Chapter 4

Updating by Conditionalization

Up to this point we have discussed *synchronic* credence constraints—rationally-required relations among the degrees of belief an agent assigns at a given time. This chapter introduces the fifth (and final) core normative Bayesian rule, Conditionalization. Conditionalization is a *diachronic* rule, requiring an agent's degrees of belief to line up in particular ways across times.

I begin by laying out the rule and some of its immediate consequences. We will then practice applying Conditionalization using Bayes' Theorem. Some of Conditionalization's consequences will prompt us to ask what notions of learning and evidence pair most naturally with the rule. I will also explain why it's important to attend to an agent's *total* evidence in evaluating her responses to learning.

Finally, we will see how Conditionalization helps Bayesians distinguish two influences on an agent's opinions: the content of her evidence, and her tendencies to respond to evidence in particular ways. This will lead to Chapter 5's discussion of how many distinct responses to the same evidence could be rationally permissible. Differing answers to that question provide a crucial distinction between Subjective and Objective Bayesianism.

4.1 Conditionalization

Suppose I tell you I just rolled a fair 6-sided die, and give you no further information about how the roll came out. Presumably you assign equal unconditional credence to each of the 6 possible outcomes, so your credence that the die came up 6 will be 1/6. I then ask you to suppose that the roll came up even (while being very clear that this is just a supposition—I'm still not revealing anything about the outcome). Applying the Ratio Formula to your unconditional distribution, we find that rationality requires your credence in 6 conditional on the supposition of even to be 1/3. Finally, I break down and tell you that the roll actually did come up even. Now how confident should you be that it came up 6?

I hope the obvious answer is 1/3. When you learn that the die actually came up even, the effect on your confidence in a 6 is identical to the effect of merely supposing evenness. This relationship between learning and supposing is captured in Bayesians' credence-updating rule:

Conditionalization: For any time t_i and later time t_j, if proposition E in $\mathcal{L}$ represents everything the agent learns between t_i and t_j and $\text{cr}_i(E) > 0$, then for any H in $\mathcal{L}$,

$$\text{cr}_j(H) = \text{cr}_i(H \,|\, E)$$

where cr_i and cr_j are the agent's credence distributions at the two times. Conditionalization captures the idea that the agent's credence in H at t_j—after *learning* E—should equal her earlier t_i credence in H had she merely been *supposing* E. If we label the two times in the die-roll case t_1 and t_2, Conditionalization tells us that

$$\text{cr}_2(6) = \text{cr}_1(6 \,|\, E) \tag{4.1}$$

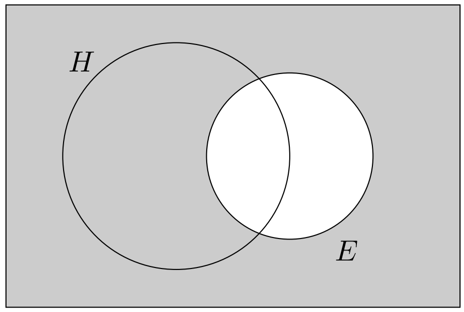
which equals 1/3 (given your unconditional distribution at t_1).

Warning: Some theorists take Conditionalization to *define* conditional credence. For them, to assign the conditional credence $\text{cr}_i(H \,|\, E) = r$ *just is* to be disposed to assign $\text{cr}_j(H) = r$ should you learn E. As I said in Chapter 3, I take conditional credence to be a genuine mental state, manifested by the agent in various ways at t_i (what she'll say in conversation, what sorts of bets she'll accept, etc.) beyond just her dispositions to update. For us, Conditionalization represents a *normative* constraint relating the agent's unconditional credences at a later time to her conditional credences earlier on.

Combining Conditionalization with the Ratio Formula gives us

$$\text{cr}_j(H) = \text{cr}_i(H \,|\, E) = \frac{\text{cr}_i(H \,\&\, E)}{\text{cr}_i(E)} \tag{4.2}$$

Figure 4.1: Updating H on E



(when $\text{cr}_i(E) > 0$). A Venn diagram shows why dividing these particular t_i credences should yield the agent's credence in H at t_j. In Chapter 3 we used a diagram like Figure 4.1 to understand conditional credences. There the white circle represented a set of possibilities to which the agent had temporarily narrowed her focus in order to entertain a supposition.

Now let's imagine the rectangle represents all the possible worlds the agent entertains at t_i (her doxastically possible worlds). The size of the H-circle represents the agent's unconditional t_i credence in H. Between t_i and t_j the agent learns that E is true. Among the worlds she had entertained before, the agent now excludes all the non-E worlds. Her set of doxastic possibilities narrows down to the E-circle; in effect, *the E-circle becomes the agent's new rectangle*. How unconditionally confident is the agent in H now? That depends what fraction of her new doxastic space is occupied by H-worlds. And this is what Equation (4.2) calculates: it tells you what fraction of the E-circle is occupied by H & E worlds.

As stated, the Conditionalization rule is useful for calculating a single unconditional credence value after an agent has gained evidence. But what if you want to generate the agent's entire t_j credence distribution at once? We saw in Chapter 2 that the agent's entire t_i credence distribution can be specified by a stochastic truth-table, which gives the agent's unconditional credence in each state-description of $\mathcal{L}$. To satisfy the probability axioms, the credence values in a stochastic truth-table must be non-negative and sum to 1. The agent's unconditional credence in any (non-contradictory) proposition can then be determined by summing her credences in the state-descriptions on which that proposition is true.

When an agent updates her credence distribution by applying Conditionalization to some learned proposition E, we say that she "conditionalizes on E". To calculate the stochastic truth-table values resulting from such an update, we apply a two-step process:

1. Give credence 0 to all state-descriptions inconsistent with the evidence learned.

2. Multiply each remaining nonzero credence by the *same* constant so that they all sum to 1.

As an example, let's consider what happens to your confidence that the fair die roll came up prime[1] when you learn that it came up even.

P	E	cr_1	cr_2
T	T	1/6	1/3
T	F	1/3	0
F	T	1/3	2/3
F	F	1/6	0

Here we've used a language $\mathcal{L}$ with atomic propositions P and E representing "prime" and "even". The cr_1 column represents your unconditional credences at time t_1, while the cr_2 column represents your t_2 credences. Between t_1 and t_2 you learn that the die came up even. That's inconsistent with the second and fourth state-descriptions, so in the first step of our update process their cr_2-values go to 0. The cr_1-values of the first and third state-descriptions (1/6 and 1/3 respectively) add up to only 1/2. So we multiply both of these values by 2 to obtain unconditional t_2-credences summing to 1.[2]

In this way, we generate your unconditional state-description credences at t_2 from your state-description credences at t_1. We can then calculate cr_2-values for other propositions. For instance, adding up the cr_2-values on the lines that make P true, we find that

$$\text{cr}_2(P) = 1/3$$

Given your initial distribution, your credence that the die came up prime after learning that it came up odd is required to be 1/3. Hopefully that squares with our intuitions about rational requirements in this case!

One final note: Our two-step process for updating stochastic truth-tables leads to a handy trick. Notice that in the second step of the process, every state-description that hasn't been set to zero is multiplied by the *same* constant. When two values are multipled by the same constant, the ratio between them remains intact. This means that if two state-descriptions have nonzero credence values after an update by Conditionalization, those values will stand in the same ratio as they did before the update. This fact will prove useful for problem-solving later on. (Notice that it applies only to *state-descriptions*; propositions that are not state-descriptions may not maintain their credence ratios after a conditionalization.)

4.1.1 Consequences of Conditionalization

If we adopt Conditionalization as our updating norm, what follows? When an agent updates by conditionalizing on E, her new credence distribution is just her earlier distribution conditional on E. In Section 3.1.2 we saw that if an agent's credence distribution obeys the probability axioms and Ratio Formula, then the distribution she assigns conditional on any particular proposition (in which she has nonzero credence) will be probabilistic as well. This yields the important result that if an agent starts off obeying the probability axioms and Ratio Formula and then updates by Conditionalization, her resulting credence distribution will satisfy the probability axioms as well.[3]

The process may then iterate. Having conditionalized her probabilistic cr_1 distribution on some evidence E to obtain probabilistic credence distribution cr_2, the agent may then gain further evidence E', which she conditionalizes upon to obtain cr_3 (and so on). Moreover, Conditionalization has the elegant mathematical property of being **cumulative**: Instead of obtaining cr_3 from cr_1 in two steps—first conditionalizing cr_1 on E to obtain cr_2, then conditionalizing cr_2 on E' to obtain cr_3—we can generate the same cr_3 distribution by conditionalizing cr_1 on $E \,\&\, E'$, a conjunction representing all the propositions learned between t_1 and t_3. (You'll prove this in Exercise 4.3.) Because Conditionalization is cumulative it is also **commutative**: Conditionalizing first on E and then E' has the same effect as conditionalizing in the opposite order.

An agent might learn many things between times t_i and t_j. Conditionalization requires E to represent *everything* the agent learned between those times, so we usually make E a conjunction of all the propositions learned. Suppose that A is one of the propositions the agent learns between t_i and t_j, and is therefore a conjunct of E. Clearly $E \vDash A$. Applying Conditionalization and our other core rules, we obtain

$$\mathrm{cr}_j(A) = \mathrm{cr}_i(A \,|\, E) = 1 \tag{4.3}$$

This follows from Equation (3.21), in which we showed that anything entailed by the condition in a conditional credence expression gets conditional credence 1. Equation (4.3) also applies if we replace A with E; conditionalizing on E always yields $\text{cr}_j(E) = 1$.

Thus Conditionalization creates certainties; any proposition learned between two times becomes certain at the later time. Conditionalization also maintains certainties. If an agent is certain of a proposition at t_i and updates by Conditionalization, she will remain certain of that proposition at t_j. That is, if $\text{cr}_i(H) = 1$ then Conditionalization yields $\text{cr}_j(H) = 1$ as well. On a stochastic truth-table, this means that once a state-description receives credence 0 at a particular time (the agent's certainties at that time rule that state of the world out for her), it will receive credence 0 at all subsequent times as well.

In Exercise 4.2 you'll prove that Conditionalization retains certainties from the probability axioms and Ratio Formula. But it's easy to see why this occurs on a Venn diagram. You're certain of H at t_i when H is true in every world you consider a live doxastic possibility. Conditionalizing on E strictly narrows the set of possible worlds you entertain. So if H was true in every world you entertained before conditionalizing, it'll be true in every world you entertain afterwards as well.

Combining these consequences of Conditionalization yields a somewhat counterintuitive result, to which we'll return in later discussions. Conditionalizing on E between two times makes that proposition (and any conjunct it contains) certain. Future updates by Conditionalization will then retain that certainty. So if an agent updates by conditionalizing throughout her life, any piece of evidence she learns at any point will remain certain for her ever after.

What if an agent doesn't learn *anything* between two times? Bayesians represent an empty evidence set as a tautology. So when an agent gains no information between t_i and t_j, Conditionalization yields

$$\text{cr}_j(H) = \text{cr}_i(H \mid \mathsf{T}) = \text{cr}_i(H) \tag{4.4}$$

for any H in $\mathcal{L}$. (The latter half of this equation comes from Equation (3.6), in which we showed that credences conditional on a tautology equal unconditional credences.) If an agent learns nothing between two times and updates by Conditionalization, her degrees of confidence will remain unchanged.

4.1.2 Probabilities are weird! The Base Rate Fallacy

Some authors make the mistake of thinking Bayes' Theorem follows from Conditionalization. As we saw in Section 3.1.3, this is incorrect: Bayes' Theorem can be derived exclusively from the probability calculus and Ratio Formula. Still, Conditionalization gives Bayes' Theorem added significance. Conditionalization tells us that your unconditional credence in a hypothesis H after updating on some evidence E should equal the posterior you assigned before updating—that is, $\mathrm{cr}_i(H \mid E)$. Bayes' Theorem is a tool for calculating this posterior from other credences you assign at t_i. As new evidence comes in over time and we repeatedly update by conditionalizing, Bayes' Theorem can be a handy tool for generating new credences from old.

For example, we could've used Bayes' Theorem to answer our earlier question of what happens to your credence in 6 when you learn that a fair die roll has come up even. The hypothesis is 6, and the evidence is E (for even). By Conditionalization and then Bayes' Theorem,

$$\mathrm{cr}_2(6) = \mathrm{cr}_1(6 \mid E) = \frac{\mathrm{cr}_1(E \mid 6) \cdot \mathrm{cr}_1(6)}{\mathrm{cr}_1(E)} \tag{4.5}$$

$\mathrm{cr}_1(6)$, your prior credence in 6, is 1/6, and $\mathrm{cr}_1(E)$, your prior credence in E, is 1/2. The likelihood of E, $\mathrm{cr}_1(E \mid 6)$, is easy—it's 1. So the numerator is 1/6, the denominator is 1/2, and the posterior $\mathrm{cr}_2(6) = 1/3$ as we saw before.[4]

Let's apply Bayes' Theorem to a more interesting case:

> 1 in 1,000 people have a particular disease. You have a test for the presence of the disease that is 90% accurate, in the following sense: If you apply the test to a subject who has the disease it will yield a positive result 90% of the time, and if you apply the test to a subject who lacks the disease it will yield a negative result 90% of the time.
>
> You randomly select a person and apply the test. The test yields a positive result. How confident should you be that this subject actually has the disease?

Most people—including trained medical professionals!—answer this question with a value around 80% or 90%. But if you set your credences by the statistics given in the problem, the rationally-required degree of confidence that the subject has the disease is less than 1%.

We'll use Bayes' Theorem to work that out. Let D represent the proposition that the subject has the disease and P the proposition that when applied to the subject, the test yields a positive result. Here D is our hypothesis, and P is the evidence acquired between t_1 and t_2. At t_1 (before applying the test) we take the subject to be representative of the population, giving us priors for the hypothesis and the catchall:

$$\mathrm{cr}_1(D) = 0.001 \qquad \mathrm{cr}_1(\sim D) = 0.999$$

The accuracy profile of the test gives us likelihoods for the hypothesis and catchall:

$$\mathrm{cr}_1(P \mid D) = 0.9 \qquad \mathrm{cr}_1(P \mid \sim D) = 0.1$$

In words, if the subject has the disease it's 90% probable the test will yield a positive result, while if he lacks the disease there's still a 10% chance we'll get a "false positive" indicating that he has it.

Now we'll apply a version of Bayes' Theorem from Section 3.1.3, in which the Law of Total Probability has been used to expand the denominator:

$$\begin{aligned} \mathrm{cr}_2(D) &= \frac{\mathrm{cr}_1(P \mid D) \cdot \mathrm{cr}_1(D)}{\mathrm{cr}_1(P \mid D) \cdot \mathrm{cr}_1(D) + \mathrm{cr}_1(P \mid \sim D) \cdot \mathrm{cr}_1(\sim D)} \\ &= \frac{0.9 \cdot 0.001}{0.9 \cdot 0.001 + 0.1 \cdot 0.999} \\ &\approx 0.009 = 0.9\% \end{aligned} \tag{4.6}$$

So there's the calculation. After learning of the positive test result, your credence that the subject has the disease should be a little bit less than 1%. But even having seen this calculation, most people find it hard to believe. Why shouldn't we be more confident that the subject has the disease? Wasn't the test 90% accurate?

Tversky and Kahneman (1974) suggested that in cases like this one, people's intuitive responses ignore the "base rate" of a phenomenon. The base rate in our example is the prior credence that the subject has the disease. In this case, that base rate is extremely low. But people tend to forget about that fact and be overwhelmed by accuracy statistics (such as likelihoods) concerning the test. This is known as the **Base Rate Fallacy**.

Why is the base rate so important? To illustrate, let's suppose you applied the test to 10,000 people. Using the base rate statistics, we would expect about 10 of those people to have the disease. Since the test gives a positive result for 90% of people who have the disease, we would expect these 10 diseased people to yield about 9 positive results—so-called "true positives". Then there would be about 9,990 people lacking the disease.

Since $\mathrm{cr}_i(P \mid {\sim}D)$—the false positive rate—is 10%, we'd expect to get about 999 false positive results. Out of 1,008 positive results the test would yield, only 9 of those subjects (or about 0.9%) would actually have the disease. This particular disease is so rare—its base rate is so tiny—that even with an accurate test we should expect the false positives to swamp the true positives. So when a single individual takes the test and gets a positive result, we should be much more confident that this is a false positive than a true one.

Another way to see what's going on is to consider the **Bayes factor** of the evidence you receive in this case. The idea of the Bayes factor is to compare your prior credence that the hypothesis is true with your prior credence that it's false, then make that comparison again after you've conditionalized on the evidence. Using Conditionalization and the Ratio Formula, we can derive

$$\frac{\mathrm{cr}_j(H)}{\mathrm{cr}_j({\sim}H)} = \frac{\mathrm{cr}_i(H \mid E)}{\mathrm{cr}_i({\sim}H \mid E)} = \frac{\mathrm{cr}_i(H)}{\mathrm{cr}_i({\sim}H)} \cdot \frac{\mathrm{cr}_i(E \mid H)}{\mathrm{cr}_i(E \mid {\sim}H)} \tag{4.7}$$

That last fraction on the right—the ratio of the likelihood of the hypothesis to the likelihood of the catchall—is the Bayes factor. It gives you the value by which your prior ratio of hypothesis to catchall is multiplied to get the posterior ratio. This is one way of measuring how much the evidence changes your opinions about the hypothesis.

In our disease example, the Bayes factor is

$$\frac{\mathrm{cr}_1(P \mid D)}{\mathrm{cr}_1(P \mid {\sim}D)} = \frac{0.9}{0.1} = 9 \tag{4.8}$$

At t_1 your priors in D and ${\sim}D$ have the ratio 1/999. The positive test has a substantial influence on this ratio; as the Bayes factor reveals, evidence of a positive test result multiplies the ratio by 9. Yet since the ratio was so small initially, multiplying it by 9 only brings the posterior ratio to 9/999. So even after seeing the test outcome, you should be much more confident that the subject doesn't have the disease than you are that he does.[5]

4.2 Evidence and Certainty

Combining Conditionalization with the probability axioms and Ratio Formula creates a Bayesian approach to evidence that many have found troubling. Conditionalization works with a proposition E representing everything the agent learns between two times. (If many propositions are learned, E is their conjunction.) We also speak of E as the evidence the agent gains between those two times. Yet Conditionalization gives E properties that epistemologists don't typically attribute to evidence.

We've already seen that a piece of evidence E (as well as all of its conjuncts, and anything else it entails) becomes certain once conditionalized upon. When an agent learns E, the set of doxastically possible worlds she entertains shrinks to a set of worlds that all make E true; on the Venn diagram, what once was merely an E-circle within her rectangle of worlds now becomes the entire rectangle. And as we saw in Section 4.1.1, this change is permanent: as long as the agent keeps updating by Conditionalization, any evidence she once learned remains certain and possible worlds inconsistent with it continue to be ignored.

What realistic conception of evidence—and of learning—meets these requirements? When I learn that my sister is coming over for Thanksgiving dinner, I become highly confident in that proposition. But do I become 100% certain? Do I *rule out* all possible worlds in which she doesn't show, refusing to consider them ever after? To do so seems not only odd but positively irrational, in violation of the

Regularity Principle: In a rational credence distribution, no logically contingent proposition receives unconditional credence 0.

The Regularity Principle captures the common-sense idea that one's evidence is never so strong as to *entirely* rule out any logical possibility. (Recall that a logically contingent proposition is neither a logical contradiction nor a logical tautology.[6]) As damning evidence against a contingent proposition mounts up, we can keep decreasing and decreasing our credence in it, but our unconditional credence distribution should always remain **regular**—it should assign each contingent proposition at least a tiny bit of confidence.

The Regularity Principle adds to the synchronic Bayesian rules we have seen so far—it is not entailed by the probability axioms, the Ratio Formula, or any combination of them. As our Contradiction result showed in Section 2.2.1, those rules do entail that all logical contradictions receive credence 0. But Regularity is the converse of Contradiction; instead of saying that *all* contradictions receive credence 0, it entails that *only* contradictions do. Similarly, Regularity (along with the probability axioms) entails the converse of Normality: instead of saying that *all* tautologies receive credence 1, it entails that *only* tautologies do. (The negation of a contingent proposition is contingent; if we were to assign a contingent proposition credence 1 its negation would receive credence 0, in violation of Regularity.) This captures the common-sense idea that one should never be absolutely certain of a proposition that's not logically true.[7]

Conditionalization conflicts with Regularity; the moment an agent conditionalizes on contingent evidence, she assigns credence 1 to a non-tautology. As we saw earlier, Conditionalization on contingent evidence rules out doxastic possibilities the agent had previously entertained; on the Venn diagram, it narrows the set of worlds under consideration. Regularity, on the other hand, fixes an agent's doxastic possibility set as the full set of logical possibilities. While evidence might shift the agent's credences around among various possible worlds, an agent who satisfies Regularity will never eliminate a possible world outright.

We might defend Conditionalization by claiming that whenever agents receive contingent evidence, it is of a highly specific kind, and Regularity is false for this kind of evidence. Perhaps I don't actually learn that my sister is coming over for Thanksgiving—I learn that she *told* me she's coming; or that it *seemed* to me that she said that; or that I had a phenomenal experience as of.... Surely I can be certain what phenomenal experiences I've had, or at least what experiences I'm having right now. When in the midst of having a particular phenomenal experience, can't I entirely rule out the logical possibility that I am having a different experience instead? Suffice it to say that the existence of such indubitable phenomenal evidence is a highly fraught topic, debated throughout the history of epistemology, that I will not attempt to adjudicate here.

There are other ways to make sense of Conditionalization's conception of evidence. Levi (1980) took credence-1 propositions to represent "standards of serious possibility":

> When witnessing the toss of a coin, [an agent] will normally envisage as possibly true the hypothesis that the coin will land heads up and that it will land tails up. He may also envisage other possibilities—e.g., its landing on its edge. However, if he takes for granted even the crudest folklore of modern physics, he will rule out as impossible the coin's moving upward to outer space in the direction of Alpha Centauri. He will also rule out the hypothesis that the Earth will explode. (p. 3)

Yet Levi was careful to formalize standards of serious possibility so that they could change—growing either stronger or weaker—for a given agent over time.

Alternatively, we could represent agents as ruling out contingent possibilities only relative to a particular inquiry. Consider a scientist who has just received a batch of experimental data and wants to weigh its import for a set of hypotheses. There are always outlandish possibilities to consider: the data might have been faked; the laws of physics might have changed a moment ago; she might be a brain in a vat. But to focus on the problem at hand, she might conditionalize on the data and see where that takes her credences in the hypotheses. Updating by Conditionalization might fail as a big-picture, permanent strategy, but nevertheless could be useful in carefully-delimited contexts. (I mentioned this possibility in Section 2.4.1.)

Perhaps these interpretations of evidence conditionalized-upon remain unsatisfying. We will return to this problem in Chapter 5, considering an alternative updating rule (Jeffrey Conditionalization) that allows agents to redistribute their credences over contingent possibilities without eliminating any of them entirely. For the rest of this chapter we will simply assume that Conditionalization on some kind of contingent evidence is a rational updating rule, so as to draw out further features of the rule.

4.2.1 Probabilities are weird! The Monty Hall Problem

Classical entailment is **monotonic** in the following sense: If a piece of evidence E you have received entails that H, any augmentation of that evidence (any conjunction that includes E as a conjunct) will continue to entail H as well. Probabilistic relations, however, can be **nonmonotonic**: H might be highly probable given E, but improbable given $E \,\&\, E'$. For this reason, it's important when recommending rational credences for an agent to consider *all* of the evidence she possesses, or has acquired during a particular learning experience. Carnap (1950) called this the **Principle of Total Evidence**.

We sometimes violate the Principle of Total Evidence by failing to note the *manner* in which an agent gained particular information.[8] If the agent is aware of the mechanism by which a piece of information was received, it can be important to recognize facts about that mechanism as a component of her total evidence (along with the information itself). In Eddington's (1939) classic example, you draw a sample of fish from a lake, and all the fish are longer than six inches. Normally, updating on this information would increase your confidence that every fish in the lake is at least that long. But if you know the net used to draw the sample has big holes through which shorter fish fall, the confidence increase is unwarranted. Here it's important to conditionalize not only on the lengths of the fish but also on how they were caught.

The process by which information is obtained is also crucial to a famously counterintuitive probability puzzle, the **Monty Hall Problem** (Selvin 1975):

> In one of the games played on *Let's Make a Deal*, a prize is randomly hidden behind one of three doors. The contestant selects one door, then the host (Monty Hall) opens one of the doors the contestant didn't pick. Monty always opens a door that doesn't have the prize behind it. (If both the unselected doors are empty, he randomly chooses which one to open.) After he opens an empty door, Monty asks the contestant if she wants what's behind the door she initially selected, or what's behind the other remaining closed door. Assuming she understands the details of Monty's procedure, how confident should the contestant be that the door she initially selected contains the prize?

Most people's initial reaction is to answer 1/2: the contestant originally spread her credence equally among the three doors; one of them has been revealed to be empty; so she should be equally confident that the prize is behind each of the remaining two. This analysis can be backed up by the following stochastic truth-table:

	cr_1	cr_2
Prize behind door A	1/3	1/2
Prize behind door B	1/3	0
Prize behind door C	1/3	1/2

Here we've used the obvious partition of three locations where the prize might be. Without loss of generality, I've imagined that the contestant initially selects door A and Monty then opens door B. At time t_1—after the contestant has selected door A but before Monty has opened anything—she is equally confident that the prize is hidden behind each of the three doors. When Monty opens door B at t_2, the contestant should conditionalize on the prize's not being behind that door. It looks like this yields the cr_2 distribution listed above, which matches most people's intuitions.

Yet the contestant's *total* evidence at t_2 includes not only the fact that the prize isn't behind door B, but also the fact that Monty opened that one. These two propositions aren't equivalent among the agent's doxastically possible worlds; there are possible worlds consistent with what the contestant knows about Monty's procedure in which door B is empty yet Monty opens door C. That door B was not only empty but was *revealed* to be so is not expressible in the partition used above. So we need a richer partition, containing information both about the location of the prize and about what Monty does:

	cr_1	cr_2
Prize behind door A & Monty reveals B	1/6	1/3
Prize behind door A & Monty reveals C	1/6	0
Prize behind door B & Monty reveals C	1/3	0
Prize behind door C & Monty reveals B	1/3	2/3

Given what the agent knows of Monty's procedure, these four propositions partition her doxastic possibilities at t_1. At that time she doesn't know where the prize is, but she has initially selected door A (and Monty hasn't opened anything yet). If the prize is indeed behind door A, Monty randomly chooses whether to open B or C. So the contestant divides her 1/3 credence that the prize is behind door A equally between those two options. If the prize is behind door B, Monty is forbidden to open that door as well as the door the contestant selected, so Monty must open C. Similarly, if the prize is behind door C, Monty must open B.

At t_2 Monty has opened door B, so the contestant sets her credence in the second and third partition elements to 0, then multiplies the remaining values by a constant so that they sum to 1. This maintains the ratio between her credences on the first and fourth lines; initially she was twice as confident of the fourth as the first, so she remains twice as confident after the update. She is now 2/3 confident that the prize isn't behind the door she initially selected, and 1/3 confident that her initial selection was correct. If she wants the prize, the contestant should switch doors.

This is the correct analysis. If you find that surprising, the following explanation may help: When the contestant originally selected her door, she was 1/3 confident that the prize was behind it and 2/3 confident that the prize was somewhere else. If her initial pick was correct, it makes sense to stick with that pick after Monty opens a door. But if her initial selection was wrong, she should switch to the other remaining closed door, because it must contain the prize. So there's a 1/3 chance that sticking is the best strategy, and a 2/3 chance that switching will earn her the prize. Clearly switching is a better idea.

When I first heard the Monty Hall Problem, even that explanation didn't convince me. I only became convinced after I simulated the scenario over and over and found that sticking made me miss the prize roughly 2 out of 3 times. If you're not convinced, try writing a quick computer program or finding a friend with a free afternoon to act as Monty Hall for you a few hundred times. You'll eventually find that the stochastic truth-table taking *total* evidence into account provides the correct analysis.

4.3 Hypothetical Priors and Evidential Standards

Suppose we are talking to an agent who has always been perfectly rational with respect to the Bayesian rules—she has always assigned credence distributions satisfying the probability axioms and Ratio Formula, and she has always updated those distributions by Conditionalization. At this advanced stage of her life, the credences she assigns will have been affected a great deal by the empirical evidence she's gained over time. But will there have been any other influences on her credences? Might another rational agent, receiving the same course of evidence over her lifespan, have assigned different credences in response? What determines how an agent responds to her evidence?

We will return to many of these questions in our discussion of Subjective vs. Objective Bayesianism in Chapter 5. At this point I want to develop a mathematical tool that helps Bayesians distinguish the credal influence of an agent's evidence from that of the **evidential standards** by which she responds to her evidence. We begin with the following theorem:

Hypothetical Priors Theorem: Given any finite series of credence distributions $\text{cr}_1, \text{cr}_2, \ldots, \text{cr}_n$ satisfying the probability axioms and Ratio Formula, let E_i be a conjunction of the agent's total evidence at t_i. If the cr update by Conditionalization, then there exists a regular probability distribution cr_H such that for all $1 \leqslant i \leqslant n$,

$$\text{cr}_i(\cdot) = \text{cr}_H(\cdot \,|\, E_i)$$

I will refer to the distribution cr_H whose existence is guaranteed by this theorem as a **hypothetical prior distribution**. (Other authors call it an "ur-prior".)

I want to first illustrate what the Hypothetical Priors Theorem says mathematically, then explain how Bayesians interpret its significance. Suppose that at t_1 I show you a 6-sided die and tell you it has just been rolled. Moreover, I tell you that I took this die from the craps table at a well-known Las Vegas Casino. Five minutes later, at t_2, I tell you that the die roll came up even. Finally, at t_3, I tell you that it also came up prime. Let's say that in response to this evidence, you assign the credences expressed by this stochastic truth-table at those three times:

C	P	E	cr_1	cr_2	cr_3
T	T	T	1/6	1/3	1
T	T	F	1/3	0	0
T	F	T	1/3	2/3	0
T	F	F	1/6	0	0
F	T	T	0	0	0
F	T	F	0	0	0
F	F	T	0	0	0
F	F	F	0	0	0

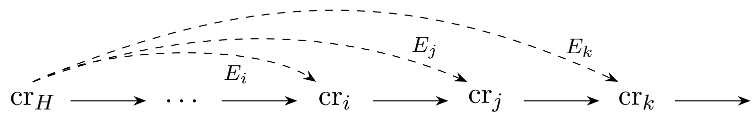
Here P stands for the roll's coming up prime, E stands for even, and C represents the die's origin in a reputed casino. At t_1 your total evidence (relevant to the roll outcome) is C; we'll call this E_1. At t_2 your total evidence E_2 is $C \,\&\, E$. E_3 is $C \,\&\, E \,\&\, P$. Since C is part of your evidence at all times reflected in this table, you assign 0 credence throughout the table to any state-description on which C is false.

Since your credence distributions c_1 through c_3 are probabilistic, and update in the manner dictated by Conditionalization, the Hypothetical Priors Theorem guarantees the existence of at least one hypothetical prior cr_H standing in a special relation to those distributions. I've added a column to the stochastic truth-table below representing one such cr_H-distribution:

C	P	E	cr_H	cr_1	cr_2	cr_3
T	T	T	1/12	1/6	1/3	1
T	T	F	1/6	1/3	0	0
T	F	T	1/6	1/3	2/3	0
T	F	F	1/12	1/6	0	0
F	T	T	1/12	0	0	0
F	T	F	1/6	0	0	0
F	F	T	1/6	0	0	0
F	F	F	1/12	0	0	0

As the Hypothetical Priors Theorem requires, cr_H is regular—it doesn't assign credence 0 to any contingent propositions—and it satisfies the probability axioms. It also stands in the relation to each of cr_1, cr_2, and cr_3 that each of those distributions can be obtained from cr_H by conditionalizing it on your total evidence at the relevant time. To take one example, consider cr_2. E_2 is $C \& E$. To conditionalize cr_H on $C \& E$, we write a zero on each line whose state-description is inconsistent with $C \,\&\, E$. That puts zeroes on the second and fourth through eighth lines of the truth-table. We then multiply the cr_H values on the first and third lines of the table by a constant (in this

Figure 4.2: An initial credence distribution?



case, 4) so that the results sum to 1. This yields the cr_2 distribution above. With a bit of work you can verify that cr_1 results from conditionalizing cr_H on E_1, and cr_3 is the result on conditionalizing cr_H on E_3.

So that's how the math works. But what does the hypothetical prior cr_H represent? Many Bayesians interpret the hypothetical prior as an **initial credence distribution**. The idea is that an agent begins her rational life possessing no contingent evidence; she therefore has no contingent certainties and adopts a regular credence function. This is the moment in her doxastic life represented by her initial credence distribution cr_H.[9] She then learns pieces of contingent information over the course of her life, to which she responds by conditionalizing. If we want to know what credence distribution she will assign at some later time t_k, we could start with cr_H, conditionalize on the first evidence piece of evidence she gains, then conditionalize on the second, and so on until we get to cr_i, conditionalize it to get cr_j, and finally conditionalize our way to cr_k. But since Conditionalization is cumulative (Section 4.1.1), there's a much more direct way to proceed. We can generate cr_k by simply conditionalizing cr_H on E_k, the agent's total evidence at t_k. This explains why cr_H (construed as an initial credence function) stands in the relation to the various cr_i, cr_j, cr_k, etc. described in the Hypothetical Priors Theorem.[10]

Figure 4.2 depicts the initial credence function interpretation of cr_H. Each distribution in the series is generated from the previous one by conditionalizing (solid arrows), but we can also derive each distribution directly (dashed arrows) by conditionalizing cr_H on the agent's total evidence at the relevant time.

Yet ultimately this interpretation must be a myth. Given the Hypothetical Priors Theorem, any agent who obeys the core Bayesian rules can be furnished with a hypothetical prior. But must any such agent have had some point in her life at which she lacked *all* contingent information? And even if there was such a point in our intellectual prehistory, is it plausible that at such a time we assigned degrees of belief satisfying the probability calculus? David Lewis used to refer to such highly intelligent, blank creatures in conversation as "superbabies"; sadly, I doubt the world has ever seen their like.[11]

I prefer an alternate interpretation of hypothetical priors. To understand that interpretation, start with the possibility of having two agents, both of whom satisfy all the core Bayesian rules, both of whom have the same total evidence at every time in their conditionalizing sequence, and yet who assign different credences to a wide variety of propositions. The core Bayesian rules do nothing to rule this possibility out—they do not force an agent's credences to supervene on her evidence. To give a simple example, suppose you have a friend who is suspicious of Vegas casinos and think they all weight their dice to produce extra snake-eyes (double 1s). At t_1, when you and he have both been told proposition C about the casino origin of the die that was rolled, you assign 1/2 credence that it came up even. But he will assign a lower credence to E, because he will assume the die is weighted towards the number 1. So despite having the same total evidence as you at t_1, he will assign a different cr_1 distribution than the ones described in our tables above. Yet it's perfectly possible that he will still satisfy the probability axioms and Ratio Formula, and as he gains the same evidence as you at t_2 and t_3 he will still be able to update by Conditionalization.

When we go to construct a hypothetical prior for your friend, we will find that it contains different values than yours. Your hypothetical prior assigns

$$\text{cr}_H(E \mid C) = 1/2 \tag{4.9}$$

This is why, when you conditionalize your cr_H on C (your total evidence at t_1) to obtain cr_1, you wind up with a $\text{cr}_1(E)$ value of 1/2. Yet since your friend has a different value for $\text{cr}_1(E)$, while having the same total evidence as you at t_1, he must have a different $\text{cr}_H(E \mid C)$ value than you. Moreover, your differing cr_H values cannot be the result of your responding to different evidence, since by stipulation cr_H is regular; it's a distribution without contingent certainties, and therefore a distribution that does not contain any evidence to which it might respond.

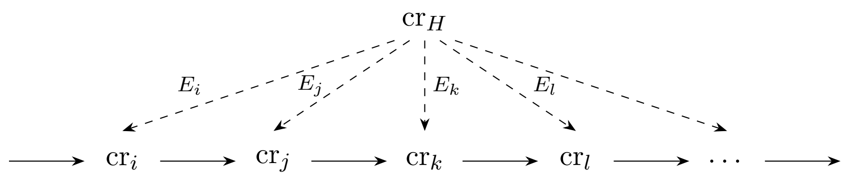
In the example at hand this probably seems preposterous. Your friend probably responds differently than you to news that the die came from Vegas because he has some evidence you don't about the way Vegas casinos run their games. (Why, after all, does he focus on *snake-eyes*?) Perhaps he lacks any extra evidence that can be expressed in our toy $C/P/E$ language, but his *total* evidence is surely different from yours, and that must be what drives the difference in your credences.

But we can extend the example to a much more general case. Imagine two agents who have literally identical total evidence over the course of their entire lives—perhaps two twins raised together at all times. Perhaps one of them has an inherently more trusting character than the other, and so (despite having been exposed to all the same stories of Vegas malfeasance) is much more willing to treat casino dice as fair than her sister. There is nothing in our core Bayesian norms that rules this out. And even if it's impossible to have two distinct agents who literally share all the same evidence, the example is merely meant to be illustrative. We could always focus on a particular agent and the total evidence she (alone) possesses, then point out that she could have assigned different credences than she actually does without violating the core Bayesian rules.[12]

Given this fact, it can be useful to isolate two independent aspects of an agent's credal history. Presented with a series of credence distributions, each of which satisfies the probability axioms and Ratio Formula, and each of which is generated from the last by Conditionalization, we can first determine the total evidence possessed by the agent at each time. To do so, we simply look at the contingent propositions of which she is certain (to which she assigns unconditional credence 1) at each time, and construct their conjunction.[13] But second, the Hypothetical Priors Theorem guarantees that we can abstract a hypothetical prior from this agent's series of distributions as well. The hypothetical prior represents whatever non-evidential influences combined with the agent's evidence to generate her credences at each time in the series. By plugging the agent's total evidence at a given time into this hypothetical prior, we can re-create her credence distribution at that time. And if we imagine a counterfactual case in which her total evidence remains the same at a given time yet she assigns different credences to particular propositions, any hypothetical prior consistent with her attitudes in that counterfactual situation must conflict with the one we've constructed for her actual credences. (For the same reason that your suspicious friend's cr_H values must be different from your own for him to have a different cr_1.)[14]

I refer to the non-evidential influences represented by an agent's hypothetical prior cr_H as her **evidential standards**. Some people are more skeptical than others, and so require more evidence to become confident in particular propositions (that Americans actually landed on the moon, that a lone gunman shot JFK, that a material world exists). Some people are interested in avoiding high confidence in falsehoods, while others are more interested in obtaining high confidence in truths. Some people are more inclined to believe elegant scientific theories, while others incline towards the theory that hews closest to the data. All these differences may be reflected

Figure 4.3: Hypothetical priors as evidential standards



in an agent's evidential standards, which combine with the evidence they receive to furnish their views about the world.

An agent's hypothetical prior function cr_H is regular; it contains no contingent certainties and therefore no evidence. That means it doesn't reflect changes in an agent's evidence as she learns about the world and updates by Conditionalization. The agent's hypothetical prior remains constant as different batches of total evidence are plugged into it over time. Instead of being the initial member of an agent's Conditionalization series, or identical to any credence distribution she actually assigns at a particular time, the hypothetical prior "hovers above" the agent's ongoing series of distributions, as depicted in Figure 4.3. Again, the solid arrows represent ongoing conditionalizations, while the dashed arrows represent the possibility of generating an agent's distribution at a given time by conditionalizing cr_H on her total evidence at that time.

But shouldn't an agent's evidential standards change over time, as she gains information? Here it's important to distinguish *ultimate* from *ongoing* evidential standards. Your ultimate evidential standards, represented in your hypothetical priors, determine attitudes when combined with bodies of *total* evidence. Your ongoing evidential standards at a particular time determine attitudes in combination with *further* pieces of evidence you may receive going forward. As you go through life you gain evidence that not only changes your attitudes towards particular propositions, but also changes the significance you attach to further pieces of evidence that might arrive. Yet we can represent all these changes as proceeding from an ultimate set of standards authorizing each individual stance in light of the total evidence possessed at the relevant time.

An analogy may help here. Suppose you're playing five-card stud, a poker game in which each player receives one card at a time and four-of-a-kind is an excellent hand. Before you receive any cards, learning that your last card will be the three of hearts would leave you with a low credence that you will win the hand. Now suppose your first card dealt is the Jack of spades. At this point learning that your last card will be the three of hearts would leave you with an even worse attitude toward your prospects. But when your second, third, and fourth cards turn out to be the three of spades, the three of diamonds, and the three of clubs, your assessment of the three of hearts changes. Now learning that your last card will be the three of hearts would give you a very high credence in victory.

As the game progresses and your evidence about your hand increases, you change your views on whether the three of hearts should make you confident in a win. These evolving views are like ongoing evidential standards. Yet underlying those developing standards lie a stable set of views about what cards the deck contains, the probabilities of various five-card hands if one deals from scratch, and which hands are stronger or weaker in the game. Instead of walking you through the deal one card at a time, I could have dropped you into a given point in the middle of the game and told you all the cards that had been seen. Combining this total evidence with your underlying knowledge of the game would generate a particular degree of confidence that you're going to win. Though this underlying knowledge is not strictly speaking a hypothetical prior (since it contains contingent information about decks, five-card stud, etc.), it plays a similar role to ultimate evidential standards in the context of this game.

Evidence and evidential standards come up in a variety of contexts in epistemology, many having nothing to do with degrees of belief. Bayesian epistemology provides a particularly elegant formal apparatus for isolating each of these elements, by way of contingent credence-1 propositions and hypothetical priors. Once we have the ability to separate an agent's evidence from her evidential standards, an obvious question arises: What rational constraints are there on evidential standards? We've already stipulated that hypothetical priors must obey the probability axioms and Ratio Formula.[15] But are there more stringent requirements than that? Some probabilistic hypothetical priors will be anti-inductive, or will recommend highly skeptical attitudes in the face of everyday bundles of total evidence. Can we rule out such hypothetical priors as rationally impermissible? And once we start adding constraints beyond the core Bayesian rules, how many distinct hypothetical priors will wind up rationally allowed? This will be our first topic in Chapter 5, as we distinguish between Objective and Subjective Bayesianism.

4.4 Exercises

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

Problem 4.1. Galileo intends to determine whether gravitational acceleration is independent of mass by dropping two cannonballs of differing mass off the Leaning Tower of Pisa. Conditional on the quantities' being independent, he is 95% confident that the cannonballs will land within 0.1 seconds of each other. (The experiment isn't perfect—one ball might hit a bird.) Conditional on the quantities' being dependent, he is 80% confident that the balls *won't* land within 0.1 seconds of each other. (There's some chance that although mass affects acceleration, it doesn't have *much* of an effect.)*

(a) Before performing the experiment, Galileo is 30% confident that mass and gravitational acceleration are independent. How confident is he that the cannonballs will land within 0.1 seconds of each other?

(b) After Galileo conditionalizes on the evidence that the cannonballs landed within 0.1 seconds of each other, how confident is he in each hypothesis?

Problem 4.2. Prove that Conditionalization retains certainties. In other words, prove that if $\text{cr}_i(H) = 1$ and cr_j is generated from cr_i by Conditionalization, then $\text{cr}_j(H) = 1$ as well.

Problem 4.3. Prove that Conditionalization is cumulative. That is, prove that for any cr_i, cr_j, and cr_k, conditions 1 and 2 below entail condition 3.

1. For any proposition X in $\mathcal{L}$, $\text{cr}_j(X) = \text{cr}_i(X \mid E)$.
2. For any proposition Y in $\mathcal{L}$, $\text{cr}_k(Y) = \text{cr}_j(Y \mid E')$.
3. For any proposition Z in $\mathcal{L}$, $\text{cr}_k(Z) = \text{cr}_i(Z \mid E \,\&\, E')$.

Problem 4.4. (a) Provide an example in which an agent conditionalizes on new evidence, yet her credence in a proposition *compatible with* the evidence decreases. That is, provide an example in which H and E are consistent, yet $\text{cr}_2(H) < \text{cr}_1(H)$ when E is learned between t_1 and t_2.

*This is a version of a problem from Julia Staffel.

(b) Prove that when an agent conditionalizes on new evidence, her credence in a proposition that *entails* the evidence cannot decrease. That is, when $H \vDash E$, it must be the case that $\text{cr}_2(H) \geqslant \text{cr}_1(H)$ when E is learned between t_1 and t_2.

(c) Prove that as long as $\text{cr}_1(H)$ and $\text{cr}_1(E)$ are both nonextreme, conditionalizing on E increases the agent's credence in H when $H \vDash E$.[†]

Problem 4.5. Reread the details of the Base Rate Fallacy example in Section 4.1.2. After you apply the diagnostic test once and it yields a positive result, your credence that the subject has the disease should be about 0.009.

(a) Suppose you apply the test a second time to the same subject, and it yields a positive result once more. How confident should you now be that the subject has the disease? (Assume that D and $\sim D$ each screen off the results of the first test from the results of the second.)

(b) How many consecutive tests (each independent of the results of prior tests conditional on both D and $\sim D$) would have to yield positive results before your confidence that the subject has the disease exceeded 50%?

(c) Does this shed any light on why patients diagnosed with rare diseases are often advised to seek a second opinion? Explain.

Problem 4.6. Your friend Jones has a gambling problem. His problem is so bad that he gambles on whether to gamble. Every time he goes to the track, he flips a fair coin to determine whether to bet that day. If it comes up heads he bets on his favorite horse, Speedy. If it comes up tails he doesn't bet at all.

On your way to the track today, you were 1/6 confident that out of the six horses running, Speedy would win. You were 1/2 confident that Jones's coin would come up heads. And you considered the outcome of the horse race independent of the outcome of the coin flip. But then you saw Jones leaving the track with a smile on his face. Clearly either Jones bet on Speedy and won, or Jones didn't bet and Speedy didn't win.

(a) Using a language with the atomic propositions H (for heads on the coin) and S (for a Speedy win), express the information you learn when you see Jones smiling.

[†]This problem was inspired by a problem of Sarah Moss'.

(b) After updating on this information by conditionalizing, how confident are you that Speedy won? How confident are you that the coin came up heads?

(c) Explain why one of the unconditional credences you calculated in part (b) differs from its prior value and the other one doesn't. Be sure to include an explanation of why *that* unconditional credence was the one that changed out of the two. ("Because that's what the math says" is not an adequate explanation—we want to know why the mathematical outcome *makes sense*.)

Problem 4.7. At t_1, t_2, and t_3, Jane assigns credences over the language $\mathcal{L}$ constructed from atomic propositions P and Q. Jane's distributions satisfy constraints 1 through 6:

1. At t_1, Jane is certain of $Q \supset P$, anything that proposition entails, and nothing else.
2. Between t_1 and t_2 Jane learns P and nothing else. She updates by conditionalizing between those two times.
3. $\mathrm{cr}_1(Q \mid P) = 2/3$.
4. $\mathrm{cr}_3(Q \mid {\sim}P) = 1/2$.
5. $\mathrm{cr}_3(P \supset Q) = \mathrm{cr}_2(P \supset Q)$.
6. At t_3, Jane is certain of ${\sim}(P \,\&\, Q)$, anything that proposition entails, and nothing else.

(a) Completely specify Jane's credence distributions at t_2 and t_3.

(b) Create a hypothetical prior for Jane. In other words, specify a regular probabilistic distribution cr_H over $\mathcal{L}$ such that cr_1 can be generated from cr_H by conditionalizing on Jane's set of certainties at t_1; cr_2 is cr_H conditionalized on Jane's certainties at t_2; and cr_3 is cr_H conditionalized on Jane's t_3 certainties.

(c) Does Jane update by Conditionalization between t_2 and t_3? Explain how you know.

(d) The Hypothetical Priors Theorem says that *if* an agent always updates by conditionalizing, *then* her credences can be represented by a hypothetical prior distribution. Is the converse of this theorem true?

Problem 4.8. Suppose you have a finite partition $\{B_1, B_2, \dots B_n\}$. Suppose also that between t_1 and t_2 you conditionalize on evidence equivalent to a disjunction of some of the Bs. Prove that for any A in $\mathcal{L}$ and any B_i such that $\text{cr}_2(B_i) > 0$,

$$\text{cr}_2(A \mid B_i) = \text{cr}_1(A \mid B_i)$$

Problem 4.9. Do you think only one set of evidential standards is rationally permissible? Put another way: If two agents' series of credence distributions cannot be represented by the same hypothetical prior distribution, must at least one of them have assigned irrational credences at some point?

4.5 Further reading

Introductions and Overviews

Ian Hacking (2001). *An Introduction to Probability and Inductive Logic*. Cambridge: Cambridge University Press

Chapter 15 works through many excellent examples of applying Bayes' Theorem to manage complex updates.

Classic Texts

Rudolf Carnap (1950). *Logical Foundations of Probability*. Chicago: University of Chicago Press

Section 45B of Chapter IV contains Carnap's discussion of the Principle of Total Evidence.

Extended Discussion

Paul Teller (1973). Conditionalization and Observation. *Synthese* 26, pp. 218–258

Offers a number of arguments for the Conditionalization updating norm. (We'll discuss the Dutch Book argument for Conditionalization that Teller provides in Chapter 9.)

Isaac Levi (1980). *The Enterprise of Knowledge*. Boston: The MIT Press

Though Levi's notation and terminology are somewhat different from mine (my "evidential standards" are his "confirmational commitments"), Chapter 4 thoroughly works through the mathematics of hypothetical priors. Levi also discusses various historically-important Bayesians' positions on how many distinct hypothetical priors are rationally permissible.

Notes

[1]Remember that 1 is not a prime number, while 2 is!

[2]A bit of reflection on Equation (4.2) will reveal that the constant we multiply by in the second step of our stochastic truth-table updating process—the **normalization factor**—will always be the reciprocal of the agent's earlier unconditional credence in the evidence. In other words, the second step divides all nonzero state-description credences by $\mathrm{cr}_i(E)$.

[3]We can also now see an alternate explanation for steps (3.40) and (3.42) of Lewis's triviality proof from Section 3.3. The proposal assessed there is that for some conditional $\rightarrow$, the agent's conditional credence $\mathrm{cr}(Z \mid Y)$ for any Y and Z in $\mathcal{L}$ equals her unconditional credence in $Y \rightarrow Z$. Whatever motivates that proposal, we should want the proposal to remain true even after the agent learns some information X. If the relevant values are going to match after conditionalization on X, it must be true before conditionalization that $\mathrm{cr}(Y \rightarrow Z \mid X) = \mathrm{cr}(Z \mid Y \,\&\, X)$, which is just Equation (3.48).

[4]For reasons we are now in a position to understand, the term "posterior" is sometimes used ambiguously in the Bayesian literature. I have defined "posterior" as an agent's credence in the hypothesis given the evidence—$\mathrm{cr}(H \mid E)$. If the agent updates by conditionalizing on E, this will equal her credence in the hypothesis after the update. The terms "prior" and "posterior" come from the fact that on an orthodox Bayesian position, those quantities pick out the agent's unconditional credences in the hypothesis before and after the update. But unorthodox Bayesians who prefer an alternative updating rule to Conditionalization nevertheless sometimes refer to an agent's post-update credence in a hypothesis as her "posterior". As I've defined the term, this is strictly speaking incorrect.

[5]Someone involved in neuroscience recently told me that when a prisoner in the American penal system comes up for parole, a particular kind of brain scan can predict with greater than 90% accuracy whether that prisoner will, if released, be sent back to jail within a specified period of time. He suggested that we use this brain scan in place of traditional parole board hearings, whose predictive accuracy is much lower. I asked why we don't just apply the brain scan to everyone in American society, rather than wait to see if a person commits a crime worth sending them to jail. He replied that the base rates make this impractical: While the recidivism rate among prisoners is fairly high, the percentage of ordinary Americans committing crimes is low, so the scan would generate far too many false positives if used on the general population.

[6]In Section 2.4.1 I mentioned that Bayesians often work with an agent's set of doxastically possible worlds instead of the full set of logically possible worlds, understanding "mutually exclusive" and "tautology" in the Kolmogorov axioms in terms of the restricted doxastic set. The Regularity Principle concerns the *full* set of logically possible worlds—it forbids assigning credence 0 to any proposition that is true in at least one of them. So for the rest of this section, references to "contingent propositions", "tautologies", etc. should be read against that full logical set of possibilities.

[7]Throughout this section I identify credence 1 with absolute certainty in a proposition and credence 0 with ruling that proposition out. This becomes more complicated when we consider events with infinitely many possible outcomes; we'll consider the relevant complications in Chapter 5.

[8]Consider a person who thinks the refrigerator light is always on, because it's on whenever she opens the refrigerator to look.

[9]Normality makes every probabilistic credence function—even cr_H—assign unconditional credence 1 to tautologies. Depending on how one thinks of evidence, this might

mean that all probabilistic credence functions contain some form of tautological evidence. In what follows we will be interested only in agents' acquisition and response to contingent evidence; for ease of locution I will often leave the "contingent" implied.

[10]I have selected a capital "H" for the subscript of the hypothetical prior distribution cr_H so it is not confused with distribution cr_h occurring later in the Conditionalization series (presumably just after cr_g and just before cr_i).

[11]I learned of Lewis's "superbaby" talk from Alan Hájek. Susan Vineberg suggested to me that Lewis's inspiration for the term may have been I.J. Good's (1968) discussion of "an infinitely intelligent newborn baby having built-in neural circuits enabling him to deal with formal logic, English syntax, and subjective probability"—to which we shall return in Chapter 6.

[12]On some (but certainly not all) mentalist views of evidence, the credences one assigns count as part of one's evidence. Thus any change in an agent's credences counts as a change in her evidence, and we cannot create a counterfactual that holds an agent's total evidence fixed while varying her credences. I think that even on a view of evidence like that it would be worthwhile to distinguish evidence from evidential standards, perhaps by focusing on the evidence *relevant* to a particular proposition and suggesting that one's credence in that proposition isn't relevant evidence. But the distinction would certainly be more difficult to work out.

[13]Strictly speaking, if a probabilistic agent is certain of one contingent proposition at a given time she will be certain of infinitely many, since any contingent proposition entails infinitely many contingent propositions. (Think, for instance, of the infinitely-many disjunctions one can form by disjoining the contingent proposition with itself finitely-many times.) But as long as we are using a language with finitely-many atomic propositions, one will always be able to construct a finite proposition that is equivalent to the conjunction of all the contingent propositions of which the agent is certain. One strategy is to use the negation of the disjunction of all the state-descriptions to which the agent assigns credence 0.

[14]The Hypothetical Priors Theorem says that if an agent always updates by Conditionalization, there will be *at least one* cr_H distribution consistent with her series of credence distributions. While future chapters will suggest rational constraints on hypothetical priors beyond the probability axioms and Ratio Formula, if we demand only that hypothetical priors satisfy those norms then for any realistic series of agent credence distributions (in particular, any series whose first element contains at least some contingent certainties), there will be many cr_H distributions consistent with that series. For instance, in the earlier table with propositions C, P, and E, we could've filled in the bottom four lines of cr_H with any nonnegative values summing to 1/2 and still made the function consistent with cr_1, cr_2, and cr_3 in the relevant fashion. (I filled in those four lines as if, should you have learned that the die didn't come from a casino, you nevertheless would have assumed it to be fair.)

Any hypothetical prior that is consistent with all the distributions in a conditionalizing series will generate the same credences should that series be further extended by Conditionalization, so the differences among such hypothetical priors need not make much difference. The crucial point is that if two agents assign different credences at a given time despite sharing the same total evidence, there will be no hypothetical prior that is consistent with the entirety of both their distribution series.

[15]The Hypothetical Priors Theorem requires hypothetical priors to be regular, and you might think that's because we have endorsed Regularity as a rational constraint on evidential standards. Yet hypothetical priors are required to be regular not because of

any allegiance to the Regularity norm, but instead because we need to cleanse them of contingent evidence for them to represent truly non-evidential influences on an agent's attitudes. Remember, the Hypothetical Priors Theorem applies only to an agent who updates by Conditionalization, and Conditionalization is in tension with the Regularity Principle.