Studies in History and Philosophy of Biological and Biomedical Sciences xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

journal homepage: www.elsevier.com/locate/shpsc

Causal structure and hierarchies of models

Kevin D. Hoover *

Department of Economics, Duke University, Durham, NC, USA
Department of Philosophy, Duke University, Durham, NC, USA

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Causal structure
Identity
Independence
Modularity
James Woodward
Herbert Simon

ABSTRACT

Economics prefers complete explanations: general over partial equilibrium, microfoundational over aggregate. Similarly, probabilistic accounts of causation frequently prefer greater detail to less as in typical resolutions of Simpson's paradox. Strategies of causal refinement equally aim to distinguish direct from indirect causes. Yet, there are countervailing practices in economics. Representative-agent models aim to capture economic motivation but not to reduce the level of aggregation. Small structural vector-autoregression and dynamic stochastic general-equilibrium models are practically preferred to larger ones. The distinction between exogenous and endogenous variables suggests partitioning the world into distinct subsystems. The tension in these practices is addressed within a structural account of causation inspired by the work of Herbert Simon's, which defines cause with reference to *complete* systems adapted to deal with incomplete systems and piecemeal evidence. The focus is on understanding the constraints that a structural account of causation places on the freedom to model complex or lower-order systems as simpler or higher-order systems and on to what degree piecemeal evidence can be incorporated into a structural account.

1. Models, piecemeal knowledge, and aggregation

The history of science is a history of piecemeal additions to human understanding of the natural and social world. Yet the image of science portrayed not only in philosophy but by scientists themselves is frequently totalizing. The aim of science is to discover universal laws; but have we ever actually discovered such laws, and, if we had, would we know that we had? The world is ruled by an exceptionless determinism; but when have we ever seen a rule that held invariably? Even those who acknowledge that we have always fallen short of the whole truth frequently hold a "perfect-model model" (Teller, 2001) as the standard to which we ought to aspire.

Economics is by no means exempt. The Laplacian fantasy of a clockwork world in which a statement of initial conditions and the laws of nature would allow us to write the history of the future has an analogue in the Walrasian fantasy of an agent-by-agent description of the economy:

> Thus the system of the economic universe reveals itself, at last, in all its grandeur and complexity: a system at once vast and simple, which, for sheer beauty, resembles the astronomic universe. (Walras, 1954, p. 374)

Neither Laplace's nor Walras's vision is practical; yet they each guide the imagination in a manner that shapes scientific practice.

The alternative to the perfect-model model is a vision of scientific practice that embraces the piecemeal acquisition of knowledge (Wimsatt, 2007). Although knowledge was actually acquired piece by piece, economics—and perhaps other sciences—has reflected relatively little on what an effective methodology of the piecemeal acquisition of knowledge would look like. This is not to say that there have not been efforts to formulate such a methodology. Milton Friedman (1949) attempted to revive Alfred Marshall's (1885) methodology (see Hoover, 2006, 2009a), rejecting the notion that the contrast between Marshall and Walras was one between *partial* and *general* equilibrium rather than one between an approach in which empirical knowledge is built up

* Address: Department of Economics, Duke University, Box 90097, Durham, NC 27278, USA.
E-mail address: kd.hoover@duke.edu

1369-8486/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.shpsc.2012.05.007

brick by brick and one that demands an architectonic theoretical account as a starting point.

The grip of the Walrasian approach is clear in the macroeconomic modeling tradition that started with Jan Tinbergen (1939) and was largely guided by Lawrence Klein (1950; Duesenberry, Fromm, Klein, & Kuh 1965; Klein & Goldberger 1955). Looking back, Klein describes his modeling approach:

> In contrast with the parsimonious view of natural simplicity, I believe that economic life is enormously complicated and that the successful model will try to build in as much of the complicated interrelationships as possible. That is why I want to work with large econometric models and a great deal of computer power. Instead of the rule of parsimony, I prefer the following rule: the largest possible system that can be managed and that can explain the main economic magnitudes as well as the parsimonious system is the better system to develop and use. (Klein, 1992)

It was a view that Friedman had criticized when Klein's program was still in the cradle. He suggested a Marshallian alternative: "the focus should be on the analysis of parts of the economy in the hope that we can find bits of order here and there and gradually combine these bits into a systematic picture of the whole" (Friedman, 1951, p. 114). And he summed this view up in his famous methodological essay:

> A hypothesis is important if it 'explains" much by little, that is, if it abstracts the common and crucial elements from the mass of complex and detailed circumstances surrounding the phenomena to be explained and permits valid predictions on the basis of them alone. (Friedman, 1953, p. 14)

As Friedman himself acknowledged, Klein's argument carried the day: "We curtsy to Marshall, but we walk with Walras" (Friedman, 1949, p. 489).

Modern macroeconomics has rejected Klein's modeling program, though not for any lack of sympathy with the ideal of the perfect model. Rather they have rejected it, first, in the belief that it compromises on fundamental and essential economic principles—namely that economics is only economics if it is fundamentally grounded in individual agents described as making choices under constraints. This is the essentially false charge that the Klein's macroeconomics—in line with most of the macroeconomics of the 1930s through 1960s—lacked *microfoundations* (see Hoover, 2012a). Second, they rejected it not for adopting the Walrasian ideal but for failing to be thoroughly enough committed to that ideal that the "deep parameters"—that is, the parameters that reflect the preferences and constraints of the individual agents—could be *identified* empirically (Hoover, 1988; Lucas, 1976, ch. 8, section 3). In practice, the second concern dominated the first. After 1970, macroeconomics worked mainly with smaller models, but the ideal remained to model the individual agent, so that increasingly complex models beckoned from an ever receding horizon of technical possibility.

All parties to these debates agree that economics is a science of models. Through most of this period—although it is changing recently—economists were deeply uncomfortable with talk of causes, even though most of their principal interests, such prediction and control, are thoroughly causal (Hoover, 2004). Economics, then, should be seen not only as a science of models but as a science of causes.

The history of postwar macroeconomics just sketched suggests that to understand economics as a science of models we must understand the relationships among models of different levels of complexity. Roughly speaking, we must understand the relationship between aggregated and disaggregated models of the same phenomena. Aggregation in economics has been traditionally viewed through a reductionist lens—a question related to the microfoundations of macroeconomics. This is indeed an important issue (Hoover, 2001a, 2001b, ch. 5, 2009b, 2010b). Yet it does not exhaust the questions that; arise with respect to hierarchies of models in economics. Consider three common issues:

The relationship of larger to smaller models of the same phenomena on the same level of aggregation. For example, empirical macroeconomics frequently employs vector autoregressions (VARs) that attempt to model the interrelated dynamics of sets of variables. One investigator interested in the Federal Reserve's monetary policy might employ a three-variable VAR with, say, GDP, inflation, and the Federal funds rate. Another investigator addressing the same issue might instead employ a five-variable VAR using consumption, investment, and government expenditure (accounting components of GDP) and a long-term government bond rate, as well as the Federal funds rate. How do empirical results gathered with one such model relate to those of the other?

The relationship of models at different levels of temporal aggregation. The VAR analysts using GDP in a three-variable model typically use a quarterly time unit, since that is the way that the GDP data are published. But if industrial production, which theoretically should, and practically does, track GDP is used instead, then monthly data is available.

The relationship of dynamic to static models. Empirical economic data stand in complex intertemporal relationships. The most widely accepted theoretical economic models refer principally to static (or steady-state) equilibria. How then are static theoretical models to be related to dynamic empirical models?

Traditionally, the problem of aggregation was posed in two ways: First, when does the aggregate behave just like its constituent units? Or, second, under what circumstances are the behaviors of the aggregate and its constituents consistent? (See Janssen (1993) for a discussion of approaches to aggregation.) I want to take a different approach here, one that shifts the focus from what might be seen as accounting relationships to one that considers different degrees of causal articulation.

2. Perspectival realism and causation

Philosopher most often—though not exclusively—analyze cause as a relationship among token events or facts in a background of scientific laws (see, for example, Lewis, 1973). Such an approach is not suitable to a piecemeal approach to scientific knowledge as it is effectively question-begging—relying on universal laws, which stand at the end of inquiry. Some philosophers have turned this common view on its head: causal relationships precedes even conjectured universal laws, which may be regarded as abstractions from causal knowledge. For example, the ideal gas laws can be seen as a functional relationship that is common to various apparatuses in some of which temperature causes pressure and others of which pressure causes volume or temperature, and so forth (Hoover, 2001b, pp. 81–87).

Causal relationships on this view share the generality of scientific laws and are more naturally expressed as type relationships among variables rather than token relationships among events or facts. The variables that characterize such relationships are necessarily always less than fully descriptive of the concrete situations in which the causal relationships are embedded. Even selecting a set of variables to describe a concrete relationship must, therefore, impose a perspective and constrains what causal relationships are possible within a model built on those variables.

Please cite this article in press as: Hoover, K. D. Causal structure and hierarchies of models. *Studies in History and Philosophy of Biological and Biomedical Sciences* (2012), http://dx.doi.org/10.1016/j.shpsc.2012.05.007

The familiar notion of cause is, as reflected in the title of Woodward's (2003) book, that causes make things happen. Cause in our everyday experience is typically about control. But we must be careful. First, there is no need to anthropomorphize: by control we generally do not require human manipulation or intervention, although that may well be the origin of our causal concepts. We have no trouble with the idea that a lightening strike caused a fire, and have not for millennia believed that a sentient agent must lie behind each particular natural phenomenon. Second, it is not particular token interventions that are essential. The issue is less making any particular thing happen than in identifying how things could happen. In contrast to Woodward (2003, p. 39), the issue is not particular interventions, but rather what is the scope and topology of interventions. Again, in contrast to some philosophers (e.g., Lewis, 1973) causal knowledge supports counterfactual analysis, not the other way round (Woodward, 2003, p. 16, ch. 6; Hoover, 2011).

Our approach is broadly realist in the sense that it is predicated on the belief that models are used to assert true general claims about causal relationships, the truth-status of which is determined by the world external to our mental constructions—that is, external to our models. Such a realism is compatible with models viewing the world from different perspectives (Giere, 2006; Hoover, 2012b). The truth that we seek is what the world is actually like *when seen this way*.

Such a perspectival causal realism has implications for aggregation: causal claims may appear to differ in different models and may appear to conflict if we fail to notice the particular perspective of the model. Which causal relationships we see depend on which model we use and its conceptual/causal articulation; which model is best depends on our purposes and pragmatic interests.

Take the case of Simpson's paradox, which can be described as the situation in which conditional probabilities (often related to causal relations) are opposite for subpopulations than for the whole population. Let academic salaries be higher for economists than for sociologists, and let salaries within each group be higher for women than for men. But let there be twice as many men as women in economics and twice as many women as men in sociology. By construction, the average salary of women is higher than that for men in each group; yet, for the right values of the different salaries, women are paid less on average, taking both groups together.

An aggregate model leads to the conclusion that that being female causes a lower salary.[1] We might feel an uneasiness with such a model, since I have already filled in the details that show more precisely why the result comes about. The temptation is to say that the aggregate model shows that being female *apparently* causes lower salaries; but the more refined description of a disaggregated model shows that *really* being female causes higher salaries. A true paradox, however, is not a contradiction, but a seeming contradiction. Another way to look at it is to say that the aggregate model is really true at that level of aggregation and is useful for policy and that equally true more disaggregated model gives an explanation of the mechanism behind the true aggregate model.

It is not wrong to take an aggregate perspective and to say that being female causes a lower salary. We may not have access to the refined description. Even if we do, we may as matter of policy think (a) that the choice of field is not susceptible to useful policy intervention, and (b) that our goal is to equalize income by sex and not to enforce equality of rates of pay. That we may not believe the factual claim of (a) nor subscribe to the normative end of (b) is immaterial. The point is that that they mark out a perspective in which the aggregate model suits both our purposes and the facts: it tells the truth as seen from a particular perspective.

3. Modeling causal structure

If models give us a perspective on the causal structure of the world, the next question is surely what property is it of models that captures causal relationships. An influential approach to this question is given by Herbert Simon in various papers, starting with "Causal Order and Identifiability" in 1953. Simon considers causal relationships in complete systems of equations. I have previously presented a formal generalization of Simon's approach (Hoover, 2001b, ch. 3), and I will not present the formalism, but instead exposit the key points with examples. It is worth noting that, while I am prepared to defend my reading of Simon against critics who interpret him differently, it is the approach itself rather than its pedigree that concerns us here (cf. Cartwright, 2007, chs. 13, 14; Fennell, 2005, ch. 4; Hoover, 2010a, 2011, 2012c).

Simon started with a *self-contained structure*—that is, a system of equations in which variables have a unique solution conditional on the values of parameters and the particular functional forms.[2] While his illustrations are all linear in variables, there is nothing in the general approach that prevents the systems of equations from being nonlinear. Furthermore, variables may be continuous or discrete. He then focused on *self-contained subsets* of the self-contained structure.

To illustrate, consider the following system of equations (S1):

Structure S1

$$A = \alpha_A, \tag{1}$$

$$B = \alpha_{BA}A, \tag{2}$$

$$C = \alpha_{CA}A + \alpha_{CB}B, \tag{3}$$

where, for the moment, we regard the α_{ij} as fixed coefficients.

System S1 is a complete system, and a complete system is itself self-contained. System S1 is not, however, a *minimal self-contained subset*, as it contains subsystems, which are themselves self-contained. Equation (1) is minimal a self-contained subset: it determines the value of *A* without reference to any other equation and it contains no subsystems that are similarly self-contained. Equations (2) and (3) considered separately are not self-contained subsets as they do not contain enough information to determine *B* or *C*. In contrast, equations (1) and (2) together form a self-contained subset, since they determine the values of *A* and *B* without reference to Equation (3)—though the subset is not minimal, since it contains the subsystem of Equation (1), which is self-contained.

Simon's conception is closely related to his later work on hierarchies of systems (Simon, 1996, chs. 1 and 8). Causes are the outputs of lower-level systems and the inputs to higher-level systems. The relationship is closely connected to the solution algorithms for systems of equations. In S1, *A* is determined entirely by (1) and can be regarded as an output. If we know *A*, we do not need to know (1) to determine *B*; a specific value for *A* forms an input that, in effect, turns the non-self-contained subsystem (2) into a self-contained subset in which the value of *A* is given parametrically. Its output is, of course, *B*. Knowing *B* alone, however, does not turn (3) into a self-contained subset. Substituting its value into (3) leaves the variable *A* in place (despite the fact that *B* cannot have a well-defined value unless *A* also has a well-defined value) and we have to substitute *A* directly from (1). Thus, *A* directly causes *B*, and *A* and *B* directly cause *C*; so, *A* is both a direct and an indirect cause of *C*. This, of course, is the causal structure of Fig. 1, where the arrows indicate the relationship of *direct causation*.

[1] Woodward (2003), p. 113, argues that essential characteristics, such as species, sex, or race, are not properly causes. For a counterargument, see Hoover (2012c).

[2] A *linear structure* in Simon's (1953) terminology.

Please cite this article in press as: Hoover, K. D. Causal structure and hierarchies of models. *Studies in History and Philosophy of Biological and Biomedical Sciences* (2012), http://dx.doi.org/10.1016/j.shpsc.2012.05.007

While the solution to a system of equations is specific, the causal relationship for Simon is generic. Specific systems differ only in taking different admissible values of parameters. The set of such specific systems is a *model*.[3] Causal structure is a property of models that is instantiated in specific systems. Implicit in Simon's account, though never explicitly articulated, is that parameters are *variation-free*—that is, the values of different parameters are mutually unconstrained. To put it anthropomorphically, the value of any parameter can be changed without necessarily changing the values of any other parameter. I will argue presently that this is, in fact, a useful formal requirement but one that in no way constrains what causal relations a model can represent.

The distinction between *parameters* and *variables* is a key one. Causal relationships for Simon exist only among variables. Parameters are not causes, but indicators of the scope of possible interventions from outside the system. There is a temptation—partly driven by a desire to make Simon's formal language conform to other usages—to regard "parameter" as a synonym for "exogenous variable" and, therefore, in diagrams to draw causal arrows from parameters to other parameters and variables (e.g., Cartwright, 2007, ch. 14). But this is not Simon's usage. Simon models an exogenous variable, for example, as $X = \alpha_X$, where X is a variable and α_X a parameter. Interventions are mediated only through parameters, so that an exogenous change in X is modeled as a change in α_X. It does little harm to treat exogenous variables and their governing parameters as synonyms, so long as the parameters are uniquely assigned to the exogenous variable; but, as we shall see, this is not the general case; so, it is better to nip that practice in the bud.

The analysis of the causal structure of a complete structure proceeds by articulating the hierarchy of self-contained subsets in the manner illustrated with structure S1. It can be proved that the causal order of a complete structure is unique (Iwasaki & Simon, 1994, p. 156). The fly in the ointment is that uniqueness depends on privileging the manner in which the equations are written. But the same functional relations can be written in a variety of ways. For example, the self-contained subset (1) and (2) could be replaced by S2:

Structure S2

$$A = \beta_A + \beta_{AB}B, \tag{4}$$

$$B = \beta_B, \tag{5}$$

which has the same numerical solution as (1) and (2) provided that:

$$\beta_A = \alpha_A/(1 - \alpha_{BA}), \tag{6}$$

$$\beta_{AB} = -1/(1 - \alpha_{BA}), \tag{7}$$

and

$$\beta_B = \alpha_A\alpha_{BA} \tag{8}$$

(Nothing depends on the fact that the β_{ij} are defined in terms of the α_{ij}. We could as easily have started with equations (4) and (5) and derived an analogous set of restrictions defining the α_{ij} in terms of the β_{ij} to guarantee identical solutions.) The two sets of equations have the same solution, but under Simon's analysis B causes A in (4) and (5), whereas A causes B in (1) and (2). Indeed, since every linear combination of equations (1) and (2) is functionally equivalent, we can easily write down systems that would be interpreted as having no causal connections or as displaying mutual causation. The different functionally equivalent sets of equations are all *observationally equivalent*.

As long as we think of the causal relationship embedded in a model as merely related to the solution for the variables,

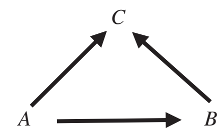

Fig. 1. Structure S1.

observational equivalence will undermine the utility of the models. Simon offers two different strategies for dealing with the problem of observational equivalence. Simon's first strategy is simply to forbid linear combinations of equations on the grounds that each equation represents a distinct mechanism and that a linear combination creates a mongrel without causal significance (Simon & Rescher, 1966; Simon & Iwasaki 1988).[4] Such a strategy amounts to assigning equations to variables. The relationship is then easily indicated quasi-graphically by introducing a new "causal equality" operator ($\Leftarrow$).[5] The subset of equations (1) and (2) could then be written as:

$$A \Leftarrow \alpha_A, \tag{1′}$$

$$B \Leftarrow \alpha_{BA}A, \tag{2′}$$

with the convention that the arrowhead indicates the direction of causation.

In his 1953 paper, however, Simon offers a *second strategy* for resolving observational equivalence. While he considers his formal account of causal order syntactic, he suggests that we adopt a higher-order semantic relation of *direct control* over parameters (Simon, 1953, pp. 24–27). He invites us (and nature) to experiment on a system by directly controlling the value of its parameters (the coefficients now being thought of as parameters that can take different values). The privileged parameterization is the one in which such experiments can be conducted independently. Thus, if one represents a causal system by equations (1) and (2) and can control A directly by choosing α_A and thereby control B indirectly without altering the functional form of equation (2), then the parameter set $\{\alpha_A, \alpha_{BA}\}$ and its associated functional form is privileged.

If for example, (1) and (2) represented the true causal order, but we instead modeled the causal relationships with (4) and (5), our control of A and B would not show the same sort of functional invariance. In fact, the only way to achieve the same values for A and B would be for the coefficient values of $\{\beta_A, \beta_B, \beta_{AB}\}$ to shift according to the restrictions (6)–(8). In effect, the decision that $\{\alpha_A, \alpha_{BA}\}$ is *the* parameter set—and that any other set of coefficients (e.g., $\{\beta_A, \beta_B, \beta_{AB}\}$) are simply functions of those parameters—determines the causal direction among the variables: it puts the arrowheads on the shafts. No other functionally equivalent system shares this invariance property. In fact, the uniqueness of the causal order determined on this basis can be proved (Iwasaki & Simon, 1994, p. 156).

I have previously supported and elaborated Simon's second strategy. Yet, why should we prefer the idea of privileging certain sets of parameters as capturing the possibilities for interventions in the world over the idea of privileging certain equations as capturing the causal mechanisms in the world? It is instructive to notice that Simon's original analysis of causal order was offered in support of the Cowles Commission's (1953) econometric program. While he talked about the structure of formal models, the point was ultimately to aid the progress of an epistemological project. Cartwright (2007, p. 81 *passim*) has stigmatized causal analysis that supposes modularity (of which more in due course) as making

[3] A *linear model* in Simon's (1953) terminology.

[4] The strategy is first in our exposition, but not first in its appearance in Simon's work.

[5] Cartwright uses "c$^{=}$ " to serve the same purpose.

Please cite this article in press as: Hoover, K. D. Causal structure and hierarchies of models. *Studies in History and Philosophy of Biological and Biomedical Sciences* (2012), http://dx.doi.org/10.1016/j.shpsc.2012.05.007

assumptions because they are "epistemically convenient." In a sense, I would turn her evaluation on its head: Simon's second strategy makes weaker assumptions—that is, assumptions that make smaller epistemological demands on the econometric modeler or scientist in the construction of his models. Weaker assumptions have considerable *epistemological utility*, which is an excellent methodological reason to prefer them.

Simon appears insensitive to the difference between his two strategies—mostly likely because they coincide in his examples. Simon implicitly supposes that parameters are uniquely associated with particular variables or mechanisms—for example, if a parameter shows up in one equation, it does not show up in any other equation. Such a supposition implies that it is easy to analyze a system of equations into distinct parts. Many economic models, however, violate Simon's supposition and comprise equations in which the same parameter might appear in multiple equations. For example, we might have

Structure S3

$$A = \alpha_A, \tag{9}$$

$$B = \alpha_A + \alpha_{BA}A. \tag{10}$$

Hoover (2001b, ch. 3) suggests that such systems of equations require modification of Simon's formalism for causal structure. We must require Simon's hierarchy condition: a direct cause must belong to a lower-order self-contained subset with no self-contained subset intervening between the direct cause and the self-contained subset that determines the effect. But in addition we require that the set of parameters associated with the subset that determines the cause be a proper subset of the set of parameters that determine the effect. This condition is met in equations (9) and (10) in which *A* causes *B*. It would not be met in the following system of equations:

Structure S4

$$A = \alpha_A, \tag{11}$$

$$B = \alpha_A A. \tag{12}$$

An advantage of this extension of Simon's approach is that it inherits the property of his second strategy that causal order is uniquely defined by the functional relations among variables (that is, no matter what equivalent form the equations take) so long as we can say which objects are parameters and which are coefficients the values of which are functions of the true parameters. It does not matter, for example, whether we describe the relationship among the variables by structure S1 or by structure S2 as long as we know that the α's represent the actual scope of interventions and the β's are, at best, functions of the α's.

A consequence of our parameter-nesting condition is that every effect must have some parameter distinct from its causes—that is, there is some means of intervening directly on every effect independently of its causes. This is one version of a "modularity" assumption. Cartwright (2007, chs. 7 and 8) regards modularity as an ungrounded assumption that is made to render systems of equations into an "epistemically convenient" form. In contrast, I would like to distinguish different types of modularity and to suggest that understanding them and their relationship is a key to understanding the relationship of models at different levels of description or aggregation.

4. Modularity and identity

The general idea of modularity is that a system is modular if it can broken down into parts that retain their integrity and functionality. There are at least three types of modularity at issue in the formal account of causal order under discussion.

First, is the assumption that parameters are variation-free. Recall that a parameter is called *variation-free* when it can be set to any value in its admissible range independently of the setting of any other parameter. Equivalently, parameters are variation-free when there are no constraints operating among them.[6] The requirement that parameters be variation-free is a useful convention that helps to keep our reasoning clear about the causal structure of models, but it is essentially trivial in the sense that any formal model that incorporated non-variation-free parameters could be turned into one in which the constraints among parameters is shifted into the functional form as a constraint among variables.

An example should make this clear. Consider the system

Structure S5

$$X = \alpha \tag{13}$$

$$Y = \beta X + \gamma, \tag{14}$$

subject to the parameter constraint $\beta \leqslant \alpha$. The parameters are not variation-free, since the value of α restricts the range of admissible values of β. However, this system can be reformulated into a related system with the same solutions in which the parameters are variation-free:

Structure S5′

$$X = \alpha \tag{15}$$

$$Y = \begin{cases} c\beta X + \gamma, \text{ if } \alpha \geqslant \beta \\ \alpha X + \gamma, \text{ if } \alpha < \beta \end{cases} \tag{16}$$

A model that was linear in variables with a nonlinear parameter constraint has been transformed into a model that is nonlinear in variables with unconstrained parameters. Applying our extension of Simon's causal formalism, *X* causes *Y*.

A second form of modularity has been the subject of vigorous debate (see Cartwright, 2007, chs. 7 and 8; Hausman & Woodward, 1999, 2004). To understand this form of modularity consider Woodward's (2003, ch. 2) *manipulability account* of causation. The essence of the approach is conveyed in his definition of a *direct cause*:

> (**DC**) A necessary and sufficient condition for *X* to be a direct cause of *Y* with respect to some variable set **V** is that there be a possible intervention on *X* that will change *Y* (or the probability distribution of *Y*) when all other variables in **V** besides *X* and *Y* are held fixed at some value by interventions. (Woodward, 2003, p. 55)

Despite Woodward's (2003, p. 39) regarding causation as fundamentally a type-level relationship among variables, (**DC**) defines direct cause in terms of a token-level action—an intervention. Suppose we wish to assess whether *B* is a direct cause of *C* in system S1 (i.e., Fig. 1 or equations (1)–(3)). Following Pearl (2000), Woodward suggests that we do this by setting the other variables (here only *A*) to token values and, in effect, "breaking" (or "wiping out") the causal connections between variables wherever needed to achieve this. Thus Fig. 2 would replace Fig. 1 in which the lower-case letter indicates the token value of the correlative upper-case variable and in which the causal arrow into *B* are removed. *B* causes *C*, then, if a change in *B*, say, from *b* to $\acute{b}$ changes (or would change, the intervention being conceived of counterfactually) in a change in *C*, say, from *c* to *c*′.

Pearl (2000, pp. 70) represents interventions by the operators "set(*X*)" or "do(*X*)". Woodward (2003, pp. 47–48) notes "*X* and set *X* are not really different variables, but rather the same variable

[6] Fennell (2005, p. 50) claims that parameters can both be variation-free and mutually constraining. This shows that he simply fails to grasp the meaning of "variation-free," and results, in the particular example offered, from failing to draw a clear distinction between parameters and variables.

Please cite this article in press as: Hoover, K. D. Causal structure and hierarchies of models. *Studies in History and Philosophy of Biological and Biomedical Sciences* (2012), http://dx.doi.org/10.1016/j.shpsc.2012.05.007

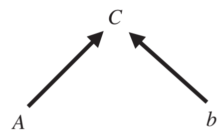

Fig. 2. An intervention on S1.

embedded in different causal structures…" The transition from Fig. 1 to Fig. 2—from one causal structure to another—presupposes that the wiping out of causal arrows without affecting other parts of the graph makes sense. Woodward refers to the property that warrants such an intervention as *modularity*:

> a system of equations will be modular if it is possible to disrupt or replace (the relationships represented by) any one of the equations in the system by means of an intervention on (the magnitude corresponding to) the dependent variable in that equation, without disrupting any of the other equations. (Woodward, 2003, p. 48)

And while he recognizes that representations of causal relationships may not always display modularity, he assumes

> that when causal relationships are correctly and fully represented by systems of equations, each equation will correspond to a distinct causal mechanism and that the equation system will be modular. (Woodward, 2003, p. 49)

The modularity of the system consists in the independence of the causal linkage between *B* and *C* from modifications that set other variables to particular values or break causal linkages. While Woodward promises us that a full representation of causal system is necessarily modular, he neither demonstrates the basis for that claim nor gives us any guidance about representing causality in less than full representations—another tug of the perfect-model model. As the following example shows economic models are frequently not modular in Woodward's sense, but nonetheless are causally ordered.

Consider a commonplace model of supply and demand supplemented with a tax on supply:

Structure S6

$$R = \rho, \tag{17}$$
$$N = \nu, \tag{18}$$
$$Q = \alpha(1-\tau)P + \beta R, \tag{19}$$
$$P = \gamma Q + \delta N, \tag{20}$$
$$T = \tau PQ + \phi \tag{21}$$

For concreteness, we follow Simon and Rescher (1966) in thinking of the commodity as wheat (where *Q* is its quantity and *P* is its price), *R* as rainfall, and *N* as population. Lower-case Greek letters are parameters. We can think of equation (17) as a supply equation, where the parameter τ represents an *ad valorem* tax on the sales of farmers that generates revenues of *T*.

Applying our version of Simon's formal analysis of causal order (his second strategy), we get the causal structure shown in Fig. 3. Some points to notice: First, a failure of modularity shows up in the fact that we cannot set *Q* to some arbitrary value ($Q = q$), as required in Woodward's definition of direct cause, without destroying the tax relationship (21), since the tax rate (τ) appears in (21) only because it also appears in (19). A strength of our extension of Simon's second strategy is that it, nevertheless, determines the causal order without any appeal to modularity of Woodward's type.

Second, when a subsystem involving more than one variable cannot be separated into distinct self-contained subsystems, then

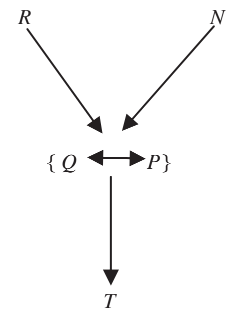

Fig. 3. Structure S6 on Simon's second strategy.

Simon regards them as standing in relationship of mutual causation—indicated both by the brackets and by the double-headed arrow. Such mutual causation underlines the difference between Simon's first and second strategies. On the first strategy, which identifies distinct mechanisms, we might wish to read equations (19) and (20) as distinct mechanisms and therefore to draw the causal graph somewhat differently as in Fig. 4.

To clarify what is at stake between the two strategies, consider, a pair of models:

Structure S7

$$A = \alpha, \tag{22}$$
$$B = \beta A + \delta D, \tag{23}$$
$$C = \delta B, \tag{24}$$
$$D = \phi C \tag{25}$$

Structure S8

$$A = \alpha, \tag{26}$$
$$B = \beta A + \theta C, \tag{27}$$
$$C = \delta B + \lambda D, \tag{28}$$
$$D = \phi C \tag{29}$$

On Simon's first strategy, we might see these as causally distinct models with the graphical representations shown in Figs. 5 and 6. But Simon's second strategy is unable to draw any distinction between the two models and would represent the causal structure as in Fig. 7. The variables *B*, *C*, and *D* are so thoroughly entangled in a web of mutual causality that we cannot articulate their internal structure on the basis of the topology of interventions.

The inability to distinguish these cases is not a weakness of Simon's second strategy. Rather it points to an important characteristic of models operating on different levels of causal explanation. It is elementary that causes must be distinct from their effects. It is not enough that they be conceptually distinct; they must be causally distinct. My favorite example is provided by bond prices and yields. The price of a bond is the amount one pays, say, in pounds sterling for a bond. The yield of the bond is the equivalent annual percentage return for holding the bond. Price and yield are conceptually distinct. They are not measured in the same units, and they play different roles in evaluating various economic and financial situations. They are, however, connected by an identity. The simplest example is the consol, a perpetual bond invented by the British government in the 18th century. A 3-percent consol pays its holder £3 each year. Its yield (*R*) and its price (*P*) are connected by an identity:

$$R \equiv \frac{3}{P} \tag{30}$$

or equally

$$P \equiv \frac{3}{R} \tag{30'}$$

Please cite this article in press as: Hoover, K. D. Causal structure and hierarchies of models. *Studies in History and Philosophy of Biological and Biomedical Sciences* (2012), http://dx.doi.org/10.1016/j.shpsc.2012.05.007

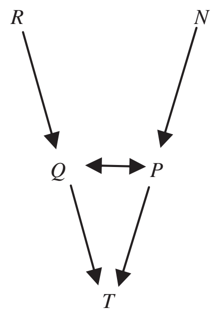

Fig. 4. Structure S6 on Simon's first strategy.

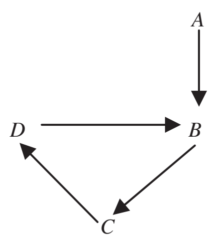

Fig. 5. Structure S7 (first strategy).

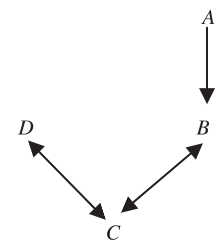

Fig. 6. Structure S8 (first strategy).

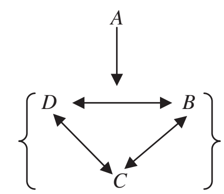

Fig. 7. Structures S7 and S8 (second strategy).

One sometimes reads in the newspapers that bond yields rose causing bond prices to fall. But this is wrong. The identity means that a rise in bond yields is just the very same thing as a fall in bond prices.

Causally, there is no way to draw a wedge between the two variables—no way to intervene independently on them—even though they are conceptually distinct. And this, I suggest, is the standard for causal identity. In terms of Simon's second strategy we can say that two variables are *causally identical* if, and only if, they share the same parameter set—that is, they are determined in a subsystem that cannot be separated into smaller subsystems. Causal identity defined in this way implies that the parameter-nesting condition that we added to Simon's formalism is a logical requirement, given that causes must not be identical with their effects.

Causal identity has some bearing on aggregation. Whenever, variables are causally identical, it is possible without loss of information to let one serve as index for the others and to simplify the model. For example, in the supply-and-demand model (S6), Q and P are causally identical and we could omit one equation with no loss of causal information. If the tax function (21) were replaced by a flat tax without the parameter ϕ (i.e., by (21′) $T = \tau PQ$), then that equation could be omitted as well. The tax function would still give useful information to the accountant, but it would provide no causally salient information for the policymaker.

We can relate this notion of causal identity to the issue of modularity. Cartwright offers the operation of "a well-made toaster" as a counterexample to Woodward's modularity requirement:

> The expansion of the sensor due to the heat produces a contact between the trip plate and the sensor. This completes the circuit, allowing the solenoid to attract the catch, which releases the lever. The lever moves forward and pushes the toast rack open.
>
> I would say that the bolting of the lever causes the movement of the rack. It also causes a break in the circuit. Where then is the special cause that affects only the movement of the rack? Indeed, where is there space for it? The rack is bolted to the lever. The rack must move exactly as the lever dictates. So long as the toaster stays intact and operates as it is supposed to, the movement of the rack must be fixed by the movement of the lever to which it is bolted.
>
> Perhaps, though, we should take the movement of the lever to the rack as an additional cause of the movement of the rack? In my opinion we should not. To do so is to mix up causes that produce effects within the properly operating toaster with the facts responsible for the toaster operating in the way it does; that is, to confuse the causal laws at work with the reason those are the causal laws at work. (Cartwright, 2007, pp. 85–86.)

Cartwright's toaster, I believe, does provide an illustration of the failure of modularity in the sense that if we take the position of the bolt not as a parameter subject to various settings, but as a constant, then there is no intervention on the lever in a model of a properly operating toaster that does also alter the rack. Another way to look it, however, is that Cartwright has described to us what it means to be a module. The lever and the rack are causally identical in a model formulated from a perspective in which the bolt is neither a variable nor a parameter subject to intervention. But in drawing a distinction between "causes that produce effects within the properly operating toaster" and "facts responsible for the toaster operating in the way it does," she also hints that a model of a well-made toaster is not the only possible model of the same toaster. We might need a model of the *causal* facts responsible for the toaster operating properly—for example, if we were in the business of designing toasters. Then the fact that a bolt can determine that a lever and rack remain tightly joined is a salient causal fact. Similarly, the model of the properly operating toaster, leaves out of the model many facts that are true in the world. If the toaster is knocked off the counter and the bolt is loosened (or the lever is bent), the invariant connection between lever and rack would fail. The model that would be of use to the repairman is one that admits that failure as a causal possibility.

The models here are not entirely separate; rather they form a family with different levels of aggregation. A model in which the bolt is considered only in its tightened state is a special case or an aggregated version of a model in which the position of the bolt is a variable. We might—to take an economic case—elaborate the supply-and-demand model by treating the tax rate τ as a variable causally determined by connections to other variables governed by other parameters. In this case, the model that takes τ to be a variation-free parameter would be a special case or an aggregated version of the more general model. Formally, it is easy to construct

Please cite this article in press as: Hoover, K. D. Causal structure and hierarchies of models. *Studies in History and Philosophy of Biological and Biomedical Sciences* (2012), http://dx.doi.org/10.1016/j.shpsc.2012.05.007

models in related families at different levels of aggregation; but, substantively, facts about the world constrain whether such families are natural or useful or concocted and misleading. The variation-freeness of parameters is, as we have seen, trivially constructed within formal models. Yet, it is anything but trivial that a model in which the parameters have been specified as variation-free adequately captures the functional relationships among variables. It is only when it does that it can adequately represent the causal facts from any given perspective or at any given level of aggregation.

Cartwright's example of the toaster illustrates that the second type of modularity is not an essential feature of causation. But more importantly, it illustrates what it takes to construct a module. The reason that the lever-rack ensemble cannot be intervened upon separately is that they are, at the level of a properly operating toaster, in fact a module. A module is constructed by establishing conditions in which the conceptually distinct parts are causally identical. When the only admissible interventions necessarily alter each of a set of variables, the variables act as a unit—a *module*.

There is, then, a third type of modularity. When a unit can be constructed out of parts, the parts can be considered modules at one level and the constructed unit a module at higher level. Modularity of this type is essential to anyone who builds new mechanisms and devices out of parts with established properties. Notice, however, that the parts of a modular unit need not display modularity of the second type—that is, that the properties that they underwrite at the higher level need not be robust to interventions on the distinct parts. For example, airplane fuselages are sometimes monocoque constructions in which the stressed skin of the airplane and underlying structural members are mutually supporting without a connected framework under the skin. The skin and the framework function properly only in a relationship of mutual support; they do not display the second type of modularity any more than do Cartwright's lever and rack. Yet they are built out of distinct pieces or modules of the third type. Having a model of the capacities of different pieces—a model less aggregated that a model of the properly functioning toaster or airplane is essential to engineers and designers.

At the level of the properly operating market, the same situation arises in the supply-and-demand model (S6). Equations (19) and (20) are a module and, from one causal level, can be treated as unit. The way that the equations are written and Simon's first strategy to causal order suggests (see Fig. 4) that the module is, in fact, constructed out of parts. The second strategy does not, however, provide the basis for analyzing the internal structure of a module. For that we have to appeal to knowledge other than the scope of interventions—e.g., knowledge of the parts and their capacities—or we have to analyze a model on a deeper level in which aggregated relationships are articulated into additional variables and parameters which permit us to define the distinction between the parts of the module by means of the scope of the interventions allowed by the parameters. In the economic framework, the implicit equilibrium assumption embedded in supply always equaling demand is an aggregation that elides an underlying mechanism (Arrow, 1959). A model that specified that mechanism may well permit us to analyze the module using Simon's second strategy into the very parts (among others) that are implicit in his first strategy.

It is frequently a goal of science to break modules down into component parts and to provide a model of how the parts interact to form a module. While it is a good heuristic to try to push to a deeper level of analysis, there can be no guarantee that one will succeed.

The account of modularity and causal identity provided here relates to another, and frequently neglected (at least among economists) aspect of Simon's causal analysis that he sees as the groundwork for what he calls the "sciences of the artificial" (Simon, 1996, especially chs. 1 and 8). Simon focuses on the fact that complex systems are frequently *decomposable* into units. Some units may have an internal structure, which is irrelevant to their interactions with the other units. They could, in fact, be replaced with units with different internal structures, so long as the alternative units each process inputs to outputs in the same manner. For example, the subsystems determining *B*, *C*, and *D* in structures S1 and S2 (see Figs. 4 and 5) have different internal structures; yet as shown in Fig. 6, their external causal relations (or external environment) are identical. The subsystems thus form "black boxes," interchangeable with other black boxes that function externally in the same way.

For example, if the lever-rack assembly in Cartwright's toaster were made of metal, we may well be able to replace it with one made of some type of plastic or we may be able to replace the assembly that is bolted together out of two parts with a single part by, say, welding rather than bolting the parts together. We need not imagine that the inner and outer environments are completely separated in fact. There may be boundary conditions that, if breached, connect the two environments. We may, for example, generally ignore the distinction between the metal and the plastic assemblies in the toaster; but, in an environment with particularly high heat, the plastic assembly may fail.

For this reason, Simon frequently emphasizes *near decomposability*—the situation in which some parts are connected by strong or highly stable linkages and others by much weaker or more contingent linkages (Simon 1996, ch. 8; Iwasaki & Simon 1994; Simon & Iwasaki 1988). The highly connected parts will act modularly with respect to the parts (or other modules) to which they are more weakly linked. Simon's most common examples of near decomposability are drawn from dynamic models. When variables adjust to each other very quickly, they may be treated as in constant static equilibrium with respect to each other, ignoring their own dynamics. When variables that adjust very slowly to others, they can be treated as exogenous over shorter time horizons. Dynamic analysis in the short run can then be simplified to focusing on the relationships of variables or ensembles of variables that adjust neither very quickly nor very slowly to each other.

This brings us back to perspectival realism. Whether we find a model adequate that ignores the very fast or very slow adjustments among variables depends on pragmatic considerations—our purposes or interests. A macroeconomist makes no important error when he takes the yields on financial assets as standing in a state of constant equilibrium (i.e., models financial markets with a "no-arbitrage condition"). A trader whose function is to conduct arbitrage would find the macroeconomist's model useless. Equally, a macroeconomist makes no important error in assuming that population in the short run is exogenous relative to economic variables. A demographer whose preferred time horizon is decades would miss an essential causal mechanism by making the same assumption.

Acknowledgements

This paper was presented under the title "Causal Structure and Levels of Explanation" at the conference on Causality in the Biomedical and Social Sciences" at Erasmus University, Rotterdam, 6-8 October 2010. I acknowledge the support of the U.S. National Science Foundation (grant no. NSF SES-1026983).

References

Arrow, K. J. (1959). Toward a theory of price adjustment. In M. Abramovitz, Moses et al. *The allocation of economic resources: Essays in honor of Bernard Francis Haley* (pp. 41–51). Stanford: Stanford University Press.

Cartwright, N. (2007). *Hunting causes and using them: Approaches in philosophy and economics*. Cambridge: Cambridge University Press.

Duesenberry, J. S., Fromm, G., Klein, L. R., & Kuh, E. (Eds.). (1965). *The Brookings quarterly econometric model of the United States*. Chicago: Rand McNally.

Fennell, D. J. (2005). *A philosophical analysis of causality in econometrics*. Ph.D dissertation, London School of Economics and Political Science.

Friedman, M. (1949). The Marshallian demand curve. *Journal of Political Economy, 57*, 463–495.

Friedman, M. (1951). Comment on "A Test of an econometric model for the United States, 1921–1947,". In C. Christ (Ed.), *Conference on business cycles* (pp. 107–114). New York: National Bureau of Economic Research.

Friedman, M. (1953). The methodology of positive economics. In *Essays in positive economics* (pp. 3–43). Chicago: University of Chicago Press.

Giere, R. N. (2006). *Scientific perspectivism*. Chicago: University of Chicago Press.

Hausman, D. M., & Woodward, J. (1999). Independence, invariance, and the causal Markov condition. *British Journal for the Philosophy of Science, 50*, 521–583.

Hausman, D. M., & Woodward, J. (2004). Modularity and the causal Markov condition: A restatement. *British Journal for the Philosophy of Science, 55*, 147–161.

Hoover, K. D. (1988). *The new classical macroeconomics: A skeptical inquiry*. Oxford: Blackwell.

Hoover, K. D. (2001a). Is macroeconomics for real? In U. Mäki (Ed.), *The economic world view*. Cambridge: Cambridge University Press.

Hoover, K. D. (2001b). *Causality in macroeconomics*. Cambridge: Cambridge University Press.

Hoover, K. D. (2004). Lost causes. *Journal of the History of Economic Thought, 26*, 149–164.

Hoover, K. D. (2006). The Past as future: The Marshallian approach to post-Walrasian economics. In D. Colander (Ed.), *Post-Walrasian macroeconomics: Beyond the dynamic stochastic general equilibrium model* (pp. 239–257). Cambridge: Cambridge University Press.

Hoover, K. D. (2009a). Milton Friedman's stance. The methodology of causal realism. In U. Mäki (Ed.), *The methodology of positive economics: Milton Friedman's essay fifty years later* (pp. 303–320). Cambridge: Cambridge University Press.

Hoover, K. D. (2009b). Microfoundations and the Ontology of macroeconomics. In H. Kincaid & D. Ross (Eds.), *Oxford handbook of the philosophy of economic science* (pp. 386–409). Oxford: Oxford University Press.

Hoover, K. D. (2010a). Causal pluralism and the limits of causal analysis: A review of Nancy Cartwright's Hunting causes and using them: Approaches in philosophy and economics. In B. E. Ross & E. B. Jeff (Eds.), *Research in the history of economic thought and methodology*, (Vol. 28A) (pp. 381–395).

Hoover (2010b). Idealizing reduction: The microfoundations of macroeconomics. *Erkenntnis, 73*, 329–347.

Hoover, K. D. (2011). Counterfactuals and causal structure. In P. M. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 338–360). Oxford University Press.

Hoover, K. D. (2012a). Microfoundational programs. In P. G. Duarte & G. T. Lima (Eds.), *Microfoundations reconsidered: The relationship of micro and macroeconomics in historical perspective* (pp. 19–61). Cheltenham: Edward Elgar.

Hoover, K. D. (2012b). Pragmatism, perspectival realism, and econometrics. In A. Lehtinen, J. Kuorikoski, & P. Ylikoski (Eds.), *Economics for real: Uskali Mäki and the place of truth in economics* (pp. 223–240). London: Routledge.

Hoover, K. D. (2012c). Identity, structure, and causation in economic models. In H.-K. Chao, S.-T. Chen, & R. A. Millstein (Eds.), *Mechanism and causality in biology and economics*. Berlin: Springer.

Iwasaki, Y., & Simon, H. A. (1994). Causality and model abstraction. *Artificial intelligence, 67*, 143–194.

Janssen, M. C. W. (1993). *Microfoundations: A critical inquiry*. London: Routledge.

Klein, L. R. (1950). *Economic fluctuations in the United States, 1921–1941*. New York: Wiley.

Klein, L. R. (1992). My professional life philosophy. In M. Szenberg (Ed.), *Eminent economists: Their life philosophies* (pp. 180–189). Cambridge: Cambridge University Press.

Klein, L. R. & Goldberger, A. S. (1955). *An econometric model of the United States: 1929–1952*. North-Holland.

Lewis, D. (1973). Causation. *Journal of Philosophy, 70*, 556–567.

Lucas, R. E. Jr., (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester conference series on public policy, 1*, 19–46.

Marshall, A. (1885). The present position of economics. In A. C. Pigou (Ed.), *Memorials of Alfred Marshall* (pp. 152–174). London: Macmillan.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge: Cambridge University Press.

Simon, H. A. (1953). Causal order and identifiability. In W. C. Hood & T. C. Koopmans (Eds.), *Studies in econometric method*. *Cowles Commission Monograph* (pp. 49–74). New York: Wiley (Page numbers refer to the reprint in Simon, *Models of man* (pp. 10–36). New York: Wiley, 1957.

Simon, H. A. (1996). *The sciences of the artificial* (3rd ed). Cambridge, MA: MIT Press.

Simon, H. A., & Iwasaki, Y. (1988). Causal ordering, comparative statics, and near decomposability. *Journal of Econometrics, 39*, 149–173.

Simon, H. A., & Rescher, N. (1966). Cause and counterfactual. *Philosophy of science*, 323–340.

Teller, P. (2001). Twilight of the perfect model model. *Erkenntnis, 55*, 393–415.

Tinbergen, J. (1939). *Business cycles in the United States of America. 1919–32*. Geneva: League of Nations.

Wimsatt, W. C. (2007). *Re-engineering philosophy for limited beings: Piecewise approximations to reality*. Cambridge, MA: Harvard.

Woodward, J. B. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.

Walras, L. (1874[1954]). *Elements of pure economics*, translated by William Jaffe. Homewood, IL: Richard Irwin, 1954.

Please cite this article in press as: Hoover, K. D. Causal structure and hierarchies of models. *Studies in History and Philosophy of Biological and Biomedical Sciences* (2012), http://dx.doi.org/10.1016/j.shpsc.2012.05.007