2 Causal Models

This section reviews the definition of causal models as understood in the structural modeling tradition started by Pearl [22], where I use the notation from Halpern [16].

Definition 1. *A signature $\mathcal{S}$ is a tuple $(\mathcal{U}, \mathcal{V}, \mathcal{R})$, where $\mathcal{U}$ is a set of* exogenous *variables, $\mathcal{V}$ is a set of* endogenous *variables, and $\mathcal{R}$ a function that associates with every variable $Y \in \mathcal{U} \cup \mathcal{V}$ a nonempty set $\mathcal{R}(Y)$ of possible values for Y (i.e., the set of values over which Y* ranges*). If $\vec{X} = (X_1, \dots, X_n)$, $\mathcal{R}(\vec{X})$ denotes the crossproduct $\mathcal{R}(X_1) \times \dots \times \mathcal{R}(X_n)$.*

Exogenous variables represent unobserved factors whose causal origins are outside the scope of the causal model, such as background conditions and noise. The values of the endogenous variables, on the other hand, are causally determined by other variables within the model.

Definition 2. *A* causal model *M is a pair $(\mathcal{S}, \mathcal{F})$, where $\mathcal{S}$ is a signature and $\mathcal{F}$ defines a function that associates with each endogenous variable Y a* structural equation *F_Y giving the value of Y in terms of the values of other endogenous and exogenous variables. Formally, the equation F_Y maps $\mathcal{R}(\mathcal{U} \cup \mathcal{V} - \{Y\})$ to $\mathcal{R}(Y)$, so F_Y determines the value of Y, given the values of all the other variables in $\mathcal{U} \cup \mathcal{V}$.*

We usually write the equation for an endogenous variable as $Y = f(\vec{X})$, where $\vec{X}$ are called the *parents* of Y (and Y is called a *child* of each variable in $\vec{X}$), and the function f is such that it only depends on the values of $\vec{X}$. The *ancestor* relation is the transitive closure of the parent relation. In this paper we restrict attention to *acyclic* models, that is, models where no variable is an ancestor of itself. A (directed) *path* is a sequence of variables in which each element is a child of the previous element.In this manner an acyclic causal model induces a unique *DAG*, i.e., a Directed Acyclic Graph, which is simply a graphical representation of all the ancestral relations.

An *intervention* has the form $\vec{X} \leftarrow \vec{x}$, where $\vec{X}$ is a set of endogenous variables. Intuitively, this means that the values of the variables in $\vec{X}$ are set to the values $\vec{x}$. The equations define what happens in the presence of interventions. The intervention $\vec{X} \leftarrow \vec{x}$ in a causal model $M = (\mathcal{S}, \mathcal{F})$ results in a new causal model, denoted $M_{\vec{X} \leftarrow \vec{x}}$, which is identical to M, except that $\mathcal{F}$ is replaced by $\mathcal{F}^{\vec{X} \leftarrow \vec{x}}$: for each variable $Y \notin \vec{X}$, $F_Y^{\vec{X} \leftarrow \vec{x}} = F_Y$ (i.e., the equation for Y is unchanged), while for each X' in $\vec{X}$, the equation $F_{X'}$ for X' is replaced by $X' = x'$ (where x' is the value in $\vec{x}$ corresponding to X').

Given a signature $\mathcal{S} = (\mathcal{U}, \mathcal{V}, \mathcal{R})$, an *atomic formula* is a formula of the form $X = x$, for $X \in \mathcal{V}$ and $x \in \mathcal{R}(X)$. A *causal formula (over $\mathcal{S}$)* is one of the form $[Y_1 \leftarrow y_1, \dots, Y_k \leftarrow y_k]\phi$, where

- ϕ is a Boolean combination of atomic formulas,
- $Y_1, \dots, Y_k$ are distinct variables in $\mathcal{V}$, and
- $y_i \in \mathcal{R}(Y_i)$ for each $1 \leq i \leq k$.

Such a formula is abbreviated as $[\vec{Y} \leftarrow \vec{y}]\phi$. The special case where $k = 0$ is abbreviated as ϕ. Intuitively, $[Y_1 \leftarrow y_1, \dots, Y_k \leftarrow y_k]\phi$ says that ϕ would hold if Y_i were set to y_i, for $i = 1, \dots, k$.

We call a setting $\vec{u} \in \mathcal{R}(\mathcal{U})$ of values of exogenous variables a *context*. A causal formula ψ is true or false in a *causal setting*, which is a causal model given a context. As usual, we write $(M, \vec{u}) \models \psi$ if the causal formula ψ is true in the causal setting $(M, \vec{u})$. The $\models$ relation is defined inductively. $(M, \vec{u}) \models X = x$ if the variable X has value x in the unique (since we are dealing with recursive models) solution to the equations in M in context $\vec{u}$ (i.e., the unique vector of values that simultaneously satisfies all equations in M with the variables in $\mathcal{U}$ set to $\vec{u}$). The truth of conjunctions and negations is defined in the standard way. Finally, $(M, \vec{u}) \models [\vec{Y} \leftarrow \vec{y}]\phi$ if $(M_{\vec{Y} \leftarrow \vec{y}}, \vec{u}) \models \phi$.

In addition to the causal setting $(M, \vec{u})$ that describes both the objective causal relations and their actual realization, we also need to represent the agent's beliefs regarding what could possibly happen in order to fill in the **Epistemic Condition**. I do so in the same manner as proposed by HK: we take $\Pr$ to be a probability distribution over a set of causal settings $\mathcal{K}$, so that $\Pr$ expresses the agent's subjective probabilities before the agent performs their action. As do HK, I assume for simplicity that all the causal models appearing in $\mathcal{K}$ have the same signature (i.e., the same exogenous and endogenous variables). We define an *epistemic state* of an agent to consist of a pair $\mathcal{E} = (\Pr, \mathcal{K})$, and define a *responsibility setting* $(M, \vec{u}, \mathcal{E})$ as the combination of a causal setting and an epistemic state.