## 2.4 Stable Distributions

Although the minimality principle is sufficient for forming a normative theory of inferred causation, it does not guarantee that the structure of the actual data-generating model would be minimal, or that the search through the vast space of minimal structures would be computationally practical. Some structures may admit peculiar parameterizations that would render them indistinguishable from many other minimal models that have totally disparate structures. For example, consider a binary variable $C$ that takes the value 1 whenever the outcomes of two fair coins ($A$ and $B$) are the same and takes the value 0 otherwise. In the trivariate distribution generated by this parameterization, each pair of variables is marginally independent yet is dependent conditional on the third variable. Such a dependence pattern may in fact be generated by three minimal causal structures, each depicting one of the variables as causally dependent on the other two, but there is no way to decide among the three. In order to rule out such "pathological" parameterizations, we impose a restriction on the distribution called *stability*, also known as DAG-isomorphism (Pearl 1998b, p. 128) and faithfulness Spirtes et al. 1993). This restriction conveys the assumption that all the independencies embedded in $P$ are stable; that is, they are entailed by the structure of the model $D$ and hence remain invariant to any change in the parameters $\Theta_D$. In our example, only the correct structure (namely, $A \rightarrow C \leftarrow B$) will retain its independence pattern in the face of changing parameterizations—say, when one of the coins becomes slightly biased.

**Definition 2.4.1 (Stability)**
*Let $I(P)$ denote the set of all conditional independence relationships embodied in $P$. A causal model $M = \ <D, \Theta_D>$ generates a* stable *distribution if and only if $P(<D, \Theta_D>)$ contains no extraneous independences—that is, if and only if $I(P(<D, \Theta_D>)) \subseteq I(P(<D, \Theta'_D>))$ for any set of parameters $\Theta'_D$.*

The stability condition states that, as we vary the parameters from $\Theta$ to $\Theta'$, no independence in $P$ can be destroyed; hence the name "stability." Succinctly, $P$ is a stable distribution if there exists a DAG $D$ such that

$(X \perp\!\!\!\perp Y|Z)_P \Leftrightarrow (X \perp\!\!\!\perp Y|Z)_D$ for any three sets of variables $X, Y$, and $Z$ (see Theorem 1.2.5).

The relationship between minimality and stability can be illustrated using the following analogy. Suppose we see a picture of a chair and that we need to decide between two theories as follows.

$T_1$: The object in the picture is a chair.

$T_2$: The object in the picture is either a chair or two chairs positioned such that one hides the other.

Our preference for $T_1$ over $T_2$ can be justified on two principles, one based on minimality and the other on stability. The minimality principle argues that $T_1$ is preferred to $T_2$ because the set of scenes composed of single objects is a proper subset of scenes composed of two or fewer objects and, unless we have evidence to the contrary, we should prefer the more specific theory. The stability principle rules out $T_2$ a priori, arguing that it would be rather unlikely for two objects to align themselves so as to have one perfectly hide the other. Such an alignment would be *unstable* relative to slight changes in environmental conditions or viewing angle.

The analogy with independencies is clear. Some independencies are *structural*, that is, they would persist for every functional-distributional parameterization of the graph. Others are sensitive to the precise numerical values of the functions and distributions. For example, in the structure $Z \leftarrow X \rightarrow Y$, which stands for the relations

$$z = f_1(x, u_1), \qquad y = f_2(x, u_2), \tag{2.1}$$

the variables $Z$ and $Y$ will be independent, conditional on $X$, for all functions $f_1$ and $f_2$. In contrast, if we add an arrow $Z \rightarrow Y$ to the structure and use a linear model

$$z = \gamma x + u_1, \qquad y = \alpha x + \beta z + u_2, \tag{2.2}$$

with $\alpha = -\beta\gamma$, then $Y$ and $X$ will be independent. However, the independence between $Y$ and $X$ is unstable because it disappears as soon as the equality $\alpha = -\beta\gamma$ is violated. The stability assumption presumes that this type of independence is unlikely to occur in the data, that all independencies are structural.

To further illustrate the relations between stability and minimality, consider the causal structure depicted in Figure 2.1(c). The minimality principle rejects this structure on the ground that it fits a broader set of distributions than those fitted by structure (a). The stability principle rejects this structure on the ground that, in order to fit the data (specifically, the independence $(a \perp\!\!\!\perp b)$), the association produced by the arrow $a \rightarrow b$ must cancel precisely the one produced by the path $a \leftarrow c \rightarrow b$. Such precise cancelation cannot be stable, for it cannot be sustained for all functions connecting variables $a$, $b$, and $c$. In structure (a), by contrast, the independence $(a \perp\!\!\!\perp b)$ is stable.