Chapter 5

Further Rational Constraints

The previous three chapters have discussed the five core normative Bayesian rules: Kolmogorov's probability axioms, the Ratio Formula, and Conditionalization. Bayesians offer these rules as necessary conditions for an agent's credences to be rational. We have not discussed whether these five rules are jointly sufficient for rational credence.

Agents can satisfy the core rules and still have wildly divergent attitudes. If you and I draw 1,000 balls from an urn and every one of them is black, I might be highly confident that the next ball is black. But there is a credence distribution satisfying all the rational constraints we have mentioned so far that would allow you to be 99% confident the next ball will be white. Similarly, if I tell you I have rolled a fair die but don't tell you how the roll came out, there is a probabilistic distribution that will allow you to assign credence 0.8 that it came up 3.

If we think these credence assignments are irrational, we need to identify additional rational requirements beyond the Bayesian core that rule them out. We have already seen one potential requirement that goes beyond the core: the Regularity Principle (Section 4.2) prohibits assigning credence 0 to logically contingent propositions. What other requirements on rational credence might there be? And could they be strong enough to limit us to exactly one rationally-permissible credence assignment for each body of total evidence?

The answer to this last question is sometimes taken to separate Subjective from Objective Bayesians. Unfortunately, "Objective/Subjective Bayesian" terminology is used ambiguously, so this chapter begins by distinguishing two different ways in which that distinction is used. In the course of doing so we'll cover various interpretations of probability, including frequency and propensity views.

Then we will consider a number of additional rational credence constraints proposed in the Bayesian literature. We will begin with synchronic constraints: the Principal Principle (relating credences to chances); the Reflection Principle (concerning one's current credences about one's future credences); principles for deferring to experts; indifference principles (for distributing credences in the absence of evidence); and principles for distributing credences over infinitely many possibilities. Finally, we will turn to Jeffrey Conditionalization, an alternative diachronic updating principle to standard Conditionalization.

5.1 Subjective and Objective Bayesianism

When a weather forecaster comes on television and says, "The probability of snow tomorrow is 30%," what does that mean? What exactly has the weather forecaster communicated to the audience? Such questions have been asked throughout the history of mathematical probability theory; in the twentieth century, rival answers became known as **intepretations of probability**. There is an excellent literature devoted to this topic and its history (see the Further Readings of this chapter for recommendations), so I don't intend to let it take over this book. But for our purposes it's important to touch on some of the main interpretations, and at least mention some of their advantages and disadvantages.

5.1.1 Frequencies and Propensities

If an event has a 30% probability of producing a certain outcome, we expect that were the event repeated it would produce that type of outcome about 30% of the time. The **frequency theory** of probability offers this fact as an *analysis* of "probability". On this interpretation, when the weather forecaster says the probability of snow tomorrow is 30%, she means that days like tomorrow produce snow 30% of the time. According to the frequency theory, the probability is x that event A will have outcome B just in case proportion x of events like A have outcomes like B.[1] The frequency theory originated in work by Robert Leslie Ellis (1849) and John Venn (1866), then was famously developed by the logical positivist Richard von Mises (1928/1957).

The frequency theory has a number of problems; I will mention only a few.[2] Suppose my sixteen-year-old daughter asks for the keys to my car; I wonder what the probability is that she will get into an accident should I give her the keys. According to the frequency theory, the probability that the event of my giving her the keys will have the outcome of an accident is determined by how frequently this type of event leads to accidents. But what type of event is it? Is the frequency in question how often people who go driving get into accidents? How often sixteen-year-olds get into accidents? How often sixteen-year-olds with the courage to ask their fathers for the keys get into accidents? How often my daughter gets into accidents? Presumably these frequencies will differ—which one is *the* probability that if I give my daughter the keys on this occasion she will get into an accident?

Any event can be subsumed under many types, and the frequency theory leaves it unclear which event-types determine probability values. Event types are sometimes known as reference classes, so this is the **reference class problem**. In response, one might suggest that outcomes have frequencies—and therefore probabilities—only *relative* to the specification of a particular reference class (either implicitly or explicitly). But it seems we can meaningfully inquire about the probabilities of particular event outcomes (or of propositions simpliciter) without specifying a reference class. I need to decide whether to give the keys to my daughter; I want to know how probable it is that she will crash. To which reference class should I relativize?

Frequency information about specific event-types seems more relevant to determining probabilities than information about general types. (The probability that my daughter will get into an accident on this occasion seems much closer to *her* frequency of accidents than to the accident frequency of drivers in general.) Perhaps probabilities are frequencies in the maximally specific reference class? But the *maximally* specific reference class containing a particular event contains only that individual event. The frequency with which my daughter gets into an accident when I give her my keys *on this occasion* is either 0 or 1—but we often think probabilities for such events have nonextreme values.

This raises another problem for frequency theories. Suppose I have a fair coin and am considering the probability that if I flip it the outcome will be heads. We'll just grant that the correct reference class for this event is flips of fair coins. According to the frequency theory, the probability that this fair coin flip will come up heads is the fraction of all fair coin flips that ever occur which come up heads. While I'd be willing to bet that number is *close* to 1/2, I'd be willing to bet even more that the fraction is not *exactly* 1/2.[3] So why do we say the probability that a fair coin comes up heads is 1/2?

One might respond that the probability of heads on a fair coin flip is not the frequency with which fair coin flips *actually* come up heads over the finite history of our universe; instead, it's the frequency *in the limit*—were such coins to continue being flipped forever. But now consider events that couldn't possibly be repeated so many times, or even events that couldn't be repeated once. Before the Large Hadron Collider was switched on, physicists were asked for the probability that doing so would destroy the Earth. Were that to have happened, switching on the Large Hadron Collider would not have been a repeatable event. Scientists also sometimes discuss the probability that our universe began with a Big Bang; arguably, that's not an event that will happen over and over or even could happen over and over. The problem of assigning meaningful nonextreme probabilities to individual, perhaps non-repeatable events is called the **problem of the single case**.

The frequency theory can address such problems, but at a significant cost. We first move from frequencies of outcomes in finite, actual repetitions of an event to frequencies that would be approached as the number of repetitions tended towards the limit. This gives us **hypothetical frequency theory**. Yet this move undermines one of the original appeals of the frequency theory: its empiricism. The proportion of event repetitions that produce a particular outcome in the actual world is the sort of thing that could be observed (at least in principle)—providing a sound empirical base for otherwise-mysterious probability talk. Empirically grounding hypothetical frequencies is a much more difficult task. Matters become worse if we address non-repeatable events by considering not multiple event occurrences within one world, but instead single event occurrences across multiple possible worlds. It's awfully difficult to tally up the outcomes of such event sets empirically.[4]

An alternate interpretation of probability admits that probabilities are related to frequencies, but draws our attention to the features that *cause* particular outcomes to appear with the frequencies that they do. When we have a fair coin, what is it about the coin that makes it fair? Something about its physical attributes, the symmetries with which it interacts with surrounding air as it flips, etc. These traits lend the coin a certain tendency to come up heads, and an equal tendency to come up tails. This quantifiable disposition—or **propensity**—gives rise to facts about frequencies in hypothetical long-run trials. But the propensity is also at work in each individual flip of the coin, whether that flip is ever repeated or could ever be repeated. Even a non-repeatable event can have a nonextreme propensity to generate a particular outcome.

While an early propensity theory appeared in the work of Charles Sanders Peirce (1910/1932), its most famous champion was Karl Popper (1957). Popper was especially driven to this view by developments in quantum mechanics. In quantum theory the Born rule calculates probabilities of experimental outcomes from a particular quantity (the amplitude of the wave-function) that has independent significance in the theory's dynamics. Moreover, this quantity can be determined for a particular experimental setup even if that setup is never to be repeated (or *couldn't* be repeated) again. This gives propensities a respectable place within an empirically-established scientific theory. Propensities (sometimes called **objective chances**) also seem to play an important role in such theories as thermodynamics and population genetics.

Yet even if there are propensities in the world, that doesn't mean all probabilities are propensities. Suppose we're discussing the likelihood that a particular outcome occurs given that a quantum experiment is set up in a particular fashion. This is a conditional probability, and it has a natural interpretation in terms of physical propensities. But where there is a likelihood, probability mathematics suggests there will also be a posterior—if there's a probability of outcome given setup, there should also be a probability of setup given outcome. Yet the latter hardly makes sense as a physical propensity—does an experimental outcome have a quantifiable tendency to set up the experiment that produces it in a particular way?[5]

Further, propensities are generated by arrangements of matter and energy as governed by the physical laws of our world. Since physical laws give rise to the propensities, there can be no propensity for the physical laws to be one way or another. (What set of laws beyond the physical might determine those propensities?) Yet it seems physicists can meaningfully discuss the probability that the physical laws of the universe are one way rather than another. While the propensity interpretation may make sense of some of our probability talk, it nevertheless seems to leave a remainder.

5.1.2 Two Distinctions

So what *are* physicists talking about when they discuss the probability that the physical laws of the universe are one way rather than another? Perhaps they are expressing their degrees of confidence in alternative physical hypotheses. Perhaps there are no probabilistic facts out in the world, about which our opinions change as we gain evidence. Instead, it may be that facts in the world are just facts, true or false, probability-free, and our probability talk records our changing confidences in those facts in the face of changing evidence.

Bayesian theories are often characterized as "Subjective" or "Objective", but this distinction is used in two ways. One of them is about the interpretation of probability talk. On this distinction—which I'll call the **semantic distinction**—Subjective Bayesians adopt the position I proposed in the previous paragraph. For them, probability talk expresses or reports the degrees of confidence of the individuals doing the talking, or perhaps of communities to which they belong. Objective Bayesians, on the other hand, interpret probability assertions as having truth-conditions independent of the attitudes of particular agents.[6] In the twentieth century, talk of "Objective" and "Subjective" Bayesianism was usually used to draw a semantic distinction.[7]

More recently "Subjective Bayesian" and "Objective Bayesian" have been used to draw a different distinction, which I will call the **normative distinction**. However we interpret the meaning of probability talk, we can grant that agents assign different degrees of confidence to different propositions (or, more weakly, that it is at least useful to model agents as if they do). Once we grant that credences exist and are subject to rational constraints, we may inquire about the stringency of those constraints.

On one end of the spectrum, Objective Bayesians (in the normative sense) endorse what Richard Feldman (2007) and Roger White (2005) have called the

Uniqueness Thesis: Given any proposition and body of total evidence, there is exactly one attitude it is rationally permissible for agents with that body of total evidence to adopt towards that proposition.

Assuming the attitudes in question are degrees of belief, the Uniqueness Thesis says that in any situation there's exactly one credence an agent is rationally permitted to take towards a given proposition, and the value of that credence is determined by her evidence. The Uniqueness Thesis therefore entails **evidentialism**, according to which the attitudes rationally permissible for an agent supervene on her evidence.

Suppose we have two agents with identical total evidence who adopt different credences in some propositions. Because they endorse the Uniqueness Thesis, Objective Bayesians (in the normative sense) hold that at least one of these agents is responding to her evidence irrationally. Notice that whatever is causing the difference in these agents' attitudes, it cannot be the contents of their evidence (because we stipulated that their total evidence is identical). In Section 4.3 we identified the extra-evidential factors that determine an agent's attitudes in light of her evidence as her "ultimate evidential standards". These evidential standards might reflect pragmatic influences, predilections for certain kinds of hypotheses, a tendency towards mistrust or skepticism, etc.

In a credal context, the Hypothetical Priors Theorem tells us that whenever an agent's credence distributions over time satisfy the probability axioms, Ratio Formula, and Conditionalization, her evidential standards can be represented by a hypothetical prior distribution. This regular, probabilistic distribution stays constant as the agent gains evidence over time. Yet we can always recover the agent's credence distribution at a given time by conditionalizing her hypothetical prior on her total evidence at that time.

The core Bayesian rules (probability axioms, Ratio Formula, Conditionalization) leave a wide variety of hypothetical priors available. Assuming they satisfy the core rules, our two agents who assign different credences in response to the same total evidence must have different hypothetical priors. According to the Objective Bayesian, any time such a situation arises at least one of the agents must be violating rational requirements. Thus the Objective Bayesian (in the normative sense) thinks there is exactly one set of rationally permissible hypothetical priors—one set of correct evidential standards that embodies all rational agents' common responses to evidence. If we think of hypothetical priors as the input that, given a particular evidential situation, produces an agent's credence distribution as the output, then the only way for Objective Bayesians to secure unique outputs in every situation is to demand a universal unique input.

How might the unique correct hypothetical prior be generated, and how might it be justified? Our ongoing evidential standards for responding to particular pieces of evidence are often informed by other pieces of evidence we have received in the past. I take a fire alarm to support a particular hypothesis about what's going on in my building because I have received past evidence about the import of such alarms. But if we go back far enough this process must stop somewhere; our *ultimate* evidential standards, represented by our hypothetical priors, encode responses to our *total* evidence, and so cannot be responses to elements of that evidence. If we are to select and justify a unique set of such ultimate evidential standards, we must do so *a priori*.

Keynes (1921) and Carnap (1950) thought that just as there are objective facts about which propositions are logically entailed by a given body of evidence, there are objective logical facts about the degree to which a body of evidence probabilifies a particular proposition. Carnap went on to offer a mathematical algorithm for calculating the uniquely logical hypothetical priors from which these facts could be determined; we will discuss that algorithm in Chapter 6. (The **logical interpretation** of probability holds that an agent's probability talk concerns logical probabilities relative to her current total evidence.[8]) Many recent theorists, while backing away from Keynes' and Carnap's position that these values are *logical*, nevertheless embrace the idea of **evidential probabilities** reflecting the degree to which a proposition is probabilified by a body of evidence. If you think that rationality requires an agent to assign credences equal to the true evidential probabilities on her current total evidence, you have an Objective Bayesian view in the normative sense.[9]

At the other end of the spectrum from Objective Bayesians (in the normative sense) are theorists who hold that the probability axioms and Ratio Formula are the only rational constraints on hypothetical priors.[10] The literature often defines "Subjective Bayesians" as people who hold this view. But that terminology leaves no way to describe theorists in the middle of the spectrum—the vast majority of Bayesian epistemologists who believe in rational constraints on hypothetical priors that go beyond the probability axioms but are insufficient to narrow us down to a single viable standard. I will use the term "Subjective Bayesian" (in the normative sense) to refer to anyone who thinks more than one hypothetical prior is rationally permissible. I will call people who think the Ratio Formula and probability axioms are the only rational constraints on hypothetical priors "extreme Subjective Bayesians".

Subjective Bayesians allow for what White calls **permissive cases**: examples in which two agents reach different conclusions on the basis of the same evidence without either party's making a rational mistake. This is because each agent interprets the evidence according to different (yet equally rational) evidential standards, which allows them to draw different conclusions.

I have distinguished the semantic and normative Objective/Subjective Bayesian distinctions because they can cross-cut one another. Historically, Ramsey (1931) and de Finetti (1931/1989) reacted to Keynes' Objective Bayesianism with groundbreaking theories that were Subjective in both the semantic and normative senses. But one could be a Subjective Bayesian in the semantic sense—taking agents' probability talk to express their own current credences—while maintaining that strictly speaking only one credence distribution is rationally permitted in each situation (thereby adhering to Objective Bayesianism in the normative sense). Going in the other direction, one could admit the existence of degrees of belief while holding that they're not what probability talk is about. This would give an Objective Bayesian semantic view that combined with either Subjective or Objective Bayesianism in the normative sense. Finally, probability semantics need not be monolithic; many Bayesians now think that some probability assertions express credences, others report objective chances, while still others indicate what would be reasonable to think given one's evidence.

Regardless of your position on the semantics, as long as you aren't an extreme Subjective Bayesian in the normative sense you will concede that there are rational constraints on agents' hypothetical priors beyond the probability axioms and Ratio Formula. What might those constraints be? We will now investigate some proposals.

5.2 Deference Principles

5.2.1 The Principal Principle

Suppose it is now 1pm on a Monday. I tell you that over the weekend I found a coin from a foreign country that is somewhat irregular in shape. Despite being foreign, one side of the coin is clearly the "Heads" side and the other is "Tails". I also tell you that I flipped the foreign coin today at noon.

Let H be the proposition that the noon coin flip landed heads. Consider each of the propositions below one at a time, and decide what your credence in H would be if that proposition was *all* you knew about the coin besides the information in the previous paragraph:

- **E_1:** After discovering the coin I spent a good part of my weekend flipping it, and out of my 100 weekend flips 64 came up heads.
- **E_2:** The coin was produced in a factory that advertises its coins as fair, but also has a side business generating black-market coins biased towards tails.
- **E_3:** The coin is fair.
- **E_4:** Your friend Amir was with me at noon when I flipped the coin, and he told you it came up heads.

Hopefully it's fairly clear how to respond to each of these pieces of evidence, taken singly. For instance, in light of the frequency information in E_1, it seems rational to have a credence in H somewhere around 0.64. We might debate whether precisely 0.64 is required, but certainly a credence in H of 0.01 (assuming E_1 is your *only* evidence about the coin) seems unreasonable.

This point generalizes to a rational principle that whenever one's evidence includes the frequency with which events of type A have produced outcomes of type B, one should set one's credence that the next A-event will produce a B-outcome equal to (or at least in the vicinity of) that frequency.[11] While some version of this principle ought to be right, working out the specifics creates problems like those faced by the frequency interpretation of probability. For instance, we have a reference class problem: Suppose my evidence includes accident frequency data for drivers in general, sixteen-year-old drivers in general, and my sixteen-year-old daughter in particular. Which value should I use to set my credence that my daughter will get in a car accident tonight? The more specific data seems more relevant, but the more general data contains a larger sample size.

There are well-known statistical tools for dealing with these problems, some of which we will discuss in Chapter XXX. But for now let's focus on a different question about frequency data: *Why* do we use known flip outcomes to predict the outcome of unobserved flips? Perhaps because known outcomes indicate something about the physical properties of the coin itself; they help us figure out its propensity (or objective chance) of coming up heads. In Section 5.1 I raised some problems for thinking that objective chances provide a universal semantics for probability-talk. But one can believe that chances exist and are often important to our reasoning about probabilities without thinking that probability talk always references chance. It's very plausible that known flip data influences our unknown flip predictions because it makes us think our coin has a particular chance profile. In this case, frequency data influences predictions *by way of* our opinions about chances.

This relationship between frequency and chance is revealed when we *combine* pieces of evidence listed above. We've already said that if your only evidence about the coin is E_1—it came up heads on 64 of 100 known tosses—then your credence that the noon toss (of uncertain outcome) came up heads should be around 0.64. On the other hand, if your only evidence is E_3, that the coin is fair, then I hope it's plausible that your credence in H should be 0.5. But what if you're already certain of E_3, and then learn E_1? In that case your credence in heads should still be 0.5.

Keep in mind we're imagining you're *certain* that the coin is fair before you learn the frequency data; we're not concerning ourselves with the possibility that, say, learning about the frequencies makes you suspicious of the source from which you learned the coin is fair. If it's a fixed, unquestionable truth for you that the coin is fair, then learning it came up 64 heads on 100 flips will not change your credence in heads. If *all* you had was the frequency information, that would support a different hypothesis about the chances. But it's not as if 64 heads on 100 flips is *inconsistent* with the coin's being fair—even a fair coin usually won't come up heads on exactly half the flips in a given sample. So once you're already certain of heads, the frequency information becomes redundant, irrelevant to your opinions about unknown flips. Frequencies help you learn about chances, so if you are already certain of the chances there's nothing more for frequency information to do.

David Lewis (1980) called information that changes your credences about an event *by way of* changing your opinions about the chances **admissible** information. His main insight about admissible information was that when the chance values for an event have already been established, admissible information becomes irrelevant to a rational agent's opinions about the outcome.

Here's another example: Suppose your only evidence about the noon flip outcome is E_2, that the coin was produced in a factory that advertises its coins as fair but has a side business in tails-biased coins. Given only this information your credence in H should be somewhere below 0.5. (Exactly how far below depends on how extensive you estimate the side business to be.) On the other hand, suppose your learn E_2 after already learning E_3, that the coin is fair. E_2 then becomes unimportant information, at least with respect to predicting flips of the coin. E_2 is relevant in isolation because it informs you about the propensities of the coin. But once you're certain that the coin is fair, further possessing information E_2 only teaches you that you happened to get lucky not to have a black-market coin; it doesn't do anything to push your credence in H away from 0.5. E_2 is admissible information.

Contrast that with E_4, your friend Amir's report that he observed the flip landing heads. Assuming you trust Amir, E_4 should make you highly confident in H. And this should be true even if you already possess information E_3 that the coin is fair. Notice that E_3 and E_4 are consistent; the coin's being fair is consistent with its having landed heads on this particular flip, and with Amir's reporting that outcome. But E_4 trumps the chance information; it moves your credence in heads away from where it would be (0.5) if you knew only E_3. Information about this particular flip's outcome does not change your credences about the flip *by way of* influencing your opinions about the chances. You still think the coin is fair, and was fair at the time it was flipped. You just know now that the fair coin happened to come up heads on this occasion. Information about this flip's outcome is inadmissible with respect to H.

Lewis expressed his insight about the irrelevance of admissible information in a principle relating chance and rational credence, which he called the

Principal Principle: Let cr_H be any reasonable initial credence function. Let t_i be any time. Let x be any real number in the unit interval. Let $\mathrm{Ch}_i(A) = x$ be the proposition that the chance, at time t_i, of A's holding equals x. Let E be any proposition compatible with $\mathrm{Ch}_i(A) = x$ that is admissible at time t_i. Then

$$\mathrm{cr}_H(A \,|\, \mathrm{Ch}_i(A) = x \;\&\; E) = x$$

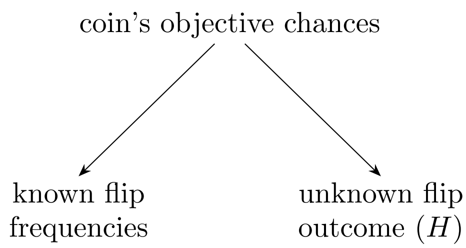
(I have copied this principle verbatim from (1980, p. 266), though I have altered Lewis' notation to match our own.) There's a lot to unpack in the Principal Principle, so we'll take it one step at a time. By "reasonable initial credence function" Lewis means what we have been calling a "rational hypothetical prior" (hence my use of cr_H in the equations). So the Principal Principle is proposed as a rational constraint on hypothetical priors, one that goes beyond the probability axioms and Ratio Formula. In particular, the Principal Principle is a principle of direct inference (Section 3.1.3), constraining credences in an experimental outcome relative to a chance hypothesis.

Why frame the Principal Principle around hypothetical priors, instead of focusing on the credences of rational agents at particular times? One advantage is that this makes the total evidence in question explicit, and therefore easy to reference in the principle. Recall from Section 4.3 that a hypothetical prior is a probabilistic, regular distribution containing no contingent evidence. A rational agent is associated with a particular hypothetical prior, in the sense that if you conditionalize that hypothetical prior on the agent's total evidence at any given time, you get the agent's credence distribution at that time.

In the Principal Principle, we imagine that a real-life agent is considering some proposition A about the outcome of a chance event. She has some information about the chance of A, $\mathrm{Ch}_i(A) = x$, and then some further evidence E. So her total evidence is $\mathrm{Ch}_i(A) = x \;\&\; E$, and by the definition of a hypothetical prior her credence in A equals $\mathrm{cr}_H(A \,|\, \mathrm{Ch}_i(A) = x \;\&\; E)$. Lewis claims that as long as E is both admissible for A, and is compatible (which we can take to mean "logically consistent") with $\mathrm{Ch}_i(A) = x$, E should make no difference to the agent's credence in A. In other words, as long as E is admissible and compatible, the agent should be just as confident in A as she would be if all she knew were $\mathrm{Ch}_i(A) = x$. That is, her credence in A should be x.

Return to our example about the noon coin flip, and the relationship between chance and frequency information. Suppose that at 1pm your total

Figure 5.1: Chances screen off frequencies



evidence about the flip outcome consists of E_1 and E_3. E_3, the chance information, says that $\text{Ch}(H) = 0.5$. E_1, the frequency information, supplies your additional total evidence, which will play the role of E in the Principal Principle. Because this additional evidence is both consistent with $\text{Ch}(H) = 0.5$ and admissible for H, the Principal Principle says your 1pm credence in H should be 0.5. Which is exactly the result we came to before.

Yet the Principal Principle provides further insight into this result by connecting it to our earlier (Section 3.2.4) discussion of causation and screening off. Figure 5.1 illustrates the causal relationship between chances, frequencies, and unknown results in the coin example. The coin's physical propensity to come up heads causally influenced the frequency with which it came up heads in the observed trials. The coin's physical makeup also affects the outcome of the unknown flip. Thus frequency information is relevant to the unknown flip, but only by way of the chances. We saw in Section 3.2.4 that when this kind of causal fork structure obtains, the common cause screens its effects off from each other. Conditional on the chances, frequency information becomes irrelevant to flip predictions. That is,

$$\text{cr}_H(H \mid \text{Ch}(H) = 0.5 \,\&\, E) = \text{cr}_H(H \mid \text{Ch}(H) = 0.5) \tag{5.1}$$

and intuitively the expression on the right should equal 0.5.

A similar analysis applies if your total evidence about the coin flip contains only $\text{Ch}(H) = 0.5$ and E_2, the evidence about the coin factory. This time our structure is a causal chain, as depicted in Figure 5.2. The situation in the coin factory causally affects the chance profile of the coin, which in turn causally affects the unknown flip outcome. Thus the coin factory information affects opinions about H by way of the chances, and if the chances are already determined then factory information becomes irrelevant. Letting the factory information play the role of E in the Principal Principle,

Figure 5.2: Chance in a causal chain

$$\begin{array}{c}\text{coin factory details}\\ \downarrow \\ \text{chance of heads } (\mathrm{Ch}(H))\\ \downarrow \\ \text{unknown flip outcome } (H)\end{array}$$

the chances screen off E from H and we have the relation in Equation (5.1).

Finally, information E_4, your friend Amir's report, is not admissible information about H. E_4 affects your opinions about H, but not by way of affecting your opinions about the chances. The Principal Principle applies only when E, the information possessed in addition to the chances, is admissible. Since E_4 is inadmissible, the Principal Principle supplies no guidance about setting your credences in light of it.

There are still a few details in the principle to unpack. For instance, you'll notice that the chance expression $\mathrm{Ch}_i(A)$ is indexed to a time t_i. That's because the chance that a particular proposition will obtain can change as time goes on. For instance, suppose that at 11am our foreign coin was fair, but at 11:30 I stuck a particularly large, non-aerodynamic wad of chewing gum to one of its side. In that case, the proposition H that the coin comes up heads at noon would have a chance of 0.5 at 11am but might have a different chance after 11:30. The physical details of an experimental setup determine its chances, so as physical conditions change chances may change as well.[12]

Finally, the Principal Principle's formulation in terms of conditional credences allows us to apply it even when an agent doesn't have full information about the chances. Suppose your total evidence about the outcome A of some chance event is E. E influences your credences in A by way of informing you about A's chances (so E is admissible), but E does not tell you what the chances are exactly. Instead, E tells you that the chance of A (at some time, which I'll suppress for the duration of this example) is either 0.7 or 0.4. E also supplies you with a favorite among these two chance hypotheses: it sets your credence that 0.7 is the true chance at 2/3, and your credence that 0.4 is the true chance at 1/3.

How can we analyze this situation using the Principal Principle? Since your total evidence is E, the definition of a hypothetical prior tells us that your current credences cr should be related to your hypothetical prior cr_H as follows:

$$\mathrm{cr}(A) = \mathrm{cr}_H(A \mid E) \tag{5.2}$$

This value is not dictated directly by the Principal Principle. However, the Principal Principle does set

$$\mathrm{cr}_H(A \mid \mathrm{Ch}(A) = 0.7 \,\&\, E) = 0.7 \tag{5.3}$$

because we stipulated that E is admissible. (For simplicity's sake I'm not worrying about time-indexing the chances.) Similarly, the Principal Principle sets

$$\mathrm{cr}_H(A \mid \mathrm{Ch}(A) = 0.4 \,\&\, E) = 0.4 \tag{5.4}$$

Since E narrows the possibilities down to two mutually exclusive chance hypotheses, those hypotheses ($\mathrm{Ch}(A) = 0.7$ and $\mathrm{Ch}(A) = 0.4$) form a partition relative to E. Thus we can apply the Principle of Total Evidence (in its conditional credence form) to obtain

$$\begin{aligned}\mathrm{cr}_H(A \mid E) = &\mathrm{cr}_H(A \mid \mathrm{Ch}(A) = 0.7 \,\&\, E) \cdot \mathrm{cr}_H(\mathrm{Ch}(A) = 0.7 \mid E) + \\ &\mathrm{cr}_H(A \mid \mathrm{Ch}(A) = 0.4 \,\&\, E) \cdot \mathrm{cr}_H(\mathrm{Ch}(A) = 0.4 \mid E)\end{aligned} \tag{5.5}$$

By Equations (5.3) and (5.4), this is

$$\mathrm{cr}_H(A \mid E) = 0.7 \cdot \mathrm{cr}_H(\mathrm{Ch}(A) = 0.7 \mid E) + 0.4 \cdot \mathrm{cr}_H(\mathrm{Ch}(A) = 0.4 \mid E) \tag{5.6}$$

As Equation (5.2) suggested, $\mathrm{cr}_H(\cdot \mid E)$ is just $\mathrm{cr}(\cdot)$. So this last equation becomes

$$\mathrm{cr}(A) = 0.7 \cdot \mathrm{cr}(\mathrm{Ch}(A) = 0.7) + 0.4 \cdot \mathrm{cr}(\mathrm{Ch}(A) = 0.4) \tag{5.7}$$

Finally, we fill in the values stipulated in the problem to conclude

$$\mathrm{cr}(A) = 0.7 \cdot 2/3 + 0.4 \cdot 1/3 = 0.6 \tag{5.8}$$

This is a lot of calculation, but the overall lesson comes to this: When your total evidence is admissible and restricts you to a finite set of chance values for A, the Principal Principle sets your credence in A equal to a weighted average of those chance values (where each chance value is weighted by your credence that it's the true chance).

This is an extremely useful conclusion, *provided* we can tell when our evidence is admissible. Lewis writes that, "Admissible propositions are the sort of information whose impact on credence about outcomes comes entirely by way of credence about the chances of those outcomes." (1980, p. 272) He then sketches out some categories of information we should expect to be admissible, and inadmissible. For example, evidence about events causally upstream from the chances will be admissible, because those events will form the first link in a causal chain like Figure 5.2. This includes information about the physical laws that give rise to chances; such information affects our credences about experimental outcomes by affecting our views about their chances. On the other hand, evidence that is caused by the outcome of the chance event is inadmissible, as we saw in the example of Amir's report. Generally, then, it's a good rule of thumb that facts concerning events temporally *before* the chance outcome are admissible, and inadmissible information is always about events *after* the outcome. (Though Lewis does remark at one point (1980, p. 274) that if backward causation is possible, seers of the future or time-travelers might give us inadmissible information about chance events yet to come.)

We'll end our discussion of the Principal Principle with a couple of caveats. First, I have been talking about coin flips, die rolls, etc. as if their outcomes have non-extreme objective chances. If you think that these outcomes are fully determined by the physical state of the world prior to such events, you might think these examples aren't really chancy at all—or if there are chances associated with their outcomes, the world's determinism makes those chances either 1 or 0. There are authors who think non-extreme chance assignments are compatible with an event's being fully deterministic. This will be especially plausible if you think a single phenomenon may admit of causal explanations at multiple levels of description. (Though the behavior of a gas is fully determined by the positions and velocities of its constituent particles, we might still use a thermodynamical theory that treats the gas's behavior as chancy.) In any case, if the compatibility of determinism and non-extreme chance concerns you, you can always replace my coin-flipping and die-rolling examples with genuinely indeterministic quantum events.

Second, you might think frequency data can affect rational credences without operating through opinions about chances. Suppose a new patient walks into a doctor's office, and the doctor assigns a credence that the patient has a particular disease equal to that disease's frequency in the general population. In order for this to make sense, need the doctor assume the patient was brought to her through some genuinely chancy physical process?

(That is, need the frequency affect her credences by way of informing some opinions about chances?) This will probably depend on how broadly we are willing to interpret macroscopic events as having objective chances. But unless chances are literally everywhere, inferences governed by the Principal Principle form a proper subset of legitimate instances of inductive reasoning. To move from frequencies in an observed population to predictions about the unobserved when chances are not present, we may need something like the frequency-credence principle (perhaps made more plausible by incorporating statistical tools) with which this section began. Or we may need a theory of inductive confirmation in general—something we will try to construct in Chapter 6.

For the time being, the lesson of the Principal Principle is clear: Where there are objective chances in the world, we should align our credences with them to the extent we can determine what they are. While there are exceptions to this rule, they depend on the causal relation between our information and the chances of which we're aware.

5.2.2 Expert principles and Reflection

The Principal Principle is sometimes described as a **deference principle**: to the extent you can determine what the objective chances are, the principle directs you to defer to them by making your credences match. In a certain sense, you treat the chances as authorities on what your credences should be. Might other sorts of authorities demand such rational deference?

Testimonial evidence plays a large role in how we learn about the world. Suppose an expert on some subject reveals her credences to you. Instead of coming on television and talking about the "probability" of snow, the weather forecaster simply tells you she's 30% confident that it will snow tomorrow. It seems intuitive that—absent other evidence about tomorrow's weather—you should set your credence in snow to 0.30 as well.

We can generalize this intuition with a principle for deference to experts modeled on the Principal Principle:

$$\mathrm{cr}_H(A \,|\, \mathrm{cr}_E(A) = x) = x \tag{5.9}$$

Here cr_H is a rational agent's hypothetical prior, A is a proposition within some particular subject matter, and $\mathrm{cr}_E(A) = x$ is the proposition that an expert on that subject matter assigns credence x to A. Equation (5.9) has consequences similar to the Principal Principle's: When a rational agent is *certain* that an expert assigns credence x to A, conditionalizing a hypothetical prior satisfying Equation (5.9) on that certainty will leave her with unconditional credence $\text{cr}(A) = x$. On the other hand, an agent who is uncertain of the expert's opinion will use Equation (5.9) to calculate a weighted average of all the values she thinks the expert might assign.

This principle tells us how to defer to someone we've identified as an expert. But it doesn't say anything about how to make that identification! Ned Hall helpfully distinguishes two kinds of experts we might look for:

> Let us call the first kind of expert a *database-expert*: she earns her epistemic status simply because she possesses more information. Let us call the second kind an *analyst-expert*: she earns her epistemic status because she is particularly good at evaluating the relevance of one proposition to another. (Hall 2004, p. 100)

A database expert possesses strictly more evidence than me. While she may not reveal the contents of that evidence, I can still take advantage of it by assigning the credences she assigns on its basis. On the other hand, I defer to an analyst expert not because of her superior evidence but because she is particularly skilled at forming opinions from the evidence we share. Clearly these categories can overlap; relative to me, a weather forecaster is probably both an analyst expert and a database expert with respect to the weather.

One particular database expert has garnered a great deal of attention in the deference literature: an agent's future self. Because Conditionalization retains certainties (Section 4.1.1), at any given time a conditionalizing agent will possess all the evidence possessed by each of her past selves—and typically quite a bit more. So an agent who is certain she will update by conditionalizing should treat her future self as a database expert. On the supposition that her future self will assign credence x to a proposition A, she should now assign credence x to A as well. This is van Fraassen's (1984)

Reflection Principle: For any proposition A in $\mathcal{L}$, real number x, and times t_i and t_j with $j \geqslant i$, rationality requires

$$\text{cr}_i(A \,|\, \text{cr}_j(A) = x) = x$$

Although the Reflection Principle mentions the agent's credences at both t_i and t_j, note that strictly speaking it is a synchronic principle, relating various credences the agent assigns at a given time. If we apply the Ratio Formula to the principle's equation and then cross-mutiply, we obtain:

$$\text{cr}_i[A \,\&\, \text{cr}_j(A) = x] = x \cdot \text{cr}_i[\text{cr}_j(A) = x] \tag{5.10}$$

The two credences related by this equation are both assigned *at* t_i; they just happen to be credences *in* some propositions about t_j.

Yet despite this synchronic nature, Reflection bears an intimate connection to Conditionalization. If an agent is certain she will update by conditionalizing between t_i and t_j—and meets a few other side conditions—Reflection follows. For instance, the Reflection Principle can be proven from the following set of conditions:

1. The agent is certain at t_i that cr_j will result from conditionalizing cr_i on the total evidence she learns between t_i and t_j (call it E).

2. The agent is certain at t_i that E (whatever it may be) is true.

3. $\mathrm{cr}_i(\mathrm{cr}_j(A) = x) > 0$

4. At t_i the agent can identify a set of propositions S in $\mathcal{L}$ such that:

   (a) The elements of S form a partition relative to the agent's certainties at t_i.

   (b) At t_i the agent is certain that E is one of the propositions in S.

   (c) For each element in S, the agent is certain at t_i what degree of belief she assigns to A conditional on that element.

References to the proofs described in this section can be found in the Further Readings. Here I'll simply provide an example that illustrates the connection between Conditionalization and Reflection. Suppose that I've rolled a die you're certain is fair, but as of t_1 have told you nothing about the outcome. However, at t_1 you're certain that between t_1 and t_2 I'll reveal to you whether the die came up odd or even. The Reflection Principle suggests you should assign

$$\mathrm{cr}_1(6 \,|\, \mathrm{cr}_2(6) = 1/3) = 1/3 \tag{5.11}$$

Assuming the enumerated conditions hold in this example, we can reason to Equation (5.11) as follows: In this case the partition S contains the proposition that the die came up odd and the proposition that it came up even. You are certain at t_1 that one of these propositions will provide the E you learn before t_2. You're also certain that your $\mathrm{cr}_2(6)$ value will result from conditionalizing your t_1 credences on E. So you're certain at t_1 that

$$\mathrm{cr}_2(6) = \mathrm{cr}_1(6 \,|\, E) \tag{5.12}$$

Equation (5.11) involves your t_1 credence in 6 conditional on the supposition that $\mathrm{cr}_2(6) = 1/3$. To determine this value, let's see what conditional reasoning *you* could do at t_1 on the supposition that $\mathrm{cr}_2(6) = 1/3$. We just said that at t_1 you're certain of Equation (5.12), so you could conclude that $\mathrm{cr}_1(6 \,|\, E) = 1/3$. Then you could examine your current t_1 credences conditional on both odd and even, and find that $\mathrm{cr}_1(6 \,|\, E)$ will equal 1/3 only if E is the proposition that the die came up even. (Conditional on the die's coming up odd, your credence in a 6 would be 0.) Thus you could conclude that E is the proposition that the die came up even. You're also certain at t_1 that E (whatever its content) is true, so concluding that E says the die came up even would allow you to conclude that the die did indeed come up even. And on the condition that the die came up even, your t_1 credence in a 6 is 1/3.

All of the reasoning in the previous paragraph was conditional, starting with the supposition that $\mathrm{cr}_2(6) = 1/3$. We found that conditional on this supposition, your rational credence in 6 would be 1/3. And that's exactly what the Reflection Principle gave us in Equation (5.11).[13] Information about your future credences tells you something about what evidence you'll receive between now and then. And information about what evidence you'll receive in the future should be incorporated into your credences in the present.

But how often do we really get information about our future opinions? Approached the way I've just done, the Reflection Principle seems to have little usable content. But van Fraassen originally proposed Reflection in a very different spirit. He saw the principle stemming from basic commitments we undertake when we form opinions.

van Fraassen drew an analogy to making promises. Suppose I make a promise at a particular time, but at the same time admit to being unsure whether I will actually carry it out. van Fraassen writes that "To do so would mean that I am now less than fully committed (a) to giving due regard to the felicity conditions for this act, or (b) to standing by the commitments I shall overtly enter." (1984, p. 255) To fully stand behind a promise requires full confidence that you will carry it out. And what goes for current promises goes for future promises as well: if you know you'll make a promise later on, failing to be fully confident *now* that you'll enact the future promise is a betrayal of solidarity with your future promising self.

Now apply this lesson to the act of making judgments: assigning a different credence *now* to a proposition than you will in the future is a failure to stand by the commitments implicit in that future opinion. As van Fraassen put it in a later publication, "Integrity requires me to express my commitment to proceed in what I now classify as a rational manner, to stand behind the ways in which I shall revise my values and opinions." (1995, pp. 25–26) This is his motivation for endorsing the Reflection Principle.[14] For van Fraassen, Reflection brings out a substantive commitment inherent in judgment, which underlies various other rational requirements. For instance, since van Fraassen's argument for Reflection does not rely on Conditionalization, van Fraassen at one point (1999) uses Reflection to argue for Conditionalization![15]

Of course, one might not agree with van Fraassen that credence assignment necessarily involves such strong commitments. And even if Reflection can be supported as van Fraassen suggests, moving from that principle to Conditionalization is going to require serious further premises. As we've seen, Reflection itself is a synchronic principle, relating an agent's attitudes at one time to other attitudes she assigns at the same time. At best, it will support the conclusion that an agent with certain attitudes at a given time is required to *predict* that she will update by Conditionalization. To actually establish Conditionalization as a diachronic norm, we would need the further assumption that rationally requires agents to update in the manner they antecedently predict.

5.3 The Principle of Indifference

The previous section discussed various deference principles (the Principal Principle, expert principles, the Reflection Principle) that place additional rational constraints on credence beyond the probability axioms, Ratio Formula, and Conditionalization. Yet each of those deference principles works with a particular kind of evidence—evidence about the chances, about an expert's credences, or about future attitudes. When an agent lacks these sorts of evidence about a proposition she's considering, the deference principles will do little to constrain her credences. If an Objective Bayesian (in the normative sense) wants to narrow what's rationally permissible to a single hypothetical prior, he is going to need a stronger principle than these three.

The Principle of Indifference is often marketed to do the trick. This principle has forerunners dating back to a proposal of Laplace's (1814/1995) that came to be known as the "principle of insufficient reason". But it was first called "the Principle of Indifference" by John Maynard Keynes, who wrote,

> The Principle of Indifference asserts that if there is no *known* reason for predicating of our subject one rather than another

> of several alternatives, then relatively to such knowledge the assertions of each of these alternatives have an *equal* probability. (Keynes 1921, p. 42, emphasis in original)

Applied to degrees of belief, the **Principle of Indifference** holds that if an agent has no evidence favoring any particular proposition in a partition over any other, she should spread her credence equally over the members of the partition. If I tell you I have painted my house one of the seven colors of the rainbow but tell you nothing more about my selection, the Principle of Indifference requires 1/7 credence that my house is now violet.

The Principle of Indifference looks like it could settle all open questions about rational credence. An agent could assign specific credences to doxastic possibilities when portions of her evidence require it (say, when one of the deference principles applies); her remaining credence would be spread equally among the remaining possibilities by the Principle of Indifference. For example, suppose I tell you that I flipped a fair coin to decide on a house color—heads meant gray, while tails meant a color of the rainbow. You could follow the Principal Principle and assign credence 1/2 to my house's being gray, while the Principle of Indifference directed you to distribute the remaining 1/2 credence equally among each of the rainbow colors (so each would receive credence 1/14). This plan seems to dictate a unique rational credence for every proposition in every evidential situation, thereby specifying a unique hypothetical prior distribution.[16]

Unfortunately, the Principle of Indifference has a serious flaw, one that was recognized by Keynes himself.[17] Suppose I tell you just that I painted my house some color—I don't tell you what selection I chose from—and you wonder whether it was violet. You might partition the possibilities into the proposition that I painted the house violet and the proposition that I didn't. In that case, lacking further information the Principle of Indifference would require you to assign credence 1/2 that the house is violet. But if you use the seven colors of the rainbow as your partition, you will assign 1/7 credence that my house is now violet. And if you use the colors in a box of crayons. . . . The trouble is that faced with the same evidential situation and same proposition to be evaluated, the Principle of Indifference will recommend different credences depending on which partition you consider.

Might one partition be superior to all the others, perhaps on grounds of the naturalness with which it divides the space of possibilities? (The selection of colors in a crayon box is pretty arbitrary!) Well, consider this example: I just drove 80 miles to visit you. I tell you it took between 2 and 4 hours to make the trip, and ask how confident you are that it took less than 3.

3 hours seems to neatly divide the possibilities in half, so by the Principle of Indifference you assign credence 1/2. Then I tell you I maintained a constant speed throughout the drive, and that speed was between 20 and 40 miles per hour. You consider the proposition that I drove faster than 30mph, and since that neatly divides the possible speeds the Indifference Principle again recommends a credence of 1/2. But these two credence assignments conflict. I drove over 30mph just in case it took me less than two hours and forty minutes to make the trip. So are you 1/2 confident that it took me less than 3 hours, or that it took me less than 2 hours 40 minutes? If you assign any positive credence that my travel time fell between those durations, the two answers are inconsistent. But thinking about my trip in velocity terms is just as natural as thinking about how long it took.[18]

This example is different from the painting example, in that time and speed require us to consider continuous ranges of possibilities. Infinite possibility spaces introduce a number of complexities we will discuss in the next section, but hopefully the intuitive problem here is clear. Joseph Bertrand (1888/1972) produced a number of infinite-possibility paradoxes for principles like Indifference. His most famous puzzle (now usually called **Bertrand's Paradox**) asks how probable it is that a chord of a circle will be longer than the side of an inscribed equilateral triangle. Indifference reasoning yields conflicting answers depending on how one specifies the chord in question—by specifying its endpoints, by specifying its orientation and length, by specifying its midpoint, etc.

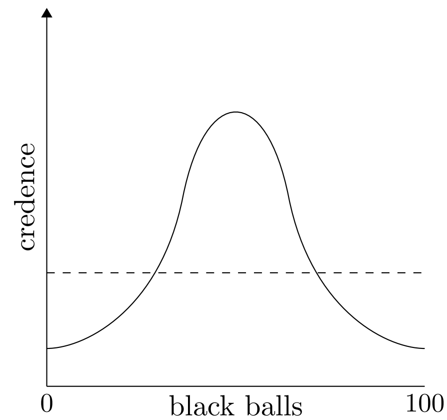
Since Keynes's discussion, a number of authors have modified his Indifference Principle. Chapter 6 will look in detail at Carnap's proposal. Another well-known suggestion is E.T. Jaynes' (1957a,b) **Maximum Entropy Principle**. Given any set of constraints on allowable credence distributions, the Maximum Entropy Principle selects the allowable distribution with the highest entropy. Over a finite partition of propositions $Q_1, Q_2, \ldots, Q_n$, the entropy of a distribution is calculated as

$$-\sum_{i=1}^{n} \mathrm{cr}(Q_i) \cdot \log \mathrm{cr}(Q_i) \tag{5.13}$$

The technical details of Jaynes' proposal are beyond the level of this book. The intuitive idea, though, is that by *maximizing* entropy in a distribution we *minimize* information.

To illustrate, suppose you know an urn contains 100 balls, each of which is either black or white. Initially, you assign an equal credence to each available hypothesis about how many black balls are in the urn. This "flat"

Figure 5.3: Possible urn distributions



distribution over the urn hypotheses is reflected by the dashed line in Figure 5.3. Then I tell you that the balls were created by a process that tends to produce as many white balls as black. This moves you to the more "peaked" distribution of Figure 5.3's solid curve. The peaked distribution reflects the fact that at the later time you had more information about the contents of the urn. There are various mathematical ways to measure the informational content of a distribution, and it turns out that a distribution's information content goes up as its entropy goes down.

Maximizing entropy is thus a strategy for selecting the lowest-information distribution consistent with what we already know. Jaynes' principle says that within the bounds imposed by your evidence, you should select the "flattest" credence distribution available. In a sense, this is a directive not to make any presumptions beyond what you know. As van Fraassen puts it, "one should not jump to unwarranted conclusions, or add capricious assumptions, when accommodating one's belief state to the deliverances of experience." (1981, p. 376) If *all* your evidence about my urn is that it contains 100 black or white balls, it would be strange for you to peak your credences around any particular number of black balls. What in your evidence would justify such a maneuver? The flat distribution seems the most rational option available.[19]

The Maximum Entropy approach has a number of its advantages. First, it extends easily from cases of finitely many possibilities to cases of infinitely many. (The summation in Equation (5.13) becomes an integral.) Second, for cases in which an agent's evidence simply delineates a space of doxastic possibilities (without favoring particular possibilities over particular others), the Principle of Maximum Entropy yields the same results as the Principle of Indifference. But Maximum Entropy also handles cases involving more complicated sorts of information. Besides restricting the set of possibilities, an agent's evidence might require her credence in one possibility to be twice that of another, or might require a particular conditional credence between two propositions. No matter the constraints, Maximum Entropy chooses the "flattest" (most entropic) distribution consistent with those constraints. Third, probability distributions selected by the Maximum Entropy Principle have been highly useful in various scientific applications, ranging from statistical mechanics to CT scans to natural language processing.

Yet the Maximum Entropy Principle also has flaws. It suffers from a version of the Indifference Principle's partitioning problem, assigning different credences to the same proposition depending on which partition is selected. Also, in some evidential situations satisfying the Maximum Entropy Principle both before and after an update requires agents to assign credences that violate Conditionalization.

5.4 Credences for Infinite Possibilities

Suppose I tell you a positive integer was just selected by some process, and tell you nothing more about that process. You need to distribute your credence across all the possible integers that might have been selected. Let's further suppose that you want to do so in such a way that each positive integer receives the same credence. In the last section we asked whether, given your scant evidence in this case about the selection process, you're required to assign each positive integer an equal credence. In this section I want to set aside the question of whether an equal distribution is *required*, and ask whether it's even *possible*.

We're going to have a small, technical problem here with the propositional language over which you credence distribution is assigned. In Chapter 2 we set up propositional languages with a *finite* number of atomic propositions, while a distribution over every positive integer requires infinitely many atomic propositions. Yet there are standard logical methods for dealing with languages containing infinitely many atomic propositions, and even for representing them using a finite number of symbols. For example, we could use "1" to represent the proposition that the number 1 was selected, "2" to represent 2's being selected, etc. This will allow us to represent infinitely many atomic propositions with only the standard 10 Arabic digits.

So the language isn't the real problem; the real problem is what credence value you could possibly assign to each and every one of those positive integers. To start seeing the problem, imagine you pick some positive real number r and assign it as your unconditional credence in each positive integer's being picked. For any positive real r you pick, there exists an integer n such that $r > 1/n$. So now consider the proposition that the positive integer selected was less than or equal to n. By repeated applications of finite additivity,

$$\mathrm{cr}(1 \vee 2 \vee \ldots \vee n) = \mathrm{cr}(1) + \mathrm{cr}(2) + \ldots + \mathrm{cr}(n) \tag{5.14}$$

Each of the credences on the righthand side equals r, so your credence in the disjunction is $r \cdot n$. But we selected n such that $r > 1/n$, so $r \cdot n > 1$. And now you've violated the probability rules.

This rules out assigning the same positive real value to each and every positive integer. What other options are there? Historically the most popular proposal has been to assign each positive integer a credence of 0. Yet this proposal creates its own problems.

The first problem with assigning each integer zero credence is that we must reconceive what an unconditional credence of 0 means. So far in this book we have equated assigning credence 0 to a proposition with ruling that proposition out as a live possibility. In this case, though, we've proposed assigning credence 0 to each positive integer while still treating each as a live possibility. So while we will still assign credence 0 to propositions that have been ruled out, there will now be other types of propositions that receive credence 0 as well. Similarly, we may assign credence 1 to propositions of which we are not certain.

Among other things, this reconception of credence 0 will undermine arguments for the Regularity Principle. As stated (Section 4.2), Regularity forbids assigning credence 0 to any logically contingent proposition. The argument there was that one should never entirely rule out a proposition that's logically possible, so one should never assign such a proposition 0 credence. Now we've opened up the possibility of assigning credence 0 to a proposition without having ruled it out. So while we can endorse the idea that no contingent proposition should be ruled out, Regularity no longer follows. Moreover, the current proposal provides infinitely-many explicit counterexamples to Regularity: we have proposed assigning credence 0 to the contingent proposition that the positive integer selected was 1, to the proposition that the integer was 2, that it was 3, etc.

Once we've decided to think about credence 0 in this new way, we encounter a second problem: the Ratio Formula. In Section 3.1.1 I framed the Ratio Formula as follows:

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

$$\mathrm{cr}(P \,|\, Q) = \frac{\mathrm{cr}(P \,\&\, Q)}{\mathrm{cr}(Q)}$$

This constraint relates an agent's conditional credence $\mathrm{cr}(P \,|\, Q)$ to her unconditional credences *only when* $\mathrm{cr}(Q) > 0$. As stated, it remains silent on how an agent's conditional and unconditional credences relate when $\mathrm{cr}(Q) = 0$.

Yet we surely want to have some rational constraints on that relation for cases in which an agent assigns credence 0 to a contingent proposition that she hasn't ruled out.[20] For example, in the positive integer case consider your conditional credence $\mathrm{cr}(2 \,|\, 2)$. Surely this conditional credence should equal 1. Yet because the current proposal sets $\mathrm{cr}(2) = 0$, the Ratio Formula cannot tell us anything about $\mathrm{cr}(2 \,|\, 2)$. And since we've derived all of our rational constraints on conditional credence from the Ratio Formula, the Bayesian system we've set up isn't going to deliver a requirement that $\mathrm{cr}(2 \,|\, 2) = 1$.[21]

There are various ways to respond to this problem. One interesting suggestion is to reverse the order in which we proceeded with conditional and unconditional probabilities: We began by laying down fairly substantive constraints (Kolmogorov's probability axioms) on *un*conditional credences, then tying conditional credences to those via the Ratio Formula. On the reverse approach, substantive constraints are first placed on conditional credences, then some further rule relates unconditional to conditional. The simplest such rule is that for any proposition P, $\mathrm{cr}(P) = \mathrm{cr}(P \,|\, \mathsf{T})$.

Some advocates of this technique describe it as making conditional credence "basic". The way we've approached conditional and unconditional credences, neither is more fundamental than the other in any sense significant to metaphysics or the philosophy of mind. Each is an independently existing type of doxastic attitude, and any rules we offer relating them are strictly *normative* constraints. The only sense in which the conditionals-first technique makes conditional credences prior to unconditional is in its order of normative explanation. Just as the Ratio Formula helped us transform constraints on unconditional credences into constraints on conditional credences, the rule that $\mathrm{cr}(P) = \mathrm{cr}(P \,|\, \mathsf{T})$ transforms constraints on conditionals into constraints on unconditionals.

Examples of the conditionals-first technique include (Popper 1955), (Renyi 1970), and (Roeper and Leblanc 1999). Like many of these approaches, Popper includes an axiom that directly addresses $\mathrm{cr}(Q \mid Q)$ for any Q that the agent deems possible, regardless of its unconditional credence value. This solves the problem of ensuring that $\mathrm{cr}(2 \mid 2) = 1$.

The final problem I want to address with assigning each positive integer 0 unconditional credence of being selected has to do with your unconditional credence that any integer was selected at all. The proposition that some integer was selected is a disjunction of the proposition that 1 was selected, the proposition that 2 was selected, the proposition that 3 was selected, etc. Finite Additivity directly governs unconditional credences in disjunctions of two (mutually exclusive) disjuncts; iterating that rule extends it to disjunctions with finitely many disjuncts. But this case concerns an *infinite* disjunction, and none of the constraints we've seen so far relates the unconditional credence of an infinite disjunction to the credences of its disjuncts.

It might seem natural to supplement our credal constraints with the following:

Countable Additivity: For any countable partition $Q_1, Q_2, Q_3, \ldots$ in $\mathcal{L}$,

$$\mathrm{cr}(Q_1 \vee Q_2 \vee Q_3 \vee \ldots) = \mathrm{cr}(Q_1) + \mathrm{cr}(Q_2) + \mathrm{cr}(Q_3) + \ldots$$

Notice that Countable Additivity does not apply to *every* partition of infinite size; it applies only to partitions of *countably many* members. The set of positive integers is countable, while the set of real numbers is not. (If you are unfamiliar with infinite sets of differing sizes, I would suggest reading the brief explanation referenced in this chapter's Further Readings.)

Countable Additivity naturally extends the idea behind Finite Additivity to sets of (countably) infinite size. Many authors have found it attractive. Yet in our example it rules out assigning credence 0 to each proposition stating that a particular positive integer was selected. Taken together, the proposition that 1 was selected, the proposition that 2 was selected, the proposition that 3 was selected, etc. form a countable partition (playing the role of Q_1, Q_2, Q_3, etc. in Countable Additivity). Countable Additivity therefore requires your credence in the disjunction of these propositions to equal the sum of your credences in the individual disjuncts. Yet the latter credences are each 0, while your credence in their disjunction (namely, the proposition that *some* positive integer was selected) must be 1.

So perhaps Countable Additivity wasn't such a good idea after all. The trouble is, without Countable Additivity we lose a very desirable property:

Conglomerability: For each proposition P and partition $Q_1, Q_2, Q_3, \ldots$ in $\mathcal{L}$, $\text{cr}(P)$ is no greater than the largest $\text{cr}(P \mid Q_i)$ and no less than the least $\text{cr}(P \mid Q_i)$.

In other words, if Conglomerability holds then finding the largest $\text{cr}(P \mid Q_i)$ and the smallest $\text{cr}(P \mid Q_i)$ create a set of bounds into which $\text{cr}(P)$ must fall.

In defining Conglomerability I didn't say how large the Q-partitions in question are allowed to be. We might think of breaking up the general Conglomerability principle into a number of sub-cases: Finite Conglomerability applies to finite partitions, Countable Conglomerability applies to countable partitions, Continuous Conglomerability applies to partitions of continuum-many elements, etc. Finite Conglomerability is guaranteed by the standard probability axioms. You'll prove this in Exercise 5.5, but the basic idea is that by the Law of Total Probability $\text{cr}(P)$ must be a weighted average of the various $\text{cr}(P \mid Q_i)$, so it can't be greater than the largest of them or less than the smallest. With the standard axioms in place, Countable Conglomerability then stands or falls with our decision about Countable Additivity; without Countable Additivity, Countable Conglomerability is false.[22]

We've already seen that the strategy of assigning equal, 0-credence to each positive integer's being selected violates Countable Additivity; let's see how it violates (Countable) Conglomerability as well.[23] Define "the 1-set" as the set of positive integers $\{1, 10, 100, 1000, \ldots\}$; define "the 2-set" as the set of positive integers $\{2, 20, 200, 2000, \ldots\}$; etc. Now take the proposition that the integer selected was a member of the 1-set, and the proposition that the integer selected was a member of the 2-set, and the proposition that the integer selected was a member of the 3-set, etc. Collect all these propositions all the n-sets where n is *not* a multiple of 10. The set of these propositions forms a partition. (If you think about it carefully, you'll see that any positive integer that might have been selected will belong to exactly one of these sets.)

The distribution strategy we're considering is going to want to assign

$$\begin{aligned}&\text{cr(the selected integer is not a multiple of } 10 \mid \\ &\qquad\qquad \text{the selected integer is a member of the 1-set)} = 0\end{aligned} \tag{5.15}$$

Why is that? Well, the only number in the 1-set that is not a multiple of 10 is the number 1. The 1-set contains infinitely many positive integers; on the supposition that one of those integers was selected you want to assign equal credence to each one's being selected; so you assign 0 credence to each one's being selected (including the number 1) conditional on that supposition. This gives us Equation (5.15). This argument generalizes; for any positive integer n that is not a multiple of 10, you'll have

$$\text{cr(the selected integer is not a multiple of } 10 \mid \text{the selected integer is a member of the } n\text{-set}) = 0 \tag{5.16}$$

Yet unconditionally it seems rational to have

$$\text{cr(the selected integer is not a multiple of 10)} = 9/10 \tag{5.17}$$

Conditional on any particular member of our partition, your credence that the selected integer isn't a multiple of 10 is 0. Yet unconditionally, you're highly confident that the integer selected is not a mutiple of ten. This is a flagrant violation of (Countable) Conglomerability—your credences conditional on the members of the (countable) partition are all the same, yet your unconditional credence has a very different value!

Why is violating Conglomerability a problem? Well, imagine I'm about to give you some evidence on which you're going to conditionalize. In particular, I'm about to tell you to which of the n-sets the selected integer belongs. Whichever piece of evidence you're about to get, your credence that the integer isn't a multiple of ten conditional on that evidence is 0. So you can be certain right now that immediately after receiving the evidence, your credence that the integer isn't a multiple of ten will be 0. Yet despite being certain that your better-informed future self will assign a particular proposition a credence of 0, you continue to assign that proposition a credence of 9/10 right now. This is a flagrant violation of the Reflection Principle, as well as generally good principles for evidence management. We usually adopt a particular opinion while awaiting evidence because we think that some potential pieces of evidence will pull us away from that opinion in one direction while other potential pieces will pull us away in the other. If we know that no matter what evidence comes in we're going to be pulled away from our current opinion in a particular direction, it seems irrationally stubborn to maintain our current opinion and not move in that direction right now. Conglomerability embodies these principles of good evidential hygiene; without Conglomerability our evidential interactions can begin to look absurd.

Where does this leave us? We wanted to find a way to assign an equal credence to each positive integer's being the one selected. We quickly concluded that equal credence could not be a positive real number. So we considered assigning credence 0 to each integer's being selected. Doing so violates Countable Additivity (a natural extension of our finite principles for calculating credences in disjunctions) and Conglomerability, which looks desirable for a number of reasons. Are there any *other* options?

I will briefly mention two further possibilities. The first possibility is to assign each positive integer an **infinitesimal** credence of having been selected. To work with infinitesimals, we extend the standard real-number system to include numbers that are greater than 0 but smaller than all the positive reals. If we assign each integer an infinitesimal credence of having been picked, we avoid the problems with assigning a positive real and also the problems of assigning 0. (For instance, if you pile enough infinitesimals together they can sum to 1.) Yet infinitesimals involve a great deal of advanced mathematics, seem implausible candidates for credence values agents might assign, and introduce troubles of their own (see Further Readings). So perhaps only one viable option remains: Perhaps if you learn a positive integer was just selected, it's *impossible* to assign equal credence to each of the possibilities consistent with what you know.

5.5 Jeffrey Conditionalization

Section 4.1.1 showed that conditionalizing on new evidence creates and retains certainties; evidence gained between two times becomes certain at the later time and remains so ever after. Contraposing, if an agent updates by Conditionalization and gains no certainties between two times, it must be because she gained no evidence between those times. In that section we also saw that if an agent gains no evidence between two times, Conditionalization keeps her credences fixed. Putting all this together, we see that under Conditionalization an agent's credences change just in case she gains new certainties. C.I. Lewis affirmed this point as follows:

> If anything is to be probable, then something must be certain. The data which themselves support a genuine probability, must themselves be certainties. We do have such absolute certainties, in the sense data initiating belief and in those passages of experience which later may confirm it. (1946, p. 186)

As we noted in Section 4.2, many contemporary epistemologists are uncomfortable with this kind of foundationalism (and even more so with appeals to sense data). Richard C. Jeffrey, however, had a slightly different concern, which he expressed with the following example and analysis:

> The agent inspects a piece of cloth by candlelight, and gets the impression that it is green, although he concedes that it might be blue or even (but very improbably) violet. If G, B, and V are the propositions that the cloth is green, blue, and violet, respectively, then the outcome of the observation might be that, whereas originally his degrees of belief in G, B, and V were .30, .30, and .40, his degrees of belief in those same propositions after the observation are .70, .25, and .05. If there were a proposition E in his preference ranking which described the precise quality of his visual experience in looking at the cloth, one would say that what the agent learned from the observation was that E is true. . . .
>
> But there need be no such proposition E in his preference ranking; nor need any such proposition be expressible in the English language. Thus, the description "The cloth looked green or possibly blue or conceivably violet," would be too vague to convey the precise quality of the experience. Certainly, it would be too vague to support such precise conditional probability ascriptions as those noted above. It seems that the best we can do is to describe, not the quality of the visual experience itself, but rather its effects on the observer, by saying, "After the observation, the agent's degrees of belief in G, B, and V were .70, .25, and .05." (1965, pp. 165–6)

Jeffrey's concern was that even if we granted the existence of a sense datum for each potential learning experience, the quality of that sense datum might not be representable in a proposition to which the agent could assign certainty, or at least might not be representable in a precise-enough proposition to differentiate that sense datum from other nearby data with different effects on the agent's credences.

At the time Jeffrey was writing the standard Bayesian updating norm (Conditionalization) relied on the availability of such propositions. Jeffrey therefore felt the need to provide a new updating rule, capable of handling examples like the cloth one above. He proposed what he called a **probability kinematics**, now universally known as

Jeffrey Conditionalization: Given any t_i and t_j with $i < j$, any A in $\mathcal{L}$, and a finite partition $B_1, B_2, \ldots, B_n$ in $\mathcal{L}$ whose elements each have nonzero cr_i,

$$\mathrm{cr}_j(A) = \mathrm{cr}_i(A \mid B_1) \cdot \mathrm{cr}_j(B_1) + \mathrm{cr}_i(A \mid B_2) \cdot \mathrm{cr}_j(B_2) + \ldots + \mathrm{cr}_i(A \mid B_n) \cdot \mathrm{cr}_j(B_n)$$

Let's apply Jeffrey Conditionalization to the cloth example. Suppose I'm fishing around in a stack of my family's clean laundry looking for any shirt of mine, but the lighting is dim because I don't want to turn on the overheads and awaken my wife. The color of a shirt in the stack would be a strong clue as to whether it was mine, as reflected by my conditional credences:

$$\begin{aligned} \mathrm{cr}_1(\mathrm{mine} \,|\, G) &= 0.80 \\ \mathrm{cr}_1(\mathrm{mine} \,|\, B) &= 0.50 \\ \mathrm{cr}_1(\mathrm{mine} \,|\, V) &= 0 \end{aligned} \tag{5.18}$$

For simplicity's sake we imagine green, blue, and violet are the only color shirts I imagine I might fish out of the stack. At t_1 I pull out a shirt. Between t_1 and t_2 I take a glimpse of the shirt. According to Jeffrey's story, my unconditional credence distributions across the $G/B/V$ partition are:

$$\begin{array}{lll} \mathrm{cr}_1(G) = 0.30 & \mathrm{cr}_1(B) = 0.30 & \mathrm{cr}_1(V) = 0.40 \\ \mathrm{cr}_2(G) = 0.70 & \mathrm{cr}_2(B) = 0.25 & \mathrm{cr}_2(V) = 0.05 \end{array} \tag{5.19}$$

Applying Jeffrey Conditionalization, I find my credence in the target proposition at the later time by combining my post-update unconditional credences across the partition with my pre-update credences in the target proposition conditional on elements of the partition. This yields:

$$\begin{aligned} &\mathrm{cr}_2(\mathrm{mine}) = \\ &\mathrm{cr}_1(\mathrm{mine} \,|\, G) \cdot \mathrm{cr}_2(G) + \mathrm{cr}_1(\mathrm{mine} \,|\, B) \cdot \mathrm{cr}_2(B) + \mathrm{cr}_1(\mathrm{mine} \,|\, V) \cdot \mathrm{cr}_2(V) = \\ &0.80 \cdot 0.70 + 0.50 \cdot 0.25 + 0 \cdot 0.05 = \\ &0.685 \end{aligned} \tag{5.20}$$

The low-light glimpse makes me fairly confident that the shirt I've selected is mine. (A quick calculation with the Law of Total Probability reveals that before the update I was 0.39 confident that the shirt was mine.)

Jeffrey Conditionalization allows us to represent evidential experiences that redistribute unconditional credence over a partition in virtually any way. (The only constraints are the probability axioms' demands that the resulting values be non-negative and sum to 1.) No proposition need go to certainty in this process. This means that unlike traditional Conditionalization, Jeffrey Conditionalization is perfectly consistent with the Regularity Principle (which forbids logically contingent propositions from receiving credence 0). An agent may Jeffrey Conditionalize as many times as she likes without any contingent proposition's credence going to 1 or 0. In fact, so long as the agent sends no propositions to certainty, Jeffrey Conditionalization is *reversible*: a proposition sent to high unconditional credence by a Jeffrey update may be sent back to low credence by a subsequent update. This contrasts with Conditionalization's irreversible setting of evidence to certainty.

These properties have made Jeffrey Conditionalization amenable to Regularity theorists, whose insistence on regular credence distributions forces them to reject Conditionalization. But Jeffrey Conditionalization can also allow learning by conditionalizing. In arguing for his new kinematics, Jeffrey was concerned only to establish that *some* learning experiences send no proposition to certainty; he didn't need to argue that *no* learning experiences do so. While Jeffrey's updating rule can be usefully applied even if no one ever conditionalizes (as the Regularity theorist supposes), it is also consistent with Conditionalization.

Here's how. Jeffrey Conditionalization's key feature is that for the selected partition $B_1, B_2, \ldots, B_n$, the following condition is maintained:

Rigidity: For any A in $\mathcal{L}$ and any B_m,

$$\mathrm{cr}_j(A \mid B_m) = \mathrm{cr}_i(A \mid B_m)$$

In the cloth example, my credence that the shirt I've selected is mine is a function of two kinds of credences: for each color, my unconditional credence that the selected shirt is that color; and for each color, my credence that the shirt is mine conditional on its being that color. My overall credence that the shirt is mine changes when I get a glimpse of its color, but *only because the first kind of credence changes*. I change my opinion about what color the shirt is, but I don't change my confidence that it's my shirt given that (say) it's green. I know what percentage of the green shirts in the house are mine; I just don't know if *this* is a green shirt. The rigidity condition points out which values remain fixed in this example, with A representing the proposition that the shirt is mine and the B_m representing the colors in the partition.

Jeffrey thought rigidity was appropriate for updates that "originated" in the B_m partition.[24] In the cloth example, my credal changes in non-color propositions from t_1 to t_2 are driven by changes in my color credences caused by my experience. Using the probability axioms and Ratio Formula, one can prove that Jeffrey Conditionalization holds over a partition just in case rigidity is maintained for that partition. (See Exercise 5.7.) So Jeffrey thought his updating rule applied whenever a change in credence originated in the partition $B_1, B_2, \ldots, B_n$. In that case, given the agent's full cr_i distribution and her cr_j credences in the B_m, Jeffrey Conditionalization will output a full cr_j distribution over $\mathcal{L}$.

Traditional Conditionalization (sometimes called "strict Conditionalization" to contrast with Jeffrey's rule) is consistent with Jeffrey Conditionalization because it also maintains a version of rigidity. One way for an update to originate in a partition $B_1, B_2, \ldots, B_n$ is for the agent to become certain that a particular disjunction of the B_m is true. When an agent conditionalizes on such a disjunction, it will turn out that for any A in $\mathcal{L}$ and any B_i with nonzero credence at t_j (so that $\mathrm{cr}_j(A \mid B_i)$ is defined under the Ratio Formula),

$$\mathrm{cr}_j(A \mid B_i) = \mathrm{cr}_i(A \mid B_i) \tag{5.21}$$

(This was proven in exercise 4.8.) When an agent learns evidence expressible in terms of a particular partition, conditionalizing on that evidence maintains rigidity in that partition. In these cases, strict Conditionalization is a special case of Jeffrey Conditionalization.

While Jeffrey Conditionalization is an exceedingly flexible tool, it does have some drawbacks. Unlike Conditionalization, Jeffrey Conditionalization is neither cumulative nor commutative. Jeffrey himself recognized these features, offering an example in which we first change an agent's unconditional credence in B to p (using $B/{\sim}B$ as our partition) and then change the unconditional credence in C to q (using the $C/{\sim}C$ partition). In some such cases there will be no proposition such that these two Jeffrey updates have the same cumulative result as a single Jeffrey update changing the credence of that proposition to some r. Further, we can consider the simple situation in which B and C are identical but $p \neq q$; that is, we perform two Jeffrey updates in succession that directly adjust the agent's unconditional credence in the *same* proposition. Clearly adjusting $\mathrm{cr}(B)$ to p and then adjusting $\mathrm{cr}(B)$ to some different q will leave the agent with a different final credence distribution than first adjusting $\mathrm{cr}(B)$ to q and then adjusting it to p. Jeffrey Conditionalization does not commute, which is problematic if you think that the effects of evidence on an agent should not depend on the order in which pieces of evidence arrive.

Finally, Jeffrey Conditionalization can be seen as a generalization of Conditionalization to broader sorts of evidential experiences. Conditionalization handles only evidence that sets unconditional credences to certainty. Jeffrey Conditionalization expands that reach by covering evidence that sets unconditional credences to nonextreme values. But what if an agent receives evidence that directly alters her *conditional* credences? How can we calculate the effects of such evidence on her other degrees of belief? van Fraassen (1981) describes a "Judy Benjamin Problem" in which direct alteration of conditional credences plausibly occurs, and which cannot be solved by Jeffrey Conditionalization.[25]

5.6 Exercises

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

Problem 5.1. At noon I rolled a 6-sided die. It came from either the Fair Factory (which produces exclusively fair dice), the Snake-Eyes Factory (which produces dice with a 1/2 chance of coming up 1 and equal chance of each other outcome), or the Boxcar Factory (which produces dice with a 1/4 chance of coming up 6 and equal chance of each other outcome).

(a) Suppose you use the Principle of Indifference to assign equal credence to each of the three factories from which the die might have come. Applying the Principal Principle, what is your credence that my die roll came up 3?

(b) Maria tells you that the die I rolled didn't come from the Boxcar Factory. If you update on this new evidence by Co nditionalization, how confident are you that the roll came up 3?

(c) Is Maria's evidence admissible with respect to the outcome of the die roll? Explain.

(d) After you've incorporated Maria's information into your credence distribution, Ron tells you the roll didn't come up 6. How confident are you in 3 after conditionalizing on Ron's information?

(e) Is Ron's evidence admissible with respect to the outcome of the die roll? Explain.

Problem 5.2. The expert deference principle in Equation (5.9) resembles the Principal Principle in many ways. Yet that expert deference principle makes no allowance for anything like inadmissible information. What kind of information should play the role for expert deference that inadmissible information plays for deference to chances? How should Equation (5.9) be modified to take such information into account?

Problem 5.3. Suppose it is currently t_1, and t_2 and t_3 are times in the future (with t_3 after t_2). At t_1, you satisfy the probability axioms, Ratio Formula, and Reflection Principle. You are also certain at t_1 that you will satisfy these constraints at t_2. However, for some proposition X your t_1 credences are equally divided between the following two (mutually exclusive and exhaustive) hypotheses about what your t_2 self will think of your t_3 credences:

Y: $(\mathrm{cr}_2[\mathrm{cr}_3(X) = 1/10] = 1/3)$ & $(\mathrm{cr}_2[\mathrm{cr}_3(X) = 2/5] = 2/3)$

Z: $(\mathrm{cr}_2[\mathrm{cr}_3(X) = 3/8] = 3/4)$ & $(\mathrm{cr}_2[\mathrm{cr}_3(X) = 7/8] = 1/4)$

Given all this information, what is $\mathrm{cr}_1(X)$? (Be sure to explain your reasoning clearly.)

Problem 5.4. Can you think of any kind of real-world situation in which it would be rationally permissible to violate the Reflection Principle? Explain the situation you're thinking of, and why it would make a Reflection violation okay.

Problem 5.5. Using Non-Negativity, Normality, Finite Additivity, the Ratio Formula, and any results we've proven from those four, prove Finite Conglomerability. (Hint: The Law of Total Probability may be useful here.)

Problem 5.6. Suppose that at t_1 you assign a "flat" credence distribution over language $\mathcal{L}$ whose only two atomic propositions are B and C—that is, you assign equal credence to each of the four state-descriptions of $\mathcal{L}$. Between t_1 and t_2 you perform a Jeffrey Conditionalization that originates in the $B/{\sim}B$ partition and sets $\mathrm{cr}_2(B) = 2/3$. Between t_2 and t_3 you perform a Jeffrey Conditionalization that originates in the $C/{\sim}C$ partition and sets $\mathrm{cr}_3(C) = 3/4$.

(a) Calculate your cr_2 and cr_3 distributions.

(b) Show that there is no proposition in $\mathcal{L}$ such that cr_3 could be generated directly from cr_1 by a Jeffrey Conditionalization that originated in the partition consisting of that proposition and its negation. (This demonstrates that Jeffrey Conditionalization is not cumulative, as suggested in the text.)

Problem 5.7. Prove that Jeffrey Conditionalization is equivalent to Rigidity. That is: Given any times t_i and t_j, proposition A in $\mathcal{L}$, and finite partition $B_1, B_2, \ldots, B_n$ in $\mathcal{L}$ whose elements each have nonzero cr_i, the following two conditions are equivalent:

1. $\mathrm{cr}_j(A) = \mathrm{cr}_i(A \,|\, B_1) \cdot \mathrm{cr}_j(B_1) + \mathrm{cr}_i(A \,|\, B_2) \cdot \mathrm{cr}_j(B_2) + \ldots + \mathrm{cr}_i(A \,|\, B_n) \cdot \mathrm{cr}_j(B_n)$
2. For all B_m in the partition, $\mathrm{cr}_j(A \,|\, B_m) = \mathrm{cr}_i(A \,|\, B_m)$.

(Hint: Complete two proofs—first condition 2 from condition 1, then *vice versa*.)

Problem 5.8. Suppose we apply Jeffrey Conditionalization over a finite partition $B_1, B_2, \ldots, B_n$ in $\mathcal{L}$ to generate cr_2 from cr_1. Show that we could have obtained the same cr_2 from cr_1 in the following way: start with cr_1; Jeffrey Conditionalize it in a particular way over a partition containing only two propositions; Jeffrey Conditionalize the result of *that* operation in a particular way over a partition containing only two propositions (possibly different from the ones used the first time); repeat this process a finite number of times until cr_2 is eventually obtained.*

5.7 Further reading

Subjective and Objective Bayesianism

Alan Hájek (2011b). Interpretations of Probability. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2011. URL: http://plato.stanford.edu/archives/win2011/entries/probabil
interpret/

Survey of the various historical interpretations of probability, with extensive references.

Bruno de Finetti (1931/1989). Probabilism: A Critical Essay on the Theory of Probability and the Value of Science. *Erkenntnis* 31. Translation of B. de Finetti, *Probabilismo*, Logos 14: 163–219., pp. 169–223

Classic paper critiquing objective interpretations of probability and advocating a Subjective Bayesian (in the semantic sense) approach.

*I owe this problem to Sarah Moss.

Donald Gillies (2000). Varieties of Propensity. *British Journal for the Philosophy of Science* 51, pp. 807–835

Reviews different versions of the propensity theory and their motivations. Focuses at the end on how propensity theories might respond to Humphreys' Paradox.

<u>Deference Principles</u>

David Lewis (1980). A Subjectivist's Guide to Objective Chance. In: *Studies in Inductive Logic and Probability.* Ed. by Richard C. Jeffrey. Vol. 2. Berkeley: University of California Press, pp. 263–294

Lewis's classic article laying out the Principal Principle and its consequences for theories of credence and chance.

Adam Elga (2007). Reflection and Disagreement. *Noûs* 41, pp. 478–502

Offers principles for deferring to many different kinds of agents, including experts, gurus (individuals with good judgment who lack some of your evidence), past and future selves, and peers (whose judgment is roughly as good as your own).

Bas C. van Fraassen (1984). Belief and the Will. *The Journal of Philosophy* 81, pp. 235–256

Article in which van Fraassen proposes and defends the Reflection Principle.

Jonathan Weisberg (2007). Conditionalization, Reflection, and Self-Knowledge. *Philosophical Studies* 135, pp. 179–197

Discusses conditions under which Reflection can be derived from Conditionalization, and *vice versa*.

<u>The Principle of Indifference</u>

E. T. Jaynes (1957a). Information Theory and Statistical Mechanics I. *Physical Review* 106, pp. 620–30

E. T. Jaynes (1957b). Information Theory and Statistical Mechanics II. *Physical Review* 108, pp. 171–90

E.T. Jaynes introduces the Maximum Entropy approach.

Teddy Seidenfeld (1986). Entropy and Uncertainty. *Philosophy of Science* 53, pp. 467–491

Explains the flaws with Jaynes's Maximum Entropy approach discussed at the end of Section 5.3, along with several others. Also contains useful references to Jaynes's many defenses of Maximum Entropy over the years and to the critical discussion that has ensued.

<u>Credences for Infinite Possibilities</u>

David Papineau (2012). *Philosophical Devices: Proofs, Probabilities, Possibilities, and Sets*. Oxford: Oxford University Press

Chapter 2 offers a highly accessible introduction to the cardinalities of various infinite sets. (Note that Papineau uses "denumerable" where we use the term "countable".)

Alan Hájek (2003). What Conditional Probability Could Not Be. *Synthese* 137, pp. 273–323

Assesses the viability of the Ratio Formula as a definition of conditional probability in light of various infinite phenomena and plausible violations of Regularity.

Timothy Williamson (2007). How Probable Is an Infinite Sequence of Heads? *Analysis* 67, pp. 173–80

Brief introduction to the use of infinitesimals in probability distributions, followed by an argument against using infinitesimals to deal with infinite cases.

Kenny Easwaran (2014). Regularity and Hyperreal Credences. *Philosophical Review* 123, pp. 1–41

Excellent, comprehensive discussion of the motivations for Regularity, the mathematics of infinitesimals, arguments against using infinitesimals to secure Regularity (including Williamson's argument), and an alternative approach.

<u>Jeffrey Conditionalization</u>

Richard C. Jeffrey (1965). *The Logic of Decision*. 1st. McGraw-Hill series in probability and statistics. New York: McGraw-Hill

Chapter 11 contains Jeffrey's classic presentation of his "probability kinematics", now universally known as "Jeffrey Conditionalization".

Notes

[1]The frequency theory is sometimes referred to as "frequentism" and its adherents as "frequentists". However "frequentism" more often refers to a school of statistical practice at odds with Bayesianism (which we'll discuss in Chapter XX). The ambiguity probably comes from the fact that most people in that statistical school also adopt the frequency theory as their interpretation of probability. But the positions are logically distinct and should be denoted by different terms. So I will use "frequency theory" here, and reserve "frequentism" for my later discussion of the statistical school.

[2]For many, many more see (Hájek 1996) and its sequel (Hájek 2009b).

[3]For one thing, the number of fair coin flips that will ever occur in the history of the universe is finite, and it might very well be an odd number!

[4]The frequency theory will also need to work with counterfactuals if nonextreme probabillities can be meaningfully ascribed to *a priori* truths, or to metaphysical necessities. (Might a chemist at some point have said, "It's highly probable that water is H_2O"?) Assigning nonextreme frequencies to such propositions' truth involves possible worlds far away from the actual.

[5]This difficulty for the propensity theory is often known as **Humphreys' Paradox**, since it was proposed in (Humphreys 1985).

One might respond by suggesting that propensities don't follow the standard mathematical rules of probability. And in fact, it's not obvious why they should. The frequency theory clearly yields probabilistic values: in any sequence of event repetitions a given outcome has a non-negative frequency, the tautologous outcome has a frequency of 1, and mutually exclusive outcomes have frequencies summing to the frequency of their disjunction. But establishing that propensity values (objective chances) satisfy the probability axioms takes *argumentation* from one's metaphysics of propensity. Nevertheless, most authors assume that propensities do satisfy the axioms; if they didn't, the propensity interpretation's probabilities wouldn't count as probabilities in the mathematician's sense (Section 2.2).

[6]One could focus here on a metaphysical distinction rather than a semantic one—instead of asking what probability talk *means*, I could ask what probabilities *are*. But some of the probability interpretations we will discuss don't have clear metaphysical commitments. The logical interpretation, for instance, takes probability to be a logical relation, but need not go on to specify an ontology for such relations. So I will stick with a semantic distinction, which in any case matches how these questions were discussed in much of twentieth-century analytic philosophy.

[7]In the twentieth century Subjective Bayesianism was also typically read as a form of expressivism; individuals' probability talk *expressed* their credal attitudes towards propositions without having truth-conditions. Nowadays there are other Subjective Bayesian

semantics that interpret probability talk in a more cognitivist mode, while still seeing it as reflecting agents' subjective degrees of belief.

[8]Carnap himself did not believe all probability talk picked out the logical values just described. Instead, he thought "probability" was ambiguous between two meanings, one of which was logical probability and the other of which had more of a frequency interpretation.

[9]There is disagreement about whether the logical and evidential interpretations of probability should be considered Objective Bayesian in the semantic sense. Popper (1957) says that objective interpretations make probability values objectively *testable*. Logical and evidential probabilities don't satisfy that criterion, and Popper seems to class them as subjective interpretations. Yet other authors (such as (Galavotti 2005)) distinguish between logical and subjective interpretations. I have defined the semantic Subjective/Objective Bayesian distinction such that logical and evidential interpretations count as Objective; while they may be normative for the attitudes of agents, logical and evidential probabilities do not vary with the attitudes particular agents possess.

[10]As I explained in Chapter 4, note 15, defining hypothetical priors as regular does not commit anyone to Regularity as a rational constraint.

[11]Since the ratio of B-outcomes to A-events must always fall between 0 and 1, this principle sheds some light on why credence values are usually scaled from 0 to 1. (Compare note 5 above.)

[12]Notice that the time t_i to which the chance in the Principal Principle is indexed need not be the time at which the agent in question assigns her credence concerning the experimental outcome A. In our coin example, the agent forms her credence at 1pm about the coin flip outcome at noon using information about the chances *at noon*. This is significant because on some metaphysical theories of chance, once the coin flip lands heads (or tails) the chance of H goes to 1 (or 0) forevermore. Yet even if the chance of H has become extreme by 1pm, the Principal Principle may still direct an agent to assign a nonextreme 1pm credence to H if all she knows are the chances from an earlier time.

I should also note that because chances are time-indexed, the notion of admissibility must be time-indexed as well. The information about the wad of chewing gum is admissible relative to 11:30am chances—learning about the chewing gum affects your credence about the flip outcome by way of your opinions about the 11:30am chances. But the information that chewing gum was stuck to the coin after 11 is *in*admissible relative to the 11:00am chances. (Chewing gum information affects your credence in H, but not by influencing your opinions about the chances associated with the coin at 11:00am.) So strictly speaking we should ask whether a piece of information is admissible *for* a particular proposition *relative* to the chances at a given time. I have suppressed this complication in the main text.

[13]The justification I've just provided for Equation (5.11) uses explicitly every one of the enumerated conditions except Condition 3. Condition 3 is necessary so that the conditional credence in Equation (5.11) is well-defined according to the Ratio Formula.

[14]One complication here is that van Fraassen sometimes describes Reflection as relating attitudes, but other times portrays it as being about various *acts* of commitment, and therefore more directly concerned with assertions and avowals than with particular mental states.

[15]Earlier we saw that under the Reflection Principle, opinions about your future credences may influence other credences you assign now. van Fraassen's argument for Conditionalization runs in the opposite direction, from credences you assign now to what you'll do in the future.

[16]Recall from Chapter 4 that conditionalizing keeps intact credence ratios among state-descriptions that are not eliminated by the evidence conditionalized upon. So if the Principle of Indifference requires an agent to assign state-descriptions equal credence in a particular evidential situation, that will be possible only if the agent's hypothetical prior assigns those state-descriptions equal credence as well. In general, once the various deference principles we've discussed have been satisfied by a rational hypothetical prior, that prior will have to distribute equal credence among all the possibilities that remain. (This will introduce complications if the hypothetical prior distributes credence over infinitely many possibilities; we'll come to those in the next section.)

[17]Joyce (2005)Joyce, James M. reports that this sort of problem was first identified by John Venn in the 1800s.

[18]This example is adapted from one in (Salmon 1966, pp. 66-7). A related example is van Fraassen's (1989) Cube Factory, which describes a factory making cubes of various sizes and asks how confident I should be that a given manufactured cube has a size falling within a particular range. The Principle of Indifference yields conflicting answers depending on whether cube size is described in terms of side length, face area, or volume.

[19]In Chapter XXX we will discuss a different credal response to this kind of ignorance.

[20]What about cases in which an agent *has* ruled out the proposition Q? Should rational agents assign credences conditional on conditions that they've ruled out? For discussion and references on this question, see (Titelbaum 2013, Ch. 5).

[21]I was careful to define the Ratio Formula so that it simply goes silent when $\mathrm{cr}(Q) = 0$, and is therefore in need of *supplementation* if we want to constrain values like $\mathrm{cr}(2 \,|\, 2)$. Other authors define the Ratio Formula so that it contains the same equation as ours but leaves off the restriction to $\mathrm{cr}(Q) > 0$ cases. This forces an impossible calculation when $\mathrm{cr}(Q) = 0$. Alternatively, one can leave the Ratio Formula unrestricted but make its equation $\mathrm{cr}(P \,|\, Q) \cdot \mathrm{cr}(Q) = \mathrm{cr}(P \,\&\, Q)$. This has the advantage of being *true* even when $\mathrm{cr}(Q) = 0$ (because $\mathrm{cr}(P \,\&\, Q)$ will presumably equal 0 as well), but does no better than our Ratio Formula on constraining the value of $\mathrm{cr}(2 \,|\, 2)$. (Any value we fill in for that conditional credence will make the relevant multiplication-equation true.)

[22]Seidenfeld/Schervish/Kadane [CITE] shows that this pattern generalizes: At each infinite cardinality, we cannot secure the relevant Conglomerability principle with Additivity principles at lower cardinalities; Conglomerability at a particular level requires Additivity at that same level.

[23]I owe the example that follows to Brian Weatherson.

[24]Actually, Jeffrey's original proposal was a bit more complicated than that. In (Jeffrey 1965) he began with a set of propositions $B_1, B_2, \ldots, B_n$ in which the credence change originated, but did not require the B_m to form a partition. Instead, he constructed a set of "atoms", which we can think of as state-descriptions constructed from the B_m. (Each atom was a consistent conjunction in which each B_m appeared exactly once, either affirmed or negated.) The rigidity condition (which Jeffrey sometimes called "invariance") and Jeffrey Conditionalization were then applied to these atoms rather than directly to the B_m in which the credence change originated.

Notice that in this construction the atoms form a partition. Further, Jeffrey recognized that if the B_m themselves formed a partition, the atoms wound up in a one-to-one correspondence with the B_m to which they were logically equivalent. I think it's for this reason that Jeffrey later (2004, Ch. 3) dropped the business with "atoms" and applied his probability kinematics directly to any finite partition.

[25]Interestingly, the main thrust of van Fraassen's article is that while Maximum Entropy *is* capable of providing a solution to the Judy Benjamin Problem, that solution is

intuitively unappealing.