# Chapter 3

# Conditional Credences

Chapter 2's discussion was confined to unconditional credence, an agent's outright degree of confidence that a particular proposition is true. This chapter takes up conditional credence, an agent's credence that one proposition is true on the supposition that another is.

The main focus of this chapter is our fourth core Bayesian rule: the Ratio Formula. This rational constraint on conditional credences has a number of important consequences, including Bayes' Theorem (which gives Bayesianism its name).

Conditional credences are also central to the way Bayesians understand evidential relevance. I will define relevance as positive correlation, then explain how this notion has been used to investigate causal relations through the concept of screening off.

Having achieved a deeper understanding of the mathematics of conditional credences, I return at the end of the chapter to what exactly a conditional credence is. In particular, I discuss an argument by David Lewis that a conditional credence can't be understood as an unconditional credence in a conditional.

## 3.1 Conditional credences and the Ratio Formula

Andy and Bob know that two events will occur simultaneously in separate rooms: a fair coin will be flipped, and a clairvoyant will predict how it will land. Let $H$ represent the proposition that the coin comes up heads, and $C$ represent the proposition that the clairvoyant predicts heads. Suppose Andy and Bob each assign an unconditional credence of $1/2$ to $H$ and an unconditional credence of $1/2$ to $C$.

Although Andy and Bob assign the same unconditional credences as each other to $H$ and $C$, they still might take these propositions to be *related* in different ways. We could tease tease out those differences by saying to each agent, "I have no idea how the coin is going to come up or what the clairvoyant is going to say. But suppose for a moment the clairvoyant predicts heads. On this supposition, how confident are you that the coin will come up heads?" If Andy says 1/2 and Bob says 99/100, that's a good indication that Bob has more faith in the mystical than Andy.

The quoted question in the previous paragraph elicits Andy and Bob's *conditional* credences, as opposed to the *unconditional* credences discussed in Chapter 2. An unconditional credence is a degree of belief assigned to a single proposition, indicating how confident the agent is that that proposition is true. A **conditional credence** is a degree of belief assigned to an ordered pair of propositions, indicating how confident the agent is that the first proposition is true on the supposition that the second is. We symbolize conditional credences as follows:

$$\mathrm{cr}(H \mid C) = 1/2 \tag{3.1}$$

This equation says that a particular agent (in this case, Andy) has a 1/2 credence that the coin comes up heads conditional on the supposition that the clairvoyant predicts heads. The vertical bar indicates a conditional credence; to the right of the bar is the proposition supposed; to the left of the bar is the proposition evaluated in light of that supposition. The proposition to the right of the bar is sometimes called the **condition**; I am not aware of any generally-accepted name for the proposition on the left.

Note that a conditional credence is assigned to an *ordered* pair of propositions. It makes a difference which proposition is supposed and which is evaluated. Consider a case in which I'm going to roll a fair die and you have various credences involving the proposition $E$ that it comes up even and the proposition 6 that it comes up six. Compare:

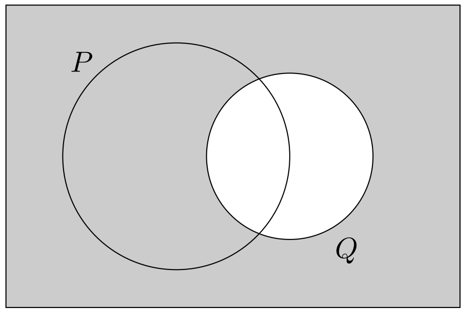
$$\mathrm{cr}(6 \mid E) = 1/3 \tag{3.2}$$

$$\mathrm{cr}(E \mid 6) = 1 \tag{3.3}$$

### 3.1.1 The Ratio Formula

Section 2.2 described Kolmogorov's probability axioms, which Bayesians take to represent rational constraints on an agent's unconditional credences. Bayesians then add a constraint relating conditional to unconditional credences:

Figure 3.1: $\text{cr}(P \mid Q)$




**Ratio Formula:** For any $P$ and $Q$ in $\mathcal{L}$, if $\text{cr}(Q) > 0$ then

$$\text{cr}(P \mid Q) = \frac{\text{cr}(P \,\&\, Q)}{\text{cr}(Q)}$$

Stated this way, the Ratio Formula remains silent on the value of $\text{cr}(P \mid Q)$ when $\text{cr}(Q) = 0$. There are various positions on how one should assign conditional credences when the condition has credence 0; we'll cover some of them in our discussion of the infinite in Chapter 5.

Why should an agent's conditional credences equal the ratio of those unconditionals? Consider Figure 3.1. The rectangle represents all the possible worlds the agent entertains. The agent's unconditional credence in $P$ is the fraction of that rectangle taken up by the $P$-circle. (The area of the rectangle is stipulated to be 1, so that fraction is the area of the $P$-circle divided by 1, which is just the area of the $P$-circle.) When we ask the agent to evaluate a credence conditional on the supposition that $Q$, she temporarily narrows her focus to just those possibilities that make $Q$ true. In other words, she excludes from her attention the worlds I've shaded in the diagram, and considers only what's in the $Q$-circle. The agent's credence in $P$ conditional on $Q$ is the fraction of the $Q$-circle occupied by $P$-worlds. So it's the area of the $PQ$ overlap divided by the area of the entire $Q$-circle, which is $\text{cr}(P \,\&\, Q)/\text{cr}(Q)$.

In the scenario in which I roll a fair die, your initial doxastic possibilities include all six outcomes of the die roll. If I ask you to evaluate $\text{cr}(6 \mid E)$, you exclude from consideration all the odd outcomes. That doesn't mean you've actually *learned* that the die outcome is even; I've just asked you to suppose momentarily that it comes up even and assign a confidence to other propositions in light of that supposition. You distribute your credence equally over the outcomes that remain under consideration (2, 4, and 6), so your credence in 6 conditional on even is 1/3.

We get the same result from the Ratio Formula:

$$\mathrm{cr}(6 \,|\, E) = \frac{\mathrm{cr}(6 \,\&\, E)}{\mathrm{cr}(E)} = \frac{1/6}{1/2} = \frac{1}{3} \tag{3.4}$$

The Ratio Formula allows us to calculate your conditional credences (confidences under a supposition) in terms of your unconditional credences (confidences assigned when no suppositions are made). Hopefully it's obvious why $E$ gets an unconditional credence of 1/2 in this case; as for $6 \,\&\, E$, that's equivalent to just 6, so it gets an unconditional credence of 1/6.[1]

---

> Warning: Mathematicians often take the Ratio Formula to be a *definition* of conditional probability. From their point of view, a conditional probability has the value it does *in virtue of* two unconditional probabilities' standing in a certain ratio. But I do not want to reduce the possession of a conditional credence to the possession of two unconditional credences standing in a particular relation. I take a conditional credence to be a genuine mental state (an attitude towards an ordered pair of propositions) capable of being elicited in various ways, such as by asking an agent her confidence in a proposition given a supposition. The Ratio Formula is a rational constraint on how an agent's conditional credences should relate to her unconditional credences, and as a normative *constraint* (rather than a *definition*) it can be violated—by assigning a conditional credence that doesn't equal the specified ratio.

---

The point of the previous warning is that the Ratio Formula is a rational constraint, and not all agents meet all the rational constraints on their credences. Yet for an agent who does satisfy the Ratio Formula, there can be no difference in her conditional credences without a difference in her unconditional credences as well. (We say that a rational agent's conditional credences **supervene** on her unconditional credences.) Fully specifying an agent's unconditional credence distribution suffices to specify her conditional credences as well. For instance, we might specify Andy's and Bob's credence distributions using the following stochastic truth-table:

| $C$ | $H$ | $\mathrm{cr}_A$ | $\mathrm{cr}_B$ |
|---|---|---|---|
| T | T | 1/4 | 99/200 |
| T | F | 1/4 | 1/200 |
| F | T | 1/4 | 1/200 |
| F | F | 1/4 | 99/200 |

Here $\mathrm{cr}_A$ represents Andy's credences and $\mathrm{cr}_B$ represents Bob's. Andy's unconditional credence in $C$ is identical to Bob's—the values on the first two rows sum to 1/2 for each of them. Similarly, Andy and Bob have the same unconditional credence in $H$ (the sum of the first and third rows). Yet Andy and Bob disagree in their confidence that the coin will come up heads given that the clairvoyant predicts heads. Using the Ratio Formula, we calculate this conditional credence by dividing the value on the first row of the table by the sum of the values on the first two rows. This yields:

$$\mathrm{cr}_A(H \mid C) = \frac{1/4}{1/2} = \frac{1}{2} \neq \frac{99}{100} = \frac{99/200}{100/200} = \mathrm{cr}_B(H \mid C) \tag{3.5}$$

Bob has high confidence in the clairvoyant's abilities. So on the supposition that the clairvoyant predicts heads, Bob is almost certain that heads will come up on the flip. Andy, on the other hand, is skeptical, so supposing that the clairvoyant predicts heads leaves his opinions about the flip outcome unchanged.

### 3.1.2 Consequences of the Ratio Formula

Combining the Ratio Formula with the probability axioms yields further useful probability rules. First we have the

**Law of Total Probability:** For any finite partition $Q_1, Q_2, \ldots, Q_n$ in $\mathcal{L}$,

$$\mathrm{cr}(P) = \mathrm{cr}(P \mid Q_1) \cdot \mathrm{cr}(Q_1) + \mathrm{cr}(P \mid Q_2) \cdot \mathrm{cr}(Q_2) + \\ \ldots + \mathrm{cr}(P \mid Q_n) \cdot \mathrm{cr}(Q_n)$$

Suppose you're trying to predict whether I will bike to work tomorrow, but you're unsure if the weather will rain, hail, or be clear. The Law of Total Probability allows you to systematically work through the possibilities in that partition. You multiply your confidence that it will rain by your confidence that I'll bike should it rain. Then you multiply your confidence that it'll hail by your confidence in my biking given hail. Finally you multiply your unconditional credence that it'll be clear by your conditional credence that I'll bike given that it's clear. Adding these three products together gives your unconditional credence that I'll bike. (In the formula the proposition that I'll bike plays the role of $P$ and the three weather possibilities are $Q_1$, $Q_2$, and $Q_3$.)

Next, the Ratio Formula makes any conditional credence distribution itself a probability distribution. To understand what that means and how it works in general, we'll start with a very special example. Let's say I ask you to report your unconditional credences in some propositions. Then I ask you to assign credences to those propositions conditional on the supposition of... nothing. I give you nothing more to suppose. Clearly you'll just report back to me the same credences. Bayesians represent vacuous information as a tautology, so what we've just seen is that a rational agent's credences conditional on a tautology equal her unconditional credences.[2] In other words, for any $P$ in $\mathcal{L}$

$$\text{cr}(P \mid \mathsf{T}) = \text{cr}(P) \tag{3.6}$$

Since we're assuming the agent's unconditional credences satisfy the probability axioms, the credences that result if we ask her to suppose a tautology also satisfy the probability axioms. The distribution $\text{cr}(\cdot \mid \mathsf{T})$ (where the "$\cdot$" is a blank to be filled by a proposition) must be a probability distribution. Remarkably, this turns out to be true for non-tautologous conditions as well. Pick any proposition $R$ in $\mathcal{L}$ such that $\text{cr}(R) > 0$, and the function you get by taking various credences conditional on the supposition of $R$ will satisfy the probability axioms. That is,

- For any proposition $P$ in $\mathcal{L}$, $\text{cr}(P \mid R) \geqslant 0$.
- For any tautology $\mathsf{T}$ in $\mathcal{L}$, $\text{cr}(\mathsf{T} \mid R) = 1$.
- For any mutually exclusive propositions $P$ and $Q$ in $\mathcal{L}$, $\text{cr}(P \vee Q \mid R) = \text{cr}(P \mid R) + \text{cr}(Q \mid R)$.

In other words, $\text{cr}(\cdot \mid R)$ satisfies Kolmogorov's probability axioms. (You'll prove this in Exercise 3.3.)

Knowing that a conditional credence distribution is a probability distribution can be a handy shortcut. (It also has a significance for updating credences that we'll discuss in Chapter 4.) Because it's a probability distribution, a conditional credence distribution must satisfy all the consequences of the probability axioms we saw in Section 2.2.1. If I tell you that $\text{cr}(P \mid R) = 0.7$, you know that $\text{cr}({\sim}P \mid R) = 0.3$, by the following implementation of the Negation rule:

$$\text{cr}({\sim}P \mid R) = 1 - \text{cr}(P \mid R) \tag{3.7}$$

Similarly, by Entailment if $P \vDash Q$ then $\text{cr}(P \mid R) \leqslant \text{cr}(Q \mid R)$.

### 3.1.3 Bayes' Theorem

The most famous consequence of the Ratio Formula and Kolmogorov's axioms is

**Bayes' Theorem:** For any $H$ and $E$ in $\mathcal{L}$,

$$\mathrm{cr}(H \mid E) = \frac{\mathrm{cr}(E \mid H) \cdot \mathrm{cr}(H)}{\mathrm{cr}(E)}$$

The first thing to say about Bayes' Theorem is *that it is a theorem*—it can be proven straightforwardly from the axioms and Ratio Formula. This is worth remembering, because there is a great deal of controversy about how Bayesians *apply* the theorem. (The significance they attach to this theorem is how Bayesians came to be called "Bayesians".)

What philosophical significance could attach to an equation that is, in the end, just a truth of mathematics? The theorem was first articulated by the Reverend Thomas Bayes in the 1700s.[3] Prior to Bayes, much of probability theory was concerned with problems of **direct inference**. Direct inference starts with the supposition of some probabilistic hypothesis, then asks how likely that hypothesis makes a particular experimental result. You probably learned to solve many direct inference problems in school, such as "Suppose I flip a fair coin 20 times; how likely am I to get exactly 19 heads?" Here the probabilistic hypothesis $H$ is that the coin is fair, and the experimental result $E$ is exactly 19 heads. Your credence that the experimental result will occur on the supposition that the hypothesis is true—$\mathrm{cr}(E \mid H)$—is called the **likelihood**.[4]

Yet Bayes was also interested in **inverse inference**. Instead of making suppositions about hypotheses and determining probabilities of courses of evidence, his theorem allows us to calculate probabilities of hypotheses from suppositions about evidence. Instead of calculating the likelihood $\mathrm{cr}(E \mid H)$, Bayes' Theorem shows us how to calculate $\mathrm{cr}(H \mid E)$. A problem of inverse inference might ask, "Suppose a coin comes up heads on exactly 19 of 20 flips; how probable is it that the coin is fair?"

Assessing the significance of Bayes' Theorem, Hans Reichenbach wrote,

> The *method of indirect evidence*, as this form of inquiry is called, consists of inferences that on closer analysis can be shown to follow the structure of the rule of Bayes. The physician's inferences, leading from the observed symptoms to the diagnosis of a specified disease, are of this type; so are the inferences of the historian determining the historical events that must be assumed

> for the explanation of recorded observations; and, likewise, the inferences of the detective concluding criminal actions from inconspicuous observable data.... Similarly, the general inductive inference from observational data to the validity of a given scientific theory must be regarded as an inference in terms of Bayes' rule. (Reichenbach 1935/1949, pp. 94–5)[5]

Here's an example of inverse inference: You're a biologist studying a particular species of fish, and you want to know whether the genetic allele coding for blue fins is dominant or recessive. Intially you assign 50-50 credence to each possibility. A simple direct inverse from the theory of genetics tells you that if the allele is dominant, roughly 3 out of 4 species members will have blue fins; if the allele is recessive blue fins will appear on roughly 25% of the fish. But you're going to perform an inverse inference, making experimental observations to decide between genetic hypotheses. You will capture fish from the species at random and examine their fins. How significant will your first observation be to your credences in dominant vs. recessive? When you contemplate various ways that observation might turn out, how should supposing one outcome or the other affect your credences about the allele? Before we do the calculation, try estimating how confident you should be that the allele is dominant on the supposition that the first fish you observe has blue fins.

In this example our hypothesis $H$ will be that the blue-fin allele is dominant. The evidence $E$ to be supposed is that a randomly-drawn fish has blue fins. We want to calculate the **posterior** value $\mathrm{cr}(H \mid E)$. This value is called the "posterior" because it's your credence in the hypothesis $H$ *after* the evidence $E$ has been supposed. In order to calculate this posterior, Bayes' Theorem requires the values of $\mathrm{cr}(E \mid H)$, $\mathrm{cr}(H)$, and $\mathrm{cr}(E)$.

$\mathrm{cr}(E \mid H)$ is the likelihood of drawing a blue-finned fish on the hypothesis that the allele is dominant. On the supposition that the allele is dominant, 75% of the fish have blue fins, so your $\mathrm{cr}(E \mid H)$ value should be 0.75. The other two values are known as **priors**; they are your unconditional credences in the hypothesis and the evidence *before* anything is supposed. The prior in the hypothesis $H$ is easy—we said you initially split your credence 50-50 between dominant and recessive. So $\mathrm{cr}(H)$ is 0.5. But what about the prior in the evidence? How confident are you before observing any fish that the first one you draw will have blue fins?

Here we can apply the Law of Total Probability to the partition consisting of $H$ and $\sim H$. This yields:

$$\mathrm{cr}(E) = \mathrm{cr}(E \mid H) \cdot \mathrm{cr}(H) + \mathrm{cr}(E \mid \sim H) \cdot \mathrm{cr}(\sim H) \tag{3.8}$$

The values on the righthand side are all either priors in the hypothesis or likelihoods. These values we can easily calculate. So

$$\mathrm{cr}(E) = 0.75 \cdot 0.5 + 0.25 \cdot 0.5 = 0.5 \tag{3.9}$$

Plugging all these values into Bayes' Theorem gives us

$$\mathrm{cr}(H \mid E) = \frac{\mathrm{cr}(E \mid H) \cdot \mathrm{cr}(H)}{\mathrm{cr}(E)} = \frac{0.75 \cdot 0.5}{0.5} = 0.75 \tag{3.10}$$

Observing a single fish has the potential to change your credences substantially. On the supposition that the fish you draw has blue fins, your credence that the blue-fin allele is dominant goes from its prior value of 1/2 to a posterior of 3/4.

Again, all of this is strictly mathematics from a set of axioms that are rarely disputed. So why has Bayes' Theorem been the focus of controversy? One issue is the role Bayesians see the theorem playing in *updating* our attitudes over time; we'll return to that application of the theorem in Chapter 4. But the main idea that Bayesians take from Bayes—and that has proven controversial—is that probabilistic inverse inference is the key to induction. Bayesians think the primary way we ought to draw conclusions from data—how we ought to reason about scientific hypotheses, say, on the basis of experimental evidence—is by calculating posterior credences using Bayes' Theorem. This view stands in direct conflict with other statistical methods, such as frequentism and likelihoodism. Once we have considerably deepened our understanding of Bayesian Epistemology, we will discuss this conflict in Chapter XXX.

Before moving on, I'd like to highlight two useful alternative forms of Bayes' Theorem. We've just seen that calculating the prior of the evidence—$\mathrm{cr}(E)$—can be easier if we break it up using the Law of Total Probability. Incorporating that manuever into Bayes' Theorem yields

$$\mathrm{cr}(H \mid E) = \frac{\mathrm{cr}(E \mid H) \cdot \mathrm{cr}(H)}{\mathrm{cr}(E \mid H) \cdot \mathrm{cr}(H) + \mathrm{cr}(E \mid {\sim}H) \cdot \mathrm{cr}({\sim}H)} \tag{3.11}$$

When a particular hypothesis $H$ is under consideration, its negation ${\sim}H$ is known as the **catchall** hypothesis. So this form of Bayes' Theorem calculates the posterior in the hypothesis from the priors and likelihoods of the hypothesis and its catchall.

In other situations we have multiple hypotheses under consideration instead of just one. Given a finite partition of $n$ hypotheses $H_1, H_2, \ldots, H_n$, the Law of Total Probability transforms the denominator of Bayes' Theorem to yield

$$\text{cr}(H_i \mid E) = \frac{\text{cr}(E \mid H_i) \cdot \text{cr}(H_i)}{\text{cr}(E \mid H_1) \cdot \text{cr}(H_1) + \text{cr}(E \mid H_2) \cdot \text{cr}(H_2) + \ldots + \text{cr}(E \mid H_n) \cdot \text{cr}(H_n)} \tag{3.12}$$

This version allows you to calculate the posterior of one particular hypothesis $H_i$ in the partition from the priors and likelihoods of all the hypotheses.

## 3.2 Relevance and independence

Andy doesn't believe in hocus pocus; from his point of view, information about what a clairvoyant predicts is irrelevant to determining how a coin flip will come out. So supposing that a clairvoyant predicts heads makes no difference to Andy's confidence in a heads outcome. If $C$ says the clairvoyant predicts heads, $H$ says the coin lands heads, and $\text{cr}_A$ is Andy's credence distribution, we have

$$\text{cr}_A(H \mid C) = 1/2 = \text{cr}_A(H) \tag{3.13}$$

Generalizing this idea yields a key definition: Proposition $P$ is **probabilistically independent** of proposition $Q$ relative to distribution cr just in case

$$\text{cr}(P \mid Q) = \text{cr}(P) \tag{3.14}$$

In this case Bayesians also say that $Q$ is **irrelevant** to $P$. When $Q$ is irrelevant to $P$, supposing $Q$ leaves an agent's credence in $P$ unchanged.

Notice that probabilistic independence is always relative to a credence distribution cr. The very same propositions $P$ and $Q$ might be independent relative to one credence distribution but dependent relative to another. (Relative to Andy's credences the clairvoyant's prediction is irrelevant to the flip outcome, but relative to the credences of his friend Bob—who believes in psychic powers—it is not.) In what follows I may omit reference to a particular credence function when context makes it clear, but you should keep the relativity of independence to probability distribution in the back of your mind.

While Equation (3.14) will be our official definition of probabilistic independence, there are many equivalent tests for independence. Given the probability axioms and Ratio Formula, the following equations are all true just when Equation (3.14) is:[6]

$$\mathrm{cr}(P) = \mathrm{cr}(P \mid {\sim}Q) \tag{3.15}$$

$$\mathrm{cr}(P \mid Q) = \mathrm{cr}(P \mid {\sim}Q) \tag{3.16}$$

$$\mathrm{cr}(Q \mid P) = \mathrm{cr}(Q) = \mathrm{cr}(Q \mid {\sim}P) \tag{3.17}$$

$$\mathrm{cr}(P \,\&\, Q) = \mathrm{cr}(P) \cdot \mathrm{cr}(Q) \tag{3.18}$$

The equivalence of Equations (3.14) and (3.15) tells us that if supposing $Q$ makes no difference to an agent's confidence in $P$, then supposing ${\sim}Q$ makes no difference as well. The equivalence of (3.14) and (3.17) shows us that independence is symmetric: if supposing $Q$ makes no difference to an agent's credence in $P$, supposing $P$ won't change the agent's attitude towards $Q$ either. Finally, Equation (3.18) embodies a useful probability rule:

**Multiplication:** $P$ and $Q$ are probabilistically independent relative to cr if and only if $\mathrm{cr}(P \,\&\, Q) = \mathrm{cr}(P) \cdot \mathrm{cr}(Q)$.

(Some authors *define* probabilistic independence using this biconditional, but we will define independence using Equation (3.14) and then treat Multiplication as a consequence.)

Notice that we can calculate the credence of a conjunction by multiplying the credences of its conjuncts only when those conjuncts are *independent*. This trick will not work for any arbitrary propositions. The general formula for credence in a conjunction can be derived quickly from the Ratio Formula:

$$\mathrm{cr}(P \,\&\, Q) = \mathrm{cr}(P \mid Q) \cdot \mathrm{cr}(Q) \tag{3.19}$$

When $P$ and $Q$ are probabilistically independent, the first term on the right-hand side equals $\mathrm{cr}(P)$.

It's important not to get Multiplication and Finite Additivity confused. Finite Additivity says that the credence of a *disjunction* is the *sum* of the credences of its *mutually exclusive* disjuncts. Multiplication says that the credence of a *conjunction* is the *product* of the credences of its *independent* conjuncts. If I flip two fair coins in succession, heads on the first and heads on the second are independent, while heads on the first and tails on the first are mutually exclusive.

Probabilistic independence fails to hold when one proposition is **relevant** to the other. Replace the "=" signs in Equations (3.14) through (3.18) with ">" signs and you have tests for $Q$'s being **positively relevant** to $P$. Once more the tests are equivalent—if any of the resulting inequalities is true, all of them are. $Q$ is positively relevant to $P$ when assuming $Q$ makes you more confident in $P$. For example, since Bob believes in mysticism he takes the clairvoyant's predictions to be highly relevant to the outcome of the coin flip—supposing that the clairvoyant has predicted heads takes him from equanimity to near-certainty in a heads outcome. Bob assigns

$$\mathrm{cr}_B(H \mid C) = 99/100 > 1/2 = \mathrm{cr}_B(H) \tag{3.20}$$

Like independence, positive relevance is symmetric. Given his high confidence in the clairvoyant's accuracy, supposing that the coin came up heads will make Bob highly confident that the clairvoyant predicted it would.

Similarly, replacing the "=" signs with "<" signs above yields tests for **negative relevance**. For Bob, the clairvoyant's predicting heads is negatively relevant to the coin's coming up tails. Like positive correlation, negative correlation is symmetric (supposing a tails outcome makes Bob less confident in a heads prediction). "Relevance" terms have a number of synonyms. Instead of finding "positively/negatively relevant" terminology, you'll sometimes find "positively/negatively dependent", "positively/negatively correlated", or even "correlated/anti-correlated" used.

The strongest forms of positive and negative relevance are entailment and refutation. Suppose a hypothesis $H$ has nonextreme prior credence. If a particular piece of evidence $E$ *entails* the hypothesis, the probability axioms and Ratio Formula tell us

$$\mathrm{cr}(H \mid E) = 1 \tag{3.21}$$

Supposing $E$ takes $H$ from a middling credence to the highest credence allowed. Similarly, if $E$ refutes $H$ (what philosophers of science call **falsification**), then

$$\mathrm{cr}(H \mid E) = 0 \tag{3.22}$$

Relevance will be most important to us because of its connection to confirmation, the Bayesian notion of evidential support. A piece of evidence confirms a hypothesis only if it's relevant to that hypothesis. Put another way, learning a piece of evidence changes a rational agent's credence in a hypothesis only if that evidence is relevant to the hypothesis. (Much more on all this later.)

### 3.2.1 Conditional independence and screening off

The definition of probabilistic independence compares an agent's conditional credence in a proposition to her unconditional credence in that proposition.

But we can also compare conditional credences. When Bob, who believes in the occult, hears a clairvoyant's prediction about the outcome of a fair coin flip, he takes it to be highly correlated with the true flip outcome. But what if we ask Bob to suppose that this particular clairvoyant is an impostor? Once he supposes the clairvoyant is an impostor, Bob may see the clairvoyant's predictions as completely irrelevant to the flip outcome. Let $C$ be the proposition that the clairvoyant predicts heads, $H$ be the proposition that the coin comes up heads, and $I$ be the proposition that the clairvoyant is an impostor. It's possible for Bob's credences to satisfy both of the following equations at once:

$$\text{cr}(H \mid C) > \text{cr}(H) \tag{3.23}$$

$$\text{cr}(H \mid C \,\&\, I) = \text{cr}(H \mid I) \tag{3.24}$$

Equation (3.23) tells us that unconditionally, Bob takes $C$ to be relevant to $H$. But conditional on the supposition of $I$, $C$ becomes independent of $H$; supposing $C \,\&\, I$ gives Bob the same confidence in $H$ as supposing $I$ alone.

Generalizing this idea yields the following definition of **conditional independence**: $P$ is probabilistically independent of $Q$ conditional on $R$ just in case

$$\text{cr}(P \mid Q \,\&\, R) = \text{cr}(P \mid R) \tag{3.25}$$

If this equality fails to hold, we say that $P$ is relevant to (or dependent on) $Q$ conditional on $R$.

One more piece of terminology: We will say that $R$ **screens off** $P$ from $Q$ when $P$ is unconditionally dependent on $Q$ but independent of $Q$ conditional on $R$. In other words, supposing $R$ makes the correlation between $P$ and $Q$ disappear. Equations (3.23) and (3.24) demonstrate that for Bob, $I$ screens off $H$ from $C$.[7] We'll now consider a number of challenging and puzzling probabilistic phenomena whose explanations involve the notions of conditional dependence and screening off.

### 3.2.2 The Gambler's Fallacy

People often act as if future chancy events will "compensate" for unexpected past results. When a good hitter strikes out many times in a row, someone will say he's "due" for a hit. If a fair coin comes up heads 19 times in a row, many people become more confident that the next outcome will be tails.

This mistake is known as the **Gambler's Fallacy**.[8] A person who makes the mistake is thinking along something like the following lines: In twenty flips of a fair coin, it's more probable to get 19 heads and 1 tail than it is to get 20 heads. So having seen 19 heads, it's much more likely that the next flip will come up tails.

This person is providing the right answer to the wrong question. The answer to the question "When a fair coin is flipped 20 times, is 19 heads and 1 tail more likely than 20 heads?" is yes—in fact, it's 20 times as likely! But that's the wrong question to ask in this case. Instead of wondering what sorts of outcomes are probable when one flips a fair coin 20 times in general, it's more appropriate to ask of this specific case: *given* that the coin has already come up heads 19 times, how confident are we that the *next* flip will be tails? This is a question about our conditional credence

$$\text{cr}(\text{next flip heads} \,|\, \text{previous 19 flips heads}) \tag{3.26}$$

How should we calculate this conditional credence? Ironically, it might be more reasonable to make a mistake in the *opposite* direction from the Gambler's Fallacy. If I see a coin come up heads 19 times, shouldn't that make me suspect that it's biased towards heads? If anything, shouldn't supposing 19 consecutive heads make me more confident that the next flip will come up heads than tails?

This line of reasoning would be appropriate to the present case if we hadn't stipulated in setting things up that the coin is fair. The fact that the coin is fair screens off information about the first 19 flips from the outcome of the 20th. That is

$$\begin{aligned} \text{cr}(\text{next flip heads} \,|\, \text{previous 19 flips heads} \,\&\, \text{fair coin}) = \\ \text{cr}(\text{next flip heads} \,|\, \text{fair coin}) \end{aligned} \tag{3.27}$$

We can justify this equation as follows: Typically, information that a coin came up 19 times in a row would alter your opinion about whether it's a fair coin. Changing your opinion about whether it's a fair coin would then affect your prediction for the 20th flip. So typically, information about the first 19 flips alters your credences about the 20th flip *by way of* your opinion about whether the coin is fair. But if you've already established that the coin is fair, information about the first 19 flips has no further significance for your prediction about the 20th. So conditional on the coin's being fair, the first 19 flips' outcomes are irrelevant to the outcome of the 20th flip.

The lefthand side of Equation (3.27) captures the correct question to ask about the Gambler's Fallacy case. The righthand side is easy to calculate; it's 1/2. So after seeing a coin known to be fair come up heads 19 times, we should be 1/2 confident that the next flip will be heads.[9]

### 3.2.3 Probabilities are weird! Simpson's Paradox

Perhaps you're too much of a probabilistic sophisticate to ever commit the Gambler's Fallacy. Perhaps you successfully navigated Tversky and Kahneman's Conjunction Fallacy (Section 2.2.3) as well. But even probability experts sometimes have trouble with countertuitive relations between conditional and unconditional dependence.

Here's an example of how odd things can get: In a famous case, the University of California, Berkeley's graduate departments were investigated for gender bias in admissions. The concern arose because in 1973 about 44% of overall male applicants were admitted to graduate school at Berkeley, while only 35% of female applicants were. Yet when the graduate departments (where admissions decisions are made) were studied one at a time, it turned out that individual departments either were admitting men and women at roughly equal rates, or in some cases were admitting a higher percentage of female applicants.

This is an example of **Simpson's Paradox**, in which probabilistic dependencies (or independencies) that hold conditional on each member of a partition nevertheless fail to hold unconditionally. A Simpson's Paradox case involves a collection with a number of subgroups. Each of the subgroups displays the same probabilistic correlation between two traits. Yet when we examine the collection as a whole that correlation disappears—or is even reversed!

To see how this can happen, consider another example: In 1995, David Justice had a higher batting average than Derek Jeter. In 1996, Justice also had a higher average than Jeter. Yet over that entire two-year span, Jeter's average was better than Justice's.[10]

Here are the data for the two hitters:

| | **1995** | | **1996** | | **Combined** | |
|---|---|---|---|---|---|---|
| Jeter | 12/48 | .250 | 183/582 | .314 | 195/630 | .310 |
| Justice | 104/411 | .253 | 45/140 | .321 | 149/551 | .270 |

The first number in each box is the number of hits; the second is the number of at-bats; the third is the batting average (hits divided by at-bats). Looking at the table, you can see how Justice managed to beat Jeter for average in each individual year yet lose to him overall. In 1995 Justice beat Jeter but both batters hit in the mid-.200s; in 1996 Justice beat Jeter while both hitters had a much better year. Jeter's trick was to have fewer at-bats than Justice during the off year and many more at-bats when both hitters were going well. Totaling the two years, many more of Jeter's at-bats produced hits at the over-.300 rate, while the preponderance of Justice's at-bats came while he was toiling in the .200s.[11]

Scrutiny revealed a similar effect in Berkeley's 1973 admissions data. Bickel, Hammel, and O'Connell (1975) concluded, "The proportion of women applicants tends to be high in departments that are hard to get into and low in those that are easy to get into." Although individual departments were just as willing to admit women as men, female applications were less successful overall because more were directed at departments with low admission rates.

How can we express these examples using conditional probabilities? Suppose you select a Jeter or Justice at-bat at random from the 1,181 at-bats in the combined 1995 and 1996 pool, making your selection so that each of the 1,181 at-bats is equally likely to be selected. How confident should you be that the selected at-bat is a hit? How should that confidence change if you suppose a Jeter at-bat is selected, or an at-bat from 1995?

Below is a stochastic truth-table for your credences, assembled from the real-life statistics above. Here $E$ says that it's a Jeter at-bat; 5 says it's from 1995; and $H$ says it's a hit. (Given the pool from which we're sampling, $\sim E$ means a Justice at-bat and $\sim 5$ means it's from 1996.)

| $E$ | 5 | $H$ | cr |
|---|---|---|---|
| T | T | T | 12/1181 |
| T | T | F | 36/1181 |
| T | F | T | 183/1181 |
| T | F | F | 399/1181 |
| F | T | T | 104/1181 |
| F | T | F | 307/1181 |
| F | F | T | 45/1181 |
| F | F | F | 95/1181 |

A bit of calculation with this stochastic truth-table reveals the following:

$$\mathrm{cr}(H \mid E) > \mathrm{cr}(H \mid \sim E) \tag{3.28}$$

$$\mathrm{cr}(H \mid E \,\&\, 5) < \mathrm{cr}(H \mid \sim E \,\&\, 5) \tag{3.29}$$

$$\mathrm{cr}(H \mid E \,\&\, \sim 5) < \mathrm{cr}(H \mid \sim E \,\&\, \sim 5) \tag{3.30}$$

If you're selecting an at-bat from the total sample, Jeter is more likely to get you a hit than Justice. Put another way, Jeter batting is unconditionally positively relevant to an at-bat's being a hit. But Jeter batting is negatively relevant to a hit conditional on each of the two years in the sample. If you're selecting from only the at-bats associated with a particular year, you're more likely to get a hit if you go with Justice.

### 3.2.4 Correlation and causation

You may have heard the expression "correlation is not causation." People typically use this expression to point out that just because two events have both occurred—and maybe occurred in close spatio-temporal proximity—that doesn't mean they had anything to do with each other. But "correlation" is a technical term in probability discussions. The propositions describing two events may both be true, or you might have high credence in both of them, yet they still might not be probabilistically correlated. For the propositions to be correlated, supposing one to be true must *increase* the probability of the other. I'm confident that I'm under 6 feet tall and that my eyes are blue, but that doesn't mean I see those facts as correlated.

So does probabilistic correlation always indicate a causal relationship? Perhaps not. If I suppose that the fiftieth Fibonacci number is even, that makes me highly confident that it's the sum of two primes. But being even and being the sum of two primes are not *causally* related; Goldbach's Conjecture that every even number greater than 2 is the sum of two primes is an *arithmetic* fact (if it's a fact at all).[12] On the other hand, most correlations we encounter in everyday life are due to empirical conditions. When two propositions are correlated due to empirical facts, must the event described by one cause the event described by the other?

Hans Reichenbach offered a classic counterexample. He wrote,

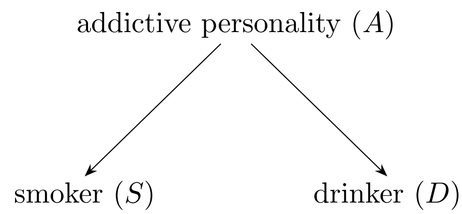
> Suppose two geysers which are not far apart spout irregularly, but throw up their columns of water always at the same time. The existence of a subterranean connection of the two geysers with a common reservoir of hot water is then practically certain. (1956, p. 158)

If you've noticed that two nearby geysers always spout simultaneously, seeing one spout will increase your confidence that the other is spouting as well. So your credences about the geysers are correlated. But you don't think one geyser's spouting *causes* the other to spout. Instead, you hypothesize an unobserved reservoir of hot water that is the **common cause** of both spouts.

Reichenbach proposed a famous principle about empirically correlated events:

**Principle of the Common Cause:** When event outcomes are probabilistically correlated, either one causes the other or they have a common cause.[13]

Figure 3.2: A causal fork



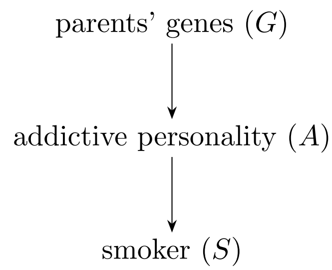

Along with this principle, he offered a key mathematical insight about causation: a common cause screens its effects off from each other.

Let's carefully work through an example. Suppose the proposition that a particular individual is a drinker is positively relevant to the proposition that she's a smoker. This may be because drinking causes smoking—the kinds of places and social situations in which one drinks may encourage smoking. But there's another possible explanation: being a smoker and being a drinker may both be promoted by an addictive personality, which we can imagine results from a genetic endowment unaffected by one's behavior. In that case, an addictive personality would be a common cause of both being a drinker and being a smoker. (See Figure 3.2; the arrows indicate causal influence.)

Imagine the latter explanation is true, and moreover is the *only* true explanation of the correlation between drinking and smoking. That is, being a smoker and being a drinker are positively correlated only due to their both being caused by an addictive personality. Given this assumption, let's take a particular subject whose personality you're unsure about, and consider what happens to your credences when you make various suppositions about her.

If you begin by supposing that the subject drinks, this will make you more confident that she smokes—*but only because it makes you more confident that the subject has an addictive personality.* On the other hand, you might start by supposing that the subject has an addictive personality. That will certainly make you more confident that she's a smoker. But once you've made that adjustment, going on to suppose that she's a drinker won't affect your confidence in smoking. Information about drinking affects your smoking opinions only *by way of* helping you detect an addictive personality, and the answer to the personality question was filled in by your initial supposition. Once an addictive personality is supposed, drinking has

Figure 3.3: A causal chain




no further relevance to smoking. (Compare: Once a coin is supposed to be fair, the outcomes of its first 19 flips have no relevance to the outcome of the 20th.) Drinking is probabilistically independent of smoking conditional on an addictive personality. That is,

$$\mathrm{cr}(S \mid D \,\&\, A) = \mathrm{cr}(S \mid A) \tag{3.31}$$

Causal forks (as in Figure 3.2) give rise to screening off. $A$ is a common cause of $S$ and $D$, so $A$ screens off $S$ from $D$.

But that's not the only way screening off can occur. Consider Figure 3.3. Here we've focused on a different portion of the causal structure. Imagine that the subject's parents' genes causally determine whether she has an addictive personality, which in turn causally promotes smoking. Now her parents' genetics are probabilistically relevant to the subject's smoking, but that correlation is screened off by facts about her personality. Again, if you're uncertain whether the subject's personality is addictive, facts about her parents' genes will affect your opinion of whether she's a smoker. But once you've made a firm supposition about the subject's personality, suppositions about her parents' genetics have no further influence on your smoking opinions. In equation form:

$$\mathrm{cr}(S \mid G \,\&\, A) = \mathrm{cr}(S \mid A) \tag{3.32}$$

$A$ screens off $S$ from $G$.[14]

Relevance, conditional relevance, and causation can interact in very complex ways.[15] My goal here has been to introduce the main ideas and terminology employed in their analysis. The state of the art in this field has come a long way from Reichenbach; computational tools now available can look at statistical correlations among a large number of variables and hypothesize a causal structure lying beneath them. The resulting causal diagrams are known as **Bayes Nets**, and have practical applications from satellites to health care to car insurance to college admissions.

And as for Reichenbach's Principle of the Common Cause? It remains highly controversial.

## 3.3 Conditional credences and conditionals

I want to circle back and get clearer on the nature of conditional credence. First, it's important to note that the conditional credences we've been discussing are indicative, not subjunctive. The distinction is familiar from the theory of conditional propositions. Compare:

If Shakespeare didn't write *Hamlet*, someone else did.
If Shakespeare hadn't written *Hamlet*, someone else would have.

The former conditional is indicative, while the latter is subjunctive. Typically one evaluates the truth of a conditional by considering possible worlds in which the antecedent is satisfied, then seeing if those worlds make the consequent true as well. When you evaluate an indicative conditional, you're restricted to considering worlds among your doxastic possibilities. Evaluating a subjunctive conditional, on the other hand, permits you to engage in *counterfactual* reasoning involving worlds you've actually ruled out. So for the subjunctive conditional above, you can consider worlds that make the antecedent true because *Hamlet* never exists. But for the indicative conditional, you have to take into account that *Hamlet* does exist, and entertain only worlds in which that's true. So you consider bizarre "author-conspiracy" worlds which, while far-fetched, satisfy the antecedent and are among your current doxastic possibilities. In the end, I'm guessing you take the indicative conditional to be true but the subjunctive to be false.

Now suppose I ask for your credence in the proposition that someone wrote *Hamlet*, conditional on the supposition that Shakespeare didn't. This value will be high, again because *Hamlet* exists. In assigning this conditional credence, you aren't bringing into consideration possible worlds you'd otherwise ruled out (such as *Hamlet*-free worlds). Instead, you're focusing in on the narrow set of author-conspiracy worlds you currently entertain. As we saw in Figure 3.1, assigning a conditional credence strictly narrows the worlds under consideration; it's doesn't expand your attention to worlds previously ruled out. Thus the conditional credences discussed in this book—and typically discussed in the Bayesian literature—are indicative rather than subjunctive.[16]

Are there more features of conditional propositions that can help us understand conditional credences? Might we understand conditional credences *in terms of* conditionals? Initiating his study of conditional degrees of belief, F.P. Ramsey warned against assimilating them to conditional propositions:

> We are also able to define a very useful new idea—"the degree of belief in p given q". This does not mean the degree of belief in "If p then q", or that in "p entails q", or that which the subject would have in p if he knew q, or that which he ought to have. (1931, p. 82)

Yet many authors failed to heed Ramsey's warning. It's very tempting to equate conditional credences with some simple combination of conditional propositions and unconditional credences. For example, when I ask, "How confident are you in $P$ given $Q$?", it's easy to hear that as "Given $Q$, how confident are you in $P$?" or just "If $Q$ is true, how confident are you in $P$?" This simple slide might suggest that

$$\text{cr}(P \mid Q) = r \text{ is equivalent to } Q \rightarrow \text{cr}(P) = r \tag{3.33}$$

Here I'm using the symbol "$\rightarrow$" to represent some kind of conditional. For the reasons discussed above, it should be an indicative conditional. But it need not be the material conditional symbolized by "$\supset$"; many authors think the material conditional's truth-function fails to accurately represent the meaning of natural-language indicative conditionals.

There are two problems with the proposal of Equation (3.33). First, it gets the logic of conditional credences wrong. On most theories of the indicative conditional (and certainly if $\rightarrow$ is the material conditional),

$$X \rightarrow Z \text{ and } Y \rightarrow Z \text{ jointly entail } (X \vee Y) \rightarrow Z \tag{3.34}$$

for any propositions $X$, $Y$, and $Z$. Thus for any propositions $A$, $B$, and $C$ and constant $k$ we have

$$A \rightarrow [\text{cr}(C) = k] \text{ and } B \rightarrow [\text{cr}(C) = k] \text{ entail } (A \vee B) \rightarrow [\text{cr}(C) = k] \tag{3.35}$$

Combining Equations (3.33) and (3.35) yields

$$\text{cr}(C \mid A) = k \text{ and } \text{cr}(C \mid B) = k \text{ entail } \text{cr}(C \mid A \vee B) = k \tag{3.36}$$

which is false. Not only can one design a credence distribution satisfying the probability axioms and Ratio Formula such that $\text{cr}(C \mid A) = k$ and

$\text{cr}(C \mid B) = k$ but $\text{cr}(C \mid A \vee B) \neq k$; one can even describe real-life examples in which it's rational for an agent to assign such a distribution. (See Exercise 3.12.) The failure of Equation (3.36) is another case in which credences confound expectations developed by our experiences with classificatory terms.

The second problem with the equivalence proposed by Equation (3.33) is that it's just bizarre. $\text{cr}(P \mid Q) = r$ says that when the agent supposes proposition $Q$ is true, her confidence in $P$ is $r$. $Q \rightarrow \text{cr}(P) = r$ says that if $Q$ is *actually* true, the agent's unconditional credence in $P$ is *actually* $r$. These claims are oddly mismatched, as brought out by the following example: Right now you're highly confident that you'll live for the next year (proposition $P$). Conditional on the supposition that the sun exploded ten seconds ago (proposition $Q$), you are considerably less confident about your life expectancy. But this doesn't mean that if (unbeknownst to you) the sun did explode ten seconds ago, you are right now unconfident that you'll be alive in a year.

Perhaps we've mangled the transition from conditional credences to conditional propositions. Perhaps we should hear "How confident are you in $P$ given $Q$?" as "How confident are you in '$P$, given $Q$'?" which is in turn "How confident are you in 'If $Q$, then $P$'?" Maybe a conditional credence is a credence in a conditional. Or perhaps more weakly: an agent assigns a particular conditional credence value whenever she unconditionally assigns that value to a conditional. In symbols, the proposal is

$$\text{cr}(P \mid Q) = r \text{ is equivalent to } \text{cr}(Q \rightarrow P) = r \tag{3.37}$$

for any propositions $P$ and $Q$, any credence distribution cr, and some indicative conditional $\rightarrow$. Reading Equation (3.37) left-to-right offers a possible analysis of conditional credences. On the other hand, some philosophers of language have read the right-to-left direction as a key to analyzing indicative conditionals.[17]

We can quickly show that Equation (3.37) fails if "$\rightarrow$" is read as the material conditional $\supset$. Under the material reading, the proposal entails that

$$\text{cr}(P \mid Q) = \text{cr}(Q \supset P) \tag{3.38}$$

Using the probability calculus and Ratio Formula, we can show that Equation (3.38) holds only when $\text{cr}(P) = 1$ (and $Q$ is nonzero). This is a *triviality result*: it shows that Equation (3.38) can hold only for the narrow range of propositions $P$ of which the agent is absolutely certain. Equation (3.38) does not express a truth that holds for *all* conditional credences in *all* propositions; nor does Equation (3.37) when "$\rightarrow$" is read materially.

Equation (3.37) can be saved from this objection by construing its "→" as something other than a material conditional. But Lewis (1976) provided a clever objection that works whichever conditional → we choose. Begin by selecting arbitrary propositions $P$ and $Q$. We then derive the following from the proposal on the table:

$$\mathrm{cr}(Q \rightarrow P) = \mathrm{cr}(P \mid Q) \qquad \text{[from Equation (3.37)]} \tag{3.39}$$

$$\mathrm{cr}(Q \rightarrow P \mid P) = \mathrm{cr}(P \mid Q \,\&\, P) \qquad \text{[see below]} \tag{3.40}$$

$$\mathrm{cr}(Q \rightarrow P \mid P) = 1 \qquad [Q \,\&\, P \text{ entails } P] \tag{3.41}$$

$$\mathrm{cr}(Q \rightarrow P \mid \sim\!P) = \mathrm{cr}(P \mid Q \,\&\, \sim\!P) \qquad \text{[see below]} \tag{3.42}$$

$$\mathrm{cr}(Q \rightarrow P \mid \sim\!P) = 0 \qquad [Q \,\&\, \sim\!P \text{ refutes } P] \tag{3.43}$$

$$\mathrm{cr}(Q \rightarrow P) = \mathrm{cr}(Q \rightarrow P \mid P) \cdot \mathrm{cr}(P) \quad + \quad \mathrm{cr}(Q \rightarrow P \mid \sim\!P) \cdot \mathrm{cr}(\sim\!P) \qquad \text{[Law of Tot. Prob.]} \tag{3.44}$$

$$\mathrm{cr}(Q \rightarrow P) = 1 \cdot \mathrm{cr}(P) + 0 \cdot \mathrm{cr}(\sim\!P) \qquad \text{[(3.41), (3.43), (3.44)]} \tag{3.45}$$

$$\mathrm{cr}(Q \rightarrow P) = \mathrm{cr}(P) \tag{3.46}$$

$$\mathrm{cr}(P \mid Q) = \mathrm{cr}(P) \qquad \text{[(3.39)]} \tag{3.47}$$

Some of these lines require explanation. The idea of lines (3.40) and (3.42) is this: We've already seen that a credence distribution conditional on a particular proposition satisfies the probability axioms. This suggests that we should think of a distribution conditional on a proposition as being just like any other credence distribution. (We'll see more reason to think this in Chapter 4, note 3.) So a distribution conditional on a proposition should satisfy the proposal of Equation (3.37) as well. If you conditionally suppose $X$, then under that supposition you should assign $Y \rightarrow Z$ the same credence you would assign $Z$ were you to *further* suppose $Y$. In other words,

$$\mathrm{cr}(Y \rightarrow Z \mid X) = \mathrm{cr}(Z \mid Y \,\&\, X) \tag{3.48}$$

In line (3.40) the roles of $X$, $Y$, and $Z$ are played by $P$, $Q$, and $P$; in line (3.42) it's $\sim\!P$, $Q$, and $P$.

Lewis has offered us another triviality result. Assuming the probability axioms and Ratio Formula, the proposal in Equation (3.37) can hold only for propositions $P$ and $Q$ such that $\mathrm{cr}(P \mid Q) = \mathrm{cr}(P)$. In other words, it can hold only for propositions the agent takes to be independent. Or (taking things from the other end), the proposed equivalence can hold for all the conditionals an agent entertains only if the agent takes every pair of propositions in $\mathcal{L}$ to be independent!

So a rational agent's conditional credence will not in general equal her unconditional credence in a conditional. This is not to say that conditional credences have *nothing* to do with conditionals. A popular idea now usually called "Adams' Thesis" (Adams 1965) holds that an indicative conditional $Q \rightarrow P$ is *acceptable* to a degree equal to $\mathrm{cr}(P \mid Q)$.[18] But we cannot maintain that an agent's conditional credence is equal to her credence that some conditional is *true*.

This brings us back to a proposal I discussed in Chapter 1. One might try to relate degrees of belief to binary beliefs by suggesting that whenever an agent has an $r$-valued credence, she has a (binary) belief with $r$ as part of its content. Working out this proposal for conditional credences reveals how hopeless it is. Suppose an agent assigns $\mathrm{cr}(P \mid Q) = r$. Would we suggest that the agent believes that if $Q$, then the probability of $P$ is $r$? This gets the logic of conditional credences wrong. Perhaps the agent believes that the probability of "if $P$, then $Q$" is $r$? Lewis's argument dooms this idea.

I said in Chapter 1 that the numerical value of an unconditional degree of belief is an attribute of the *attitude taken* towards the proposition, not a *constituent of* that proposition itself. As for conditional credences, $\mathrm{cr}(P \mid Q) = r$ does not say that an agent takes some attitude towards a conditional with a probability value in its consequent. Nor does it say that the agent takes some attitude towards a single, conditional proposition composed of $P$ and $Q$. $\mathrm{cr}(P \mid Q) = r$ says that the agent takes an $r$-valued attitude towards an *ordered pair* of propositions—neither of which need refer to the number $r$.

## 3.4 Exercises

Unless otherwise noted, you should assume when completing these exercises that the cr-distributions under discussion satisfy the probability axioms and Ratio Formula. You may also assume that whenever a conditional cr expression occurs, the condition has a nonzero unconditional credence so that the conditional credence is well-defined.

**Problem 3.1.** Suppose there are 30 people in a room. For each person, you're equally confident that their birthday falls on any of the 365 days in a year. (You're certain none of them was born in a leapyear.) Your credences about each person's birthday are independent of your credences about all the other people's birthdays. How confident are you that at least two people in the room share a birthday? (Hint: First calculate your credence that *no* two people in the room share a birthday.)

**Problem 3.2.** One might think that real humans only assign credences that are rational numbers—and perhaps only rational numbers involving relatively small whole-number numerators and denominators. But we can write down simple conditions that *require* an irrational-valued credence function. For example, take these three conditions:

1. $\text{cr}(Y \mid X) = \text{cr}(X \vee Y)$
2. $\text{cr}(X \,\&\, Y) = 1/4$
3. $\text{cr}(\sim X \,\&\, Y) = 1/4$

Show that there is exactly one credence distribution over language $\mathcal{L}$ with atomic propositions $X$ and $Y$ that satisfies all three of these conditions, and that that distribution contains irrational-valued credences.*

**Problem 3.3.** Prove that credences conditional on a particular proposition form a probability distribution. That is, prove that for any proposition $R$ in $\mathcal{L}$ such that $\text{cr}(R) > 0$, the following three conditions hold:

(a) For any proposition $P$ in $\mathcal{L}$, $\text{cr}(P \mid R) \geqslant 0$.

(b) For any tautology $\mathsf{T}$ in $\mathcal{L}$, $\text{cr}(\mathsf{T} \mid R) = 1$.

(c) For any mutually exclusive propositions $P$ and $Q$ in $\mathcal{L}$, $\text{cr}(P \vee Q \mid R) = \text{cr}(P \mid R) + \text{cr}(Q \mid R)$.

**Problem 3.4.** Pink gumballs always make my sister sick. (They remind her of Pepto Bismol.) Blue gumballs make her sick half of the time (they just look unnatural), while white gumballs make her sick only one-tenth of the time. Yesterday, my sister bought a single gumball from a machine that's one-third pink gumballs, one-third blue, and one-third white. The gumball made her sick. Applying the version of Bayes' Theorem in Equation (3.12), how confident should I be that my sister got a pink gumball yesterday?

**Problem 3.5.** (a) Prove Bayes' Theorem from the probability axioms and Ratio Formula. (Hint: Start by using the Ratio Formula to write down expressions involving $\text{cr}(H \,\&\, E)$ and $\text{cr}(E \,\&\, H)$.)

(b) Exactly which unconditional credences must we assume to be positive in order for your proof to go through?

(c) Where exactly does your proof rely on the probability axioms (and not just the Ratio Formula)?

*I owe this problem to Branden Fitelson.

**Problem 3.6.** Once more, consider the probabilistic credence distribution specified by this stochastic truth-table (from Exercise 2.5):

| $P$ | $Q$ | $R$ | cr |
|---|---|---|---|
| T | T | T | 0.1 |
| T | T | F | 0.2 |
| T | F | T | 0 |
| T | F | F | 0.3 |
| F | T | T | 0.1 |
| F | T | F | 0.2 |
| F | F | T | 0 |
| F | F | F | 0.1 |

Answer the following questions about this distribution:

(a) What is $\text{cr}(P \mid Q)$?

(b) Is $Q$ positively relevant to $P$, negatively relevant to $P$, or probabilistically independent of $P$?

(c) What is $\text{cr}(P \mid R)$?

(d) What is $\text{cr}(P \mid Q \,\&\, R)$?

(e) Conditional on $R$, is $Q$ positively relevant to $P$, negatively relevant to $P$, or probabilistically independent of $P$?

(f) Does $R$ screen off $P$ from $Q$? Explain why or why not.

**Problem 3.7.** Prove that all the alternative statements of probabilistic independence in Equations (3.15) through (3.18) follow from our original independence definition. That is, prove that each Equation (3.15) through (3.18) follows from Equation (3.14). (Hint: Once you prove that a particular equation follows from Equation (3.14), you may use it in subsequent proofs.)

**Problem 3.8.** Show that probabilistic independence is not transitive. That is, provide a single probability distribution on which all of the following are true: $X$ is independent of $Y$, and $Y$ is independent of $Z$, but $X$ is not independent of $Z$. Show that your distribution satisfies all three conditions. (For an added challenge, have your distribution assign every state-description a nonzero unconditional credence.)

**Problem 3.9.** In the text we discussed what makes a *pair* of propositions probabilistically independent. If we have a larger collection of propositions, what does it take to make them all independent of each other? You might think all that's necessary is *pairwise independence*—for each pair within the set of propositions to be independent. But pairwise independence doesn't guarantee that each proposition will be independent of *combinations* of the others.

To demonstrate this fact, describe a real-world example (spelling out the propositions represented by $X$, $Y$, and $Z$) in which it would be rational for an agent to assign credences meeting all four of the following conditions:

1. $\mathrm{cr}(X \mid Y) = \mathrm{cr}(X)$
2. $\mathrm{cr}(X \mid Z) = \mathrm{cr}(X)$
3. $\mathrm{cr}(Y \mid Z) = \mathrm{cr}(Y)$
4. $\mathrm{cr}(X \mid Y \,\&\, Z) \neq \mathrm{cr}(X)$

Show that your example satisfies all four conditions.

**Problem 3.10.** Using the program PrSAT referenced in the Further Readings for Chapter 2, find a probability distribution satisfying all the conditions in Exercise 3.9, plus the following *additional* condition: Every state-description expressible in terms of $X$, $Y$, and $Z$ must have a non-zero unconditional probability.

**Problem 3.11.** After laying down probabilistic conditions for a causal fork, Reichenbach demonstrated that a causal fork induces correlation. Consider the following four conditions:

1. $\mathrm{cr}(A \mid C) > \mathrm{cr}(A \mid {\sim}C)$
2. $\mathrm{cr}(B \mid C) > \mathrm{cr}(B \mid {\sim}C)$
3. $\mathrm{cr}(A \,\&\, B \mid C) = \mathrm{cr}(A \mid C) \cdot \mathrm{cr}(B \mid C)$
4. $\mathrm{cr}(A \,\&\, B \mid {\sim}C) = \mathrm{cr}(A \mid {\sim}C) \cdot \mathrm{cr}(B \mid {\sim}C)$

(a) Assuming $C$ is the common cause of $A$ and $B$, explain what each of the four conditions means in terms of relevance, independence, conditional relevance, or conditional independence.

(b) Prove that if all four conditions hold, then $\mathrm{cr}(A \,\&\, B) > \mathrm{cr}(A) \cdot \mathrm{cr}(B)$. (This is a tough one!)

**Problem 3.12.** In Section 3.3 I pointed out that the following statement (labeled Equation (3.36) there) is false:

$$\text{cr}(C \mid A) = k \text{ and } \text{cr}(C \mid B) = k \text{ entail } \text{cr}(C \mid A \vee B) = k$$

(a) Describe a real-world example (involving dice, or cards, or something more interesting) in which it's rational for an agent to assign $\text{cr}(C \mid A) = k$ and $\text{cr}(C \mid B) = k$ but $\text{cr}(C \mid A \vee B) \neq k$. Show that your example meets this description.

(b) Prove that if $A$ and $B$ are mutually exclusive, then whenever $\text{cr}(C \mid A) = k$ and $\text{cr}(C \mid B) = k$ it's also the case that $\text{cr}(C \mid A \vee B) = k$.

**Problem 3.13.** Fact: For any propositions $P$ and $Q$, if $\text{cr}(Q) > 0$ then $\text{cr}(Q \supset P) \geqslant \text{cr}(P \mid Q)$.

(a) Use a stochastic truth-table built on propositions $P$ and $Q$ to prove this fact.

(b) Show that Equation (3.38) in Section 3.3 entails that $\text{cr}(P) = 1$.

## 3.5 Further reading

INTRODUCTIONS AND OVERVIEWS

Todd A. Stephenson (2000). *An Introduction to Bayesian Network Theory and Usage*. Tech. rep. 03. IDIAP

Section 1 provides a nice, concise overview of what a Bayes Net is and how it interacts with conditional probabilities. (Note that the author uses $A, B$ to express the *conjunction* of $A$ and $B$.) Things get fairly technical after that as he covers algorithms for creating and using Bayes Nets. Sections 6 and 7, though, contain real-life examples of Bayes Nets for speech recognition, Microsoft Windows troubleshooting, and medical diagnosis.

CLASSIC TEXTS

Hans Reichenbach (1956). The Principle of Common Cause. In: *The Direction of Time*. University of California Press, pp. 157–160

Article in which Reichenbach introduces his account of common causes in terms of screening off. (Note that Reichenbach uses a period to express conjunction, and a comma rather than a vertical bar for conditional probabilities—what we would write as $\text{cr}(A \mid B)$ he writes as $\text{P}(B, A)$.)

David Lewis (1976). Probabilities of Conditionals and Conditional Probabilities. *The Philosophical Review* 85, pp. 297–315

Article in which Lewis presents his triviality argument concerning probabilities of conditionals.

<u>Extended Discussion</u>

Frank Arntzenius (1993). The Common Cause Principle. *PSA 1992* 2, pp. 227–237

Presents a number of objections to Reichenbach's Principle of the Common Cause, with citations (when the objections aren't original to Arntzenius himself).

Alan Hájek and Ned Hall (1994). The Hypothesis of the Conditional Construal of Conditional Probability. In: *Probability and Conditionals: Belief Revision and Rational Decision.* Ed. by Ellery Eells and Brian Skyrms. Cambridge Studies in Probability, Induction, and Decision Theory. Cambridge University Press, pp. 75–112

Hájek and Hall extensively assess views about conditional credences and credences in conditionals in light of Lewis's and other triviality results.

## Notes

[1]Here's a good way to double-check that $6 \,\&\, E$ is equivalent to 6: Remember that equivalence is mutual entailment. Clearly $6 \,\&\, E$ entails 6. Going in the other direction, 6 entails 6, but 6 also entails $E$. So 6 entails $6 \,\&\, E$. When evaluating conditional credences using the Ratio Formula we'll often find ourselves simplifying a conjunction down to just one or two of its conjuncts. To make this work, the conjunct that remains has to entail each of the conjuncts that was removed.

[2]Some authors take advantage of this fact to formalize probability theory in exactly the opposite order from what I've pursued here. They begin by introducing conditional probabilites or credences and subject them to a number of constraints somewhat like Kolmogorov's axioms. All the desired rules for *un*conditional credences are then obtained by introducing the single constraint that $\text{cr}(P) = \text{cr}(P \mid \top)$. Just as the Ratio Formula helps us transform constraints on unconditional credences into constraints on conditional credences, this rule transforms constraints on conditionals into constraints on unconditionals. For examples of this conditional-probability-first approach, see (Popper 1955), (Renyi 1970), and (Roeper and Leblanc 1999).

[3]Bayes never published the theorem; Richard Price found it in Bayes' notes and published it after Bayes' death in 1761. Pierre-Simon Laplace independently rediscovered the theorem later on and was responsible for much of its early popularization.

[4]In everyday English "likely" is a synonym for "probable". Yet R.A. Fisher introduced the technical term "likelihood" to represent a particular *kind* of probability—the probability of some evidence given a hypothesis. This somewhat peculiar terminology has stuck.

[5]Quoted in (Galavotti 2005, p. 51).

[6]I'm assuming throughout this discussion that $P$, $\sim P$, $Q$, and $\sim Q$ all have non-zero unconditional credences so that the relevant conditional credences are well-defined.

[7]Different authors define "screening off" in different ways. For example, while conditional independence is interesting only when the propositions in question are unconditionally correlated, most authors leave out the requirement that $P$ be unconditionally relevant to $Q$. (I suppose one could alter my definition so that unconditionally-independent $P$ and $Q$ would count as trivially screened off by anything.)

Many authors will say that $R$ screens off $P$ and $Q$ only when $P$ and $Q$ are independent not only conditionally on $R$ but also conditionally on $\sim R$. This is part of a more general view on which propositions are simply dichotomous random variables (see Chapter 2, note 6) and a random variable $X$ screens off $Y$ from $Z$ only if $Y$ and $Z$ are independent conditional on every possible value of $X$. If one takes this kind of position, then in Bob's credence function $I$ does *not* count as screening $H$ off from $C$, because $H$ and $C$ are correlated conditional on $\sim I$. Adopting this alternate definition would lead us to recharacterize some of the examples I'll soon discuss, but would not make any difference to the causal examples we'll eventually consider.

[8]Not to be confused with the Rambler's Fallacy: I've said so many false things in a row, the next one must be true!

[9]20 flips of a fair coin provide a good example of what statisticians call **IID trials**. "IID" stands for "independent, identically distributed." Each of the coin flips is probabilistically independent of all the others; information about the outcomes of other coin flips doesn't change the probability that a particular flip will come up heads. The flips are identically distributed because each has the same probability of producing a heads outcome.

Anyone who goes in for the Gambler's Fallacy and thinks that future flips will make up for past outcomes is committed to the existence of some *mechanism* by which future flips can respond to what happened in the past. Understanding that no such mechanism exists leads one to treat repeated flips of the same coin as IID.

[10]I learned about the Jeter/Justice example from the Wikipedia page on Simpson's Paradox. (The batting data for the two hitters is widely available.) The UC Berkeley example was brought to the attention of philosophers by (Cartwright 1979).

[11]An analogy: Suppose we each have some gold bars and some silver bars. Each gold bar you're holding is heavier (and therefore more valuable) than each of my gold bars. Each silver bar you're holding is heavier (and more valuable) than each of my silver bars. Then how could I possibly be richer than you? If I have many more gold bars than you, while you have more silver than I.

[12]You may be concerned that arithmetic facts are true in every possible world, and so cannot rationally receive nonextreme credences, and so cannot be probabilistically correlated. We'll come back to that concern in Chapter XXX.

[13]I'm playing a bit fast and loose with the objects of discussion here. Throughout this chapter we're considering correlations in an agent's credence distribution. Reichenbach

was concerned not with probabilistic correlations in an agent's credences but instead with correlations in objective frequency or chance distributions (more about which in Chapter 5). But presumably if the Principle of the Common Cause holds for objective probability distributions, that provides an agent who views particular propositions as empirically correlated some reason to suppose that the events described in those propositions either stand as cause to effect or share a common cause.

[14] You might worry that Figure 3.3 presents a counterexample to Reichenbach's Principle of the Common Cause, because $G$ and $S$ are unconditionally correlated yet $G$ doesn't cause $S$ and they have no common cause. It's important to the principle that the causal relations need not be *direct*; for Reichenbach's purposes $G$ counts as a cause of $S$ even though it's not the immediate cause of $S$.

[15] Just to indicate a few more complexities that can arise: One can have a common cause (an "indirect" common cause) that doesn't screen off its effects from each other. For example, if we imagine merging Figures 3.2 and 3.3 to show how the subject's parents' genes are a common cause of both smoking and drinking by way of her addictive personality, it is possible to arrange the numbers so that her parents' genetics don't screen off smoker from drinker. Even more complications arise if some causal arrows do end-arounds past others—what if in addition to the causal structure just described, the parents' genetics tend to make them smokers which in turn influences the subject's smoking behavior?

[16] One *could* study a kind of attitude different from the conditional credences in this book, and to which the Ratio Formula applies—something like a subjunctive degree of belief. Joyce (1999) does exactly that, but is careful to distinguish his analysandum from standard conditional degrees of belief.

[17] While Equation (3.37) has often been read as an *analysis* in one direction or another, it could also be read as a normative constraint: If an agent is rational, she will assign the same value to $\mathrm{cr}(P \mid Q)$ as she does to $\mathrm{cr}(Q \rightarrow P)$ for any $P$ and $Q$. Since our arguments against Equation (3.37) will be based on the probability axioms and Ratio Formula, and since we assume that rational credences also satisfy those constraints, such a normative interpretation would ultimately be vulnerable to our arguments as well.

[18] Interestingly, this idea is often traced back to a suggestion in Ramsey, known as "Ramsey's test". (Ramsey 1929/1990, p. 155n)