History, Philosophy and Theory of the Life Sciences

Volume 3

Editors:
Charles T. Wolfe, Ghent University, Belgium
Philippe Huneman, IHPST (CNRS/Université Paris I Panthéon-Sorbonne), France
Thomas A.C. Reydon, Leibniz Universität Hannover, Germany

Editorial Board:
Marshall Abrams (University of Alabama at Birmingham)
Andre Ariew (Missouri)
Minus van Baalen (UPMC, Paris)
Domenico Bertoloni Meli (Indiana)
Richard Burian (Virginia Tech)
Pietro Corsi (EHESS, Paris)
François Duchesneau (Université de Montréal)
John Dupré (Exeter)
Paul Farber (Oregon State)
Lisa Gannett (Saint Mary's University, Halifax)
Andy Gardner (Oxford)
Paul Griffiths (Sydney)
Jean Gayon (IHPST, Paris)
Guido Giglioni (Warburg Institute, London)
Thomas Heams (INRA, AgroParisTech, Paris)
James Lennox (Pittsburgh)
Annick Lesne (CNRS, UPMC, Paris)
Tim Lewens (Cambridge)
Edouard Machery (Pittsburgh)
Alexandre Métraux (Archives Poincaré, Nancy)
Hans Metz (Leiden)
Roberta Millstein (Davis)
Staffan Müller-Wille (Exeter)
Dominic Murphy (Sydney)
François Munoz (Université Montpellier 2)
Stuart Newman (New York Medical College)
Frederik Nijhout (Duke)
Samir Okasha (Bristol)
Susan Oyama (CUNY)
Kevin Padian (Berkeley)
David Queller (Washington University, St Louis)
Stéphane Schmitt (SPHERE, CNRS, Paris)
Phillip Sloan (Notre Dame)
Jacqueline Sullivan (Western University, London, ON)
Giuseppe Testa (IFOM-IEA, Milano)
J. Scott Turner (Syracuse)
Denis Walsh (Toronto)
Marcel Weber (Geneva)

For further volumes:
http://www.springer.com/series/8916

Hsiang-Ke Chao • Szu-Ting Chen
Roberta L. Millstein
Editors

Mechanism and Causality in Biology and Economics

Springer

Kerem, Bat-sheva, et al. 1989. Identification of the cystic fibrosis gene: Genetic analysis. *Science* 245: 1073–1080.

Kirk, Kevin L., and David C. Dawson (eds.). 2003. *The cystic fibrosis transmembrane conductance regulator*. New York: Kluwer.

Knol, K. 1995. Cystic fibrosis: The past 25 years. *The Netherlands Journal of Medicine* 46: 266–270.

Konstan, Michael W., P.J. Byard, C.L. Hoppel, and P.B. Davis. 1995. Effect of high-dose ibuprofen in patients with cystic fibrosis. *The New England Journal of Medicine* 332(13): 848–854.

Lindee, Susan, and Rebecca Mueller. 2011. Is Cystic Fibrosis Genetic Medicine's Canary? *Perspectives in Biology and Medicine* 54(3): 316–331.

Liu, X., et al. 2002. Partial correction of endogenous delta F508 CFTR in human cystic fibrosis airway epithelia by spliceosome-mediated RNA trans-splicing. *Nature Biotechnology* 20(1): 47–52.

Liu, X., et al. 2005. Spliceosome-mediated RNA trans-splicing with recombinant adeno-associated virus partially restores cystic fibrosis transmembrane conductance regulator function to polarized human cystic fibrosis airway epithelial cells. *Human Gene Therapy* 16(9): 1116–1123.

Machamer, Peter. 2004. Activities and causation: The metaphysics and epistemology of mechanisms. *International Studies in the Philosophy of Science* 18: 27–39.

Machamer, Peter, Lindley Darden, and Carl F. Craver [MDC]. 2000. Thinking about mechanisms. *Philosophy of Science* 67: 1–25. Reprinted in Darden (2006, Ch. 1).

Moghaddam-Taaheri, Sara. 2011. Understanding pathology in the context of physiological mechanisms: The practicality of a broken-normal view. *Biology and Philosophy* 26: 603–611.

Morgan, Thomas H., A.H. Sturtevant, H.J. Muller, and C.B. Bridges. 1915. *The mechanism of Mendelian heredity*. New York: Henry Holt & Company.

Pearson, Helen. 2009. Human genetics: One gene, twenty years. *Nature* 460: 164–169.

Quinton, Paul. 1983. Chloride impermeability in cystic fibrosis. *Nature* 301: 421–422.

Riordan, John R., et al. 1989. Identification of the cystic fibrosis gene: Cloning and characterization of complementary DNA. *Science* 245(4922): 1066–1073.

Rommens, Johanna M., et al. 1989. Identification of the cystic fibrosis gene: Chromosome walking and jumping. *Science* 245(4922): 1059–1065.

Rowe, Steven M., Stacey Miller, and Eric J. Sorscher. 2005. Cystic fibrosis: Review article on mechanisms of disease. *The New England Journal of Medicine* 352(19): 1992–2001.

Skipper, Robert A. Jr., and Roberta L. Millstein. 2005. Thinking about evolutionary mechanisms: Natural selection. In ed. Carl F. Craver and Lindley Darden, Special Issue: "Mechanisms in Biology." *Studies in History and Philosophy of Biological and Biomedical Sciences* 36: 327–347.

Thagard, Paul. 1998. Explaining disease: Causes, correlations, and mechanisms. *Minds and Machines* 8: 61–78.

Thomson, Jeremy. 2002. Gene therapy hope for cystic fibrosis: SMaRT treatment for fatal disease shows promise. *Nature News*. doi:10.1038/news020101-6. Published online 4 January 2002. http://www.nature.com/news/2002/020104/full/news020101-6.html. Accessed 22 Sept 2011.

Wang, X., et al. 2006. Hsp90 cochaperone aha1 downregulation rescues misfolding of CFTR in cystic fibrosis. *Cell* 127: 803–815.

Watanabe, Akira, Seungho Choe, Vincent Chaptal, John M. Rosenberg, Ernest M. Wright, Michael Grabe, and Jeff Abramson. 2010. The mechanism of sodium and substrate release from the binding pocket of vSGLT. *Nature* 468: 988–992.

Watson, James D., Tania A. Baker, Stephen P. Bell, and Alexander Gann. 2007. *Molecular biology of the gene*, 6th ed. San Francisco: Benjamin Cummings.

Widdicombe, J.H. 2003. CFTR and airway pathophysiology. In *The cystic fibrosis transmembrane conductance regulator*, ed. Kevin L. Kirk and David C. Dawson, 137–159. New York: Kluwer.

Woodward, James. 2002. What is a mechanism? A counterfactual account. *Philosophy of Science* 69(4): S366–S377.

Chapter 3
Identity, Structure, and Causal Representation in Scientific Models

Kevin D. Hoover

Abstract Recent debates over the nature of causation, casual inference, and the uses of causal models in counterfactual analysis, involving *inter alia* Nancy Cartwright (*Hunting Causes and Using Them*), James Woodward (*Making Things Happen*), and Judea Pearl (*Causation*), hinge on how causality is represented in models. Economists' indigenous approach to causal representation goes back to the work of Herbert Simon with the Cowles Commission in the early 1950s. The paper explicates a scheme for the representation of causal structure, inspired by Simon, and shows how this representation sheds light on some important debates in the philosophy of causation. This structural account is compared to Woodward's manipulability account. It is used to evaluate the recent debates – particularly, with respect to the nature of causal structure, the identity of causes, causal independence, and modularity. Special attention is given to modeling issues that arise in empirical economics.

1 Models and Causes

Formal scientific models possess some distinct advantages over verbal accounts. (There are, to be sure, disadvantages as well.) All representations (formal or verbal) are partial: they omit, simplify, approximate, and idealize; they fall short of saying

---

Prepared for the *International Conference on the Philosophy of Economics and Biology* at the National Tsing Hua University, Hsinchu, Taiwan, 24–25 March 2011; it is a substantial revision of a paper first presented at the conference on *Modeling the World: Perspectives from Biology and Economics*, Helsinki, 28–30 May 2009. I am grateful to François Claveau and two anonymous referees for comments on an earlier draft. I acknowledge the support of the US National Science Foundation (grant no. NSF SES-1026983).

K.D. Hoover (✉)
Departments of Economics and Philosophy, Duke University,
213 Social Sciences Building, Box 90097, Durham, NC 27708-0097, USA
e-mail: kd.hoover@duke.edu

H.-K. Chao et al. (eds.), *Mechanism and Causality in Biology and Economics*,
History, Philosophy and Theory of the Life Sciences 3, DOI 10.1007/978-94-007-2454-9_3,

everything that could be truly said and short of saying everything that we might like to say. Recognition of the gap between the representation and the world leads various philosophers – among them, Paul Teller (2001) and Ronald Giere (2006) – to reconceptualize scientific knowledge as perspectival. Part of the reconceptualization is a repudiation of the view that *omni*science sets the standard for the worthiness of scientific knowledge.

The truth in a once-common vision of science is that the only fully adequate scientific knowledge trades in exceptionless, universal generalizations – scientific laws. All more specific knowledge is, in principle at least, derivable from these laws. Recognizing – as indeed any serious philosopher or scientist must – that we do not, in fact, possess all the laws simply meant that what we did possess was a slightly shabby, deficient version of what we wanted. We do not stand on Olympus, but science was nonetheless to be judged from the Olympian heights.

An alternative vision of science championed by Giere and Teller, as well as by Nancy Cartwright (1999), and William Wimsatt (2007), among others, starts lower and builds upward. The standards of good or successful science are partial and local, and science itself is constructed, to use an apt term from the subtitle of Wimsatt's (2007) book, in a *piecewise* manner.

The local knowledge that grounds science in this vision is often causal knowledge. Yet, like other parts of science, causation has often been analyzed top down. Many accounts of causation – for example, those of David Lewis (1973) and Daniel Hausman (1998) – explicate causes against a background of universal laws. In contrast, piecewise accounts of science typically take the causal relation as primitive or, at least, built from something more local and specific than universal laws.

A piecewise approach is especially suited to economics and other social sciences, biology, and areas of physical sciences, such as climatology – fields that would be hard to analyze from a small set of universal laws on the model of Newtonian mechanics. Economists, for example, increasingly conceptualize economics causally, as evident in the work of Clive Granger in time-series econometrics, James Heckman in microeconometrics, recent developments in "natural experiments" in economics, and counterfactual analysis.[1]

Causal realism is the doctrine that causal relationships exist in the world and that the role of causal models is to represent them adequately for some purpose. Not all scientists (nor all philosophers) who talk about causes are realists, but that is a question for another day. Here, I want to focus on *representation* of causes and not on fundamental ontology or epistemology. It is a commonplace that different representations or notational schemes allow us to see different things and that some schemes are more effective than others – consider Arabic numerals or Feynman diagrams. The main goal of this chapter is to develop a scheme for representing causal relationships and to consider the light that it sheds on how we should understand causation generally. The roots of the approach advocated here are found in my own work as a practitioner of economics and draw on sources, such as the work of Herbert Simon, that were originally aimed at problems that arose in economic and econometric analysis. The application is much broader than these origins might suggest.

---

[1] See Hoover (2008 and 2012a) on the place of causal analysis in economics and Reiss (2007) on natural experiments and counterfactual analysis. Hoover (2004) documents the fall and rebirth of causal analysis and language in economics.

In part, this chapter reacts to Cartwright's (2007) "pluralistic" account of causation. In stressing plurality, Cartwright fails to illuminate the close relationships among a number of approaches to causality that are hidden in alternative schemes of representing causal relations. In part, the chapter reacts to James Woodward's (2003) "manipulability" account of causation – an account which is much criticized by Cartwright. Woodward's understanding of causation appears to be driven by particular schemes of representing causes. A more effective scheme of representation suggests different conclusions with respect to several important issues. The account proposed here in no way fundamentally conflicts with the general approach of modeling causal relationships graphically, developed especially by Judea Pearl (2000) and Peter Spirtes et al. (2000) and used by Woodward. Rather it clarifies the relationship between graphical representations and systems of equations in a manner that both enriches the graphical approach and demonstrates the fundamental kinship of the two approaches.

2 Representing Causal Structure

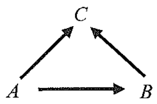
2.1 *Graphs and Equations*

While many philosophers understand causal relations as holding fundamentally among particular events, occurrences, or properties (i.e., among *tokens*), Woodward and most economists understand causal relations as holding among variables (i.e., among *types*). Token-level relationships for Woodward and the economists are causal to the degree that they instantiate a type-level relationship. In stochastic cases, token-level relationships are seen as the realization of random processes. Relations among variables are often expressible in the form of systems of equations. Equality is a symmetrical relationship, and the most distinctive characteristic of causal relations is their asymmetry: A causes B gives no ground for holding that B causes A (although we must not rule out mutual causation without further consideration). Woodward, in common with Pearl (2000), Spirtes et al. (2000), and other advocates of graph-theoretic or Bayes net methods of causal inference, represents causal relations by graphs in conjunction with equations.

Figure 3.1 shows a typical causal graph (uppercase letters represent variables) that corresponds to a system of equations:

$$A = \alpha_A, \tag{3.1}$$

Fig. 3.1 Causal graph of the system (3.1)–(3.3)



$$B = \alpha_{BA}A, \tag{3.2}$$

$$C = \alpha_{CA}A + \alpha_{CB}B, \tag{3.3}$$

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

Systems of equations are causally ambiguous. In stochastic cases, we generally recognize that correlation is not causation; in nonstochastic cases, the analogue is that functional relations are not causation. The arrows in the graph represent the primitive notion of the asymmetry of causation.

Graphs and equations interpreted causally both have a long history in economics (see Hoover 2004). But it is fair to say that equations have gained the upper hand and that, for many years, causation itself was rarely referred to directly, but at best was implicit in distinctions between dependent and independent (or endogenous and exogenous) variables and in synonyms and circumlocutions: instead of "*A* causes *B*," *A produces*, *influences*, *engenders*, *affects*, or *brings about B*, or *B reflects*, *is a consequence of*, *is a result of*, or *is an effect of A* (see Hoover 2009). A growing wariness of causal language went hand in hand with a wariness of graphical representation. As Pearl puts it:

> Early econometricians were very careful mathematicians; they fought hard to keep their algebra clean and formal, and they could not agree to have it contaminated by gimmicks such as diagrams. (Pearl 2000, p. 347)

Equations alone are causally ambiguous, since in themselves they do not represent causal asymmetry. But graphs are themselves causally ambiguous, because quite different functional relationships can be represented by the same graph (Woodward 2003, p. 44). Just as economists found circumlocutions to express "cause," they have typically – although not necessarily consistently – represented causal asymmetry by the convention of writing causes on the right-hand side and effects on the left-hand side of equations. Various devices have been suggested for explicitly combining the functional detail of systems of equations with the asymmetries of the graph. In lieu of the equal sign, Cartwright (2007, p. 13) suggests a causal equality ($c^{=}$) which Hoover (2001, p. 40) writes as ($\Leftarrow$). With a new notational device, the graph in Fig. 3.1 could be omitted and the system (3.1), (3.2), and (3.3) could, then, be rewritten as

$$A \Leftarrow \alpha_A, \tag{3.1$'$}$$

$$B \Leftarrow \alpha_{BA}A, \tag{3.2$'$}$$

$$C \Leftarrow \alpha_{CA}A + \alpha_{CB}B. \tag{3.3$'$}$$

Cartwright (2007, p. 16, *passim*) refers to such equations as "causal laws" – that is, laws that connect specific causes to specific effects. Woodward (2003), as well as most of the literature on graphical causal models, considers one-way causation only. Economists refer to such systems as *recursive*, while the graphical representations are often known as *directed acyclical graphs* (*DAG*s). Cyclical graphs (e.g., $A \rightarrow B \rightarrow C \rightarrow A$) are sometimes entertained, but the tight cycle of the simultaneous system ($A \rightarrow B \rightarrow A$ or $A \leftrightarrow B$), a bread-and-butter system in economics, is encountered far less frequently. The equations themselves are generally taken to be linear – especially linear in parameters. While these restrictions are by no means necessary, they highlight the inadequacy of the graphs fully to represent various levels of causal complexity. Economists avoiding graphs (*pace* Pearl) are perhaps partly motivated by an appreciation of the subtlety of causal representation and not some intuitive revulsion toward graphical gimmickry.

2.2 Simon on Causal Order

Following Haavelmo's "The Probability Approach in Econometrics" (1944), econometricians focused on what Frisch had called the "inversion problem" – namely, how to infer the original structure from passive observation of the data that it generates (Louçã 2007, p. 95). Later dubbed the "identification problem," a detailed account of the mathematics was for a time the central focus of the Cowles Commission (Koopmans 1950; Hood and Koopmans 1953). Identification naturally requires something to identify. Simon's contribution to the 1953 Cowles Commission volume sought to characterize the causal order of a system of equations.

Simon started with a *complete* system of equations – that is, a system that could be represented as a multivariate function with a well-defined solution. He then focused on *self-contained subsystems* of the complete system. To illustrate, Eqs. (3.1), (3.2), and (3.3) form a complete system. Equation (3.1) is a self-contained subsystem in that it determines the value of *A* without reference to any other equation. Equations (3.2) and (3.3) considered separately are not self-contained subsystems as they do not contain enough information to determine *B* or *C*. In contrast, Eqs. (3.1) and (3.2) together are a self-contained subsystem, since they determine the values of *A* and *B* without reference to Eq. (3.3).

Simon's conception is closely related to his later work on hierarchies of systems (Simon 1996; see also Hoover 2012c). 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 system (3.1), (3.2), and (3.3), *A* is determined entirely by (3.1) and can be regarded as an output. If we know *A*, we do not need to know (3.1) to determine *B*; a specific value for *A* forms an input that, in effect, turns the non-self-contained subsystem (3.2) into a self-contained subsystem. Its output is, of course, B. Knowing B alone, however, does not turn (3.3) into a self-contained subsystem. Substituting its value into (3.3) 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 (3.1). 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. 3.1.

Suppose, however, that we modify the system slightly by replacing Eq. (3.3) with (3.3″):

$$C = \alpha_{CB} B. \tag{3.3''}$$

Then, A directly causes B, and B directly causes C, but A only *indirectly* causes C: Eq. (3.1) is nested in the complete system (3.1), (3.2), (3.3″), but the self-contained subsystem (3.1) and (3.2) intervenes between the self-contained subsystem (3.1) and the self-contained complete system (3.1), (3.2), (3.3″).

Simon's analysis assumes that the original way of writing the equations is canonical. But he notices that the same functional relationships can be represented by other sets of equations. So, for example, the self-contained subsystem (3.1) and (3.2) could be replaced by

$$A = \beta_A + \beta_{AB} B, \tag{3.4}$$

$$B = \beta_B, \tag{3.5}$$

which has the same numerical solution as (3.1) and (3.2) provided that

$$\beta_A = \frac{\alpha_A}{(1 - \alpha_{BA})}, \tag{3.6}$$

$$\beta_{AB} = \frac{-1}{(1 - \alpha_{BA})}, \tag{3.7}$$

and

$$\beta_B = \alpha_A \alpha_{BA}. \tag{3.8}$$

(Nothing depends on the fact that the β_{ij} are defined in terms of the α_{ij}. We could as easily have started with Eqs. (3.4) and (3.5) 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 (3.4) and (3.5), whereas A causes B in (3.1) and (3.2). Indeed, since every linear combination of Eqs. (3.1) and (3.2) is functionally equivalent, we can easily write down systems that would be interpreted as having no causal connections or as displaying mutual causation. This is the sense in which systems

of equations are causally ambiguous, which is the rationale for supplementing them with graphs.

Simon does not appeal to graphs. Instead, he considers a higher-order 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 Eqs. (3.1) and (3.2) and can control A directly by choosing α_A and thereby control B indirectly without altering the functional form of Eq. (3.2), then the parameter set $\{\alpha_A, \alpha_{BA}\}$ is privileged. No other functionally equivalent system shares this property.

If, for example, (3.1) and (3.2) represented the true causal order, but we instead modeled the causal relationships with (3.4) and (3.5), 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 (3.6), (3.7), and (3.8). 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.

2.3 *The Structural Account of Causal Order*

I refer to an account of causal order based on Simon's seminal analysis as the *structural account*.[2] It is structural in the sense that what matters for determining the causal order is the relationship among the parameters and the variables and among the variables themselves. The parameterization – that is, the identification of privileged set of parameters that govern the functional relationships – is the source of the causal asymmetries that define the causal order. The idea of a privilege parameterization can be made more precise, by noting that a set of parameters is privileged when its members are, in the terminology of the econometricians, variation-free. A parameter is *variation-free* if, and only if, the fact that other parameters take some particular values in their ranges does not restrict the range of admissible values for that parameter.

Defining parameters as variation-free variables has a similar flavor to Hans Reichenbach's (1956) *Principle of the Common Cause*: any genuine correlation among variables has a causal explanation – either one causes the other, they are mutual causes, or they have a common cause. Since we represent causal connections as obtaining only between variables *simpliciter*, we insist that parameters not display any mutual constraints. Whereas, the Principle of the Common Cause is a metaphysical or methodological presupposition with significant bite, the variation-freeness of

---

[2] A more formal presentation of the structural account is given in Hoover (2001, Chap. 3).

parameters is only a representational convention. Any situation in which it appears that putative parameters are mutually constraining can always be rewritten so that the constraints are moved into the functional forms that connect variables to each other.

For example, in the system

$$X = a \tag{3.9}$$

$$Y = bX, \quad b \leq a \tag{3.10}$$

the parameters are not variation-free, since the choice of a constrains the value of b. However, this system can be reformulated into a related (nonlinear) system with the same solutions in which the parameters are variation-free:

$$X = a \tag{3.11}$$

$$Y = \begin{cases} bX, & \text{if } a \geq b \\ \textit{undefined}, & \text{if } a < b. \end{cases} \tag{3.12}$$

Because of its analogy with the Principle of the Common Cause, we refer to the stipulation that parameters be variation-free as the *Reichenbach Convention.*

Except for the system of Eqs. (3.11) and (3.12), we have considered only linear equations. But the structural account can accommodate nonlinearity quite generally. The key step is that parameters are not defined as coefficients uniquely associated with particular variables, as they are, for example, in path analysis, in which the parameters are merely the regression weights associated with each causal arrow.

To see the role of nonlinearity, consider a simplified example of a two-equation system from a macroeconomic model with rational expectations[3]:

$$m_t = \lambda + m_{t-1} + \varepsilon_t, \tag{3.13}$$

$$p_t = m_t + \alpha\lambda - \delta + \nu_t. \tag{3.14}$$

The subscripts are time indices. Our concern is only with the causal relationship between m_t and p_t, so the lagged value of m can be regarded as a constant. While it is not vital for our purposes, (3.13) is interpreted as a rule for fixing the money supply, while (3.14) determines the price level.

It is obvious that in Simon's framework m_t directly causes p_t. In our earlier examples, there was a simple, natural association of individual parameters with individual variables in equations written in a canonical form (causes on the right-hand side; the effect on the left-hand side; the two sides connected by an asymmetrical assignment operator ("$\Leftarrow$")).[4] But here we cannot associate the parameter λ exclusively with either equation. Indeed, the notion of a canonical form of equations is merely heuristic and must be abandoned in this case.

---

[3] The model is drawn from Hamilton (1995).

Equations displaying this sort of nonlinearity in parameters are referred to in the macroeconometrics literature as subject to "cross-equation restrictions." Suppose that the monetary authority wants to loosen monetary policy; it would increase λ. Because of the cross-equation restriction, in addition to the direct causal effect of m_t on p_t, there is a change in the functional relationship between p_t and m_t. In a stochastic version of the model, the conditional probability distribution of p_t on m_t would not be invariant to changes in m_t. This striking conclusion is well known to economists as the "Lucas critique" (Lucas 1976).[5] Economists often discuss it in terms of "deep parameters" (here α and λ) versus empirically observable coefficients (say, a regression coefficient Π, which in fact equals $\alpha\lambda - \delta$, but which is estimated as a unit). In terms of our account of causal representation, the deep parameters are just the parameters that define causal order.

While the Lucas critique is not unknown to philosophers, it is not always appreciated that it undermines any necessary connection between a well-defined causal relationship and the invariance of the probability of an effect *conditional* on its causes. Indeed, our account of causal order suggests that it is the invariance of the probability distribution of the cause (the *marginal* probability distribution) to independent changes of other causes of the effect that is the empirical hallmark of a causal relation (see Hoover 2001, Chap. 8). This claim amounts to saying that it is not the conservation of the functional relationship of causes to effect as causes vary that is most characteristic of causal relations; rather it is that effects do not flow backward against the causal arrow.

2.4 *Causal Identity*

Implicit in our discussion so far is the notion that variables in causal relationships must be causally distinct. Let us make this notion more explicit. Variables are distinguishable when we have some independent means of measuring, observing, or characterizing them. Yet variables that are distinguishable in this general way need not be causally distinct.

To take an economic example, prices (P) are distinguishable from quantities (Q), but consider the simple supply and demand model in which quantities and prices are mutually determined:

---

[4] Which in fact suggested the scheme of distinguishing parameters by subscripts: for example, α_{BC} was the parameter multiplying the variable C in the canonical equation for B.

[5] For expositions of the Lucas critique, see Hoover (1988, Chap. 8, section 8.3; 2001, Chap. 7, section 7.4).

$$Q = \alpha + \beta P \tag{3.15}$$

$$P = \delta + \gamma Q. \tag{3.16}$$

Solving (3.15) and (3.16) yields

$$Q = \frac{\alpha + \beta\delta}{1 - \beta\gamma}, \tag{3.17}$$

$$P = \frac{\delta + \alpha\gamma}{1 - \beta\gamma}. \tag{3.18}$$

Both variables are determined by the same set of parameters. It would be impossible, therefore, that we alter the value of one of them without also altering the value of the other. We might, then, regard the two variables as having a two-way or mutual causal relationship. But should we really call variables that have no causal relationships distinguishable from one another as standing in a causal relationship with each other? It would be more to the point to say that, causally speaking, there is no difference between them.

The issue arises not only in simultaneous systems of equations. Consider instead the following system:

$$A = \alpha, \tag{3.19}$$

$$B = \beta A, \tag{3.20}$$

$$C = \beta A. \tag{3.21}$$

On Simon's criterion, A clearly causes both B and C, but what is the causal relationship between B and C? It might appear to be mutual, since there is no intervention on either variable that does not alter the other. But this seems counterintuitive, because the connection is through the parameter β rather than through the variables; yet our presumption is that causal relationships are mediated only through variables. If we imagine the variable B and Eq. (3.20) eliminated, nothing would change for C.

We see, then, that some systems with or without mutual or simultaneous causation are problematic, but there is no reason to believe that problematic cases arise inevitably in simultaneous systems. We need a way of characterizing problematic and unproblematic systems. This suggests that we characterize causal identity and causal distinctiveness:

Causal Identity: Two variables are *causally identical* if aside from their mutual relationship, they have all the same causes and effects.

Causal Distinctiveness: Variables that are not causally identical are *causally distinct*.

In invoking causes and effects, these definitions are not circular, since whether or not the relevant causal relationships exist can be determined from the parameterization of the system in line with our elaboration of Simon's structural account.[6]

When variables are distinguishable because we possess independent means of measuring, observing, or characterizing them, a failure also to be causally distinct will be rare. Causal identity is more likely to be a property of an impoverished representation of the world, arising most naturally in cases in which a few variables stand in a tight relationship. Causally, identity will rarely arise in nondeterministic cases, as the variables that describe such cases are, in general, subject to "shocks" that distinguish one from another. However, models are nearly always highly simplified, and shocks that are small enough in the world may be neglected in a model, so that, if the world produces "near causal identity," a good model of the world may produce exact causal identity (cf. Suppes 1970, p. 33 on ε-direct cause). Similarly, in selecting a simplified representation of the world, we may choose to ignore some ways in which variables could be causally distinguished – again, producing causal identity in the model.

There is one type of case of causal identity that is not rare, but a pitfall to be carefully avoided. Conceptual identities or variables that are "equal by definition" belong to a special class of causal identity that does not depend on modeling choices but on the meaning of the variables. For example, the price (per dollar of coupon payment) of a perpetual bond or consol (P_C) is, by definition, the inverse of its yield (R): $P_C \equiv 1/R$. Anything that affects the yield affects the price; yet we should not regard these variables – conceptually different and with different units of measurement – as causally related. A system of equations that embedded this identity would, according to our definitions, find P_C and R to be causally identical and, therefore, not stand in any other sort of causal relationship with each other. This is exactly as it should be.

3 The Structural Account Versus the Manipulability Account of Causation

So far the discussion of causal order has been formal. But the importance of a representational scheme arises from its power to illuminate genuine scientific issues, a matter to which we now turn.

---

[6] Causal identity can be thought of as a metaphysical property of the world and as a property of a model or representation. Elsewhere I have argued in favor of a *perspectival realism* in which a successful model tells us the truth about the world from a particular point of view (Hoover 2012b, see also 2012c), which reduces the force of a distinction between the metaphysics and the properties of the model.

3.1 *Modularity*

I have referred to the development of Simon's scheme of causal representation as the *structural account* of causation. While I will not discuss it in detail here, a causal structure is, I believe, essentially what some philosophers mean when they refer to *mechanisms*. A causal model is thus the representation of the workings of a mechanism. The account is "structural," in a formal sense, in that causal order depends on nonunique functional relationships among variables acquiring a unique form or structure through the specification of the parameter space. But what are the appropriate semantics? That is, how in reference to the world – that is, in reference not to the representation but to what it represents – should we understand the parameters? A natural reading, suggested by Simon's notion of an experimenter's ability to control or intervene directly to set parameter values, is to regard parameters as the loci of interventions. Such an interpretation points to a similarity to Woodward's (2003) *manipulability account* of causation. While the similarity is genuine, we should distinguish the structural account from the manipulability account.

Although Woodward (2003, Chap. 2) provides a detailed and nuanced development of the manipulability account, the essential point is conveyed in his definition of a *direct cause*:

> **(DC)** A necessary and sufficient condition for X to be a direct causes 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)[7]

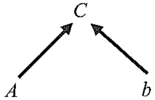
Despite Woodward (2003, p. 39) regarding causation as fundamentally a type-level relationship among variables, **(DC)** defines direct cause in terms of a token-level action – an intervention. For example, an intervention on the variable B in Fig. 3.1 would set B to a particular value, say, $B = b$, and holding it fixed at that value amounts to wiping out or breaking the arrow from A to B, indicating that no change in A is allowed to affect B. B is a direct cause of C, according to **(DC)** if C changes (or would change, the intervention being conceived of counterfactually) as a result of this intervention.

Pearl (2000, p. 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 embedded in different causal structures..." After the intervention, we can represent Fig. 3.1 with a new graph as in Fig. 3.2. The transition from one graph to the other – 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*:

---

[7] I have written V where Woodward writes **V**, to remain consistent with the notation of Sect. 2 above.

Fig. 3.2 Causal graph of Fig. 3.1 after the intervention set (B= b)



> 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)

Cartwright (2007, part II) objects to modularity as an essential feature of causation.[8] The structural account illuminates both what is right and what is wrong in Cartwright's objections. Cartwright denies that all well-defined causal systems are modular. And she is correct. We should notice, first, that the system defined by Eqs. (3.13) and (3.14), which has a well-defined causal order on the structural account, is itself not modular, since the individual equations do not represent distinct mechanism, but can function only as a pair (see Hoover, 2011, section 16.3.2). Cartwright herself argues largely through counterexamples. The first example is a carburetor (Cartwright 2007, pp. 15–16). Cartwright describes the operation of the carburetor through a system of equations in which key coefficients depend on the geometry of its chamber[9]:

> we can see a large number of [functional] laws all of which depend on the same physical features – the geometry of the carburettor. So no one of these laws can be changed on its own. To change any one requires a redesign of the carburettor, which will change the others in train. By design the different causal laws are harnessed together and cannot be changed singly. So modularity fails. [Cartwright 2007, p. 16]

Cartwright illustrates her point with a set of causal laws from which I reproduce two (in an altered notation):

$$X \Leftarrow h(G, A; \gamma), \tag{3.22}$$

$$A \Leftarrow j(S, T; \sigma), \tag{3.23}$$

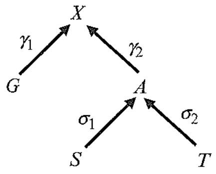
where X = gas exiting the emulsion tube; G = gas in the emulsion tube; A = air pressure in the chamber; S = suck of the pistons; T = throttle valve; and

---

[8] Cartwright's chapters are her side of a vigorous debate over modularity carried on with Hausman and Woodward (1999, 2004).

[9] I write "coefficients," not "parameters" as Cartwright does, since she assumes that they are functions of other things, violating the usage established in Sect. 2.3 above.

Fig. 3.3 The causal structure of Cartwright's carburetor



$$\gamma = \gamma(\text{geometry of chamber}, \ldots), \tag{3.24}$$

$$\sigma = \sigma(\text{geometry of chamber}, \ldots). \tag{3.25}$$

The system of equations can be represented in a causal graph as in Fig. 3.3. An oddity of Cartwright's exposition is that each of the principal equations has only one coefficient. I do not think, however, that any of her critical points hang on that, so I will treat them as vectors: $\gamma = [\gamma_1\, \gamma_2]$ and $\sigma = [\sigma_1\, \sigma_2]$. In Fig. 3.3, the elements of these vectors are listed alongside the appropriate causal arrows as indices of the strength of the influence of each cause over its effect.[10]

Cartwright's point is that modularity requires that each cause can be intervened upon separately. So, for example, if we wish to change A, we have to change σ. A change in σ can be achieved through a change in the geometry of the chamber, but that necessarily changes γ as well; so the causal relationship of G and A to X is neither distinct nor invariant with respect to that of S and T to A. Modularity fails. (As indeed it does in the closely analogous monetary-policy example of Eqs. (3.13) and (3.14).)

The representational conventions of the structural account force us to take a stand on some of the details of Cartwright's example. First, γ and σ are not parameters as we have defined the term (see fn. 8). Their interdependence violates the Reichenbach Convention. We must decide, then, whether they are variables or simply coefficients (a shorthand way of grouping parameters that interact with a variable when writing a function). Second, the "geometry of the chamber" is unlikely to be characterized by a single variable or parameter. In practice, the most natural way of representing it would be as a set of interrelated variables governed by parameters that conform to the Reichenbach Convention. Imagine representing the geometry in a computer-automated design program. The designer can set various parameters independently to generate various shapes that constitute the "geometry of the chamber." Aspects of that geometry (which can be represented as causally salient variables) are what feed into Cartwright's "causal laws" through γ and σ.

Figure 3.3, or even a more elaborate diagram, is too coarse to represent the refinements to the causal structure of the carburetor needed adequately to flesh out

---

[10] This makes the inessential, but in this case harmless, assumption that the equations are linear in variables.

Cartwright's account of its working. The rough sketch of a structural causal representation of the carburetor supports Cartwright's view that in this case, as in many others, a causally unambiguous system need not be modular.

Where Cartwright goes astray is in her belief that the failure of a well-defined causal system to be modular in Woodward's sense threatens an interventionist account and not just Woodward's particular formulation of it (the manipulability account). The structural account is a type of interventionist account that relies on a different sort of modularity – that is, modularity at the level of parameters. By definition, parameters can change independently of each other. Cartwright might object to the assumption that parameters are necessarily independent (or variation-free). But as I previously argued, this is a matter of convention; representations can always be formulated with variation-free parameters, constraints having been moved into the functional relationship of variables. As a *conventional* restriction on causal representation, however, modularity of the parameters does not pack any punch: it does not tell us that a causal mechanism can be disassembled into parts that operate independently of each other. Indeed, modularity of that sort is not conventional but substantial and highly special. The modularity of parameters is, nonetheless, the sort of modularity that we need to *define* causal structure. The structural account allows us to see that Woodward's definition of direct cause is too strong; it rules out too many relationships that are clearly causal in an obvious and practical sense.

Woodward may object to Cartwright's implicit, and the structural account's explicit, characterization of an intervention. For Woodward, an intervention is setting a variable to a value come what may – a severing of its relations to its own causes. In the structural account, an intervention is a more delicate matter of influencing a variable in some particular way by changing one or more parameters in a context in which multiple parameters connected to the variable under some functional constraints are the rule.

The failure of modularity does not depend on which notion of intervention we employ. Wiping out a causal arrow (or equation) does not necessarily leave other causal arrows intact. Consider the monetary-policy system (3.13) and (3.14) referred to in our earlier discussion of the Lucas critique. The cross-equation restriction (i.e., the appearance of λ in both equations) arises because of the assumption that agents form expectations of the path of the money supply (m_t) based on knowledge of the policy rule. Woodward's type of intervention would amount to setting m_t to a definite value independent of its past value – essentially wiping out the causal arrow from m_{t-1} to m_t. Eliminating that causal arrow does not merely imply a change in the values of the parameters of (3.14), which would be a failure of invariance of the sort highlighted by the Lucas critique and implicit in Cartwright's carburetor example, it would in fact render the parameter λ meaningless as it would undercut any basis for forming a rational expectation of the path of m_t. In effect, the wiping out of the causal arrow from m_{t-1} to m_t does not merely alter the causal arrow from m_t to p_t; it smashes it. There are plenty of real-world examples of devices in which one part cannot be removed without breaking others, which nonetheless possess well-defined causal structure.

Cartwright objects to Woodward defining direct cause by interventions that set a variable to a value come what may precisely because, as we already observed, such interventions alter the causal system (e.g., moving from Fig. 3.1 to Fig. 3.2), even when the other causal arrows (and the equations to which they correspond) are left intact. We should be concerned with the normal workings of a causal system, and the workings of some other causal system are irrelevant to them (Cartwright 2007, p. 107; Cartwright and Jones 1991).

Another of Cartwright's counterexamples to modularity – the operation of "a well-made toaster" – helps to clarify the point:

> 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)

What, we may ask, is *proper* operation? To ask such a question requires that we can distinguish changes in its state that constitute its proper operation from changes that undermine the proper operation or destroy the mechanism. Surely, such a distinction is partly a matter of perspective and often driven by pragmatic considerations. It requires that we be able to decide when the mechanism has been so altered that it is effectively a new mechanism and when the mechanism is preserved – that is, we need identity conditions for a causal system (cf. Woodward 2003, pp. 108–109).

Cartwright uses the toaster example to argue that the relevant interventions operate only within a context of a preserved mechanism and that Woodward's come-what-may interventions generally break the mechanism. While she does not provide the necessary identity conditions, they are evident in the structural account. Two mechanisms are causally the same when they have the same parameterization (i.e., the same privileged set of variation-free parameters) and differ only in the particular values that the individual parameters take within their admissible ranges. In other words, two mechanisms are causally identical when they differ not in their parameters or variables but in the token instantiations of their parameters and the token consequences of those instantiations for the variables.

Another example illustrates both the pragmatic and the conceptual issue. In the movie *The African Queen*, Charlie Allnut (Humphrey Bogart's character) runs a steamboat. From time to time the pressure in the boiler of the steam engine gets dangerously high. He hits a particular valve with a hammer, which frees up the valve, and allows the steam to escape. Later in the film, as part of his general effort to make himself and the boat more presentable to Rose Sayer (Katharine Hepburn's character), he cleans and lubricates the valve so that the pressure is released automatically without the use of the hammer.

In each case, with or without the sticky valve, the steam engine has a *typical* operation. Which is proper depends as much on Charlie's relationship with Rose as on any fact about the steam engine. Charlie cleaned up is a different man; the steam engine cleaned up is different engine. But there are important senses in which both are still the same. In the case of the engine, if not the man, we can represent the preserved mechanism as one in which there is a parameter (or parameters) that governs whether the engine is in its clean or dirty state.

We may prefer one state for pragmatic reasons and, therefore, wish to analyze the workings only within the one state. Here, John Anderson's (1938, p. 128) notion of a causal field is helpful (see also Mackie 1980, p. 35; Hoover 2001, pp. 41–49). The *causal field* consists of background conditions that, for analytical or pragmatic reasons, we would like to set aside in order to focus on some more salient causal system. We are justified in doing so when, in fact, they do not change or when the changes are causally irrelevant. In terms of representation within the structural account, setting aside causes amounts to fixing certain parameters to constant values. The effect is not unlike Pearl's or Woodward's wiping out of a causal arrow, though somewhat more delicate. The replacement of a parameter by a constant amounts to absorbing that part of the causal mechanism into the functional form that connects the remaining parameters and variables. We might, for instance, wish to conduct our analysis of the *African Queen's* steam engine entirely in the spit-and-polished state, by setting the parameter governing the state of the valve to *clean* and holding it there.

Cartwright's toaster can be treated in the same manner. A parameter might represent the state of the bolt holding the rack to the lever: when it takes the value *tight*, the operation of the toaster is "as advertized"; when *loose*, it is a little wonky; when *missing*, it does not pop up the toast at all. While there are purposes for which only *tight* matters and for which we can treat the bolt parameter as a constant with that value, impounding the state of the bolt to the causal field, it would miss a critical point not to notice that broken mechanisms are mechanisms of the same type as well-functioning mechanisms or that less refined descriptions of mechanisms are special cases of more refined descriptions. Recognizing the first is essential to the repairman; recognizing the second is essential to the design engineer.

The structural account supplemented with the notion of the causal field provides a tool through which different models, different perspectives on phenomena, may be brought into systematic relationship one to another. It also allows us to understand hierarchical relationships among causal systems stressed by Simon (1996) and Wimsatt (2007). While the examples so far involve physical mechanisms, economic examples abound. Cochrane (1998, p. 283) points out in a monetary-policy system similar to (3.13) and (3.14) that α is interpreted as the slope of the aggregate-supply curve and is typically treated by monetary economists as a parameter; yet a body of economic theory and empirical analysis treats α as a variable, determined by "deeper" parameters (Lucas 1972, 1973; Hoover 1988, Chap. 2). The monetary-policy system impounds these deeper parameters in the causal field.

The monetary-policy rule in (3.13) offers another example. As written, we can analyze the effects of different settings of the parameter λ. We can also consider an "institutional" change in which the rule is altered to depend on different conditioning variables or in different ways. As it stands, these alternatives have been impounded in the causal field. When released and represented in a model in which (3.13) is a special case, we can consider which is the best rule within the now wider class of rules, which contains the current rule as one parameterization (see Woodford 2003 for an extensive discussion of optimal monetary rules).

3.2 *Interventions and Identity*

The structural account of causal order is similar to Woodward's account in a number of ways. A key difference, however, is that direct cause is not defined with respect to a token-level notion, such as Woodward's come-what-may intervention. Direct cause is expressed instead entirely with respect to the type-level relationship between a privileged set of parameters and a functional relation representing the interrelationship among variables. The structural account, as we showed in the last section, supports a notion of the identity of causal systems: two causal systems are identical if, and only if, they differ at most by their parameterization (i.e., they differ only in the token settings of the parameters). This understanding of causal identity also suggests a different conception of an intervention.

The notion of a parameter developed in the structural account was inspired by Simon's notion of direct control and the notion that the parameter space could be thought of as the loci of direct control. Direct control is virtually indistinguishable from Woodward's notion of intervention. The only question that separates them is direct control of what? For Woodward, it is direct control of a variable; for the structural account, it is direct control of a parameter. But a parameter was defined to be a variable subject to some additional constraints; so the difference seems small. I have no doubt that the experience of manipulation and control are the source of our original intuitions about causal powers and, therefore, are important in the way that we learn about causes and learn to use causal language. Nonetheless, the structural account does not actually *use* the notion of direct control in any physical or metaphysical sense to define causal order. For Woodward an intervention involves a change to the bearer of a variable – a real entity. In contrast, the notion of a parameterization does not require that we change any parameter in a temporal or genetic sense, but merely that we consider different settings of parameters in otherwise causally identical systems.

Woodward accepts that the relevant intervention could be hypothetical and certainly need not be practically implementable (e.g., removing the moon to discover its effect on the tides would be an acceptable, but hypothetical, intervention). Nonetheless, the counterfactual that is entertained is still a particular token change to a particular entity. In contrast, the structural account is a thoroughly type-level account. The parameter space is the loci of possible interventions; nevertheless, it is the topology that the parameterization imposes on the variables, rather than any – even hypothetical – selection of particular parameter values that defines the causal structure.

The difference between Woodward's account and the structural account is illustrated in comparative static analysis – a technique familiar to economists. For example, we might ask, what is the effect *ceteris paribus* of a higher rate of inflation on the level of prices? The answer given by the quantity theory of money is that the level of prices is lower for the same quantity of money in circulation when their rate of change is higher (Cagan 1956). The experiment cannot be conducted on a single, actual economy since an increase in the rate of inflation necessarily increases the price level. In principle, we could address the question by considering two economies with identical causal structures differing only in the parameterization necessary to produce distinct inflation rates and evaluate their price levels at the time that their money stocks happen to be equal. (This is, of course, a practical impossibility; we do not have a box of causally identical economies to draw from, but difficulty is different from the in-principle impossibility of changing the inflation rate without changing the price level.)[11] Such comparative static analysis may be relevant to actual economies in just those circumstances that some causal channels are so weak that we can neglect them (or impound them in the causal field) or that we can account for them by conditioning. These are exactly the strategies that Cagan (1956) attempts to implement in his famous paper on hyperinflations.

Comparative statics in well-formulated models are examples of a kind of possible-worlds analysis in which the connection to our world – or at least to the model that we take to best represent our world – is substantially more precise than the metrics proposed by Lewis (1973, 1979). Counterfactual analysis and which counterfactuals are sensible to address which particular problems is straightforward in such models (see Hoover 2011).

Something like comparative static analysis is critical to design and engineering. One might, for example, want to understand the difference in behavior of cars of the same model, differing only in having either a four- or a six-cylinder engine.

Motivated in part by his particular conception of an intervention and in part by an essentialist ontology, Woodward rejects the notion that such questions are properly causal with respect to some properties, including race, sex, and species:

> the notion of an intervention will not be well-defined if there is no well-defined notion of changing the values of that variable. Suppose that we introduce a variable "animal" which takes the values {*lizard, kitten, raven*}... we have no coherent idea of what it is to change a raven into lizard or kitten. Of course, we might keep a raven in a cage and replace it with a

---

[11] My view here as a shift from my earlier understanding of the causal significance of comparative static analysis (Hoover 2001, p. 102).

> lizard or a kitten, but this is not to change one of these animals into another. What is changed in this case is the content of the cage, not the animals themselves. (Woodward 2003, p. 113)

The issue arises in economic and sociological research when, for example, race discrimination in mortgage applications or sex discrimination in employment is assessed by sending applicants who have been matched as thoroughly as feasible for salient characteristics, differing significantly only in race or sex, through a mortgage qualification or hiring process. With respect to sex discrimination, Woodward (2003, p. 115) rejects the claim "[b]eing female causes one to be discriminated against in hiring and/or salary" as "fundamentally unclear" since "we lack any clear idea of what it would be like to manipulate it." Woodward argues that it is not the applicant's sex but the employer's beliefs about them that is causal.

How are sex and race different from considering the causal outcomes of cars that differ only by their engine type? Consider a coin-sorting machine – coin discrimination being less emotive than sex or race discrimination. Different coins can be placed into the machine and fall into different slots depending on their shapes.

The mechanically relevant description of a coin is as a vector of variables – for example [*diameter*, *thickness*, *weight*]. What else is it to be a coin other than having the right values in such a vector? The social metaphysical answer might be that the essence of being, say, a nickel is to have the imprimatur of the government (to have been dubbed legal tender for $0.05 by the United States Mint) and to have an appropriate standing in the social practice of money. In other contexts, such considerations may be of genuine interest, but they are not *mechanically* salient.

For a causal understanding for purposes of designing, building, operating, maintaining, and repairing the coin sorter, we do not need a penny to change into a nickel. Each penny, nickel, dime, and quarter is just *the coin in the machine* – an instantiation of a vector-valued variable and its causal fate is a realization of the causal process represented by the causal connections among that variable and those that describe the state of the machine.

We need to distinguish between existential and causal identity. Each may be salient in different contexts. The critical question is which context is relevant for what purposes. We have no notion of how to turn a penny into a nickel, but we have a clear notion of how to change a coin from being a nickel to being a penny in the context of the operation of a coin sorter, and it seems perfectly sensible to say that it is "being a nickel" that is a cause of the coin falling into slot 3. Slot 3 being configured in the right way is, of course, another cause, the causes interacting according to the design of the machine.

How is sex discrimination different in principle from coin discrimination? In the hiring process, we can represent people as a vector-valued variable *Applicant* = [*Race, Sex, Age, Employment Status, Wealth,...*], which interacts with variables describing the other factors and processes related to hiring. We do not need to change a particular person from male to female to understand this causal process. It is enough to reparameterize the vector. Much of the effort in conducting discrimination research using such techniques goes in to establishing the relevant causal identity by closely matching relevant characteristics of different applicants – other than the target characteristic of sex or race. Significant knowledge is obtainable in such ways and there is no need to stigmatize its causal *bona fides*.

Woodward's argument that it is the beliefs of the employer not the being female that is the relevant causal variable does not persuade. The beliefs of employers may be causes of the detailed outcomes of the discrimination – for example, not hiring or paying a lower salary. The relevant question here, however, is what causes those beliefs. The most common reason that we believe someone to be female is that she *is* female. Woodward could object that it is not being female that causes the belief but the appearance of being female. But this is just the analogue of saying that it is not the "moneyness" of the nickel, but its physical characteristics that cause it to fall into slot 3. The physical characteristics *are* the causally relevant ones. What is it to be female in a causally relevant sense? It is to have a sufficient number of stereotypical female characteristics. For purposes of the employment process, a sufficiently plausible transvestite counts as female. Characteristics of actual females determine the female stereotype – that is, why the values of certain variables bundled in a certain way convey the appearance of femaleness. And without the stereotype, variables with those values would not be causally salient. (The same issue arises with the coin sorter: a slug is to a nickel as a transvestite is to a female.)

Some entities may have an essence that cannot be changed while maintaining existential identity. But if, as the structural account would have it, token manipulations are not essential to defining cause, then it is better in these cases to say that causal identity is determined by possession of the explicitly causally relevant characteristics and not by some *sine qua non*. It is also pragmatically superior if we accept the view that knowledge is acquired in a piecemeal fashion. As we have seen, variables such as sex or race are avoided through substantial extra articulation and refinement of causal mechanisms, which may lack clear conceptual or evidential foundations. What, for example, are the detailed mechanisms by which the appearance of being female translates into beliefs and how are they causally relevant? We may not have grounds for knowing such details and we may not need to know them to know what is pragmatically relevant about discrimination. And if I am correct that the coin sorter is really no different in principle than the sex discrimination case, then we would be forced to look for such extra articulation in a vast array of cases. But if the structural account is plausible, we can do happily without it and without the troubling requirement of token intervention.

4 Causation and Representation

The focus of this chapter has been on representing causes. Of course, the ultimate importance of causal analysis is not located in its representation but in discovering what causal relationships actually obtain in the world and using those causes to control the world for desirable ends – what Cartwright (2007) refers to as "hunting" and "using" causes. Either activity, however, is greatly furthered by a good representation: think of the utility of having a good wanted poster or a good set of blueprints. What I hope to have shown in this chapter is that a relatively straightforward system of causal representation can be useful in understanding and resolving substantive debates in the philosophy of causation, as well as in the actual applications of causal inference and manipulation in economics.

References

Anderson, John. 1938. *The problem of causality*. Reprinted in *Studies in empirical philosophy*. Sydney: Angus Robertson, 1962.

Cagan, Philip. 1956. The monetary dynamics of hyperinflation. In *Studies in the quantity theory of money*, ed. Milton Friedman. Chicago: University of Chicago Press.

Cartwright, Nancy. 1999. *The dappled world: A study of the boundaries of science*. Cambridge: Cambridge University Press.

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

Cartwright, Nancy, and Martin Jones. 1991. How to hunt quantum causes. *Erkenntnis* 35: 205–231.

Cochrane, John H. 1998. What do the VARs mean? Measuring the output effects of monetary policy. *Journal of Monetary Economics* 41(2): 277–300.

Giere, Ronald N. 2006. *Scientific perspectivism*. Chicago: University of Chicago Press.

Haavelmo, Trgyve. 1944. The probability approach in econometrics. *Econometrica* 12 (supplement): 1–115.

Hamilton, James D. 1995. Rational expectations and the econometric consequences of changes in regime. In *Macroeconometrics: Developments, tensions, and prospects*, ed. Kevin D. Hoover, 325–345. Boston: Kluwer.

Hausman, Daniel M. 1998. *Causal asymmetries*. Cambridge: Cambridge University Press.

Hausman, Daniel M., and James Woodward. 1999. Independence, invariance, and the causal Markov condition. *The British Journal for the Philosophy of Science* 50(4): 521–583.

Hausman, Daniel M., and James Woodward. 2004. Modularity and the causal Markov condition: A restatement. *The British Journal for the Philosophy of Science* 55: 147–161.

Hood, William C., and Tjalling C. Koopmans (eds.). 1953. *Studies in econometric method*, Cowles Commission Monograph 14. New York: Wiley.

Hoover, Kevin D. 1988. *The new classical macroeconomics: A sceptical inquiry*. Oxford: Blackwell.

Hoover, Kevin D. 2001. *Causality in macroeconomics*. Cambridge: Cambridge University Press.

Hoover, Kevin D. 2004. Lost causes. *Journal of the History of Economic Thought* 26(2): 149–164.

Hoover, Kevin D. 2008. Causality in economics and econometrics. In *The New Palgrave dictionary of economics*, 2nd ed, ed. Steven N. Durlauf and Lawrence E. Blume. New York: Palgrave Macmillan.

Hoover, Kevin D. 2009. Milton Friedman's stance: The methodology of causal realism. In *The methodology of positive economics: Milton Friedman's essay fifty years later*, ed. Uskali Mäki, 303–320. Cambridge: Cambridge University Press.

Hoover, Kevin D. 2011. Counterfactuals and causal structure. In *Causality in the sciences*, ed. Phyllis McKay Illari, Federica Russo, and Jon Williamson, 338–360. New York: Oxford University Press.

Hoover, Kevin D. 2012a. Economic theory and causal inference. In *Handbook of the philosophy of economics*, ed. Uskali Mäki. Amsterdam: Elsevier/North-Holland.

Hoover, Kevin D. 2012b. Pragmatism, perspectival realism, and econometrics. In *Economics for real: Uskali Mäki and the place of truth in economics*, ed. Aki Lehtinen, Jaakko Kuorikoski, and Petri Ylikoski. London: Routledge.

Hoover, Kevin D. 2012c. Causal structure and hierarchies of models. In *Studies in History and Philosophy of Science (Part A)* 43(4): 778–786.

Koopmans, Tjalling C. 1950. *Statistical inference in dynamic economic models*, Cowles Commission Monograph 10. New York: Wiley.

Lewis, David. 1973. Causation. *Journal of Philosophy* 70: 556–567.

Lewis, David. 1979. Counterfactual dependence and time's arrow. *Noûs* 13(4): 455–476.

Louçã, Francisco. 2007. *The years of high econometrics: A short history of the generation that reinvented economics*. London: Routledge.

Lucas Jr., Robert E. 1972. Expectations and the neutrality of money. *Journal of Economic Theory* 4: 103–124.

Lucas Jr., Robert E. 1973. Some international evidence on output-inflation tradeoffs. *American Economics Review* 63: 326–334.

Lucas Jr., Robert E. 1976. *Econometric policy evaluation: A critique*. Reprinted in Lucas, *Studies in business cycle theory*, 104–130. Oxford: Blackwell.

Mackie, John L. 1980. *The cement of the universe: A study in causation*, 2nd ed. Oxford: Clarendon.

Pearl, Judea. 2000. *Causality: Models, reasoning, and inference*. Cambridge: Cambridge University Press.

Reichenbach, H. 1956. *The direction of time*. Berkeley: University of California Press.

Reiss, Julian. 2007. *Error in economics: The methodology of evidence-based economics*. London: Routledge.

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

Simon, Herbert A. 1953. Causal order and identifiability. in Hood and Koopmans (1953). Page numbers refer the reprint in Simon (1957). *Models of man*. New York: Wiley.

Spirtes, Peter, Clark Glymour, and Richard Scheines. 2000. *Causation, prediction, and search*, 2nd ed. Cambridge, MA: MIT Press.

Suppes, Patrick. 1970. *A probabilistic theory of causality*. Amsterdam: North-Holland.

Teller, Paul. 2001. Twilight of the perfect model model. *Erkenntnis* 55(3): 393–415.

Wimsatt, William C. 2007. *Re-engineering philosophy for limited beings: Piecewise approximations to reality*. Cambridge, MA: Harvard University Press.

Woodford, Michael. 2003. *Interest and prices: Foundations of a theory of monetary policy*. Princeton: Princeton University Press.

Woodward, James 2003. *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.