Action as a Local Surgery

Judea Pearl
Cognitive Systems Laboratory
Computer Science Department
University of California, Los Angeles, CA 90024
judea@cs.ucla.edu

Introduction

What gives us the audacity to expect that actions should have neat and compact representations? Why did the authors of STRIPS [Fikes & Nilsson, 1971] and BURIDAN [Kushmerick et al., 1993] believe they could get away with such short specification for actions?

Whether we take the probabilistic paradigm that actions are transformations from probability distributions to probability distributions, or the deterministic paradigm that actions are transformations from states to states, such transformations could in principle be infinitely complex. Yet, in practice, people teach each other rather quickly what actions normally do to the world, people predict the consequences of any given action without much hustle, and AI researchers are writing languages for actions as if it is a God given truth that action representation should be compact, elegant and meaningful. Why?

The paradigm I wish to explore in this paper is that these expectations are not only justified but, mainly, that once we understand the justification, we will be in better shape to craft effective representations for actions.

Mechanisms and surgeries

Why are the expectations justified? Because the actions we normally invoke in common reasoning tasks are *local surgeries*. The world consists of a huge number of autonomous and invariant linkages or mechanisms (to use Simon's word), each corresponding to a physical process that constrains the behavior of a relatively small groups of variables. In principle, then, the formalization of actions should not be difficult. If we understand how the linkages interact with each other, usually they simply share variables, we should also be able to understand what the effect of an action would be: Simply re-specify those few mechanisms that are perturbed by the action, then let the modified population of mechanisms interact with one another, and see what state will evolve at equilibrium. If the new specification is complete, a single state will evolve. If the specification is probabilistic, a new probability distribution will emerge and, if the specification is logical (possibly incomplete) a new, mutilated logical theory will then be created, capable of answering queries about post-action states of affair.

If this sounds so easy, why did AI ever get into trouble in the arena of action representation? The first answer I wish to explore is that what is local in one space may not be local in another. A speck of dust, for example, appears extremely diffused in the frequency (or Fourier) representation and, vice versa, a pure musical tone requires a long stretch of time to be appreciated. It is important therefore to emphasize that actions are local in the space of mechanisms and not in the space of variables or sentences or time slots. For example, tipping the left-most object in an array of domino tiles does not appear "local" in the spatial representation, because, in the tradition of domino theories, every tile might be affected by such action. Yet the action is quite local in the mechanism domain: Only one mechanism gets perturbed, the gravitational restoring force which normally keeps the left-most tile in a stable erect position. It takes no more than a second to describe this action on the phone, without enumerating all its ramifications. The listener, assuming she shares our understanding of domino physics, can figure out for herself the ramifications of this action, or any action of the type: "tip the *ith* domino tile to the right". By representing the domain in the form of an assembly of stable mechanisms, we have in fact created an oracle capable of answering queries about the effects of a huge set of actions and action combinations, without us having to explicate those effects.

Laws vs. facts

This surgical procedure still sounds easy and does not explain why AI got into trouble with action representation. The trouble begins with the realization that in order to implement surgical procedures in mechanism space, we need a language in which some sentences are given different status than others; sentences describing mechanisms should be treated differently than those describing other facts of life, such as observations, assumption and conclusions, because the former are presumed stable, while the latter are transitory. Indeed the mechanism which couples the state of the $(i+1)th$ domino tile to that of the *ith* domino tile remains unaltered (unless we set them apart by some action) whereas the states of the tiles themselves are free to vary with circumstances.

Admitting the need for this distinction has been a difficult cultural transition in the logical approach to actions, perhaps because much of the power of classical logic stems from its representational uniformity and syntactic invariance, where no sentence commands special status. Probabilists were much less reluctant to embrace the distinction between laws and facts, because this distinction has already been programmed into probability language by Reverend Bayes in 1763: Facts are expressed as ordinary propositions, hence they can obtain probability values and they can be conditioned on; laws, on the other hand, are expressed as conditional-probability sentences (e.g., $P(accident|careless\text{-}driving) = high$), hence they should not be assigned probabilities and cannot be conditioned on. It is due to this tradition that probabilists have always attributed nonpropositional character to conditional sentences (e.g., birds fly); refusing to allow nested conditionals [Levi, 1988], and insisting on interpreting probabilities of conditionals as conditional probabilities [Adams, 1975, Lewis, 1976]. Remarkably, these constraints, which some philosophers view as limitations, are precisely the safeguards that have kept probabilists from confusing laws and facts, and have protected from some of the traps that have lured logical approaches.[1]

Mechanisms and causal relationships

The next issue worth discussing is how causality enters into this surgical representation of actions. To understand the role of causality, we should note that most mechanisms do not have names in common everyday language. In the domino example above I had to struggle hard to name the mechanism which would be perturbed by the action "tip the left-most tile to the right". And there is really no need for the struggle; instead of telling you the name of the mechanism to be perturbed by the action, I might as well gloss over the details of the perturbation processes and summarize its net result in the form of an *event*, e.g., "the left-most tile is tipped to right", which yields equivalent consequences as the perturbation summarized. After all, if you and I share the same understanding of physics, you should be able to figure out for yourself which mechanism it is that must be perturbed in order to realize the specified new event, and this should enable you to predict the rest of the scenario.

This linguistic abbreviation defines a new relation among events, a relation we normally call "causation": Event A causes B, if the perturbation needed for realizing A entails the realization of B.[2] Causal abbreviations of this sort are used very effectively for specifying domain knowledge. Complex descriptions of what relationships are stable and how mechanisms interact with one another are rarely communicated explicitly in terms of mechanisms . Rather, they are communicated in terms of cause-effect relationships between events or variables. We say, for example: "If tile i is tipped to the right, it causes tile $i+1$ to tip to the right as well"; we do not communicate such knowledge in terms of the tendencies of each domino tile to maintain its physical shape, to respond to gravitational pull and to obey Newtonian mechanics.

A formulation of action as a local surgery on causal theories has been developed in a number of recent papers [Goldszmidt & Pearl, 1992; Pearl, 1993a; Pearl, 1993b; Darwiche & Pearl, 1994; Pearl, 1994a; Goldszmidt & Darwiche, 1994]. The appendix provides a brief summary of this formulation, together with a simple example that illustrates how the surgery semantics generalizes to nonprobabilistic formalisms.

Causal ordering

Our ability to talk directly in terms of one event causing another, (rather than an action altering a mechanism and the alteration, in turn, having an effect) is computationally very useful, but, at the same time it requires that the assembly of mechanisms in our domain satisfy certain conditions. Some of these conditions are structural, nicely formulated in Simon's "causal ordering" [Simon, 1953], and others are substantive - - invoking relative magnitudes of forces and powers.

The structural requirement is that there be a one-to-one correspondence between mechanisms and variables – a unique variable in each mechanism is designated as the output (or effect), and the other variables, as inputs (or causes). Indeed, the definition of causal theories given in the appendix assumes that each equation is associated with a unique variable, situated on its left hand side. In general, a mechanism may be specified as a function

$$G_i(X_1, ..., X_n; U_1, \ldots, U_m) = 0$$

without identifying any so called "dependent" variable X_i. Simon's causal ordering is a procedure for deciding whether a collection of such G functions has a unique preferred way of associating variables with mechanisms, based on the requirement that we should be able to solve for the *ith* variable without solving for its successors in the ordering.

In certain structures, called *webs* [Dalkey, 1994, Dechter & Pearl, 1991], Simon's causal ordering determines a unique one-to-one correspondence, but in others, such as those involving feedback, the correspondence is not unique. Yet in examining feedback circuits, for example, people can assert categorically that the flow of causation goes clockwise, rather than counterclockwise. They make such assertions on the basis of relative magnitudes of forces; for example, it takes very little energy to make an input of a gate change its output, but no force applied to the output can influence the input. When such considerations are available, causal directionality can be determined by appealing again to the notion of hypothetical intervention and asking whether an external control over one variable in the mechanism necessarily affects the others. The variable which does not affect any of the others is the dependent variable. This then constitutes the operational semantics for identifying the dependent variables X_i in nonrecursive causal theories (see appendix).

[1]The distinction between laws and facts has been proposed by Poole (1985) and Geffner (1992) as a fundamental principle for nonmonotonic reasoning. It seems to be gaining broader support recently as a necessary requirement for formulating actions (see title of Gelfond's paper, this volume).

[2]The word "needed" connotes minimality and can be translated to: "...if every minimal perturbation realizing A, entails B".

Imaging vs. conditioning

If action is a transformation from one probability function to another, one may ask whether every transformation corresponds to an action, or are there some constraints that are peculiar to exactly those transformations that originate from actions. Lewis (1976) formulation of counterfactuals indeed identifies such constraints: the transformation must be an *imaging* operator (Imaging is the probabilistic version of Winslett-Katsuno-Mendelzon possible worlds representation of "update").

Whereas Bayes conditioning $P(s|e)$ transfers the entire probability mass from states excluded by e to the remaining states (in proportion to their current $P(s)$), imaging works differently; each excluded state s transfers its mass individually to a select set of states $S^*(s)$, which are considered "closest" to s. The reason why imaging is a more adequate representation of transformations associated with actions can be seen more clearly through a representation theorem due to Gardenfors [1988, Theorem 5.2 pp.113] (strangely, the connection to actions never appears in Gardenfors' analysis). Gardenfors' theorem states that a probability update operator $P(s) \rightarrow P_A(s)$ is an imaging operator iff it preserves mixtures, i.e.,

$$[\alpha P(s) + (1-\alpha)P'(s)]_A = \alpha P_A(s) + (1-\alpha)P'_A(s) \quad (1)$$

for all constants $1 > \alpha > 0$, all propositions A , and all probability functions P and P'. In other words, the update of any mixture is the mixture of the updates[3].

This property, called homomorphism, is what permits us to specify actions in terms of *transition probabilities*, as it is usually done in stochastic control and Markov decision process. Denoting by $P_A(s|s')$ the probability resulting from acting A on a known state s', homomorphism (1) dictates:

$$P_A(s) = \sum_{s'} P_A(s|s')P(s')$$

saying that, whenever s' is not known with certainty, $P_A(s)$ is given by a weighted sum of $P_A(s|s')$ over s', with the weight being the current probability function $P(s')$.

This characterization, however, is too permissive; while it requires any action-based transformation to be describable in terms of transition probabilities, it also accepts any transition probability specification, however whimsical as a descriptor of some action. The valuable information that actions are defined as *local* surgeries, is totally ignored in this characterization. For example, the transition probability associated with the atomic action $A_i = do(X_i = x_i)$ originates from the deletion of just one mechanism in the assembly. Hence, one would expect that the transition probabilities associated with the set of atomic actions would not be totally arbitrary but would constrain one another.

We are currently exploring axiomatic characterizations of such constraints which we hope to use as a logic for sentences of the type: "X affects Y when we hold Z fixed". With the help of such logic we hope to be able to derive, refute or confirm sentences such as "If X has no effect on Y and Z affects Y, then Z will continue to affect Y when we fix X." The reader might find some challenge proving or refuting the sentence above, that is, testing whether it holds in any causal theory, when "affecting" and "fixing" are interpreted by the local-surgery semantics described in this paper.

[3]Assumption (1) is reflected in the (U8) postulate of [Katsuno & Mendelzon, 1991]: $(K_1 \vee K_2) o \mu = (K_1 o \mu) \vee (K_2 o \mu)$, where o is an update operator.

Appendix: Causal theories, subtheories, actions and causal effects

Definition 1 *A causal theory is a n-tuple*

$$T = < V, U, P(\mathbf{u}), \{f_i\} >$$

where

(i) $V = \{X_1, \ldots, X_n\}$ *is a set of observed variables*

(ii) $U = \{U_1, \ldots, U_m\}$ *is a set of unobserved variables which represent disturbances, abnormalities or assumptions,*

(iii) $P(\mathbf{u})$ *is a distribution function over* $U_1, \ldots, U_m$, *and*

(iv) $\{f_i\}$ *is a set of n deterministic functions, each of the form*

$$X_i = f_i(X_1, \ldots, X_n, U_1, \ldots, U_m) \quad i = 1, \ldots, n$$

We will assume that the set of equations in (iv) has a unique solution for $X_i, \ldots, X_n$, *given any value of the disturbances* $U_1, \ldots, U_m$. *Therefore the distribution* $P(\mathbf{u})$ *induces a unique distribution on the observables, which we denote by* $P_T(\mathbf{v})$.

We will consider concurrent actions of the form $do(X = x)$, where $X \subseteq V$ is a set of variables and x is a set of values from the domain of X. In other words, $do(X = x)$ represents a combination of actions that forces the variables in X to attain the values x.

Definition 2 *(Effect of actions) The effect of the action* $do(X = x)$ *on a causal theory* T *is given by a subtheory* T_x *of* T, *where* T_x *obtains by deleting from* T *all equations corresponding to variables in* X *and substituting the equations* $X = x$ *instead.*

Definition 3 *(causal effect) Given two disjoint subsets of variables,* $X \subseteq V$ *and* $Y \subseteq V$, *the causal effect of* X *on* Y, *denoted* $P_T(y|\hat{x})$, *is a function from the domain of* X *to the space of probability distributions on* Y, *such that*

$$P_T(y|\hat{x}) = P_{T_x}(y)$$

for each realization x *of* X. *In other words for each* $x \in dom(X)$, *the causal effect* $P_T(y|\hat{x})$ *gives the distribution of* Y *induced by the action* $do(X = x)$.

Note that causal effects are defined relative to a given causal theory T; *the subscript* T *is often suppressed for brevity.*

Figure 1 illustrates a simple causal theory in the form of a diagram. It describes the causal relationships among the season of the year (X_1), whether rain falls (X_2), whether the sprinkler is on (X_3), whether the pavement would get wet (X_4), and whether the pavement would be slippery (X_5). All variables in this figure are binary, taking a value of either true or false, except the root variable X_1 which can take one of four values: Spring, Summer, Fall, or Winter. Here, the absence of a direct link between X_1 and X_5, for example, captures our understanding that the influence of seasonal variations on the slipperiness of the pavement is mediated by other conditions (e.g., the wetness of the pavement).

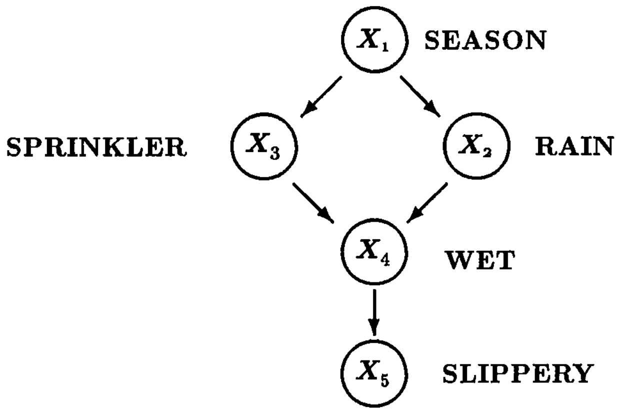

Figure 1: A diagram representing a causal theory on five variables.

The theory corresponding to Figure 1 consists of five functions, each representing an autonomous mechanism:

$$
\begin{aligned}
X_1 &= U_1 \\
X_2 &= f_2(X_1, U_2) \\
X_3 &= f_3(X_1, U_3) \\
X_4 &= f_4(X_3, X_2, U_4) \\
X_5 &= f_5(X_4, U_5) \qquad\qquad (A-1)
\end{aligned}
$$

The probabilistic analysis of causal theories becomes particularly convenient when two conditions are satisfied:

1. The theory is recursive, i.e., there exists an ordering of the variables $V = \{X_1, \ldots, X_n\}$ such that each X_i is a function of a subset $\mathbf{pa}_i$ of its predecessors

$$X_i = f_i(\mathbf{pa}_i, U_i), \qquad \mathbf{pa}_i \subseteq \{X_1, \ldots, X_{i-1}\}$$

2. The disturbances $U_1, \ldots, U_n$ are mutually independent and exogenous, i.e.,

$$U_i \perp\!\!\!\perp \{X_1, \ldots, X_{i-1}\}$$

These two conditions, also called Markovian, are the basis of Bayesian networks [Pearl, 1988] and they enable us to compute causal effects directly from the conditional probabilities $P(x_i|\mathbf{pa}_i)$, without specifying the functional form of the functions f_i, or the distributions $P(u_i)$ of the disturbances.

In our example, the distribution induced by the theory T is given by the product

$$P_T(x_1, \ldots, x_n) = \prod_i P(x_i|\mathbf{pa}_i)$$

where $\mathbf{pa}_i$ are the direct predecessors (called *parents*) of X_i in the diagram, giving

$$
\begin{aligned}
&P_T(x_1, x_2, x_3, x_4, x_5) \\
&\quad = P(x_1)P(x_2|x_1)P(x_3|x_1)P(x_4|x_2, x_3)P(x_5|x_4)
\end{aligned}
$$

To represent the action "turning the sprinkler ON", $do(X_3 = \text{ON})$, we delete the equation $X_3 = f_3(x_1, u_3)$ from the theory of Eq. (A-1), and replace it with $X_3 = \text{ON}$. Graphically, this surgery corresponds to deleting the link $X_1 \rightarrow X_3$ and fixing the value of X_3 to ON. The resulting joint distribution on the remaining variables will be

$$
\begin{aligned}
&P_T(x_1, x_2, x_4, x_5|do(X_3 = \text{ON})) \\
&\quad = P(x_1)\ P(x_2|x_1)\ P(x_4|x_2, X_3 = \text{ON})\ P(x_5|x_4)
\end{aligned}
$$

Note the difference between the action $do(X_3 = \text{ON})$ and the observation $X_3 = \text{ON}$. The latter is encoded by ordinary Bayesian conditioning, while the former by conditioning a mutilated graph, with the link $X_1 \rightarrow X_3$ removed. This mirrors indeed the difference between seeing and doing: after observing that the sprinkler is ON, we wish to infer that the season is dry, that it probably did not rain, and so on; no such inferences should be drawn in evaluating the effects of the deliberate action "turning the sprinkler ON". The amputation of $X_3 = f_3(X_1, U_3)$ from (A-1) ensures the suppression of any abductive inferences from any of the action's consequences.

The surgical procedure described above is not limited to probabilistic analysis. The causal knowledge represented in Figure 1 can be captured by logical theories as well, for example,

$$
\begin{aligned}
x_2 &\iff [(X_1 = \text{Winter}) \vee (X_1 = \text{Fall}) \vee ab_2] \wedge \neg ab'_2 \\
x_3 &\iff [(X_1 = \text{Summer}) \vee (X_1 = \text{Spring}) \vee ab_3] \wedge \neg ab'_3 \\
x_4 &\iff (x_2 \vee x_3 \vee ab_4) \wedge \neg ab'_4 \\
x_5 &\iff (x_4 \vee ab_5) \wedge \neg ab'_5
\end{aligned}
$$

where x_i stands for $X_i = true$, and ab_i and ab'_i stand, respectively, for trigerring and inhibiting abnormalities. The double arrows represent the assumption that the events on the r.h.s. of each equation are the *only* causes for the l.h.s.

It should be emphasized though that the models of a causal theory are not made up merely of truth value assignments which satisfy the equations in the theory. Since each equation represents an autonomous process, the scope of each individual equation must be specified in any model of the theory, and this can be encoded using either the graph (as in Figure 1) or the generic description of the theory, as in (A-1). Alternatively, we can view a model of a causal theory to consist of a mutually consistent set of submodels, with each submodel being a standard model of a single equation in the theory.

Historical background

An explicit translation of interventions to "wiping out" equations from linear econometric models was first proposed by Strotz & Wold (1960) and later used in Fisher (1970) and Sobel (1990) . Extensions to action representation in nonmonotonic systems are reported in [Goldszmidt & Pearl, 1992, Pearl, 1993a]. Graphical ramifications of this translation were explicated first in Spirtes et al. (1993) and later in Pearl (1993b). Calculi for actions and counterfactuals based on this interpretation are developed in [Pearl, 1994b] and [Balke & Pearl, 1993], respectively.

References

[Adams, 1975] Adams, E., *The Logic of Conditionals,* Chapter 2, D. Reidel, Dordrecht, Netherlands, 1975.

[Balke & Pearl, 1993] Balke, A. and Pearl, J., "Probabilistic evaluation of counterfactual queries," in *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-94)*, Seattle, WA, Volume I, 230-237, July 31 - August 4, 1994. See also Balke, A., and Pearl, J., "Counterfactual probabilities: Computational methods, bounds, and applications," in R. Lopez de Mantaras and D. Poole (Eds.), *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI-94)*, Morgan Kaufmann, San Mateo, CA, 46-54, July 29-31, 1994.

[Dalkey, 1994] Dalkey, N., "Webs," UCLA Cognitive Systems Laboratory, *Technical Report (R-166),* Computer Science Department, University of California, Los Angeles, March 1994.

[Darwiche & Pearl, 1994] Darwiche, A., and Pearl, J., "Symbolic causal networks for planning under uncertainty," In *Symposium Notes of the 1994 AAAI Spring Symposium on Decision-Theoretic Planning*, Stanford, CA, 41-47, March 21-23, 1994.

[Dechter & Pearl, 1991] Dechter, R. and Pearl, J., "Directed constraint networks: A relational framework for Causal Modeling," in *Proceedings, 12th International Joint Conference of Artificial Intelligence (IJCAI-91)*, Sydney, Australia, 1164-1170, August 24-30, 1991,

[Fikes & Nilsson, 1971] Fikes, R.E. and Nilsson, N.J., "STIRPS: A new approach to the application of theorem proving to problem solving," *Artificial Intelligence* 2(3/4), 189–208, 1971.

[Fisher, 1970] Fisher, F.M., "A correspondence principle for simultaneous equation models," *Econometrica*, 38, 73–92, 1970.

[Gardenfors, 1988] Gardenfors, P., *Knowledge in Flux: Modeling the Dynamics of Epistemic States*, MIT Press, Cambridge, MA, 1988.

[Geffner, 1992] Geffner, H.A., *Default Reasoning: Causal and Conditional Theories*, MIT Press, Cambridge, MA, 1992.

[Goldszmidt & Darwiche, 1994] Goldszmidt, M. and Darwiche, A., "Action networks: A framework for reasoning about actions and change under uncertainty," in R. Lopez de Mantaras and D. Poole (Eds.), *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence (UAI-94)*, Morgan Kaufmann, San Mateo, CA, 136–144, 1994.

[Goldszmidt & Pearl, 1992] Goldszmidt, M. and Pearl, J., "Rank-based systems: A simple approach to belief revision, belief update, and reasoning about evidence and actions," in B. Nebel, C. Rich, and W. Swartout (Eds.), *Proceedings of the Third International Conference on Knowledge Representation and Reasoning,* Morgan Kaufmann, San Mateo, CA, 661-672, October 1992.

[Katsuno & Mendelzon, 1991] Katsuno, H. and Mendelzon, A., "On the difference between updating a knowledge base and revising it," in *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, Boston, MA, 387–394, 1991.

[Kushmerick et al., 1993] Kushmerick, N., Hanks, S., and Weld, D., "An algorithm for probabilistic planning," *Technical Report 93-06-03*, Department of Computer Science and Engineering, University of Washington, 1993.

[Lewis, 1976] Lewis, D., "Probabilities of conditionals and conditional probabilities," *Philosophical Review*, Vol. 85, 297–315, 1976.

[Levi, 1988] Levi, I., "Iteration of conditionals and the Ramsey test," *Synthese*, 76, 49–81, 1988.

[Pearl, 1988] Pearl, J., *Probabilistic Reasoning in Intelligence Systems*, Morgan Kaufmann, San Mateo, CA, 1988.

[Pearl, 1993a] Pearl, J., "From Conditional Oughts to Qualitative Decision Theory" in D. Heckerman and A. Mamdani (Eds.), *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence*, Washington, D.C., Morgan Kaufmann, San Mateo, CA, 12–20, July 1993.

[Pearl, 1993b] Pearl, J., "Graphical models, causality, and intervention," *Statistical Science*, 8 (3), 266–273, 1993.

[Pearl, 1994a] Pearl, J., "From Adams' conditionals to default expressions, causal conditionals, and counterfactuals," in E. Eells and B. Skyrms (Eds.), *Probability and Conditionals*, Cambridge University Press, New York, NY, 47-74, 1994.

[Pearl, 1994b] Pearl, J., "A Probabilistic calculus of actions," in R. Lopez de Mantaras and D. Poole (Eds.), *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence (UAI-94)*, Morgan Kaufmann, San Mateo, CA, 454-462, 1994.

[Poole, 1985] Poole, D., "On the comparison of theories: Preferring the most specific explanations," in *Proceedings of International Conference on*

Artificial Intelligence (IJCAI-85), Los Angeles, CA, 144–147, 1985.

[Simon, 1953] Simon, H., "Causal ordering and identifiability," in W.C. Hood and T.C. Koopmans (Eds.), *Studies in Econometric Method*, New York, NY, Chapter 3, 1953.

[Sobel 1990] Sobel, M.E., "Effect analysis and causation in linear structural equation models," *Psychometrika*, 55(3), 495–515, 1990.

[Spirtes et al., 1993] Spirtes, P., Glymour, C., and Schienes, R., *Causation, Prediction, and Search*, Springer-Verlag, New York, 1993.

[Strotz & Wold, 1960] Strotz, R.H. and Wold, H.O.A., "Recursive versus nonrecursive systems: An attempt at synthesis," *Econometrica* 28, 417–427, 1960.