

## 2.8 Nontemporal Causation and Statistical Time

Determining the direction of causal influences from nontemporal data raises some interesting philosophical questions about the relationships between time and causal explanations. For example, can the orientation assigned to the arrow  $X \rightarrow Y$  in Definitions 2.7.2 or 2.7.4 ever clash with the available temporal information (say, by a subsequent discovery that  $Y$  precedes  $X$ )? Since the rationale behind Definition 2.7.4 is based on strong intuitions about the statistical aspects of causal relationships (e.g., no correlation without some causation), it is apparent that such clashes, if they occur, are rather rare. The question then arises: Why should orientations determined solely by statistical dependencies have anything to do with the flow of time?

In human discourse, causal explanations satisfy two expectations, temporal and statistical. The temporal aspect is represented by the understanding that a cause should precede its effect. The statistical aspect expects a complete causal explanation to screen off its various effects (i.e., render the effects conditionally independent);<sup>11</sup> explanations that do not screen off their effects are considered “incomplete,” and the residual dependencies are considered “spurious” or “unexplained.” The clashless coexistence of these two expectations through centuries of scientific observations implies that the statistics of natural phenomena must exhibit some basic temporal bias. Indeed, we often encounter phenomenon where knowledge of a present state renders the variables of the future state conditionally independent (e.g., multivariate economic time series as in (2.3)). However, we rarely find the converse phenomenon, where knowledge of the present state would render the components of the past state conditionally independent. Is there any compelling reason for this temporal bias?

A convenient way to formulate this bias is through the notion of

---

<sup>11</sup>This expectation, known as Reichenbach’s “conjunctive fork” or “common-cause” criterion (Reichenbach 1956; Suppes and Zaniotti 1981; Sober and Barrett 1992) has been criticized by Salmon (1984), who showed that some events qualify as causal explanations though they fail to meet Reichenbach’s criterion. However, Salmon’s examples involve incomplete explanations, as they leave out variables that mediate between the cause and its various effects (see Section 2.9.1).

statistical time.

**Definition 2.8.1 (Statistical Time)**

*Given an empirical distribution  $P$ , a statistical time of  $P$  is any ordering of the variables that agrees with at least one minimal causal structure consistent with  $P$ .*

We see, for example, that a scalar Markov chain process has many statistical times; one coinciding with the physical time, one opposite to it, and others that correspond to orderings that agree with any orientation of the Markov chain away from one of the nodes (arbitrarily chosen as a root). On the other hand, a process governed by two coupled Markov chains, such as

$$\begin{aligned} X_t &= \alpha X_{t-1} + \beta Y_{t-1} + \xi_t, \\ Y_t &= \gamma X_{t-1} + \delta Y_{t-1} + \eta_t, \end{aligned} \tag{2.3}$$

has only one statistical time—the one coinciding with the physical time.<sup>12</sup> Indeed, running the IC algorithm on samples taken from such a process—while suppressing all temporal information—quickly identifies the components of  $X_{t-1}$  and  $Y_{t-1}$  as genuine causes of  $X_t$  and  $Y_t$ . This can be seen from Definition 2.7.1 (where  $X_{t-2}$  qualifies as a potential cause of  $X_{t-1}$  using  $Z = Y_{t-2}$  and  $S = \{X_{t-3}, Y_{t-3}\}$ ) and Definition 2.7.2 (where  $X_{t-1}$  qualifies as a genuine cause of  $X_t$  using  $Z = X_{t-2}$  and  $S = \{Y_{t-1}\}$ ).

The temporal bias postulated earlier can be expressed as follows.

**Conjecture 2.8.2 (Temporal Bias)**

*In most natural phenomenon, the physical time coincides with at least one statistical time.*

Reichenbach (1956) attributed the asymmetry associated with his conjunctive fork to the second law of thermodynamics. It is doubtful that the second law can provide a full account of the temporal bias just

---

<sup>12</sup>Here  $\xi_t$  and  $\eta_t$  are assumed to be two independent, white-noise time series. Also,  $\alpha \neq \delta$  and  $\gamma \neq \beta$ .

described, since the influence of the external noise  $\xi_t$  and  $\eta_t$  renders the process in (2.3) nonconservative.<sup>13</sup> Moreover, the temporal bias is language-dependent. For example, expressing (2.3) in a different coordinate system—say, using a linear transformation

$$\begin{aligned} X'_t &= aX_t + bY_t, \\ Y'_t &= cX_t + dY_t \end{aligned}$$

—it is possible to make the statistical time in the  $(X', Y')$  representation run contrary to the physical time; that is,  $X'_t$  and  $Y'_t$  will be independent of each other conditional on their future values ( $X'_{t+1}$  and  $Y'_{t+1}$ ) rather than their past values. This suggests that the consistent agreement between physical and statistical times is a byproduct of the human choice of linguistic primitives and not a feature of physical reality. For example, if  $X_t$  and  $Y_t$  stand for the positions of two interacting particles at time  $t$ , with  $X'_t$  the position of their center of gravity and  $Y'_t$  their relative distance, then describing the particles' motion in the  $(X, Y)$  versus  $(X', Y')$  coordinate system is (in principle) a matter of choice. Evidently, however, this choice is not entirely whimsical; it reflects a preference toward coordinate systems in which the *forward* disturbances ( $\xi_t$  and  $\eta_t$  in (2.3)) are orthogonal to each other, rather than the corresponding backward disturbances ( $\xi'_t$  and  $\eta'_t$ ). Pearl and Verma (1991) speculated that this preference represents survival pressure to facilitate predictions of future events, and that evolution has evidently ranked this facility more urgent than that of finding hindsight explanations for current events. Whether this or some other force has shaped our choice of language remains to be investigated (see discussions in Price 1996), which makes the statistical-temporal agreement that much more interesting.

---

<sup>13</sup>I am grateful to Seth Lloyd for this observation.