Finally, certain concepts that are ubiquitous in human discourse can be defined only in the Laplacian framework. We shall see, for example, that such simple concepts as "the probability that event B occured *because* of event A" and "the probability that event B would have been *different* if it were not for event A" cannot be defined in terms of purely stochastic models. These so-called *counterfactual* concepts will require a synthesis of the deterministic and probabilistic components embodied in the Laplacian model.

1.4.1 Structural Equations

In its general form, a functional causal model consists of a set of equations of the form

$$x_i = f_i(pa_i, u_i), \quad i = 1, \ldots, n, \tag{1.40}$$

where pa_i (connoting *parents*) stands for the set of variables that directly determine the value of X_i and where the U_i represent errors (or "disturbances") due to omitted factors. Equation (1.40) is a nonlinear, nonparametric generalization of the linear structural equation models (SEMs)

$$x_i = \sum_{k \neq 1} \alpha_{ik} x_k + u_i, \quad i = 1, \ldots, n, \tag{1.41}$$

which have become a standard tool in economics and social science (see Chapter 5 for a detailed exposition of this enterprise). In linear models, pa_i corresponds to those variables on the r.h.s. of (1.41) that have nonzero coefficients.

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 represents an autonomous mechanism is called a *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.~~

To illustrate, Figure 1.5 depicts a canonical econometric model relating price and demand through the equations

$$q = b_1 p + d_1 i + u_1, \tag{1.42}$$

$$p = b_2 q + d_2 w + u_2, \tag{1.43}$$

where Q is the quantity of household demand for a product A, P is the unit price of product A, I is household income, W is the wage rate for producing product A, and U_1 and

cannot be ignored when the meaning of the concept is in question. Indeed, compliance with human intuition has been the ultimate criterion of adequacy in every philosophical study of causation, and the proper incorporation of background information into statistical studies likewise relies on accurate interpretation of causal judgment.

[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).

the formers change meaning under solution-preserving algebraic operation, such as moving terms from one side of an equation to the other.

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

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.

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 all mediating factors constant at whatever value they would have obtained under $do(x)$. This hypothetical change, which Robins and Greenland (1991) called "pure" and Pearl (2001c) called "natural," is precisely what lawmakers 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 equation 3.51, Pearl (2001c) gave the following definition for the "natural direct effect":

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

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 natural 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 natural 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 (2001c) has nevertheless shown that, if certain assumptions of "no confounding" are deemed valid,[9] the natural direct effect can be reduced to

$$DE_{x,x'}(Y) = \sum_z [E(Y \mid do(x', z)) - E(Y \mid do(x, z))] P(z \mid do(x)). \tag{4.12}$$

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

$DE_{x,x'}$ is

In particular, ~~expression (4.12) is both valid and~~ identifiable in Markovian models, through:

~~where all *do*-operators can be eliminated using Corollary 3.2.6, for example,~~

$$DE_{x,x'}(Y) = \sum_{zw} [E(Y \mid x', z, w) - E(Y \mid x, z, w)] P(z \mid x, w) P(w)$$

~~$$P(z \mid do(x)) = \sum_t P(z \mid x, pa_X = t) P(pa_X = t)$$~~ (4.13)

where W satisfies the back-door criterion relative to both $X \to Z$ and $(X,Z) \to Y$. (See Pearl (2001c) and Shpitser and vanderWeele (2011).)

[9] One sufficient condition is that $Z(x) \perp\!\!\!\perp Y(x', z) \mid W$ holds for some set W of measured covariates. See details and graphical criteria in Pearl (2001c, 2005a) and in Petersen et al. (2006).

This model is as compact as (5.7)–(5.9) and is covariance equivalent to M with respect to the observed variables X, Y, Z. Upon setting $\alpha' = \alpha$, $\beta' = \beta$, and $\delta = \gamma$, model M' will yield the same probabilistic predictions as those of the model of (5.7)–(5.9). Still, when viewed as data-generating mechanisms, the two models are not equivalent. Each tells a different story about the processes generating X, Y, and Z, so naturally their predictions differ concerning the changes that would result from subjecting these processes to external interventions.

5.3.3 Causal Effects: The Interventional Interpretation of Structural Equation Models

The differences between models M and M' illustrate precisely where the structural reading of simultaneous equation models comes into play, and why even causally shy researchers consider structural parameters more "meaningful" than covariances and other statistical parameters. Model M', defined by (5.12)–(5.14), regards X as a direct participant in the process that determines the value of Y, whereas model M, defined by (5.7)–(5.9), views X as an indirect factor whose effect on Y is mediated by Z. This difference is not manifested in the data itself but rather in the way the data would change in response to outside interventions. For example, suppose we wish to predict the expectation of Y after we intervene and fix the value of X to some constant x; this is denoted $E(Y \mid do(X = x))$. After $X = x$ is substituted into (5.13) and (5.14), model M' yields

$$E[Y \mid do(X = x)] = E[\beta'\alpha' x + \beta'\varepsilon_2 + \delta x + \varepsilon_3] \tag{5.15}$$

$$= (\beta'\alpha' + \delta)x; \tag{5.16}$$

model M yields

$$E[Y \mid do(X = x)] = E[\beta\alpha x + \beta\varepsilon_2 + \gamma u + \varepsilon_3] \tag{5.17}$$

$$= \beta\alpha x. \tag{5.18}$$

Upon setting $\alpha' = \alpha$, $\beta' = \beta$, and $\delta = \gamma$ (as required for covariance equivalence; see (5.10) and (5.11)), we see clearly that the two models assign different magnitudes to the (total) causal effect of X on Y: model M predicts that a unit change in x will change $E(Y)$ by the amount $\beta\alpha$, whereas model M' puts this amount at $\beta\alpha + \gamma$.

At this point, it is tempting to ask whether we should substitute $x - \varepsilon_1$ for u in (5.9) prior to taking expectations in (5.17). If we permit the substitution of (5.8) into (5.9), as we did in deriving (5.17), why not permit the substitution of (5.7) into (5.9) as well? After all (the argument runs), there is no harm in upholding a mathematical equality, $u = x - \varepsilon_1$, that the modeler deems valid. This argument is fallacious, however.[15] Structural equations are not meant to be treated as immutable mathematical equalities. Rather, they are meant to define a state of equilibrium – one that is *violated* when the equilibrium is perturbed by outside interventions. In fact, the power of structural equation models is

δ (lowercase delta)

[15] Such arguments have led to Newcomb's paradox in the so-called evidential decision theory (see Section 4.1.1).

that they encode not only the initial equilibrium state but also the information necessary for determining which equations must be violated in order to account for a new state of equilibrium. For example, if the intervention consists merely of holding X constant at x, then the equation $x = u + \varepsilon_1$, which represents the preintervention process determining X, should be overruled and replaced with the equation $X = x$. The solution to the new set of equations then represents the new equilibrium. Thus, the essential characteristic of structural equations that sets them apart from ordinary mathematical equations is that the former stand not for one but for many sets of equations, each corresponding to a subset of equations taken from the original model. Every such subset represents some hypothetical physical reality that would prevail under a given intervention.

If we take the stand that the value of structural equations lies not in summarizing distribution functions but in encoding causal information for predicting the effects of policies (Haavelmo 1943; Marschak 1950; Simon 1953), it is natural to view such predictions as the proper generalization of structural coefficients. For example, the proper generalization of the coefficient β in the linear model M would be the answer to the control query, "What would be the change in the expected value of Y if we were to intervene and change the value of Z from z to $z + 1$?", which is different, of course, from the observational query, "What would be the difference in the expected value of Y if we were to *find* Z at level $z + 1$ instead of level z?" Observational queries, as we discussed in Chapter 1, can be answered directly from the joint distribution $P(x, y, z)$, while control queries require causal information as well. Structural equations encode this causal information in their syntax by treating the variable on the left-hand side of the equality sign as the effect and treating those on the right as causes. In Chapter 3 we distinguished between the two types of queries through the symbol $do(\cdot)$. For example, we wrote

$$E(Y \mid do(x)) \triangleq E\,[Y \mid do(X = x)] \tag{5.19}$$

for the controlled expectation and

$$E(Y \mid x) \triangleq E(Y \mid X = x) \tag{5.20}$$

for the standard conditional or observational expectation. That $E(Y \mid do(x))$ does not equal $E(Y \mid x)$ can easily be seen in the model of (5.7)–(5.9), where $E(Y \mid do(x)) = \alpha\beta x$ but $E(Y \mid x) = r_{YX}x = (\alpha\beta + \gamma)\,x$. Indeed, the passive observation $X = x$ should not violate any of the equations, and this is the justification for substituting both (5.7) and (5.8) into (5.9) before taking the expectation.

(lowercase gamma)

In linear models, the answers to questions of direct control are encoded in the path (or structural) coefficients, which can be used to derive the total effect of any variable on another. For example, the value of $E(Y \mid do(x))$ in the model defined by (5.7)–(5.9) is $\alpha\beta x$, that is, x times the product of the path coefficients along the path $X \rightarrow Z \rightarrow Y$. Computation of $E(Y \mid do(x))$ would be more complicated in the nonparametric case, even if we knew the functions f_1, f_2, and f_3. Nevertheless, this computation is well defined; it requires the solution (for the expectation of Y) of a modified set of equations in which f_1 is "wiped out" and X is replaced by the constant x:

$$z = f_2(x, \varepsilon_2), \tag{5.21}$$

$$y = f_3(z, u, \varepsilon_3). \tag{5.22}$$

any sentence of the form $P(A \mid B) < p$, where A and B are Boolean expressions representing events. A *causal model*, naturally, should encode the truth values of sentences that deal with causal relationships; these include action sentences (e.g., "A will be true if we do B"), counterfactuals (e.g., "A would have been different were it not for B"), and plain causal utterances (e.g., "A may cause B" or "B occurred because of A"). Such sentences cannot be interpreted in standard propositional logic or probability calculus because they deal with changes that occur in the external world rather than with changes in our beliefs about a static world. Causal models encode and distinguish information about external changes through an explicit representation of the mechanisms that are altered in such changes.

structural

Definition 7.1.1 (Causal Model)
A causal model *is a triple*

$$M = \langle U, V, F \rangle,$$

where:

(i) *U is a set of* background *variables, (also called* exogenous),[2] *that are determined by factors outside the model;*

(ii) *V is a set $\{V_1, V_2, \ldots, V_n\}$ of variables, called* endogenous, *that are determined by variables in the model – that is, variables in $U \cup V$; and*

(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 $PA_i \subseteq V \backslash V_i$ and the entire set F forms a mapping from U to V. In other words, each f_i in*

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

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]

Every causal model M can be associated with a directed graph, $G(M)$, in which each node corresponds to a variable and the directed edges point from members of PA_i and U_i toward V_i. We call such a graph the *causal diagram* associated with M. This graph merely identifies the endogenous and background variables that have direct influence on each V_i; it does not specify the functional form of f_i. The convention of confining the parent set PA_i to variables in V stems from the fact that the background variables are often unobservable. In general, however, we can extend the parent sets to include observed variables in U.

---

[2] We will try to refrain from using the term "exogenous" in referring to background conditions, because this term has acquired more refined technical connotations (see Sections 5.4.3 and 7.4). The term "predetermined" is used in the econometric literature.

[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 nonrecursive systems.

Task 3

Compute $P(Y_x = y)$ (i.e., the causal effect of smoking on cancer).

For any variable Z, by composition we have

$$Y_x(u) = Y_{xz}(u) \quad \text{if} \quad Z_x(u) = z.$$

Since $Y_{xz}(u) = Y_z(u)$ (from (7.29)),

lowercase x Y_x

$$Y_x(u) = Y_{xz_x}(u) = Y_z(u), \quad \text{where } z_x = Z_x(u). \tag{7.35}$$

Thus,

$$\begin{aligned} P(Y_x = y) &= P(Y_{zx} = y) && \text{from (7.35)} \\ &= \sum_z P(Y_{zx} = y \mid Z_x = z)\, P(Z_x = z) \\ &= \sum_z P(Y_z = y \mid Z_x = z)\, P(Z_x = z) && \text{by composition} \\ &= \sum_z P(Y_z = y)\, P(Z_x = z). && \text{from (7.30)} \end{aligned} \tag{7.36}$$

The probabilities $P(Y_z = y)$ and $P(Z_x = z)$ were computed in (7.34) and (7.31), respectively. Substituting gives us

$$P(Y_x = y) = \sum_z P(z \mid x) \sum_{x'} P(y \mid z, x')\ P(x'). \tag{7.37}$$

The right-hand side of (7.37) can be computed from $P(x, y, z)$ and coincides with the front-door formula derived in Section 3.4.3 (equation (3.42)).

Thus, $P(Y_x = y)$ can be reduced to expressions involving probabilities of observed variables and is therefore identifiable. More generally, our completeness result (Theorem 7.3.5) implies that *any* identifiable counterfactual quantity can be reduced to the correct expression by repeated application of composition and effectiveness (assuming recursiveness).

7.3.3 Axioms of Causal Relevance

In Section 1.2 we presented a set of axioms for a class of relations called *graphoids* (Pearl and Paz 1987; Geiger et al. 1990) that characterize informational relevance.[16] We now develop a parallel set of axioms for *causal relevance*, that is, the tendency of certain events to affect the occurrence of other events in the physical world, independent of the observer–reasoner. Informational relevance is concerned with questions of the form: "Given that we know Z, would gaining information about X give us new information

[16] "Relevance" will be used primarily as a generic name for the relationship of being relevant or irrelevant. It will be clear from the context when "relevance" is intended to negate "irrelevance."

seen that the meaning of the error term u_Y in the equation $Y = f_Y(pa_Y, u_Y)$ is captured by the counterfactual variable Y_{pa_Y}. In other words, the variable U_Y can be interpreted as a modifier of the functional mapping from PA_Y to Y. The statistics of such modifications is observable when pa_Y is held fixed. This translation into counterfactual notation may facilitate algebraic manipulations of U_Y without committing to the functional form of f_Y. However, from the viewpoint of model specification, the error terms should still be viewed as (summaries of) omitted factors.

Armed with this interpretation, we can obtain graphical and counterfactual definitions of causal concepts that were originally given error-based definitions. Examples of such concepts are causal influence, exogeneity, and instrumental variables (Section 5.4.3). In clarifying the relationships among error-based, counterfactual, and graphical definitions of these concepts, we should first note that these three modes of description can be organized in a simple hierarchy. Since graph separation implies independence, but independence does not imply graph separation (Theorem 1.2.4), definitions based on graph separation should imply those based on error-term independence. Likewise, since for any two variables X and Y the independence relation $U_X \perp\!\!\!\perp U_y$ implies the counterfactual independence $X_{pa_X} \perp\!\!\!\perp Y_{pa_Y}$ (but not the other way around), it follows that definitions based on error independence should imply those based on counterfactual independence. Overall, we have the following hierarchy:

$$\text{graphical criteria} \implies \text{error-based criteria} \implies \text{counterfactual criteria.}$$

The concept of exogeneity may serve to illustrate this hierarchy. The pragmatic definition of exogeneity is best formulated in counterfactual or interventional terms as follows.

Exogeneity (Counterfactual Criterion)
A variable X is exogenous relative to Y if and only if the effect of X on Y is identical to the conditional probability of Y given X – that is, if

$$P(Y_x = y) = P(y \mid x) \tag{7.45}$$

or, equivalently,

$$P(Y = y \mid do\,(x)) = P(y \mid x); \tag{7.46}$$

this in turn is ~~equivalent to~~ implied by the independence condition $Y_x \perp\!\!\!\perp X$, named "weak ignorability" in Rosenbaum and Rubin (1983).[26]

This definition is pragmatic in that it highlights the reasons economists should be concerned with exogeneity by explicating the policy-analytic benefits of discovering that a variable is exogenous. However, this definition fails to guide an investigator toward

[26] We focus the discussion in this section on the causal component of exogeneity, which the econometric literature has unfortunately renamed "superexogeneity" (see Section 5.4.3). Epidemiologists refer to (7.46) as "no-confounding" (see (6.10)). We also postpone discussion of "strong ignorability," defined as the joint independence $\{Y_x, Y_{x'}\} \perp\!\!\!\perp X$, to Chapter 9 (Definition 9.2.3).

a mean difference (using $P(z_1) = 0.50$) of

$$P(y_1 \mid x_1) - p(y_1 \mid x_0) = \frac{0.473}{0.473 + 0.139} - \frac{0.073 + 0.081}{1 + 0.315 + 0.073} = 0.662,$$

and an encouragement effect (intent to treat) of

$$P(y_1 \mid z_1) - P(y_1 \mid z_0) = 0.073 + 0.473 - 0.081 = 0.465.$$

According to (8.17), ACE($X \rightarrow Y$) can be bounded by

$$\text{ETT}(X \rightarrow Y) \geq 0.465 - 0.073 - 0.000 = 0.392,$$
$$\text{ETT}(X \rightarrow Y) \leq 0.465 + 0.315 + 0.000 = 0.780.$$

These are remarkably informative bounds: although 38.8% of the subjects deviated from their treatment protocol, the experimenter can categorically state that, when applied uniformly to the population, the treatment is guaranteed to increase by at least 39.2% the probability of reducing the level of cholesterol by 28 points or more.

The impact of treatment "on the treated" is equally revealing. Using equation (8.20), ETT($X \rightarrow Y$) can be evaluated precisely (since $P(x_1 \mid z_0) = 0$):

$$\text{ETT}(X \rightarrow Y) = \frac{0.465}{0.610} = 0.762.$$

In words, those subjects who stayed in the program are much better off than they would have been if not treated: the treatment can be credited with reducing cholesterol levels by at least 28 units in 76.2% of these subjects.

8.3 COUNTERFACTUALS AND LEGAL RESPONSIBILITY

Evaluation of counterfactual probabilities could be enlightening in some legal cases in which a plaintiff claims that a defendant's actions were responsible for the plaintiff's misfortune. Improper rulings can easily be issued without an adequate treatment of counterfactuals (Robins and Greenland 1989). Consider the following hypothetical and fictitious case study, specially crafted in Balke and Pearl (1994a) to accentuate the disparity between causal effects and causal attribution.

The marketer of PeptAid (antacid medication) randomly mailed out product samples to 10% of the households in the city of Stress, California. In a follow-up study, researchers determined for each individual whether they received the PeptAid sample, whether they consumed PeptAid, and whether they developed peptic ulcers in the following month.

The causal structure for this scenario is identical to the partial compliance model given by Figure 8.1, where z_1 asserts that PeptAid was received from the marketer, x_1 asserts that PeptAid was consumed, and y_1 asserts that peptic ulceration occurred. The data showed the following distribution:

Table 9.3. *PN as a Function of Assumptions and Available Data*

Assumptions			Data Available		
Exogeneity	Monotonicity	Additional	Experimental	Observational	Combined
+	+		ERR	ERR	ERR
+	−		bounds	bounds	bounds
−	+	covariate control	—	corrected ERR	corrected ERR
−	+		—	—	corrected ERR
−	−		—	—	bounds

Note: ERR stands for the excess risk ratio, $1 - P(y \mid x')/P(y \mid x)$; corrected ERR is given in (9.31).

can be ascertained: exogeneity (i.e., no confounding) and monotonicity (i.e., no prevention). When monotonicity does not hold, ERR provides merely a lower bound for PN, as shown in (9.13). (The upper bound is usually unity.) The nonentries (—) in the right-hand side of Table 9.3 represent vacuous bounds (i.e., $0 \leq \text{PN} \leq 1$). In the presence of confounding, ERR must be corrected by the additive term $[P(y \mid x') - P(y_{x'})]/P(x, y)$, as stated in (9.31). In other words, when confounding bias (of the causal effect) is positive, PN is higher than ERR by the amount of this additive term. Clearly, owing to the division by $P(x, y)$, the PN bias can be many times higher than the causal effect bias $P(y \mid x') - P(y_{x'})$. However, confounding results only from association between exposure and other factors that affect the outcome; one need not be concerned with associations between such factors and susceptibility to exposure (see Figure 9.2).

The last row in Table 9.3, corresponding to no assumptions whatsoever, leads to vacuous bounds for PN, unless we have combined data. This does not mean, however, that justifiable assumptions *other* than monotonicity and exogeneity could not be helpful in rendering PN identifiable. The use of such assumptions is explored in the next section.

9.4 IDENTIFICATION IN NONMONOTONIC MODELS

In this section we discuss the identification of probabilities of causation without making the assumption of monotonicity. We will assume that we are given a causal model M in which all functional relationships are known, but since the background variables U are not observed, their distribution is not known and the model specification is not complete.

Our first step would be to study under what conditions the function $P(u)$ can be identified, thus rendering the entire model identifiable. If M is Markovian, then the problem can be analyzed by considering each parents–child family separately. Consider any arbitrary equation in M,

$$\begin{aligned} y &= f(pa_Y, u_Y) \\ &= f(x_1, x_2, \ldots, x_k, u_1, \ldots, u_m), \end{aligned} \tag{9.55}$$

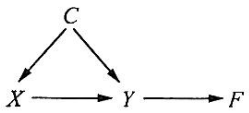

Figure 11.6 Graph applicable for accessing the effect of X on Y.

because the graph applicable for this task is given in Figure 11.6; F becomes a descendant of X, and is excluded by the back-door criterion.

2. If the explanation of confounding and sufficiency sounds at variance with traditional epidemiology, it is only because traditional epidemiologists did not have proper means of expressing the operations of blocking or creating dependencies. They might have had a healthy intuition about dependencies, but graphs translate this intuition into a formal system of closing and opening paths.

   We should also note that before 1985, causal analysis in epidemiology was in a state of confusion, because the healthy intuitions of leading epidemiologists had to be expressed in the language of associations – an impossible task. Even the idea that confounding stands for "bias," namely, a "difference between two dependencies, one that we wish to measure, the other that we do measure," was resisted by many (see Chapter 6), because they could not express the former mathematically.[3]

   Therefore, instead of finding "something in graph language that is closer to traditional meaning," we can do better: explicate what that "traditional meaning" ought to have been.

   In other words, traditional meaning was informal and occasionally misguided, while graphical criteria are formal and mathematically proven.

   Chapter 6 (pp. 183, 194) records a long history of epidemiological intuitions, some by prominent epidemiologists, that have gone astray when confronted with questions of confounding and adjustment (see Greenland and Robins 1986; Wickramaratne and Holford 1987; Weinberg 1993). Although most leading epidemiologists today are keenly attuned to modern developments in causal analysis, (e.g., Glymour and Greenland 2008), epidemiological folklore is still permeated with traditional intuitions that are highly suspect. (See Section 6.5.2.)

   In summary, graphical criteria, as well as principles 1–3 above, give us a sensible, friendly, and unambiguous interpretation of the "traditional meaning of epidemiological concepts."

11.3.2 Demystifying "Strong Ignorability"

Researchers working within the confines of the potential-outcome language express the condition of "zero bias" or "no-confounding" using an independence relationship called

[3] Recall that Greenland and Robins (1986) were a lone beacon of truth for many years, and even they had to resort to the "black-box" language of "exchangeability" to define "bias," which discouraged intuitive interpretations of confounding (see Section 6.5.3). Indeed, it took epidemiologists another six years (Weinberg 1993) to discover that adjusting for factors affected by the exposure (~~a~~ in Figure 11.5) would introduce bias.

mapping variables, 204