The Three Layer Causal Hierarchy

Recent discussions concerning causal mediation gave me the impression that many researchers in the field are not familiar with the ramifications of the Causal Hierarchy, as articulated in Chapter 1 of *Causality* (2000, 2009). This note represents the Causal Hierarchy in table form (Fig. 1) and discusses the distinctions between its three layers: 1. Association, 2. Intervention, 3. Counterfactuals.

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y\mid x)$	Seeing	What is? How would seeing X change my belief in$Y$?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y\mid do(x), z)$	Doing	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x\mid x', y')$	Imagining, Retrospection	Why? Was it X that caused Y? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past 2 years?

Figure 1: The ladder of causation

An extremely useful insight unveiled by the logic of causal reasoning is the existence of a sharp classification of causal information, in terms of the kind of questions that each class is capable of answering. The classification forms a 3-level hierarchy in the sense that questions at level i ($i = 1, 2, 3$) can only be answered if information from level j ($j \geq i$) is available.

Figure 1 shows the 3-level hierarchy, together with the characteristic questions that can be answered at each level. The levels are titled 1. Association, 2. Intervention, and 3. Counterfactual. The names of these layers were chosen to emphasize their usage. We call the first level Association, because it invokes purely statistical relationships, defined by the naked data. For instance, observing a customer who buys toothpaste makes it more likely that he/she buys floss; such association can be inferred directly from the observed data using conditional expectation. Questions at this layer, because they require no causal information, are placed at the bottom level on the hierarchy. The second level, Intervention, ranks higher than Association because it involves not just seeing what is, but changing what we see. A typical question at this level would be: What happens if we double the price? Such questions cannot be answered from sales data alone, because they involve a change in customers behavior, in reaction to the new pricing. Customer choices under the new price structure may differ substantially from that prevailing in the past. Finally, the top level is called Counterfactuals, a term that goes back to the philosophers David Hume and John Stewart Mill, and which has been given structural semantics in the SCM framework. A typical question in the counterfactual category is “What if I were to act differently,” thus necessitating retrospective reasoning.

Counterfactuals are placed at the top of the hierarchy because they subsume interventional and associational questions. If we have a model that can answer counterfactual queries, we can also answer questions about interventions and observations. For example, the interventional question, What will happen if we double the price? can be answered by asking the counterfactual question: What would happen had the price been twice its current value? Likewise, associational questions can be answered once we can answer interventional questions; we simply ignore the action part and let observations take over. The translation does not work in the opposite direction. Interventional questions cannot be answered from purely observational information (i.e., from statistical data alone). No counterfactual question involving retrospection can be answered from purely interventional information, such as that acquired from controlled experiments; we cannot re-run an experiment on subjects who were treated with a drug and see how they behave had then not given the drug. The hierarchy is therefore directional, with the top level being the most powerful one.

Counterfactuals are the building blocks of scientific thinking as well as legal and moral reasoning. In civil court, for example, the defendant is considered to be the cause of an injury to the plaintiff if, *but for* the defendant’s action, it is more likely than not that the injury would not have occurred. The computational meaning of *but for* calls for comparing the real world to an alternative world in which the defendant action did not take place.

Each layer in the hierarchy has a syntactic signature that characterizes the the sentences admitted into that layer. For example, the association layer is characterized by conditional probability sentences, e.g., $P(y|x) = p$ stating that: the probability of event $Y = y$ given that we observed event $X = x$ is equal to p. In large systems, such evidential sentences can be computed efficiently using Bayesian Networks, or any of the graphical models that support deep-learning systems.

At the interventional layer we find sentences of the type $P(y|do(x), z)$, which denotes “The probability of event $Y = y$ given that we intervene and set the value of X to x and subsequently observe event $Z = z$. Such expressions can be estimated experimentally from randomized trials or analytically using Causal Bayesian Networks (Pearl, 2000, Chapter 1).

Finally, at the counterfactual level, we have expressions of the type $P(y_x|x', y')$ which stand for “The probability of event $Y = y$ had X been x, given that we actually observed X to be x' and and Y to be y'. Such sentences can be computed only when we possess functional or Structural Equation models, or properties of such models.

References

PEARL, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York. 2nd edition, 2009.