What is Probability?

Contents

Critical Thinker Academy

Introduction

"But for us, probability is the very guide of life."

- (Bishop Joseph Butler)

Probability concepts are important in everyday reasoning about chance and uncertainty, in the formal methods of inductive logic and scientific reasoning, and in philosophical arguments of many different kinds.

This course focuses on the MEANING of probability, how to understand the different things that people mean, or what scientists or mathematicians mean, when they use expressions like "the odds of getting a 2 on a dice roll is 1/6", or "the probability of precipitation is 60%", or "the probability of the atom decaying in one hour is 50%".

There are, in fact, several different views of how such language should be interpreted. Becoming familiar with these views will help you to think more clearly and critically about situations where probability concepts arise.

Part 1: Introduction

1. Probability: Why Learn This Stuff?

Welcome to this lecture series on reasoning with probabilities. I want to start off by acknowledging that studying probability theory isn't high on most people's "bucket lists" of things to do before they die, so we should probably spend some time talking about why this stuff is important from a critical thinking standpoint.

Here are **five reasons to study probability**.

1. It's an essential component of so-called "**inductive logic**", which is the branch of logic that deals with risky inferences. Inductive logic is arguably more important for critical thinking purposes than deductive logic.
2. It's an essential component of **scientific reasoning,** so if you want to understand scientific reasoning, you need to understand something about probability.
3. There are many interesting **fallacies** associated with probabilistic reasoning, and critical thinkers should be aware of at least some of these fallacies.
4. Human beings suffer from what some have called **"probability blindness"**. On our own, we're very bad at reasoning with probabilities and uncertainty. Or to put it another way, we're very susceptible to probabilistic fallacies. This fact about us is absolutely essential to understand if we're going devise strategies for avoiding these fallacies.
5. Finally, probability is **philosophically very interesting**, and a lot of important philosophical debates turn on the interpretation of probabilistic statements, so some grounding in the philosophy of probability can be very helpful in both understanding those debates and making informed critical judgments about those issues.

Just to give an example of the fifth point , the so-called “fine-tuning” argument for the existence of God is based on the premise that we live in a universe that is probabilistically very unlikely if it wasn’t the product of some kind of intelligent design, and therefore the best explanation for our existence in this universe is that it was, in fact, a product of intelligent design. But this kind of argument turns on what it means for something to be “probabilistically unlikely”, and whether it’s even meaningful to talk about the universe in this way. I won’t say any more about that here, but that’s just one example of an interesting philosophical debate where probability plays an important role.

2. What is Inductive Logic?

In this lecture I want to revisit the first point we raised, which is about **inductive logic**. I want to lay out some terms here so that it's clear what we're talking about, and the role that probability concepts play in inductive reasoning.

Deductive vs Inductive Logic

We distinguish deductive logic from inductive logic. Deductive logic deals with deductive arguments, inductive logic deals with inductive arguments. So what's the difference between a deductive argument and an inductive argument?

The difference has to do with the logical relationship between the premises and the conclusion. Here we've got a schematic representation of an argument, a set of premises from we infer some conclusion.

1. Premise
2. Premise
:
n. Premise
∴ Conclusion

That three-point triangle shape is the mathematicians symbol for "therefore", so when you see that just read it as "premise 1, premise 2, and so on, THEREFORE, conclusion.

Now, in a **deductive argument**, the intention is for **the conclusion to follow from the premises with CERTAINTY**. And by that we mean that IF the premises are all true, the conclusion could not possibly be false. So the inference isn't a risky one at all — if we assume the premises are true, we're guaranteed that the conclusion will also be true.

For those who've worked through the course on "Basic Concepts in Logic and Argumentation", you'll recognize this as the definition of a logically VALID argument. A deductive argument is one that is intended to be valid. Here's a simple example, well-worn example.

1. All humans are mortal.
2. Socrates is human.
∴ Socrates is mortal.

If we grant both of these premises, it follows with absolute deductive certainty that Socrates must be mortal.

Now, by contrast, with **inductive arguments**, we don't expect the conclusion to follow with certainty. **With an inductive argument, the conclusion only follows with some probability, some likelihood**. This makes it a "risky" inference in the sense that, even if the premises are all true, and we're 100 % convinced of their truth, the conclusion that we infer from them could still be false. So there's always a bit of a gamble involved in accepting the conclusion of an inductive argument.

Here's a simple example.

1. 90% of humans are right-handed.
2. John is human.
∴ John is right-handed.

This conclusion obviously doesn't follow with certainty. If we assume these two premises are true, the conclusion could still be false, John could be one of the 10% of people who are left-handed. In this case it's highly likely that John is right-handed, so we'd say that, while the inference isn't logically valid, it is a logically STRONG inference. On the other hand, an argument like this …

1. Roughly 50% of humans are female.
2. Julie has a new baby.
∴ Julie's baby is female.

... is not STRONG. In this case the odds of this conclusion being correct are only about 50%, no better than a coin toss. Simply knowing that the baby is human doesn't give us good reasons to infer that the baby is a girl; the logical connection is TOO WEAK to justify this inference.

These two examples show how probability concepts play a role in helping us distinguish between logically strong and logically weak arguments.

Now, I want to draw attention to two different aspects of inductive reasoning.

Two Different Questions to Consider

When you're given an inductive argument there are two questions that have to be answered before you can properly evaluate the reasoning.

The **first** question is this: **How strong is the inference from premises to conclusion? In other words, what is the probability that the conclusion is true, given the premises?**

This was easy to figure out with the previous examples, because the proportions in the population were made explicit, and we all have at least some experience with reasoning with percentages — if 90% of people are right-handed, and you don't know anything else about John, we just assume there's a 90% chance that John is right-handed, and 10% chance that he's left-handed. We're actually doing a little probability calculation in our head when we draw this inference.

This is where probability theory can play a useful role in inductive reasoning. For more complicated inferences the answers aren't so obvious. For example, if I shuffle a deck of cards and I ask you what are the odds that the first two cards I draw off the top of the deck will both be ACES, you'll probably be stumped. But you actually do have enough information to answer this question, assuming you're familiar with the layout of a normal deck of cards. It's just a matter of using your background knowledge and applying some simple RULES for reasoning with probabilities.

Now, the **other question** we need to ask about inductive arguments isn't so easy to answer.

The question is, **how high does the probability have to be before it's rational to accept the conclusion?**.

This is a very different question. This is a question about **thresholds for rational acceptance**, how high the probability should be before we can say "okay, it's reasonable for me to accept this conclusion — even though I know there's still a chance it's wrong". In inductive logic, **this is the threshold between STRONG and WEAK arguments** — strong arguments are those where the probability is high enough to warrant accepting the conclusion, weak arguments are those where the probability isn't high enough.

Now, I'm just going to say this up front. THIS is an unresolved problem in the philosophy of inductive reasoning. Why? Because it gets into what is known as the "problem of induction".

This is a famous problem in philosophy, and it's about how you justify inductive reasoning in the first place. The Scottish philosopher David Hume first formulated the problem and there's no consensus on exactly how it should be answered. And for those who do think there's an answer and are confident that we are justified in distinguishing between strong and weak inductive arguments, the best that we can say is that it's at least partly a conventional choice where we set the threshold.

To refine our reasoning on this question we need to get into rational choice theory where we start comparing the costs and benefits of setting the bar too low versus the costs and benefits of setting it high, and to make a long story short, that's an area that I'm not planning on going into in this course.

In this course we're going to stick with the first question, and look at how probability theory, and different interpretations of the probability concept, can be used to assign probabilities to individual claims AND to logical inferences between claims.

With this under our belt we'll then be in a good position to understand the material on probabilistic fallacies and probability blindness, which is really, really important from a critical thinking standpoint.

3. Probability as a Mathematical Object vs What That Object Represents

The first thing I want to do is **distinguish probability as a mathematical object from the various things that this object is used to represent**. This distinction helps to frame what we're doing in this first tutorial course in the probability series, which is the *meaning of probability*, and how it differs from what we're doing in the second tutorial course, which is on *the rules for reasoning with probabilities.*

First thing to note is that **modern probability theory is really a branch of mathematics**. The first formal work on the subject is from the 17th century in France by mathematicians Pierre de Fermat and Blaise Pascal, who were trying to figure out whether, if you throw a pair of dice 24 times, you should bet even money on getting at least one double-six over those 24 throws. They had an exchange of letters, and out of this exchange grew the first mathematical description of the rules for reasoning with probabilities.

Some Key Ideas of Probability Theory

Modern probability theory is a complicated beast, but here some are the key ideas.

i. Elementary Events

You imagine some set of **elementary events** or **outcomes**. Let's assume there are only six, so there are six elementary outcomes or events, {1, 2, 3, 4, 5, 6}. These could be the six sides of a dice.

ii. Probabilities of Elementary Events

We want to associate a probability with each elementary outcome — rolling a one, or a two, or a three, etc. We'll call these P(1), P(2), etc.

In this case it's pretty obvious, the odds for each of these elementary outcomes is just 1/6, i.e. P(1) = 1/6, P(2) = 1/6, etc. These are the probabilities of these elementary events.

iii. Probabilities of Logical Combinations of Elementary Events

But we also want to be able to figure out the odds of **different logical combinations of these elementary outcomes**. Like for example, the odds of rolling an even number, or a number less than 4, or a number that's either a 1 or a 5, or a number that's not a 6.

- P(even) = ?
- P(<4) = ?
- P(1 or 5) = ?
- P(not-6) = ?

We'd like to be able to calculate these probabilities.

iv. A Probability is a Mathematical Function

So, we've got these expressions that read "the probability of event A equals some number", P(A) = n, and this event is an an elementary outcome or some logical combination of elementary outcomes.

Mathematically, what we have here is a **function** that assigns to each event a number. That symbol, P, represents a mathematical function.

More specifically, this function takes as input some description of a possible event, and maps it onto the real number line.

P: elementary outcomes → real number

e.g.

P(3) = 1/6

v. P(E) is a Real Number Between 0 and 1

The value of this number is going to lie between 0 and 1, where 0 represents events that can't happen, that have probability 0, and 1 represents events that MUST happen, that have probability 1.

So, the probability of rolling a 1 is just 1/6, which is about 0.17. The probability of rolling an *even number* is just 1/2, or 0.5, because the even numbers include 2, 4 and 6, which make up half of all the possible outcomes.

vi. Events are Subsets of the Set of Elementary Outcomes

The other thing to note here is that, mathematically, the way we represent these different events is in terms of **subsets of the space of all possible events**.

> Example: The event "roll an even number" is represented by the set {2, 4, 6}, which is a subset of the set of elementary outcomes, {1, 2, 3, 4, 5, 6}.

That's how a description of an event gets translated into mathematical form. So, **a probability function is a mapping between the subsets of this larger set and the real numbers between 0 and 1**.

Now, we're not doing formal probability theory here. This is just about all I want to say about probability as a mathematical concept, since for critical thinking purposes this is about all you need to know. Mathematicians will use all kinds of terminology to really specify what's going on here. They'll talk about "sigma-algebras" and "structures that satisfy the Kolmogorov axioms", but all of this is stuff that we don't need to worry about.

What Formal Probability Theory Can Tell Us

The one thing I want you to note about probability theory is this. Given an assignment of probabilities to events A and B, the mathematics of probability gives us rules for figuring out the probabilities of various other events:

- P(not-A)
- P(A and B)
- P(A or B)
- P(A given B)

We'll look at the rules for calculating these probabilities in the course on "Rules for Reasoning with Probabilities".

Interpretations of Probability: Why They're Needed

The rules for reasoning with probabilities tell us how, given P(A) and P(B), we can work out P(A and B), P(A or B), etc.

Here's a question: **how exactly do we assign values to P(A) and P(B) in the first place?**

The mathematics of probability doesn't really address this question.

Why not? Because this is really a question about **what it *means to say* that the probability of an event is such-and-such**; this is about what probability, as a concept, represents in the world OUTSIDE of mathematics.

This is the question that different INTERPRETATIONS of probability try to answer. We're going to look at a range of these and their variations in the next section of the course:

- the classical interpretation
- the logical interpretation
- frequency interpretations
- subjective interpretations
- propensity interpretations

They each represent a distinct way of thinking about chance and uncertainty in the world.

The mathematics of probability puts some constraints on what can count as a viable interpretation of probability, but it allows for more than one interpretation. **The question isn't which interpretation is correct, but rather which interpretation is suitable or appropriate for a given application.**

That's why, as critical thinkers, it helps to be familiar with these different interpretations, because no single interpretation is suitable for every situation. There may even be situations where NO interpretation is suitable, and we have to conclude that it's simply a mistake to apply probabilistic concepts to situations like this.

Part 2: Interpretations of the Concept of Probability

In Part 2 we're going to be looking at different ways that mathematicians and philosophers have interpreted the concept of probability, what it means to say that "the probability of rolling a six is 1 in 6", or "there's a sixty percent chance of rain today". We'll see that there are several different ways of interpreting this language, and for this reason, these are sometimes called different "interpretations" or different "theories" of probability.

1. Classical Probability

The first interpretation we're going to look at is also one of the earliest and most important, and it's come to be called the "**classical"** interpretation of probability.

The classical interpretation of probability comes from the work of mathematicians in the 17th and 18th century — people like Laplace, Pascal, Fermat, Huygens and Leibniz. These guys were trying to work out the principles that governed games of chance and gambling games, like dice and cards and roulette, and in particular they were interested in working out the best betting strategies for different games. This was where the modern mathematical theory of probability was born.

The main idea behind the classical interpretation is very straightforward. Given some random trial with a set of possible outcomes, like tossing a coin or rolling a dice, we say that **the probability of any particular outcome is just the ratio of the favorable cases to the total number of equally possible cases**. Here a "favorable" case is just a case where the outcome in question occurs.

$$P(A) = \frac{\#\ \text{favorable cases}}{\#\ \text{total equally possible cases}}$$

So, if we're talking about a coin toss, the probability of it landing heads is obviously 1/2 on this interpretation. There are only two possible outcomes, heads or tails, so the denominator is 2. And of those two there's only one case where it lands heads, so the numerator is a 1.

Let's look at a dice example. What's the probability of rolling a 2 on a six-sided die? Well, there are 6 equally possible outcomes, and only one outcome where it lands 2, so the numerator is 1 and the denominator is 6, so the answer is 1/6, or 0.17, or about 17 percent.

$$P(2) = \frac{(2)}{(1),(2),(3),(4),(5),(6)} = \frac{1}{6}$$

If we want to know the probability of rolling an even number, then our situation is a bit different. Now our favorable cases include three of the six possible outcomes — 2, 4 and 6 — which are even numbers. So the probability is just 3 out of 6, or 1/2.

$$P(\text{even}) = \frac{(2),(4),(6)}{(1),(2),(3),(4),(5),(6)} = \frac{3}{6} = \frac{1}{2}$$

Note: The Requirement of Equally Possible Outcomes

These results are all correct, and the reasoning seems intuitively right. But it's clear that **this only works if each of the elementary outcomes is equally possible**. The classical interpretation is especially well suited to **games of chance** that are designed precisely to satisfy this condition — this is an interpretation of probability that was born in casinos and gambling halls and card tournaments.

However, it's not at all clear that this interpretation of probability is adequate as a *general* interpretation of the probability concept. In particular, this condition that all the outcomes be equally possible has been a cause for concern. What exactly does "equally possible" mean, in general? If we just mean "equally probable", then there's a risk of circularity, since our definition of probability is now invoking the concept of probability in the definition.

The French mathematician Laplace famously tried to clarify this idea. He says that we should treat a set of outcomes as "equally possible" **if we have no reason to consider one more probable than the other**. This is known as Laplace's **"principle of indifference"** (though it was John Maynard Keynes who coined this expression). The idea is that if we have no reason to consider one outcome more probable than the other, then we shouldn't arbitrarily choose one outcome to favor over another, that doing so would be *irrational*.

We're intended to use this principle of indifference in cases when we have *no evidence at all* for what the elementary probabilities might be, and cases where we have *symmetrically balanced evidence*, like in the case of coin tosses and dice rolls, where you know, given the geometry and symmetries of a cubical dice, that each side is as equally like to land as any other.

So, a strength of the classical interpretation is that **it gives intuitively satisfying answers to a wide variety of cases where these conditions apply**, like games of chance. But it has a lot of weaknesses as a general theory of probability.

Objections

Let me just lay out a couple of objections to the theory.

1. The requirement of "equally possible outcomes".

Consider for example how we might use this interpretation to assign a probability value to the question, "what are the odds that it's going to rain today?"

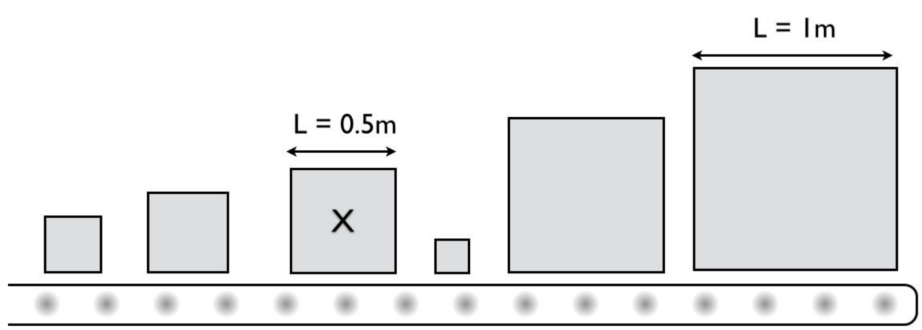
Okay, what's the favorable outcome? It rains. What's the set of possible alternative outcomes? It rains or it doesn't? But if we're forced to assign equal probabilities to each outcome in order to use the classical definition of probability, as the principle of indifference suggests, and these are the two elementary outcomes, then according to this definition *the probability of it raining is always going to be 1/2*.

That makes no sense, something is clearly not right. This is an example of a situation where it's very hard to see how the necessary conditions for the use of the classical definition could apply and make intuitive sense of the question. In this case it's not obvious how to define the set of alternative outcomes that are supposed to be equally possible.

2. Consistency objections

The most serious objections to the classical interpretation of probability are **consistency** objections. It seems that under this interpretation it's possible to come up with **contradictory probability assignments** depending on how you describe the favorable outcomes relative to the space of possible outcomes, and **the interpretation doesn't have the resources to resolve these contradictions without smuggling in other concepts of probability**.

Here's a well known example from the literature that illustrates the problem. Suppose a factory produces cubes with a side-length between 0 and 1 meter. We don't know anything about the production process.



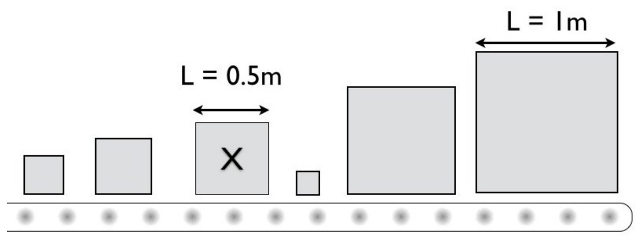
Question: What is the probability that a **randomly** chosen cube has a side-length $0 \leq L < 0.5$? (i.e **smaller than box X**)

Question: **what is the probability that a randomly chosen cube has a side-length between 0 and half a meter?** In other words, what is the probability that a randomly chosen cube is smaller than that box marked X in the figure above?

Well, given this phrasing of the question, it's natural to spread the probability evenly over these two event types: picking a cube that has a side length between 0 and half a meter, and picking a cube that has a side length between half a meter and 1 meter.

Why? Because we don't have any reason to think that one outcome is more probable than the other.

So the classical interpretation would give an answer of 1/2 to this question, and we can see why.

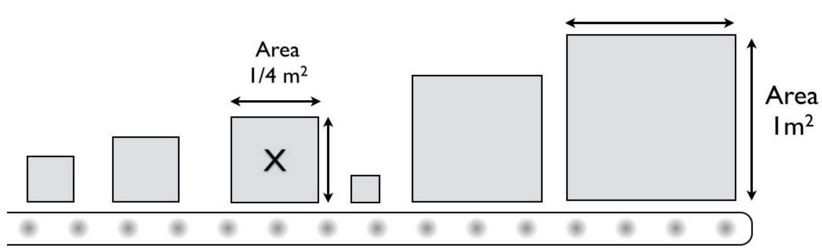


Two elementary outcomes: $(0 \leq L < 0.5)$ and $(0.5 < L \leq 1)$

$$P\,(0 \leq L < 0.5) = \frac{(0 \leq L < 0.5)}{(0 \leq L < 0.5)\,,\,(0.5 < L \leq 1)} = \frac{1}{2} = 0.5$$

Since we've got two equally possible outcomes (the box length is between 0 and 0.5, OR the box length is between 0.5 and 1 — and that number goes in the denominator) and only one favored outcome (the box length is between 1 and 0.5 — and that number goes in the numerator) this gives us probability one half, or 0.5.

Now, to see how the consistency problem arises, let's take the exact same setup, but let's phrase the question slightly differently. Suppose our factory produces cubes with face-area (not side-length, but the area of the face of a cube) between 0 and 1 square meter. So the area of the face of every cube is randomly between 0 and 1 square meter.

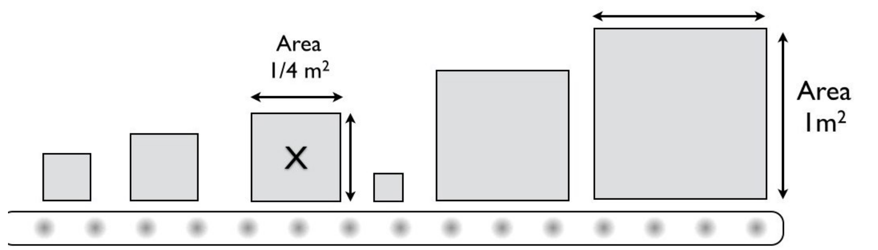


Question: What is the probability that a **randomly** chosen cube has a **side-AREA** $0 \leq A < 1/4$? (i.e. smaller than box X)

Question: **What is the probability that a randomly chosen cube has a face-area between 0 and 1/4 square meters?**

Now, phrased this way, the natural answer, using the classical interpretation, is going to end up being 1/4, instead of 1/2.

Why? Because it's natural now to consider **four** equally possible event types: picking a cube with an area between 0 and 1/4 square meters, picking a cube between 1/4 and 1/2 square meters, picking a cube between 1/2 and 3/4 square meters, and picking a cube between 3/4 and 1 square meter.



FOUR elementary outcomes:

$$\underset{a}{(0 \leq A < 0.25)}\ ,\ \underset{b}{(0.25 \leq A < 0.5)}\ ,\ \underset{c}{(0.5 < A \leq 0.75)}\ ,\ \underset{d}{(0.75 < A \leq 1)}$$

$$P(a) = \frac{\{a\}}{\{a, b, c, c\}} = \frac{1}{4} = 0.25$$

We don't have any reason to think that one of these outcomes is more likely than any other, so the principle of indifference will tell us to assign equal probabilities to each.

Our favorable outcome is just one out of these four equally possible outcome, so the numerator is 1 and the denominator is 4, giving 1/4.

I hope that's clear enough. In the diagram I've just labeled the outcomes a, b, c and d to help make the point, all we're doing is calculating the ratio of the number of favorable outcomes to the total number of possible outcomes.

Now, here's the point. I want you to see that these two questions:

1) what is the probability of randomly choosing a cube with side-length between 0 and a half meter?

and

(2) what is the probability of randomly choosing a cube with face-area between 0 and 1/4?

are asking for the probability of the SAME EVENT.

Why? Because the cubes with a side-length of 1/2 are ALSO the cubes with a face-area of 1/4, since the area of the face is just 1/2 times 1/2, which is 1/4 (or 0.5 times 0.5, which is 0.25). So, **all the cubes that satisfy the first description also satisfy the second description. The events are just described differently.** In other words, that "box X" is the same box in both cases.

And here's the problem: the classical interpretation of probability lets you assign *different* probabilities to the *same* event, *depending on how you formulate the question*. It turns out that there's literally an infinite number of different ways of reformulating this particular question, and the classical interpretation gives different answers for every formulation.

Now, this might not seem like a big deal to you, but this is regarded by mathematicians and philosophers as a fatal flaw in the theory, and it's one of the reasons why you won't find any experts today who defend the classical interpretation of probability as a general theory of probability. It gives the right answers in a bunch of special cases, and there's something about the reasoning in those special cases that is intuitively compelling, but that's about the most you can say for it.

2. Logical Probability

When you're reading about different interpretations of the probability concept, you might encounter the term "**logical probability**" used basically as a synonym for "classical probability", which we discussed in the previous tutorial. There's nothing wrong with this usage if it's clear what you're talking about, but there's potential for confusion because the term "logical probability" is also used to refer to a broader 20th century research program in the foundations of probability theory and inductive reasoning, and there are significant differences between this interpretation of probability and the classical interpretation of the 17th and 18th century.

What is Logical Probability?

The basic idea behind logical probability is to treat it as a generalization of the concept of *logical entailment* in deductive logic.

Just to refresh your memory, in a deductively valid argument the premises logically entail the conclusion in the sense that, if the premises are all true, the conclusion can't possibly be false — the truth of the premises guarantees the truth of the conclusion. If the argument is deductively invalid then the premises do not logically entail the conclusion, which simply means that even if the premises are all true, it's still possible for the conclusion to be false.

Logical entailment in this sense is a bivalent or binary property, it only has two values, "yes" or "no", like a digital switch — every argument is either deductively valid or it isn't. There are no degrees of validity, no degrees of logical entailment.

However, we can all recognize examples where logical support seems to come in degrees, and we can sometimes quantify the strength of the logical entailment between premises and a conclusion.

Here’s an example:

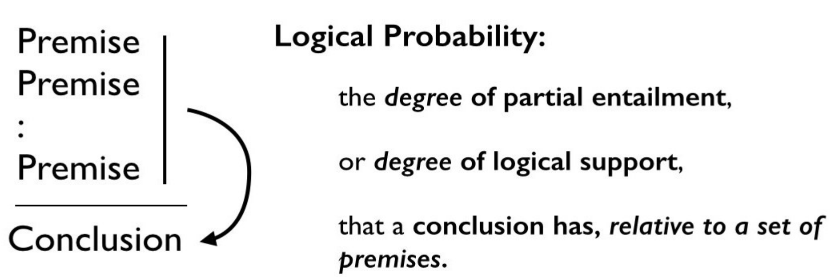
1. There are 10 students in this room.
2. 9 of them are wearing green shirts.
3. 1 is wearing a red shirt.
4. One student is chosen randomly from this group.
Therefore, the student that was chosen is wearing a green shirt.

This is a case where the conclusion doesn’t follow with deductive certainty, the argument isn’t deductively valid in this sense, but our intuition is that the conclusion does follow *with a high probability*.

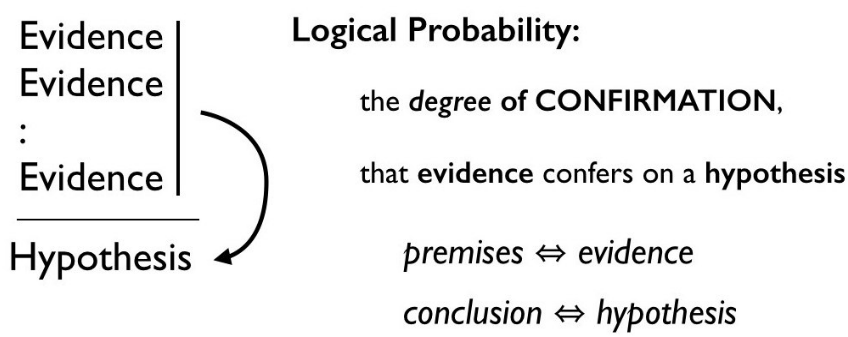
How high? Well, we’re about 90% sure that the conclusion is true — not 100%, but still something that a reasonable person might bet on.

Another way to say this is that the premises don’t *completely* entail the conclusion, but they *partially* entail the conclusion. So on this reading, the statement “The student is probably wearing a green shirt”, or more precisely, “it is 90% likely that the student is wearing a green shirt” — can be read as making a claim about **the degree of partial entailment or logical support that the premises confer on the conclusion**.

The logical approach to probability *defines probability in these terms*, as a measure of the *degree of partial entailment*, or *degree of logical support* that a conclusion has, given certain premises. When you think of the probability of a statement as ranging in value between 0 and 1, then 1 represents classical logical entailment, where the premises guarantee the truth of that statement, and values less than 1 represent greater and lesser degrees of partial entailment.



Another way this is often framed as in terms of the *degree of confirmation* that *evidence confers on a hypothesis*, where the evidence is identified with the premises of an argument, and the hypothesis is identified with the conclusion. When you phrase it this way you're making explicit the connection between logical probability and some very basic issues in scientific reasoning, like how to estimate how likely it is that a scientific theory is true, given all the evidence we have so far.



One of the interesting features of this approach to probability is that it makes no sense to talk about the probability of the conclusion all by itself; you're always talking about conditional probability, the probability of the conclusion given the premises. **Unconditional probability makes no sense** — *probability is always relative to some body of evidence, or some background assumptions*. Many critics of logical probability take issue with this point, they think it should be perfectly reasonable to talk about unconditional probabilities.

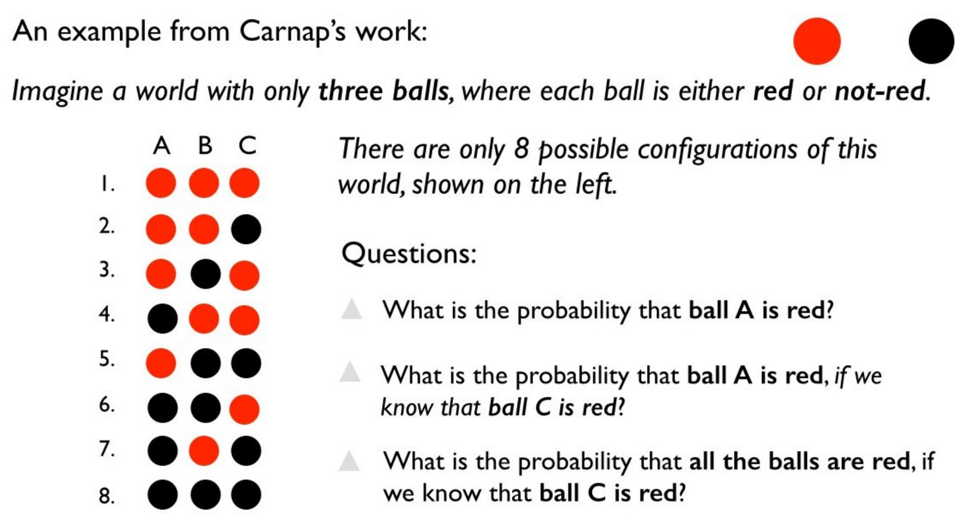
Trying to Work This Out in Detail

So, that's the basic idea behind logical probability. The difficulty with this approach comes when you try to work out this idea in detail.

For this to be a candidate for a general theory of probability you'll need to work out a general account of what this logical relationship is and how to operationalize it, how to actually assign a value to the strength of the logical support that evidence confers on a hypothesis. And this has proven to be a very tricky problem to solve. It has preoccupied some of the smartest minds of the 20th century, including the British economist John Maynard Keynes in the 1920s, and most notably the philosopher Rudolph Carnap in the 1950s. It's fair to say that no one has yet come up with a satisfactory way of defining and operationalizing logical probability.

The difficulty of the problem arises in part from the desire to have a genuinely logical definition of partial entailment, or degree of confirmation. That means that this relation should only depend on logical features of the world, or more accurately, logical properties of our descriptions of the world. So, for example, it shouldn't depend on specific knowledge we may have about the particular world that we find ourself in, it should be independent of that kind of substantive empirical knowledge.

So, Carnap for example tries to define a "confirmation function" that applies to formal languages, and to illustrate the idea he uses little toy models of worlds with, say, only three objects in them, and only one property, that each object either has or doesn't have.



In these toy worlds you can list all the possible states that such a world can be in, and you can define different event-types as subsets on this state space. And when you do this you can show how, given information about one of the objects, you can formally define how likely it is that certain facts will be true of the other objects. So in this example, if you have evidence that Ball C is red, then this will change your estimation of the likelihood that, say, ball A is red, or that all the balls are red.

Now, Carnap was trying to generalize this procedure in a way that would give a general definition of logical confirmation in the form of a confirmation function that would apply to all cases where evidence has a logical relationship to a hypothesis. But Carnap himself realized that there's more than one way to define the confirmation function in his system, and logic alone can't tell us which of these to choose.

The details don't matter too much for our purposes, but Carnap's system runs into technical and philosophical problems when you try to work it out.

There are other objections to the whole program of logical probability that I won't go into here, again, but most of them arise, as I said, from the constraint that the definition of probability be a logical or formal one that doesn't rely on specific knowledge about the world.

It's fair to say that today, this program is mostly of academic interest to certain philosophers and people working in the foundations of probability; it's not where the cutting edge of the discussion is among scientists or the mainstream of people working on probability.

Today, most of the discussion is between proponents of frequency-based approaches to probability, and proponents of subjective or Bayesian approaches, which is what we'll turn to in the next couple of tutorials.

3. Frequency Interpretations

In the last two lectures we've looked at the classical and the logical interpretations of probability. Now let's turn to one of the most widely used interpretations of probability in science, the **"frequency"** interpretation.

The frequency interpretation has a long history. It goes back to Aristotle, who said that "the probable is that which happens often". It was elaborated with greater precision by the British logician and philosopher (1834-1923) John Venn, in his 1866 book *The Logic of Chance*, and there are many important 20th century figures who have elaborated or endorsed some version of the frequency interpretation (e.g. Jerzy Neyman, Egon Pearson, Ronald Fisher, Richard von Mises, etc.)

The basic idea behind the frequency approach, as always, is pretty straightforward. Let's turn once again to coin tosses. **How does the frequency interpretation define the probability of a coin landing heads on a coin toss?**

Well, let's start flipping the coin, and let's record the sequence of outcomes. And for each sequence we'll write down the number of heads divided by the total number of tosses.

First toss, heads. So that's 1/1.

Second toss is a tail. So that's 1/2.

Third toss is a head, so now we've got 2/3.

Fourth is a tail, so now it's 2/4.

Fifth is a tail, so know it's 2/5.

Let's cycle through the next five tosses quickly and see the result after ten tosses:

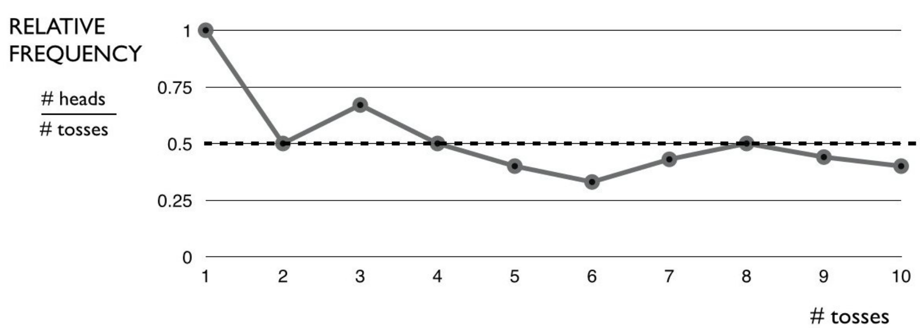
1. H
2. H T
3. H T H
4. H T H T
5. H T H T T
6. H T H T T T
7. H T H T T T H
8. H T H T T T H H
9. H T H T T T H H T
10. H T H T T T H H T T

$$\frac{\#\text{ heads}}{\#\text{ tosses}} = \frac{1}{1}, \frac{1}{2}, \frac{2}{3}, \frac{2}{4}, \frac{2}{5}, \frac{2}{6}, \frac{3}{7}, \frac{4}{8}, \frac{4}{9}, \frac{4}{10}$$

These ratios on the bottom are called **"relative frequencies"**, and a sequence like this is called a **relative frequency sequence**.

There are a few obvious observations we can make about this sequence. First, we see that it jumps around the value of 1/2, sometimes exactly 1/2, sometimes higher, sometimes lower.

It might be easier to look at the sequence in decimal notation to see this more clearly. And it to make it even more clear, let's graph this sequence.



$$\frac{\#\text{ heads}}{\#\text{ tosses}} = 1,\ 0.5,\ 0.67,\ 0.5,\ 0.4,\ 0.33,\ 0.43,\ 0.5,\ 0.44,\ 0.4$$

It's more obvious now that the sequence bounces around the 0.5 mark. Three times it's exactly 0.5, but we know it can't stay at 0.5, since the next toss will move the ratio either above or below 0.5.

What's also obvious, I think, is that *the range of variation gets smaller as the number of tosses increases*, and if we were to continue tossing this coin and recording the relative frequency of heads, we would expect that this number would get closer and close to 0.5 the more tosses we added.

The Frequency Definition of Probability

Now, none of this is surprising, but **what does it have to do with the definition of probability**?

Everyone agrees that there are important relationships between probability and the relative frequencies of events. This is exactly the sort of behavior you'd expect if this was a fair coin that was tossed in an unbiased manner. We assume the probability of landing heads is 1/2, so the fact that the relative frequency approaches 1/2 isn't surprising.

But what's distinctive about frequency interpretations of probability is that they want to IDENTIFY probabilities WITH relative frequencies. On this interpretation, to say that the probability of landing heads is 1/2 IS JUST TO SAY that if you were to toss it, it would generate a sequence of relative frequencies like this one. Not exactly like this one, but similar.

For a case like this one, **the frequency interpretation will DEFINE the probability of landing heads as the relative frequency of heads that you would observe in the long run, as you kept tossing the coin**.

To be even more explicit, this long-run frequency is defined as the **limit** that the sequence of relative frequencies approaches, as the number of tosses goes to infinity. In this case it's intuitive that the sequence will converge on 1/2 in the limit. And if that's the case, then, according to this approach, we're justified in saying that the probability of landing heads is exactly 1/2.

Actually, what we've done here is introduce two different relative frequency definitions:

finite relative frequency

H T H T T T H H T T $\left(\frac{\text{\# favored outcomes}}{\text{\# trials}}\right)$

limiting relative frequency

H T H T T T H H T T H T H T T T H H T T . . . ⟶

$\left(\frac{\text{\# favored outcomes}}{\text{\# trials}}\right)$ as the number of trials approaches infinity

You can talk about probabilities in terms of **finite relative frequencies**, where we're only dealing with an actual finite number of observed trials; or we can talk about probabilities in terms of **limiting relative frequencies**, where we're asked to consider what the relative frequency would converge to in the long run as the number of trials approaches infinity.

Some cases are more suited to one definition than the other.

Batting averages in baseball, for example, are based on actual numbers of hits over actual numbers of times at bat. It doesn't make much sense to ask what Ty Cobb's batting average would be if he had kept playing forever, since (a) we'd expect his performance to degrade as he got older, and (b) in the long run, to quote John Maynard Keynes, we're all dead!

Coin tosses, on the other hand (and other games of chance) look like suitable candidates for a limiting frequency analysis. But it's clear that more work needs to be done to specify just what the criteria are and what cases lend themselves to a limiting frequency treatment, and this is something that mathematicians and philosophers have worked on and debated over the years.

Frequency Interpretations and Statistical Sampling

I've said a couple of times that frequency interpretations are widely used in science, and I'd like to add a few words now to help explain this statement. There's a version of the frequency approach that shows up in ordinary statistical analysis, and it's arguably what most of us are more familiar with. It's based on the fact that sequences of random trails are formally related to proportions in a random sampling of populations.

Just to make the point obvious, when it comes to relative frequencies, there's **no real difference** between **flipping a single coin ten times in a row** and **flipping ten coins all at once**. In either case some fraction of the tosses will come up heads.

In the **single coin case**, as you keep tossing the coins, we expect the relative frequency of heads to converge on 1/2.

In the **multiple coin case**, as you increase the number of coins that you toss at once — from ten to twenty to a hundred to a thousand — we expect the ratio of heads to number of coins to converge on 1/2.

This fact leads to an obvious connection between relative frequency approaches and standard statistical sampling theory, such as what pollsters use when they try to figure out the odds that a particular candidate will win an election. You survey a representative sampling of the population, record proportions of "Yes" or "No" votes, and these become the basis for an inference about the proportions one would expect to see if you surveyed the whole population.

All I'm drawing attention to here is the fact that frequency approaches to probability are quite commonly used in standard statistical inference and hypothesis testing.

Objections to the Frequency Interpretation

Let's move on to some possible objections to the frequency interpretation of probability. Let me reiterate that my interest here is not to give a comprehensive tutorial on the philosophy of probability. My goal, as always, is nothing more than **probability literacy** — we should all understand that probability concepts can be used and interpreted in different ways, and some contexts lend themselves to one interpretation better than another. These objections lead some to believe that the frequency interpretation just won't cut it as a general theory of probability, but for my purposes I'm more concerned about developing critical judgment, knowing when a particular interpretation is appropriate and when it isn't.

i. How do we know what the limiting frequencies will be?

Let's start with this objection. If probabilities are limiting frequencies, then how do we know what the limiting frequencies are going to be? The problem arises from the fact that these limiting behaviors are supposed to be inferred from the patterns observed in actual, observed, finite sequences, they're not defined beforehand, like a mathematical function. So we can't deductively PROVE that the relative frequencies of a coin toss will converge on 0.5. Maybe the coin is biased, and it's going to converge on something else? Or let's say that we now get a series of ten heads in a row? Does that indicate that the coin is biased and that it won't converge on 0.5? But isn't a series of ten heads in a row still consistent with it being a fair coin, since if you tossed the coin long enough you'd eventually get ten in a row just by chance.

I'm not saying these questions can't be worked out in a satisfying why, I'm just pointing out one of the ways that the application of the limiting frequency approach to concrete cases can be difficult, or can be challenged.

ii. The reference class problem

Let's move on to another objection, which is sometimes called the "reference class problem". And this one applies both to finite and limiting frequency views.

Let's say I want to know the probability that I, 46 years old at the time of writing this, will live to reach 80 years old. One way to approach this is to use historical data to see what proportion of people who are alive at 46, also survive to 80. The question is, how do we select this group of people from which to measure the proportion? A random sample of people will include men and women, smokers and non-smokers, people with histories of heart disease and people who don't, people of different ethnicities, and so on. Presumably the relative frequency of those who live to age 80 will vary across most of these reference classes. Smokers as a group are less likely to survive than non-smokers, all other things being equal, right?

The problem for the frequency interpretation is that it doesn't seem to give a single answer to the question "What is the probability that I will live to 80?". Instead, what it'll give me is **a set of answers relative to a particular reference class** — my probability *as a male*, my probability *as a non-smoker*, my probability *as a male non-smoker*, and so on.

To zero in on a probability specific to me, it seems like you need to define a reference class that is so specific that it may only apply to a single person, me. But then you don't have a relative frequency anymore, what you've got is a "single-case" probability.

iii. Single-case probabilities

Single-case probabilities are another category of objection to frequency interpretations.

When I toss a coin, it doesn't seem completely crazy to think that for this one, single coin toss, there's an associated probability of that toss landing heads. But frequency interpretations have a hard time justifying this intuition. This is important to see: **on the frequency interpretation, probabilities aren't assigned to single trials, they're assigned to actual or hypothetical *sequences* of trials**. For a strict frequentist, it doesn't make any sense to ask, "what is the probability of a single-case event?" But a lot of people think this concept should make sense, and so they reject frequency interpretations in favor of interpretations that do make sense of single-case probabilities.

So, for these and other reasons, many believe that the frequency interpretation just can't function as a truly general interpretation of probability.

In the next two lectures we'll look at interpretations of probability that, as we'll see, are much better at handling single-case probabilities. These are the subjective interpretations and the propensity interpretations, respectively.

4. Subjective (Bayesian) Probability

In this lecture we're going to look at so-called **"subjective"** or **"Bayesian"** interpretations of the probability concept. By the end you should have a better idea what this approach is all about and why many people find it an attractive framework for thinking about probabilities.

Probability as "Degree of Belief"

In the previous lecture we looked at "relative frequency" interpretations of probability. On this view, when we say that the probability of a fair coin landing heads is 1/2, what we're really saying is that if you were to toss the coin repeatedly, in the long run half of the tosses would land heads, and the probability is just identified with the long-run relative frequency behavior of the coin.

But there are lots of cases where the relative frequency interpretation just doesn't make intuitive sense of how we're using a probability concept.

For example, let's say I'm at work and I remember that I left the back door of my house unlocked, and I'm worried about the house being robbed because there have been a rash of robberies in my area over the past two weeks. So I'm driving home from work and I'm asking myself, what are the odds that when I get home I'll discover that my house has been robbed?

This is an example of a **single-case probability**. I'm not interested in any kind of long-run frequency behavior, I'm interested in the odds of my house being robbed *on this specific day, on this one, single occasion*. Examples like these are very hard for frequency approaches to analyze.

The more natural way to think of this case is this. I have a certain DEGREE OF BELIEF in whether this event occurred. If someone asks me how likely it is that I've been robbed today, and I say I think there's at least a 10% chance I was robbed, what I'm doing is reporting on **the strength of my subjective degree of confidence in this outcome**. What I'm reporting on is a subjective attitude I have toward the belief that I've been robbed — if my degree of belief is very low, that's a low probability;

if it's moderate, that's a moderate probability; if it's high, that's a high probability.

This is what it means to say that probability is SUBJECTIVE. What you're saying is that what probabilities represent are not features of the external world, but rather features of your personal subjective mental states, namely, your degree of belief that a given event will occur or that a given statement is true.

Subjectivists about probability want to generalize this idea and say that **probability in general should be interpreted as a measure of degree of belief.**

This conceptual framework can be applied even to cases where the frequency interpretation also works. If we're talking about the probability of a coin landing heads being equal to 1/2 , the subjectivist will interpret this as saying that you're 50% confident in your belief that the coin will land heads.

The Obvious Objection: We're Bad at Reasoning with Probabilities

Okay, at this point there's an obvious objection to interpreting probability in this way. The objection is this: People are notoriously BAD at reasoning with probabilities, our degrees of belief routinely violate the basic mathematical rules for reasoning with probabilities.

If probabilities are interpreted as mere subjective degrees of belief, then in what sense can we possibly have a *theory* of probability, a theory that distinguishes good reasoning from bad reasoning?

Subjectivists, or Bayesians, as they're often called, have an answer to this question.

They argue that the only logically consistent way of reasoning with subjective degrees of belief is if those degrees of belief satisfy the basic mathematical rules for reasoning with probabilities. **All the action in the subjective interpretation lies in the details of this argument.** What I'm going to give here is just a rough sketch of the reasoning, which I'll break down into three steps.

Step 1: Find a way of assigning a number to a person's degree of belief

One of the challenges of reasoning with subjective probabilities is that because they're a feature of our inner mental states, they're hard to access. We need some way of assigning a number to represent a degree of belief. How do we do this?

In the early part of the 20th century Frank Ramsey and Bruno de Finetti independently came up with the basic solution, which exploits the fact that there is a close relationship between belief and action, and in this case, between degrees of belief and one's betting behavior when one is asked to choose between different bets or gambles or lotteries.

I'm going to use a visual device to help illustrate the idea.

Let's say I want to measure the strength of your belief that a coin will land heads on the next toss. I know we know the probability in this case, it's 50%, but let's just use this case to illustrate the procedure. Then we can use an example that isn't so obvious.

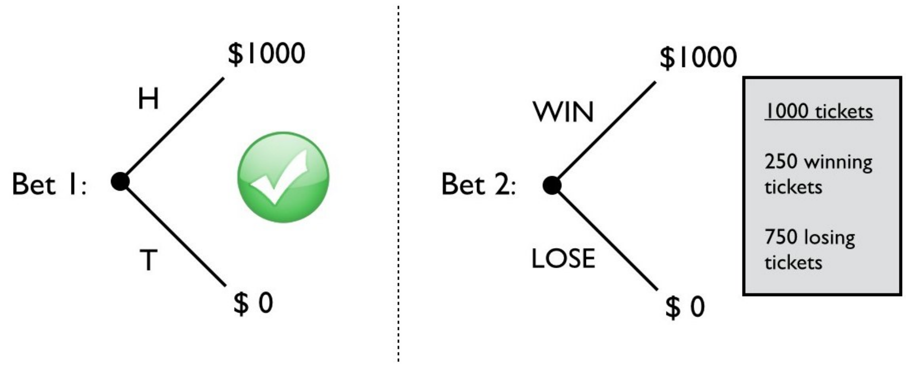
Okay, we imagine that you're faced with a choice, to select between two different bets.

Bet 1: The bet is whether you think it's true or not that the coin will land heads. If it lands heads, you win $1000. If it lands tails, you win nothing.

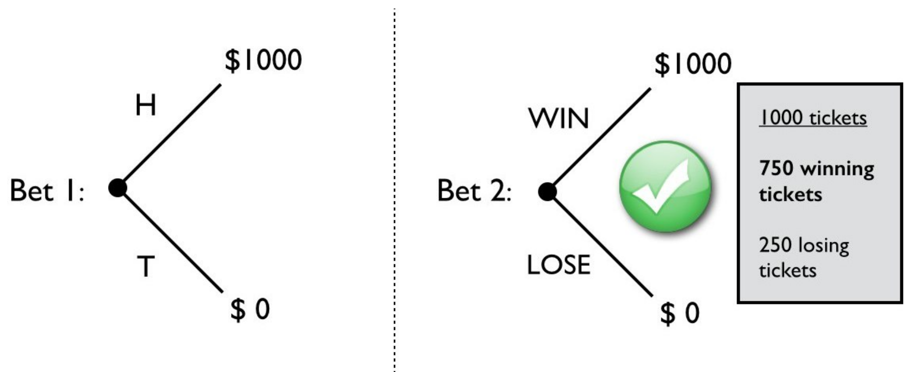
Bet 2: The bet is whether you should play a lottery, where the lottery has a thousand tickets. And in this lottery there are 250 winning tickets. If you draw a winning ticket you win $1000. If you don't draw a winning ticket you win nothing.

So the question is, **which bet would you prefer to take, Bet 1 or Bet 2**?

This is easy, we're all going to pick Bet 1, right? Because we think the odds of the coin landing heads are higher than the odds of winning the lottery, which is 25%. We believe that we're more likely to win Bet 1 than Bet 2.



Now imagine that Bet 2 was different. Imagine that the lottery in Bet 2 has 750 winning tickets, so you win if you draw any of those 750 winning tickets. Now which bet would you pick?

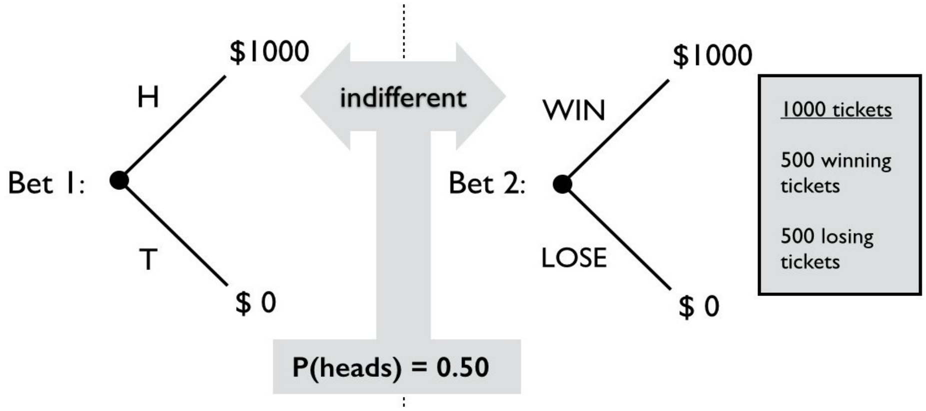


Well, this time we'd all pick Bet 2, because now the odds of winning the lottery is 75%, so we're more confident that we'd win this bet than Bet 1.

So, **what we've established here**, by examining your preferences between different bets, is that **your degree of confidence that the coin will land heads lies somewhere between 0.25 and 0.75**.

We can narrow this range by selecting different bets with different numbers of winning lottery tickets.

Now, what will happen if we're offered a lottery with exactly 500 winning tickets?

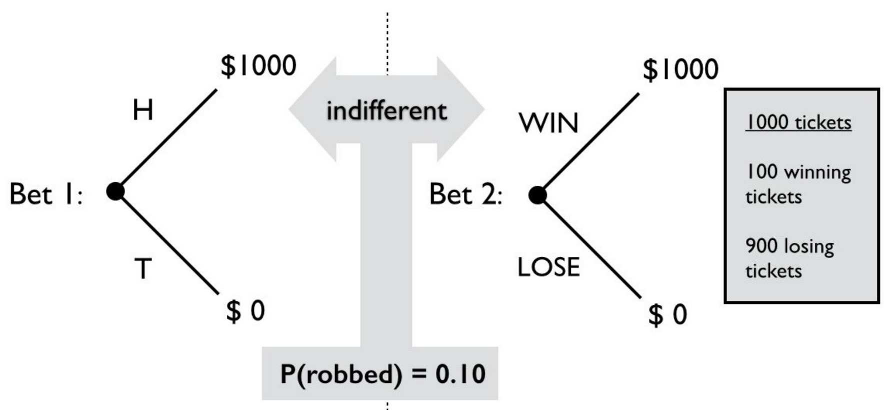


In this case, we should be **indifferent between these two bets**, since we think the odds of winning the 1000 dollars are the same in both cases. We wouldn't prefer one bet over the other.

And THIS is the behavioral fact that fixes your degree of belief in the proposition at hand. When you're indifferent between these two choices, **the percentage of winning tickets in the lottery can function as a numerical measure of the strength of your belief in the proposition that you're betting on in Bet 1**.

We can use this imaginary procedure to measure the strength of your belief in any proposition. Like my belief that when I get home I'll discover that my house was robbed.

If I end up being indifferent between a bet that my house was robbed and a bet that I'll win the lottery with, say 100 winning tickets, then we can say that I'm about 10% confident that my house was robbed.



Now, if you ever find yourself reading the subjective probability literature, the more common language you'll encounter is the language of "betting ratios" and "betting rates", but the main idea is the same. The procedure I'm describing here using lotteries is more commonly used in decision theory, but it's inspired by the same body of work by Ramsey and de Finetti.

Step 2: Show that a rational betting strategy must satisfy the rules of probability theory

So, we now have a way of representing our personal degrees of belief by betting rates on imaginary gambles. This gives us an operational procedure for assigning a real number to a degree of belief. But we still don't have any rules for how to reason with these degrees of belief.

The next step in the subjectivist program is to show that **a rational betting strategy will automatically satisfy the basic mathematical rules for reasoning with probabilities.**

In this context, all we mean by a rational betting strategy is this: no rational person will willingly agree to a bet that is guaranteed to lose them money. A bet that is guaranteed to lose you money is called a "sure-loss contract".

If someone's personal degrees of belief are open to a sure-loss contract, then that person can become a "money pump"— a bookmaker could exploit this knowledge to sell you betting contracts that you will accept, but that you will never win, you'll always lose money. Not a good thing.

These sure-loss contracts are also known as "Dutch book" contracts, and this kind of argument is called a "Dutch book" argument. Ramsey was the one who introduced this language but I don't know why he called it a Dutch book (I'm not sure what being Dutch has to do with it), but the term Dutch book is now standard in probability theory and economics. I'm going to follow Ian Hacking and just call it a "sure-loss contract".

We can now define an important concept: **If a set of personal degrees of belief is not open to a sure-loss contract, then the set of beliefs is called COHERENT.**

In other words, if your set of beliefs is coherent, then *by definition* you can't be turned into a money pump for an unscrupulous bookie. Note that this is a technical sense of "coherence" specific to this context. It's intended as an extension of the logical concept of consistency, applied to partial degrees of belief.

Now, the main theoretical result that Ramsey and de Finetti developed was this: **A set of personal degrees beliefs is coherent if and only if the set satisfies the basic rules of probability theory.** And here we're just talking about the standard mathematical rules.

The details of this theorem aren't important, what's important is what it represents for the subjectivist program. Our original concern, remember, was that personal degrees of beliefs are unconstrained, they don't follow any rules. What Ramsey and de Finetti and others have shown is that if one adopts this very pragmatic and self-interested concept of rationality — namely, that a rational person won't willingly adopt a set of subjective degrees of belief that is guaranteed to lose them money — then it follows that this person's belief set will satisfy all the basic rules of probability theory. And it is in this sense that the subjective approach to probability brings with it a normative theory of probability.

Step 3: Adopt Bayes' Rule as a general principle for how to learn from experience

Now, in the literature, people who work within the subjective framework I'm describing here are often called "Bayesians", and this approach is called "Bayesianism". So let's say a few words about this language.

Bayes' Rule can be derived from the basic mathematical rules of probability, it's basically just a way of calculating conditional probabilities given certain information. Here's the simplest form of Bayes' Rule.

Bayes' Rule:

$$P(H|E) = \frac{P(H) \times P(E|H)}{P(E)}$$

H and E can stand for any two propositions, but in practice we often use Bayes' Rule to evaluate how strongly a bit of evidence supports a hypothesis, so let H be a hypothesis and let E be some bit of evidence. Maybe H is the hypothesis that a patient has the HIV virus, and E is a positive blood test for the virus.

We read this term, P(H|E), as the probability that H is true, *given that* E is true. On the subjectivist reading, this is the degree of belief that we should have in hypothesis H, once we've learned about the evidence E. This is also called the **"posterior probability"** of the hypothesis.

P(H), all by itself, is called the **"prior probability"** of the hypothesis. This is the degree of belief we had in H before ever learning about the new evidence E. In our example, this would be the probability that the patient has HIV, before learning the results of the blood test.

P(E|H) is called the **"likelihood"** of the evidence, given the hypothesis. This is how likely it is that we would observe evidence E, if the hypothesis H was in fact true. So in our example, this is the probability that someone will test positive for the HIV virus, given that they actually have the virus.

The term in the denominator is called the **"total probability"** of the evidence E. In our example, this term is going to represent the probability of testing positive for the HIV virus, whether or not the patient actually has the virus. So this term will also depend on information about the false-positive rate for the rest, the percentage of times a patient will test positive, even when the don't have the virus.

I'm not going to spend any more time explaining how this calculation will go right here, because it's not vital to the point I'm making, and I've got a whole other course on the rules of probability theory that explains it in more detail.

The point I want to make here is that when you interpret probabilities the way that subjectivists do, *Bayes' Rule gives us a model for how we ought to learn from experience*, how we ought to update our degrees of belief in a hypothesis in light of new evidence. Bayes' Rule has lots of important applications in statistical inference theory, but in the hands of subjectivists it also functions as the central principle of a theory of rational belief formation and rational inference.

So this is why subjectivists are often called "Bayesians" — it's because within this interpretation, Bayes' Rule takes on great importance as part of a general theory of rationality. For frequency theorists, Bayes' rule is just another useful formulation of conditional probability, and its use is restricted to cases where relative frequencies can be defined. For subjectivists, Bayes' Rule is fundamental to their approach to rationality, and it can be used in a much wider range of applications, since they're not restricted to applications using relative frequencies.

There's also a whole field of philosophical work that you could describe as falling under the label of "Bayesian epistemology", which applies Bayesian principles to various problems in the philosophy of knowledge, the philosophy of science, in decision theory and learning theory, and so forth.

Regardless of what you think of it, this approach to probability has had a huge impact on philosophy and science.

Summary

Here's the summary of what we've been talking about.

- Step 1 in the Bayesian program is to find a way of numerically representing a person's degree of belief. We use betting rates to do this. Once we've got this, we can talk about rational and irrational betting strategies.

- In Step 2 we show that if our degrees of belief are "coherent", then they'll automatically satisfy the basic mathematical rules of probability theory.
- Step 3 involves the use of Bayes' Rule as a guide for how we ought to update our beliefs based on evidence.

Objections

We've been looking at objections to all the previous interpretations of probability that we've covered, so it's only fair to mention that of course there are objections to the kind of subjective Bayesianism that I've been describing here. The best I can do here is just name a few, since it would take too long to try to explain them all in detail here.

Here we go:

First objection: Bayesianism assumes logical omniscience

The claim is that if our beliefs satisfy the basic rules of probability, then the rules require that all beliefs about logical truths have probability 1, and beliefs about logical contradictions have probability 0. So on this view, if our beliefs are coherent then we can never believe a contradiction. The objection is that this is just false of human beings, none of us are logically omniscient in this way, and so it's unreasonable standard to impose on our beliefs.

Second objection: Bayesianism assumes that classical logic is the only logic

This follows from the bit about logical omniscience. We've never talked about non-classical logics before, but there are such things — logical theories that use different fundamental rules of inference from standard classical logic. We've largely moved away from the days when everyone thought that classical logic was the only possible logic one could use. The objection is that Bayesianism presupposes that the rules of classical logic are correct, and makes them immune to revision based on empirical evidence, and consequently it grants them a kind of a priori status that few people actually think it has anymore.

Third objection: The problem of old evidence

From Bayes' rule it follows that if the probability of a piece of evidence is 1, then the likelihood of the evidence given some hypothesis is also 1. But if this is so, then *such evidence can never raise the probability of a hypothesis*: the posterior probability will always be just the same as the prior probability.

This poses a problem for Bayesian views on how so-called "old evidence" might support a new scientific theory. For example, Newton's theory of gravity doesn't completely predict the orbit of Mercury, it doesn't adequately account for the precession of Mercury's orbit around the sun. This behavior of Mercury's orbit was known in the mid 19th century. 60 years later Einstein comes up with the general theory of relativity and his new theory accurately predicts this piece of "old evidence", the precession of Mercury's orbit. The objection is that this is rightly viewed as an empirical success of Einstein's theory, it should lend support to his theory; but the Bayesian has a hard time explaining HOW this old evidence can give us additional reason to believe the theory.

Fourth objection: The problem of new theories

It seems intuitive that sometimes, the invention of a new theory, all by itself, can influence our confidence in an old theory, especially when the old theory didn't have any rivals. Imagine the old earth-centered cosmology of Ptolemy, where all the heavenly bodies move around a motionless Earth. This theory had no competition for a long time. Then along comes Copernicus with this Sun-centered cosmology, that can explain everything that Ptolemy's theory did. Wouldn't this fact alone lead some people to reconsider their support for Ptolemy's theory, to lower their conviction in the truth of this theory? The objection to Bayesianism is that it's not clear how this kind of shift in support can be explained or justified in the Bayesian framework.

Fifth objection: Additional constraints on prior probabilities are needed

This is sometimes just called "the problem of the priors". The issue is this. What we're calling "subjective" Bayesianism doesn't place any restrictions on the values of the prior probabilities, beyond the requirement of coherence. In other words, it doesn't constrain your beliefs, beyond the requirement that they be consistent with the rules of probability, and when you learn new evidence, you update your beliefs according to Bayes' rule.

The objection is that this is just way too permissive. You can have literally crazy views of the world that would be permitted by these rules. Within those belief sets, you'd be updating your beliefs rationally when you encountered new evidence, but the belief sets themselves would be wildly different.

So different constraints on prior probabilities have been proposed. One proposal, for example, is that subjective degrees of belief should, at the very least, track the relative frequencies that are known. So, for example, if it's known that a baseball player is hitting .350 this season, then all other things being equal, it seems reasonable to assign a degree of belief of .35, or 35%, as the prior probability that he'll get a hit the next time at bat. Now, if you're a Bayesian and you think along these lines, then you're not a strict, subjective Bayesian — you're what's called an **"objective" Bayesian**, because you think there are objective features of the world, like relative frequencies, that should restrict the probabilities you assign to your beliefs. What we end up with is really **a family of Bayesian approaches to probability**, that range from more subjective to more objective varieties, and where you fall on this range depends on how many and what sorts of additional constraints you're willing to place on the prior probabilities.

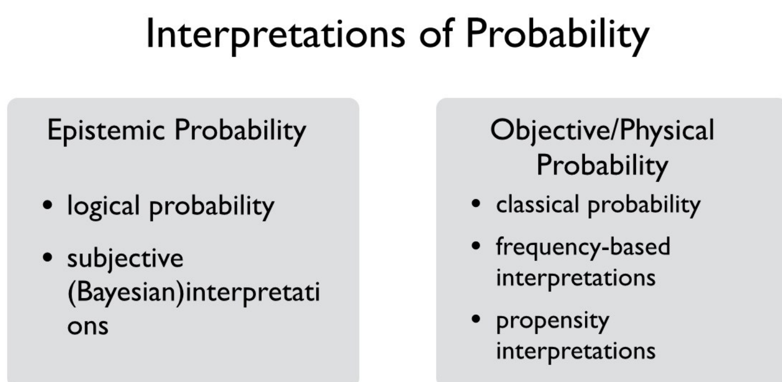
Okay, I think that's more than enough for this introduction to subjective probability. The are other objections of course, but these are some of the main ones.

To wrap things up, I'll just conclude with this: There are a lot of smart people working today in philosophy, statistics, applied math and science, computer science and artificial intelligence, who are engaged in research within what can be described as a broadly Bayesian framework. That doesn't mean that there aren't a lot of open problems with the framework that need to be solved. But it does mean that this is a framework that people are willing to openly endorse without embarrassment.

5. Propensity Interpretations

The last interpretation of probability that we're going to look at is known as the **propensity interpretation**. The term was coined by the philosopher Karl Popper in the 1950s.

Actually before we get into the concept of a propensity, let me just back up and situate this discussion a little bit. We've looked at a number of different interpretations of the probability concept, but you can carve up interpretations into roughly two camps, corresponding to two larger, umbrella concepts of probability.



The first concept is sometimes called **"epistemic probability"** or **"inductive probability"**. Ian Hacking calls it "belief-type" probability. This kind of probability is about how probable a statement is, or **how strongly we should hold a belief, *given certain facts or evidence*.** Given such-and-such evidence, what is the probability that the Big Bang Theory is true?, or that it'll rain tomorrow?, and so on.

The key thing about this kind of probability is that it doesn't depend on unknown facts about the world, it only depends on our available evidence — **probability judgments of this kind are always relative to the evidence that is available to some agent.**

Looking back at the probability concepts that we've discussed, it's clear that LOGICAL probability and SUBJECTIVE probability belong in this camp.

But there's another probability concept that we've also been discussing, which some call **"objective probability"**, or **"physical probability"**. This kind of probability is associated with **properties of the world itself, independent of available evidence, independent of what anyone happens to believe about the world**.

So, for example, when we hear reports of an outbreak of a new flu virus, we're told that in certain regions there's an increased chance of contracting the virus, and if this is true, it's true independently of what anyone happens to believe about the world.

Or think about radioactive decay, where there's, say, a 50% chance that a particular atom of some element will decay in the next hour. The half-life of a radioactive element is an objective feature of the world that we discover, it's not something that depends on the evidence or beliefs that we have about it.

Of the probability concepts that we've looked at, CLASSICAL probability and FREQUENCY interpretations of probability belong more to this camp. Now, admittedly the physical properties that are associated with probabilities in these theories are a little weird and abstract. In the classical theory they're ratios of favored outcomes over all equally possible outcomes; in frequency theories they're relative frequencies of observed or hypothetical trials. But in both cases, the probability of a coin toss landing heads is identified with a feature of coin tossings, not with our beliefs about coin tossings.

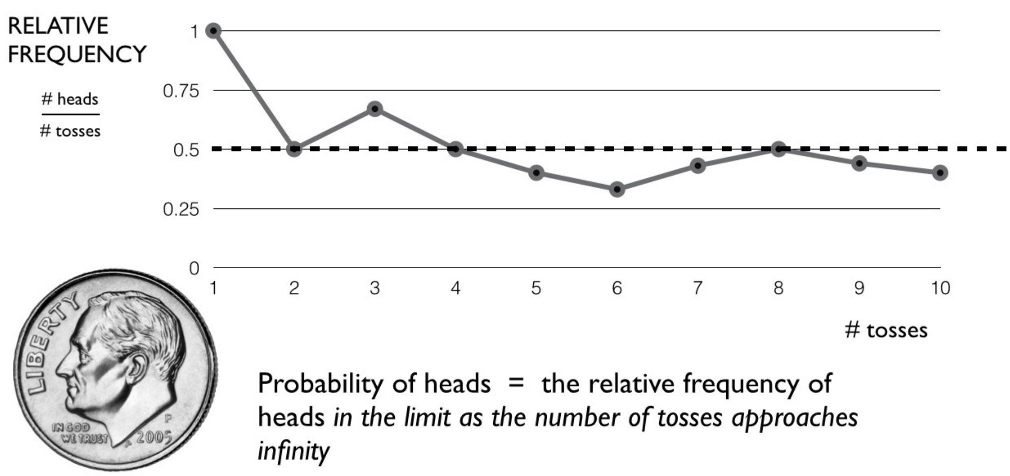
So what does this have to do with Popper and the propensity interpretation?

Well, **propensity interpretations land squarely in the objective, physical probability camp**. Popper introduced this concept because he thought that frequency-style interpretations of physical probability weren't adequate, so he's trying to articulate the concept of a physical probability in a better way.

Relative Frequencies vs Propensities

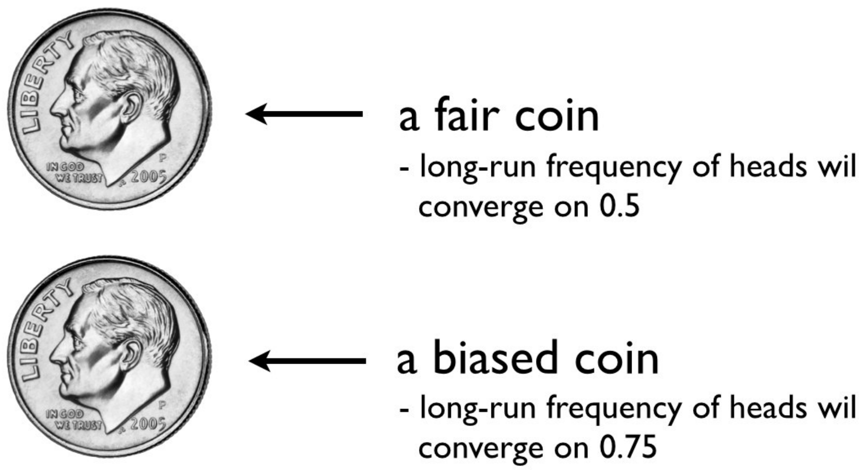
So, what's the difference between relativity frequencies and propensities? Let's consider our coin tossing example again.

On a relative frequency interpretation, the probability of the coin landing heads is identified with the long-run relative frequency of heads; so in the long run this frequency will converge on 0.5 for a fair coin, and this limiting frequency, this ratio, is what we're referring to when we say that P(H) = 0.5.



In other words, a probability, on this view, isn't a property of any individual coin toss — it's a property of a (potentially infinite) *sequence* of coin tosses.

Now, Popper thinks there's something incomplete about this approach. Consider two coins. The first is a fair, unbiased coin. The other is a biased coin, it's weighted more on one side than the other; so that when you toss it, it lands heads more often than tails. Let's say that in the long run it lands heads 3/4 of the time.



Popper asks us to consider these two coins, sitting in front of us on the table. These coins will generate different long-run frequencies when you toss them. **Why? What explains this difference in behavior?**

The obvious answer, says, Popper, is that the two coins **have different physical characteristics that are causally responsible for their long-term frequency behavior.**

It's these different physical characteristics that Popper calls "propensities". It's their different propensities to land heads that account for the differences in their frequency behavior. **And this is what numerical probabilities are taken to represent, *propensities* of an experimental setup to generate these different relative frequencies of outcomes.**

Now, an important feature of these propensities is that they belong to individual coin tosses, not to sequences of coin tosses. Propensities are supposed to be causally responsible for the patterns you see in sequences of coin tosses, but propensities themselves are properties of individual coin tosses.

So on a propensity interpretation, if you toss both of these coins just once, you can say of *this singular event*, *this individual coin toss*, that the unbiased coin has a probability = 0.5 of landing heads, and the biased coin has a probability = 0.75 of landing heads.

Popper and other propensity theorists take it as a major advantage of this approach that it lets us talk about **single-case probabilities**, and it has a theoretical advantage in that it **explains** the long-run frequency behavior of chance setups, rather than just treat them as brute empirical facts, as frequency approaches tend to.

Popper also thought that a propensity interpretation was the **only** way to interpret the probabilities associated with **quantum mechanical properties**, like the decay rates of atoms. He interprets quantum mechanical probabilities as measuring genuine indeterminacies in the world, not just our ignorance of the physical details that actually determine when the atom decays. According to standard interpretations of quantum mechanics, there are no such details, the quantum probabilities represent genuinely indeterministic processes, an objective chanciness in the laws of nature itself. So Popper thinks that the propensity interpretation is the most natural way to interpret these kinds of physical probabilities.

Objections

That's the basic idea behind propensities. As you might suspect by now, this is just the tip of the iceberg. We haven't said anything yet about possible objections to propensity interpretations, or even whether they're a viable interpretation of the probability calculus. Maybe propensities can help us understand objective indeterminacy in the world, but what's the guarantee that they'll obey the mathematical rules of probability theory? And how exactly do propensities relate to relative frequencies? And what exactly *are* propensities, metaphysically speaking?

All of these questions are interesting, and in the decades since Popper introduced this approach various different theories of propensity and objective chance have been developed to help answer these questions. I'm a little hesitant to get into this literature because, (a) it's mostly of philosophical interest — this is something that scientists or statisticians tend not to have much interest in; and (b) I don't want this introduction to be any longer than it has to be. In an introductory classroom discussion I would probably stop right here.

However, since I talked about objections for all the other interpretations I might as well say something about how this approach has been developed and the sorts of challenges it faces.

First of all, there really are **two kinds of propensity theories** in circulation, and these theories differ in how they view the relation between propensities and relative frequencies.

For Popper, for example, the probability of landing a 2 on a dice roll is interpreted as a propensity of a certain kind of repeatable experimental setup — in this case, the dice-rolling setup — to produce a sequence of dice rolls where, in the long-run, the dice lands a 2 with relative frequency 1 in 6.

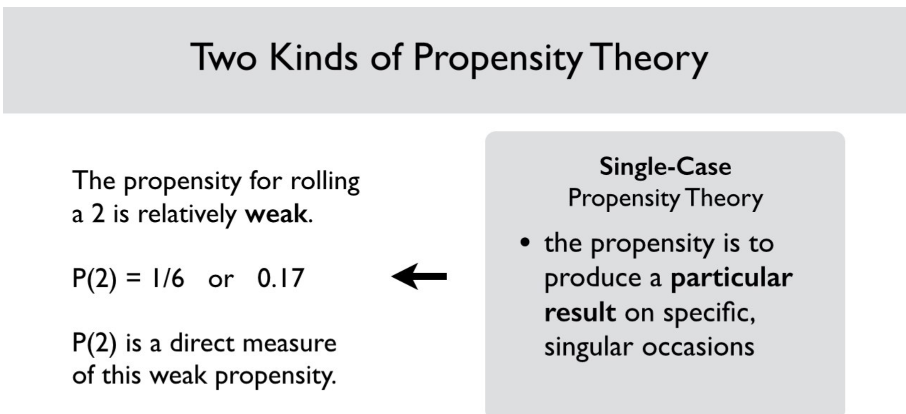
So on this view, propensities are **always associated with long-run relative frequencies**. They're precisely the physical features of the experimental setup that are causally responsible for those long-run frequencies.

So on Popper's view, even though he talks about single-case propensities, these propensities are only defined for single cases that involve some repeatable experimental setup, and the physical property associated with the propensity is defined in terms of its ability to generate these long-run frequencies, if you were to repeat the experiment over and over. Notice that this is not a propensity to produce a particular result on a particular occasion; this is a propensity to produce a sequence of results over repeated trials.

For this reason, some people call this kind of propensity theory a **"long-run" propensity theory**, and they distinguish it from a genuinely **single-case propensity theory**, which treats propensities as propensities to produce particular results on specific, singular occasions.

I know this might seem like just a verbal distinction, but metaphysically the two views really are quite different.

For example, for a single-case propensity theory, the propensity for rolling a 2 on a fair dice roll is relatively weak, it's measured by the ratio of 1 in 6, or about 0.17. That's a low number. The probability is a direct measure of this weak tendency, or propensity, to land '2' on a single dice roll.



For a long-run propensity theorist like Popper, on the other hand, the propensity for rolling a 2 on a dice roll is NOT measured by this low number, it's not identified with the probability of rolling a 2. The propensity for rolling a 2 is a very strong, extremely strong tendency, but not for rolling a 2. The propensity is the tendency of the dice rolling setup to land '2' *with a long-run relative frequency of 1/6*, and THAT is a VERY, VERY strong tendency.

Two Kinds of Propensity Theory

Long-Run
Propensity Theory

- the propensity is to produce a **sequence of results** over repeated trials

→

The propensity is VERY STRONG, but not for rolling a 2.

The propensity is for generating a sequence of trials where the relative frequency of 2 converges on 1/6 in the long run.

We have the same outcome as with the single-case propensity approach — the probability of rolling a '2' is defined as 1/6, but the interpretation of the physical property that is responsible for this outcome is very different.

Now, I draw this distinction because **objections to propensities theories differ between these two types of theory.**

As I mentioned earlier, one concern that all interpretations of probability face is whether they can function as a suitable interpretation of the probability calculus, the mathematical theory of probability.

Long-run propensity approaches tie propensities to relative frequencies, which is good in one sense, since it can piggy-back on the widespread use of relative frequencies in science. But from a foundational standpoint it's not so good, since — as we saw in the tutorial on frequency interpretations — there are reasons to question whether relative frequencies can provide a suitable interpretation of the probability calculus.

With single-case propensities it's even less clear why we should think they would obey the laws of probability theory. Of course if we wanted to we could DEFINE single-case propensities in such a way that they necessarily satisfy the laws, but as Alan Hajek puts it, simply defining what a witch is doesn't show us that witches exist; so simply defining propensities in this way doesn't give us any additional reason to think they exist.

Another class of objections focuses precisely on this question of existence. Unlike relative frequencies or subjective degrees of belief, which aren't metaphysically mysterious to most of us, **it's not at all clear what propensities are, metaphysically speaking**. The closest category we have to describe physical tendencies of things is "dispositions" — certain kinds of physical properties are dispositional properties. Think of a property like "fragility", which we can think of as a disposition to break when subjected to a suitably strong and sudden stress. So maybe a propensity is a probabilistic disposition of some kind. But making this idea clear is more challenging than it looks.

Some people object that, in the absence of a proper metaphysical theory, the term "propensity" is an empty concept. If believing in propensities amounts to nothing more than believing there is SOME property of this dice-rolling setup which entails that the dice will land '2' with a certain long-run frequency, then this is fine as far as it goes, but it doesn't add to our understanding of what generates those frequencies. It's like saying that I understand how it is that birds know how to build nests, by saying that they have an "instinct" for next-building, and defining this instinct as "an innate ability to build nests". This language only tricks us into thinking we understand something when we really don't.

So these are some objections to propensity interpretations. But the story isn't all bad. If you survey the literature you'll see there's been quite a bit of work on propensity interpretations that have been re-branded as theories of "objective chance". Here I'm thinking of work by David Lewis and Isaac Levi and Hugh Mellor and including more recent work by people like Carl Hoefer and Michael Strevins and others. These folks are trying to fit the concept of objective chance into a broader theory of probability that integrates elements of subjective and frequentist approaches, and to show how these various probability concepts are implicitly defined by their relationships to one another. From a philosophical standpoint I think this work is very interesting, but it's still very much a heterogeneous research program with a lot of unresolved problems.

From a critical thinking standpoint, however, I don't think that much of this matters. What matters are the broad distinctions, like the distinction between epistemic probability and physical probability, or the distinction between subjective or Bayesian approaches and frequency approaches.

Critical thinking about probabilities and probabilistic fallacies requires a certain level of basic philosophical literacy, but I don't think it requires anything beyond what we've covered here.