# Determinism and Indeterminism in Causation

---

## 1000: Introduction

---

Suppose you have agreed to buy a house for $100,000, but you are still negotiating with the current owners about what needs to be fixed before they move out. As you walk through the house with the inspector you've hired, you notice that it has really old heaters called Taylor burners in really old fire places. The inspector says they need to come out because they have asbestos lining, and asbestos causes cancer. Mrs. Smith, the current owner, is offended and doesn't want to pay for professionals to remove the asbestos. "I raised five kids here, and it didn't do them any harm," she says. "My relatives lived in houses with these heaters for 40 years, and none of **them** got cancer," she continues. How should you reply? If asbestos really causes cancer, then why didn't her children or her relatives get it?

So far, all the causal relationships we have studied are "deterministic," that is, the effect is always completely determined by the values assigned to the causes. In an "indeterministic" causal relationship, the effect is not always determined by the causal assignments. For example, smoking causes cancer, but doesn't determine it. Some people who smoke live to 100 and never get lung cancer.

Almost every causal relationship we study in the social, behavioral, and medical sciences at least **appears** to be indeterministic. Having wealthy parents is a cause of wealth, but it does not guarantee it. Eating well and exercising regularly cause good health, but they don't guarantee it. We thus cannot settle for an account of causation in which causes must completely determine their effects.

There are three core ideas in this module. The first is that causal relations are sometimes indeterministic. The second is that sometimes they are indeterministic in appearance only. In most cases we don't explicitly **mention** all of the causes that act on an effect of interest, but only some of them. If we did list **all** of the causes of a particular effect, then the relationship between the causes and the effect might well be deterministic. When we leave some out, however, the relationship will **appear** to be **indeterministic**.

 

The third core idea is this: even though a particular causal assignment might not always determine whether a given effect will or will not occur, it can determine the **chances**, or the **probability** that the effect will occur. At the least, switching from one causal assignment to another can change the chances that the effect will occur. Being exposed to asbestos, for example, does not guarantee that you will get cancer, but it does raise the **chances** that you will get it compared with not being exposed. Even though Mrs. Smith's children and relatives got lucky, you don't want to expose your children to asbestos, because to do so would increase their chances for getting cancer later in their life.

The idea that causes only determine the **probability** of their effects is why causal reasoning needs statistics. If smoking had to lead to lung cancer in every single case before we agreed to say that smoking causes cancer, then a single anecdote about some old man who smoked until 99 but died of old age with no trace of lung cancer would be enough to defeat the causal claim and thus alter our policy of labelling cigarette packs with warnings like: "Smoking causes Cancer." We wouldn't need to keep data around about thousands of smokers or lung cancer patients. Saying that smoking changes the **chances** of lung cancer is another matter entirely. We need to know the chances of getting lung cancer if you are made to smoke, and the chances of getting it if you are made not to smoke, and to answer these questions we need to collect data. Whether or not the data we measure confirm that the chances of lung cancer are **really** different is the province of statistics.

We begin this module by interactively illustrating the first core idea: that in cases where we don't explicitly mention all the causes of an effect, the causes we do mention do not appear to determine the effect. Later in the module we delve into how causes can change the **chances** of their effects. We leave all the statistics out - that comes much later.

---

## 2000: Exploration

---

### 2100: A Concrete Example: Cell Phones

---

Suppose you are traveling in a car and you want to make several calls on your cell phone. In the simulation below, click on the "SEND" button to place a call, and click on the "END" button to end a call (you must click on END before you can try another call). The phone on the right will ring if your call got through. Attempt at least 10 calls.

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

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

In real life, calls from cell phones in cars do not always go through. So this is clearly a case in which the cause, clicking on the "SEND" button, does not guarantee that the effect, the phone ringing, will occur. Here, and in the next few sections that follow, we abstract from this particular case to the general concept of indeterministic causation. First, we translate this case into the language of variables and response structures. This will allow us to discuss indeterministic causation via indeterministic response structures.

The cell phone system, as we have presented it to you, includes two variables, each of which has two values:

**TABLE 2100-1: CELL PHONE SYSTEM**

| Variable | Values |
|---|---|
| **BUTTON PUSHED** | [Send, End] |
| **CALL CONNECTED** | [Yes, No] |

Consider the effect: **CALL CONNECTED**. What is the **response structure** for this effect? The response structure for an effect lists the values that the effect takes on for every possible causal assignment of the other variables in the system. So to build the response structure, first enumerate all the possible causal assignments, and then fill in the value of the effect for each. In this case there are exactly two causal assignments, one for each value we might give to the **Button Pushed** variable:

**TABLE 2100-2: RESPONSE STRUCTURE FOR THE CELL PHONE SYSTEM**

| Causal Assignment | Effect: Call Connected |
|---|---|
| **BUTTON PUSHED** = Send | ??? |
| **CALL CONNECTED** = End | No |

Hitting the End button **does** determine the effect: the call is not connected. The problem is assigning a value to the effect for the first causal assignment: **BUTTON PUSHED** = Send. As you experienced in the simulation - we don't know what will happen to the effect in this causal assignment. Sometimes the call goes through, and sometimes not. Because of this, we say that this response structure is indeterministic, and we separate deterministic from indeterministic causal relationships by separating deterministic vs. indeterministic response structures.

**Definition: Deterministic Response Structure**

In a **deterministic response structure**, every causal assignment completely determines the value of the effect.

**Definition: Indeterministic Response Structure**

In an **indeterministic response structure**, there is at least one causal assignment that does not determine the value of the effect.

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

### 2200: An Abstract Example: Red and Blue Squares

Having translated the cell phone example into the language of variables and response structures, we can now discuss indeterministic causation abstractly. We will use an example that has the same structure as the cell phones, but that involves abstract red and blue squares that have no connection to objects with which you are familiar.

In the following simulation, the state of the red square causes the state of the blue square. In the "Experiment" section of the simulation, you can change the causal assignment for the red square from solid to hatched (dots), or vice versa, by clicking on it. The "Results Table" will record how many times the effect (blue square) came out solid or hatched in response to the causal assignment you gave to the red square: 1) red square = solid, 2) red square = hatched. Click on the red square enough times to bring about causal assignment 1) (solid red square) at least 10 times and causal assignment 2) (hatched red square) at least 10 times.

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

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

Since we don't know what will happen to the blue square in one of the causal assignments for the red square, the response structure is indeterministic. The red square is a cause of the blue square, but an indeterministic cause.

### 2300: Individuals and Populations

In the modules on Event Causation and Variable Causation, we discussed three different levels of causal relationships:

1. causal relations between particular events (e.g., the Titanic sank when it hit an iceberg),
2. causal relations between variables in an individual (e.g., poison ivy contact events cause rashes in Smith)
3. causal relations between variables in a population (e.g., poison ivy contact events cause rashes among people in general)

So far, we have described indeterministic causation as a feature of a response structure, which applies to particular individuals. In the cell phone and red-blue-square simulations, we were dealing with a single individual. We discovered that the causal relation between placing a call (hitting "Send") and having it go through is indeterministic by repeatedly placing the same phone in the same causal assignment and observing different effects at different times. Sometimes the call got through, and sometimes it didn't. In the red-square blue-square example, we did the same thing. The simulation allowed us to repeat the same causal assignment many times for the same **individual**. When we assigned the red square to solid, sometimes we observed a solid blue square, and sometimes we observed a hatched blue square. Thus, we discovered that, for this individual, the response structure for the blue square is indeterministic.

As we stressed in the module on Variable Causation, scientists often cannot examine the response structure of particular individuals - they can only gather data on a population of indviduals. So first, what is it for causation in a population to be indeterministic?

**Definition: Apparent Indeterministic Causation in a Population**

If there are two individuals in a population who exhibit different effects even though they have the same causal assignment, then there is **apparent indeterministic causation in the population.**

Recall the idea of response structure uniformity in a population:

**Definition: Response Structure Uniformity**

A population has **response structure uniformity** for a given effect if every individual in the population has the same response structure for that effect.

Notice that this definition doesn't say anything about whether each individual has a deterministic or indeterministic response structure, it only insists that they all have the **same** response structure.

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

We can tell if a population of individuals with deterministic response structures violates response structure uniformity if we give two individuals the same causal assignment but then observe them to have different effects. This won't always happen, because two individuals that have different response structures might have the same response for all of their causal assignments but one, and if we don't put them in the causal assignment that shows the difference, we might never observe their difference.

So what happens to this problem if we don't assume that every individual in a population has a deterministic response structure. What happens if we allow for indeterministic response structures?

In the simulation above, you performed the same experiment over and over on the same pair of red and blue squares. In the simulation below, we give you a population of 20 **distinct** individual pairs. In each of the individual pairs in the Experiment section, you can click on the red square to change its causal assignment. Click on all the red squares **once**, and notice how many of the blue squares match. Now repeat the experiment, and again notice how many of the blue squares change.

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

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

Causation in a population can **appear** indeterministic for three reasons:

1. The population is response structure uniform, but each individual in the population has an indeterministic response structure.
2. Each individual in the population has a deterministic response structure, but the population is not response structure uniform.
3. The population is not response structure uniform, and some individuals in the population have an indeterministic response structure.

If, when we programmed the population simulation of the 20 pairs of red and blue squares, we took the individual red-blue square simulation (2200-1) and duplicated it 20 times, then the population would appear indeterministic for the first reason.

If, when we programmed the population simulation of the 20 pairs of red and blue squares, we created 20 different individual red-blue pairs, some of which had one deterministic response structure and some of which had another, then the population would appear indeterministic for the second reason.

If, when we programmed the population simulation of the 20 pairs of red and blue squares, we created 10 different red-blue pairs, some of which had one deterministic response structure and some of which had another, and we took 10 that had identical but indeterministic response structures, then the population would appear indeterministic for the third reason, which is really a combination of the first two.

In fact we took the first option, and programmed the population simulation with 20 duplicates of the single pair simulation. Could you tell from the experiment you did on the population simulation, or would you have to do more?

---

**2400: Cell Phones Revisited**

---

You should now have a reasonable grasp of indeterministic response structures. Both the cell phone and red-blue-square simulations you used in the previous sections are instances. In this section we reveal how an indeterministic response structure can only **appear** to be indeterministic because some of the causes for an effect were left out.

In the original simulation on the cell phone, you probably only managed to get about half of your calls to go through. Why? Isn't hitting the "SEND" button a cause of a call getting through? The answer is yes, it is a cause, but not the **only** cause.

We only showed you part of the story in the cell phone simulation. Here we uncover another cause of getting the call to go through -- whether you are in range of a cell phone tower or not. You cannot control whether you are in range of the tower -- but in this simulation you can observe it (your location appears as a small square).

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

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

---

### 2500: Red and Blue Squares Revisited

---

In this section we abstract away from cell phones to red, blue, and now green squares.

What happened in the cell phones is the same as what happened in the red and blue square simulation, we only gave you part of the story. In the following applet you get the full picture:

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

Change the state of the red square, and **notice** the state of the green square. When you change the state of the red square, the program assigns a value at random to the green square. Sometimes it assigns it hatched green and sometimes solid green. The red and green squares are both causes of the blue square. Like the cell phone situation, once you know about the hidden factor (the state of the green square in this case and whether the car was in range of the cell tower in the cell phone case), you see that the causal assignment (of both the red and green squares) fully determines the effect. As we have programmed it, you have no control over the green square -- only the red one. Nevertheless, the causal assignment among the red and green square totally determines the state of the blue square, which is the effect in this system. In this **expanded** system, the effect now has a deterministic response structure.

This table summarizes the response structure for the blue square in the expanded system, but leaves one response blank for you to complete:

**TABLE 2500-1: RESPONSE STRUCTURE FOR THE SYSTEM OF SQUARES**

| Causal Assignment | Causal Factor 1: RED | Causal Factor 2: GREEN | Effect: BLUE |
|---|---|---|---|
| 1 | Solid | Solid | Solid |
| 2 | Solid | Hatched | Hatched |
| 3 | Hatched | Solid | |
| 4 | Hatched | Hatched | Hatched |

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

The simulation that included no green square is programmed identically to the one here, except that you are not **shown** the green square. In the previous simulation, it was a **hidden variable**. This is exactly the same as the cell phone simulations. In both cases, the simulations are identical, but in one you are shown the range from the tower, and in the other it is kept hidden from your view.

---

## 3000: Indeterminism and Pseudo-Indeterminism

---

### 3100: Introduction

---

When the response structure for the cell phones includes **BUTTON PUSHED** and **TOWER RANGE**, i.e., **all** the causes of whether the call goes through, it is a deterministic response structure. When it includes only **BUTTON PUSHED**, it is an indeterministic response structure. Since the indeterminism in the second case is the result of leaving out a variable, we call it **pseudo-indeterminism**. In this section, we make the idea of pseudo-indeterminism a little more precise, contrast it with true indeterminism, and examine the practical consequences of the distinction.

---

### 3200: Pseudo-Indeterminism

---

Concretely, when the response structure for the cell phones includes all the causes of the call going through: **BUTTON PUSHED** and **TOWER RANGE**, it is deterministic. When it includes only **BUTTON PUSHED**, it is indeterministic. Since placing a call on the cell phone only appears to be an indeterministic cause of the call getting through when **TOWER RANGE** is left out of the story, we label the relationship pseudo-indeterministic. Here we explain pseudo-indeterminism and make it a little more precise.

Consider the case of malaria in the context of deterministic vs. indeterministic causation. Leaving out the irrelevant factor of whether the individual drank gin and tonics regularly, we represented the causal structure governing whether or not Sam (or others) would get malaria with the following response structure:

**TABLE 3200-1: RESPONSE STRUCTURE FOR THE MALARIA CASE**

| Causal Assignment | Causal Factor 1: BITTEN | Causal Factor 2: HAS GENE | Causal Factor 3: INOCULATED | Effect: MALARIA |
|---|---|---|---|---|
| 1 | True | True | True | False |
| 2 | True | True | False | False |
| 3 | True | False | True | False |
| 4 | True | False | False | True |
| 5 | False | True | True | False |
| 6 | False | True | False | False |
| 7 | False | False | True | False |
| 8 | False | False | False | False |

Since each of the eight possible causal assignments in the table are different with respect to the variables: **BITTEN**, **HAS GENE**, and **INOCULATED**, and only one possible outcome for the effect is listed for each assignment, the effect has a deterministic response structure.

Suppose, however, that the system was again the same, but we had no knowledge of the genetic make-up of each individual, e.g., Table 3200-2.

**TABLE 3200-2: RESPONSE STRUCTURE FOR MALARIA -- WITH GENE UNOBSERVED**

| Causal Assignment | Causal Factor 1: BITTEN | Causal Factor 2: HAS GENE | Cusal Factor 3: INOCULATED | Effect: MALARIA |
|---|---|---|---|---|
| 1 | True | Unobserved | True | False |
| 2 | True | Unobserved | False | False |
| 3 | True | Unobserved | True | False |
| 4 | True | Unobserved | False | True |
| 5 | False | Unobserved | True | False |
| 6 | False | Unobserved | False | False |
| 7 | False | Unobserved | True | False |
| 8 | False | Unobserved | False | False |

Suppose that we only observed whether each person was bitten by a malaria carrying mosquito and whether each person was inoculated. Even though there are in fact eight different possible causal assignments, observationally we can now only distinguish four. In Table 3200-2, assignments 1 and 3 now appear identical to us, as do assignments 2 and 4, 5 and 7, and 6 and 8.

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

We say that the response structure for malaria in Table 3000-2 is pseudo-indeterministic.

**Definition: Pseudo-indeterministic Response Structure**

---

A response structure for an effect E is **pseudo-indeterministic** when i) it is an indeterministic response structure for E, and ii) there are other causes of E that, when added, make the response structure deterministic.

In a pseudo-indeterministic system, the indeterminism arises only because we have not observed all the relevant causal factors. If we had, then the system would appear deterministic. So even if we believe that the world is at bedrock deterministic, whether it appears deterministic is relative to what we are or are not able to observe.

Until the 20th century, almost all scientists believed that the world was fundamentally deterministic. Pierre Simon de Laplace, one of the greatest French mathematicians and astronomers of the 18th century, claimed that if we could know the current state of the universe down to the smallest detail, we could predict exactly how it will be at any time in the future and how it was at any time in the past. He was well aware that most things **appeared** indeterministic, but he was convinced that it was the limitations on what we could observe and measure and not the nature of the world that gives rise to the indeterminism we observe. Laplace believed that we live in a pseudo-indeterministic world.

---

## 3300: True Indeterminism

---

In this century, the theory of Quantum Mechanics, which describes the world at very small scales (an atom and smaller), has convinced most scientists that Laplace was wrong -- the world is not pseudo-indeterministic. In the standard interpretation of Quantum Mechanics, the world is **truly indeterministic**, especially at the sub-atomic scale. For example, the element Carbon 14 is radioactive. Carbon atoms that have a total of 14 protons and neutrons in their nucleus sometimes radioactively decay. If we had a trillion Carbon 14 atoms in a box, only a small percentage of them would decay within an hour, and nothing in physical theory can tell us which atoms will decay in that time and which will not. A disciple of Laplace might argue that the indeterminism among Carbon 14 radioactive decay is only in appearance -- if we just knew more about each Carbon 14, i.e., other causes of radioactive decay, then we could indeed predict perfectly whether it would or would not decay in the first hour. Most modern physicists disagree -- they believe that there is nothing else about the atoms we could learn that would help in predicting whether or not a Carbon 14 atom decays in any specific time period. They believe that radioactive decay is a truly indeterministic phenomenon.

There remain to this day a minority of physicists that refuse to believe that "God plays dice" with the universe, and contend that we simply don't understand the "hidden" variables in cases such as atomic decay. These scientists pursue "hidden variable theories" that try to describe the atomic world as pseudo-indeterministic, not truly indeterministic.

**Definition: Truly Indeterministic Response Structure**

---

A response structure for an effect E is **truly-indeterministic** when i) it is an indeterministic response structure for E, and ii) there is no set of causes we can add that will make it a deterministic response structure.

---

### 3400: The Practical Situation

---

As you can see, whether or not a response structure for an effect is indeterministic is sometimes a question of what we observe. If a system appears indeterministic, it might be because it is pseudo-indeterministic and we have left out some important causes in our description of the response structure, or it might be because the system is truly indeterministic.

For virtually any effect that we care to study in Biology, Psychology, or any Social Science, we are not even close to being able to identify and measure **all** of its causes. Thus the distinction between pseudo-indeterministic and truly-indeterministic systems turns out to be more of philosophical than practical interest.

What are the causes of depression? Brain chemistry, emotional trauma in childhood, poor relationships, death of a loved one? Perhaps, but surely the response structure for depression with these variables is not deterministic. Adding other causes besides these might help us predict depression more reliably, but we are unlikely to ever be in a position to predict or prevent depression with perfect certainty. We are unlikely to ever be able to even identify a fraction of all the causes of depression. The same is true of conflict, crime, and economic health. Given this, causal reasoning in any domain is by necessity about indeterministic relationships. Whether they are pseudo-indeterministic or truly indeterministic matters little. What matters is how to understand and reason about indeterministic causal relationships.

---

## 4000: Indeterministic Causation

---

### 4100: Introduction

---

Having equipped ourselves with plenty of terminology, we now need to include indeterminism as part of an account of variable causation. What do mean when we say that smoking causes lung cancer, even though it does not determine it?

Roughly, we mean that the chances of getting lung cancer are different among people who are made to smoke than they are among people made not so smoke. A more refined answer involves a combination of statistical ideas (for example, relative frequency, independence, and random assignment) and causal ideas (such as ideal manipulation and unobserved confounding), all of which we will cover in the modules that follow. In this section, however, we give a basic account of indeterministic causation that will be enough for now, and contains the core of the intuitions behind a more sophisticated account.

---

### 4200: Cell Phones and Indeterminism

---

Consider the causal system for the cell phone again.

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

In a response structure for **CALL CONNECTED** in the system in which the only cause under consideration is **BUTTON PUSHED**, there are exactly two causal assignments, one for each value we might give to **BUTTON PUSHED**.

**TABLE 4200-1: RESPONSE STRUCTURE FOR THE SMALL CELL PHONE SYSTEM**

| Causal Assignment | BUTTON PUSHED | Effect: CALL CONNECTED |
|---|---|---|
| 1 | Send | ??? |
| 2 | End | No |

The problem is that we don't know what value to write in for the effect **CALL CONNECTED** when the causal assignment is **BUTTON PUSHED** = Send.

When we considered the full Cell Phone system, however, we included the variable: **IN TOWER RANGE**.

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

In this full version of the system the Response Structure has four different causal assignments, and is deterministic.

**TABLE 4200-2: RESPONSE STRUCTURE FOR THE FULL CELL PHONE SYSTEM**

| Causal Assignment | BUTTON PUSHED | IN TOWER RANGE | Effect: CALL CONNECTED |
|---|---|---|---|
| 1 | Send | Yes | Yes |
| 2 | Send | No | No |
| 3 | End | Yes | No |
| 4 | End | No | No |

The reason the response structure for the **small** system appears indeterministic is that in the causal assignment labelled: **BUTTON PUSHED**=Send, we don't know whether the hidden variable **IN TOWER RANGE** is Yes or No. We represented this uncertainty by putting "???" in for the value of the effect **CALL CONNECTED**.

As it turns out, we can often do better at representing our uncertainty than just writing "???." In all the Cell Phone simulations, we had the computer set the value of **IN TOWER RANGE** by flipping a computer's version of a fair coin. About 50% of the time, **IN TOWER RANGE** should equal Yes, and about 50% of the time it should equal No. That means that about 50% of the time you pushed the Send button, the call went through, and 50% it didn't. If you only did a few trials, you might have gotten a proportion quite different from 50% by luck, but if you did alot of trials its very unlikely you got something very different from 50%. In any event, instead of putting "???" into the response structure for the small system, we could have put something more informative about our uncertainty, like: Yes = 50%, No = 50%.

**TABLE 4200-3: RESPONSE STRUCTURE FOR THE SMALL CELL PHONE SYSTEM**

| Causal Assignment | BUTTON PUSHED | Effect: CALL CONNECTED |
|---|---|---|
| 1 | Send | Yes = 50%, No = 50% |
| 2 | End | No |

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

In our Cell Phone simulations, although pushing the Send button doesn't guarantee that your call will go through, it does change the chances of the call getting through. If you don't push the Send button, the chances of the call getting through are 0. If you do push it, they change to 50%. Thus pushing the send button is a cause of the call going through.

---

### 4300: Quantifying Indeterministic Causation

---

Although we must accept that some response structures are indeterministic, we don't have to accept ones in which there are completely vague entries for the effect like "???". We can at least quantify our uncertainty about whether the effect will occur in each causal assignment. We do so with probabilities, or chances.

For example, consider the following (fictitious) indeterministic response structure for smoking and lung cancer.

**TABLE 4300-1: RESPONSE STRUCTURE FOR SMOKING AND LUNG CANCER**

| Causal Assignment | SMOKER | Effect: LUNG CANCER = Yes | Effect: LUNG CANCER = No |
|---|---|---|---|
| 1 | Yes | 15% | 85% |
| 2 | No | 1% | 99% |

Instead of a standard response structure table, in which there is only one column for the effect, we list, for each causal assignment, the chances the effect will be cancer and the chances it will be not cancer. In this example, we don't know with certainty what the effect will be in any causal assignment, but using a table like this we can express how the **chances** of the effect change as we change the causal assignment.

If the effect is a variable that can take on more than two values, then the response structure is more complicated. For example, suppose that we want to represent how the variable **INCOME** depends on the variable **EDUCATION**. Suppose that**EDUCATION** can take on the values High School, College, or Graduate Degree, and **INCOME** can take on the values Low, Medium, or High. Below is an example of an indeterministic response structure in which we give the chances for achieving different levels of **INCOME** as a function of the causal assignment for **EDUCATION**.

**TABLE 4300-2: RESPONSE STRUCTURE FOR EDUCATION AND INCOME**

| Causal Assignment | EDUCATION | Effect: INCOME = Low | Effect: INCOME = Medium | Effect: INCOME = High |
|---|---|---|---|---|
| 1 | High School | 60% | 35% | 5% |
| 2 | College | 20% | 70% | 10% |
| 3 | Grad Degree | 5% | 70% | 25% |

According to this table, the chances of earning a sizable amount of money increase if we increase the amount of education in the causal assignment. The probabilities for the values of the effect must add to 1, and in general the way the total of 1 is divided up among the different values is called a "probability distribution."

So in general, if we can't construct a deterministic response structure for an effect variable, we want a response structure in which we list the probability distribution over the values the effect might take on. This gives us a quantitative sense of how the chances of the effect depend on the causal assignment given.

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

---

### 4400: Defining Indeterministic Causation

---

We now have all the pieces we need to include indeterminism in an account of variable causation. Recall the two key definitions we gave in the module on Variable Causation.

**Definition: Test Pair of Causal Assignments**

If two causal assignments C1 and C2 are identical except for the values assigned to variable **X**, then C1 and C2 are a **test pair of causal assignments** for **X**.

**Definition: Direct Cause**

If, in a system of variables **S** there are any test pair of causal assignments for **X** in which there is a difference in the effect **Y**, then **X** is a **direct cause** of **Y** relative to **S**.

The idea behind test pairs of causal assignments is to keep everything in the system the same except for one variable, which we make different between one causal assignment and another. If the effect changes, then we have causation.

What is it for the effect to **change**, however? Does it mean that the effect must take on a different value in the two causal assignments? Or is it enought that the **chances** for the effect change? Clearly, in the case of indeterministic causation, only the chances need to change. So let's modify the definition of cause ever so slightly to accomodate indeterminism:

**Definition: Direct Indeterministic Cause**

If, in a system of variables **S** there are any test pair of causal assignments for **X** in which there is a difference in the probability of the effect **Y**, then **X** is an **direct indeterministic cause** of **Y** relative to **S**.

Just as we argued that the effect need not be different across **every** test pair when we defined cause, the parallel holds for our definition of indeterministic cause. The chances of the effect don't have to change across every test pair of causal assignments, only across at least one pair. For example, consider an indeterministic response structure that depicts the effect of **INCOME** on **HAPPINESS**.

**TABLE 4400-1: RESPONSE STRUCTURE FOR INCOME AND HAPPINESS**

| Causal Assignment | INCOME | Effect: HAPPINESS = Low | Effect: HAPPINESS = High |
|---|---|---|---|
| 1 | Low | 60% | 40% |
| 2 | Medium | 45% | 55% |
| 3 | High | 45% | 55% |

According to this response structure, changing **INCOME** from Low to Medium improves the chances for high **HAPPINESS**, but changing from Medium to High **INCOME** doesn't have any effect. After you earn more than low income, extra happiness depends more on other factors.

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

So, having come this far, how do we now answer Mrs. Smith, who doesn't want to pay for removing the Taylor burners that have asbestos because she says: "My relatives lived in houses with these heaters for 40 years, and none of **them** got cancer"?

Perhaps we respond as so: "Asbestos exposure need not guarantee cancer to be a cause of it. It need only increase the chances of getting cancer, and scientists believe it does. Your relatives were lucky to avoid cancer, even though they increased their chances of getting it by living with Taylor burners. I don't want to expose my own children to asbestos and increase their risk of getting cancer, even if it is possible that they will be as lucky as your relatives, so please remove them before we move in." Flustered, Mrs. Smith sees the irrefutable logic in your answer and agrees immediately to all of your demands. You promptly send a donation to the authors of Causal and Statistical Reasoning, and all are happy. Only kidding about the donation.

---

## 5000: Summary

---

The **response structure** for an effect lists the values the effect takes on for every possible causal assignment of the causes mentioned in the system. In a **deterministic response structure**, every causal assignment completely determines the value of the effect. In an **indeterministic response structure**, there is at least one causal assignment that does not determine the value of the effect.

For many effects of interest, like cancer, a response structure that includes causal assignments for all the causes we know about now still **appears indeterministic**. Someday in the future, we will perhaps actually be able to measure **all** the causes of cancer. Suppose that when we do, it turns out the full response structure for cancer turns out to be deterministic. In that case the current response structure for cancer is **pseudo-indeterministic**. The response structure we currently know about for cancer is indeterministic, but it is embedded within a bigger response structure that is deterministic.

A response structure for an effect E is **pseudo-indeterministic** when: it is an indeterministic response structure for E, and there are other causes of E that, when added, make the response structure deterministic.

Unfortunately, the system that we specify and observe in practice is rarely more than a partial version of the causal system that governs the world. For almost no medical or social or psychological system, for example, do we believe we can identify and observe all the causal factors that contribute to producing an effect. For example, although the Surgeon General concluded in 1965 that smoking causes lung cancer, he did not conclude (and it is not true) that all smokers get lung cancer. Even considering dietary habits, family history, and other environmental and life-style factors, we cannot identify any causal assignment that is sufficient to determine whether someone will or will not get lung cancer. We can identify many situations that result in lung cancer more frequently, but none that guarantee it, and none that guarantee preventing it.

Will it ever be possible to specify all the causes of lung cancer, so that we can build a deterministic response structure? No one knows. If it is not possible, then the response structure for cancer is truly indeterminisitic. A response structure for an effect E is **truly-indeterministic** when: it is an indeterministic response structure for E, and there is no set of causes we can add that will make it a deterministic response structure.

Whether or not the world is fundamentally deterministic or indeterministic, in practice we can only specify and measure some of the relevant factors that cause an effect of interest. Indeed, if we don't make prior decisions about which factors might be relevant, the number of potential causal assignments is practically infinite. In almost every case, however, no matter how many factors we include in the causal system we **specify**, we will leave out some that are relevant. As a result, in practice our representations of causal structures are always partial, and we have to accept the fact that indeterministic response structures are the rule, not the exception.

To deal with indeterministic response structures, we quantify our uncertainty about the effect for each assignment we might give to the causes. In a quantified indeterminisitic response structure, for each causal assignment we give the chances for each possible value of the effect variable, that is, we give the probability distribution over the effect.

 

One variable **X** is a cause of another **Y** if there is are two causal assignments identical except for the value of **X**, that is, if there is a **test pair** of causal assignments for **X** such that **Y** is different across the two assignments.

One variable **X** is an indeterministic cause of another **Y** if there is are two causal assignments identical except for the value of **X**, that is, if there is a **test pair** of causal assignments for **X** such that the probability distribution over **Y** is different across the two assignments.

---