Interventions

1000: Introduction

In previous modules, we have been building an account of causation through ideas like response structures and the causal assignments that they include. A crucial feature of a causal assignment is that the values of variables are assigned, and not observed. The difference between assigning and observing is at the root of causation, and is the topic of this module.

For example, consider two variables that relate to traveling on an airplane: {Seat Belt Sign [lit, unlit], Turbulence [yes, no]}. It is quite often the case that the seat belt light goes on a few moments before turbulence. If we build a response structure for the effect: Turbulence, however, we must consider two causal assignments:

TABLE 1000-1: RESPONSE STRUCTURE FOR TURBULENCE

Assignment	Seat Belt Sign	Turbulence
1	Lit	??
2	Unlit	??

Assuming we take causal assignment 1 seriously, then the causal question is: if we **intervene** to light up the Seat Belt signs, will Turbulence follow? Not usually, unless we only light up the sign when we have other information, like a weather report about turbulence in the area. Normally, during flights we observe the sign light up, and then observe that turbulence follows. To claim that the Seat Belt Sign being lit is a "cause" of turbulence, we must force the sign to light up, and then see Turbulence as an effect.

In this module, we consider what an **intervention** is, and how we can act to determine causation from association. An **ideal intervention** allows investigators to break the causal relationship between an effect and its causes. In this way it is possible to manipulate the causal relations among the variables within a causal system.

The essential ideas are these:

+ An **ideal intervention** directly fixes the value of a variable within a causal system.
+ Ideally intervening to change the value of a cause does not change the relationship between the cause and its effects.
+ Ideally intervening to change the value of an effect does change the relationship between the effect and its causes.

Because ideal interventions can change the relationship between an effect and its causes, we distinguish between the causal system as it occurs **before** an intervention and the system as it occurs after the intervention. We call the pre-intervention system the "natural" or "pre-manipulated" system, and the post-intervention system the "experimental," "post-manipulation," or "manipulated" system.

When you finish this module, you should understand what an ideal intervention is and how it changes a causal system. You should be able to take 1) a causal graph that describes a "natural" or "pre-manipulated" system and 2) an intervention that we might perform on that system, and produce the "experimental," or "manipulated" causal graph that would result.

2000: The Idea of Intervention

2100: Interventions as Actions

Sometimes a variable changes from naturally occurring causes, and sometimes we **act** to change it. For example, the noise level in my study sometimes changes when my children come in, or when the neighbor begins to mow his lawn. Sometimes, however, I act to make it quiet by kicking my kids out and shutting the window.

Another example is the amount of money in my checking account, which can be categorized as low or high, depending, let's say, on whether it has under $2000 or over $2000. The level naturally varies quite a lot during the course of a month, depending on the deposits or debits that come in daily. Suppose the most important factor for me is whether my paycheck has been deposited recently (within 3 days) or not. So consider the causal system: {**CHECK DEPOSITED RECENTLY**, **CHECKING ACCOUNT LEVEL**}, which we will represent with his causal graph:

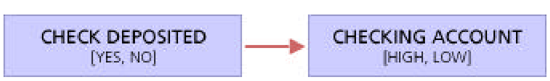

FIGURE 2100-1

There are other factors that will affect my checking account, e.g., whether I buy an airplane ticket for my whole family that month or not, but I won't include any of these in my system.

At most times during the month, I am capable of intervening from outside this system to change the level of my checking account. I can:

+ go into the bank and withdraw my entire account in cash (the total withdrawal intervention), or
+ deposit $2,500 in cash that I have stashed in my attic (the $2500 deposit intervention).

An intervention is an **action**. Deciding to withdraw all the money from my account is not an action, it is a decision. Going to the bank and filling out the appropriate paperwork for a total withdrawal of my account is an intervention. In addition, an intervention is not just an observation. Looking at my account balance and seeing that it is low is an observation. Actually withdrawing all my money is an intervention.

An intervention does not necessarily change the value of the variable. For example, if I already have over $2000 in my checking account, depositing $2500 in cash will not change the value of the variable **Checking Account** -- the value will still be High. The act is still an ideal intervention, because it **determines** the value of the variable, regardless of its value before the intervention.

In practical cases we usually restrict our attention to interventions that are actions actually under our direct control. For example, "getting my friend to stop smoking" is not a practical intervention per se, because it doesn't clearly specify an action under your control. "Lecturing my friend for half an hour on the perils of smoking" is an action under your control, as is "taking my friend to the cancer ward." "Reducing crime" is not an action under the government's actual control, it is an outcome that candidates hope you believe they can indirectly control. "Increasing the FBI's budget by $100 million" is an action that the government can actually perform, as is "increasing the budget for crime prevention programs."

Even if a hypothetical intervention is not practically possible, we can still often sensibly reason about it. For example, it is not practically possible to give each resident of Iowa a million dollars, but it is perfectly clear what it is we are considering doing, and we can reason about the economic consequences in Iowa were we to actually execute this interventionIn contrast, some interventions are too vague to even consider hypothetically. For example: "making all relationships loving," or "creating equal opportunity for all."

< A link to exercises in the interactive version of this module. >

2200: Ideal Interventions

In this section, we want to discuss the simplest sorts of interventions, which we will call **ideal interventions**.

Definition: Ideal Intervention

An **ideal intervention** is an action which i) directly targets exactly one variable, and ii) is completely effective in determining the value of that variable

.
For example, the total withdrawal intervention only directly targets the level of my checking account and does not influence whether my paycheck is deposited recently or not. It is also completely effective in determining that the level of my checking account is "low." Even if my checking account was already low, this act would still count as an ideal intervention, because it **determines** the level of my checking account.

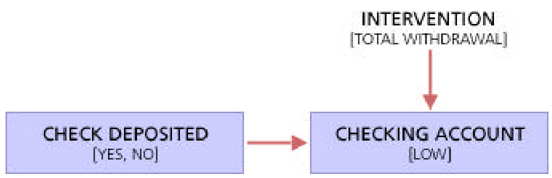

FIGURE 2200-1

Contrast this with an intervention that withdraws $1,000 from my account (the $1,000 intervention). Although this does directly target only the checking account level, it is not completely effective in determining whether my checking account is high (above $2000), or low (below $2,000). It will certainly change the chances that my account is high or low, but not determine it.

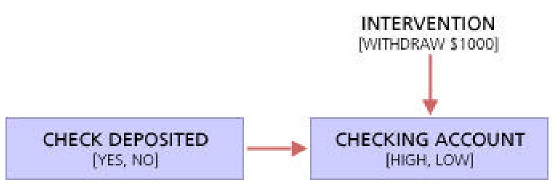

FIGURE 2200-2

< A link to exercises in the interactive version of this module. >

Other interventions might intend to target only one variable, but actually directly effect others in the system. Consider the following causal system involving cholesterol, the condition or your arteries, and the condition of your heart:

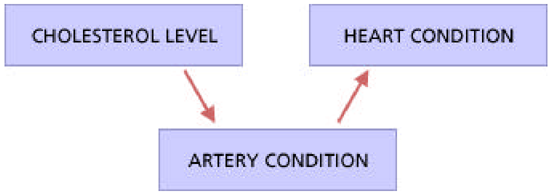

FIGURE 2200-3

Suppose your doctor decides to intervene to reduce your cholesterol level, intending to improve your heart condition indirectly. Suppose she prescribes a drug that reduces your body's production of cholesterol, has no direct effect on the condition of your arteries, but that independently induces a chronic arrhythmia which seriously damages your heart.

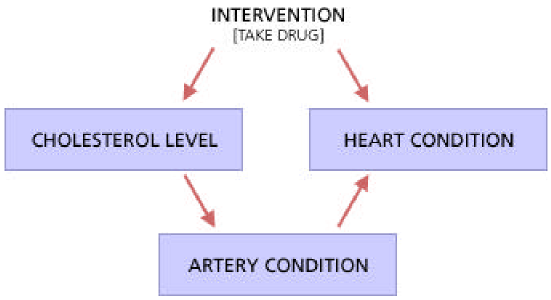

FIGURE 2200-4

Although the intervention was intended to target only cholesterol, we say that it had a "fat hand" because it inadvertently also directly affected the condition of your heart.

If, on the other hand, the drug does have the direct desired effect on your cholesterol, then your heart condition would be improved. It would seem then, that the intervention, even if it worked the way we wanted, would still have an effect on more than one variable. Wouldn't it still be a "fat hand"? The answer is no, because in this second case the drug only indirectly affects your heart condition. An ideal intervention must not affect more than one variable directly, otherwise it's a "fat hand."

Not all interventions are ideal, and the theory of interventions and causal graphs in no way depends on the interventions being ideal. For this module and those that follow, however, we will assume that the interventions we consider are ideal. That is, we will assume that interventions are actions that directly and effectively change the value of one and only one variable in a causal system.

< A link to exercises in the interactive version of this module. >

3000: Interventions and Causal Relationships

3100: Values and Relationships

Interventions directly change the value of a variable in a causal system, but they can also change the causal **relationships** themselves. In general, interventions that change the value of a cause leave intact the relationship between the cause and its effects, but interventions that change the value of an effect change the relationship between the effect and its causes. In this section we explore how this happens and its consequences for causal reasoning.

3200: Intervening on Causes

Example 1

Consider again the simple causal system involving my checking account:

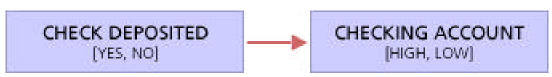

FIGURE 3200-1

In this system, there is a causal relationship posited to exist between my paycheck being deposited and the state of my checking account. Suppose we intervene to prevent my paycheck from being deposited, which we model with a new variable coming in from the outside:

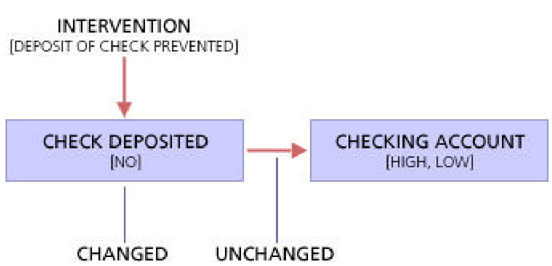

FIGURE 3200-2

The intervention changes the value of the paycheck variable, but it does not affect the relationship between the Paycheck variable and the Checking Account variable. When my paycheck is not deposited recently, that will still cause my checking account to be low.

In general, if there is no cycle between a cause and its effect, then intervening to change the value of a cause does not change the relationship between the cause and its effects.

Example 2

Consider again the simple causal system involving cholesterol and your heart:

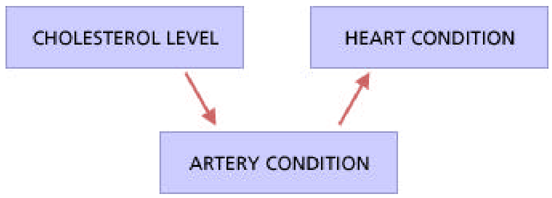

FIGURE 3200-3

In this system, there is a causal relationship posited to exist between cholesterol and artery condition.

< A link to exercises in the interactive version of this module. >

3300: Intervening on Effects

Example 1

Ideal interventions on an effect do change the relationship between the effect and its causes. Consider again the simple causal system involving my checking account:

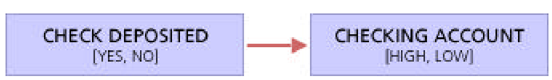

FIGURE 3300-1

Whereas before we intervened on my paycheck, which is the cause, now let's intervene on the effect (the checking account) and see what happens. Consider the $2500 deposit intervention (I deposit $2,500 in cash). Again, we model the intervention with a new variable coming in from the outside:

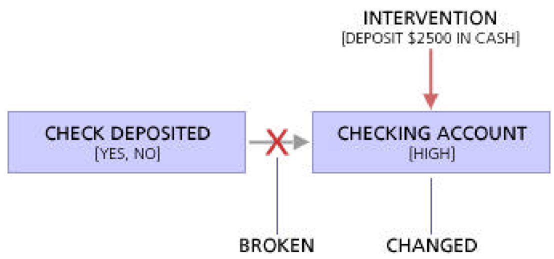

FIGURE 3300-2

The intervention changes the value of the Checking Account variable, but it also severs the relationship between the Paycheck variable and the Checking Account variable. The Paycheck deposit no longer has any influence on whether my checking account is high. My intervention took over all influence, and in so doing changed the relationship between the effect and its causes.

In general, if there is no cycle between a cause and its effect, then ideally intervening to change the value of the effect **breaks the relationship between the effect and its causes.**

Example 2

Consider again the simple causal system involving cholesterol and your heart:

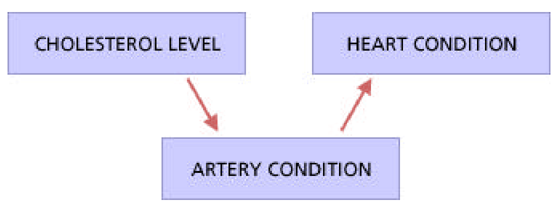

FIGURE 3300-3

In this system, there is a causal relationship posited to exist between cholesterol and artery condition.

< A link to exercises in the interactive version of this module. >

3400: Natural and Experimental Causal Systems

Because interventions can change the relationship between an effect and its causes, to keep things clear we distinguish between the causal system as it occurs **before** an intervention and the system as it occurs **after** the intervention. We call the pre-intervention system the "natural" or "pre-manipulated" system, and the post-intervention system the "experimental" or "manipulated" system.

For example, consider smoking and nicotine stains. In the natural course of things, smoking is a cause of nicotine stained fingers. Thus we draw the following causal graph to represent the "natural" or "pre-manipulated" system:

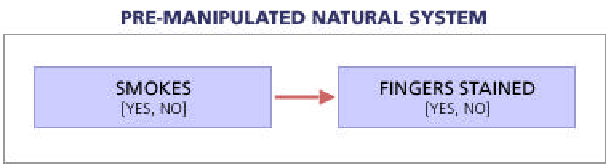

FIGURE 3400-1

Consider two ideal interventions, one to ensure that someone smokes, and another to ensure they have no nicotine stains on their fingers. In the first case, consider the intervention that makes everyone smoke. The post-manipulated system is the same as the pre-manipulated one, because the relationship between smoking and finger stains is unchanged, and it's the only relationship in the natural system.

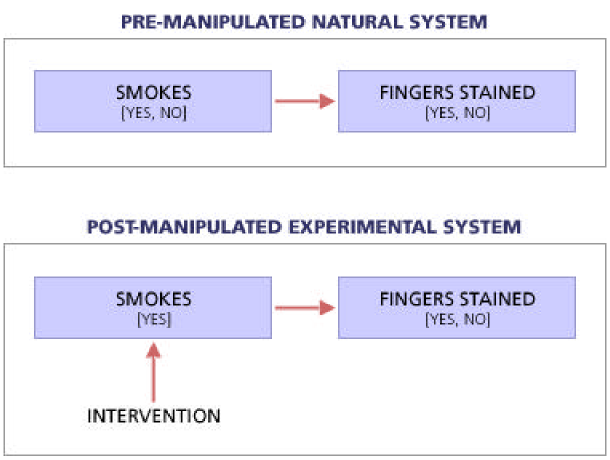

FIGURE 3400-2

In the second case, however, in which we intervene to prevent nicotine stains, the post-manipulated system is different. Consider an intervention in which we scrub everyone's hands with nicotine solvent soap until they are without stains. In that case the influence of smoking on finger stains is broken. It has been usurped by our intervention:

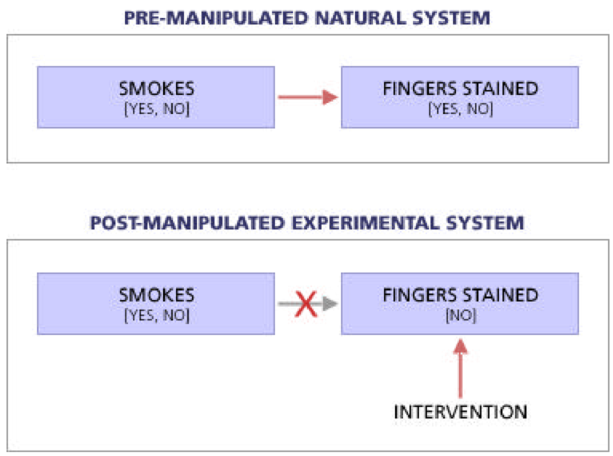

FIGURE 3400-3

It is important to note here that even though we have been talking about intervening on the entire population -- i.e. making everyone smoke, or scrub his or her hands -- an intervention is defined as an action performed on a single individual. Thus, we can talk about intervening on a population to make everyone smoke only if we realize what that is: the same intervention on each individual in the population. If we keep this in mind, then we can also talk about intervening on a population, but not making everyone do the same thing. We could, for instance, intervene on the variable **SMOKES** by making half the population smoke, and preventing the other half from smoking. Or we could intervene on the variable **FINGERS STAINED** by making half the population scrub their fingers and forcing half the population to rub nicotine on their fingers. In either case, whether we intervene on **FINGERS STAINED** by making everyone have the same value or different values, the post-manipulation graph is the same. In both cases we had an ideal intervention, so in both cases the arrow from **SMOKES** to **FINGERS STAINED** will be broken.

In general, moving from the pre- to the post-manipulated causal graphs representing the system is simple. To form the post-manipulated graph, begin with the pre-manipulated graph, and:

+ Add the intervention as an extra variable(s) to the causal graph with an arrow into the variable(s) it directly changes.
+ Break or erase all the arrows pointing into variables that are directly intervened upon.

< A link to exercises in the interactive version of this module. >

3500: Exogenous and Endogenous Variables and Causal Structure

---

It is useful to separate variables in causal systems into two sorts:

+ Independent, or exogenous variables, and
+ Dependent, or endogenous variables

Definition: Independent, or Exogenous Variable

A variable is **independent, or exogenous**, if it is the effect of no other variable in the system.

Definition: Dependent, or Endogenous Variable

A variable is **dependent, or endogenous**, if it is the effect of at least one other variable in the system.

< A link to exercises in the interactive version of this module. >

Because interventions only change the relationship between an effect and its causes, an intervention on an exogenous variable does not change the causal graph at all. Except for adding the intervention variable, the pre-manipulation and post-manipulation causal graphs are the same. We say that such interventions are **structure-preserving**.

Definition: Structure Preserving Intervention

An intervention is **structure preserving** if it is an intervention on exogenous variables only.

An (ideal) intervention on an endogenous variable does change the relationship between it and its causes, so we call interventions on endogenous variables **structure-altering**.

Definition: Structure Altering Intervention

An intervention is **structure altering** if it is an intervention on at least one endogenous variable.

< A link to exercises in the interactive version of this module. >

4000: Exploration

Explore the causal system that consists just of **SWITCH M**, **LIGHT 1** and **LIGHT 2**. The square in the upper right corner is a kind of photoelectric cell that turns light into electricity. It is positioned so as to only receive light from **LIGHT 2**.

< A simulation in the interactive version of this module. >

< A link to exercises in the interactive version of this module. >

Now explore a similar system, but this time with two more switches. Again the square in the upper left corner is a kind of photoelectric cell that turns light into electricity. It is positioned so as to only receive light from **LIGHT 2**.

< A simulation in the interactive version of this module. >

< A link to exercises in the interactive version of this module. >

5000: Interventions on Cyclic Causal Systems

So far, we have only considered acyclic causal systems, that is, cases in which no variable is a direct or indirect cause of itself. The same rules for changing the causal graph from the pre-manipulated to post-manipulated system apply to cyclic cases, however.

In general, moving from the pre- to the post-manipulated system is simple:

+ Add the intervention as an extra variable(s) to the causal graph.
+ Break or erase all the arrows pointing into variables that are directly intervened upon.

For example, consider a bicycle as a causal system:

< A simulation in the interactive version of this module. >

< A link to exercises in the interactive version of this module. >

In this case, the causal system for the bicycle has no cycles. The rear wheel is turned by pedaling and by setting a gear shift, but the motion of the rear wheel has no causal influence on the motion of the pedals. Contrast this to an older style bicycle in which the pedals and rear wheel are more tightly linked:

< A simulation in the interactive version of this module. >

< A link to exercises in the interactive version of this module. >

In this case the causal relationship between the pedals and the real wheel is cyclic. If we intervene and change the state of the rear wheel, however, we break the influence of the pedals on the wheel.

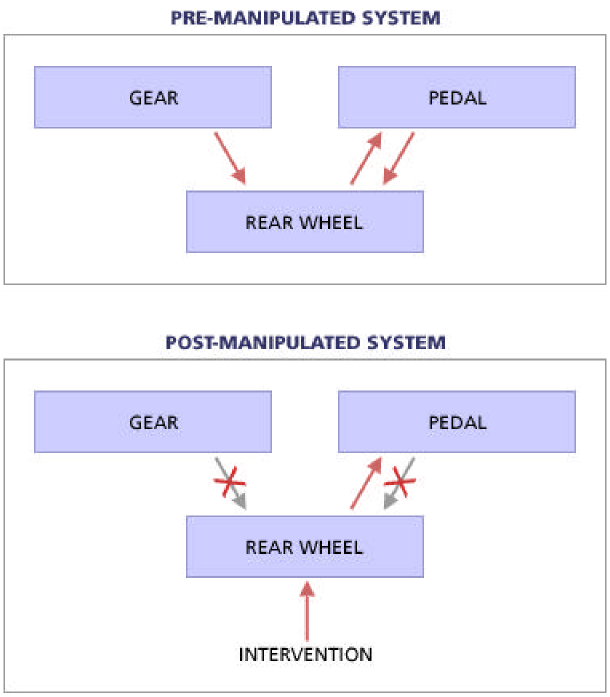

FIGURE 5000-1

< A link to exercises in the interactive version of this module. >

6000: Summary

Causal knowledge informs us about what would happen if we intervened to act in certain ways, but associational knowledge does not. Causal systems involve a set of variables and causal relations among those variables.

The simplest kinds of interventions, which we call **ideal interventions**, are clearly defined actions from outside the causal system that directly determine (fix) the value of one variable in the causal system. In this module we have only covered ideal interventions.

Ideal interventions not only change the value of a variable in a causal system, they sometimes change the causal relationships that obtain in a causal system. Interventions that change the value of a variable that is a cause in the system do not change the relationship between the cause and its effects. Interventions that change the value of an effect do change the relationship between the effect and its causes; they break the relationship between the effect and its causes -- because, by setting the value of the effect the intervention prevents the causes from having any influence on the effect.

The causal graph that describes the relations among the variables in a causal system before we intervene on it is called the **pre-manipulation** graph, or the natural system. The causal graph that describes the causal relations that obtain among the variables in the same system after an intervention is called the **post-manipulation**, or experimental graph.

A variable that is only a cause in a system is called **exogenous**, or independent. A variable that is an effect of at least one other variable in the system is called **endogenous**, or dependent. Interventions that change the value of an exogenous variable are said to be structure-preserving, and interventions that change the value of an endogenous variable are said to be structure- altering. If an intervention is structure-preserving, the pre-manipulation and post-manipulation causal graphs are the same. If an intervention is structure-altering, the pre-manipulation and post-manipulation graphs are different.
