may learn from experience. We illustrate using examples from microeconomics, option pricing, game theory, and recursive estimation. Among the features of settable systems that distinguish it from Pearl's causal models and that help deliver its capabilities are its countable rather than finite structure, its explicit use of attributes, and the introduction of partitions and partition-specific response functions.

Keywords: structural modeling; mutual causality; economic agents; game theory

1 Introduction

Pearl's work on causality, especially as embodied in his landmark book (Pearl, 2000), represents a rich framework in which to understand, analyze, and explain causal relations. In particular, Pearl's definition of a Causal Model (Pearl, 2000, Def. 7.1.1, p. 203) provides a formal statement of the elements essential to causal reasoning. According to this definition, a Causal Model is a triple $M \equiv (U, V, F)$, where $U \equiv \{u_1, \dots, u_m\}$ is a collection of background or "exogenous" variables determined outside the model, $V \equiv \{v_1, \dots, v_n\}$ is a collection of "endogenous" variables determined within the model, and $F \equiv \{f_1, \dots, f_n\}$ is a collection of "structural" functions that specify how each endogenous variable is determined by the other variables of the model, so that

$$v_i = f_i(v_{(i)}, u), \quad i = 1, \dots, n.$$

Here $v_{(i)}$ denotes the vector containing every element of V except v_i ; we also write $u \equiv (u_1, \dots, u_m)$. The integers m and n are finite. Finally, the definition requires that F yields a unique fixed point for each u , so that there exists a unique collection of functions $G \equiv \{g_1, \dots, g_n\}$ such that for each u ,

$$v_i = g_i(u) = f_i(g_{(i)}(u), u), \quad i = 1, \dots, n. \tag{1}$$

In stating the elements of this definition, we have adapted Pearl's original notation somewhat to facilitate the discussion to follow, but all essential elements of the definition are present and complete. For ease of reference, we refer to this as the Pearl Causal Model (PCM). A variant of this model analyzed by Halpern (2000) does not require a fixed point, but if any exist, there may be multiple collections of functions G yielding a fixed point. We refer to such a model as a Generalized Pearl Causal Model (GPCM).

White and Chalak (2006) (WC) introduce the settable system framework, explicitly constructed to facilitate causal analysis of the systems that are the central concern of economics, namely systems of optimizing, interacting agents. The settable system framework builds on a number of different prominent approaches to the study of causal relations, specifically the classical structural equations framework of the Cowles Commission, the methods of modern labor econometrics developed by Heckman and his colleagues, the methods of the treatment effect literature developed by Rubin and his colleagues, and the causal model of Pearl and his colleagues. (See WC

for references and further discussion.) In particular, WC claim that settable systems encompass the PCM. Given such a claim, it is fair to ask whether settable systems in fact contain anything beyond the PCM or GPCM, and, if so, whether the additional features of this framework are of any use or significance. Our purpose here is to address these questions.

In what follows, we first show that the (G)PCM is strictly contained in the settable system framework, so that all causal analysis possible within these systems is admitted by settable systems. Any result of this sort is of little consequence if the additional flexibility of settable systems has no tangible benefits. To address this, we proceed to show how the settable system framework facilitates causal analysis of the behavior of systems of one or more optimizing and interacting agents. We contrast this with the performance of the PCM or GPCM for these systems, describing a number of ways in which these are not well-suited to their analysis. Because optimization is important well beyond economics, because the systems of interacting agents we examine are standard in game theory, which has important applications well beyond economics, and because, as we show, settable systems readily admit learning, which is of general interest, the useful features of settable systems may also be claimed to extend well beyond economics. We pay particular attention to the opportunities afforded by settable systems for obtaining satisfying explanations, both causal and non-causal, of the behavior of optimizing, interacting agents.

2 Settable Systems and the PCM

In this section we compare the elements of WC's settable system framework to their PCM analogs, commenting briefly on areas of contrast. Subsequent sections develop some implications of these contrasts. For ease of reference, we provide a formal definition of WC's attribute-indexed partitioned settable systems in the appendix.

2.1 Correspondences between Settable System and PCM Elements

The settable system analogs of the "nodes" indexed in the PCM as $i \in \{1, \dots, n\}$ are agent-specific pairs (h, j) . Each of a countable number of agents ($h \in \{1, 2, \dots\}$) may govern a countable number of variables, indexed by $j \in \{1, 2, \dots\}$. As WC discuss, the notion of "agent" should be construed broadly. It could be a person, firm, or market,

or, more broadly, a neuron, cell, electronic circuit, or government. One important role of the variables indexed by j is to represent decisions under the control of the agent. These decisions should also be construed broadly; they may be active (e.g., a choice) or passive (e.g., release of a hormone). Another role of the agent-governed variables is to represent elements of the agent's knowledge or beliefs. This role is particularly important in systems involving learning.

For concreteness and to facilitate comparisons with the PCM in the discussion to follow, we focus primarily on the role of the agent-governed variables as decisions. We thus refer to "agent-*decision* pairs" (h, j) to make this role clear. In Section 6.2, however, we discuss learning systems in which both roles, as decisions and as knowledge elements, are explicit, exploiting the general interpretation of the variables indexed by agent-specific pairs (h, j) .

To map settable systems to the PCM, it suffices to restrict attention in the settable system to a finite number of agents, each having a finite number of decisions, so that there is a finite total number n of agent-*decision* pairs. Then each node i corresponds to a distinct agent-*decision* pair, $(h, j) = (h_i, j_i)$.

As settable systems are countable and the PCM is finite, settable systems encompass this aspect of the PCM. We examine the practical value of countability later, in Section 6.

In settable systems, each agent h has associated characteristics or "attributes," $a_h \in \mathcal{A}$. The attribute space \mathcal{A} permits a_h to be a countably-dimensional vector. Components of this vector may be binary, categorical, integer-valued, or real-valued. Attribute components may also be decision-specific. In the settable systems framework, attributes are fixed, that is, not subject to counterfactual variation. The PCM references no attribute other than the node index i , so only settable systems with $\mathcal{A} = \emptyset$ can correspond to the PCM. In Sections 4 and 5 we discuss how explicitly defined attributes help in the process of explaining empirical phenomena.

In settable systems, the "fundamental" variables $Z_0 \equiv (Z_{0,j}, j = 1, 2, \dots)$ are random variables (i.e., measurable functions) whose realizations correspond to the exogenous variables U of the PCM. For conciseness, in settable systems the fundamental variables are associated with "agent" $h = 0$. A difference between settable systems and the PCM is that there may be a countable infinity of fundamental variables, whereas there is a finite number m of exogenous variables in the PCM. To map settable systems to the PCM, it suffices to restrict the number of fundamental

variables to m .

The settable system analogs of the endogenous variables V of the PCM are "responses," $Y_{h,j}$, determined as

$$Y_{h,j} = r_{h,j}(Z_{(h,j)}, a), \quad j = 1, 2, \dots; h = 1, 2, \dots,$$

where the "response function" $r_{h,j}$ is the analog of the structural function f_i in the PCM. The argument a appearing in the response function represents the vector of attributes for all agents of the system, $a \equiv (a_1, a_2, \dots)$. The variables $Z_{(h,j)}$ appearing as function arguments are random variables called "settings" in the settable system framework; these have no direct analog in the PCM. For every agent-decision pair (h, j) , $Z_{h,j}$ is a random variable, $j = 1, 2, \dots; h = 1, 2, \dots$. We write $Z_{(h,j)}$ to denote every setting of the system except $Z_{h,j}$, including Z_0 . The term "setting" is used to suggest that these variables are not determined by their governing agent, but are instead set arbitrarily, taking the values $Z_{h,j}$. By analogy to the $h > 0$ case, we call the elements of Z_0 "fundamental settings." An underlying probability measure P governs the joint behavior of $\{Z_{h,j}, j = 1, 2, \dots; h = 0, 1, 2, \dots\}$, in a manner analogous to the way a probability measure P governs the behavior of U in the probabilistic version of the PCM (Pearl, 2000, p. 706).

The settable system implementation of the "setting" concept formalizes the Strotz and Wold (1960) device of "wiping out" the structural equation that otherwise determines a given response and replacing this response with an arbitrary value. This same idea was used by Fisher (1970). The importance of this device has been perceptively and repeatedly emphasized by Pearl, although the PCM implements this notion in a manner distinct from that of settable systems. We provide a detailed discussion in Section 2.2 below.

Despite the presence of elements with no direct analog in the PCM (e.g., settings for $h > 0$), we can nevertheless restrict the settable system to correspond to the PCM by enforcing the final PCM requirement of a unique fixed point for the structural equations. In settable systems, this is the requirement that $r \equiv \{r_{h,j}, j = 1, 2, \dots; h = 1, 2, \dots\}$ is such that for each (z_0, a) , there exists a unique fixed point $y \equiv (y_{h,j}, j = 1, 2, \dots; h = 1, 2, \dots)$ such that

$$y_{h,j} = r_{h,j}(y_{(h,j)}, z_0, a), \quad j = 1, 2, \dots; h = 1, 2, \dots,$$

where $y_{(h,j)}$ denotes the vector containing every element of y except $y_{h,j}$. We emphasize that although this restriction may be imposed within the settable system

framework, it is by no means required. The absence of a fixed point requirement is an important distinguishing feature of settable systems. We examine its consequences in several of the sections that follow.

It follows that the PCM is strictly nested within the settable system framework. That is, the PCM is a settable system that is restricted to a finite number of agent-decision pairs, suppresses attributes ($\mathcal{A} = \emptyset$), and enforces a fixed point condition on r . The GPCM is similarly nested, being a settable system that is restricted to a finite number of agent-decision pairs and suppresses attributes. Consequently, any causal analysis possible with the PCM or GPCM is also possible within the settable system framework.

As emphasized in the introduction, however, this fact is of little consequence if the additional flexibility of settable systems has no useful or significant implications. We devote Sections 3 through 6 to an examination of the consequences of this flexibility.

2.2 The *do* Operator, Potential Responses, Partitioning, and Partition-Specific Response Functions

In the PCM, the Strotz-Wold "wiping out" device is formalized in Pearl's (2000) Definitions 7.1.2 (Submodel), 7.1.3 (Effect of Action), 7.1.4 (Potential Response), and 7.1.5 (Counterfactual). These notions are essential to causal discourse in the PCM. Specifically, if $V = (y, x)$, where x is a given $\ell \times 1$ subvector of V (so that y is $(n - \ell) \times 1$), Pearl defines the "*do* operator" as "the minimal change in M required to make $X = x$ hold true under any u ," so that the "effect of action $do(X = x)$ " on M is given by the "submodel" denoted $M_x \equiv (U, V, F_x)$, where F_x is the collection of functions F with $f_i, i = n - \ell + 1, \dots, n$, replaced by the constant vector x .

Pearl's "potential response" $Y_x(u)$ to action $do(X = x)$ can then be represented as the unique fixed point for the system of equations

$$y_i = f_i(y_{(i)}, x, u), \quad i = 1, \dots, n - \ell, \quad (2)$$

where $y_{(i)}$ is the vector containing all but the i th element of y , assuming this fixed point exists. Equivalently, we can write

$$y_i = g_{x,i}(x, u) = f_i(g_{x,(i)}(x, u), x, u), \quad i = 1, \dots, n - \ell, \quad (3)$$

where $g_x \equiv (g_{x,1}, \dots, g_{x,n-\ell})$ is the $(n - \ell) \times 1$ vector function delivering the unique fixed point of the system of equations (2), so that $g_x(x, u)$ corresponds to Pearl's potential

response, $Y_x(u)$. Equations (1) are the special case of equations (3) with x taken to be null (i.e., of dimension zero).

We emphasize that without potential response functions, causal discourse involving endogenous variables is not possible in the PCM. Indeed, for every choice of subvector and value x , the existence of a unique fixed point for equations (1) is a necessary but not a sufficient condition for the existence of a unique fixed point for equations (2). Thus, even if the PCM is well defined, there is no guarantee that the potential response function is well defined for any subvector or values of the endogenous variables, in which case causal analysis may be restricted to discussions involving only exogenous variables. Further, elements of the GPCM that are not also PCMs have no unique fixed point for equations (1), so even when x is null, one cannot define the potential response function for these systems. The GPCM is thus not a particularly powerful framework for analyzing causal relations among endogenous variables when equations (1) do not have a unique fixed point.

Settable systems do not impose a fixed point requirement, so in this aspect they resemble GPCMs, readily accommodating systems with no fixed point, a unique fixed point, or multiple fixed points. On the other hand, they implement the Strotz-Wold device in a manner that does not preclude causal analysis involving endogenous variables in general systems. The key to this is that in contrast to PCMs, where the focus is on which variables are set to arbitrary values, settable systems instead place emphasis on which variables are jointly free to respond. The partitioning device that implements this emphasis (the counterpart of the *do* operator) accommodates either the absence or presence of fixed points in a way that nevertheless interlinks causal relations across different partitions. As we see below, this interlinkage can identify and eliminate irrelevant fixed points that otherwise invalidate the PCM or the potential response functions and that cannot be formally identified in either the PCM or the GPCM.

Thus, in settable systems, the purpose of partitions is to specify groups of variables that are free to respond jointly to arbitrary settings of all other variables of the system. Formally, a partition $\Pi \equiv \{\Pi_b, b = 1, 2, \dots\} = \{\Pi_b\}$ is a countable, mutually exclusive, exhaustive collection of sets Π_b ("blocks") of agent-decision indexes for $h > 0$. That is, $\Pi_a \cap \Pi_b = \emptyset$ for $a \neq b$ (mutual exclusivity), and $\cup_b \Pi_b = \{(h, j), j = 1, 2, \dots; h = 1, 2, \dots\}$ (exhaustion). Examples are the elementary partition, $\Pi^e \equiv \{\Pi_{h,j}\}$, where $\Pi_{h,j} \equiv \{(h, j)\}$; the agent partition $\Pi^a = \{\Pi_h\}$, where $\Pi_h \equiv \{(h, j), j = 1, 2, \dots\}$, and

the fundamental partition $\Pi^f \equiv \{(h, j), j = 1, 2, \dots; h = 1, 2, \dots\}$. Π^e is the "finest" partition; Π^f is the "coarsest." By convention, we take $\Pi_0 \equiv \{(h, j), j = 1, 2, \dots; h = 0\}$. Variables in a given block Π_b are those jointly responding to arbitrary settings for all other variables ($Z_{i,k}$ for $(i, k) \notin \Pi_b$).

The *partition-specific* response functions $r_{h,j}^\Pi$ specify how the joint responses of the agents in a given block b depend on the settings of variables outside the block, $Z_{(b)}$, as

$$Y_{h,j} = r_{h,j}^\Pi(Z_{(b)}, a), \quad (h, j) \in \Pi_b.$$

This response may be the "partial equilibrium" fixed point that holds among the agents belonging to the given block when presented with arbitrary settings $Z_{(b)}$. That is, if indeed there is a unique fixed point $y_b \equiv (y_{h,j}, (h, j) \in \Pi_b)$ such that

$$y_{h,j} = r_{h,j}(y_{b,(h,j)}, z_{(b)}, a), \quad (h, j) \in \Pi_b,$$

where $y_{b,(h,j)}$ denotes the vector containing every element of y_b except $y_{h,j}$, then $r_b^\Pi \equiv (r_{h,j}^\Pi, (h, j) \in \Pi_b)$ can represent this fixed point, so that

$$y_{h,j} = r_{h,j}^\Pi(z_{(b)}, a) = r_{h,j}(r_{b,(h,j)}^\Pi(z_{(b)}, a), z_{(b)}, a), \quad (h, j) \in \Pi_b.$$

Here, $r_{b,(h,j)}^\Pi$ denotes the vector function containing all elements of r_b^Π except $r_{h,j}^\Pi$. In this case, r_b^Π corresponds precisely to g_x defined above.

On the other hand, $r_b^\Pi \equiv (r_{h,j}^\Pi, (h, j) \in \Pi_b)$ may represent one of several possible such fixed points, determined by some well-motivated equilibrium selection operator, thereby eliminating irrelevant fixed points that pose difficulties for the (G)PCM. Alternatively, even when there is no fixed point, r_b^Π may represent some meaningful default joint response. The key property is that r_b^Π specifies the joint responses for the agent-decision pairs of the block to arbitrary settings of variables outside the block, specific to the group of jointly responding variables, *whatever* those responses might be. It thus forms a valid basis for counterfactual and therefore causal analysis, specific to the partition.

A useful aspect of the potential response functions of the PCM is that when they exist, they create opportunities for structural identification of effects of interest, useful in statistical estimation. An example is the identification of structural coefficients from reduced form coefficients exploited by Haavelmo's (1943) method of indirect least squares. Because of the flexibility of settable systems, they preserve these opportunities when they exist, and create further opportunities when the PCM or its

potential responses are not well defined. For brevity, we do not further consider these opportunities here but take them up elsewhere.

We emphasize that although the "partitioned settable system" that arises from this approach is tailored to a collection of specific counterfactual exercises of interest, it is not just a different custom-tailored PCM. First, there is still no fixed point requirement, attributes explicitly appear, and there may be a countable number of agent-decision pairs; thus, the partitioned settable system is not a PCM. Nor is it a GPCM, due to the presence of attributes and countable dimensionality. More importantly, however, the flexibility afforded by partition-specific responses plays a key role in permitting the elimination of fixed points that pose difficulties for PCMs or GPCMs. We give examples below.

3 Fixed Points and Partition-Specific Responses

3.1 Inadequacy of the PCM under Optimizing Behavior

In this section, we begin to examine the question of whether the additional flexibility of settable systems offers any tangible benefits. Here, we analyze the optimizing behavior of a single firm. We show how, in a certain precise sense, the PCM is inadequate for describing the behavior of the firm, a consequence of the PCM's lack of partition-specific response functions, whereas the firm's behavior is readily accommodated by the partition-specific response functions of the settable system framework.

Consider a profit-maximizing price-taking firm that produces a single good with a given technology using capital (i.e., physical capital, such as plant and equipment) and labor as input factors of production. In the short run, the firm operates with one input fixed, choosing the level of the other to maximize profits. If capital is fixed, the firm's short-run labor demand (optimal labor choice) is a function of prices and the given level of capital. If labor is fixed, the firm's short-run capital demand (optimal capital choice) is a function of prices and the given level of labor. Only one of these short-run demands can operate for any given firm; the other is counterfactual. An important feature of the short run is that the firm will operate at prices resulting in a loss (negative profits), whenever the loss from shutting down entirely would be greater than that from continuing to operate. (The firm must pay the fixed input regardless of whether it is productively employed.)

In the long run, the firm operates with both inputs chosen optimally. The resulting

long-run labor and capital demands are functions of prices. An important feature of the long run is that the firm will never operate at a loss, because it can always shut down and earn zero profit. These descriptions of the short-run and long-run behavior of the firm are standard in any modern microeconomics textbook (e.g., Nicholson, 2005, ch. 8).

The short-run behavior of the firm maps directly to the PCM. The exogenous variables U are the output and input prices ($p =$ output price, $r =$ rental cost of capital, $w =$ wage of labor). The endogenous variables V are the levels of capital (k) and labor (l). Under standard conditions, the short-run capital and labor demand functions can be written

$$\begin{aligned} k &= f_1(l, p, r, w) \\ l &= f_2(k, p, r, w), \end{aligned}$$

so we can take $F \equiv \{f_1, f_2\}$ for the PCM. Under some further standard conditions, for each (p, r, w) , this set of structural equations has a unique solution that we represent as

$$\begin{aligned} k &= g_1(p, r, w) \\ l &= g_2(p, r, w). \end{aligned} \tag{4}$$

As all conditions of the PCM are satisfied, the causal model $M \equiv (U, V, F)$ should provide all information needed to answer any causal question.

Nevertheless, as we now show, M does not correctly describe the behavior of the firm in the long run. Economic theory dictates that long-run input demands are given by

$$\begin{aligned} k &= h_1(p, r, w) \\ l &= h_2(p, r, w), \end{aligned} \tag{5}$$

where $H \equiv \{h_1, h_2\}$ are the long-run capital and labor demand functions. The issue is whether or not $h_1 = g_1$ and $h_2 = g_2$, as both equations (4) and (5) are supposed to describe the causal outcomes for the endogenous variables given the exogenous variables. We show that these functions differ by finding (p^*, r^*, w^*) such that $h_1(p^*, r^*, w^*) \neq g_1(p^*, r^*, w^*)$ or $h_2(p^*, r^*, w^*) \neq g_2(p^*, r^*, w^*)$.

For this, fix r and w , and let $p_{r,w}$ denote the firm's minimum long-run average cost for given r and w . This exists and is positive under mild conditions, for example,

those generating the classical U -shaped average cost function of the textbooks. Let $p^* \equiv p_{r,w} - \epsilon$ for some small $\epsilon > 0$. The firm's long-run capital and labor demands must satisfy

$$\begin{aligned} h_1(p^*, r, w) &= 0 \\ h_2(p^*, r, w) &= 0, \end{aligned}$$

because any other input choice would yield losses. On the other hand, losses are permitted in the short run. Thus, by choosing ϵ sufficiently small to ensure that operating is less unprofitable than shutting down, we can ensure that the firm continues to operate at p^* in the short run. This implies either or both of the conditions

$$\begin{aligned} g_1(p^*, r, w) &> 0 \\ g_2(p^*, r, w) &> 0. \end{aligned}$$

It follows that $h_1 \neq g_1$ or $h_2 \neq g_2$ or both. Because of this disparity, M cannot provide the information needed to correctly answer questions about how the firm behaves in the long run, when k and l can both be freely chosen. When disparities of this sort arise, we say that M is "inadequate." Nor can recourse to a GPCM resolve the difficulty, as here the problem is not the absence of a unique fixed point, but the inability of the unique fixed point to properly describe the behavior of the firm.

Such inadequacies are removed by permitting the response functions for a group of variables jointly responding to settings of other variables to depend on the grouping. This first requires a convenient means for keeping track of which variables are responding and which are set. The partitioning device described in the previous section accomplishes this. The partition-specific response functions then provide the group-dependent joint response functions. With this, we have sufficient structure to provide correct answers to causal questions under optimizing behavior in both the short and the long run.

In terms of the example above, the short-run behavior of the firm above corresponds to the elementary partition, $\Pi^e \equiv \{\{1\}, \{2\}\}$, where the endogenous variables respond separately: capital ($\{1\}$) responds to exogenous variables ($\{0\}$) and labor ($\{2\}$); and labor ($\{2\}$) responds to exogenous variables ($\{0\}$) and capital ($\{1\}$). The response functions here are f_1 and f_2 . The long-run behavior of the firm corresponds to the fundamental partition, $\Pi^f \equiv \{\{1, 2\}\}$, where the endogenous variables ($\{1, 2\}$) respond jointly to the exogenous variables ($\{0\}$). The response functions here are

h_1 and h_2 . As the example illustrates, these need not be identical to their short-run analogs, g_1 and g_2 . We note for completeness, however, that for all r, w , and $p \geq p_{r,w}$, we do have $h_1(p, r, w) = g_1(p, r, w)$ and $h_2(p, r, w) = g_2(p, r, w)$.

3.2 Resolution of Multiple Fixed Points by Optimizing Behavior

Now consider the special case in which the firm's technology is given by

$$q(k, l) = k^\alpha l^\beta,$$

where $\alpha > 0$, $\beta > 0$, $\alpha + \beta < 1$. This is the standard Cobb-Douglas production function with decreasing returns to scale. The short-run capital and labor demands in this case are

$$\begin{aligned} k &= f_1(l, p, r, w) = l^{\beta/(1-\alpha)} (\alpha p/r)^{1/(1-\alpha)} \\ l &= f_2(k, p, r, w) = k^{\alpha/(1-\beta)} (\beta p/w)^{1/(1-\beta)}. \end{aligned} \quad (6)$$

This system of structural equations has two fixed points:

$$\begin{aligned} k &= g_1(p, r, w) = 0 \\ l &= g_2(p, r, w) = 0, \end{aligned} \quad (7)$$

and

$$\begin{aligned} k &= h_1(p, r, w) \\ &= \alpha^{(1-\beta)/(1-\alpha-\beta)} \beta^{\beta/(1-\alpha-\beta)} r^{-(1-\beta)/(1-\alpha-\beta)} w^{-\beta/(1-\alpha-\beta)} p^{1/(1-\alpha-\beta)} \\ l &= h_2(p, r, w) \\ &= \alpha^{\alpha/(1-\alpha-\beta)} \beta^{(1-\alpha)/(1-\alpha-\beta)} r^{-\alpha/(1-\alpha-\beta)} w^{-(1-\alpha)/(1-\alpha-\beta)} p^{1/(1-\alpha-\beta)}. \end{aligned} \quad (8)$$

The PCM requires a unique fixed point, so this system falls outside the PCM. Although Halpern's (2000) GPCMs cover systems with multiple fixed points, in this example not all fixed points are causally relevant. Instead, the firm's optimizing behavior eliminates one of the fixed points from consideration.

Specifically, equations (8) uniquely describe the firm's profit-maximizing long-run capital and labor demands. Equations (7) represent a profit-*minimizing* zero profit

solution; the firm's optimizing behavior ensures that this fixed point never describes the firm's behavior. Although the conditions of the PCM do not hold, the fact that only one of the fixed points is relevant effectively restores the insights of the PCM. Nevertheless, neither the PCM nor the GPCM have a formal mechanism for removing such irrelevant fixed points. On the other hand, WC's partition-specific response functions are explicitly designed to formally permit optimization to eliminate the profit-minimizing fixed point $l = k = 0$ in describing the firm's long-run behavior.

Note that here the profit-maximizing fixed point of equations (6) does deliver the long-run capital and labor demands. The difficulties of the previous section do not arise, as here the minimum long-run average cost $p_{r,w}$ is zero, due to decreasing returns to scale. In both cases, however, the use of partition-specific response functions avoids the obstacles faced by the PCM.

4 The Role of Attributes in Explanation

Explanations provide answers to the questions "why?" and "how?" According to the deductive-nomological theory of explanation of Hempel and Oppenheim (1948) and Popper (1959), one explains a given phenomenon by demonstrating that it is the logical deductive consequence of a universal law or principle applied to a set of premises or given conditions. Nagel (1961, ch.2) provides discussion.

In economics, the phenomena to be explained are the decisions of economic agents and the results of agent interactions. Agent decisions are explained as the logical consequence of the principle of optimizing behavior (e.g., profit maximization) in the face of conditions represented by constraints (e.g., on available physical capital) and variables outside the agent's control, such as prices. The outcomes of agent interactions are explained as the mutually consistent outcomes of optimizing agents. The theory of optimal choice of factor inputs described above is an example of a formal explanatory theory of agent behavior. Below we discuss examples of formal explanations of agent interactions based on game theory.

Whenever causal relations governed by the PCM or the settable system framework emerge from such an explanatory theory, one has an explanation of those causal relations. For example, the theory of the profit-maximizing firm explains why and how in the short run capital inputs causally affect labor demand and why and how in the long run factor prices causally affect labor and capital demand. Further, as Pearl

(2000, pp. 221-223) insightfully discusses, the operations of such causal relations can themselves act as universal principles that can support explanations of consequent phenomena.

Typically, the premises relevant to a given phenomenon involve the identity and characteristics or attributes of the entities involved in the phenomenon. As emphasized by Holland (1986) in his formalization of Rubin's (1974) treatment effect framework, the distinguishing feature of an attribute is that it is an inherent characteristic of the entity, not subject to counterfactual variation within the explanatory system. Entities with different attributes are necessarily different entities, although different entities may have the same attributes. In economics, examples of such entities are individuals, firms, or markets. We refer to these entities generically as "agents." In the example of the last section, the technology of the firm, represented by the quantities α and β , provides an example of an attribute of the firm.

Given that attributes form part of the premises of the explanatory theory, they necessarily play a key role in explaining any given phenomenon. Specifically, they permit us to make potentially falsifiable predictions about outcomes for different entities or about interactions between different entities. Observation of these outcomes may enable us to reject a given explanatory theory. Consequently, not only causal variables but also attributes play a central role in constructing explanations; we may thus justifiably refer to both as "explanatory factors."

The PCM expresses attributes in their most fundamental form. Specifically, by indexing the structural functions f_i with index i , the PCM signals that nodes with different indexes are distinct; that is, each node possesses the fundamental attribute of individuality. In the PCM, it is significant that the node indexes are not subject to counterfactual variation. That is, we could equally well write

$$v_i = f(v_{(i)}, u; i), \quad i = 1, \dots, n,$$

for some unique function f , but it is understood that variation in i belongs to a different category than variation in $v_{(i)}$. The latter is the counterfactual variation that underlies the analysis of causal effects. The former is a non-counterfactual and therefore acausal or non-causal variation that permits us to direct our attention to one specific node or another.

In the settable system framework, attributes are expressed in their most flexible form. The response functions defined by $r_{h,j}(z_{(h,j)}, a)$ express responses as a function

not necessarily just of the agent's own attributes a_h , but as a function of the attributes of all agents in the system. As we see below, this latter dependence is not an empty generality, but is rather a necessary consequence and feature of systems of interacting agents. Moreover, the agent attributes a_h are specified as vectors (not necessarily finite in dimension) of component attributes, possibly real-valued. This enables us to explain variation in responses between nodes with identical causal inputs as due to variation in distinct attribute components; it also permits us to form equivalence classes of agents defined in terms of shared attributes or subsets of attributes. Just as is true for the fundamental attribute of individuality, the general attributes of the settable system framework are not subject to counterfactual variation. Instead they act as non-causal response and effect modifiers.

Failing to recognize attributes and their unique properties can lead to explanatory fallacies and confusion. Faced with different responses from two nodes that have identical causal inputs, an attribute-blind observer might ascribe the difference to unobserved causes, i.e. "random effects." An observer not clear on the distinction between causes and attributes might ascribe causal properties to what should properly be recognized as attributes, e.g., "fixed effects" (an oxymoron so entrenched that it likely can never be eradicated).

The Nobel prize-winning work of Black and Scholes (1973) provides a useful illustration. Black and Scholes explain the price of an option to buy a given stock as the deductive consequence of premises involving the properties of the stock, the option, and a risk-free bond and the universal law of absence of arbitrage, that is, that there can be no riskless opportunities to make a positive profit. According to Black and Scholes (1973), the price p of a European option to buy a given stock (a "call" option) at a given price (the "exercise price") on a given date (the "expiration date"), is determined by the stock price (s), the volatility of stock returns (σ), the risk-free interest rate (r), the time remaining to expiration (t), and the exercise price (e):

$$p = f(s, \sigma, r, t, e).$$

The structural function f provides an explicit formula (the Black-Scholes option-pricing formula) that contains no quantities other than the listed arguments, and in particular, no unknown coefficients. The exogenous variables here are s and r , as these may vary counterfactually. The remaining quantities σ , t , and e are attributes. The exercise price e is clearly an attribute; options with different exercise prices are

different options. The time to expiration t is an attribute determined as the difference between the attribute of expiration date (options with different expiration dates are different options) and the current date (taken as given). Because the formula only gives the correct option price when volatility is a constant, the volatility of returns σ is necessarily an attribute. A much more elaborate and still developing theory (stochastic volatility option pricing) is required to describe option prices if volatility is allowed to vary. The volatility σ is an attribute of the stock, itself an attribute of the option.

Two complementary confusions arise in option-pricing theory and practice. The first is that in their theory Black and Scholes treat volatility as an attribute of the given stock, even though volatility is well known to vary. The second is that practitioners treat volatility in the Black-Scholes pricing formula as causal by ascribing effects to it, even though these are meaningless within the constant volatility theory. In particular, the derivative $\partial f/\partial\sigma$ (known as "vega") is commonly (mis)interpreted as the marginal effect of an increase in volatility.

The option-pricing example demonstrates that attributes are not just "fixed causes," that is, variables that could be subject to counterfactual variation but are for whatever reason held fixed. The fact that the option-pricing formula is valid only when volatility is constant eliminates any possibility of counterfactual variation for volatility within the Black-Scholes framework. The option-pricing example further illustrates that some number of precisely defined attributes may fully determine the identity of a response of interest, in the sense that nodes sharing these attributes belong to an equivalence class of nodes that have identical responses to all admissible causal input values.

5 Game Theory, Settable Systems, and the PCM

5.1 Games and Attributes

The importance of attributes in economics is underscored by the fact that deep interest attaches to situations involving interactions of agents in which attributes alone matter and the only exogenous variable is the "constant": the trivial exogenous variable always equal to one. A wealth of such cases arises in game theory, the study of multi-person decision problems. Game theory provides a rich formal framework in which to attempt to understand and explain the actions or behavior of interacting

agents. Gibbons (1992) is an excellent reference.

The simplest games are static games of complete information (Gibbons, 1992, ch.1), mentioned by WC (2006, p. 12). In these games, each of n players has: (i) a number of playable strategies (let player i have K_i playable strategies, $s_{i,1}, \dots, s_{i,K_i}$); and (ii) a utility (or "payoff") function u_i that describes the payoff π_i to that player when each player plays one of their given strategies. That is,

$$\pi_i = u_i(s_1, \dots, s_n),$$

where $s_j \in S_j \equiv \{s_{j,1}, \dots, s_{j,K_j}\}$, $j = 1, \dots, n$. The players simultaneously choose their strategies; then each receives the payoff specified by the collection of the jointly chosen strategies and the players' payoff functions. Such games are "static" because of the simultaneity of choice. They are "complete information" games because the players' possible strategies and payoff functions are known to all players. (This enables each player to assess the game from every other player's point of view.) An n -player static game of complete information is formally represented in "normal form" as $\mathcal{G} = \{S_1, \dots, S_n; u_1, \dots, u_n\}$.

These games map directly and explicitly to WC's settable system framework. Specifically, the players correspond to agents $h = 1, \dots, n$. The agent attributes a_h consist of the number of strategies K_h available to player h and the player's utility function u_h , represented by the vector containing the elements $u_h(s_1, \dots, s_n)$ as s_1, \dots, s_n ranges over all elements of $S_1 \times \dots \times S_n$. We write $a_h \equiv (K_h, u_h)$ and $a \equiv (a_1, \dots, a_n)$. When a strategy for player h is set arbitrarily, we denote its value as $z_h \in S_h$; when player h chooses a strategy (a response) we represent its value as $y_h \in S_h$. For concreteness and without loss of generality we take $S_h \equiv \{1, \dots, K_h\}$. We implicitly understand that strategy 1 for player h may represent a different action than that of strategy 1 for player i . The players' utility functions account for these differences.

Player choices are assumed to be rational: each player seeks to maximize their payoff given the strategies of the other players, so that

$$y_h = r_h(z_{(h)}, a), \quad \text{where}$$

$$r_h(z_{(h)}, a) = \arg \max_{z_h \in S_h} u_h(z_1, \dots, z_n).$$

For clarity, we assume for now that for each player there is a unique utility maximizing response, given the other players' strategies. We relax this assumption later.

In game theory, r_h is called a "best-response" function. WC call this a "response" function, in part motivated by this usage. Because in game theory the specific game \mathcal{G} under consideration is almost always clear, there is usually no need to explicitly reflect its elements in the players' best response functions. We reflect these elements here by including the players' attributes a (which characterize the game) to emphasize the role of attributes in determining player responses. Certainly, a player's own attributes $a_h \equiv (K_h, u_h)$ determine the response, as K_h determines S_h and u_h appears explicitly. We reference a and not just a_h , as the dimensionality of u_h depends on $K \equiv \{K_1, \dots, K_n\}$. The response for player h does not depend on $u_i, i \neq h$, but for simplicity we do not modify our representation to reflect this. So far, we have not required fundamental variables, Z_0 . These are present, although in a trivial sense: we can take Z_0 to be a scalar always taking the value unity ($Z_0 \equiv \mathbf{1}$).

Now consider the PCM representation. This requires the constant $u = 1$ to serve as the single exogenous variable, so $U = \{u\} = \{1\}$. Endogenous variables $V = \{s_1, \dots, s_n\}$ represent agent strategies, and the structural functions $F = \{f_1, \dots, f_n\}$ represent the best response for agent i as

$$s_i = f_i(s_{(i)}), \quad i = 1, \dots, n.$$

Although this captures essential features of the game, it also suppresses attribute dependence, which is central to understanding or explaining the outcome of the game. We shall shortly discuss this aspect of the PCM. First, however, we impose the final condition required for the PCM to apply, namely that there exists a unique fixed point defined by functions g_i such that

$$s_i = s_i^* \equiv g_i(u), \quad i = 1, \dots, n.$$

To preserve the form of the representation, we have written g_i as a function of the exogenous variable $u = 1$, but because u is the constant, $g_i(u) = g_i(1)$ is simply a constant, denoted s_i^* . When such a unique fixed point exists, it represents a *pure-strategy Nash equilibrium* (Nash, 1950), which by definition satisfies

$$u_i(s_1^*, \dots, s_i^*, \dots, s_n^*) \geq u_i(s_1^*, \dots, s_i, \dots, s_n^*)$$

for all $s_i \in S_i, i = 1, \dots, n$ (see Gibbons, 1992, p. 8).

The above PCM fixed-point representation of the unique pure-strategy Nash equilibrium also suppresses the attribute dependence of the equilibrium. To appreciate the

limitations of the PCM framework in this regard, consider the consequences of replacing player 2 in a two-person pure-strategy game with an alternate, say player 3. Thus, the original best-response functions $F = \{f_1, f_2\}$ are replaced with $\tilde{F} = \{f_1, f_3\}$. Assuming that the replacement preserves the uniqueness of the pure-strategy Nash equilibrium, we can ask within the PCM whether or not the new Nash equilibrium will be the same as the old one. That is, letting (s_1^*, s_2^*) denote the Nash equilibrium with players 1 and 2 and letting $(\tilde{s}_1^*, \tilde{s}_3^*)$ denote the Nash equilibrium with players 1 and 3, do we have $(s_1^*, s_2^*) = (\tilde{s}_1^*, \tilde{s}_3^*)$?

In the absence of knowledge of player attributes (and in particular of players 2 and 3), there is no way to provide a definite answer. On the other hand, the knowledge of attributes for players 1, 2, and 3 represented in the settable system framework permits us to specify whether the identical Nash equilibrium will be observed and why. That is, we can provide a complete explanation of what happens: If players 2 and 3 have identical attributes, the identical Nash equilibrium will be observed. If players 2 and 3 are not identical, different outcomes can easily arise. Nevertheless, it may happen that players 2 and 3 are not identical (suppose they have identical strategies but different known payoff functions), yet the outcome will be known to remain the same. This possibility can easily arise in the famous *Prisoners' Dilemma* game (e.g., Gibbons, 1992, pp. 2-5).

We emphasize that here we are not asking causal questions by inquiring about a counterfactual and therefore unobservable outcome, as when one asks what would have happened to a patient if they had not received a drug. Here we are asking about the observable outcome that arises when player 1 plays player 3 instead of player 2. That is, we are asking what we will see when we compare the outcome of connecting nodes 1 and 2 to the outcome of connecting nodes 1 and 3. This provides a further concrete example of the non-causal manner in which attributes explain (or, in the presence of causal variables, help explain) system outcomes.

5.2 Nash Equilibria and Partition-Specific Responses

Apart from the difficulties that arise from the lack of explicit attributes, an interesting aspect of the application of the PCM to models of game theory is that, just as in Section 3, the PCM faces difficulties arising from its requirement of a unique fixed point and its lack of flexibility in specifying the response functions for groups of jointly responding endogenous variables. As these difficulties contrast in interesting

ways with what happens in the analogous settable systems framework, we examine the possibilities in some detail.

A first difficulty for the PCM is that there are important games for which a pure-strategy Nash equilibrium does not exist; the PCM therefore has nothing to say about such games. A leading example of such games is known as *Matching Pennies* (Gibbons, 1992, p. 29). In this game, each of two players has a penny that they can choose to display face up (heads) or face down (tails). If the pennies match, player 2 gets both; otherwise player 1 gets both. This game has applicability in any situation in which one player would like to outguess the other, as in poker (bluffing), baseball (pitcher vs. hitter), and battle.

Given the interest attaching to such games, one would like to have an applicable causal model. This need is met by the settable system framework. Because this framework imposes no fixed point requirement, it applies regardless of the existence of a unique pure-strategy Nash equilibrium. For games with no pure strategy Nash equilibrium, the response functions $r_h(z_{(h)}, a)$ of the elementary partition $\Pi^e \equiv \{\{1\}, \dots, \{n\}\}$ readily provide complete information about the best response for all counterfactual strategy combinations of the other players. On the other hand, the fundamental partition $\Pi^f \equiv \{\{1, \dots, n\}\}$ need not yield a valid settable system for such games. This provides an interesting example in which we have a well-defined settable system for the elementary partition, but not for the fundamental partition. In contrast, the PCM does not apply at all.

When a unique pure-strategy Nash equilibrium exists, response functions for both the elementary and fundamental partitions are well defined. The fundamental partition represents the unique Nash equilibrium as

$$y_h = r_h^f(a), \quad h = 1, \dots, n,$$

for unique functions r_h^f obeying the fixed point requirement

$$y_h = r_h(r_{(h)}^f(a), a), \quad h = 1, \dots, n.$$

Another difficulty for the PCM is that the unique fixed point requirement prevents it from applying to games with multiple pure-strategy Nash equilibria. An example is the game known as *Battle of the Sexes* (Gibbons, 1992, p. 11). In this game, two players (Ralph and Alice) are trying to decide on what to do on their next night out: attend a boxing match or attend an opera. Each would rather spend the evening

together than apart, but Ralph prefers boxing and Alice prefers the opera. With the payoffs suitably arranged (symmetric), there is a unique best response for each player, given the strategy of the other. Nevertheless, this game has two pure-strategy Nash equilibria: (i) both select boxing; (ii) both select the opera. Thus, the PCM does not apply.

In contrast, the settable system framework does apply, as it does not impose a unique fixed point requirement. The elementary partition describes each agent's unique best response to a given strategy of the other. Further, when multiple Nash equilibria exist, the fundamental partition can yield a well-defined settable system by selecting one of the possible equilibria. As Gibbons (1992, p. 12) notes, "In some games with multiple Nash equilibria one equilibrium stands out as the compelling solution to the game," leading to the development of "conventions" that provide standard means for selecting a unique equilibrium from the available possibilities. An example is the classic *Coordination Game*, in which there are two pure-strategy Nash equilibria, but one yields greater payoffs to both players. The convention is to select the higher payoff equilibrium. Whenever such a convention exists, the fundamental partition can specify the response functions r_h^f to deliver this equilibrium. In such cases, the responses for the fundamental partition satisfy not only a fixed-point property, but also embody an equilibrium selection mechanism. Interestingly, *Battle of the Sexes* is not a game with such an outcome, as both equilibria appear equally compelling. A more elaborate version of this game, involving incomplete information, does possess a unique equilibrium, however (Gibbons, 1992, pp. 152-154).

The possibility of elaborating a game not only provides opportunities for modifying the character of the game's equilibrium set, but also yields further interesting insights into the contrasts between the PCM and settable systems. Specifically, consider "mixed-strategy" static games of complete information. Instead of optimally choosing a pure strategy, each player i now chooses a vector of probabilities $p_i \equiv (p_{i,1}, \dots, p_{i,K_i})$ (a mixed strategy) over their available pure strategies $S_i \equiv \{s_{i,1}, \dots, s_{i,K_i}\}$, so that p_{i,s_i} is the probability that player i plays strategy $s_i \in S_i$. For example, the probability vector $(1, 0, \dots, 0)$ for player i represents playing the pure strategy $s_{i,1}$. Each player i now behaves rationally by choosing their mixed-strategy probabilities to maximize their expected payoff,

$$\bar{\pi}_i = v_i(p^n) \equiv \sum_{s^n \in S^n} u_i(s^n) \Pr(s^n; p^n),$$

where for compactness we now write $s^n \equiv (s_1, \dots, s_n)$, $S^n \equiv S_1 \times \dots \times S_n$, and $p^n \equiv (p_1, \dots, p_n)$. The strategies are chosen independently, so that $\Pr(s^n; p^n)$, the probability that the agents jointly choose the configuration of strategies s^n , is given by

$$\Pr(s^n; p^n) = \prod_{j=1}^n p_{j,s_j}.$$

It is a famous theorem of Nash (1950) that if n is finite and if K_i is finite, $i = 1, \dots, n$, then there must exist at least one Nash equilibrium for \mathcal{G} , possibly involving mixed strategies (e.g., Gibbons, 1992, p. 45).

We map mixed-strategy games to WC's settable systems as follows. Agents again correspond to players, and attributes a_h for agent h are exactly as previously specified. Now, however, there is a vector of settings and responses for each agent. We denote the probabilities of the mixed strategy for agent h as $z_{h,j}, j = 1, \dots, K_h$, when these are set, and as $y_{h,j}, j = 1, \dots, K_h$, when these constitute the agent's best response. Let z_h be the vector with elements $z_{h,j}$, and let y_h be the vector with elements $y_{h,j}$. Given all other player's mixed strategies $z_{(h)}$, agent h 's rational strategy is

$$y_h = r_h(z_{(h)}, a), \quad \text{where}$$

$$r_h(z_{(h)}, a) = \sigma_h(\arg \max_{z_h \in \mathbf{S}_h} v_h(z_1, \dots, z_n)),$$

and the maximization is taken over the simplex $\mathbf{S}_h \equiv \{z_h \in [0, 1]^{K_h} : \sum_{j=1}^{K_h} z_{h,j} = 1\}$. Again we make explicit the dependence of the best response on a .

An interesting feature of this mixed-strategy game is that the set of maximizers $\arg \max_{z_h \in \mathbf{S}_h} v_h(z_1, \dots, z_n)$ can easily fail to have a unique element. This set thus defines the player's best-response *correspondence*, rather than simply giving a best-response function. We obtain a best-response function by applying a measurable selection operator σ_h to the set of maximizers. By definition, the agent is indifferent between elements of this set; the choice of selection operator is not crucial. In fact, the selection may be random, implemented by introducing a fundamental variable $z_{0,h}$ into the list of arguments of σ_h , so that one has

$$r_h(z_{(h)}, x_{01}, a) = \sigma_h(\arg \max_{z_h \in \mathbf{S}_h} v_h(z_1, \dots, z_n), z_{0,h}).$$

We view $z_{0,h}$ as the realization of a random variable generated by the stochastic structure of the settable system.

Observe that the response functions just defined are those for the agent partition $\Pi^a \equiv \{(1, 1), \dots, (1, K_1)\}, \dots, \{(n, 1), \dots, (n, K_n)\}$, which is the natural partition here, rather than the elementary partition $\Pi^e \equiv \{(h, j)\}, j = 1, \dots, K_h; h = 1, \dots, n\}$ previously examined.

Now consider how this game maps to the PCM. Again, the only exogenous variable is the constant, so $U = \{1\}$. The endogenous variables are most appropriately represented as vectors such that $V = \{p_1, \dots, p_n\}$. The elements of $F \equiv \{f_1, \dots, f_n\}$ are correspondingly vector-valued. These must satisfy

$$p_i = f_i(p_{(i)}) \equiv \sigma_i(\arg \max_{p_i \in \mathbf{S}_i} v_i(p_1, \dots, p_n)).$$

In order to apply the PCM, we must have a unique fixed point. Even when a unique Nash equilibrium exists, to obtain this as the fixed point requires choosing the selection operator σ_i so that it specifically produces the Nash equilibrium response. Any other selection will fail to produce the unique fixed point. In the usual situation, the properties of F determine whether or not a fixed point exists. Here, however, knowledge of the unique fixed point is required to properly specify σ_i , hence f_i , an awkward reversal signaling that the PCM is not well-suited to this application. In particular, the selection cannot be random, which is a plausible response when the player is indifferent between different possible strategies.

An interesting feature of this example is that when the PCM applies, it does so with vector-valued nodes rather than the scalar-valued nodes formally treated by Pearl (2000) or Halpern (2000). This means that the PCM is necessarily silent about what happens when components of an agent's strategy are arbitrarily set. In contrast, the setttable system readily applies to partitions both finer and coarser than the agent partition.

As before, the PCM suppresses attributes and therefore does not facilitate explanations. Unlike the case of pure-strategy games, there must always be at least one mixed-strategy Nash equilibrium, so the PCM does not run into the difficulty that there may be no equilibrium. Nevertheless, mixed-strategy games can also have multiple Nash equilibria, so the PCM does not apply there. For a given game, the GPCM does apply to the agent partition, but it does not incorporate equilibrium selection mechanisms. In contrast, the setttable system framework permits explanations at the level of the agent partition (as well as coarser or finer partitions); represents the unique Nash equilibrium at the level of the fundamental partition without requiring

a selection operator when a unique equilibrium exists; and otherwise represents the desired responses when a unique mixed-strategy Nash equilibrium does not exist but conventions or other plausible selection mechanisms apply.

Static games of complete information are the beginning of a sequence of increasingly richer games, including dynamic games of complete information, static games of incomplete information, and dynamic games of incomplete information. As Gibbons (1992, p. 173) notes, each of these games employs progressively stronger equilibrium concepts in order to rule out implausible equilibria in the richer games that would survive if one applied equilibrium concepts suitable for simpler games. Among other things, these implausible equilibria satisfy fixed-point (simple Nash equilibrium) requirements. The unique fixed point requirement of the PCM thus acts to severely limit its applicability across the game-theoretic spectrum, due to the many opportunities for multiple Nash equilibria. Although GPCMs formally apply, they cannot support discourse about causal relations between endogenous variables, due to the lack of an analog of the potential response function. In contrast, by taking advantage of partition-dependent response functions, the settable system framework permits implementations of whichever stronger and/or more refined equilibria criteria are natural for a given game, together with any natural equilibrium selection mechanism. Its explicit accommodation of attributes further supports more thorough explanations than are possible within the (G)PCM.

6 Countable vs. Finite Systems

So far, our discussion has been entirely in terms of finite systems, that is, systems in which the number of nodes, n , is finite. As we saw in Section 2, however, one of the differences between the PCM and the settable system framework is that whereas the PCM has a finite number of nodes, settable systems permit a countable infinity of nodes, admitting (among other things) a countable number of agents, a countable number of decisions for a given agent, or both. Here we address the issue of whether there is any benefit to the additional flexibility offered by working with a countable system.

As we now discuss, there are in fact substantial benefits. We illustrate this by discussing two examples: a dynamic game and a system involving learning, which can be either parametric or nonparametric. Because the behavior of interest in these

examples cannot be analyzed in finite systems, neither the PCM nor the generalized PCM applies. We see, however, that these examples fit comfortably into the settable system framework.

6.1 Infinitely Repeated Dynamic Games of Complete and Perfect Information

One class of games mentioned above is that of dynamic games of complete information. In these games, there is an element of sequential play. For example, two players can repeatedly play the *Prisoner's Dilemma* game. Game theorists have paid particular attention to infinitely repeated games, in which play can proceed indefinitely, as the equilibria emerging from such games generally exhibit rich and subtle behavior distinct from what may emerge in static games or finite horizon dynamic games. Clearly, infinite repetition cannot be handled in a finite system, so the PCM cannot apply.

In infinitely repeated dynamic games of complete and perfect information (see Gibbons, 1992, ch. 2.3.B), players play a given static game in "stages" or periods $t = 1, 2, \dots$, in which the period t payoff to player i is given by

$$\pi_{i,t} = u_{i,t}(\alpha_{1,t}, \dots, \alpha_{n,t}),$$

where $u_{i,t}$ is the i th player's payoff function for period t , taking as arguments the "actions" $\alpha_{j,t}$ at time t of all n players. (The strategies of the static games correspond to the actions of the dynamic games.) The game is "infinitely repeated," as it is played in a countable number of periods $t = 1, 2, \dots$. It is "dynamic," as the game is played sequentially, first in $t = 1$, then in $t = 2$, and so forth. Information is "complete," as every player knows every other player's possible actions and payoff functions. Information is "perfect," as at every period t , each player knows the entire history of play up to that period.

Players act rationally by maximizing their average present discounted value of payoff,

$$\bar{\pi}_i = \bar{u}_i(\alpha_1, \dots, \alpha_n) \equiv (1 - \delta)^{-1} \sum_{t=1}^{\infty} \delta^{t-1} u_{i,t}(\alpha_{1,t}, \dots, \alpha_{n,t}),$$

where $\alpha_j \equiv \{\alpha_{j,t}\}$ denotes player j 's countable sequence of actions, and $0 < \delta < 1$ is a "discount rate" (common across players, for simplicity) that converts payoffs $\pi_{i,t}$ in

period t to a value in period 1 as $\delta^{t-1}\pi_{i,t}$. (Because $\delta < 1$, a future payoff of a given amount is worth less than a payoff now of the same amount.)

A player's best response to any collective sequence of actions by the other players is therefore a solution to the problem

$$\begin{aligned} & \max_{s_i \in S_i} \bar{u}_i(\alpha_1, \dots, \alpha_n) \\ \text{s.t. } & \alpha_{i,t} = s_{i,t}(\alpha_1^{t-1}, \dots, \alpha_n^{t-1}), \quad t = 1, 2, 3, \dots \end{aligned}$$

Here, the maximization is performed over player i 's set S_i of all admissible sequences $s_i \equiv \{s_{i,t}\}$ of "strategy functions" $s_{i,t}$. These represent player i 's action in period t as a function only of the prior history of player actions, $\alpha_1^{t-1}, \dots, \alpha_n^{t-1}$. (When $t = 1$, we interpret $s_{i,t}$ as a constant function.) Player i 's best responses can therefore be written as

$$\alpha_{i,t}^* = s_{i,t}^*(\alpha_1^{t-1}, \dots, \alpha_i^{*t-1}, \dots, \alpha_n^{t-1}), \quad t = 1, 2, \dots,$$

where $s_i^* \equiv \{s_{i,t}^*\}$ represents a sequence of best response strategy functions. (These need not be unique, so $s_{i,t}^*$ may be a correspondence. In this case, however, the player is indifferent among the different possibilities.)

Due to the keen interest in such games, much is known of their properties. Specifically, it is known that such games generally have multiple Nash equilibria. Further, many of these are implausible, as they involve non-credible threats, and credibility is the central issue in all dynamic games (see Gibbons, p. 55). Such non-credible Nash equilibria can be removed from consideration by requiring Nash equilibria to be "sub-game perfect." In the present context, a sub-game perfect Nash equilibrium is one that solves not only the game beginning at time 1, but also the same game beginning at any time $t > 1$ (Gibbons, pp. 94-95). A celebrated result by James Friedman (1971) ensures the existence of one or more sub-game perfect Nash equilibria in such games, provided the discount factor is sufficiently close to one. (See, e.g., Gibbons, 1992, pp. 97-102.) A particularly interesting feature of such equilibria is that they permit the emergence of tacit cooperation, yielding outcomes for the players superior to what can be achieved in the static game played at each stage of the dynamic game.

Now consider how this game maps to the settable system framework. The agents are the n players, so h corresponds to i . The decisions governed by the agents are their actions in period t , so j corresponds to t . There are thus countably many decisions. The agent attributes are their admissible actions for each period and

their payoff functions for each period. That is, a_h corresponds to the sequences $\{K_{i,t}\}, \{u_{i,t}\}$ for given i , taking $K_{i,t}$ to represent the number of possible actions available to agent i in period t . These are countable sequences, so we fully exploit the ability of settable systems to handle countably dimensioned attribute vectors. Any random variables needed to implement random selections from correspondences appear in the (countably dimensioned) vector of fundamental variables Z_0 . When a player's actions $\alpha_{i,t}$ are set arbitrarily, the settable system represents them as $z_{h,j}$. When players choose their actions, they are denoted $y_{h,j}$. The agent partition $\Pi^a \equiv \{(1, 1), (1, 2), \dots\}, \dots, \{(n, 1), (n, 2), \dots\}$ delivers the agents' best responses recursively as

$$\begin{aligned} y_{h,j} &= \sigma_{h,j}(s_{h,j}^*(z_1^{j-1}, \dots, y_h^{j-1}, \dots, z_n^{j-1}), z_{0,h,j}) \\ &= r_{h,j}^a(z_{(h)}^{j-1}, z_{0,h}^j, a) \end{aligned} \quad j = 1, 2, \dots; h = 1, \dots, n,$$

where $\sigma_{h,j}$ is a measurable selection operator, $s_{h,j}^*$ is the agent's best response correspondence defined above, and $r_{h,j}^a$ gives the settable system representation of the agent-partition response function, showing its explicit dependence on the given action histories of other agents, $z_{(h)}^{j-1}$, the history of fundamental variables $z_{0,h}^j$ appearing in the measurable selections, and the attributes a of the agents determining the game.

The fundamental partition delivers a representation of whatever selection of the collection of subgame perfect Nash equilibria is natural or compelling for the given game. In this case, the agent responses are given by the fundamental-partition response functions $r_{h,j}^f$ as

$$y_{h,j} = r_{h,j}^f(z_0, a) \quad j = 1, 2, \dots; h = 1, \dots, n.$$

Observe that this example makes use of each feature of settable systems that distinguish it from the PCM: partition-specific response functions, attributes, and countably infinite dimension.

6.2 Learning in Settable Systems

A powerfully general learning algorithm introduced by Kushner and Clark (1978) (KC) has the form

$$\hat{\theta}_{t+1} = \hat{\theta}_t + a_t M_t(\hat{\xi}_t, \hat{\theta}_t, Z_{t+1}), \quad (9)$$

$$\hat{\xi}_{t+1} = R_t(\hat{\xi}^t, \hat{\theta}^{t+1}, Z_{t+1}), \quad t = 0, 1, 2, \dots, \quad (10)$$

where $\hat{\theta}_t$ and $\hat{\xi}_t$ are finitely dimensioned vector-valued random variables, a_t is a random scalar, M_t and R_t are known vector-valued functions, $\hat{\xi}^t \equiv (\hat{\xi}_0, \dots, \hat{\xi}_t)$, $\hat{\theta}^{t+1} \equiv (\hat{\theta}_0, \dots, \hat{\theta}_{t+1})$, and Z_t is a random vector. The initial values $\hat{\xi}_0$ and $\hat{\theta}_0$ are random vectors independent of $\{Z_t\}$. KC describe this as a Robbins-Monro (1951) (RM) algorithm with feedback (RMF). Equation (9) is an RM procedure; equation (10) supplies the feedback. A main focus of interest in such systems is the convergence behavior of $\hat{\theta}_t$ as $t \rightarrow \infty$.

Chen and White (1998) analyze a more general version of this algorithm in which each vector-valued object takes its values in a real separable infinite-dimensional Hilbert space. We call this version an HRMF algorithm. Because of the flexibility afforded by permitting the components of this system to be elements of a Hilbert space, the HRMF supports nonparametric learning. (Chen and White (1998) suppress Z_{t+1} in the argument list of M_t in equation (9). This is for notational convenience in their analysis; the representation above does not result in any loss of applicability.)

The pure RM procedure arises as the special case in which $\hat{\xi}_t$ is of dimension zero, in which case the recursions become

$$\hat{\theta}_{t+1} = \hat{\theta}_t + a_t M_t(\hat{\theta}_t, Z_{t+1}), \quad t = 0, 1, 2, \dots$$

This implements well-known recursive parameter estimation methods, such as recursive least squares, recursive maximum likelihood, and recursive method of moments (e.g., Ljung and Soderstrom, 1983). The estimated parameters are $\hat{\theta}_t$, $\{Z_t\}$ represents the sequence of data observations, typically $a_t = 1/t$, and the choice for M_t determines the parameter estimation method (least squares, maximum likelihood, etc.). By permitting feedback, the RMF accommodates the evolution of internal, possibly hidden states $\hat{\xi}_t$; thus, Kalman filter methods (Kalman, 1960) become a special case.

As another special case, the RMF contains learning algorithms for recurrent artificial neural networks (ANNs). (See for example, Elman, 1990; Jordan, 1992; Kuan, Hornik, and White, 1994). For this application, the input sequence is $\{Z_t\}$; after exposure to t input observations, network weights are elements of the vector $\hat{\theta}_t$, and hidden unit activations are elements of $\hat{\xi}_t$. The learning update function is M_t , the learning rate is a_t , and the hidden unit activations are determined by the function R_t .

The RMF and HRMF also contain as special cases systems with learning by optimizing and/or interacting agents. Chen and White (1998) provide conditions ensuring

that these systems converge as $t \rightarrow \infty$ to Nash equilibria or "rational expectations" equilibria. As examples, Chen and White (1998) consider, among others, a game known as *fictitious play with continuum strategies* (an infinitely repeated dynamic game of incomplete information), and an example in which a learning agent solves a stochastic dynamic programming problem. The applications of the RMF and HRMF are thus quite broad. As we now show, both the RMF and HRMF fall into the settable systems framework.

First, consider the RMF. In equations (9) and (10), we view the elements of $\hat{\theta}_t$ as representing agents' knowledge or beliefs at time t , as suggested by the parameter estimation and recurrent ANN examples. As before, we may view $\hat{\xi}_t$ as representing agents' actions or decisions at time t . (As the Kalman filter and ANN examples now suggest, however, these may also represent internal agent states.) We now have both knowledge and decisions explicitly appearing as agent-specific variables, as described in Section 2.1. Because t can take countably many values, a finite system will not suffice; however, the countably dimensioned settable system readily accommodates this. As $\hat{\theta}_t$ and $\hat{\xi}_t$ are finitely dimensioned vectors in the RMF, there must be a finite number of agents (say n), each with a finite number of time t knowledge elements (say κ_h for agent h) and a finite number of time t decisions (say d_h for agent h).

To explicitly reflect agent-specific knowledge we represent $\hat{\theta}_t$ as $\hat{\theta}_t = (\hat{\theta}'_{1,t}, \dots, \hat{\theta}'_{n,t})'$, where $\hat{\theta}_{h,t}$ is a $\kappa_h \times 1$ column vector, $h = 1, \dots, n$, and $\hat{\theta}'_{h,t}$ denotes the transpose of $\hat{\theta}_{h,t}$, the vector of knowledge elements for agent h at time t . Similarly, we write $\hat{\xi}_t = (\hat{\xi}'_{1,t}, \dots, \hat{\xi}'_{n,t})'$ to make explicit agent-specific decisions $\hat{\xi}_{h,t}$, where $\hat{\xi}_{h,t}$ is a $d_h \times 1$ column vector. The elements of $M_t = (M'_{1,t}, \dots, M'_{n,t})'$ are agent-specific learning update functions, and the elements of $R_t = (R'_{1,t}, \dots, R'_{n,t})'$ are agent-specific decision functions. For given h , we view the sequences $\{M_{h,t}\}$ and $\{R_{h,t}\}$ as attributes a_h of agent h .

In general, we may expect that $\hat{\theta}_{h,t+1}$ will depend on the immediately prior actions of all agents, $\hat{\xi}_t$, but will depend directly only on that agent's own previous knowledge, $\hat{\theta}_{h,t}$, so that $\hat{\theta}_{h,t+1} = \hat{\theta}_{h,t} + a_t M_{h,t}(\hat{\xi}_t, \hat{\theta}_{h,t}, Z_{h,t+1})$. Here we also permit the possibility that agent h may not observe the entirety of Z_{t+1} – that is, there may be agent-specific information – by writing $Z_{h,t+1}$ rather than Z_{t+1} in $M_{h,t}$. Similarly, we may expect that $\hat{\xi}_{h,t+1}$ will depend generally on the entire history of all agent actions (as in the infinitely repeated game of the previous subsection), but will depend directly only on the history of the agent's own knowledge, $\hat{\theta}_h^{t+1}$, and the information $Z_{h,t+1}$ observed

by that agent, so that $\hat{\xi}_{h,t+1} = R_{h,t}(\hat{\xi}^t, \hat{\theta}_h^{t+1}, Z_{h,t+1})$.

Equations (9) and (10) form a recursive system. As WC discuss, the responses and settings coincide in this case. Thus $z_{h,j} = y_{h,j}$ represents both, corresponding to an element of either $\hat{\theta}_t$ or $\hat{\xi}_t$. Fundamental variables here are $\hat{\xi}_0, \hat{\theta}_0, \{a_t\}$, and $\{Z_t\}$. Equations (9) and (10) provide the response functions for the partition separating knowledge and actions. Recursive substitution of equation (10) into equation (9) provides the response functions for the "time partition," $\Pi^t \equiv \{(1, 1), \dots, (\kappa + d, 1)\}, \{(1, 2), \dots, (\kappa + d, 2)\}, \dots\}$, where we assume that $\hat{\theta}_t$ and $\hat{\xi}_t$ contain a total of $\kappa + d$ elements. As we have now mapped each element of the RMF to its corresponding element of the settable system, we have established that the settable system framework contains the RMF.

Finally, consider how the settable system framework encompasses the HRMF. The distinguishing feature of the HRMF is that agent knowledge and actions are elements of real separable infinite-dimensional Hilbert spaces. Specifically, knowledge and actions may be suitably well-behaved functions. First, consider how the countably infinite dimension of the settable system framework accommodates such objects in the simple situation in which there is a single agent with a single action, a function, and a single knowledge element, also a function. The key idea is to view such functions as represented by a vector of countably infinite dimension whose elements are the coefficients of the terms in a suitable series representation of the function. The Hilbert space structure ensures a variety of workable representations of this sort. For example, a large class of well-behaved functions can be represented in this way by the coefficients of a Fourier series expansion for the function. Anywhere such a function appears in the system, we can represent it within the settable system using this device. Further, we can apply the same approach without exhausting the dimensionality of the settable system even when there is a countable infinity of agents, each of whom has a countable infinity of knowledge elements and actions, which are themselves elements of real separable infinite-dimensional Hilbert spaces.

7 Concluding Remarks

The substantial advances of Pearl and his colleagues, founded in the PCM, constitute a rich and widely applicable body of knowledge providing deep insight into the nature of causal relations. The question addressed here is whether the settable system

framework has anything to add to the PCM.

We address this question by first showing that settable systems encompass the PCM; specifically, the PCM is a settable system with a finite number of agent–decision pairs, no explicit attributes, and with a fixed point restriction imposed. Halpern’s (2000) generalized PCM, which omits the fixed point requirement, is similarly encompassed. Important distinguishing features of settable systems are its countable dimensionality, the presence of attributes, the absence of a fixed point requirement, and the use of partitioning and partition-specific response functions to accommodate the behavior of optimizing and interacting agents. A series of examples from microeconomics, option pricing, game theory, and learning theory demonstrates the utility of each of these features, alone and in combination.

A final benefit of the settable system framework is the natural way in which it yields a meaningful formal definition of causal variables. Specifically, in defining settable systems, WC (Def. 2.1) define "settable variables" as mappings

$$\mathcal{X}_{h,j} : \{0, 1\} \times \Omega \rightarrow \mathbb{R},$$

where (Ω, \mathcal{F}) is the underlying measurable space on which settings $Z_{h,j} : \Omega \rightarrow \mathbb{R}$ and the probability measure $P : \mathcal{F} \rightarrow [0, 1]$ are defined, such that

$$\begin{aligned} \mathcal{X}_{h,j}(1, \cdot) &= Z_{h,j} \\ \mathcal{X}_{h,j}(0, \cdot) &= r_{h,j}(Z_{(h,j)}, a), \quad j = 1, 2, \dots; h = 1, 2, \dots, \end{aligned}$$

with the convention that $\mathcal{X}_{0,j}(0, \cdot) = \mathcal{X}_{0,j}(1, \cdot) = Z_{0,j}$, $j = 1, 2, \dots$. As WC discuss, it is then convenient and natural to define causal relations for a given settable system as relations holding between two settable variables, say $\mathcal{X}_{h,j}$ and $\mathcal{X}_{i,k}$. For example, if the function $z_{i,k} \rightarrow r_{h,j}(z_{(h,j)}, a)$ is constant for all values of the elements of $z_{(h,j)}$ other than $z_{i,k}$, then WC say that relative to this settable system, $\mathcal{X}_{i,k}$ does not cause $\mathcal{X}_{h,j}$; otherwise $\mathcal{X}_{i,k}$ causes $\mathcal{X}_{h,j}$. (This corresponds to the notion of absence or presence of direct cause in the PCM.) With this approach, it is unnecessary to attempt to define causal relations in terms of events (elements of \mathcal{F}) or random variables (\mathcal{F} –measurable functions on Ω), which are by themselves inadequate to carry this meaning. Instead, settable variables serve as a distinct category of causal variables, naturally constructed for defining causal relations and thus for carrying causal meaning.

APPENDIX: SETTABLE SYSTEMS

For ease of reference we provide the following definition of WC's attribute-indexed partitioned settable systems:

Definition 1 Attribute-Indexed Partitioned Settable Systems *Let \mathcal{A} be a countably dimensioned Borel set. For **agents** $h = 1, 2, \dots$, let **attributes** a_h belong to \mathcal{A} , and write $a \equiv \{a_h\} \in \mathcal{A}^\infty \equiv \mathcal{A} \times \mathcal{A} \times \dots$. Let $\Pi \equiv \{\Pi_b\}$ be a partition of the ordered pairs $\{(h, j) : j = 1, 2, \dots; h = 1, 2, \dots\}$. For $b = 1, 2, \dots$, let $\ell_b \equiv \#\Pi_b$ (the cardinality of Π_b) and $\ell_{(b)} \equiv \sum_{a \neq b} \#\Pi_a$. For any double array of real scalars $\{z_{h,j}\}$, let $z_{(b)} \equiv \{z_{i,k} : (i, k) \notin \Pi_b\}$, and suppose for $b = 1, 2, \dots$, there exist measurable **response functions** $r_b^\Pi : \mathbb{R}^{\ell_b} \times \mathbb{R}^{\ell_{(b)}} \times \mathcal{A}^\infty \rightarrow \mathbb{R}^{\ell_b}$ such that for every $z_{(b)} \in \mathbb{R}^{\ell_{(b)}}$ there exists a unique $y_b \equiv \{y_{h,j} : (h, j) \in \Pi_b\} \in \mathbb{R}^{\ell_b}$ such that*

$$r_b^\Pi(y_b, z_{(b)}, a) = \mathbf{0}.$$

*Let (Ω, \mathcal{F}, P) be a measurable space, and for $h = 0, 1, \dots$, and $j = 1, 2, \dots$, let **settings** $Z_{h,j} : \Omega \rightarrow \mathbb{R}$ be measurable functions, and for $h = 1, 2, \dots$, and $j = 1, 2, \dots$, let **responses** $Y_{h,j} : \Omega \rightarrow \mathbb{R}$ be such that*

$$r_b^\Pi(Y_b, Z_{(b)}, a) = \mathbf{0}, \quad b = 1, 2, \dots .$$

*For $h = 0$, let $Y_{0,j} \equiv Z_{0,j}, j = 1, 2, \dots$. For $h = 0, 1, \dots$, and $j = 1, 2, \dots$, define the **settable variables** $\mathcal{X}_{h,j}^\Pi : \{0, 1\} \times \Omega \rightarrow \mathbb{R}$ as*

$$\begin{aligned} \mathcal{X}_{h,j}^\Pi(1, \cdot) &= Z_{h,j} \\ \mathcal{X}_{h,j}^\Pi(0, \cdot) &= Y_{h,j} \end{aligned}$$

so that

$$r_b^\Pi(\mathcal{X}_b^\Pi(0, \cdot), \mathcal{X}_{(b)}^\Pi(1, \cdot), a) = \mathbf{0}, \quad b = 1, 2, \dots .$$

*Let $r^\Pi \equiv \{r_{h,j}^\Pi\}$ and $\mathcal{X}^\Pi \equiv \{\mathcal{X}_{h,j}^\Pi\}$. Then $\mathcal{S} \equiv \{(\Omega, F, P), (A, a, \Pi, r^\Pi, \mathcal{X}^\Pi)\}$ is an **attribute-indexed partitioned settable system**.*

For simplicity, we take the domain of r_b^Π to be $\mathbb{R}^{\ell_b} \times \mathbb{R}^{\ell_{(b)}} \times \mathcal{A}^\infty$. With additional technicality, we can let the domain be $D_{y_b} \times D_{z_{(b)}} \times \mathcal{A}^\infty$, where $D_{y_b} \subset \mathbb{R}^{\ell_b}$ and $D_{z_{(b)}} \subset \mathbb{R}^{\ell_{(b)}}$. Above, the responses Y_b are specified by the implicit equations $r_b^\Pi(Y_b, Z_{(b)}, a) = \mathbf{0}$. In WC and the text, the responses are represented in explicit form

as $Y_b = r_b^\Pi(Z_{(b)}, a)$. The former is more general, as not all response relations have an explicit representation. It is also directly relevant for handling optimizing agents, as the implicit form corresponds to the first order conditions for an optimum. In the text, we use the explicit representation for the sake of convenience and familiarity. The settable system discussed in Section 2.1 is that of the elementary partition, $\Pi^e \equiv \{\Pi_{h,j}\}$, $\Pi_{h,j} \equiv \{(h, j)\}$.

References

Black, F. and Scholes, M. (1973), "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637-654.

Chen, X. and White, H. (1998), "Nonparametric Adaptive Learning with Feedback," *Journal of Economic Theory*, 82, 190-222.

Elman, J. (1990), "Finding Structure in Time," *Cognitive Science*, 14, 179-211.

Fisher, F. (1970), "A Correspondence Principle for Simultaneous Equations," *Econometrica*, 38, 73-92.

Friedman, J. (1971), "A Noncooperative Equilibrium for Supergames," *Review of Economic Studies*, 38, 1-12.

Gibbons, R. (1992). *Game Theory for Applied Economists*. Princeton: Princeton University Press.

Haavelmo, T. (1943), "The Statistical Implications of a System of Simultaneous Equations," *Econometrica*, 11, 1-12.

Halpern, J. (2000), "Axiomatizing Causal Reasoning," *Journal of Artificial Intelligence Research*, 12, 317-337.

Hempel, C. and Oppenheim, P. (1948), "Studies in the Logic of Explanation," *Philosophy of Science*, 15, 135-175.

Holland, P. (1986), "Statistics and Causal Inference," (with Discussion), *Journal of the American Statistical Association*, 81, 945-970.

Jordan, M. (1992), "Constrained Supervised Learning," *Journal of Mathematical Psychology*, 36, 396-425.

Kalman, R. (1960), "A New Approach to Linear Filtering and Prediction Problems," Trans. ASME, Series D, *Journal of Basic Engineering*, 82, 35 - 45

Kuan, C.-M., Hornik, K., and White, H. (1994), "A Convergence Result for Learning in Recurrent Neural Networks," *Neural Networks*, 6, 420-440.

Kushner, H. and Clark, D. (1978). *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Berlin: Springer-Verlag.

Ljung, L. and Soderstrom, T. (1983). *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT Press.

Nagel, E. (1961). *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt, Brace & World

Nash, J. (1950), "Equilibrium Points in n -Person Games," *Proceedings of the National Academy of Sciences*, 36, 48-49.

Nicholson, W. (2005). *Microeconomic Theory: Basic Principles and Extensions* (Ninth Edition). Mason, OH: Thomson South-Western.

Pearl, J. (2000). *Causality*. New York: Cambridge University Press.

Popper, K. (1959). *Logic of Scientific Discovery*. London: Hutchinson.

Robbins, H. and Monro, S. (1951), "A Stochastic Approximation Method," *Annals of Mathematical Statistics*, 22, 400-407.

Rubin, D. (1974), "Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies," *Journal of Educational Psychology*, 66, 688-701.

Strotz, R. and Wold, H. (1960), "Recursive vs. Nonrecursive Systems: An Attempt at Synthesis," *Econometrica*, 28, 417-427.

White, H. and Chalak, K. (2006), "A Unified Framework for Defining and Identifying Causal Effects," UCSD Department of Economics Working Paper.