foundationalism

uniformity & circularity

induction

back to Bayes

SHOULD I BELIEVE WHAT SCIENCE TELLS ME?

Last time we introduced the question of what we should believe. We began by considering this rule of belief:

No Argument → No Belief
If you can't give a good argument for P, don't believe P.

But we saw that this led to the result that you should not believe anything at all, which seems very implausible.

At the very least, it seems, we should believe things that we can be certain of:

Certainty → Belief
If you can rule out any situation which would make P false, you should believe P.

Since we can be certain that we exist, and that 2+2=4, this rule says (correctly, it seems) that we should believe these things.

Certainty → Belief
If you can rule out any situation which would make P false, you should believe P.

Since we can be certain that we exist, and that 2+2=4, this rule says (correctly, it seems) that we should believe these things.

We then considered Descartes' idea that we should believe **only the things we can be certain of:**

Doubt → No Belief
If you cannot rule out a situation which would make P false, you should not believe P.

We encountered a challenge to that rule at the end of last time. The challenge was that that rule seems to imply that I should not believe that I have hands; but surely I should be more secure in this belief than in the principle No Doubt → No Belief.

Certainty → Belief
If you can rule out any situation which would make P false, you should believe P.

We encountered a challenge to that rule at the end of last time. The challenge was that that rule seems to imply that I should not believe that I have hands; but surely I should be more secure in this belief than in the principle No Doubt → No Belief.

Suppose, for now, that we accept this. What rule of belief would explain the fact that I should have this belief?

A plausible suggestion would seem to be:

Experience → Belief
If your sense experience tells you that P, and you have no reason to think that your sense experience is misleading, you should believe P.

After all, the reason why I should believe that I have hands, it seems, is that my experience tells me that I do.

Certainty → Belief
If you can rule out any situation which would make P false, you should believe P.

Experience → Belief
If your sense experience tells you that P, and you have no reason to think that your sense experience is misleading, you should believe P.

After all, the reason why I should believe that I have hands, it seems, is that my experience tells me that I do.

Last time we also encountered one other positive rule of belief:

Good Argument → Belief
If there is a valid argument for P and you should believe each of that argument's premises, you should believe P.

Where a "good argument" in the relevant sense is a valid argument whose premises you should believe.

Certainty → Belief
If you can rule out any situation which would make P false, you should believe P.

Experience → Belief
If your sense experience tells you that P, and you have no reason to think that your sense experience is misleading, you should believe P.

Good Argument → Belief
If there is a valid argument for P and you should believe each of that argument's premises, you should believe P.

But this still leaves us without any plausible negative rules of belief: rules which tell us when to discard beliefs that we already have. And it seems like there must be some rules of this kind, since it seems like one thing we ought to be able to do is to examine our beliefs to see which ones we should discard.

But the above suggests a plausible candidate for such a rule. Maybe I should believe what (i) I can be certain of, (ii) what my senses tell me, and (iii) what I can argue for on the basis of (i) and (ii), and **that's all**.

This is a version of a view known as **foundationalism**.

Certainty → Belief
If you can rule out any situation which would make P false, you should believe P.

Experience → Belief
If your sense experience tells you that P, and you have no reason to think that your sense experience is misleading, you should believe P.

Good Argument → Belief
If there is a valid argument for P and you should believe each of that argument's premises, you should believe P.

This is a version of a view known as **foundationalism**.

It can be summed up with the following negative rule of belief:

No Foundations → No Belief
If you can't be certain that P and your senses don't tell you that P and you can't give a good argument for P, you should not believe P.

Last time we introduced the idea of a **basic belief**. Foundationalism is the view that the only basic beliefs you should have are the ones you can be certain of and ones which your senses tell you are true.

Let's now consider a belief that, I presume, all of us have:

`The sun will come up tomorrow.`

Do my senses tell me that this claim is true?

Remember that a claim you can be **certain** of is one whose falsehood you can rule out, just on the basis of thinking about it. Can I be certain that the sun will come up tomorrow?

It follows that, if Foundationalism is true, I must be able to give a good argument for it. What might the premises of this argument be?

Presumably claims like these:

`Yesterday morning, the sun came up.`

`Two mornings ago, the sun came up.`

`Three mornings ago, the sun came up.`

And so on. Let's have a look at the argument that results.

```
1. Yesterday morning, the sun came up.
2. Two mornings ago, the sun came up.
3. Three mornings ago, the sun came up.
......
N. N days ago, the sun came up.
----------------------------
C. The sun will come up tomorrow. (1-N)
```

Remember that a good argument is a valid argument whose premises you should believe. It is plausible that you should believe each of the premises of this argument. But is it valid?

Can you think of any premise which we can add to the argument which would make the argument valid?

Here's a natural choice:

```
If on every past morning the sun
   came up, then tomorrow morning
   the sun will come up.
```

```
1. Yesterday morning, the sun came up.
2. Two mornings ago, the sun came up.
3. Three mornings ago, the sun came up.
......
N. N days ago, the sun came up.
N+1. If on every past morning the sun came up,
   then tomorrow morning the sun will come up.
-----------------------------
C. The sun will come up tomorrow. (1-N+1)
```

Is this argument valid?

This looks like progress. If we should believe all of the premises of this argument, then it looks like we have an explanation of why we should believe the conclusion.

We already have an explanation of why we should believe premises 1-N. What about premise N+1?

Do my senses tell me that it is true? Can I be certain that it is true?

Then it seems that, if Foundationalism is true, I must have a good argument for it.

N+1. If on every past morning the sun came up, then tomorrow morning the sun will come up.

N+1 is an instance of a more general claim, which is sometimes called the principle of the uniformity of nature:

The Uniformity of Nature
The future will be like the past

It seems as though, if we should believe in the Uniformity of Nature, we should believe N+1. So the basic question is whether we should believe in the Uniformity of Nature. As with N+1, the Principle of the Uniformity of Nature is not a claim we can be certain of, and is not a claim my senses tell me to be true. So we have to ask how we might argue for it.

Well, why do we believe in the Uniformity of Nature? Simply because, in the past, the future has always been like the past. Yesterday the future was like the past. And the same for the day before that. And this suggests an argument for the Uniformity of Nature:

```
1. Yesterday, the future was like the past.
2. The day before yesterday, the future was
   like the past.
3. The day before the day before yesterday,
   the future was like the past.
....
N. N days ago, the future was like the past.
-----------------
C. Today, the future will be like the past.
   (1-N)
```

Is this argument valid?

What extra premise would make the argument valid?

It is hard to see how we could make the argument valid without adding a premise which was just a restatement of the very claim — the Uniformity of Nature — which we were trying to prove.

The problem we are encountering here is very general. For consider that science characteristically tells us things which go beyond what our senses tell us. For example:

```
All massive objects
attract one another.
```

```
Every 24 hours, the earth
rotates on its axis.
```

These claims are not, on a natural interpretation, claims which we can know to be true directly on the basis of sense experience: for example, though we can observe some massive objects attracting each other, we certainly have not observed this of all presently existing massive bodies, let alone all massive bodies past and future. These claims are **generalizations**.

Much of what science tells us is a matter of generalizations. Other things that science tells us are based on generalizations. An example might be

```
Halley's comet will next be visible
from earth in 2061.
```

This is not itself a generalization; but our knowledge of it depends on our accepting certain generalizations about the movement of celestial bodies.

All massive objects attract one another.

Every 24 hours, the earth rotates on its axis.

Halley's comet will next be visible from earth in 2061.

It is very plausible that these are all claims which we should believe. But we have at present no explanation of this fact. After all, the only arguments we can give for these claims will be arguments like our argument that the sun will come up tomorrow. And these arguments are invalid, and hence not good arguments. This seems to show that Foundationalism, as we have formulated it, is false.

It also raises an important question. We think that at least some scientific theories are theories we should believe, and that other generalizations are not. Consider, for example,

People born under the sign of Taurus are more likely to be stubborn than people born under other signs.

It also raises an important question. We think that at least some scientific theories are theories we should believe, and that other generalizations are not. Consider, for example,

```
People born under the sign of Taurus
are more likely to be stubborn than
   people born under other signs.
```

This is is a prediction of the theory of astrology. One shouldn't believe this prediction because one shouldn't believe the theory. But why not?

It is tempting to say something like: there's no good argument for astrology; there's no reason to think that it is true.

The problem is that we have just seen that there's also no good argument for the view that the sun will come up tomorrow. So what's the difference? What's the difference between the scientific theories we should believe and other generalizations we should not believe?

The problem is that we have just seen that there's also no good argument for the view that the sun will come up tomorrow. So what's the difference? What's the difference between the scientific theories we should believe and other generalizations we should not believe?

There is an obvious answer to this question. We should believe the theories that are well-supported by the evidence, and should not believe theories that are not well-supported by the evidence.

That leads to our central question: what does it mean for a theory to be **well-supported** by the evidence?

That leads to our central question: what does it mean for a theory to be **well-supported** by the evidence?

The examples we have already discussed give us a plausible answer to this question. Scientific theories typically involve certain generalizations. In the simplest case, they will be claims of the form

All F's are G.

These are not claims which our senses can tell us to be true. But our senses can tell us that claims like this are true:

This particular thing is an F, and it is G.

Let's call claims which are related in this way to generalizations **instances** of the generalization.

Let's call claims which are related in this way to generalizations **instances** of the generalization.

Then we might say that a generalization is well-supported by the evidence just in case the following two conditions are satisfied:

Our senses tell us that many instances of the generalization are true.

Our senses don't tell us that any instances of the generalization are false.

This fits many of the examples we have discussed very well. Reasoning in which one proceeds from a bunch of instances of a generalization to believing that generalization is often called **inductive reasoning.** So we might state our proposed rule of belief as follows:

Induction → Belief

If you know many true instances of a generalization P, and don't know of any false instances of P, you should believe P.

This fits many of the examples we have discussed very well. Reasoning in which one proceeds from a bunch of instances of a generalization to believing that generalization is often called **inductive reasoning.** So we might state our proposed rule of belief as follows:

> **Induction → Belief**
> If you know many true instances of a generalization P, and don't know of any false instances of P, you should believe P.

This looks like a good way to explain the differences between scientific claims we should believe and, e.g., generalizations from astrology.

It is obviously somewhat vague; we have not spelled out what "many" amounts to. We can ignore this issue for now. A plausible thought would be that the more instances you come to know, the more confident you should come to be in the generalization.

Many people think that forming beliefs via a distinctive 'scientific method' is a good way to form beliefs. Forming beliefs on the basis of Induction → Belief would appear to be a reasonable interpretation of at least part of what this method might involve.

Induction → Belief

If you know many true instances of a generalization P, and don't know of any false instances of P, you should believe P.

I want to now look at two problems for Induction → Belief.

The first is called **the paradox of the ravens**. Consider the following generalization:

```
All ravens are black.
```

Now notice that this generalization is equivalent to this one:

```
All non-black things are non-
ravens.
```

If you think about it for a second, you can see that if one of these is true, the other must be as well.

All ravens are black.

All non-black things are non-ravens.

If you think about it for a second, you can see that if one of these is true, the other must be as well.

So, it seems very plausible that a piece of evidence supports one just in case it supports the other, and to just the same degree.

Let's now consider two investigations that I could undertake. Here's the first:

I go out looking for ravens. I find 10 of them, and they are all black.

It looks like this provides inductive support for the generalization that all ravens are black (and so also for the other generalization). So, according to Induction → Belief, you should increase your confidence in those generalizations.

All ravens are black.

All non-black things are non-ravens.

Here's a second investigation I could undertake:

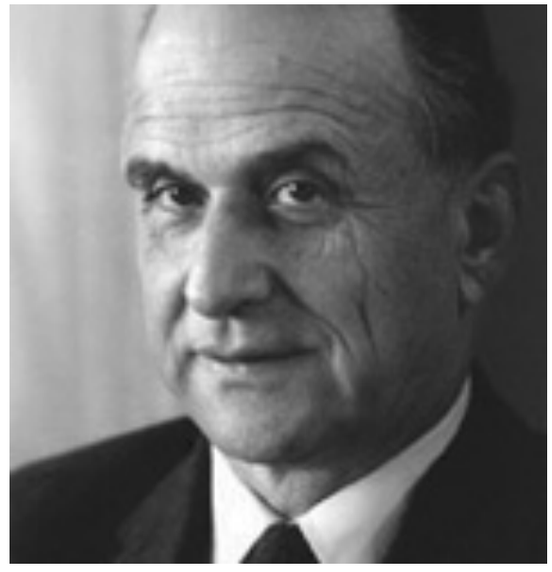
> I begin to investigate my immediate environment. I check the first ten non-black things I can find — and it turns out that none of them are ravens.

Here, as in the previous generalization, I have found ten true instances of one of the two generalizations. So it looks as though if Induction → Belief is true, I should substantially increase my confidence in the claim that all ravens are black.

But intuitively, this is bizarre. Surely the fact that a bunch of non-black things in my environment are also non-ravens should do nothing, or almost nothing, to support the generalization that all ravens are black.

Let's turn now to a different kind of worry about Induction → Belief.

This challenge is due to Nelson Goodman, one of the most important American philosophers of the 20th century.



Goodman's aim in his book *Fact, Fiction, and Forecast* was to show that rules of belief like ours are false; he did by defining a made up word, "grue," as follows:

x is **grue** if and only if either:
(i) x is green, and has been observed before 2021, or
(ii) x is blue, and has not been observed before 2021.

x is **grue** if and only if either:
(i) x is green, and has been observed before 2021, or
(ii) x is blue, and has not been observed before 2021.

It is important to see, first, that this is a perfectly legitimate definition; it succeeds in classifying all objects as either grue or non-grue.

But suppose that we enumerate all of the emeralds which have been observed so far, and consider the following pieces of data:

```
Emeralds first observed in 2021 were grue.
Emeralds first observed in 2020 were grue.
Emeralds first observed in 2019 were grue.
...
```

You have, in short, knowledge of the truth of a great many instances of the generalization

```
All emeralds are grue.
```

```
Emeralds first observed in 2021 were grue.
Emeralds first observed in 2020 were grue.
Emeralds first observed in 2019 were grue.
...
```

You have, in short, knowledge of the truth of a great many instances of the generalization

```
All emeralds are grue.
```

Further, you know of no false instances of this generalization. So, because you endorse Induction → Belief, you come to believe that the generalization is true.

It is now January 1, 2022, and you decide to go emerald hunting. Of course, since you believe that all emeralds are grue, you believe that the first emerald you discover will be grue. Since this emerald will not have been observed before 2022, this means that you believe that the first emerald you discover will be blue.

Is it reasonable for you to believe this? Of course not! But then it looks like Induction → Belief must be false.

A very natural reaction is: this is a silly example! It would be crazy just to throw out all inductive reasoning on the basis of "grue."

Perhaps what we need to do is to restrict the cases of induction that we use to avoid annoying examples like "grue;" a natural thought is that we should restrict them to cases in which only suitable scientific vocabulary is used. (Words like "grue" that we want to rule out are sometimes called "gruesome predicates.")

To pursue this thought, we need to be able to say what a gruesome predicate is - that is, we need to be able to say what, exactly, is so bad about "grue." This turns out to be harder than you might think.

A first thought is that the problem is due to "grue" being a made-up word. But this won't get us very far — after all, scientific theories introduce new scientific terms all the time, and these are "made up" in just the way that "grue" is — they are new terms defined in terms of existing vocabulary. At one time, "electron" was made up.

A more promising idea is that the problem with "grue" is that it is defined in terms of a particular threshold time. Maybe in inductive reasoning we can't use words defined in terms of a certain threshold.

A first thought is that the problem is due to "grue" being a made-up word. But this won't get us very far — after all, scientific theories introduce new scientific terms all the time, and these are "made up" in just the way that "grue" is — they are new terms defined in terms of existing vocabulary. At one time, "electron" was made up.

A more promising idea is that the problem with "grue" is that it is defined in terms of a particular threshold time. Maybe in inductive reasoning we can't use words defined in terms of a certain threshold.

The problem with this is that sometimes we want to scientifically investigate a class of things defined using a certain threshold. For example, it might be that water molecules very differently when they are at temperatures less than 0 degrees Celsius than when they are above this temperature. The category 'H2O molecules below 0 degrees Celsius' is defined in terms of a certain threshold. But surely I can pursue scientific reasoning about this category. How is this different than 'grue'?

Let's look at a different idea, which involves a more radical revision of Induction → Belief.

Consider again the case of grue. Why are we sure that the first emerald observed in 2022 will be green, rather than grue, if every emerald observed so far has been both green and grue?

One idea is that the difference comes from certain **background beliefs** we have. For example, I believe that what color something is does not depend on the time at which it was first observed. So, I think that if a certain emerald which was in fact first observed in 2021 (and is green) were to have been first observed in 2022, it would still have been green. But of course that would make it non-grue.

Perhaps, then, the question of whether some evidence supports a given theory is always relative to our background beliefs. This would make induction very different than deduction. Whether an argument is valid does **not** depend on one's background beliefs; it either is valid, or it isn't. But maybe the question of whether some evidence supports a theory is not like that.

Perhaps, then, the question of whether some evidence supports a given theory is always relative to our background beliefs. This would make induction very different than deduction. Whether an argument is valid does **not** depend on one's background beliefs; it either is valid, or it isn't. But maybe the question of whether some evidence supports a theory is not like that.

Indeed, we have already encountered a theory which provides one way of making this explicit: the theory that we should form beliefs on the basis of Bayes' theorem.

Bayes' theorem

$$P(h|e) = \frac{P(h) * P(e|h)}{P(e)}$$

How, according to Bayes' theorem, should we adjust our beliefs in generalizations on the basis of new evidence? Well, we just have to know how the prior probabilities of the theory and the evidence, and how likely the theory says that the evidence is — and then we do the math.

Bayes' theorem

$$P(h|e) = \frac{P(h)*P(e|h)}{P(e)}$$

How, according to Bayes' theorem, should we adjust our beliefs in generalizations on the basis of new evidence? Well, we just have to know how the prior probabilities of the theory and the evidence, and how likely the theory says that the evidence is — and then we do the math.

Let's look at what this says about the case of the ravens. Let the hypothesis be the claim that all ravens are black, and let the evidence be that a randomly selected raven is black. Then the probability of the evidence given the hypothesis is 1. Let's suppose that the prior probability of the generalization being true was 0.1, and that the prior probability of a randomly selected raven being black was 0.2. Then the probability of the hypothesis given our new evidence is (0.1 * 1) / 0.2 = 0.5. So, the probability of the theory goes up, given our new evidence.

With each new black raven, the probability of the theory will increase.

Bayes' theorem

$$P(h|e) = \frac{P(h) * P(e|h)}{P(e)}$$

Let's look at what this says about the case of the ravens. Let the hypothesis be the claim that all ravens are black, and let the evidence be that a randomly selected raven is black. Then the probability of the evidence given the hypothesis is 1. Let's suppose that the prior probability of the generalization being true was 0.1, and that the prior probability of a randomly selected raven being black was 0.2. Then the probability of the hypothesis given our new evidence is (0.1 * 1) / 0.2 = 0.5. So, the probability of the theory goes up, given our new evidence.

Now let the evidence be that a randomly selected non-raven thing is non-black. Again let's suppose that the prior probability of the hypothesis is 0.1. But the likelihood of a randomly selected non-raven being non-black is independent of whether all ravens are black. So here P(e|h) = P(e), and the probability of the hypothesis is unchanged. That is what we want.

How might we write this as a rule of belief?

How might we write this as a rule of belief?

Updating
When you get new evidence, you should revise the probability you assign to hypotheses in accordance with Bayes' Theorem.

We began by asking what justifies us in believing, for example, that the sun will come up tomorrow. We found that it was impossible to give valid arguments using premises which we can be certain of or which sense experience tells us are true for beliefs of this kind. This pushed us toward Induction → Belief; but we saw that this ran into troubles. Can Updating explain why we should think that it is very likely that the sun will come up tomorrow?

It can — but in a way that leaves certain fundamental questions unanswered.

Let's begin by imagining that you have complete amnesia, and don't know whether the sun comes up every morning, or just some mornings. (You've obviously also forgotten certain facts about how the solar system works.)

Suppose now that you learn that the sun has come up for the last 1000 mornings. The probability of this evidence given the hypothesis is 1. Suppose that your prior probability in the hypothesis that the sun comes up every morning is 0.5. Because it is very unlikely that the sun would have come up 1000 mornings in a row unless it came up every morning, P(e) will be very, very slightly larger than 0.5. So the new probably of the hypothesis will be very close to 1 — which is what we want.

But notice that, even if we assume that we have good reason to believe Bayes' Theorem, in order to derive the result that the probability of the hypothesis is close to 1, we need to make use of two further facts: the prior probability of the evidence, and the prior probability of the hypothesis. These are just facts about what the subject believes; they are, to a first approximation, beliefs about probabilities. How do we tell whether **these** are beliefs that the subject should have? After all, they are not given by sense experience, and they are not the sort of thing one can be certain about.

But notice that, even if we assume that we have good reason to believe Bayes' Theorem, in order to derive the result that the probability of the hypothesis is close to 1, we need to make use of two further facts: the prior probability of the evidence, and the prior probability of the hypothesis. These are just facts about what the subject believes; they are, to a first approximation, beliefs about probabilities. How do we tell whether **these** are beliefs that the subject should have? After all, they are not given by sense experience, and they are not the sort of thing one can be certain about.

But then there is a sense in which the appeal to Bayes' theorem to explain why we should have certain beliefs about (e.g.) scientific theories is cheating. In order to get the results we want, we have to make use of certain other beliefs. But we have been given no reason to think that those are beliefs we should have.

So there is a sense in which we are stuck with the same problem with which we started: the problem of explaining which basic beliefs — other than those based on sense experience and those we can be certain of — that we should have.

There is a sense in which the move from a rule like Induction → Belief to a rule like Updating gives up on the idea of trying to find a 'logic' of scientific discovery. The idea behind rules like Induction → Belief is that, given some data, we can figure out what theory is most likely to be true.

If Induction → Belief is false and something like Updating is the best we can do, then this is impossible. On this view, there is no escaping the relativity of theory confirmation to background beliefs.

The problem in the background is that this leaves us without an explanation of the difference between (intuitively) well-supported scientific theories and (intuitively) very irrational beliefs. It looks like, for virtually any bizarre hypothesis, we can imagine someone with prior beliefs which are such that Updating would lead them to believe that hypothesis. Could that really be enough to make that a belief that they should have?