

© *Jan. 2020*

# Curriculum vitae

Falnésio Ghander Soares Borges

FALNESIO.GHANDER@ECONOMIA.UFJF.BR

(33) 9 2000-5046

*19/08/1992*



# Contents

. 5

0.1 *Habilidades Virtuais* 5

*Part I* 9

1 *The Monty Hall Problem* 9

1.1 *Diagramming the Solution* 10

1.2 *Lessons Learned* 10

*Exercises* 13

2 *Logic* 15

2.1 *Validity & Soundness* 15

2.2 *Propositions* 17

2.3 *Visualizing Propositions* 17

2.4 *Strength* 19

2.5 *Forms of Inductive Argument* 20

*Exercises* 21

3 *Truth Tables* 23

3.1 *Connectives* 23

3.2 *Truth Tables* 25

3.3 *Logical Truths & Contradictions* 27

3.4	<i>Mutual Exclusivity &amp; Truth Tables</i>	28
3.5	<i>Entailment &amp; Equivalence</i>	29
3.6	<i>Summary</i>	30
	<i>Exercises</i>	31
4	<i>The Grue Paradox</i>	33
	<i>A Gruesome Concept</i>	33
	<i>The Paradox</i>	34
	<i>Grue &amp; Artificial Intelligence</i>	35
	<i>Disjunctivitis</i>	36
	<i>Time Dependence</i>	37
	<i>The Moral</i>	37
5	<i>The Problem of Induction</i>	39
	<i>The Dilemma</i>	39
	<i>The Problem of Induction vs. the Grue Paradox</i>	40
	<i>Probability Theory to the Rescue?</i>	41

•

GRADUADO em Ciências Econômicas na Universidade Federal de Juiz de Fora, Campus GV e participa do Núcleo de Agroecologia de Governador Valadares NAGÔ. Tem interesse em sistemas complexos. Acredita que modelos que incorporam as sutilezas da heterogeneidade social são úteis na produção de soluções democráticas mais eficazes e eficientes para demandas socioeconômicas e que o estudo econômico tem muito espaço para multidisciplinaridade na formulação de modelos simples mais próximos da realidade.

É um pesquisador muito dedicado e focado, é acostumado a entrar em projetos e iniciativas novas, se adaptando aos processos, metodologias e técnicas, aceitando desafios e sempre aprendendo. Trabalha bem em equipe e sempre busca passar o conhecimento que adquiriu para frente.

Informações de contato:

[falnesio.ghander@economia.ufjf.br](mailto:falnesio.ghander@economia.ufjf.br)

Telegram: (33) 9 2000-5046

LinkedIn

Lattes

GitHub

## 0.1 Habilidades Virtuais

---

**Linguagens** - Português Fluente - Inglês Fluente - Português Fluente - Inglês Fluente - Inglês Fluente - Português Fluente - Inglês Fluente

**Linguagens de Programação** - R - Python - SQL

**Softwares** - Microsoft Excel - Microsoft Word

A “cheat sheet” summarizing key definitions and formulas appears in [Appendix][Cheat Sheet] [A][Cheat Sheet]. Further appendices cover the axiomatic construction of probability theory, Hume’s problem of induction, and Goodman’s new riddle of induction.

I usually get a mix of students in my course, with different ideological inclinations and varying levels of background. For some the technical material is easy, even review. For others, a healthy skepticism about scientific methods and discourses comes naturally. My goal is to get these students all more or less on the same page.

By the end of the course, students with little formal background

have a bevy of tools for thinking about uncertainty. They can understand much more of the statistical and scientific discourse they encounter. And hopefully they have a greater appreciation for the value of formal methods. Students who already have strong formal tools and skills will, I hope, better understand their limitations. I want them to understand why these tools leave big questions open—not just philosophically, but also in very pressing, practical ways.

THE book was made with the bookdown package created by Yihui Xie. It's a wonderful tool, built on a bunch of other technologies I love, especially the R programming language and the pandoc conversion tool created by philosopher John MacFarlane. The book's visual style emulates the famous designs of Edward Tufte, thanks to more software created by Yihui Xie, J. J. Allaire, and many others who adapted Tufte's designs to HTML and PDF (via LaTeX).

If it weren't for these tools, I never would have written this book. It wouldn't have been possible to create one that does all the things this book is meant to do. I also owe inspiration to Kieran Healy's book *Data Visualization: A Practical Introduction*, which uses the same suite of tools. It gave me the idea to use those tools for an updated, open, and visually enhanced rendition of the classic material from Skyrms and Hacking.

Finally, I'm indebted to several teaching assistants and students who helped with earlier drafts. Thanks especially to Liang Zhou Koh and Meagan Phillips, who contributed several exercises; to Soroush Marouzi and Daniel Munro, who worked out the kinks during the first semester the book was piloted; and to the students in that course, who bore with us, and contributed several corrections of their own.

## **Part I**





# 1 The Monty Hall Problem

*When tackling a math problem,  
I encourage students to draw a picture.*  
—Francis Su

IMAGINE you're on a game show. There are three doors, one with a prize behind it. You're allowed to pick any door, so you choose the first one at random, door A.

Now the rules of the game require the host to open one of the other doors and let you switch your choice if you want. Because the host doesn't want to give away the game, they always open an empty door.

In your case, the host opens door C: no prize, as expected. "Do you want to switch to door B?", the host asks.

Pause a moment to think about your answer before reading on.

WHAT did you decide? Did you conclude it doesn't matter whether you stick with door A or switch to door B?

If so, you're in good company. Most people find this answer sensible, including some professors of statistics and mathematics. They figure there are only two possibilities remaining, door A and door B, each with the same one-in-two chance of being the winner. So it doesn't matter which one you pick.

But the right answer is you should switch. Door B is now twice as likely to be the winner as door A. Why?

The reason is subtle. One way to think about it is that the host's choice of which door to open is a bit of a tell. Maybe they *had* to open door C, because the prize is behind door B and they didn't want to give that away. Of course, it could be behind door A instead, so maybe they just picked door C at random. But there was only a one-in-three chance the prize would be behind door A. Which means there's a two-in-three chance they didn't really have a choice, they had to open door C to avoid showing you the prize behind door B.

Here's another way to think about it. Imagine the game had a hundred doors instead of just three. And suppose again you start by picking the first door at random. Then the host opens *all the other doors*



The Monty Hall problem is named after the creator and host of the game show *Let's Make a Deal*.



Marilyn vos Savant made the Monty Hall problem famous when she solved it correctly in her "Ask Marilyn" column for *Parade* magazine. Read more about it in *The New York Times*.

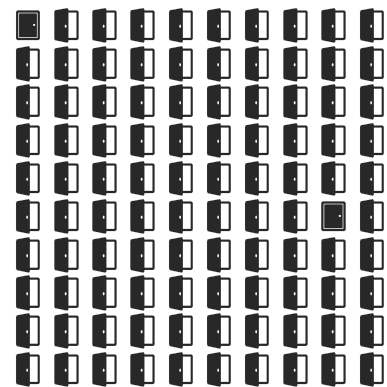


Figure 1.1: The hundred-door version of the Monty Hall problem, suggested by Marilyn vos Savant

but one, door 59 let's say. You have to ask yourself: why did they pick door 59 to leave closed?? Almost certainly because that's where the prize is hidden! Maybe you got really lucky and picked right with the first door at the beginning. But it's way more likely you didn't, and the host had to keep door 59 closed to avoid giving away the game.

### 1.1 Diagramming the Solution

A picture helps clarify things. At first the prize could be behind any of the three doors, with equal probability each way. So we draw a tree with three branches, each labeled with a probability of  $1/3$ . Figure 1.2 shows the result.

Now, which door the host opens may depend on where the prize is, i.e. which branch we're on. If it's behind door C, they won't show you by opening that door. They would have to open door B in this case.

Likewise, if the prize is behind door B, then opening door C is their only option.

Only if the prize is behind door A do they have a choice: open either door B or door C. In that case it's a tossup which door they'll open, so each of those possibilities has a  $1/2$  chance. Check out Figure 1.3.

Now imagine playing the game over and over. A third of the time things will follow the top path; a third of the time they'll follow the middle one; and the remaining third they'll follow one of the two bottom paths.

When things follow the bottom branches, half of those times the host will open door B, and half the time they'll open door C. So one in every six plays will follow the *A-and-Open-B* path. And one in every six plays will follow the *A-and-Open-C* path. See Figure ??.

Now we can understand what happens when the host opens door C. Usually it's because the prize is behind door B. Sometimes they open door C because the prize is behind door A instead. But that's only a sixth of the time, compared to a third of the time where they open door C because the prize is behind door B.

So when you see the host open door C, you should think it's more likely you're on the middle branch, with the prize behind door B. Switch!

### 1.2 Lessons Learned

TREE diagrams are a handy tool for solving probability problems. They also illustrate some central concepts of probability.

Probabilities are numbers assigned to possibilities. In the Monty Hall problem, there are three possibilities for where the prize is: door A, door B, and door C. Each of these possibilities has the same proba-

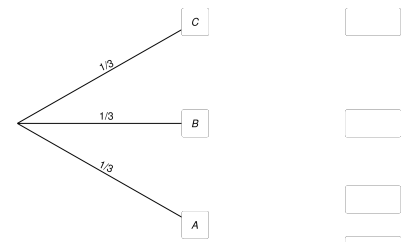


Figure 1.2: First stage of a tree diagram for the Monty Hall problem

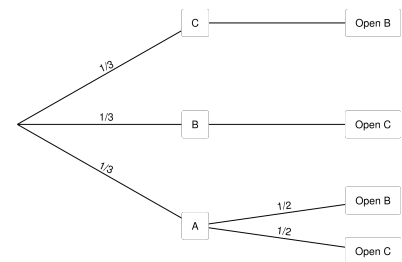


Figure 1.3: Second stage

bility:  $1/3$ .

SOME possibilities are *mutually exclusive*, meaning only one of them can obtain. The prize can't be behind door A and door B, for example. Here are more examples of mutually exclusive possibilities:

- A coin can land heads or tails, but it can't do both on the same toss.
- A card drawn from a standard deck could be either an ace or a queen, but it can't be both.
- The temperature at noon tomorrow could be 20 degrees, or it could be 25 degrees, but it can't be both.

When possibilities are mutually exclusive, their probabilities add up. For example, the initial probability the prize will be behind either door A or door B is  $1/3 + 1/3 = 2/3$ . And the probability a card drawn from a standard deck will be either an ace or a queen is  $4/52 + 4/52 = 8/52 = 2/13$ .

ANOTHER key concept is possibilities that are *exhaustive*. In the Monty Hall problem, the prize has to be behind one of the three doors, so A, B, and C "exhaust" all the possibilities. Here are more examples of exhaustive possibilities:

- A card drawn from a standard deck must be either red or black.
- The temperature at noon tomorrow must be either above zero, below zero, or zero.

IN our tree diagrams, each branch-point always uses a set of possibilities that is *both* exclusive *and* exhaustive. The first split on the three doors covers all the possibilities for where the prize might be, and only one of those possibilities can be the actual location of the prize. Likewise for the second stage of the diagram. On the bottom branch for example, the host must open either door B or door C given the rules, but he will only open one or the other.

When a set of possibilities is both exclusive and exhaustive, it's called a *partition*. A partition "carves up" the space of possibilities into distinct, non-overlapping units.

There can be more than one way to partition the space of possibilities. For example, a randomly drawn playing card could be black or red; it could be a face card or not; and it could be any of the four suits ( $\heartsuit, \diamondsuit, \clubsuit, \spadesuit$ ).

WHEN possibilities form a partition, their probabilities must add up to 1. Initially, the probability the prize will be behind one of the three



Figure 1.4: Three partitions for a card drawn from a standard deck

doors is  $1/3 + 1/3 + 1/3 = 1$ . And the probability that a card drawn from a standard deck at random will be either red or black is  $1/2 + 1/2 = 1$ .

In a way, the fundamental principle of probability is that probabilities over a partition must add up to 1.

TREE diagrams follow a few simple rules based on these concepts. The parts of a tree are called *nodes*, *branches*, and *leaves*: see Figure 1.5.

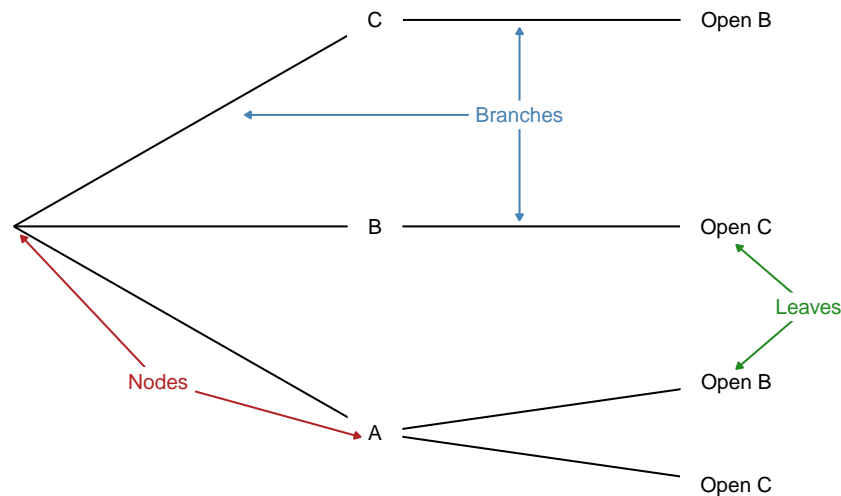


Figure 1.5: The parts of a tree diagram: nodes, branches, and leaves

The rules for a tree are as follows.

*Rule 1.* Each node must branch into a partition. The subpossibilities that emerge from a node must be mutually exclusive. And they must include every way the possibility from which they emerge could obtain.

*Rule 2.* The probabilities emerging from a node must add to 1. If we add up the numbers on the branches immediately coming out of a node, they should add to 1.

*Rule 3.* The probability on a branch is *conditional* on the branches leading up to it.

- Consider the bottom path in the Monty Hall problem. The probability the host will open door C is  $1/2$  there because we're assuming the prize is behind door A.

*Rule 4.* The probability of a leaf is calculated by multiplying across the branches on the path leading to it. This number represents the probability that all possibilities on that path occur.

Notice, Rule 4 is how we got the final probabilities (the numbers in bold) we used to solve the Monty Hall problem.

## Exercises

1. True or false: in the Monty Hall problem, it's essential to the puzzle that the host doesn't want to expose the prize. If they didn't care about giving away the location of the prize, there would be no reason to switch when they open door C.
2. In the version of the Monty Hall problem with a hundred doors, after the host opens every door except door 1 (your door) and door 59, the chance the prize is behind door 59 is:
  - a.  $1/100$
  - b.  $1/99$
  - c.  $1/2$
  - d.  $99/100$

3. Imagine three prisoners, A, B, and C, are condemned to die in the morning. But the king decides to pardon one of them first. He makes his choice at random and communicates it to the guard, who is sworn to secrecy. She can only tell the prisoners that one of them will be released at dawn, she can't say who.

Prisoner A welcomes the news, as he now has a  $1/3$  chance of survival. Hoping to go even further, he says to the guard, "I know you can't tell me whether I am condemned or pardoned. But at least one other prisoner must still be condemned, so can you just name one who is?". The guard tells him that B is still condemned. "Ok", says A, "then it's either me or C who was pardoned. So my chance of survival has gone up to  $1/2$ ".

Is prisoner A's reasoning correct? Use a probability tree to explain why/why not.

4. In a probability tree, each branch point should split into possibilities that are:
  - a. Mutually exclusive.
  - b. Exhaustive.
  - c. Both mutually exclusive and exhaustive.
  - d. None of the above.
5. Suppose you have two urns. The first has two black marbles and two white marbles. The second has three black marbles and one white marble. You are going to flip a fair coin to select one of the urns at random, and then draw one marble at random. What is the chance you will select a black marble?

Hint: draw a probability tree and ask yourself, "if I did this experiment over and over again, how often would I draw a black marble in the long run?"

- a.  $5/8$
- b.  $3/8$
- c.  $1/2$
- d.  $1/4$

## 2 Logic

*I can win an argument on any topic, against any  
opponent. People know this, and steer clear of me  
at parties. Often, as a sign of their great respect,  
they don't even invite me.*  
—Dave Barry

LOGIC is the study of what follows from what. From the information that Tweety is a bird and all birds are animals, it follows that Tweety is an animal. But things aren't always so certain. Can Tweety fly? Most birds can fly, so probably. But Tweety might be a penguin.

*Deductive* logic is the branch of logic that studies what follows with certainty. *Inductive* logic deals with uncertainty, things that only follow with high probability.

This book is about inductive logic and probability. But we need a few concepts from deductive logic to get started.

### 2.1 Validity & Soundness

In deductive logic we study “valid” arguments. An argument is *valid* when the conclusion must be true if the premises are true. Take this example again:

Tweety is a bird.  
All birds are animals.  
Therefore, Tweety is an animal.

The first two lines are called the *premises* of the argument. The last line is called the *conclusion*. In this example, the conclusion must be true if the premises are. So the argument is valid.

Here's another example of a valid argument:

Tweety is taller than Kwazi.  
Kwazi is taller than Peso.  
Therefore, Tweety is taller than Peso.

The argument is valid because it's just not possible for the premises to be true and the conclusion false.

Here's an example of an *invalid* argument:

Tweety is a bird.

Most birds can fly.

Therefore, Tweety can fly.

It's not valid because validity requires the conclusion to follow *necessarily*. If there's any way for the premises to be true yet the conclusion false, the argument doesn't count as valid. And like we said, Tweety might be a penguin.

Valid arguments are interesting because their logic is airtight. If the assumptions of the argument are correct, there's no way to go wrong accepting the conclusion. But what if the assumptions *aren't* correct? Validity isn't everything, we also want our arguments to build on true foundations.

We call an argument *sound* when it is valid *and* all the premises are true:

sound = valid + true premises.

For example, here's a sound argument:

The author of this book is human.

All humans are animals.

Therefore, the author of this book is an animal.

Sound arguments are important because their conclusions are always true. The premises of a sound argument are true by definition. And since it's valid by definition too, that guarantees the conclusion to be true as well.

Yet deductive logic studies validity, not soundness. Why?

Because logicians aren't in the business of determining when the premises of an argument are true. As a logician, I might have no idea who Tweety is, and thus no idea whether Tweety is a bird. I might not even know whether all birds fly, or just some, or even none. That's a job for an ornithologist.

A logician's job is to assess the *logic* of an argument, the connections between its assumptions and its conclusion. So a logician just takes the premises of an argument for granted and asks, how well do those assumptions support the conclusion? That's something you don't need to know any ornithology to study. Or biology, or medicine, or physics, or whatever topic a particular argument concerns.

VALIDITY is a tricky, counterintuitive concept. It's very much a hypothetical notion: it's about whether the conclusion must be true *if* the



premises are true. So when we assess an argument's validity, we ignore what we know about the truth of its premises. We pretend they're true even if they aren't. We even have to ignore what we know about the conclusion.

Instead we suspend what we know about the topic, and just imagine the premises to be true. Then we ask: in this hypothetical scenario, is there any way the conclusion could be false? If there is, the argument is invalid. Otherwise, it's valid.

## 2.2 Propositions

ARGUMENTS are made out of statements, assertions that something is true. In logic we call these statements *propositions*. And we use capital letters of the English alphabet to stand for them. For example, this argument:

If Aegon is a tyrant, then Brandon is a wizard.

Aegon is a tyrant.

Therefore, Brandon is a wizard.

can be summarized like this:

If  $A$ , then  $B$ .

$A$ .

Therefore,  $B$ .

NOT all sentences are propositions. Some are questions, some are commands, some are expressions of worry. For example:

- What time is it?
- Pass the rooster sauce!
- Uh oh.

One way to distinguish propositions from other kinds of sentences is: propositions are capable of being true or false. It wouldn't make sense to respond to someone who asks you what time it is by saying, "what you just said is false!" And you wouldn't respond to someone's request to pass the sauce with "that's true!" Except maybe as a joke.

## 2.3 Visualizing Propositions

WE learned about mutually exclusive propositions in Section 1.2. Two propositions are mutually exclusive when one of them being true means the other must be false. For example:

- $A$ : Confucius was born in the 6th Century A.D.

- *B*: Confucius was born in the 6th Century B.C.

There is no way for both of these propositions to be true, and we can visualize this relationship in a diagram (Figure 2.1).

Each circle represents a proposition. You can think of it as surrounding the possible situations where the proposition would be true. The circles don't overlap because there is no possible situation where both propositions in this example are true.

In contrast, these two propositions are not mutually exclusive:

- Confucius was born in Asia.
- Confucius was born in the 6th Century B.C.

When propositions are not mutually exclusive, we say they are *compatible*. Compatible propositions overlap (Figure 2.2). The region where the circles overlap represents the possible scenarios where both propositions are true (the "*A* & *B* region").

These are called *Euler diagrams*, after the mathematician Leonhard Euler (pronounced *oiler*).<sup>1</sup> You may have seen Venn diagrams before, which are very similar. But in an Euler diagram, the circles don't have to overlap.

SOMETIMES one circle will even contain another circle entirely. Take this example:

- Confucius was born in Asia.
- Confucius was born somewhere.

These propositions aren't just compatible. If the first is true, then the second *must* be true. Imagine an argument with the first proposition as the premise and the second proposition as the conclusion. The argument would be valid:

Confucius was born in Asia.  
Therefore, Confucius was born.

In this case we say that the first proposition *logically entails* the second. In terms of an Euler diagram, the first circle is contained entirely in the second (Figure 2.3). Because there is no possible situation where the first proposition is true yet the second false.

What if an argument has multiple premises? For example:

Zhuangzi was born in the Chinese province of Anhui.  
Zhuoru was born in the Chinese city of Beijing.  
Therefore, both Zhuangzi and Zhuoru were born in China.

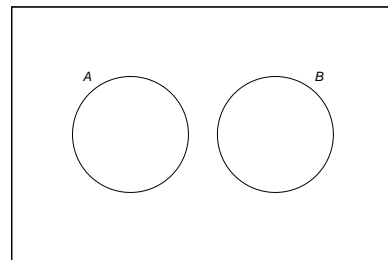


Figure 2.1: Mutually exclusive propositions

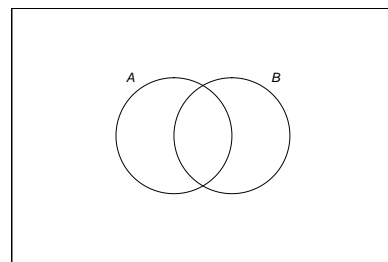


Figure 2.2: Compatible propositions

<sup>1</sup> Leonhard Euler lived from 1707 to 1783. You may have encountered some of his work before if you've worked with logarithms or taken calculus.

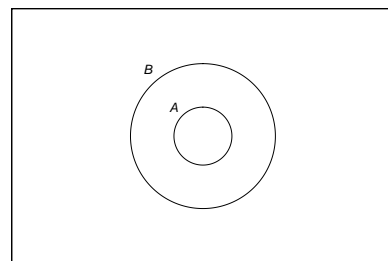


Figure 2.3: Logical entailment

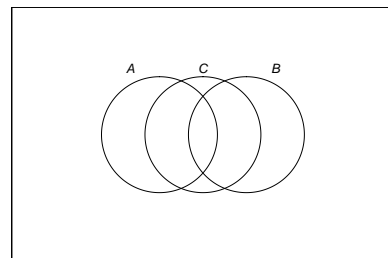


Figure 2.4: A valid argument with two premises

This argument is valid, and the diagram might look like Figure 2.4. Notice how the  $A$  &  $B$  region lies entirely within the  $C$  circle. This reflects the argument's validity: there is no way for the first two propositions to be true and the last one false.

In contrast, an invalid argument would have a diagram like Figure 2.5. This diagram allows for the possibility that  $A$  and  $B$  are both true yet  $C$  is false; part of the  $A$  &  $B$  region falls outside the  $C$  circle.

## 2.4 Strength

INDUCTIVE logic studies arguments that aren't necessarily valid, but still "strong". A **strong** argument is one where the conclusion is highly probable, if the premises are true. For example:

The sun has risen every day so far.  
Therefore, the sun will rise again tomorrow.

This argument isn't valid, because it's possible the conclusion is false even though the premise is true. Maybe the sun will explode in the night for some surprising reason. Or maybe the earth's rotation will be stopped by alien forces.

These possibilities aren't very likely, of course. So the argument is strong, even though it's not strictly valid. The premise gives us very good reason to believe the conclusion, just not a 100% guarantee.

In terms of an Euler diagram then, the premise circle isn't contained entirely within the conclusion circle (Figure 2.6). We have to leave some room for the possibility that the premise is true and the conclusion false. But we can still convey that this possibility has only a very slight chance of being true, by making it slim.

We could also label the *A-but-not-B* region with a small number, if we knew exactly how unlikely this possibility was.

STRENGTH comes in degrees. An argument's premises can make the conclusion somewhat likely, very likely, almost certain, or perfectly certain. So arguments range from weak, to somewhat strong, to very strong, etc.

Strength differs from validity here, since validity is all-or-nothing. If there is any possible way for the premises to be true and the conclusion false, the argument is invalid—no matter how remote or bizarre that possibility is.

Notice though, valid arguments are strong by definition. Since it's impossible for a valid argument's conclusion to be false if the premises are true, the premises make the conclusion 100% probable. A valid argument is the strongest possible argument.

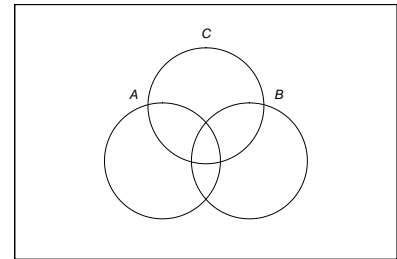


Figure 2.5: An invalid argument with two premises

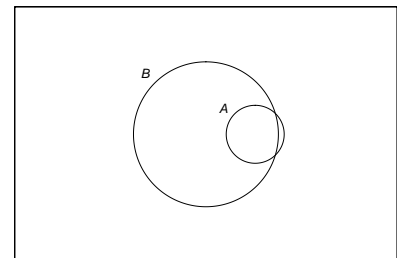


Figure 2.6: A strong argument with premise  $A$  and conclusion  $B$



Figure 2.7: Pierre Simone Laplace (1749–1827) developed a formula for calculating the probability the sun will rise tomorrow. We'll learn how to do similar calculations in the coming chapters.

## 2.5 *Forms of Inductive Argument*

WHAT kinds of strong arguments are there, and how strong are they? That's what the rest of this book is about, in a way. But we can start by identifying some common forms of inductive argument right now.

GENERALIZING from observed instances is one extremely common form of argument:

Every raven I have ever seen has been black.  
Therefore, all ravens are black.

Arguments of this kind are usually stronger the more instances you observe. If you've only ever seen two ravens, this argument won't be very compelling. But if you've seen thousands, then it's much stronger.

It also helps to observe different kinds of instances. If you've only observed ravens in your city or town, then even the thousands you've seen won't count for much. Maybe the raven population in your area is unusual, and ravens on the other side of the world are all different colours.

GOING in the opposite direction, we can use what we know about a general population to draw conclusions about particular instances. We saw an example of this earlier:

Most birds can fly.  
Tweety is a bird.  
Therefore, Tweety can fly.

Again, the strength of the inference depends on the details. If "most birds" means 99%, the argument is quite strong. If "most" only means 80%, then it's not so strong. (Usually "most" just means more than 50%.)

It also helps to know that Tweety is similar to the birds that can fly, and different from the ones that can't. If we know that Tweety is small and has feathers, that makes the argument stronger. If instead we know that Tweety is large, and coloured black and white, that makes the argument weaker. It suggests Tweety is a penguin.

INFERENCE to the best explanation is another common form of argument, quite different from the previous two. Here's an example:

My car won't start and the gas gauge reads 'empty'.  
Therefore, my car is out of gas.

An empty tank would explain the symptoms described in the premise, so the premise makes the conclusion plausible. There could be other

possible explanations, of course. Maybe the engine and the gas gauge both just happened to break at the same time. But that would be quite a coincidence, so this explanation isn't as good.

What makes one explanation better than another? That turns out to be a very hard question, and there is no generally accepted answer. We'll come back to this issue later, once we have a better grip on the basics of probability.

### *Exercises*

1. For each of the following arguments, say whether it is valid or invalid.
  - a. All cats have whiskers.  
Simba has whiskers.  
Therefore, Simba is a cat.
  - b. Ada Lovelace wrote the world's first computer program.  
Ada Lovelace was Lord Byron's daughter.  
Therefore, the first computer program was written by Lord Byron's daughter.
  - c. All Canadian residents are Russian citizens.  
Vladimir Putin is a Canadian resident.  
Therefore, Vladimir Putin is a Russian citizen.
  - d. Manitoba is located in either Saskatchewan, Ontario, or Quebec.  
Manitoba is not located in Saskatchewan.  
Manitoba is not located in Ontario.  
Therefore, Manitoba is located in Quebec.
  - e. If snow is black then pigs can fly.  
Snow is not black.  
Therefore, pigs cannot fly.
  - f. Either the moon is made of green cheese or pigs can fly.  
Pigs can't fly.  
Therefore the moon is made of green cheese.
2. For each pair of propositions, say whether they are mutually exclusive or compatible.
  - a. Regarding the roll of an ordinary die:
    - The die will land on an even number.
    - The die will land either 4 or 5.
  - b. Regarding the unemployment rate in your country tomorrow:

- The unemployment rate will be at least 5%.
  - The unemployment rate will be exactly 5%.
- c. Regarding a party tomorrow:
- Ani will be there and so will her sister PJ.
  - PJ will not be there.
3. True or false? If  $A$  and  $B$  are mutually exclusive, and  $B$  and  $C$  are also mutually exclusive, then  $A$  and  $C$  are mutually exclusive.
4. True or false? If  $A$  and  $B$  are mutually exclusive, then  $A$  logically entails that  $B$  is false.
5. True or false? It is possible for  $A$  to logically entail  $B$  even though the reverse does not hold (i.e. even though  $B$  does not logically entail  $A$ ).
6. Create your own example of each of the three types of inductive argument described in this chapter:
- a. Generalizing from Observed Instances
  - b. Inferring an Instance from a Generalization
  - c. Inference to the Best Explanation

## 3 Truth Tables

IN this chapter we'll introduce the last few concepts we need from deductive logic, and we'll learn a useful technique in the process: truth tables.

### 3.1 Connectives

COMPLEX propositions can be built up out of other, simpler propositions:

- Aegon is a tyrant **and** Brandon is a wizard.
- **Either** Aegon is a tyrant **or** Brandon is a wizard.
- **It's not true that** Aegon is a tyrant.

Here we've used two simple propositions to build up longer, more complex ones using the terms *and*, *either/or*, and *it's not true that*. Such terms are called *connectives*.

The three connectives just listed are the only ones we'll need in this book. Each has a name and a shorthand symbol:

Name	English	Symbol	Example
conjunction	and	$\&$	$A \& B$
disjunction	either/or	$\vee$	$A \vee B$
negation	it's not true that	$\sim$	$\sim A$

Notice, we call *it's not true that* a connective even though it doesn't actually connect two propositions together.

Here are some more examples of complex propositions:

- $F \& \sim G$ : Florida is warm and Geneva is not.
- $\sim J \vee \sim K$ : Either Jing won't come to the party or Kamal won't come.

SOMETIMES we also need parentheses, to avoid ambiguity. Consider an example from arithmetic:

$$4 \div 2 \times 2 = 1.$$

Is this equation true? That depends on what you mean. Does the division operation come first, or the multiplication? So we use parentheses to clarify:  $4 \div (2 \times 2) = 1$ , but  $(4 \div 2) \times 2 = 4$ .

In logic we use parentheses to prevent ambiguity similarly. Consider:

$$A \vee B \& C.$$

This proposition is ambiguous, it has two interpretations. In English we can distinguish them with a comma:

- Either Aegon is a tyrant or Brandon is a wizard, and Cerci is the queen.
- Either Aegon is a tyrant, or Brandon is a wizard and Cerci is the queen.

Notice how these statements make different claims. The first takes a definite stand on Cerci: she is the queen. It only leaves open the question whether Aegon is a tyrant or Brandon is a wizard. Whereas the second statement takes no definite stand on any of our three characters. Maybe Aegon is a tyrant, maybe not. Maybe Brandon is a wizard and Cerci is the queen, maybe not.

In logic we use parentheses to clarify which interpretation we mean:

- $(A \vee B) \& C$ .
- $A \vee (B \& C)$ .

Notice how the first statement is primarily an  $\&$  statement. It uses  $\&$  to combine the simpler statements  $C$  and  $A \vee B$  together. Whereas the second statement is primarily a  $\vee$  statement. It uses  $\vee$  to combine  $A$  with  $B \& C$ .

We call the last connective used to build up the statement the *main connective*.

- $(A \vee B) \& C$ : main connective is  $\&$ .
- $A \vee (B \& C)$ : main connective is  $\vee$ .

Two more examples:

- $\sim(A \vee B)$ : main connective is  $\sim$ .
- $\sim A \vee B$ : main connective is  $\vee$ .

Technically, the last example should have parentheses to prevent ambiguity, like so:  $(\sim A) \vee B$ . But things get cluttered and hard to read if we add parentheses around every negation. So we have a special understanding for  $\sim$  in order to keep things tidy.

- ❏ The negation symbol  $\sim$  only applies to the proposition immediately following it.

So in the proposition  $\sim A \vee B$ , the  $\sim$  only applies to  $A$ . And in  $\sim(A \& B) \vee C$ , it only applies to  $A \& B$ .

This special understanding for  $\sim$  mirrors the one for minus signs in arithmetic.



### 3.2 Truth Tables

THE truth of a complex proposition built using our three connectives depends on the truth of its components. For example,  $\sim A$  is false if  $A$  is true, and it's true if  $A$  is false:

Table 3.2: Truth table for  $\sim$

$A$	$\sim A$
T	F
F	T

Slightly more complicated is the rule for  $\&$ :

Table 3.3: Truth table for  $\&$

$A$	$B$	$A \& B$
T	T	T
T	F	F
F	T	F
F	F	F

There are four rows now because  $\&$  combines two propositions  $A$  and  $B$  together to make the more complex proposition  $A \& B$ . Since each of those propositions could be either true or false, there are  $2 \times 2 = 4$  possible situations to consider.

Notice that in only one of these situations is  $A \& B$  true, namely the first row where both  $A$  and  $B$  are true.

The truth table for  $\vee$  ("either/or") is a little more surprising:

Table 3.4: Truth table for  $\vee$

$A$	$B$	$A \vee B$
T	T	T
T	F	T
F	T	T
F	F	F

Now the complex proposition is always true, except in one case: the last row where  $A$  and  $B$  are both false. It makes sense that  $A \vee B$  is false when both sides are false. But why is it true when both sides are true? Doesn't "Either  $A$  or  $B$ " mean that *just one* of these is true?

Sometimes it does have that meaning. But sometimes it means "Ei-

ther  $A$  or  $B$ , *or both*". Consider this exchange:

X: What are you doing tomorrow night?

Y: I'm either going to a friend's house or out to a club. I might even do both, if there's time.

Person Y isn't necessarily changing their mind here. They could just be clarifying: they're doing at least one of these things, possibly even both of them.

Although it's common to use "either/or" in English to mean *just* one or the other, in logic we use the more permissive reading. So  $A \vee B$  means *either A, or B, or both*.

We can always convey the stricter way of meaning "either/or" with a more complex construction:

$$(A \vee B) \ \& \ \sim(A \ \& \ B).$$

That says:

Either  $A$  or  $B$  is true, and it's not the case that both  $A$  and  $B$  are true.

Which is just a very explicit way of saying: either one or the other, but not both.

We can even verify that the complex construction captures the meaning we want using a truth table. We start with an empty table, where the header lists all the formulas we use to build up to the final, complex one we're interested in:

---

$A$	$B$	$A \vee B$	$A \& B$	$\sim(A \ \& \ B)$	$(A \vee B) \ \& \ \sim(A \ \& \ B)$
-----	-----	------------	----------	--------------------	--------------------------------------

---

Then we fill in the possible truth values for the simplest propositions,  $A$  and  $B$ :

---

$A$	$B$	$A \vee B$	$A \& B$	$\sim(A \ \& \ B)$	$(A \vee B) \ \& \ \sim(A \ \& \ B)$
T	T				
T	F				
F	T				
F	F				

---

Next we consult the truth tables above for  $\&$  and  $\vee$  to fill in the columns at the next level of complexity:

$A$	$B$	$A \vee B$	$A \& B$	$\sim(A \& B)$	$(A \vee B) \& \sim(A \& B)$
T	T	T	T		
T	F	T	F		
F	T	T	F		
F	F	F	F		

Then move up to the next level of complexity. To fill in the column for  $\sim(A \& B)$ , we consult the column for  $A \& B$  and apply the rules from the table for  $\sim$ :

$A$	$B$	$A \vee B$	$A \& B$	$\sim(A \& B)$	$(A \vee B) \& \sim(A \& B)$
T	T	T	T	F	
T	F	T	F	T	
F	T	T	F	T	
F	F	F	F	T	

Finally, we consult the columns for  $A \vee B$  and  $\sim(A \& B)$ , and the table for  $\&$ , to fill in the column for  $(A \vee B) \& \sim(A \& B)$ :

$A$	$B$	$A \vee B$	$A \& B$	$\sim(A \& B)$	$(A \vee B) \& \sim(A \& B)$
T	T	T	T	F	F
T	F	T	F	T	T
F	T	T	F	T	T
F	F	F	F	T	F

Complex constructions like this are difficult at first, but don't worry. With practice they quickly become routine.

### 3.3 Logical Truths & Contradictions

SOME propositions come out true in every row of the truth table. Consider  $A \vee \sim A$  for example:

$A$	$\sim A$	$A \vee \sim A$
T	F	T
F	T	T

Such propositions are especially interesting because they *must* be true. Their truth is guaranteed, just as a matter of logic. So we call them *logical truths*.

The other side of this coin is propositions that are false in every row

of the truth table, like  $A \& \sim A$ :

$A$	$\sim A$	$A \& \sim A$
T	F	F
F	T	F

These propositions are called *contradictions*.

Notice that the negation of a contradiction is a logical truth. For example, consider the truth table for  $\sim(A \& \sim A)$ :

$A$	$\sim A$	$A \& \sim A$	$\sim(A \& \sim A)$
T	F	F	T
F	T	F	T

### 3.4 Mutual Exclusivity & Truth Tables

TRUTH tables can be used to establish that two propositions are mutually exclusive. A very simple example is the propositions  $A$  and  $\sim A$ :

$A$	$\sim A$
T	F
F	T

There is no row in the table where both propositions are true. And if two propositions can't both be true, they are mutually exclusive by definition.

A slightly more complex example is the propositions  $A \vee B$  and  $\sim A \& \sim B$ :

$A$	$B$	$\sim A$	$\sim B$	$A \vee B$	$\sim A \& \sim B$
T	T	F	F	T	F
T	F	F	T	T	F
F	T	T	F	T	F
F	F	T	T	F	T

Again there's no row where  $A \vee B$  and  $\sim A \& \sim B$  are both true. So they are mutually exclusive.

### 3.5 Entailment & Equivalence

TRUTH tables can also be used to establish that an argument is valid. Here's a very simple example:

$A \& B$ .  
Therefore,  $A$ .

Obviously it's not possible for the premise to be true and the conclusion false, so the argument is valid (if a bit silly). Accordingly, there is no line of the truth table where  $A \& B$  comes out true, yet  $A$  comes out false:

$A$	$B$	$A \& B$
T	T	T
T	F	F
F	T	F
F	F	F

The only line where  $A \& B$  comes out true is the first one. And on that line  $A$  is true too. So the argument from  $A \& B$  to  $A$  is valid.

One more example:

$A \vee B$ .  
 $\sim A$ .  
Therefore,  $B$ .

This argument is valid because the first premise says that at least one of the two propositions  $A$  and  $B$  must be true, and the second line says it's not  $A$ . So it must be  $B$  that's true, as the conclusion asserts. And once again there is no line of the truth table where both  $A \vee B$  and  $\sim A$  are true, yet  $B$  is false:

$A$	$B$	$\sim A$	$A \vee B$
T	T	F	T
T	F	F	T
F	T	T	T
F	F	T	F

The only line where both  $A \vee B$  and  $\sim A$  are true is the third row, and  $B$  is true on that row. So once again the truth table tells us this argument is valid.

In the previous chapter we introduced the concept of logical entailment.  $A$  logically entails  $B$  when it's impossible for  $A$  to be true and  $B$  false. When one proposition entails another, there is no line of the

truth table where the first proposition is true and the second is false.

SOMETIMES entailment goes in both directions: the first proposition entails the second *and the second entails the first*. For example, not only does  $A \& B$  entail  $B \& A$ , but also  $B \& A$  entails  $A \& B$ .

We say such propositions are *logically equivalent*. In terms of truth tables, their columns match perfectly, they are identical copies of T's and F's.

$A$	$B$	$A \& B$	$B \& A$
T	T	T	T
T	F	F	F
F	T	F	F
F	F	F	F

A more complex example is the propositions  $\sim(A \vee B)$  and  $\sim A \& \sim B$ :

$A$	$B$	$\sim A$	$\sim B$	$A \vee B$	$\sim(A \vee B)$	$\sim A \& \sim B$
T	T	F	F	T	F	F
T	F	F	T	T	F	F
F	T	T	F	T	F	F
F	F	T	T	F	T	T

Here again the columns under these two propositions are identical.

### 3.6 Summary

CONNECTIVES can be used to build more complex propositions, like  $A \& B$  or  $A \vee \sim B$ . We introduced three connectives:

- $\sim A$  means it's not true that  $A$ .
- $A \& B$  means both  $A$  and  $B$  are true.
- $A \vee B$  means either  $A$  is true, or  $B$  is true, or *both are true*.

In a complex proposition, the main connective is the last one used to build it up from simpler components. In  $A \vee \sim B$  the main connective is the  $\vee$ .

An argument's validity can be established with a truth table, if there's no row where all the premises have a T and yet the conclusion has an F.

Truth tables can also be used to establish that two propositions are mutually exclusive, if there is no row of the table where both propositions have a T.

Logically equivalent propositions entail one another. When two propositions have identical columns in a truth table, they are logically equivalent.

### Exercises

- Using the following abbreviations:

$A$  = Asha loves Cerci,

$B$  = Balon loves Cerci,

translate each of the following into logicese (e.g.  $\sim A \vee B$ ).

- Asha doesn't love Cerci.
  - Asha loves Cerci and Balon loves Cerci.
  - Asha loves Cerci but Balon does not.
  - Neither Asha nor Balon loves Cerci.
- For each pair of propositions, use a truth table to determine whether they are mutually exclusive.
    - $A \& B$  and  $A \& \sim B$ .
    - $A$  and  $\sim B$ .
    - $A \vee \sim A$  and  $A \& \sim A$ .
  - For each pair of propositions, use a truth table to determine whether they are logically equivalent.
    - $\sim A \vee B$  and  $\sim(A \& \sim B)$ .
    - $A$  and  $(A \& B) \vee (A \& \sim B)$ .
    - $A$  and  $(B \& A) \vee (B \& \sim A)$ .
  - The proposition  $A \vee (B \& C)$  features three simple propositions, so its truth table has 8 rows. Fill in the rest of the table:

$A$	$B$	$C$	$B \& C$	$A \vee (B \& C)$
T	T	T		
T	T	F		
T	F	T		
T	F	F		
F	T	T		
F	T	F		
F	F	T		
F	F	F		

5. Use a truth table to determine whether the propositions  $A \vee (B \& C)$  and  $(A \vee B) \& (A \vee C)$  are equivalent.



## 4     *The Grue Paradox*

IN Section 2.5 we noticed that many inductive arguments follow this pattern:

All observed instances of  $X$  have been  $Y$ .  
Therefore, all instances of  $X$  are  $Y$ .

All observed ravens are black, so we expect all ravens to be black. All observed emeralds are green, so we expect all emeralds to be green. And so on.

So it seems like a fundamental principle of scientific inquiry that we expect the unobserved to resemble the observed. Philosophers call this *The Principle of Induction*.

### *A Gruesome Concept*

BUT in the 1940s, Nelson Goodman discovered a problem with the Principle of Induction. To illustrate the problem he invented a very curious concept, *grue*.

There are two ways for an object to be *grue*. Some green things are *grue*, but it depends on when we first encounter them. If our first observation of a green object happens before the year 2050, then it's *grue*. So the Statue of Liberty is *grue*: it's a green object that was first observed before the year 2050 (*long* before).

But if our first encounter with a green object happens in the year 2050 or later, then it's not *grue*. The same goes if we never observe it. Objects on the far side of the universe that we'll never see, or buried deep underground, are not *grue*.

There is a second way for an object to be *grue*: some blue objects are *grue*. Not the ones observed before 2050, though. Instead it's the ones that *aren't* observed before 2050. If a blue object is observed for the first time *after* 2049, or it's never observed at all, then it's *grue*. So blue sapphires that won't be mined before the year 2050 are *grue*, for example.

As usual, it helps to have a diagram:

Here is the formal definition of *grue*:



Figure 4.1: Nelson Goodman (1906–1988) discovered the *grue* paradox in the 1940s and '50s.

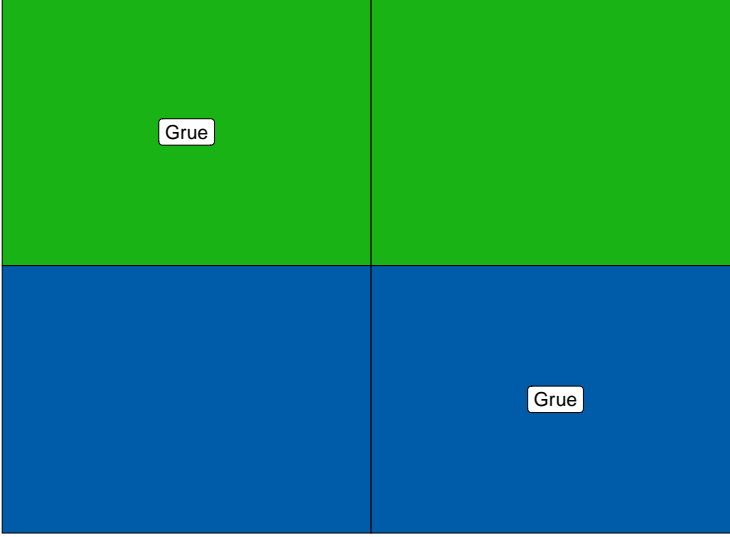
Green		
Blue		
	Observed Before 2050	Not Observed Before 2050

Figure 4.2: The definition of grue

*Grue* An object is *grue* if either (a) it is green and first observed before the year 2050, or (b) it's blue and not observed before 2050.

To test your understanding, see if you can explain why each of the following are examples of grue things: the \$20 bill in my pocket, Kermit the Frog, the first sapphire to be mined in 2050, and blue planets on the far side of the universe.

Then see if you can explain why these things aren't grue: fire engines, the Star of India, and the first \$20 bill to be printed in 2050.

Once you've got all those down, try this question: do grue objects change colour in the year 2050? It's a common mistake to think they do.

But no, grue objects don't change colour. The Statue of Liberty is green and it always will be (let's assume). So it's grue, and always will be, because it's a green thing that was first observed before the year 2050. Part (a) of the definition of grue guarantees that.

The only way time comes into it is in determining which green things are grue, and which blue things. If a green thing is first observed before 2050, then it's grue, ever and always. Likewise if a blue thing is *not* first observed before 2050. Then it's grue—and it always has been!

### *The Paradox*

Now ask yourself, have you ever seen a grue emerald? You probably have. In fact, every emerald everyone's ever seen has been grue.

Why? Because they're all green, and they've all been observed before the year 2050. So they're all *grue* the first way—they all satisfy part (a) of the definition. (Notice it's an either/or definition, so you only have to satisfy one of the two parts to be *grue*.)

So all the emeralds we've ever seen have been *grue*. Let's apply the Principle of Induction then:

All observed emeralds have been *grue*.  
Therefore *all* emeralds are *grue*.

But if all emeralds are *grue*, then the first emeralds to be mined in 2050 will be *grue*. And that means they'll be blue! Because they won't have been observed before 2050, so the only way for them to be *grue* is to be blue.

We've reached the absurd conclusion that there are blue emeralds out there, just waiting to be pulled out of the earth. Something has gone off the rails here, but what?

Here's another way to put the challenge. We have two "patterns" in our observed data. The emeralds we've seen are uniformly green, but they're also uniformly *grue*. We can't project both these patterns into the future, though. They'll contradict each other starting in 2050.

Now, obviously, common sense says the green pattern is the real one. The *grue* "pattern" is bogus, and no one but a philosopher would even bother thinking about it. But *why* is it bogus? What's so special about green?

Apparently the Principle of Induction has a huge hole in it! It says to extrapolate from observed patterns. But *which* patterns?

### *Grue & Artificial Intelligence*

Patterns are cheap, as any data scientist will tell you. Given a bunch of data points in an  $xy$ -plane, there are lots of ways to connect the dots. Even if they all lie on a straight line, you could draw an oscillating curve that passes through each point (Figure 4.3). You can probably think of even sillier curves that will fit all the points.

Designing a computer program that will know which patterns to use and which to ignore is a big part of what machine learning experts do. And it's one reason humans are still essential to designing artificial intelligence. Thanks to our experience, and our genetic inheritance, we have *lots* of information about which patterns are likely to continue, and which are bogus, like *grue*.

But how do we pass all that wisdom on to the machines? How do we teach them the difference between green and *grue*, so that they can take it from here and we can all go on permanent vacation?

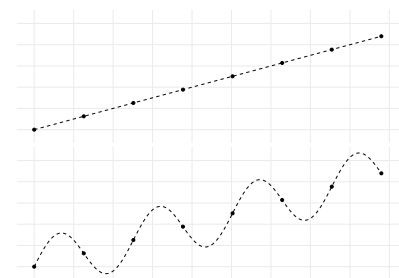


Figure 4.3: The same set of points interpreted two different ways

## Disjunctivitis

HERE'S ONE very natural answer. The problem with *grue* is that it's *disjunctive*: it's defined using either/or. It suffers from what we might call "disjunctivitis".

But the beauty of Goodman's puzzle is the neat way it exposes the flaw in this answer. It allows us to make 'green' the disjunctive concept instead! How? We start by building *grue* a friend, a concept to fill in the missing spaces in our original diagram. We'll call it *bleen*:

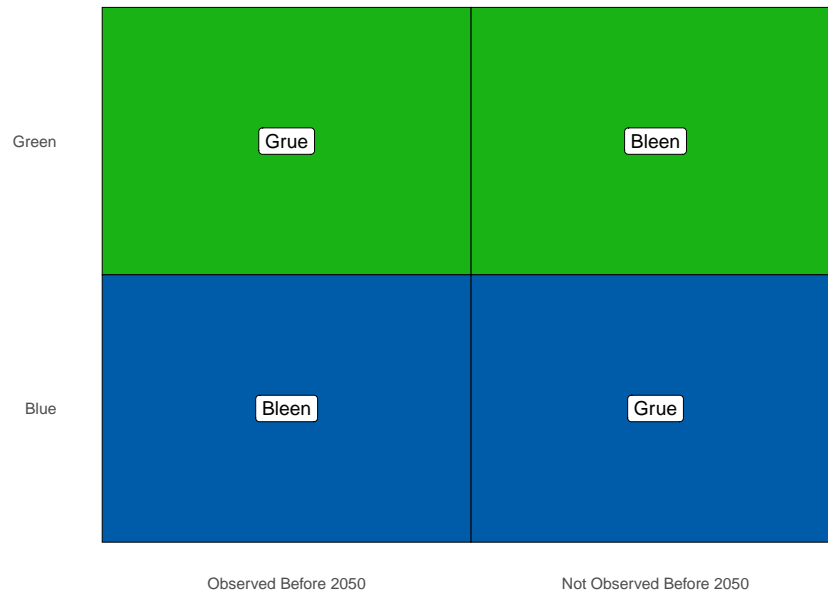


Figure 4.4: Defining bleen, a counterpart to grue

Now we can define green in terms of grue and bleen:

*Green* An object is *green* if either (a) it's grue and first observed before the year 2050, or (b) it's bleen and not observed before 2050.

Now maybe you're thinking: you *could* define green that way, but that's not how it's *actually* defined. In reality, we already understand the concept of green, and we have to learn the concept of grue from its disjunctive definition.

The problem is, that's just a fact about *us humans*, not about the concepts grue and green. That's just the way we *homo sapiens* happen to be built (or maybe socialized, or both).

Some bizarre species of alien could grow up thinking in grue/bleen terms instead. And when they landed on Earth, we'd have to explain our green/blue language to them using an either/or definition. Then *they* would be looking at *us* thinking: you guys have a very weird, disjunctive way of thinking!

What could we say to them to establish the superiority of our way of thinking? It's been more than 70 years since Goodman first posed this question. Yet no answer has emerged as the clear and decisively correct one.

### *Time Dependence*

ANOTHER natural answer to Goodman's challenge is to say that grue is defective because it's time-dependent. It means different things depending on the time an object is first observed.

But the same reversal of fortunes that toppled the "disjunctivitis" diagnosis happens here. We can define green in terms of grue and bleen. And when we do, it's green that's the time-dependent concept, not grue.

So we're left in the same spot. We need some way of showing that the "true" order of definition is the one we're used to. By what criterion can we say that green is more fundamental, more basic, than grue?

### *The Moral*

GOODMAN'S PUZZLE may seem just cute at first, a mere curiosity. But it is actually quite profound.

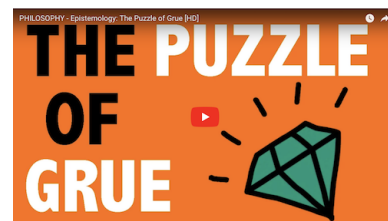
In a way, the central question of this book is: what is the logic of science? What are the correct rules of scientific reasoning?

The laws of probability seem like a good place to start. But that path led us to a dead end at the problem of priors in Chapter ???. Perhaps then we could start with the Principle of Induction instead? But then we end up in another dead end, stopped by the grue paradox.

Just as Bertrand's paradox stops us from using the Principle of Indifference to answer the problem of priors, Goodman's paradox stops us from using the Principle of Induction.

There is even more to the similarity between these two paradoxes. Both paradoxes are problems of "language dependence". Depending on what language we work in, rules like the Principle of Indifference and the Principle of Induction give different recommendations. If we apply the Principle of Indifference to length, we get one prior probability; if we apply it to area, we get another. If we apply the Principle of Induction to green, we expect all emeralds to be green; if we apply it to grue, we expect some to be blue.

To this day, we do not have an answer to the question: which language is the right one to use, and why?



For another explanation of the grue puzzle, check out this excellent Wi-Phi video.



## 5 *The Problem of Induction*

MANY inductive arguments work by projecting an observed pattern onto as-yet unobserved instances. All the ravens we've observed have been black, so all ravens are. All the emeralds we've seen have been green, so all emeralds are.

The assumption that the unobserved will resemble the observed seems to be central to induction. Philosophers call this assumption the *Principle of Induction*.<sup>1</sup> But what justifies this assumption? Do we have any reason to think the parts of reality we've observed so far are a good representative of the parts we haven't seen yet?

Actually there are strong reasons to doubt whether this assumption can be justified. It may be impossible to give any good argument for expecting the unobserved to resemble the observed.

### *The Dilemma*

WE observed previously that there are two kinds of argument, inductive and deductive. Some arguments establish their conclusions necessarily, others only support them with high probability. If there is an argument for the Principle of Induction, it must be one of these two kinds. Let's consider each in turn.

Could we give an inductive argument for the Principle of Induction? At first it seems we could. Scientists have been using inductive reasoning for millenia, often with great success. Indeed, it seems humans, and other creatures too, have relied on it for much longer, and could not have survived without it. So the Principle of Induction has a very strong track record. Isn't that a good argument for believing it's correct?

No, because the argument is circular. It uses the Principle of Induction to justify believing in the Principle of Induction. Consider that the argument we are attempting looks like this:

The principle has worked well when we've used it in the past.

Therefore it will work well in future instances.

This is an inductive argument, an argument from observed instances to

<sup>1</sup> See Section 2.5 and Appendix 4 for previous discussions of the Principle of Induction.



Figure 5.1: David Hume (1711–1776) raised the problem of induction in 1739. Our presentation of it here is somewhat modernized from his original argument.

ones as yet unobserved. So, under the hood, it appeals to the Principle of Induction. But that's exactly the conclusion we're trying to establish. And one can't use a principle to justify itself.

What about our second option: could a deductive argument establish the Principle of Induction? Well, by definition, a deductive argument establishes its conclusion with necessity. Is it necessary that the unobserved will be like the observed? It doesn't look like it. It seems perfectly possible that tomorrow the world will go haywire, randomly switching from pattern to pattern, or even to no pattern at all.

Maybe tomorrow the sun will fail to rise. Maybe gravity will push apart instead of pull together, and all the other laws of physics will reverse too. And just as soon as we get used to those patterns and start expecting them to continue, another pattern will arise. And then another. And then, just as we give up and come to have no expectation at all about what will come next, everything will return to normal. Until we get comfortable and everything changes again.

Thankfully, our universe hasn't been so mischievous. We get surprised now and again, but for the most part inductive reasoning is pretty reliable, when we do it carefully. But we're lucky in this respect, is the point.

Nature *could* have been mischievous, totally unpredictable. It is not a necessary truth that the unobserved must resemble the observed. And so it seems there cannot be a deductive argument for the Principle of Induction. Because such an argument would establish the principle as a necessary truth.

### *The Problem of Induction vs. the Grue Paradox*

If you read Appendix 4, you know of another famous problem with the Principle of Induction: the grue paradox. (If you haven't read that chapter, you might want to skip this section.)

The two problems are quite different, but it's easy to get them confused. The problem we're discussing here is about justifying the Principle of Induction. Is there any reason to believe it's true? Whereas the grue paradox points out that we don't even really know what the principle says, in a way. It says that what we've observed is a good indicator of what we haven't yet observed. But in what respects? Will unobserved emeralds be green, or will they be grue?

So the challenge posed by grue is to spell out, precisely, what the Principle of Induction says. But even if we can meet that challenge, this challenge will remain. Why should we believe the principle, once it's been spelled out? Neither a deductive argument nor an inductive argument seems possible.



### *Probability Theory to the Rescue?*

THE Problem of Induction is centuries old. Isn't it out of date? Hasn't the modern, mathematical theory of probability solved the problem for us?

Not at all, unfortunately. One thing we learn in this book is that the laws of probability are very weak in a way. They don't tell us much, without us first telling them what the prior probabilities are. And as we've seen over and over again throughout Part III, the problem of priors is very much unsolved.

For example, suppose we're going to flip a mystery coin five times. We don't know whether the coin is fair or biased, but we hope to have some idea after a few flips.

Now suppose we get through the first four flips and they've all been tails. The Principle of Induction says we should expect the next flip to be tails too. At least, that outcome should now be more probable.

Do the laws of probability agree? Well, we need to calculate the quantity:

$$Pr(T_5 \mid T_1 \& T_2 \& T_3 \& T_4).$$

The definition of conditional probability tell us:

$$Pr(T_5 \mid T_1 \& T_2 \& T_3 \& T_4) = \frac{Pr(T_1 \& T_2 \& T_3 \& T_4 \& T_5)}{Pr(T_1 \& T_2 \& T_3 \& T_4)}.$$

But the laws of probability don't tell us what numbers go in the numerator and the denominator.

The numbers have to be between 0 and 1. And we have to be sure mutually exclusive propositions have probabilities that add up, according to the Additivity rule. But that still leaves things wide open.

For example, we could finish the calculation this way:

$$\begin{aligned} Pr(T_5 \mid T_1 \& T_2 \& T_3 \& T_4) &= \frac{Pr(T_1 \& T_2 \& T_3 \& T_4 \& T_5)}{Pr(T_1 \& T_2 \& T_3 \& T_4)} \\ &= \frac{1/32}{1/16} \\ &= 1/2. \end{aligned}$$

Or we could finish it this way:

$$\begin{aligned} Pr(T_5 \mid T_1 \& T_2 \& T_3 \& T_4) &= \frac{Pr(T_1 \& T_2 \& T_3 \& T_4 \& T_5)}{Pr(T_1 \& T_2 \& T_3 \& T_4)} \\ &= \frac{5/20}{6/20} \\ &= 5/6. \end{aligned}$$

We could even do this:

$$\begin{aligned} Pr(T_5 \mid T_1 \& T_2 \& T_3 \& T_4) &= \frac{Pr(T_1 \& T_2 \& T_3 \& T_4 \& T_5)}{Pr(T_1 \& T_2 \& T_3 \& T_4)} \\ &= \frac{0}{1} \\ &= 0. \end{aligned}$$

All these options result from different choices of prior probabilities. And the laws of probability don't tell us what prior probabilities we must choose, as we learned in Part III.

So the laws of probability don't by themselves tell us what to expect. It could be undecided, with heads and tails equally probable on the final toss (1/2). Or the pattern of tails could continue into the future with high probability (5/6). There could even be no chance of the pattern continuing (0).

The laws of probability only tell us what to expect once we've specified the necessary prior probabilities. The problem of induction challenges us to justify one choice of prior probabilities over the alternatives.

In the 280 years since this challenge was first raised by David Hume, no answer has gained general acceptance.