

## 4 Further Constraints on Rational Belief

### 4.1 Belief and perception

We have looked at two assumptions about rational belief. The first, the MEU Principle, relates an agent's beliefs to her desires and choices. The second, probabilism, imposes an internal, structural constraint on rational beliefs: that they conform to the rules of probability. There is more.

#### **Example 4.1 (The Litmus Test)**

You are unsure whether a certain liquid is acidic. Remembering that acid turns litmus paper red, you dip a piece of litmus paper into the liquid. The paper turns red.

Seeing the red paper should increase your confidence that the liquid is acidic. But as far as probabilism and the MEU Principle are concerned, you could just as well remain unsure whether the liquid is acidic or even become certain that it is *not* acidic, as long as your new credences are probabilistic and your choices maximize expected utility (by the light of your beliefs and desires).

So there are further norms on rational belief. In particular, there are norms on how beliefs change in response to perceptual experience. Like the MEU Principle, and unlike probabilism, these norms state a connection between beliefs and something other than belief – here, perceptual experience. Loosely speaking, the MEU Principle describes the causal “output” of beliefs: the effects an agent's beliefs have on her behaviour. Now we turn to the “input” side. We want to know what sorts of experiences might cause a rational agent to have such-and-such beliefs.

To state a connection between perceptual experience and belief, we need a way to identify kinds of perceptual experience. How do we do that?

We could try to identify the experiences by their phenomenology, by “what it’s like” to have the experience. But there is no canonical standard for expressing phenomenal qualities. Besides, we may want our norm to handle unconscious perceptions and the perceptions of artificial agents for whom it is doubtful whether they have any phenomenal experience.

We could alternatively identify perceptions by their physiology, by the neurochemical or electrical events that take place in the agent’s sense organs. But that would go against the spirit of our general approach, which is to single out high-level patterns and remain neutral on details of biological or electrical implementation.

The usual strategy is to identify perceptual experiences by the information they provide to the agent’s belief system. In the Litmus Test, for example, we might assume that the information you receive from your visual system is that the paper has turned red. You don’t directly receive the information that the liquid is acidic. This is something you infer from the experience with the help of your background beliefs.

In the simplest and best known version of this model, we assume that the information conveyed to an agent by their perceptual experiences can always be captured by a single proposition of which the agent becomes certain. The model can be extended to allow for cases in which the perceptual information is uncertain and equivocal, but we will stick to the simplest version.

## 4.2 Conditionalization

Suppose a perceptual experience provides an agent with some information  $E$  (for “evidence”). How should the rest of the agent’s beliefs change to take into account the new information?

Return to the Litmus Test. Let  $Cr_{old}$  be your credence function before you dipped the paper into the liquid, and  $Cr_{new}$  your credence function when you see the paper turn red. If you are fairly confident that red litmus paper indicates acidity, you will also be confident, before dipping the paper, that your liquid is acidic *on the supposition that* the paper will turn red. Your initial degrees of belief might have been as follows.

$$\text{Cr}_{\text{old}}(\textit{Acid}) = 1/2.$$

$$\text{Cr}_{\text{old}}(\textit{Acid}/\textit{Red}) = 9/10.$$

What is your new credence in *Acid*, once you learn that the paper has turned red? Plausibly, it should be 9/10. Your previous conditional credence in *Acid* given *Red* should turn into your new unconditional credence in *Acid*.

This kind of belief change is called **conditionalization**. We say that you conditionalize **on** the information *Red*. Let's formulate the general rule.

#### The Principle of Conditionalization

Upon receiving information *E*, a rational agent's new credence in any proposition *A* equals her previous credence in *A* conditional on *E*:

$$\text{Cr}_{\text{new}}(A) = \text{Cr}_{\text{old}}(A/E).$$

Here it is understood that the agent's experience leaves no room for doubts about *E*, and that *E* is the *total* information the agent acquires, rather than part of her new information. If you see the paper turn red but at the same time notice a whiff of ammonium hydroxide, which you know is alkaline, your credence in the *Acid* hypothesis may not increase to 0.9.

#### Exercise 4.1 †

Assume  $\text{Cr}_{\text{old}}(\textit{Snow}) = 0.3$ ,  $\text{Cr}_{\text{old}}(\textit{Wind}) = 0.6$ , and  $\text{Cr}_{\text{old}}(\textit{Snow} \wedge \textit{Wind}) = 0.2$ . By the Principle of Conditionalization, what is  $\text{Cr}_{\text{new}}(\textit{Wind})$  if the agent finds out that it is snowing?

#### Exercise 4.2 ††

Show that conditionalizing first on  $E_1$  and then on  $E_2$  is equivalent to conditionalizing in one step on  $E_1 \wedge E_2$ . That is, if  $\text{Cr}_1$  results from  $\text{Cr}_0$  by conditionalising on  $E_1$ , and  $\text{Cr}_2$  results from  $\text{Cr}_1$  by conditionalizing on  $E_2$ , then for any *A*,  $\text{Cr}_2(A) = \text{Cr}_0(A/E_1 \wedge E_2)$ . (You may assume that  $\text{Cr}_0(E_1 \wedge E_2) > 0$ .)

**Exercise 4.3 †††**

Assume that  $Cr_{\text{new}}$  results from  $Cr_{\text{old}}$  by conditionalizing on some information  $E$  with  $Cr_{\text{old}}(E) > 0$ , and that  $Cr_{\text{old}}$  satisfies the Kolmogorov axioms. Using the probability rules, show that  $Cr_{\text{new}}$  also satisfies the Kolmogorov axioms. (You may use any of the derived rules from chapter 2. Hint for axiom (ii): if  $A$  is logically necessary, then  $A \wedge E$  is logically equivalent to  $E$ .)

When computing  $Cr_{\text{new}}(A)$ , it is often helpful to expand  $Cr_{\text{old}}(A/E)$  with the help of Bayes' Theorem. The Principle of Conditionalization then turns into the following (equivalent) norm, known as **Bayes' Rule**:

$$Cr_{\text{new}}(A) = \frac{Cr_{\text{old}}(E/A) \cdot Cr_{\text{old}}(A)}{Cr_{\text{old}}(E)}, \text{ provided } Cr_{\text{old}}(E) > 0.$$

This formulation is useful because it is often easier to evaluate  $Cr_{\text{old}}(E/A)$ , the probability of the evidence  $E$  conditional on some hypothesis  $A$ , than to evaluate  $Cr_{\text{old}}(A/E)$ , the probability of the hypothesis conditional on the evidence.

Here is an example.

**Example 4.2**

2% of women in a certain population have breast cancer. A test is developed that correctly detects 95% of cancer cases but also gives a positive result in 10% of non-cancer cases. A woman from the population comes into your practice, takes the test, and gets a positive result. How confident should you be that the woman has breast cancer?

We assume that you are aware of all the statistical facts before you learn the test result. Knowing that the woman is from a population in which 2% of women have breast cancer, your initial credence in the hypothesis, call it  $C$ , that the woman has cancer should plausibly be 0.02. So we have

$$Cr_{\text{old}}(C) = 0.02.$$

Since you know that the test yields a positive result in 95% of cancer cases, we also

have

$$\text{Cr}_{\text{old}}(P/C) = 0.95,$$

where  $P$  says that the test result is positive. Similarly, since the test yields a positive result in 10% of non-cancer cases, we have

$$\text{Cr}_{\text{old}}(P/\neg C) = 0.1.$$

Now we simply plug these numbers into Bayes' Rule, expanding the denominator by the Law of Total Probability:

$$\begin{aligned} \text{Cr}_{\text{new}}(C) &= \frac{\text{Cr}_{\text{old}}(P/C) \cdot \text{Cr}_{\text{old}}(C)}{\text{Cr}_{\text{old}}(P/C) \cdot \text{Cr}_{\text{old}}(C) + \text{Cr}_{\text{old}}(P/\neg C) \cdot \text{Cr}_{\text{old}}(\neg C)} \\ &= \frac{0.95 \cdot 0.02}{0.95 \cdot 0.02 + 0.1 \cdot 0.98} = \frac{0.019}{0.019 + 0.098} = 0.16. \end{aligned}$$

After the positive test, your degree of belief that the woman has breast cancer should be 0.16. This is lower than many people initially think – including many trained physicians. But it makes sense. Imagine we took a sample of 1000 women from the population. We would expect around 2%, or 20 women, in the sample to have breast cancer. If we tested all women in the sample, we would expect around 95% of those with cancer to test positive. That's 95% of 20 = 19 women. Of the 980 women without cancer, we would expect around 10% = 98 to test positive. The total number of positive tests would be around 19 + 98 = 117. Of these 117 women, 19 actually have cancer. So the chance that an arbitrary woman who tests positive has cancer is 19/117 = 0.16. If you look back at the above application of Bayes' Theorem, you can see that it resembles this statistical line of reasoning.

The tendency to overestimate (or underestimate) probabilities in cases like example 4.2 is known as the **base rate fallacy**, because it is assumed to arise from neglecting the low “base rate” of 2%.

**Exercise 4.4 ††**

Box A contains two black balls. Box B contains one black ball and one white ball. I choose a box at random and blindly draw a ball. The ball is black. How confident should you be that I have chosen box A?

**Exercise 4.5 (The Prosecutor's Fallacy) †††**

A murder has been committed on an island with a million inhabitants. In a database of blood donors, detectives find a record whose DNA seems to match the perpetrator's DNA from the crime scene. The DNA test is very reliable: the probability that it finds a match between distinct people is 1 in 100,000. The person with the matching DNA is arrested and brought to court. The prosecutor argues that the probability that the defendant is innocent is 1/100,000. Is this true? As a member of the jury, how confident should you be in the defendant's guilt?

### 4.3 Induction and Indifference

Suppose an agent's beliefs are probabilistic and change by conditionalization. Does this ensure that the beliefs are reasonable? No. If the agent starts out with sufficiently crazy beliefs, conditionalization will not make them sane.

**Example 4.3**

You are stranded on a remote island, which you find inhabited by a strange kind of flightless bird. During the first ten days of your stay, you see 100 birds, all of which are green.

You should be fairly confident that the next bird will also be green. The Principle of Conditionalization does not ensure this. It might even make you confident that the next bird is pink. For suppose you were born with a firm conviction that if you are ever going to see 100 green birds on an island, then the next bird you would see is pink. Your observation of 100 green birds does not challenge this conviction. After conditionalizing on your observation of the 100 green birds, you would become confident that the next bird you will encounter is pink.

What we see here is Hume's problem of induction. As Hume pointed out, there is no logical guarantee that the future will resemble the past, or that the unobserved parts of the world resemble the observed. The colour of the 101st bird is not entailed by the colour of the first 100 birds. To infer that the 101st bird is probably green we need a further premise about the "uniformity of nature". Roughly, we need to assume

that regularities in the part of the world that we have observed up to some time are likely to extend into the unobserved part of the world. If, for example, the first 100 birds we encounter on an island are all green, then other birds on the island are probably green as well. This assumption may be supported by earlier experiences. But, again, it won't be *entailed* by these experiences. Ultimately, some such premise must be accepted as bedrock.

In Bayesian terms, the problem of induction suggests that we have to put restrictions on what an agent may believe *without any relevant evidence*. Scientifically minded people sometimes feel uneasy about such restrictions, and therefore speak about the **problem of the priors**. An agent's **priors** (or "ultimate priors" or "ur-priors") are her credences at the start of her epistemic journey, before she condition-alizes on any evidence.

What should an agent believe, at the beginning of her epistemic journey? It would be irrational to be convinced, without any evidence, that the first 100 birds one might encounter on an island will be atypical in colour. Indeed, a natural thought is that without any relevant evidence, one should not be convinced of anything (except logical truths). One should be open-minded, dividing one's credence evenly between all ways the world might be:

**The (naive) Principle of Indifference**

If  $A_1, \dots, A_n$  are  $n$  propositions exactly one of which must be true, then a rational prior credence function assigns the same probability  $1/n$  to each of these propositions.

This, however, can't be right. Suppose you have no information about the colour of my hat. Here are two possibilities:

$R$ : The hat is red.

$\neg R$ : The hat is not red.

Exactly one of these must be true. By the naive Principle of Indifference, you should give credence  $1/2$  to  $R$  and  $1/2$  to  $\neg R$ . But we can also divide  $\neg R$  into several possibilities:

$R$ : The hat is red.

- B*: The hat is blue.
- G*: The hat is green.
- Y*: The hat is yellow.
- O*: The hat has some other colour.

By the naive Principle of Indifference, you should give credence  $1/5$  to each of these possibilities. The Principle entails that your credence in *R* should be  $1/2$  and also that it should be  $1/5$ !

Some have concluded that in cases like these, rationality really does require you to have *multiple* credence functions: relative to one of your credence functions, *R* has probability  $1/2$ , relative to another, it has probability  $1/5$ . I'll set this view aside for now, but we will briefly return to it in section 11.5.

A more plausible response is to restrict the propositions  $A_1, \dots, A_n$  to which the requirement of indifference applies. Intuitively, you shouldn't be indifferent between *R* and  $\neg R$  because these two propositions are not on a par. There are more ways of being non-red than of being red. Unfortunately, it is hard to turn this intuition into a consistent general rule, as the following exercise illustrates.

**Exercise 4.6 ††**

I have a wooden cube in my office whose side length is at least 2 cm and at most 4 cm. That's all you know about the cube. We can distinguish two possibilities:

- S*: The cube's side length is between 2 cm and 3 cm (excluding 3).
- L*: The cube's side length is between 3 cm and 4 cm.

The intervals have the same length, so *S* and *L* are intuitively on a par. This suggests that you should give credence  $1/2$  to each of *S* and *L*. But now observe that if a cube has side length  $x$ , then the cube's volume is  $x^3$ .

- (a) Can you restate the propositions *S* and *L* in terms of volume?
- (b) What credence do you give to *S* and *L* if you treat equally sized ranges of volume as equally likely?

There is another problem with indifference principles. Let's imagine we've found a rule for when two propositions are "on a par" so that we can consistently require



an agent's priors to be indifferent between propositions that are on a par. We should still be cautious about endorsing the requirement, for is likely to clash with the "uniformity of nature" assumption required for inductive inference.

Return to example 4.3. Assume, for simplicity, that birds can only be green or red. There are then four possibilities regarding the first two birds you might see:

*GG*: Both birds are green.

*GR*: The first bird is green, the second red.

*RG*: The first bird is red, the second green.

*RR*: Both birds are red.

Intuitively, these four possibilities are on a par. An indifference principle might say that you should give credence  $1/4$  to each.

Now what happens when you see the first bird, which is green? Your evidence rules out *RG* and *RR*. If you conditionalize on your evidence, your new credence will be divided evenly between the remaining possibilities *GG* and *GR* (as you may check). Your credence that the next bird is green will be  $1/2$ . By the same reasoning, if your prior credence is evenly divided between all possible colour distributions among the first three birds (*GGG*, *GGR*, etc.), then after having seen two green birds, your "posterior" credence that the next (fourth) bird is green will still be  $1/2$ . And so on. No matter how many green birds you see, you won't think that this tells you anything about the next bird.

If we want an observation of 100 green birds to raise your credence in the next bird being green, we have to assume that your prior credence in the "uniform" hypothesis *GGGGG.....GGGGGG* (that's 101 'G's) should be greater than your prior credence in the "non-uniform" hypothesis *GGGGG.....GGGGGR*.

Intuitively, the problem is that there are at least as many irregular worlds as regular worlds. If you spread your credence evenly over all ways the world might be, you'll end up giving too much credence to irregular worlds. You won't be able to learn by induction.

Rational priors should be open-minded, but biased towards regular worlds. There is no agreement on how to make this precise. (We will meet an intriguing partial answer in the following section.) As such, the "problem of the priors" remains open.

## 4.4 Probability coordination

We turn from the highly controversial Principle of Indifference to another norm that is almost universally accepted among Bayesians. This norm connects subjective probability with objective probability, and is often expressed as a norm on priors.

### The Probability Coordination Principle

An agent's prior credence in a proposition  $A$ , on the supposition that the objective probability of  $A$  is  $x$ , should equal  $x$ :

$$\text{Cr}_0(A/\text{Pr}(A) = x) = x.$$

Here,  $\text{Cr}_0$  is a rational prior credence function, and  $\text{Pr}$  is any kind of objective probability, such as relative frequency or quantum physical chance.

The Probability Coordination Principle implies that if a rational agent has discovered the objective probabilities – if she has conditionalized on  $\text{Pr}(A) = x$  – and she doesn't have other relevant information about  $A$ , then she will align her degrees of belief with the objective probabilities: her degree of belief in  $A$  will match the known objective probability.

We have unwittingly assumed this all along. In example 4.2, we assumed that if you know that a woman is from a population in which 2% of women have cancer, and you have no other relevant information about her, then your credence that she has cancer should be 0.02. This is not entailed by the Kolmogorov axioms. We need the Probability Coordination Principle to connect your information about relative frequency to your degree of belief.

The Probability Coordination Principle can be used even if the agent doesn't have full information about the objective probabilities. In exercise 4.4, you had to evaluate  $\text{Cr}(\textit{Black}/B)$ , where  $B$  is the hypothesis that I have drawn a ball from a box containing one black ball and one white ball, and  $\textit{Black}$  is the hypothesis that the ball is black. Assuming that the draw is random (in some objective sense),  $B$  entails that  $\text{Pr}(\textit{Black}) = 1/2$ . You don't know whether  $B$  is true, but we can infer, by the Probability Coordination Principle, that  $\text{Cr}(\textit{Black}/B) = 1/2$ .

In 1814, Pierre-Simon Laplace observed that the Probability Coordination Principle may help with Hume's problem of induction. Return to example 4.3, where

you've encountered 100 green birds in the first few days on a remote island. Suppose you think that there's a certain objective probability with which any given bird on the island is green (independently of the other birds). That probability might be 1, in which case all the birds are certain to be green. Or it might be 0. Or it might be anything in between 0 and 1, in which case you would expect to find some red birds and some green birds. Now suppose you start out maximally open-minded about this probability, giving equal credence to all values from 0 to 1. Using the Probability Coordination Principle, one can then show – the maths is beyond what we do in this course – that after observing 100 green birds, your credence that the next bird is green will be around 0.99. You have learned by induction!

In the previous section, we saw that indifference over *outcomes*, over possible sequences of *G* and *R*, makes inductive learning impossible. Laplace saw that indifference over (objective) *probabilities of outcomes* has the opposite effect. By treating the outcomes as independent matters of objective probability, and giving equal credence to the objective probability, you end up giving comparatively low credence to irregular sequences.

You may wonder where the Probability Coordination Principle comes from. Some say it is a basic norm of rationality. Others say that it must follow from more basic norms – from a restricted indifference principle, for example, or even from probabilism alone. The issue turns on deep questions about the nature of objective probability. Those who regard Probability Coordination as basic tend to believe that the ultimate fabric of the physical world includes probabilistic quantities to which rational beliefs should be aligned, for reasons nobody can explain. Those who don't regard Probability Coordination as basic see no need to posit special physical quantities with a mysterious spell on rational credence. On a simple version of the alternative view, objective probability is nothing but relative frequency and the Probability Coordination Principle follows from plausible indifference requirements. We will not look further into these debates.

**Exercise 4.7 †**

Jacob Bernoulli (an uncle of Daniel and Nicolas Bernoulli, who we've met in section 3.4) proposed the following simplified version of the Probability Coordination Principle: *If a proposition has very low objective probability, one may be certain that it is false.* What do you think of this?

## 4.5 Confirmation

An important question both in the philosophy of science and in scientific practice is how scientific hypotheses are confirmed or disconfirmed by empirical data. We can't directly observe that, say, spacetime is curved, that smoking causes cancer, or that dolphins evolved from land animals. Our evidence strongly *supports* these assumptions, but it doesn't entail them. What is this relation of evidential support? What does it take for some evidence to support a hypothesis?

Philosophers have tried to formulate general rules for evidential support, akin to the rules of deductive logic. The following rules, or "conditions" on when a hypothesis is confirmed by evidence, figure in an influential 1945 paper by Carl Hempel.

**Nicod's Condition.** Universal generalisations are confirmed by their instances: an  $F$  that is  $G$  lends support to the hypothesis that all  $F$ s are  $G$ s.

**Converse Consequence Condition.** If some evidence confirms a hypothesis then it also confirms any theory (conjunction of hypotheses) that entails the hypothesis.

**Special Consequence Condition.** If some evidence confirms a theory then it also confirms anything that is entailed by the theory.

This rule-based (or "syntactical") approach didn't work out well. Most rules that initially looked plausible turned out to have clear counterexamples. The few that remained are too weak to make sense of scientific reasoning.

Consider Nicod's Condition. Normally, observation of a black raven lends support to the hypothesis that all ravens are black. But not always. Suppose your friend is on an expedition and you've agreed that if she comes across a white raven then she is going to send you a black raven, by mail, in a cage. One day, a parcel arrives: it's a black raven. In this context, observation of a black raven is strong evidence *against* the hypothesis that all ravens are black.

**Exercise 4.8 †**

Show that the Converse Consequence Condition and the Special Consequence Condition together entail that if some evidence  $E$  confirms some hypothesis  $H$ , then  $E$  also confirms every other hypothesis  $H'$ . (*Hint*: You simply need to find a “theory” that entails both  $H$  and  $H'$ .)

A different kind of approach was suggested by Karl Popper. Popper noticed that although scientific theories are rarely entailed by empirical evidence, they can be *refuted* by the evidence. A single white raven is enough to refute (or “falsify”) the hypothesis that all ravens are black. According to Popper, a theory is confirmed (or “corroborated”, as he preferred to say) to the extent that it has withstood attempts at falsification.

One problem for this **falsificationist** approach is that many scientific theories or hypotheses can’t actually be falsified, because they don’t have directly observable consequences. The (well-confirmed) hypothesis that smoking causes cancer, for example, doesn’t imply that every single smoker gets cancer. It only predicts that smokers have a higher *probability* of getting cancer, in some objective sense of ‘probability’. (The hypothesis is not about what people believe.) We can’t directly observe that probability.

To get around this issue, falsificationism may call upon its powerful ally, “**classical**” (or “**frequentist**”) **statistics**. According to classical statistics, a hypothesis can be rejected not only if it is logically incompatible with the evidence, but also if it renders the evidence *sufficiently improbable*. Imagine, for example, that we randomly divide 1000 children into two groups. One group is instructed to take up smoking, the other to refrain from smoking. In all other respects, we force the two groups to lead similar lives. 50 years later, we find more incidents of cancer in the smoking group than in the “control group”. This could be just a coincidence. The tools of classical statistics allow us to compute the objective probability of the observed difference between the groups *on the assumption that it is a coincidence*. If this probability is sufficiently low, classical statistics tells us that we can reject the coincidence hypothesis. We can infer that smoking really does increase the risk of cancer.

One obvious problem with this move is to explain when a probability is “sufficiently low”. Just how improbable must a hypothesis render the observed evidence

to warrant rejecting the hypothesis? In the social sciences, any probability below 0.05 is usually deemed sufficiently low. In medicine, a threshold of 0.01 is preferred. Either choice looks unprincipled and arbitrary. Besides, what does it mean to “reject” a hypothesis? Should we become absolutely certain that the hypothesis is false – even though we know that low-probability events happen all the time?

Another problem with the frequentist approach is that it is only applicable to specific kinds of data. The cancer experiment I have just described has never been carried out, for obvious ethical and practical reasons. The actual data that support the link between smoking and cancer are, for the most part, of a kind for which the tools of classical statistics aren’t designed because one can’t easily compute informative objective probabilities.

A deeper problem with the falsificationist/frequentist approach is that predictive success is not the only standard by which we evaluate scientific hypotheses. Physicists, for example, favour mathematically elegant theories, like Einstein’s theory of General Relativity, that unify a diverse range of phenomena. Consider a rival hypothesis to Einstein’s according to which the laws of General Relativity hold throughout all of space and time except tomorrow afternoon in my back garden, where nature obeys the laws of Aristotelian physics. This crackpot “theory” is logically compatible with all existing observations, and it doesn’t render any of them less probable than Einstein’s. By falsificationist lights, Einstein’s theory and mine are equally well confirmed. Is that true? If you want to predict what is going to happen tomorrow afternoon in my back garden, you would surely be insane to rely on my theory.

A third approach to confirmation, besides the syntactical and the falsificationist approach, is **Bayesian Confirmation Theory**. It is by far the most popular approach in contemporary philosophy of science. (Its statistics ally is **Bayesian Statistics**.)

Why do we care about whether, or to what extent, a hypothesis is confirmed by the evidence? Ultimately, it’s because we want to know how much credence we should invest in the hypothesis. We want to know how confident we should be that smoking causes cancer, or that the laws of Aristotelian physics will be operative tomorrow in my garden.

Bayesianism offers a simple, albeit schematic, answer. If  $E$  is the relevant evidence, then the credence we should give to a hypothesis  $H$  in light of  $E$  is  $Cr_0(H/E)$ , where  $Cr_0$  is a rational prior credence function.

In fact, Bayesians distinguish two notions of evidential support. We may ask about the *absolute* degree to which a hypothesis is supported by the evidence, but we may

also ask about the *incremental* effect a single piece of data has on the credibility of the hypothesis. One black raven, for example, hardly makes it probable that all ravens are black. Still, under normal circumstances, it lends some support to the generalisation.

**The Bayesian analysis of confirmation**

*E* (**absolutely**) confirms *H* to the extent that  $Cr_0(H/E)$  is high.

*E* (**incrementally**) confirms *H* to the extent that  $Cr_0(H/E)$  exceeds  $Cr_0(H)$ .

Without more information about the prior credence  $Cr_0$  these schematic analyses may not appear terribly useful. But let's have a closer look.

On the Bayesian account, confirmation comes in degrees, and its degree is closely related to the conditional probability  $Cr_0(H/E)$ . With the help of Bayes' Theorem, we can break this conditional probability into three parts, which we may understand as three components of Bayesian confirmation:

$$Cr_0(H/E) = \frac{Cr_0(E/H) \cdot Cr_0(H)}{Cr_0(E)}.$$

The first component is  $Cr_0(E/H)$ . This is the probability of the evidence given the hypothesis. The Bayesian analysis implies that, all else equal, the more probable the evidence is in light of a hypothesis, the more the evidence support the hypothesis. Conversely, if a hypothesis renders the evidence unlikely, then (all else equal) the evidence is evidence against the hypothesis. In easy cases, we may use the tools of classical statistics to compute an objective probability for *E* given *H*, and invoke the Probability Coordination Principle to determine  $Cr_0(E/H)$ . But we don't have to go via objective probabilities. We can take into account all kinds of data. And we don't need an arbitrary cutoff at which the hypothesis is "rejected".

The second component,  $Cr_0(H)$ , is the prior probability of the hypothesis. This is where simplicity, systematicity, and other such criteria enter the picture. My crackpot theory about my Aristotelian back garden deserves negligible prior probability. (Why? Because rational priors assume that nature is "uniform", and my theory posits a bizarre kind of non-uniformity.)

The third component,  $Cr_0(E)$ , is the prior probability of the evidence. This occurs in the denominator, meaning that the *lower* the prior probability of the evidence,

the *higher* the degree of confirmation. This makes sense. Einstein's theory of Relativity predicts that light is deflected when it travels past massive objects. The first observation of this effect, in 1919, was deemed a great triumph for Einstein, because the observation was so surprising. It has low prior probability. By comparison, if an astrologer predicts that we will face personal challenges and make new acquaintances in the coming year, and the prediction comes true, this isn't a great triumph for astrology, because the prediction was highly probable all along.

So we can say a lot without knowing what  $Cr_0$  looks like. Still, it would be good to know more. This brings us back to the questions we've discussed earlier in this chapter. Should rational priors satisfy some kind of indifference requirement? If so, what does that requirement look like? How, exactly, should priors be biased towards "uniform" worlds? Should they be aligned with some basic physical quantities?

On a more general level, we may ask how tightly priors are constrained by the norms of rationality. Some hold that there is a unique rational prior credence function. Others say that rationality is "permissive", that it allows for a wide range of priors, each of which is as rational as the other. According to the permissive view, there is an irreducibly subjective element to rational credence: perfectly rational agents with the exact same evidence may arrive at different beliefs. There may, accordingly, be no objective answer to how strongly a scientific hypothesis is supported by the evidence.

**Exercise 4.9 †**

Show that if a theory  $H$  entails  $E$ , and both  $H$  and  $E$  have a prior probability that isn't 0 or 1, then  $E$  incrementally confirms  $H$ .

**Exercise 4.10 (The raven paradox) †††**

The hypothesis that all ravens are black is logically equivalent to the hypothesis that all non-black things are non-ravens. If universal generalizations are normally confirmed by their instances, and logically equivalent hypotheses are confirmed by the same data, then an observation of a white shoe ought to support the hypothesis that all ravens are black. Does it?



### Essay Question 4.1

Evaluate the hypothesis that there is a unique rational prior. Assuming that beliefs evolve by conditionalising on the evidence, this is equivalent to the hypothesis that rational agents with the same evidence should have the same degrees of belief. Can you find an argument for or against this view?

### Sources and Further Reading

Chapter 15 of Ian Hacking, *An Introduction to Probability and Inductive Logic* (2001) goes into some more details about conditionalization.

The cube exercise is due to Bas van Fraassen, *Laws and Symmetry* (1989, p.303). Similar problems for the Indifference Principle are often discussed under the heading of ‘Bertrand’s Paradox’.

The Probability Coordination Principle is best known as the ‘Principal Principle’, introduced in David Lewis, “A Subjectivist’s Guide to Objective Chance” (1980). Lewis’s formulation includes an important extra parameter for “admissible evidence” that I have omitted.

My claim that  $Cr(101 Gs/100 Gs) \approx 0.99$  is an application of Laplace’s “Rule of Succession”. Laplace’s assumptions can be weakened. For example, we don’t need to assume that you start with a uniform prior over the objective probabilities. (Search for “Bayesian convergence” if you’re interested in this.)

Hempel’s 1945 paper on confirmation is called “Studies in the Logic of Confirmation”. It comes in two parts, and also introduces the raven paradox. Popper’s falsificationist approach was first spelled out in his *The Logic of Discovery* (1935). Modern Bayesian Confirmation Theory begins with Rudolf Carnap, *Logical Foundations of Probability* (1950). Michael Strevens’s [Lecture Notes on Bayesian Confirmation Theory](#) (2017) provide a good introduction. The example of the black raven in the mail is from Strevens. For a brief comparison between the frequentist (“classical”) and the Bayesian approach to statistical inference, see Matthew Kotzen, “The Bayesian and Classical Approaches to statistical inference” (2022).

For an introduction to the debate over how wide the range of rational priors might be, see Christopher G. Meacham, “Impermissive Bayesianism” (2014).