CAUSALITY CORRECTIONS IMPLEMENTED IN 2nd PRINTING

Updated 9/26/00

page v **insert** “TO RUTH” centered in middle of page

page xv **insert** in second paragraph “David Galles” after “Dechter”

page 2 **replace** “2000” with “2004” in 2nd paragraph, line 8 of 1.1.2

page 3 **insert** “, $\Rightarrow$” after “($\neg$)” in footnote 1
replace “and *not*” with “*not*, and *implies*,” in footnote 1

page 19 *append* (continue italics) to end of Theorem 1.2.7, “(*We exclude* X_i *when speaking of its “nondescendants”*.)”

page 30 **replace** in line 7 from top “mutually” with “jointly”
insert “parental” before “Markov” in first line of 2nd paragraph after Theorem 1.4.1
append to end of footnote 16 “but I am not aware of any nonparametric version.”

page 52 **insert** “stable” after “IC*, that takes a” in 2nd paragraph after Theorem 2.6.2,
replace “sampled” with “stable” in Input line of IC* Algorithm.
append “(with respect to some latent structure).” to same line

page 52 page 55, 1st line of footnote 9
replace “Pearl (1990)” with “Pearl (1990a)”

page 68 **replace** in line 1 after Eq. (3.2), “mutually” with “jointly”

page 72 **replace** “(1990, 1999)” with “(1990, 2001)” on line 6

page 89 **replace** in paragraph starting “Indeed, if condition...”. Should be “conditions require” not “condition require”

page 126 **insert** “and Robins (1997)” after “Pearl and Robins (1995)”, line 2.

page 130 **replace** the “P” with “E” in the formula (second line of section 4.5.4.) Should read “$E(Y|\hat{x}, \widehat{pa}_{Y\setminus X})$”
replace “we can compute the difference” with “we should replace the controlled difference” in last line of page

page 131 replace from top of page through the end of section 4.5.4 with the following:

$$P(\text{admission}|\widehat{\text{male}}, \widehat{\text{dept}}) - P(\text{admission}|\widehat{\text{female}}, \widehat{\text{dept}})$$

with some average of this difference over all departments. This average should measure the increase in admission rate in a hypothetical experiment in which we instruct all female candidates to retain their department preferences but change their gender identification (on the application form) from female to male.

In general, the average direct effect is defined as the expected change in Y induced by changing X from x to x' while keeping the other parents of Y constant at whatever value they obtain under $do(x)$. This hypothetical change is what law makers instruct us to consider in race or sex discrimination cases: "The central question in any employment-discrimination case is whether the employer would have taken the same action had the employee been of a different race (age, sex, religion, national origin etc.) and everything else had been the same." (In Carson versus Bethlehem Steel Corp., 70 FEP Cases 921, 7th Cir. (1996)).

The formal expression for this hypothetical change involves probabilities of (nested) counterfactuals (see Section 7.1 for semantics and computation) that cannot be written in terms of the $do(x)$ operator.[9] Therefore, the average direct effect cannot in general be identified, even from data obtained under randomized control of all variables. However, if certain assumptions of "no confounding" are deemed valid,[10] then the average direct effect can be reduced to

$$\Delta_{x,x'}(Y) = \sum_{pa_{Y \setminus X}} [E(Y|\widehat{x'}, \widehat{pa}_{Y \setminus X}) - E(Y|\hat{x}, \widehat{pa}_{Y \setminus X})] P(pa_{Y \setminus X}|\hat{x}), \qquad \textbf{(4.11)}$$

and the techniques developed in Section 4.4 for identifying control-specific plans, $P(y|\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n)$, become applicable.

[9] Using the counterfactual notation of Section 7.1, the general expression for the average direct effect is

$$\Delta_{x,x'}(Y) = E(Y_{x'Z_x}) - E(Y_x)$$

where $Z = pa_{Y \setminus X}$. The subscript $x'Z_x$ represents the operation of setting X to x' and, simultaneously, setting Z to whatever value it would have obtained under the setting $X = x$. This general expression reduces to (4.11) if $Z_x \perp\!\!\!\perp Y_{x'z}$ holds for all z. Likewise, the average *indirect* effect is defined as $E(Y_{xZ_{x'}}) - E(Y_x)$.

[10] See details in Technical Report R-273 posted on www.cs.ucla.edu/~judea.

page 164 **replace** ... "$do(x, y, w)$" with "$do(x, z, w)$" in line 8 after Definition 5.4.3.

page 165 **replace** last two sentences of section 5.4.2 with:

The expressions corresponding to these policies are $P(y|do(x),\ do(z))$ and $P(y|do(x))$, and this pair of distributions fully represents the policy implications of indirect effects. Similar conclusions have been expressed by Robins and Greenland (1992). (But see Chapter 4, footnote 9, page 131.)

page 177 *delete* "tormented" in paragraph 3, line 2

page 184 *append* to end of Definition 6.2.1 (continue italics - except 'unbiased'): "*If* (6.10) *holds, we say that* $P(y|x)$ *is* unbiased."

page 216 In line before Eq. (7.11), **replace** "To answer the first query," with "In order to answer the first query,"

page 236 **replace** "& $(X \longrightarrow Y|$" with "& $(X \not\longrightarrow Y|$" in first formula of Theorem 7.3.8

page 240 **replace** the last sentence in the last paragraph of section 7.4.1 with:

However, this effectiveness is partly acquired by limiting the counterfactual antecedent to conjunction of elementary propositions. Disjunctive hypotheticals, such as "if Bizet and Verdi were compatriots," usually lead to multiple solutions and hence to nonunique probability assignments.

page 246 **insert** in footnote 26 after "(see Section 5.4.3)." "Epidemiologists refer to (7.46) as "no-confounding" (see (6.10))."

page 255 **replace** in the 2nd line "pregnant" with "nonpregnant"

page 259 **insert** close parentheses after "(Sections 3.2 and 7.1", line 2 of Preface

page 284 **replace** "Michie in press" with "Michie 1999" in the last line of paragraph 4

page 329 **replace** "(1999)" with "(2000)" in last line of page

page 332 **replace** in paragraph starting "Even an erratic and ...". Change "role" to "roll"

page 354 line 2 from bottom, **replace** "mediated by tar deposits" with "unmediated by tar deposits"

page 361 *update* Dawid 1997 citation. **replace** "To appear ..." with "Also [with discussion] in *Journal of the American Statistical Association* 95:407–48, 2000."

page 363 *append* to Halpern (1998) citation, "Also, *Journal of Artificial Intelligence Research* 12:317–37, 2000."

update Halpern and Pearl (1999) citation. **replace** "(1999)" with "(2000)", **replace** "Actual causality." with "Causes and explanations.", and *append* "www.cs.ucla.edu/∼judea/"

page 364 *update* Hoover 1999 citation. **replace** "(1999)" with "(2001)"

page 365 *update* Kuroki citation. *append* "29: 105–17." after "*Journal of the Japan Statistical Society*".

page 366 *update* Michie citation. **replace** "(in press)" with "(1999)" and **insert** "pp. 60-86" before "vol. 15"

page 368 *update* Pearl 1999 citation. *remove* "To appear in" and replace "121" with "121:93–149."

page 369 **insert** in Robins 1997 citation, "M. Berkane (Ed.)," before "*Latent Variable Modeling...*"

page 370 add* to Shipley 1997 citation, "Also in *Structural Equation Modelling*, 7:206–18, 2000."

page 381 **insert** "27–8", after "(examples) price and demand" and before "215-17"
replace "245" with "245–7", at end of "(exogeneity) controversies regarding...245"
combine "explanation" and "explanations" to read "explanation, 25, 58, 221–3, 285, 308–9"

page 382 **insert** "131" after "indirect effects," and before "165"

ADDENDUM TO CORRECTIONS IMPLEMENTED IN 2nd PRINTING

Updated 12/14/00

page 28 **replace** "income (Z)" with "income (I)" in the caption of Figure 1.5

page 48 **replace** in line before Definition 2.4.1, "when one of the coins becomes slightly biased." with "when the coins become slightly biased."

page 51 *append* to line 7, Rule R_4 to read:
Orient a–b into $a \longrightarrow b$ whenever there are two chains a–$c \longrightarrow d$ and $c \longrightarrow d \longrightarrow b$ such that c and b are nonadjacent and a and d are adjacent.

page 231 **replace** Definition 7.3.4 and 2 lines following to read:

Definition 7.3.4 (Recursiveness)
Let X and Y be singleton variables in a model, and let $X \longrightarrow Y$ stand for the inequality $Y_{xw}(u) \neq Y_w(u)$ for some values of x, w, and u. A model M is recursive if, for any sequence $X_1, X_2, \ldots, X_k$, we have

$$X_1 \longrightarrow X_2,\ X_2 \longrightarrow X_3, \ldots, X_{k-1} \longrightarrow X_k \Rightarrow X_k \not\longrightarrow X_1 \qquad \textbf{(7.24)}$$

Clearly, any model M for which the causal diagram $G(M)$ is acyclic must be recursive.

page 382 *change* "Markov (assumptions underlying, 30)" to "Markov (assumption, 30, 69)"

page 382 *append* "69" after "causal, 30" in "Markov condition (causal, 30)"

page 384 *add* as subentry after "structural model, 27, 44, 202" "Markovian, 30, 69".

PRINTING CORRECTIONS/ADDITIONS IN THE SECOND EDITION OF *CAUSALITY*

Updated 7/23/08

Preface

page v **replace** "TO RUTH" (centered in middle of title page) with

TO DANNY
AND THE AUDACITY OF GOODNESS
(two lines indented and flushed left on title page)

page xvi **Add** new section to Preface.

Preface to the Second Edition
It has been more than eight years since the first edition of this book presented readers with the friendly face of causation and her mathematical artistry. The popular reception of the book and the rapid growth of the field call for a new edition to assist causation through her second transformation – from a demystified wonder to a commonplace tool in research and education. This edition (1) provides technical corrections, updates and clarifications in all ten chapters of the book, (2) adds summaries of new developments and annotated bibliographical references at the end of each chapter, and (3) elucidates subtle issues that readers and reviewers have found perplexing, objectionable, or in need of elaboration. These are assembled into an entirely new chapter (11) which, I sincerely hope, clears the province of causal thinking from the last traces of controversy.

Teachers who have taught from this book before should find the revised edition more lucid and palatable, while those who have waited for scouts to carve the path, will find the road paved and tested. Supplementary educational material, slides, tutorials and homeworks can be found on my website http://www.cs.ucla.edu/∼judea/.

My main audience remain the students: students of statistics who wonder why instructors are reluctant to discuss causality in class; students of epidemiology who wonder why elementary concepts such as confounding are so hard to define mathematically; students of economics and social science who question the meaning of the parameters they estimate; and, naturally, students of artificial intelligence and cognitive science, who write programs and theories for knowledge discovery, causal explanations and causal speech.

J.P.
Los Angeles
July 2008

Chapter 1

page 9 -5 lines from bottom of page

replace "$\sum_x (x - \hat{x})^2 P(x|y)$ over all $\hat{x}$." with "$\sum_x (x - x')^2 P(x|y)$ over all possible x'."

page 11 **add** sentence to end footnote 4

"Geiger and Pearl (1993) present an in-depth analysis."

page 22 3 lines before footnote 8

replace "join distribution" with "joint distribution"

page 24 line 7

replace

(iii) $P_x(v_i|pa_i) = P(v_i|pa_i)$ for all $V_i \notin X$ whenever pa_i is consistent with $X = x$.

with:

(iii) $P_x(v_i|pa_i) = P(v_i|pa_i)$ for all $V_i \notin X$ whenever pa_i is consistent with $X = x$, i.e., each $P(v_i|pa_i)$ remains invariant to interventions not involving V_i.

page 31 footnote 18

replace "such a property." with "this property."

page 32 2 lines before footnote 20

replace 'linear programming" with "linear-programming"

page 34 in 2nd line

replace "(Dawid 1997)." with "(Dawid 2000)."

page 39 Footnote 26, last sentence

replace Causal assumptions of the type developed in Chapter 2 (see Definitions 2.4.1 and 2.7.4) must be invoked before applying such sequences in policy-related tasks.

With:

Such sequences are statistical in nature and, unless causal assumptions of the type developed in Chapter 2 (see Definitions 2.4.1 and 2.7.4) are invoked, they cannot be applied to policy-evaluation tasks.

replace first Remark

Remark: The distinction between probabilistic and statistical parameters is devised to exclude the construction of joint distributions that invoke hypothetical variables (e.g., counterfactual or theological). Such constructions, if permitted, would qualify any quantity as statistical and would obscure the distinction between causal and non-causal assumptions.

With:

Remark: The exclusion of unmeasured variables from the definition of statistical parameters is devised to rule out the construction of joint distributions that invoke counterfactual or metaphysical variables. Such constructions, if permitted, would qualify any quantity as statistical and would thus obscure the important distinction between quantities that can be estimated from statistical data alone, and those that require additional assumptions beyond the data.

insert (after "falsifiable")

Often, though not always, causal assumptions can be falsified from experimental studies, in which case we say that they are "experimentally testable." For example, the assumption that X has no effect on $E(Y)$ in model 2 of Figure 1.6 is empirically testable, but the assumption that X may cure a given subject in the population is not.

replace second Remark

Remark: The distinction between causal and statistical parameters is crisp and fundamental. Causal parameters can be discerned from joint distributions only when special assumptions are made, and such assumptions must have causal components to them. The formulation and simplification of these assumptions will occupy a major part of this book.

With:

Remark: The distinction between causal and statistical parameters is crisp and fundamental – the two do not mix. Causal parameters cannot be discerned from statistical parameters unless causal assumptions are invoked. The formulation and simplification of these assumptions will occupy a major part of this book.

page 40 **Add** to end of Chapter 1

Two Mental Barriers to Causal Analysis

The sharp distinction between statistical and causal concepts can be translated into a useful principle: behind every causal claim there must lie some causal assumption that is not discernable from the joint distribution and, hence, not testable in observational studies. Such assumptions are usually provided by humans, resting on expert judgment. Thus, the way humans organize and communicate experiential knowledge becomes an integral part of the study, for it determines the veracity of the judgments experts are requested to articulate.

Another ramification of this causal-statistical distinction is that any mathematical approach to causal analysis must acquire new notation. The vocabulary of probability calculus, with its powerful operators of expectation, conditionalization and marginalization, is defined strictly in terms of distribution functions and is insufficient therefore for expressing causal assumptions or causal claims. To illustrate, the syntax of probability calculus does not permit us to express the simple fact that "symptoms do not cause diseases," let alone draw mathematical conclusions from such facts. All we can say is that two events are dependent—meaning that if we find one, we can expect to encounter the other, but we cannot distinguish statistical dependence, quantified by the conditional probability $P(disease|symptom)$ from causal dependence, for which we have no expression in standard probability calculus.

The preceding two requirements: (1) to commence causal analysis with untested, judgmental assumptions, and (2) to extend the syntax of probability calculus, constitute the two main obstacles to the acceptance of causal analysis among professionals with traditional training in statistics (Pearl 2003b, 2008a; Cox and Wermuth 1996). This book helps overcome the two barriers through an effective and friendly notational systems based on symbiosis of graphical and algebraic approaches.

Chapter 2

page 41 replace in line 5

"is an outgrowth of Pearl (1988b, chap. 8), which describes"

with

"is an outgrowth of Rebane and Pearl (1987) (also, Pearl (1988b, Chap. 8), which describes"

replace "On the other hand, the Stanford group" with "The Stanford group"

replace "to update the posterior probabilities assigned to" with "to update prior probabilities assigned to"

page 42 replace section heading, "**2.1 INTRODUCTION**" with "**2.1 INTRODUCTION – THE BASIC INTUITIONS**"

page 43 replace last two paragraphs of Section 2.1 with

Such thought experiments tell us that certain patterns of dependency, void of temporal information, are conceptually characteristic of certain causal directionalities and not others. Reichenbach (1956) who was the first to wonder about the origin of those patterns suggested that they are characteristic of Nature, reflective of the second law of thermodynamics. Rebane and Pearl (1987) posed the question in reverse, and asked whether the distinctions among the dependencies associated with the three basic causal substructures: $X \longrightarrow Y \longrightarrow Z$, $X \longleftarrow Y \longrightarrow Z$ and $X \longrightarrow Y \longleftarrow Z$ can be used to uncover causal directionalities in the underlying data generating process. They quickly realized that the key to determining the direction of the causal relationship between X and Y lies in "the presence of a third variable Z that correlates with Y but not with X," as in the collider $X \longrightarrow Y \longleftarrow Z$, and developed an algorithm that recovers both the edges and directionalities in the class of causal graphs that they considered (i.e., a polytrees).

The investigation in this chapter formalizes these intuitions and extends the Rebane-Pearl recovery algorithm to general graphs, including graphs with unobserved variables.

page 48 replace

"also known as DAG-isomorphism" with "also known as DAG-isomorphism or perfect-mapness"

replace

Succinctly, P is a stable distribution if there exists a DAG D such that $(X \perp\!\!\!\perp Y|Z)_P \Leftrightarrow (X \perp\!\!\!\perp Y|Z)_D$

with
Succinctly, P is a stable distribution of M if it "maps" the structure D of M, that is, $(X \perp\!\!\!\perp Y|Z)_P \Leftrightarrow (X \perp\!\!\!\perp Y|Z)_D$

page 50 **replace** item #3 of IC Algorithm with
In the partially directed graph that results, orient as many of the undirected edges as possible subject to two conditions: (i) Any alternative orientation would yield a new v-structure; or (ii) Any alternative orientation would yield a directed cycle.

page 52 **replace** in line 4–5 of 2nd paragraph of Theorem 2.6.2:
"distinguished protection" with "distinguished projection"

page 64 **add** new postscript

Postscript for the second edition

Work on causal discovery has been pursued vigorously by the TETRAD group at Carengie Mellon University and reported in (Spirtes et al. 2000; Robins et al. 2003; Scheines 2002; Moneta and Spirtes 2006).

Applications of causal discovery in economics are reported in Bessler (2002), Swanson and Granger (1997), and Demiralp and Hoover (2003). Gopnik et al. (2004) applied causal Bayesian networks to explain how children acquire causal knowledge from observations and actions (see also Glymour 2001).

Hoyer et al. (2006) and Shimizu et al. (2005, 2006) have proposed a new scheme of discovering causal directionality, based not on conditional independence but on functional composition. The idea is that in a linear model $X \longrightarrow Y$ with non-Gaussian noise, variable Y is a linear combination of two independent noise terms. As a consequence, $P(y)$ is a convolution of two non-Gaussian distributions and would be, figuratively speaking, "more Gaussian" than $P(x)$. The relation of "more Gaussian than" can be given precise numerical measure and used to infer directionality of certain arrows.

Tian and Pearl (2001) developed yet another method of causal discovery based on the detection of "shocks," or spontaneous local changes in the environment which act like "Nature's interventions," and unveil causal directionality toward the consequences of those shocks.

Chapter 3

page 67 , line 2, **replace** "that connect" with "emanating from"

page 72 , footnote 4, line 3
replace "said formula" with "this formula"

page 73 **replace** equation *between* (3.11) and (3.12)

$$\begin{aligned} P(pa_i|do(x_i')) &= P(pa_i); \\ \frac{P(s_i, pa_i|do(x_i'))}{P(s_i', pa_i|do(x_i'))} &= \frac{P(s_i, pa_i)}{P(s_i', pa_i)}. \end{aligned}$$

with

$$\begin{aligned} P(pa_i|do(x_i')) &= P(pa_i); \\ \frac{P(s_i, pa_i, x_i'|do(x_i'))}{P(s_i', pa_i, x_i'|do(x_i'))} &= \frac{P(s_i, pa_i, x_i')}{P(s_i', pa_i, x_i')}. \end{aligned}$$

* Two lines before Theorem 3.2.2: **replace** “disjoint of” with “disjoint from”

page 75 **replace** the last line of Eq. (3.15)

$$\times \prod_k P^*(z_k|z_{k-1}, x_{k-1}) \prod_k P^*(x_k|x_{k-1}, z_k, z_{k-1}).$$

with

$$\times \prod_{k=1}^{n} P^*(z_k|z_{k-1}, x_{k-1}) \prod_{k=1}^{n} P^*(x_k|x_{k-1}, z_k, z_{k-1}).$$

replace the last line of Eq. (3.16)

$$\times \prod_k P(z_k|z_{k-1}, x_{k-1}) \prod_k P^*(x_k|x_{k-1}, z_k, z_{k-1}).$$

with

$$\times \prod_{k=1}^{n} P * (z_k|z_{k-1}, x_{k-1}) \prod_{k=1}^{n} P^*(x_k|x_{k-1}, z_k, z_{k-1}).$$

replace Eq. (3.17)

$$P^*(y) = \sum_{z_1, \cdots, z_n} P(y|z_1, z_2, \ldots, z_n, g_1, g_2, \ldots, g_n) \prod_k P(z_k|z_{k-1}, g_{k-1})$$

with

$$P^*(y) = \sum_{z_1, \cdots, z_n} P(y|z_1, z_2, \ldots, z_n, g_1, g_2, \ldots, g_n) \prod_{k=1}^{n} P^*(z_k|z_{k-1}, g_{k-1})$$

page 78 line 5
change “nonpositive” to “nonnegative”

page 79 **insert** sentence before Theorem 3.3.2
Chapter 11 gives the intuition for (i) and (ii).

page 82 **replace** at line 2 of paragraph 2: "since there is no back-door path from X to Z, we simply have"
with
"since there is no unblocked back-door path from X to Z in Figure 3.5, we simply have"

replace in Definition 3.3.3 (Front-Door): "(ii) there is no back-door path from X to Z; and"
with
"(ii) there is no unblocked back-door path from X to Z; and"

page 86 paragraph after Corollary 3.4.2
replace
Whether Rules 1–3 are sufficient for deriving all identifiable causal effects remains an open question. However, the task of finding a sequence of transformations (if such exists) for reducing an arbitrary causal effect expression can be systematized and executed by efficient algorithms (Galles and Pearl 1995; Pearl and Robins 1995), to be discussed in Chapter 4. As we illustrate in Section 3.4.3, symbolic derivations using the hat notation are much more convenient than algebraic derivations that aim at eliminating latent variables from standard probability expressions (as in Section 3.3.2, equation(3.24)). With
Rules 1–3 have been shown to be *complete*, namely, sufficient for deriving all identifiable causal effects (Shpitser and Pearl 2006a; Huang and Valtorta 2006). Moreover, as illustrated in Section 3.4.3, symbolic derivations using the hat notation are more convenient than algebraic derivations that aim at eliminating latent variables from standard probability expressions (as in Section 3.3.2, equation(3.24)). However, the task of deciding whether a sequence of rules exists for reducing an arbitrary causal effect expression has not been systematized, and direct graphical criteria for identification are therefore more desirable. These will be developed in Chapter 4.

page 99 line 2 of 2nd paragraph
replace "in the counterfactual versus structural" with "in the potential-outcome versus structural"

line 8-9 after Eq. (3.52)
*replace "follow as theorems from the structural interpretation." with "follow as theorems from the structural interpretation, and no other constraint need ever be considered."

page 103 **replace** last paragraph on page (including footnote 15) with:
To place this result in the context of our analysis in this chapter, we need to focus attention on condition (3.62) which facilitated Robins' derivation of (3.63) and ask whether this formal counterfactual independency can be given a meaningful graphical interpretation. The answer will be given in Chapter 4 (Theorem 4.4.1), where we derive a graphical condition for identifying the effect of a plan, i.e., a sequential set of actions. The condition reads as follows: $P(y|g = x)$ is identifiable and is given

by (3.63) if every action-avoiding back-door path from X_k to Y is blocked by some subset L_k of non-descendants of X_k. (By "action-avoiding" we mean a path containing no arrow entering an X variable later than X_k.) Chapter 11 (Section 11.4.2) further shows that the graphical criterion above is more general than that given in (3.62).

page 104 replace 1st paragraph with

The structural analysis introduced in this chapter supports and generalizes Robins's result from a new theoretical perspective. First, on the technical front, this analysis offers systematic ways of managing models where Robins's starting assumption (3.62) is inapplicable. Examples are Figures 3.8(d)–(g)).

Second, on the conceptual front, the structural framework represents a fundamental shift from the vocabulary of counterfactual independencies, to the vocabulary of processes and mechanisms, in which human knowledge resides. The former requires human assessments of esoteric relationships such as (3.62), while the latter expresses those same relationships in vivid graphical terms of missing links. Still, Robins's pioneering research has proven (i) that algebraic methods can handle causal analysis in complex multistage problems and (ii) that causal effects in such problems can be reduced to estimable quantities (see also Sections 3.6.1 and 4.4).

page 105 add new section after end of Postscript

Postscript to 2nd Edition

Complete identification results

A key identification condition, which generalizes all the criteria established in this chapter has been derived by Jin Tian. It reads:

Theorem 3.6.1 (*Tian and Pearl, 2002*)
A sufficient condition for identifying the causal effect $P(y|do(x))$ is that there exists no bi-directed path (i.e., a path composed entirely of bi-directed arcs) between X and any of its children.[15]

Remarkably, the theorem asserts that, as long as every child of X (on the pathways to Y) is not reachable from X via a bi-directed path, then, regardless of how complicated the graph, the causal effect $P(y|do(x))$ is identifiable. All identification criteria discussed in this chapter are special cases of the one defined in this theorem. For example, in Figure 3.5 $P(y|do(x))$ can be identified because the one path from X to Z (the only child of X) is not bi-directed. In Figure 3.7, on the other hand, there is a path from X to Z_1 traversing only bi-directed arcs, thus violating the condition of Theorem 3.6.1, and $P(y|do(x))$ is not identifiable.

[15] Before applying this criterion, one may delete from the causal graph all nodes that are not ancestors of Y.

Note that all graphs in Figure 3.8 and none of those in Figure 3.9 satisfy the condition above. Tian and Pearl (2002) further showed that the condition is both sufficient and necessary for the identification of $P(v|do(x))$, where V includes all variables except X. A necessary and sufficient condition for identifying $P(w|do(z))$, with W and Z two arbitrary sets, was established by Shpitser and Pearl (2006b). Subsequently, a complete graphical criterion was established for determining the identifiability of *conditional* interventional distributions, namely, expressions of the type $P(y|do(x), z)$ where X, Y and Z are arbitrary sets of variables (Shpitser and Pearl 2006a).

These results constitute a complete characterization of causal effects in graphical models. They provide us with polynomial time algorithms for determining whether an arbitrary quantity invoking the $do(*)$ operator is identified in a given semi-Markovian model and, if so, what the estimand is of that quantity. Remarkably, one corollary of these results also states that the *do*-calculus is complete, namely, a quantity $Q = P(y|do(x), z)$ is identified if and only if it can be reduced to a *do*-free expression using the three rules of Theorem 3.4.1.[16]

Applications and Critics

Gentle introductions to the concepts developed in this chapter are given in (Pearl 2003) and (Pearl 2008a). Applications of causal graphs in epidemiology are reported in Robins (2001), Hernan et al. (2002), Hernan et al. (2004), Greenland and Brumback (2002), Greenland et al. (1999) Kaufman et al. (2005), Petersen et al. (2006), and VanderWeele and Robins (2007).

Interesting applications of the front door criterion (Section 3.3.2) were noted in social science (Morgan and Winship 2007) and economics (Chalak and White 2006).

Advocates of the "potential outcome" approach have been most resistant to accepting graphs or structural equations as the basis for causal analysis and, lacking these conceptual tools, were led to dismiss important scientific concepts as "ill-defined" and "deceptive" (Holland 2001, Rubin 2004, Rubin 2005). Lauritzen (2004) and Heckman(2005) have criticized this attitude.

Equally puzzling are concerns of some philosophers (Cartwright 2007, Woodward 2003) and economists (Heckman 2005) that the *do*-operator is too local to model complex, real life policy interventions, which sometimes affect several mechanisms at once and often involve conditional decisions, imperfect control, and multiple actions. These concerns emerge from conflating the mathematical definition of a relationship (e.g., causal effect) with the technical feasibility of testing that relationship in the physical world. While the *do*-operator is indeed an ideal mathematical tool (not unlike the *derivative* in differential calculus), it nevertheless permits us to specify and analyze interventional strategies of great complexities. Readers will find

[16] This was independently established by Huang and Valtorta (2006).

examples of such strategies in Chapter 4, and a further discussion of this issue in Chapter 11 (Sections 11.4.3–11.4.7 and Section 11.5.4).

Chapter 4

page 108 paragraph 3, line 1

replace “Traditional decision theory” with “Commonsensical decision theory”

page 109 2nd paragraph, 2nd line of mnemonic rhymes

replace

On facts that preceded the act,

With

On what could have caused the act,

page 111 **replace** footnote 4:

[4]The ID literature’s insistence on divorcing the links in the ID from any causal interpretation (Howard and Matheson 1981; Howard 1990) is at odds with prevailing practice. The causal interpretation is what allows us to treat decision variables as root nodes, unassociated with all other nodes (except their descendants). with

[4]The ID literature’s insistence on divorcing the links in the ID from any causal interpretation (Howard and Matheson 1981; Howard 1990) is at odds with prevailing practice. The causal interpretation is what allows us to treat decision variables as root nodes and construct the proper decision trees for analysis; see Section 11.6 for a demonstration.

page 113 line 3

replace “Section 1.1.4” with “Section 1.4.4”

page 114 Section 4.3.1, 1st paragraph

replace

Theorem 4.3.1 characterizes the class of “*do*-identifiable” models in the form of four graphical conditions, anyone of which is sufficient for the identification of $P(y|\hat{x})$ when X and Y are singleton nodes in the graph. Theorem 4.3.2 then asserts the completeness (or necessity) of these four conditions; one of which must hold in the model for $P(y|\hat{x})$ to be identifiable in *do*-calculus. Whether these four conditions are necessary in general (in accordance with the semantics of Definition 3.2.4) depends on whether the inference rules of *do*-calculus are complete. This question, to the best of my knowledge, is still open.

With

Theorem 4.3.1 characterizes a class of “*do*-identifiable” models in the form of four graphical conditions, anyone of which is sufficient for the identification of $P(y|\hat{x})$ when X and Y are singleton nodes in the graph. Theorem 4.3.2 then states the at least one of these four conditions must hold in the model for $P(y|\hat{x})$ to be identifiable in *do*-calculus. In view of the completeness of *do*-calculus, we conclude that one of the four conditions is necessary for any method of identification compatible with the semantics of Definition 3.2.4).

page 114 Section 4.3, last sentence before Section 4.3.1

replace

In this section we establish a complete characterization of the class of models in which the causal effect $P(y|\hat{x})$ is identifiable in *do*-calculus.

With

In this section we characterize a simple class of models in which the causal effect $P(y|\hat{x})$ is identifiable in *do*-calculus. This class is subsumed by the one established by Tian and Pearl (2002) in Theorem 3.7 and the complete characterization given later in Shpitser and Pearl (2006b). It is brought here for historical purposes only.

page 118 line 1

replace $Pb = \mathrm{ClosedForm}(b|\hat{x})$ with $Pb = \mathrm{ClosedForm}(P(b|\hat{x}))$

page 121 paragraph 4, line 1

replace "Theorems 4.41 and" with "Theorems 4.4.1 and"

page 129 **add** to end of footnote 8

"Cole and Hernán (2002) present examples in epidemiology."

page 130-131 Section 4.5.4, replace entire section

replace entire section with:

4.5.4 Average Direct and Indirect Effects

Readers versed in structural equation models (SEMs) will note that, in linear systems, the direct effect $E(Y|\hat{x}, \widehat{pa}_{Y\setminus X})$ is fully specified by the path coefficient attached to the link from X to Y; therefore, the direct effect is independent of the values $pa_{Y\setminus X}$ at which we hold the other parents of Y. In nonlinear systems, those values would, in general, modify the effect of X on Y and thus should be chosen carefully to represent the target policy under analysis. For example, the direct effect of a pill on thrombosis would most likely be different for pregnant and nonpregnant women. Epidemiologists call such differences "effect modification" and insist on separately reporting the effect in each subpopulation.

Although the direct effect is sensitive to the levels at which we hold the parents of the outcome variable, it is sometimes meaningful to average the direct effect over those levels. For example, if we wish to assess the degree of discrimination in a given school without reference to specific departments, we should replace the controlled difference

$$P(\mathrm{admission}|\widehat{\mathrm{male}}, \widehat{\mathrm{dept}}) - P(\mathrm{admission}|\widehat{\mathrm{female}}, \widehat{\mathrm{dept}})$$

with some average of this difference over all departments. This average should measure the increase in admission rate in a hypothetical experiment in which we instruct all female candidates to retain their department preferences but change their gender identification (on the application form) from female to male.

Conceptually, we can define the average direct effect $DE_{x,x'}(Y)$ as the expected change in Y induced by changing X from x to x' while keeping the other parents of Y constant at whatever value they would have obtained under $do(x)$. This hypothetical change, which Robins and Greenland (1991) called "pure" and Pearl (2001) called "natural," is precisely what law makers instruct us to consider in race or sex discrimination cases: "The central question in any employment-discrimination case is whether the employer would have taken the same action had the employee been of a different race (age, sex, religion, national origin etc.) and everything else had been the same." (In Carson versus Bethlehem Steel Corp., 70 FEP Cases 921, 7th Cir. (1996)).

Using the parenthetical notation of Eq. 3.51, the formal expression for the "natural direct effect," reads:

$$DE_{x,x'}(Y) = E[(Y(x', Z(x))) - E(Y(x)] \quad \textbf{(4.11)}$$

Here, Z represents all parents of Y excluding X, and the expression $Y(x', Z(x))$ represents the value that Y would attain under the operation of setting X to x' and, simultaneously, setting Z to whatever value it would have obtained under the setting $X = x$. We see that $DE_{x,x'}(Y)$, the average direct effect of the transition from x to x', involves probabilities of *nested counterfactuals* and cannot be written in terms of the $do(x)$ operator. Therefore, the average direct effect cannot in general be identified, even with the help of ideal, controlled experiments (see Robins and Greenland (1992) and Section 7.1 for intuitive explanation). Pearl (2001) have shown nevertheless that, if certain assumptions of "no confounding" are deemed valid,[17] the average direct effect can be reduced to

$$DE_{x,x'}(Y) = \sum_z [E(Y|do(x', z)) - E(Y|do(x, z))]P(z|do(x)) \quad \textbf{(4.12)}$$

The intuition is simple; the natural direct effect is the weighted average of controlled direct effects, using the causal effect $P(z|do(x))$ as a weighing function. Under such assumptions, the techniques developed in Section 4.4 for identifying control-specific plans, $P(y|\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n)$, become applicable.

In particular, expression (4.12) is both valid and identifiable in Markovian models, where it can be written (using Corollary 3.2.6)

$$DE_{x,x'}(Y) = \sum_z [E(Y|x', z) - E(Y|x, z)] \sum_t P(z|x, pa_X = t) \quad \textbf{(4.13)}$$

[17] One sufficient condition is that $Z(x) \perp\!\!\!\perp Y(x', z)|W$ holds for some set W of measured covariates. See details and graphical criteria in (Pearl 2001) and in Petersen et al. (2006).

4.5.5 Indirect Effects

Remarkably, the definition of the natural direct effect (4.11) can easily be turned around and provide an operational definition for the *indirect effect* – a concept shrouded in mystery and controversy, because it is impossible, using the $do(x)$ operator, to disable the direct link from X to Y so as to let X influence Y solely via indirect paths.

The average indirect effect, IE, of the transition from x to x' is defined as the expected change in Y affected by holding X constant, at $X = x$, and changing Z to whatever value it would have attained had X been set to $X = x'$. Using our counterfactual notation, this reads (Pearl 2001):

$$IE_{x,x'}(Y) = E[(Y(x, Z(x'))) - E(Y(x))]$$

which is almost identical to the direct effect (Eq. (4.11)) save for a minor reversal of x and x'.

Indeed, it can be shown that, in general, the total effect TE of a transition is equal to the *difference* between the direct effect of that transition and the indirect effect of the reverse transition. Formally,

$$\begin{aligned} TE_{x,x'}(Y) &\triangleq E(Y(x) - Y(x')) \\ &= DE_{x,x'}(Y) - IE_{x',x}(Y) \end{aligned}$$

In linear systems, where reversal of transitions amounts to negating the signs of their effects, we have the standard additive formula

$$TE_{x,x'}(Y) == DE_{x,x'}(Y) + IE_{x,x'}(Y)$$

This provides a formal justification for the additive formula, since each term is based on an independent operational definition.

Note that the indirect effect has clear policy making implications. For example: in a hiring discrimination environment, a policy maker may be interested in predicting the gender mix in the work force if gender bias is eliminated and all applicants are treated equally, say the same way that males are treated today. This quantity will be given by the indirect effect of gender on hiring, mediates by factors such as education and aptitude, which may be gender-dependent.

More generally, a policy maker may be interested in the effect of issuing a directive to select set of subordinate employees, or in carefully controlling the routing of messages in a network of interacting agents. Such applications call for the definition of *path-specific effects*, that is, the causal effect of X on Y through a selected set of paths. A complete analysis for Markovian models is given in Avin et al. (2005).

Note that in all these cases, the policy intervention invokes the selection of signals to be sensed, rather than variables to be fixed. Pearl (2001) suggested that this signal selection operation are more fundamental to the notion of causation than manipulation; the latter being but a crude imitation of the former, devised for measurements.

It is remarkable that, under certain graphical conditions, counterfactual quantities like DE and ID that could not be expressed in terms of $do(x)$ operators, can nevertheless be reduced to counterfactual-free formulas involving $do(x)$ operators. In other words, quantities that at first glance appear void of empirical content, can be estimated from empirical studies involving experimental control. We shall see additional examples of this "magic of scientific analysis" in Chapters 7, 9, and 11. It constitutes an invincible argument in defense of counterfactual analysis, as expressed in (Pearl 2000) against the critical stance of Dawid (2000) and Geneletti (2007).

A general analysis of the conditions under which counterfactual quantities can be reduced to $do(x)$ expressions is given in Shpitser and Pearl (2007).

Acknowledgment

Sections 4.3 and 4.4 are based, respectively, on collaborative works with David Galles and James Robins. My interest in path-specific effects was stimulated by Jacques Hagenaars, who pointed out the importance of quantifying indirect effects in social science (see Chapter 11).

Chapter 5

page 141 2nd paragraph, line 4
insert "Edwards 2000; Cowell et al. 1999;" before "Whittaker 1990;"

page 147 line 3 of 1st paragraph (after Rule 2)
replace "would render X and Y" with "would render Z and Y"

page 171 **add** new section

Notes Added to Second Edition

5.5.1 An Econometric Awakening?

After decades of neglect of causal analysis in economics, a surge of interest seems to be in progress. In a recent series of papers, Jim Heckman (2000, 2003, 2005, 2007 (with Vytlacil)) has made great efforts to resurrect and reassert the Cowles Commission interpretation of structural equation models, and to convince economists that recent advances in causal analysis are rooted in the ideas of Haavelmo (1943) Marschak (1950), Roy (1951) and Hurwicz (1962). Unfortunately, Heckman still does not offer econometricians clear answers to the questions posed in this chapter (pp. 133, 170, 171, 215-217). In particular, unduly concerned with implementational issues, Heckman rejects Haavelmo's "equation wipe-out" as a basis for defining counterfactuals and fails to provide econometricians with an alternative definition, namely, a procedure, like that of Eq. (3.51), for computing the counterfactual $Y(x, u)$ in a well-posed economic model, with X and Y two arbitrary variables in the models. (See Sections 11.5.4–5.) Such a definition is necessary for endowing the "potential outcome" approach with a formal semantics, based on SEM, and thus unifying the two econometric camps currently working in isolation.

Another sign of positive awakening comes from the social sciences, through the publication of Morgan and Winship's book "Counterfactual and Causal Inference" (2007) in which the causal reading of SEM is clearly reinstated.[18]

5.5.2 Identification in Linear Models

In a series of papers, Brito and Pearl (2002ab, 2006) have established graphical criteria that significantly expand the class of identifiable semi-Markovian linear models beyond those discussed in this chapter. They first proved that identification is ensured in all graphs that do not contain bow-arcs, that is, no error correlation is allowed between a cause and its *direct* effect, while no restrictions are imposed on errors associated with indirect causes (Brito and Pearl 2002b). Subsequently, generalizing the concept of instrumental variables beyond the classical patterns of Figures 5.9 and 5.11, they establish a general identification condition that is testable in polynomial time and subsumes all conditions known in the literature.

5.5.3 Robustness of Causal Claims

Causal claims in SEM are established through a combination of data and the set of causal assumptions embodied in the model. For example, the claim that the causal effect $E(Y|do(x))$ in Fig. 5.9 is given by $\alpha x = r_{YZ}/r_{XZ}x$ is based on the assumptions: $cov(e_Z, e_Y) = 0$ and $E(Y|do(x,z)) = E(Y|do(z))$, both are shown in the graph. A claim is *robust* when it is insensitive to violations of some of the assumptions in the model. For example, the claim above is insensitive to the assumption $cov(e_Z, e_X) = 0$ which is shown in the model.

When several distinct sets of assumptions give rise to k distinct estimands for a parameter α, that parameter is called k-identified; the higher the k, the more robust are claims based on α, because equality among these estimands imposes $k - 1$ constraints on the covariance matrix which, if satisfied in the data, indicate an agreement among k distinct sets of assumptions, thus supporting their validity. A typical example emerges when several (independent) instrumental variables are available $Z_1, Z_2, \ldots, Z_k$ for a single link $X \longrightarrow Y$, which yield the equalities $\alpha = r_{YZ_1}/r_{XZ_1} = r_{YZ_2}/r_{XZ_Z} = \ldots = r_{YZ_k}/r_{XZ_k}$.

Pearl (2004) gives a formal definition for this notion of robustness, and established graphical conditions for quantifying the degree of robustness of a given causal claim.

[18] Though the SEM basis of counterfactuals is unfortunately not articulated.

k-identification generalizes the notion of *degree of freedom* in standard SEM analysis; the latter characterizes the entire model, while the former applies to individual parameters and, more generally, to individual causal claims.

Chapter 6

page 174 **replace** in 4th line from end of page: "is not a statement about C being a positive causal factor for E, properly written" with "is not a statement about C having a positive influence on E, properly written"

page 179 line 10
replace "$F = \mathrm{female}$ example" with "$F = \mathrm{gender}$

page 180 line 3 of footnote 6
replace "[y]ou might" with "you might"

page 195 **replace** "Figure 6.1" with "Figure 6.3" in 5th line before end of 3rd paragraph

page 200 postscript added

Postscript to Chapter 6

Readers would be amused to learn that the first printing of this chapter did not stop statisticians' fascination with Simpson's non-paradox. Textbooks continue to marvel at the phenomenon (Moore and McCabe 2005), and researchers continue to chase its mathematical intricacies (Cox and Wermuth 2003) and visualizations (Rücker and Schumacher 2008) with the passion of the 1970-90's, without once mentioning the word "cause" and without once stopping to ask: "What's the point?" In contrast, my confidence in the eventual triumph of the causal understanding of this non-paradox has been rekindled by a discussion published in the epidemiological literature, concluding in no ambiguous terms: "The explanations and solutions lie in causal reasoning which relies on background knowledge, not statistical criteria" (Arah 2008). It appears that, as we enter the age of causation, we should look to epidemiologists for guidance and wisdom.

Chapter 7

page 206 4th line after Eq. (7.5)
replace "(Dawid 1997)." with "(Dawid 2000)."

page 214–215 **replace** paragraph [through end of 7.1.4] with

The twin network reveals an interesting interpretation of counterfactuals of the form Z_{pa_Z}, where Z is any variable and PA_Z stands for the set of Z's parents. Consider the question of whether Z_x is independent of some given set of variables in the model of Figure 7.3. The answer to this question depends on whether Z^* is d-separated from that set of variables. However, any variable that is d-separated from Z^* would also be d-separated from U_Z, so the node representing U_Z can serve as a one-way proxy for the counterfactual variable Z_x. This is not a coincidence, considering that Z is governed by the equation $z = f_Z(x, u_Z)$. By definition, the probability of Z_x is equal to the probability of Z under the condition where X is held fixed at x. Under such condition, Z may vary only if U_Z varies. Therefore, if U_Z obeys a certain independence relationship then Z_x (more generally, Z_{pa_Z}) must obey that relationship as well. We thus obtain a simple graphical representation for any counterfactual variable of the form Z_{pa_Z}. Using this representation, we can easily verify from Figure 7.3 that $(U_Y \perp\!\!\!\perp X | \{Y^*, Z^*\})_G$ and $(U_Y \perp\!\!\!\perp U_Z | \{Y, Z\})_G$ both hold in the twin-network and, therefore,

$$Y_z \perp\!\!\!\perp X | \{Y_x Z_x\} \text{ and } Y_z \perp\!\!\!\perp Z_x | \{Y, Z\}$$

must hold in the model. Additional considerations involving twin networks, including generalizations to multi-networks (representing counterfactuals under different anticedants) are reported in Shpitser and Pearl (2007).

page 217 4th line from top of page
replace "observations $\{P = p_0,\ I = i,\ W = w\}$. According to Definition" with "observations $\{P = p_0,\ I = i,\ W = w\}$ (see Section 11.7.1.) According to Definition"

page 220 2nd paragraph, line 7
replace "including Dawid 1997)" with "including Dawid 2000)"

page 236 **replace** in 2nd line of footnote 17
"Bonet (1999)" with "Bonet (2001)"

page 257 (no postscript added)

Chapter 8

page 264 **replace** footnote 3 with
[3] Frangakis and Rubin (2002) dubbed them "Principal Stratification" and found them rather insightful. In the potential-outcome model (see Section 7.4.4), u stands for an experimental unit and $R(u)$ corresponds to the potential response of unit u to treatment x. The assumption that each unit (e.g., an individual subject) possesses an intrinsic, seemingly "fatalistic" response function has met with some objections (Dawid 2000), owing to the inherent unobservability of the many factors that might govern an individual response to treatment. The equivalence-class formulation of $R(u)$ mitigates those objections by showing that $R(u)$ evolves naturally and mathematically from any complex system of stochastic latent variables, provided only that we acknowledge their existence through the equation $y = f(x, u)$. Those who invoke quantum-mechanical objections to the latter step as well (e.g. Salmon 1998),

should regard the functional relationship $y = f(x, u)$ as an abstract mathematical construct, representing the extreme points (vertices) of the set of conditional probabilities $P(y|x, u)$ satisfying the constraints of (8.1) and (8.2).

page 269 **replace** 4 lines before Eq. (8.19) to read:

The analysis of $\mathrm{ACE}^*(X \rightarrow Y)$ reveals that, under conditions of *no intrusion* (i.e., $P(x_1|z_0) = 0$, as in most clinical trials), $\mathrm{ACE}^*(X \rightarrow Y)$ can be identified precisely (Bloom 1984; Heckman and Robb 1986; Angrist and Imbens 1991). The natural bounds governing $\mathrm{ACE}^*(X \rightarrow Y)$ in the general case can be obtained by similar means (Pearl 1995b), which yield

page 275 After Eq. (8.23), **replace**

However, the transition to a continuous X signals a drastic change in behavior, and it seems that the structure of Figure 8.1 induces no constraint whatsoever on the observed density (Pearl 1995c).

with

However, the transition to a continuous X signals a drastic change in behavior, and led (Pearl 1995c) to conjecture that the structure of Figure 8.1 induces no constraint whatsoever on the observed density. The conjecture was proven by (Bonet 2001).

page 280 Section 8.5.4, 2nd paragraph, 6th line

replace

the following four compliance-response populations: $\{(r_x = 0, r_y = 1),\ (r_x = 0, r_y = 2),\ (r_x = 1, r_y = 1),\ (r_x = 1, r_y = 2)\}$. Joe would have improved had he taken cholestyramine if his response behavior is either helped ($r_y = 1$) or always-recover ($r_y = 3$).

With

the four compliance-response populations:

$$\{(r_x = 0, r_y = 0),\ (r_x = 0, r_y = 1),\ (r_x = 1, r_y = 0),\ (r_x = 1, r_y = 1)\}.$$

Joe would have improved had he taken cholestyramine if and only if his response behavior is helped ($r_y = 1$).

page 281 (no postscript added)

Chapter 9

page 297 2nd paragraph, 3rd line

replace "(Dawid 1997)." with "(Dawid 2000)."

page 298 after (9.38), within the next 5 lines, **replace** $y_{x'}$ with $y'_{x'}$ (4 places) **and** correct in last line of (9.39).

replace

In order to compute PNS, we must evaluate the probability of the joint event $y_{x'} \wedge y_x$. Given that these two events are jointly true only when $U = \text{true}$, we have

$$\begin{aligned} \text{PNS} &= P(y_x, y_{x'}) \\ &= P(y_x, y_{x'}|u)P(u) + P(y_x, y_{x'}|u')P(u') \\ &= \frac{1}{2}(1+0) = \frac{1}{2}. \qquad (9.39) \end{aligned}$$

With

In order to compute PNS, we must evaluate the probability of the joint event $y'_{x'} \wedge y_x$. Given that these two events are jointly true only when $U = \text{true}$, we have

$$\begin{aligned} \text{PNS} &= P(y_x, y'_{x'}) \\ &= P(y_x, y'_{x'}|u)P(u) + P(y_x, y'_{x'}|u')P(u') \\ &= \frac{1}{2}(0+1) = \frac{1}{2}. \qquad (9.39) \end{aligned}$$

page 302 Section 9.3.4, 2nd paragraph, line 11
replace "(Dawid 1997);" with "(Dawid 2000);"

page 308 (no postscript added)

Chapter 10

page 329 **add** new postscript section

Postscript to Chapter 10

Halpern and Pearl (2001ab) discovered a need to refine the causal beam definition of 10.3.3. They retained the idea of defining actual causation by counterfactual dependency in a world perturbed by contingencies, but permitted a wider set of contingencies.

To see that the causal beam definition requires refinement, consider the following example.

Example 10.4.1 *A vote takes place, involving two people. The measure Y is passed if at least one of them votes in favor. In fact, both of them vote in favor, and the measure passes.*

This version of the story is identical to the disjunctive scenario discussed in Section 10.1.3, where we wish to proclaim each favorable vote, $V_1 = 1$ and $V_2 = 1$, a contributing cause of $Y = 1$.

However, suppose there is a voting machine that tabulates the votes. Let M represent the total number of votes recorded by the machine. Clearly $M = V_1 + V_2$ and $Y = 1$ iff $M \geq 1$. Figure 10.8 represents this more refined version of the story.

In this scenario, the beam criterion no longer qualifies $V_1 = 1$ as a contributing cause of $P = 1$, because V_2 cannot be labeled "inactive" relative to M, hence we are not

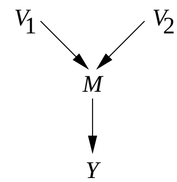

Figure 10.8: An example showing the need for beam refinement.

at liberty to set the contingency $V_2 = 0$ and test the counterfactual dependency of Y on V_1 as we did in the simple disjunctive case.

A refinement that properly handles such counterexamples was proposed in Halpern and Pearl (2001ab) but, unfortunately, Hopkins and Pearl (2002) showed that the constraints on the contingencies were too liberal. This led to a further refinement (Halpern and Pearl 2005ab). and to the definition given below:

Definition 10.4.2 *Actual Causation (Halpern and Pearl 2005)*
$X = x$ is an actual cause of $Y = y$ in a world $U = u$ if the following three conditions hold:

AC1. $X(u) = x, Y(u) = y$

AC2. *There is a partition of V into two subsets, Z and W with $X \subseteq Z$ and a setting x' and w of the variables in X and W, respectively, such that if $Z(u) = z*$, then both of the following conditions hold:*

(a) $Y_{x',w} \neq y$.

(b) $Y_{x,w,z*} = y$ *for all subsets W' of W and all subsets Z' of Z, with the setting w of W' and $z*$ of Z' equal to the setting of those variables in $W = w$ and $Z = z*$, respectively.*

AC3. *W is minimal; no subset of X satisfies conditions AC1 and AC2.*

The assignment $W = w$ acts as a contingency against which $X = x$ is given the counterfactual test, as expressed in AC2(a).

AC2 (b) limits the choice of contingencies. Roughly speaking, it says that if the variables in X are reset to their original values, then $Y = y$ must hold, even under the contingency $W = w$ and even if some variables in Z are given their original values (i.e.the values in $z*$)

In the case of the voting machine, if we identify $W = w$ with $V_2 = 0$, and $Z = z*$ with $V_1 = 1$ we see that $V_i = 1$ qualifies as a cause under AC2; we no longer require that M remains invariant to the contingency $V_2 = 0$, the invariance of $Y = 1$ suffices.

This definition, though it correctly solves most problems posed in the literature (Hitchcock 2007; Hiddleston 2005; Hitchcock 2008; Hall 2007) still suffers from one deficiency, it must rule out certain contingencies as unreasonable. Halpern (2008) has offered a solution to this problem by appealing to the notion of “normality” in default logic (Kraus et al. 1990; Spohn 1988; Pearl 1990b); only those contingencies should be considered which are at the same level of "normality" as their counterparts in the actual world.

Epilogue

No changes.

Chapter 11

New chapter – separate file.

NEW PRINTING CORRECTIONS/ADDITIONS TO BE IMPLEMENTED BY CAMBRIDGE IN THE SECOND EDITION OF *CAUSALITY*

Updated 8/20/08

Chapter 1

page 27 1st line after Eq. (1.40)

replace "the set of variables judged to be immediate causes of X_i" with
"the set of variables that directly determine the value of X_i "

page 27 3rd paragraph

replace paragraph starting "A set of equations..." (including footnote 13) with

The interpretation of the functional relationship in (1.40) is the standard interpretation that functions carry in physics and the natural sciences; it is a recipe, a strategy, or a *law* specifying what value nature would assign to X_i in response to every possible value combination that (PA_i, U_i) might take on. A set of equations in the form of (1.40) and in which each equation represent an autonomous mechanism is called *structural model*; if each variable has a distinct equation in which it appears on the left-hand side (called the *dependent* variable), then the model is called a *structural causal model* or a *causal model* for short.[13] Mathematically, the distinction between structural and algebraic equations is that any subset of structural equations is, in itself, a valid structural model – one that represents conditions under some set of interventions.

Chapter 4

page 113 -5 lines from bottom of page

insert after "hat-free expression."
Kuroki and Miyakawa (1999a, 2003) present graphical criteria.

page 126 line 2-3 after Figure 4.8

replace"Kuroki and Miyakawa (1999)." with
Kuroki (with Miyakawa 1999ab, 2003; with et al. 2003; with Cai 2004).

page 129 line 3 after Figure 4.9

replace "Z = applicant's career objectives;" with
"Z = applicant's (pre-enrollment) career objectives;"

[13] Formal treatment of causal models, structural equations, and error terms are given in Chapter 5 (Section 5.4.1) and Chapter 7 (Sections 7.1 and 7.2.5).

Chapter 6

page 194 -8 lines from bottom of page

replace "(at each level of X) and the association between Z and X may vanish as well." with "(at each level of X); similarly, the association between Z and X may vanish."

page 194 -3 lines from bottom of page

replace "is associated neither with the exposure (X) nor with the disease (Y)." with "is not associated with the disease (Y)."

page 194-196 **replace** starting with last paragraph (last 2 lines) of page 194 through last paragraph of 6.5.2 (on page 196):

The intuition behind Rothman and Greenland's statement just quoted can be explicated formally through the notion of stability: a variable that is *stably* unassociated with either X or Y can safely be excluded from adjustment. Alternatively, Rothman and Greenland's statement can be supported (without invoking stability) by using the notion of *nontrivial sufficient set* (Section 3.3)—a set of variables for which adjustment will remove confounding bias. It can be shown (see the end of this section) that each such set S, taken as a unit, must indeed be associated with X and be conditionally associated with Y, given X. Thus, Rothman and Greenland's condition is valid for nontrivial sufficient (i.e., admissible) sets but not for the individual variables in the set.

The practical ramifications of this condition are as follows. If we are given a set S of variables that is claimed to be sufficient (for removing bias by adjustment), then that claim can be given a necessary statistical test: S as a compound variable must be associated both with X and with Y (given X). In Figure 6.5, for example, $S_1 = \{A, Z\}$ and $S_2 = \{E, Z\}$ are sufficient and nontrivial; both must satisfy the condition stated.

Note that, although this test can be used for screening sets claimed to be sufficient, it does not constitute a test for detecting confounding. Even if we find a set S in a problem that is associated with both X and Y, we are still unable to conclude that X and Y are confounded. Our finding merely qualifies S as a candidate for sufficient status in case confounding exists, but we cannot rule out the possibility that the problem is unconfounded to start with. (The sets $S = \{E, A\}$ or $S = \{Z\}$ in Figure 6.1 illustrate this point.) Observing a discrepancy between adjusted and unadjusted associations (between X and Y) does not help us either, because (recalling our discussion of collapsibility) we do not know which—the preadjustment or postadjustment association—is unbiased (see Figure 6.4).

Proof of Necessity

To prove that (U_1) and (U_2) must be violated whenever Z stands for a nontrivial sufficient set S, consider the case where X has no effect on Y. In this case, confounding amounts to a nonvanishing association between X and Y. A well-known property of conditional independence, called *contraction* (Section 1.1.5), states that violation of (U_1), $X \perp\!\!\!\perp S$, together with sufficiency, $X \perp\!\!\!\perp Y|S$, implies violation of nontriviality, $X \perp\!\!\!\perp Y$:

$$X \perp\!\!\!\perp S \ \& \ X \perp\!\!\!\perp Y|S \Rightarrow X \perp\!\!\!\perp Y.$$

Likewise, another property of conditional independence, called *intersection*, states that violation of (U_2), $S \perp\!\!\!\perp Y|X$, together with sufficiency, $X \perp\!\!\!\perp Y|S$, also implies violation of nontriviality, $X \perp\!\!\!\perp Y$.

$$S \perp\!\!\!\perp Y|X \ \& \ X \perp\!\!\!\perp Y|S \Rightarrow X \perp\!\!\!\perp Y.$$

Thus, both (U_1) and (U_2) must be violated by any nontrivial sufficient set S.

Note, however, that intersection holds only for strictly positive probability distributions, which means that the Rothman-Greenland condition may be violated if deterministic relationships hold among some variables in a problem. This can be seen from a simple example in which both X and Y stand in a one-to-one functional relationship to a third variable, Z. Clearly, Z is a nontrivial sufficient set yet is not associated with Y given X; once we know the value of X, the probability of Y is determined, and would no longer change with learning the value of Z.

Chapter 11

Minor corrections made throughout Chapter 11, with exception to Section 11.3 which has major revisions – see separate file.

CORRECTIONS/ADDITIONS TO BE IMPLEMENTED BY CAMBRIDGE IN THE SECOND EDITION OF *CAUSALITY* PROOFS

Updated 11/4/08

Chapter 3

page 77 line 3 after Definition 3.2.3,
replace "samples of P" with "samples of $P(v)$"

Chapter 6

page 187 line -3 (before footnote 12)
replace "of X and E is" with "of X and $\{E, Z\}$ is"

page 195 -5 lines before "*Proof of Necessity*"
replace "Figure 6.1" with "Figure 6.3"

Chapter 7

page 251 2nd paragraph, line 4
replace "Section 7.1.2." with "Section 7.2.2."

Chapter 11

pages 333-4 **replace** paragraph starting "Divorcing simple concepts..." with

> Divorcing simple concepts from the province of statistics - the most powerful formal language known to empirical scientists – can be traumatic indeed. Social scientists have been laboring for half a century to evaluate public policies using statistical analysis, anchored in regression techniques, and only recently confessed, with great disappointment, what should have been recognized as obvious in the 1960's: "Regression analysis typically do nothing more than produce from a data set a collection of conditional means and conditional variances" (Berk 2004, p. 237). Economists have gone through a similar trauma with the concept of exogeneity (Section 5.4.3). Even those who recognized that a strand of exogeneity (i.e., superexogeneity) is of a causal variety came back to define it in terms of distributions (Maddala 1992, Hendry 1995) – crossing the demarcation line was irresistible. And we understand why; defining concepts in term of prior and conditional distributions - the ultimate oracles of empirical knowledge - was considered a mark of scientific prudence. We know better now.

page 340 insert 2 new paragraphs after line 2

It is also important to note that the danger of creating new bias by adjusting for wrong variables can threaten randomized trials as well. In such trials, investigators may wish to adjust for covariates despite the fact that, asymptotically, randomization neutralizes both measured and unmeasured confounders. Adjustment may be sought either for improving precision (Cox 1958, pp. 48–55), or to match imbalanced samples, or to obtain covariate-specific causal effects. Randomized trials are immune to adjustment-induced bias when adjustment is restricted to pre-treatment covariates, but adjustment for post-treatment variables may induce bias by the mechanism shown in Figure 11.5 or, more severely, when correlation exists between the adjusted variable Z and some factor that affects outcome (e.g., e_4 in Figure 11.5).

As an example, suppose treatment has a side effect (e.g., headache) in patients that are predisposed to disease Y. If we wish to adjust for disposition and adjust instead for its proxy, headache, a bias would emerge through the spurious path: treatment $\rightarrow$ headache $\leftarrow$ predisposition $\rightarrow$ disease. However, if we are careful never to adjust for any consequence of treatment (not only those that are on the causal pathway to disease), no bias will emerge in randomized trials.

page 342 replace footnote 4 with

[4] In fact, in those cases where "strong ignorability" is used to guide the choice of Z, the guidelines issued are invariably wrong, perpetuating myths such as: "In principle, there is little or no reason to avoid adjustment for a true covariate, a variable describing subjects before treatment." (Rosenbaum 2002, p. 76) or "a confounder must be associated with both treatment and disease," or "strong ignorability requires measurement of all covariates related to both treatment and outcome."

page 342 2nd paragraph, last sentence:
replace "(ch11-eq11x)" with "(11.4)"

page 342 5th paragraph, last sentence:
replace "$\{Y(0), Y(1)\}_s$ from X" with "W from X"

page 342 6th paragraph, 1st line:
replace "designated $\{Y(0), Y(1)\}_s$" with "designated W"

page 342 replace footnote 4 with

[4] In fact, in the rare cases where "strong ignorability" is used to guide the choice of covariates, the guidelines issued are wrong or inaccurate, perpetuating myths such as: "there is no reason to avoid adjustment for a variable describing subjects before treatment," "a confounder is any variable associated with both treatment and disease," or "strong ignorability requires measurement of all covariates related to both treatment and outcome."

page 343 replace Figure 11.7 figure with corrected figure file, "ch11-fig11y" ("$\{Y(0), Y(1)\}_s$" is replaced with "W")

page 343 Figure 11.7 caption, **replace** "ignor-ability" with "ignorability"

page 344 After last paragraph on page, **insert** new text (paragraph, new theorem, paragraph)

The following theorem generalizes the claims above, and is conjectured to be complete.

Theorem 11.3.1 (*Pearl 2008*)
Let T and Z be two arbitrary sets of covariates, and let T_m and Z_m be any two c-minimal subsets of T and Z, respectively. A sufficient condition for the c-equivalence of T and Z is that each of T_m and C_m be c-equivalent to $T_m \cup C_m$ according to the set-subset criterion of Stone (1993) and Robins (1997), that is, T and $T \cup S$ are c-equivalent if S can be partitioned into two subsets, S_1 and S_2, such that:

$$(i') \qquad S_1 \perp\!\!\!\perp X | T$$

and

$$(ii') \qquad S_2 \perp\!\!\!\perp Y | S_1, X, T$$

Having given a conditional-independence characterization of c-equivalence, does not solve or course the problem of identifying admissible sets; the latter is a causal notion and cannot be given statistical characterization.

page 345 **replace** caption of Figure 11.8 with
Figure 11.8 (a) $S_1 = \{Z_1, W_2\}$ and $S_2 = \{Z_2, W_1\}$ are each admissible yet not satisfying C_1 or C_2. (b) No subset of $C = \{Z_1, Z_2, W_1, W_2, V\}$ is admissible.

page 345 **replace** 1st paragraph following Figure 11.8 with

The graph depicted in Figure 11.8 (b) demonstrates the difficulties commonly faced by social and health scientists. Suppose our target is to estimate $P(y|do(x))$ given measurements on $\{X, Y, Z_1, Z_2, W_1, W_2, V\}$, but having no idea of the underlying graph structure. The conventional wisdom is to start with all available covariates $C = \{Z_1, Z_2, W_1, W_2, V\}$, and test if a proper subset of C would yield an equivalent estimand upon adjustment. Statistical methods for such reduction are described in Greenland et al. (1999), Geng et al. (2002), and Wang et al. (2008). For example, $\{Z_1, V\}$, $\{Z_2, V\}$ or $\{Z_1, Z_2\}$, can be removed from C by successively applying conditions C_1 and C_2. This reduction method would produce three irreducible subsets, $\{Z_1, W_1, W_2\}$, $\{Z_2, W_1, W_2\}$, and $\{V, W_1, W_2\}$ all c-equivalent to the original covariate set C. However, none of these subsets is admissible for adjustment, because none (including C) satisfies the back door criterion. While a theorem due to Tian et al. (1998) assures us that any c-equivalent subset of a set C can be reached from C by a step-at-a-time removal method, going through a sequence of c-equivalent subsets, the problem of covariate selection is that, lacking the graph structure, we do not know which (if any) of the many subsets of C is admissible. The next subsection discusses how external knowledge as well as more refined analysis of the data at hand, can be brought to bear on the problem.

page 346 replace 2nd paragraph with

For example, the model depicted in Figure 11.9 is observationally equivalent to that of Figure 11.8 (b). In particular, it satisfies all the conditional independencies implied by the latter, and none others. However, in contrast to Figure 11.8 (b), the sets $\{Z_1, W_1 W_2\}, \{V, W_1, W_2\}$, and $\{Z_2, W_1, W_2\}$ are admissible. Adjusting for the latters would remove bias if the correct model is Figure 11.9 and would produce bias if the correct model is 11.8 (b).

page 346 replace caption of Figure 11.9 with

Figure 11.9 A model that is statistically indistinguishable from that of Figure 11.8 (b), in which the irreducible sets $\{Z_1, W_1, W_2\}$, $\{W_1, W_2, V\}$, and $\{W_1, W_2, Z_2\}$ are admissible.

page 349 in last paragraph of page 349, -3 line from footnote 9, **replace** "absolves one from" with "can cause no harm (Rosenbaum 2002, p. 76) or"

page 350 1st paragraph, **replace** last sentence with

"Examples are abound (e.g., Figure 6.3) where adding a variable to the analysis, not only is not needed but would introduce irreparable bias."

page 360-1 Section 11.4.6, 1st paragraph, last sentence - **replace** with

"In so doing, she unveils several areas in need of systematic clarifications; I will address them in turn."

page 361 replace in last sentence of page:

"Thus, Cartwright is mistaken in asserting that" with "Thus, it would be a mistake to assume that"

page 361 replace in last sentence of page:

Claim (1) may apply in some cases, but certainly not in most;

page 362 replace line 3 of page with

'Claim (1) may apply in some cases, but certainly not in most; in many studies our goal is"

page 362 replace paragraph 6 with

Ironically, shunning mathematics based on ideal atomic intervention may condemn scientists to ineptness in handling realistic non-atomic interventions.

References

page 438 In [Kuroki and Miyakawa, 2003] citation, change "Journal of the Japanese Royal Statistical society" to "Journal of the Royal Statistical Society"

CORRECTIONS/QUESTIONS FOR PROOFS OF THE SECOND EDITION OF *CAUSALITY* (dated 11/13/08)

Updated 12/19/08 - still working on

Preface

page xviii stet original ending and add new heading (check copy edit format throughout for style/consistency)
J.P.
Los Angeles
August 1999

Preface to Second Edition

per judea note - "if space permits [where?] - popular reception of the unifying theory of Modifiable Structural Models (MSM) presented in this book call..."

Chapter 1

page 24 question: in (iii), should "$X = x$," be on previous line?

page 47 question: last sentence, 2nd paragraph. should it be "(i.e., polytrees)" not "a polytrees"?

Chapter 2

page 48 question: in paragraph following def 2.4.1, should it be "$(X \perp\!\!\!\perp Y|Z)_P$" (cap P)?

page 64 question: what is the format of "Postscript" heading for the Second Edition?

Chapter 3

page 105 heading for 2nd edition postscript left off after line 2. check style.

page 105 in paragraph 4 of (missing) postscirpt heading, center the asterisk in $do(*)$

page 106 insert comma between greenland:etal99 and kaufman:etal05 cites, "Greenland et al. (1999), Kaufman et al. (2005)"

Chapter 4

page 114 section 4.3, line 5: "*do* calculus" should be "*do*-calculus"

page 114 section 4.3.1, line 3: "then states the at least" should be "then states that at least" (the should be that)

page 114 section 4.3.1, line 4: "*do* calculus" should be "*do*-calculus"

page 114 section 4.3.1, line 5: "*do* calculus" should be "*do*-calculus"

page 131 section 4.5, footnote 9: replace "(Pearl 2001)" with "Pearl (2001)"

page 132 completely replace section 4.5.5 (reduced to fit page length limitaion)

Chapter 5

page 135 pages from hereon are -2 and revert back to original page numbering

page 173 (171) missing "Notes Added to Second Edition" heading before section 5.5.1

Chapter 6

page 197 (195) correct both formulas on bottom of page:
1st should be $X \perp\!\!\!\perp S \;\&\; X$... (&, not |)
2nd should be $S \perp\!\!\!\perp Y | X \;\&\; X$... (&, not $*$)

Chapter 7

page 202 (200) replace "**Postscript**" heading with "**Postscript to Second Edition**"

page 219 (217) line 4, replace "Section 11.7.1.)" with "Section 11.7.1)."

Chapter 8 & 9

no corrections.

Chapter 10

page 331 (329) replace "Postscript" heading with "Postscript to Second Edition"

Chapter 11

page 347 (345) question: -2 from bottom, should it be "solve or cause the problem" no "solve or course"?????

page 348 (346) question: in 1st paragraph, lines 11 and 15, change "is" to "are"?????

page 352 (350) correction: 2nd paragraph, line 5: replace "or the need to reason" with "and can absolve one from thinking"

page 352 (350) in footnote 9:
replace "(e.g., Rubin 2004)." with "(e.g., Rubin 2004, 2008)."

ADDITIONS TO SUBJECT INDEX TO BE IMPLEMENTED BY CAMBRIDGE IN THE SECOND EDITION OF *CAUSALITY* PROOFS

Updated 11/26/08

final page numbers to be double-checked

ADD TO SUBJECT INDEX

ADD TO AUTHOR INDEX

COPYEDIT INSERTS FOR PROOFS OF THE SECOND EDITION OF *CAUSALITY* (Updated 1/9/08)

insert-43 (sent to cup but not implemented): Such thought experiments tell us that certain patterns of dependency, void of temporal information, are conceptually characteristic of certain causal directionalities and not others. Reichenbach (1956) who was the first to wonder about the origin of those patterns suggested that they are characteristic of Nature, reflective of the second law of thermodynamics. Rebane and Pearl (1987) posed the question in reverse, and asked whether the distinctions among the dependencies associated with the three basic causal substructures: $X \longrightarrow Y \longrightarrow Z$, $X \longleftarrow Y \longrightarrow Z$ and $X \longrightarrow Y \longleftarrow Z$ can be used to uncover genuine causal influences in the underlying data generating process. They quickly realized that the key to determining the direction of the causal relationship between X and Y lies in "the presence of a third variable Z that correlates with Y but not with X," as in the collider $X \longrightarrow Y \longleftarrow Z$, and developed an algorithm that recovers both the edges and directionalities in the class of causal graphs that they considered (i.e., polytrees).

The investigation in this chapter formalizes these intuitions and extends the Rebane-Pearl recovery algorithm to general graphs, including graphs with unobserved variables.

insert-60: Rephrased as a logical guarantee, we can categorically assert that the IC* algorithm will never label an arrow $a \to b$ as genuine if in fact a has no causal influence on b and if the observed distribution is stable relative to its underlying causal model.

insert-61: and this is indeed what scientists attempt to accomplish through controlled experimentation. Absent such experimentation, the best one can do is to rely on virtual control variables, like those revealed by Nature through the dependence patterns of Definitions 2.7.1–2.7.4.

insert-63: It is people's quest for stability that explains why they find it impossible to exemplify intransitive patterns of dependencies except by envisioning a common effect of two independent causes (see page 43). Any story that convincingly exemplifies a given pattern of dependencies must sustain that pattern regardless of the numerical values of the story parameters – a requirement we called "stability."

insert-64 (new paragraph): A program of benchmarks of causal discovery algorithms, named "Causality Workbench," has been reported by Guyon et al. (2008ab; http://clopinet.com/causality). Regular contests are organized in which participants are given real data or data generated by a concealed causal model, and the challenge is to predict the outcome of select set of interventions.

insert-70: Balke and Pearl (1994ab) proposed it as the basis for defining counterfactuals, see (3.51).

insert-73 (not implemented per 3/13/07): Replace equation between (3.11) and (3.12) with:

$$\begin{aligned} P(pa_i|do(x_i')) &= P(pa_i); \\ \frac{P(s_i, pa_i, x_i'|do(x_i'))}{P(s_i', pa_i, x_i'|do(x_i'))} &= \frac{P(s_i, pa_i, x_i')}{P(s_i', pa_i, x_i')}. \end{aligned}$$

insert-79 (2nd line from top): [recasts the confounding problem in counterfactual vocabulary but falls short of providing] researchers with a workable criterion to guide the choice of covariates (see Section 11.3.2).

insert-98a: ... mathematical equivalence] – a theorem in one entails a theorem in the other. In this section we highlight the key methodological differences.

insert-98b: ; but lacking formal semantics, e.g., (3.51), did not progress beyond the stage of a "framework."

insert-104: Robins's pioneering research has shown that, to properly manage multistage problems with time-varying treatments, the opaque condition of "ignorability" (3.53) should be broken down to its sequential constituents. This has led to the sequential back-door criterion of (Theorem 4.4.5).

insert-106: [lacking these conceptual tools,] were unable to address the issue of covariate selection (Rosenbaum 2002, p. 76; Rubin 2007, 2008a) and were led to [dismiss this and other scientific questions as "ill-defined"..]

insert-106b (bold subtitle):

Chapter road map to the main results

The three key results in this chapter are: 1. The control of confounding, 2. The evaluation of policies, and 3. The evaluation of counterfactuals.

1. The problem of controlling confounding bias is resolved through the back-door condition (Theorem 3.3.2, pp. 79-80) – a criterion for selecting a set of covariates that, if adjusted for, would yield an unbiased estimate of causal effects.
2. The policy evaluation problem – to predict the effect of interventions from non-experimental data – is resolved through the *do*-calculus (Theorem 3.4.1, pp. 85-86) and the graphical criteria that it entails (Theorem 3.3.4, p. 83; Theorem 3.6.1, p. 105). The completeness of *do*-calculus implies that any (nonparametric) policy evaluation problem that is not supported by an identifying graph, or an equivalent set of causal assumptions, can be proven "unsolvable."
3. Finally, equation (3.51) provides a formal semantics for counterfactuals, through which probabilities of of counterfactuals can be defined and evaluated in the framework of scientific theories (see Chapter 7).

insert-114a: In this section we characterize a wider class of models in which the causal effect $P(y|\hat{x})$ is identifiable. This class is subsumed by the one established by Tian and Pearl (2002a) in Theorem 3.7 and the complete characterization given later in Shpitser and Pearl (2006b). It is brought here for its intuitive appeal.

insert-114b: Theorem 4.3.1 characterizes a class of models in the form of four graphical conditions, anyone of which is sufficient for the identification of $P(y|\hat{x})$ when X and Y are singleton nodes in the graph. Theorem 4.3.2 then states the at least one of these four conditions must hold in the model for $P(y|\hat{x})$ to be identifiable in *do*-calculus. In view of the completeness of *do* calculus, we conclude that one of the four conditions is necessary for any method of identification compatible with the semantics of Definition 3.2.4).

insert-118: Although the completeness results of Shpitser and Pearl (2006a) now offer a precise characterization of the boundary between the identifying and non-identifying models (see discussion following Theorem 3.6.1), the conditions of Theorem 4.3.2 may still be useful on account of their simplicity and intuitive appeal.

insert-123: Note that admissibility (4.5) requires that each subsequence $X_1, ..., X_{k-1}, Z_1, ..., Z_k$ blocks every "action-avoiding" back-door path from X_k to Y (see page 103).

insert-165: The manipulative account that we have invoked in defining the empirical content of structural equations (Definition 5.4.1) is adequate in linear systems, where most causal quantities of interest can be inferred from experimental studies at the population level (see Section 11.7.1). In nonlinear and nonparametric models, we occasionally need to go down to the individual unit level and invoke the (more fundamental) counterfactual reading of structural equations, as articulated in equation (3.51) and footnote 17, page 162. The analysis of indirect effects is a case in point; its definition (4.14) rests on nested counterfactuals and cannot be expressed in terms of population averages. Such analysis is necessary to give indirect effects operational meaning, independent of total and direct effects (see Section 11.4.2). With the help of counterfactual language, however, we can give indirect effects a simple operational definition: The indirect effect of X on Y is the increase we would see in Y while holding X constant and increasing Z to whatever value Z would have attained under a unit increase of X (see Section 4.5.5 for a formal definition). In linear systems, this definition coincides, indeed, with the difference between the total and direct effects. See chapter 11 for further discussion on the role of indirect effects in social science and policy analysis (Pearl 2005a).

insert-170: Some of these confusions are reflected in the many questions that I have received from readers (Section 11.5), to whom I dedicated a "SEM Survival Kit" (Section 11.5.3) – a set of arguments for defending the causal reading of SEM and its scientific rationale.

insert-195: Note however that, although this test can screen out some obviously bad sets S claimed to be sufficient, it has nothing to do with sufficiency or confounding; it merely tests for nontriviality, i.e., that adjusting for S would change the association between X and Y. When we find a nontrivial set S, we still cannot be sure whether the association was unbiased to start with (as in Figure 6.1) or that it turned unbiased after the adjustment.

insert-200: A notable exception is Larry Wasserman's *All Of Statistics* (Wasserman 2004) the first statistics textbook to treat Simpson's reversal in its correct causal context.

insert-203: (iii) F is a set of functions $\{f_1, f_2, \ldots, f_n\}$ such that each f_i is a mapping from (the respective domains of) $U_i \cup PA_i$ to V_i, where $U_i \subseteq U$ and $PA_i \subseteq V \setminus V_i$ and the entire set F forms a mapping from U to V. In other words, each f_i in

$$v_i = f_i(pa_i, u_i), \;\; i = 1, \ldots, n,$$

assigns a value to V_i that depends on (the values of) a select set of variables in $V \cup U$, and the entire set F has a unique solution $V(u)$.[3] [4]

insert-204 (for clarity of handwritten edit): Let X and Y be two subsets of variables in V. The counterfactual sentence "Y would be y (in situation u), had X been x" is interpreted as the equality $Y_x(u) = y$, with $Y_x(u)$ being the potential response of Y to $X = x$.

insert-206 (for clarity of handwritten edit): The definition of Y_x and Y'_x in terms of the solution for Y in two distinct submodels, governed by a standard probability space over U, neutralizes these objections by interpreting the contradictory joint statement as an ordinary event in U-space.

insert-226: If X is the set of variables directly constrained by G_k, we can ask whether there is one member of X, say X_k that accounts for the changes in all the other solutions. If the identity of that predictive member remains the same for all choices of a_k and u, then we designate X_k as the *dependent* variable in G_k.

insert-243: In contrast to the structural modeling, however, this variable is not derived from a causal model or from any formal representation of scientific knowledge, but is taken as a primitive—that is, an unobserved variable that reveals its value only when x coincides with the treatment actually received, as dictated by the consistency rule $X = x \Rightarrow Y_x = Y$ (equation (7.20)). Consequently, the potential-outcome framework does not provide a mathematical model from which such rules could be *derived* or on the basis of which an axiomatic characterization could be attempted to decide, for example, whether additional rules should be deployed, or whether a given collection of potential-outcome expressions is redundant or contradictory.

[3] The choice of PA_i (connoting *parents*) is not arbitrary, but expresses the modeller's understanding of which variables Nature must consult before deciding the value of V_i.

[4] Uniqueness is ensured in recursive (i.e., acyclic) systems. Halpern (1998) allows multiple solutions in non-recursive systems.

insert-244a: and became a mathematical fact through the explicit translation of equation (3.51) followed by the completeness result of Theorem 7.3.6.

insert-244b: Indeed, researchers who shun structural equations or graphs tend to avoid subject matter knowledge in their analyses.

insert-251: satisfy the back-door condition, because the inequality in Definition 7.5.1 can always be satisfied by conditioning on some imagined factor F (as in Figure 6.2 (c)) that generates spurious associations between C and E. nd became a mathematical fact through the explicit translation of equation (3.51) followed by the completeness result of Theorem 7.3.6.

insert-259 (clarifying end of Preface): and (iii) that prior knowledge can be harnessed effectively to obtain Bayesian estimates of those impacts.

insert-281: Methodologically, the message of this chapter has been to demonstrate that, even in cases where causal quantities are not identifiable, reasonable assumptions about the structure of causal relationships in the domain can be harnessed to yield useful quantitative information about the strengths of those relationships. Once such assumptions are articulated in graphical form and re-encoded in terms of canonical partitions, they can be submitted to algebraic methods that yield informative bounds on the quantities of interest. The canonical partition further allows us to supplement structural assumptions with prior beliefs about the population under study and invite Gibbs sampling technique to facilitate Bayesian estimation of the target quantities.

insert-302: Substituting these estimates in (9.29), which provides a lower bound on PN (see (11.42)), we obtain

insert-303: may provide a necessary test for the adequacy of the experimental procedures. For example, if the frequencies in Table 9.2 were slightly different, they could easily yield a value greater than unity for PN in (9.53) or some other violation of the fundamental inequalities of (9.33). Such violation would indicate an incompatibility of the experimental and nonexperimental groups due, perhaps, to inadequate sampling.

insert-343 (new paragraph): One application where the symbiosis between the graphical and counterfactual frameworks has been extremely useful is in estimating the effect of treatments on the treated: ETT = $P(Y_{x'} = y|x)$ (See Sections 8.2.5 and 11.9.1). This counterfactual quantity (e.g., the probability that a treated person would recover if not treated, or the rate of disease among the exposed, had the exposure been avoided) is not easily analyzed in the *do*-calculus notation. The counterfactual notation, however, allows us to derive a useful conclusion: Whenever a set of covariates Z exists that satisfies the back-door criterion, ETT can be estimated from observational studies. This follows directly from

$$(Y \perp\!\!\!\perp X|Z)_{G_{\underline{X}}} \Rightarrow Y_{x'} \perp\!\!\!\perp X|Z$$

which allows us to write

$$\begin{aligned}\text{ETT} &= P(Y_{x'} = y|x) \\ &= \sum_z P(Y_{x'} = y|x, z)P(z|x) \\ &= \sum_z P(Y_{x'} = y|x', z)P(z|x) \\ &= \sum_z P(y|x', z)P(z|x)\end{aligned}$$

The graphical demystification of "strong ignorability" also helps explain why the probability of causation $P(Y_{x'} = y'|x, y)$ and, if fact, any counterfactual expression conditioned on y, would not permit such a derivation and is, in general, non-identifiable (See Chapter 9.)

insert-345: A natural attempt would be to impose the condition that S_1 and S_2 each be c-equivalent to $S_1 \cup S_2$, and invoke the criterion of Stone (1993) and Robins (1997) for the required set-subset equivalence. The resulting criterion, while valid, is still not complete; there are cases where S_1 and S_2 are c-equivalent yet not c-equivalent to their union. A theorem by Pearl and Paz (2008) broadens this condition using irreducible sets.

insert-346 (footnote after "and no other."): Semi-Markovian models may also be distinguished by functional relationships that are not expressible as conditional independencies (Verma and Pearl, 1990; Tian and Pearl 2002b; Shpitser and Pearl 2008). We do not consider these useful constraints in this example.

insert-347: In Chapter 3, for example, we demonstrated how the measurement of an additional variable, mediating between X and Y, was sufficient for identifying the causal effect of X on Y. This facility can also be demonstrated in Figure 11.8(b); measurement of a variable Z judged to be on the pathway between X and Y would render $P(y|do(x))$ identifiable and estimable through equation (3.29). This is predicated, of course, on Figure 11.8(b) being the correct data generating model. If, on the other hand, it is Figure 11.9 that represents the correct model, the causal effect would be given by

$$\begin{aligned}P(y|do(x)) &= \sum_{pa_X} P(y|pa_X, x)P(pa_X) \\ &= \sum_{z_1,w_1,w_2} P(y|x, z_1, w_1, w_2)P(z_1, w_1, w_2)\end{aligned}$$

which might or might not agree with equation (3.29). In the latter case, we would have good reason to reject the model in Figure 11.9 as inconsistent, and seek perhaps additional measurements to confirm or refute Figure 11.8(b).

Auxiliary experiments may offer an even more powerful discriminatory tool than auxiliary observations. Consider variable W_1 in figure 11.8 (a). If we could conduct a controlled experiment with W_1 randomized, instead of X, the data obtained would enable us to estimate the causal effect of X on Y with no bias (see Section 3.4.4). At the very least, we would be able to discern whether W_1 is a parent of X, as in Figure 11.9 , or an indirect ancestor of X, as in Figure 11.8 (b).

In an attempt to adhere to traditional statistical methodology, some causal analysts have adopted an method called "sensitivity analysis" (e.g., Rosenbaum 2002, pp. 105-170), which gives the impression that causal assumptions are not invoked in the analysis. This, of course, is an illusion. Instead of drawing inferences by assuming the absence of certain causal relationship in the model, the analyst tries such assumptions and evaluates how strong alternative causal relationships must be in order to explain the observed data. The result is then submitted to a judgment of plausibility, the nature it is no different from the judgments invoked in positing a model like the one in Figure 11.9. In its richer setting, sensitivity analysis amounts to loading a diagram with causal relationships whose strengths is limited by plausibility judgments and, given the data, attempting to draw conclusions without violating those plausibility constraints. It is a noble endeavor, which thus far has been limited to problems with very small number of variables. The advent of diagrams promises to expand the applicability of this method to more realistic problems.

insert-359: The method you are proposing, to replace the current equation $x = f(pa_X)$ with $x = g(f(pa_X), I, z)$, requires that we know the functional forms of f and g, as in linear systems or, alternatively, that the parents of X are observed, as in the Process Control example on page 74. These do not hold, however, in the non-parametric, partially observable settings of Chapters 3 and 4, which might render it impossible to predict the effect of the proposed intervention from data gathered prior to the intervention, a problem we called *identification*. Because pre-intervention statistics is not available for variable I, and f is unknown, there are semi-Markovian cases where $P(y|do(x))$ is identifiable while $P(y|do(x = g(f(pa_X), I, z)))$ is not; each case must be analyzed on its own merit. It is important therefore to impose certain standards on this vast space of potential interventions, and focus attention on those that could illuminate others.

insert-360 (replace section 11.4.5):

11.4.5 Causation Without Manipulation!!!

Question to Author,

In the analysis of direct effects, Section 4.5 invokes an example of sex discrimination in school admission and, in several of the formulas, gender is placed under the "hat" symbol or, equivalently, as an argument of the *do*-operator. How can gender be placed after the *do*-operator when it is a variable that cannot be manipulated?

Author's Reply

Since Holland coined the unfortunate phrase "No Causation Without Manipulation" (Holland 1986) many good ideas were stifled or dismissed from causal analysis. To suppress talk about how gender causes the many biological, social and psychological distinctions between males and females is to suppress 90% of our knowledge about gender differences.

Surely we have causation without manipulation. The moon causes tides, race causes discrimination, and sex causes the secretion of certain hormones and not others. Nature is a society of mechanisms that relentlessly sense the values of some variables and determine the values of others, and does not wait for a human manipulator before activating those mechanisms.

True, manipulation is one way (albeit a crude one) for scientists to test the workings of mechanisms, but it should not in any way inhibit causal thoughts, formal definitions and mathematical analyses of the mechanisms that propel the phenomena under investigation. It is for that reason, perhaps, that scientists invented counterfactuals; it permit them to state and conceive the realization of antecedent conditions without specifying the physical means by which these conditions are established;

The purpose of the "hat" symbol in Definition 4.5.1 is not to stimulate thoughts about possible ways of changing applicants' gender, but to remind us that any definition concerned with "effects" should focus on causal links and filter out spurious associations from the quantities defined. True, in the case of gender, one can safely replace $P(y|do(\textit{female}))$ with $P(y|\textit{female})$, because the mechanism determining gender can safely be assumed independent of the background factors that influence Y (thus ensuring no confounding). But as a general definition, and even as part of an instructive example, mathematical expressions concerned with direct effects and sex discrimination should maintain the hat symbol. If nothing else, placing "female" under the "hat" symbol should help propagate the long-overdue counter-slogan: "Causation without manipulation? You bet!"

insert-374: ... logic] of counterfactuals, is not given an explicit formal exposition: it is relegated to a semi-formal [footnote...

insert-379: For completeness, I reiterate here explicitly (using parenthetical notation) the two fundamental connections between counterfactuals and structural equations.

1. The structural definition of counterfactuals is:

$$Y_M(x,u) = Y_{M_x}(u)$$

   Read: For any model M and background information u, the counterfactual conditional "Y if X had been x," is given by the solution for Y in submodel M_x, (i.e., the mutilated version of M with the equation determining X replaced by $X = x$.)

2. The empirical claim of the structural equation $y = f(x, e(u))$ is:

$$Y(x,z,u) = f(x,e(u)),$$

   for any set Z not intersecting X or Y.

   Read: had X and Z been x and z, respectively, Y would be $f(x, e(u))$, independently of z, and independently of other equations in the model.

insert-394 (new paragraph): It is important to mention at this point that the canonical partition conception, coupled with the linear programming method developed in Balke and Pearl (1994a, 1995ab) has turned into a powerful analytical tool in a variety of applications. Tian and Pearl (2000) applied it to bound probabilities of causation; Kaufman et al. (2005) and Cai et al. (2008) used it to bound direct effects in the presence of confounded mediation and, similarly, Imai et al. (2008) used it to bound natural direct and indirect effects. The closed-form expressions derived by this method enable researchers to assess what features of the distribution are critical for narrowing the widths of the bounds.

Rubin (2004), in an independent exploration, attempted to apply canonical partitions to the analysis of direct and indirect effects within the traditional potential-outcome framework but, lacking the graphical and structural perspectives, was led to conclude that such effects are “ill-defined” and “more deceptive than helpful.” I believe readers of this book, guided by the structural roots of potential-outcome analysis, will reach more positive conclusions (See Sections 4.5 and 11.4.2).

ANSWERS TO RUSSELL'S QUESTIONS
SECOND EDITION OF *CAUSALITY* (Updated 1/20/09)

insert-251 should be :

satisfy the backdoor condition, because the inequality in Definition 7.5.1 can always be satisfied by conditioning on some imagined factor F (as in Figure 6.2 (c)) that generates spurious associations between C and E.

insert 347 replace :

the nature it is no different from the judgments invoked in
with
the nature of which is no different from the judgments invoked in

insert-159 (replace last sentence, 1st paragraph with) :

Remarkably, graphical methods perform this computation without knowledge of f_2, f_3 and $P(\epsilon_2, \epsilon_3, u)$ (Section 3.3.2).

This is indeed the essence of identifiability in nonparametric models. The ability to answer interventional queries *uniquely*, from the data and the graph, is precisely how Definition 3.2.3 interprets the identification of the causal effect $P(y|do(x))$. As we have seen in Chapters 3 and 4, that ability can be discerned graphically, almost by inspection, from the diagrams that accompany the equations.

page 443, replace :

Pearl and Paz 2009 to
with:
Pearl and Paz 2008

page 444, replace :

Pearl 2009 to
with:
Pearl 2008
Also, Pearl 2008 should read "The mathematics of causal relations"

“HOLD-OFF” CORRECTIONS NOT IMPLEMENTED

Updated 11/6/00

page 21 **insert** “Jensen, Cowles, Haddi, Neapolitan” - add to index?

page 141 per jp - do not add (9/11/2k) **insert** in 2nd paragraph, “Edwards 1995” before “Whittaker 1990”

page 269 **insert** in paragraph starting “The analysis of ACE...”. Insert “Heckman 1986;” before “Angrist and Imbens”

page 357 **insert** either “Summary” heading (before “I now wish” paragraph) or “***” (three stars centered)

page 361 **add** to citations: Edwards, D. (1995). *Introduction to Graphical Modelling*, New York: Springer-Verlag.

page 364 per jp, “keep for 2nd edition” 9/11/2k **add** to citations: Heckman, J.J., and R.R. Robb. (1986). Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes. In H. Wainer (Ed.), *Drawing Inferences from Self-Selected Samples*, pp. 63–113. New York, NY: Springer-Verlag,

page 376 **insert** “Edwards, D., 141” before “Eells”

page 377 **insert** “Morgenstern, M., 183” before “Mueller”
insert “Poole, C., 183” before “Poole, D.”

page 384 center tables, e.g. r. 299

CORRECTIONS/ADDITIONS TO NAME INDEX IN THE SECOND EDITION OF *CAUSALITY*

2/25/09 Updated

ADD TO AUTHOR INDEX

*** designates new entry**

```
Ader, H.:  172 (not 171)
Aldrich, J.:  change 171 to 172
Angrist, J.D.:  change 243-5 to 244-5
* Arah, O.A., 200
* Avin, C., 132
```

Balke, A.: insert 70 and 257
Bayes, T.: insert 242
Bentler, P.: change 171 to 172
* Berk, _., 333 (double check insert/page number)
* Bessler, D., 64
Bollen, K.A.: remove 165, insert 172
Bonet, B.: insert 275, 308
* Brito, C., 171
* Brumback, B., 106
* Cai, Z., 126
Cartwright, N.: insert 106 and 172
* Chalak, K., 106
* Ciecro, _., 259 ????
* Cole, S.R., 129
Cowell, R.G.: insert 141
Cox, D.R.: insert 200, 340, 353
* Danks, D., 64, 253
Dawid, A.P.: insert 104-5, 111, 132, 141, check 341 epilogue
DeFenetti, B. (not De Fenetti)
* Demiralp, S., 64
Descartes, R., 335, 350 - check epilogue page #s
* Edwards, D., 141
* Elisseeff, A., 64
Fikes, R.E.: remove 725, insert 225
Fisher, R.A.: insert 127, check epilogue pages #s 340, 348, 353
* Frangakis, C.E., 264
(Fratiannea, A., 60) - remove
Freedman, D.A.: change 105 to 104-5
Galieo, G.: check epilogue page #s 334-6, 342
Galles, D.: remove 86 and 131, insert 257
Gallon, F.: check epilogue page # 339
Geiger, D.: change 12 to 11-2
* Geneletti, S., 132
* Gibbs, _., 270, 275-7, 280-1
Glymour, C.: change 63 to 63-4, insert 80, 106, 253
Goldberger, A.S.: change 171 to 172
Good, I.J.: insert 62
* Gopnik, A., 64, 253
Granger, C.W.J.: insert 64
Greenland, S.: remove 165 and 257, insert 106, 131, 271
Haavelmo, T.: insert 171
Hagenaars, J.: change 171 to 172
Hall, N.: change/insert 329 to 329-30

Halpern, J.Y.: change/insert 329 to 329-30
* Hayduk, L., 172
Heckerman, D.: insert 211
Heckman, J.J.: insert 106, 139, 171-2, 216
* Herna’n, M.A., 106, 129
* Herna’ndez-Di’az, S. , 106
* Hiddleston, E., 330
Hitchcock, C.R., insert 330
Holland, P.W.: insert 106, remove 257
(Hollander, M., 281) - remove
(Holyoak, K.J., 60) - remove
Hoover, K.: insert 64, change 171 to 172
* Hopkins, M., 329
* Hoyer, P.O., 64
* Huang, Y., 86, 105
Hume, D.: check epilogue page #s 336, 343
Hurwicz, L.: insert 171
* Hyvarinen, A., 64
Imbens, G.W.: remove 243, (stet 244-5), insert 274-5
* Kano, Y., 64
* Kaufman, S., 106
* Kaufman, J.S., 106
* Kerminen, A.J., 64

Geng, p. 345 (add etal names)
Glymour, p. 64
Gopnik, p. 64 (add etal names)
Granger, p. 64
Greenland, p. 105 (add etal names), 130ish, 194ish, 345 (etal)
Haavelmo, p. 171
Hall, p. 330ish
Halpern, p. 329, 330ish
Heckman, p. 105, 171, 269
Hendry, p. 333
Hernan, p. 105 (add etal names), 129
Hiddleston, p. 330ish
Hitchcock, p. 330ish

MISSED CORRECTIONS/ADDITIONS IN THE SECOND EDITION OF *CAUSALITY*

2/25/09 Updated

Chapter 3

page 67, calligraphed zs begin to appear in chapter, not italicized as should be

page 72, footnote 4, line 3
replace
"said formula"
with
"this formula"

page 75, Notes:

- lowercase zs appear calligraphed, not italicized.
- 1st and 2nd line in (3.16), subs under $\sum$ are too small
- in (3.17), sub under $\sum$ is too small

page 86, paragraph after Corollary 3.4.2
replace
Whether Rules 1–3 are sufficient for deriving all identifiable causal effects remains an open question. However, the task of finding a sequence of transformations (if such exists) for reducing an arbitrary causal effect expression can be systematized and executed by efficient algorithms (Galles and Pearl 1995; Pearl and Robins 1995), to be discussed in Chapter 4. As we illustrate in Section 3.4.3, symbolic derivations using the hat notation are much more convenient than algebraic derivations that aim at eliminating latent variables from standard probability expressions (as in Section 3.3.2, equation (3.24)).
with
Rules 1–3 have been shown to be *complete*, namely, sufficient for deriving all identifiable causal effects (Shpitser and Pearl 2006a; Huang and Valtorta 2006). Moreover, as illustrated in Section 3.4.3, symbolic derivations using the hat notation are more convenient than algebraic derivations that aim at eliminating latent variables from standard probability expressions (as in Section 3.3.2, equation(3.24)). However, the task of deciding whether a sequence of rules exists for reducing an arbitrary causal

effect expression has not been systematized, and direct graphical criteria for identification are therefore more desirable; see Theorem 3.6.1 and Chapter 4.

page 100, lines 1-2 after (3.53)

replace

"(an assumption that was termed "conditional ignorability" by Rosenbaum and Rubin 1983), then"

with

"(an assumption that was termed "conditional ignorability" by Rosenbaum and Rubin (1983)), then"

Chapter 4

page 108, paragraph 3, line 1

replace

"Traditional decision theory"

with

"Commonsensical decision theory"

page 109, 2nd paragraph, 2nd line of mnemonic rhymes

replace

On facts that preceded the act,

with

On what could have caused the act,

page 129, **add** sentence to end of footnote 8

"Cole and Hernán (2002) present examples in epidemiology."

page 132, in inserted page for 4.5.5, end of last sentence on page

replace

"Dawid (2000)."

with

"Dawid (2000) and Geneletti (2007)."

Chapter 8

page 264, **replace**

" Frangakis and Rubin (2002) dubbed them "Principal Stratification" and found them rather insightful. In the potential-outcome model (see Section 7.4.4),"

with

"In an experimental framework, this partition goes back to (Greenland and Robins 1986) and was dubbed "Principlal Stratification" by Frangakis and Rubin (2002). In this framework (see Section 7.4.4),"

Bibliography

page 447, correct in cites for Shimizu et al., 2005 and 2006
should be “S. Shimizu” and “A. Hyvärinen” in both references

CORRECTIONS TO THE SECOND EDITION OF *CAUSALITY*

4/15/09 Updated

throughout book: Replace “ACE*” with “ETT”

throughout book: No space between “Figure #” and (a) – i.e., Figure 11.8(a) (but not too close)

throughout book: Increase font size of subscripts, supscripts, subscript subs and sups, and supscript subs and sups

throughout book: Equalize the space between over/under bars and variables (i.e., $\overline{x}$, $\underline{x}$)

throughout book: Equalize the space between hats and variables (i.e., $\hat{x}$)

throughout book: Font size of quoted materials fluctuates – keep larger as on page 238 of the 1st edition

throughout bibliography : Numbered editions are in numerical format, not written (i.e., “2nd edition,” not “second edition”)

throughout bibliography : Numbered Proceedings are spelled out, not numerical (i.e., “Seventh Proceedings,” not “7th Proceedings”)

page xix, Add new paragraph to the end of Preface of Second Edition (marked “insert-xix”)

I hope that each of these groups will find the unified approach taken in this book to be both inspirational and instrumental in tackling new challenges in their respective fields.

page 64 add new paragraph, “insert-64” as marked

Verma and Pearl (1990) noted that two latent structures may entail the same set of conditional independencies and yet impose different equality constraints on the joint distributions. These constraints, dubbed “dormant independencies,” were characterized systematically in Tian and Pearl (2002b) and Shpitser and Pearl (2008); they promise to provide a powerful new discovery tool for structure learning.

page 80, Replace: where the set Z is called “sufficient set” (or “admissible”) for control of confounding.
With “insert-80”:
where the set Z is called “sufficient set;” *admissible* or *deconfounding* set is a better term.

page 264, "insert-264":

In an experimental framework, this partition goes back to Greenland and Robins (1986) and was dubbed "Principlal Stratification" by Frangakis and Rubin (2002). In this framework

page 106, clarification of corrections, marked "insert-106":

(2001), Hernán et al. (2002), Hernán et al. (2004), Greenland and Brumback (2002), Greenland et al. (1999ab), Kaufman et al. (2005), Petersen et al. (2006), Hernández-Díaz et al. (2006), VanderWeele and Robins (2007), and Glymour and Greenland (2008).

Interesting applications of the front-door criterion (Section 3.3.2) were noted in social science (Morgan and Winship 2007) and economics (Chalak and White 2006).

Some advocates of the "potential outcome" approach have been most resistant to accepting graphs or structural equations as the basis for causal analysis and, lacking these conceptual tools, were unable to address the issue of covariate selection (Rosenbaum 2002, p. 76; Rubin 2007, 2008a) and were led to dismiss important scientific concepts as "ill-defined," "deceptive," "confusing" (Holland 2001; Rubin 2004, 2008b), and worse (Rubin 2009). Lauritzen (2004) and Heckman(2005) have criticized this attitude.

page 131, replace last paragraph on page with

In particular, expression (4.12) is both valid and identifiable in Markovian models, where all *do*-operators can be eliminated using Corollary 3.2.6; for example

$$P(z|do(x)) = \sum_t P(z|x, pa_X = t)P(pa_X = t) \qquad (4.13)$$

page 385, paragraph 3, replace marked sentence with "insert-385":

Your earlier statement describes the habitual, uncritical use of conditiong among Bayesian researchers, which occasionally lures the unwary into blind alleys (e.g., Rubin (2009)):

Bibliography New entries:

Darwiche, 2009 A. Darwiche. *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press, New York, 2009.

McDonald, 2002a R.P. McDonald. What can we learn from the path equations?: Identifiability, constraints, equivalence. *Psychometrika*, 67(2):225-249, 2002.

Pearl, 2009 J. Pearl. Letter to the Editor: Remarks on the method of propensity scores. *Statistics in Medicine*, 28:1415-1416, 2009.

Rubin, 2009 D.B. Rubin. Author's Reply: Should observational studies be designed to allow lack of balance in covariate distributions across treatment groups? *Statistics in Medicine*, 28:1420-1423, 2009.

Shadish and Clark, 2006 W.R. Shadish and M.H. Clark. A radomized experiment comparing random to nonrandom assignment. Unpublished Paper. University of California. Merced, 2006.

Shrier, 2009 I. Shrier. Letter to the Editor: Propensity scores. *Statistics in Medicine*, 28:1317-1318, 2009.

Sjölander, 2009 A. Sjölander. Letter to the Editor: Propensity scores and M-structures. *Statistics in Medicine*, 28:1416-1423, 2009.

Name Index - additions/corrections: