Chapter 6: Belief Over Time

So far time has not entered into the discussion of degree of belief in any particular way. However, the difference between synchronic (at the same time) and diachronic (across different times) rationality is central to Bayesianism. Much of the power and appeal of Bayesianism comes from the diachronic rule of Conditionalization and the way it can be used to structure inquiry and investigation. However, the justifications for this rule are of a different character than the justifications of the synchronic norms. In this chapter, I will discuss diachronic rules of rationality that have been proposed for degrees of belief, and the justifications that have been given for them, as well as some objections. In the next chapter, I will investigate the types of normativity involved in all of these rules, using the contrast between the synchronic and diachronic rules as one starting point.

Conditionalization

The primary diachronic rule that has been proposed for Bayesianism is Conditionalization. To state this rule we need a notion not just of degree of belief, but also conditional degree of belief. It has sometimes been said (Adams, 1965, Edgington, 1995) that one’s degree of belief in \( A \) conditional on \( B \), notated as “\( P(A|B) \)”, can be identified with one’s degree of belief in the conditional, “if \( B \) then \( A \)”. While this thought is surely helpful in getting at the concept, this can’t suffice as an analysis unless we know what “if \( B \) then \( A \)” actually means, which is a notoriously difficult question. (Bennett, 2003) It further turns out that there are severe restrictions on what the logic of conditionals could be like in order for the identification of conditional probability with the probability of a conditional to hold, without putting very tight restrictions on what probabilities are possible. (The best results I know of at this point are the ones due to Bacon (2015).)

However, I think the better way to understand conditional degree of belief is to understand the two relations that it is said to enter into. One is the synchronic relation: \( P(A|B) = P(A&B)/P(B) \) — one’s conditional degree of
belief should be the ratio between one’s degree of belief in the conjunction and one’s degree of belief in the proposition being conditioned on. The other is a diachronic relation — if \( P_1 \) represents one’s degrees of belief at time \( t_1 \), and \( P_2 \) represents one’s degrees of belief at a later time \( t_2 \), and if \( B \) is everything that one learns between \( t_1 \) and \( t_2 \), then one should have \( P_2(A) = P_1(A | B) \). The term “conditionalization” is often reserved to refer to the diachronic rule, but without the synchronic rule specifying values for \( P_1(A | B) \), the force of the rule is hard to interpret.

I will describe some standard arguments for these relations later, but first I would like to illustrate them. Together, these two rules provide surprisingly strong and interesting constraints on how one’s degrees of belief should change over time as one learns about the world. For instance, let \( H \) be some hypothesis about the world, and let \( E \) be some potential evidence about it. If \( H \) entails \( E \), so that there is no epistemic possibility for the agent on which \( H \) is true without \( E \) also being true, then \( H \) and \( H & E \) are equivalent for the agent. Thus, if the agent learns \( E \) (and nothing else) between \( t_1 \) and \( t_2 \), then these rules say that

\[
P_2(H) = P_1(H | E) = \frac{P_1(H & E)}{P_1(E)} = \frac{P_1(H)}{P_1(E)}.
\]

What this means is that if a hypothesis makes a prediction, and that prediction is learned to be true, then as long as one wasn’t already quite confident in that prediction (as long as \( P_1(E) < 1 \)) then one’s confidence in the hypothesis should go up. In particular, one’s confidence in the hypothesis should go up directly in proportion to how surprising the prediction was that was verified. A hypothesis that successfully predicts that it will be warmer in Montreal than in Houston this winter will receive greater confirmation than a hypothesis that successfully predicts the reverse (though in this case only one of these predictions could be successful in any given year).

Conversely, if \( E \) is some evidence that is strong enough to entail a hypothesis, so that there is no epistemic possibility where the evidence holds while the hypothesis does not, then \( H & E \) is equivalent to \( E \), so

\[
P_2(H) = P_1(H | E) = \frac{P_1(H & E)}{P_1(E)} = \frac{P_1(E)}{P_1(E)} = 1.
\]

In fact, it is natural to propose a parallel rule for the way that epistemic possibilities should change over time. When one learns a proposition \( E \), one should eliminate all epistemic possibilities in which \( E \) doesn’t hold. The remaining set

---

1Readers might note in light of what I said in the previous chapter that this formula has problems when \( P(B) = 0 \). Hájek (2003) points out that for this and other reasons, this formula can’t be taken as the definition of conditional probability, and raises some other challenges beyond the issue of probability 0. There are many different responses to this issue, which I discuss in my [CITATION]. The main feature they have in common is that they replace the formula in the text with \( P(A & B) = P(A | B)P(B) \), which is equivalent when \( P(B) \neq 0 \) and says about \( P(A | B) \) when \( P(B) = 0 \). The responses differ in how they supplement this constraint with a theory of \( P(A | B) \) when \( P(B) = 0 \). All of these responses are mathematically quite challenging, and beyond the scope of this book.
of epistemic possibilities will not in general have probability equal to 1, so the rule of conditionalization then says to proportionally raise the probability of everything that is left in just such a way that the total probability is equal to 1. Eliminating the possibilities in which \( E \) doesn’t hold is done by replacing \( H \) with \( H \& E \), and the proportional raising is done by dividing by \( P(E) \). This is sometimes described by a metaphor of confidence as a substance that is spread out over the epistemic possibilities, with updates proceeding by removing the confidence from possibilities that are eliminated and spreading it out proportionally over what is left. (At least since Jeffrey (1965), many authors have been concerned about the fact that this update rule assumes that evidence is learned with certainty — alternatives that don’t make this assumption will be discussed later.)

Many further consequences of this rule of updating are explored by Howson and Urbach (1989). One particularly striking consequence is the way that this rule seems to vindicate a kind of inductive reasoning against Humean skepticism. Let \( H \) be some hypothesis, and let \( E_1, E_2, E_3, \ldots \) be a sequence of potential observations that would all be entailed by the truth of this hypothesis. (For instance, \( H \) might be the hypothesis that fire is always hot, and each \( E_i \) might be the claim that fire is hot on the \( i \)th time that we touch it.) Hume famously said that since there are no necessary connections between distinct matters of fact, the observations \( E_1, \ldots, E_i \) give us no reason to believe either \( H \) or even \( E_{i+1} \).

However, the Bayesian says that we should think in terms of confidence rather than belief. It may be that no number of observations of fire being hot can put us in a position to be certain that fire will still be hot on our next observation. But as long as we had some slightly positive confidence in \( H \) to begin with, we can show that there must be some number \( i \) of observations of fire such that observing it to be hot \( i \) times gives us reason to be 99% confident that the next observation will be hot as well. If this were not the case, then we could derive a violation of the Bayesian rules.

Let \( P_0 \) be the agent’s degrees of belief before any observations, and let \( P_i \) be the agent’s degrees of belief after having observed \( E_1, E_2, \ldots, E_i \). Since \( H \) entails every piece of evidence, we know that for any \( i \), we must have \( P_0(H) \leq P_0(E_1 \& \ldots \& E_i) \). But an application of the two rules of conditionalization tells us that,

\[
P_0(E_1 \& \ldots \& E_i) = P_0(E_i | E_1 \& \ldots \& E_{i-1}) P_0(E_1 \& \ldots \& E_{i-1}) = P_{i-1}(E_i) P_0(E_1 \& \ldots \& E_{i-1}).
\]

Applying this process to the probability of the slightly shorter conjunction on the right and repeating tells us that

\[
P_0(E_1 \& \ldots \& E_i) = P_{i-1}(E_i) P_{i-2}(E_{i-1}) \ldots P_1(E_2) P_0(E_1).
\]

That is, the prior degree of belief in a conjunction should be the product of the later degree of belief one has in each conjunct after learning all the prior conjuncts. If none of these later degrees of belief is above 99%, then for any positive number \( x \), there is an \( i \) such that the product of the first \( i \) of these
later degrees of belief is below $x$. In particular, if $x = P_0(H)$, then this $i$ would contradict the claim that $P_0(H) \leq P_0(E_1 \& \ldots \& E_i)$. And the same argument will work for any threshold other than 99% as well. As long as there is some general hypothesis that entails a sequence of observations, and one starts out with positive degree of confidence in that general hypothesis, then for any desired degree of confidence, there is some number of these observations that together will entail that one should be that confident in at least the next observation.

This result has some limitations. First, the result doesn’t tell us that $P_i(H)$ will ever be high — if $H$ entails the $E_i$ but also has substantial further content (for instance, $H$ might be a theory of what the underlying structure of the world is like, but the $E_i$ might just be a sequence of superficial observations of macroscale phenomena) then it could well be that no number of observations of $E_i$ can raise the probability of $H$ itself to a high value. Second, the result doesn’t apply if $P_0(H) = 0$. Karl Popper famously used a version of this argument in reverse to start from the Humean idea that no number of observations could ever give us any reason to believe that the next observation will be similar, to then argue that $P_0(H)$ must be less than the product of the probabilities of any number of independent observations, which means it must be 0, because a sufficiently large number will drive the product down arbitrarily far. But interestingly, neither this inductivist argument nor Popper’s anti-inductivist argument relies on the metaphysical notion of necessary connections that Hume was worried about — both proceed from an analysis of rational confidence by itself.

One important consequence of the rules for conditional probability is known as “Bayes’ Theorem”. This theorem was first proved by Bayes (1763), whose application of the theorem to solve a particular inference problem was later taken as the inspiration for Bayesianism as a general view. The basic observation behind the theorem is that $P_1(H \& E)$ can be related to two different conditional probabilities. On the one hand, $P_1(H \& E) = P_1(H|E)P_1(E)$. On the other, $P_1(H \& E) = P_1(E|H)P_1(H)$. This is because of the symmetry of the way in which $E$ and $H$ enter into the formula. As long as the values involved are non-zero, we can put these two equations together to get Bayes’ Theorem:

$$P_1(H|E) = P_1(E|H)\frac{P_1(H)}{P_1(E)}.$$ 

In the application to scientific reasoning and statistical methodology, $E$ is often taken to be a piece of evidence that one will observe, while we are considering several hypotheses $H_i$ that can’t be directly verified. We are interested in how the “posterior” probabilities $P_2(H_i)$ relates to the “prior” probabilities $P_1(H_i)$. The update rule tells us that $P_2(H_i) = P_1(H_i|E)$, but in general we don’t have a clear sense of how $P_1(H_i|E)$ relate to each other. However, by

\[\text{It is also true that } P_1(E|H) = P_1(H|E)\frac{P_1(E)}{P_1(H)}, \text{ but this order turns out to be less useful.}\]
Bayes’ Theorem, we see that

$$P_1(H_i|E) = \frac{P_1(E|H_i)P_1(H_i)}{P_1(E)}.$$ 

Since the denominator of this fraction doesn’t depend on $H_i$, we can see that the posterior probability of each hypothesis $H_i$ will be proportional to its prior probability multiplied by the factor $P_1(E|H_i)$, which is known as the “likelihood” of the hypothesis for the evidence. (Note that the likelihood of a hypothesis is not how likely the hypothesis is, but rather how likely the hypothesis makes the evidence. This bit of technical terminology can often be confusing.) In many cases, we have a clear sense of what the relative likelihood of various hypotheses is, because the hypothesis itself is phrased in probabilistic terms.\(^3\)

For instance, if our hypotheses are about the bias of a coin, then on seeing the coin come up heads, each hypothesis should be updated by multiplying our priors by the bias towards heads specified by the hypothesis. If we start out with degree of belief $2/3$ that the coin is fair, and $1/3$ that it is biased with a 60% chance of heads, then the posterior degrees of belief should be proportional to $2/3 \cdot 1/2$ and $1/3 \cdot 0.6$. This results in a posterior of $5/8$ that the coin is fair and $3/8$ that it is biased with a 60% chance of heads. If our hypotheses specify different potential fractions of the electorate that might favor one party over another, and our evidence is that in a random sample of 100 potential voters, 63 of them favored a given party, the calculations for how likely this evidence would be if the true fraction of the electorate is 60% or 70% will be more complicated, but again they can be worked out.

However, although Bayes’ Theorem lets us combine relatively objective likelihoods with the priors to determine posterior probabilities, differences among priors can still confound attempts to determine objectively whether a given piece of evidence counts in favor of or against a particular hypothesis. For instance, consider three hypotheses about the history of the earth. $H_1$ says that life on Earth has existed for hundreds of millions of years and that during that time there were many changes in the species that exist. $H_2$ says that life on Earth has existed for hundreds of millions of years but has been roughly similar throughout that entire period. $H_3$ says that life on Earth has existed for 6000 years, since it was created by a divine being, who also created layers of rock that simulate a long past in order to test the faith of humans. Let $E$ be the observation that some layers of rock contain the apparent skeletons of large

\(^3\)If one wants to investigate the denominator, note that if the $H_i$ are a mutually exclusive and exhaustive set of possibilities, so that the agent is certain that exactly one of them is true, then $P(E) = P(E\&H_1) + \cdots + P(E\&H_n)$. In each case, $P(E\&H_i) = P(E|H_i)P(H_i)$, so this denominator can also be worked out given the likelihoods and priors of the various hypotheses.

However, in realistic scenarios, we are often considering a few explicit hypotheses, but are not certain that one of them must be true. Thus, we must add some additional “none of the above” hypothesis, which is traditionally called the “catch-all”. Although likelihoods of explicit hypotheses can often be determined by studying the hypothesis itself, the likelihood for the catch-all is much less determinate. See for instance Maher (1995) for an approach to this issue.
reptiles that don’t exist today. Further, let us assume that all researchers agree
that \( P(E|H_1) = .05, P(E|H_2) = .005, \) and \( P(E|H_3) = .5 \). (The idea is that
under the evolutionary theory, it is likely that there were different creatures in
the past, but moderately unlikely that they were large reptiles; under the stable
theory it is quite unlikely that any fossils of this sort would exist; and under the
young creationist theory, it is just as likely as not that the deity would include
this sort of fossil.)

For someone with prior degrees of belief on which \( P(H_1) = .09, P(H_2) = .9, \) and
\( P(H_3) = .009 \) (with .001 left over for a catch-all), the posterior degrees of
belief in these theories will all be equal (because the likelihoods were in inverse
proportion to the priors). In particular, this means that learning \( E \) will increase
their degree of belief in \( H_1 \). However, for someone with prior degrees of belief
on which \( P(H_1) = .09, P(H_2) = .009, \) and \( P(H_3) = .9 \) (with .001 left over for
a catch-all), the situation will be quite different. The prior for \( H_3 \) is ten times
that of \( H_1 \), as is the likelihood of \( H_3 \). Thus, the posterior for \( H_3 \) will be one
hundred times that of \( H_1 \), so in particular \( H_1 \) must end up with a posterior of
less than .01. Even though the two researchers agreed on the likelihoods, and
had the same prior for \( H_1 \), one of them took the evidence \( E \) to be moderately
strong in favor of \( H_1 \) while the other took it to be moderately strong evidence
against \( H_1 \).

This is because the two gave different credence to various alternatives. In
general, the alternative that has the highest likelihood for the evidence will go
up in credence. But for alternatives with intermediate likelihood, it matters
whether most of the probability the agent assigns to alternatives was assigned
to alternatives with higher or lower likelihood. If there is some reason why
particular prior distributions of confidence are irrational, then perhaps we can
solve this problem, and confirmation can be determined objectively. But if every
probabilistically coherent prior distribution of confidence is equally rational,
then the only cases in which the mathematics determines whether confirmation
or disconfirmation occurs will be cases in which the evidence either entails the
hypothesis or is entailed by it.

Arguments for conditionalization

The basic argument for conditionalization was given at least as early as Teller
(1973). It proceeds by Dutch book, like the initial arguments from Chapter
2. We must first specify what role conditional degree of belief should play in
governing betting behavior. The traditional interpretation of degree of belief in
betting behavior is that if one’s degree of belief in \( A \) is \( x \), then one is willing
to pay \( x \) for a bet that pays out 1 if \( A \) is true, and nothing otherwise. That is,
one finds fair a situation in which one loses \( x \) if \( A \) is false, and gains \((1-x)\) if
\( A \) is true; if the stakes are \( S \), then one finds fair a situation in which one loses
\( xS \) if \( A \) is false, and gains \((1-x)S\) if \( A \) is true. If \( S \) is negative, then we must
switch the interpretation of gain or loss in accordance with the signs. If one or
both of these outcomes is improved, then one finds the situation favorable, and
if one or both are worsened then one finds it unfavorable.
Conditional degree of belief is then said to govern one’s acceptance of conditional bets. In an unconditional bet on \( A \), one pays a certain amount of money for a chance to win if \( A \) is true. In a conditional bet on \( A \), conditional on \( B \), one pays a certain amount of money for a chance to win if \( A \) is true, but the bet is called off (with no winning, and a refund of the initial payment) if \( B \) is false. That is, if one pays \( x \) for a bet on \( A \) conditional on \( B \), then the net result is that one loses \( x \) if \( B \) is true and \( A \) is false, one gains \( (1 - x) \) if both \( B \) and \( A \) are true, and one has no gain or loss if \( B \) is false. More generally, if one’s degree of belief in \( A \) conditional on \( B \) is \( x \), then one finds fair a situation in which one loses \( xS \) if \( A \) is false and \( B \) is true, and gains \( (1 - x)S \) if \( A \) and \( B \) are both true, and has no change if \( B \) is false. If one or more of these three outcomes is improved, then one finds the situation favorable, and if one or more are worsened then one finds it unfavorable.

This can be summarized in the following table, with the relevant degrees of belief and stakes of each bet:

<table>
<thead>
<tr>
<th></th>
<th>( B &amp; A )</th>
<th>( B &amp; \sim A )</th>
<th>( \sim B &amp; A )</th>
<th>( \sim B &amp; \sim A )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bet on ( A )</td>
<td>(+1 - P(A))S( A )</td>
<td>(-P(A)S( A )</td>
<td>(+1 - P(A))S( A )</td>
<td>(-P(A)S( A )</td>
</tr>
<tr>
<td>Bet on ( B )</td>
<td>(+1 - P(B))S( B )</td>
<td>(+1 - P(B))S( B )</td>
<td>(-P(B)S( B )</td>
<td>(-P(B)S( B )</td>
</tr>
<tr>
<td>Bet on ( A &amp; B )</td>
<td>(+1 - P(A</td>
<td>B))S( A &amp; B )</td>
<td>(-P(A</td>
<td>B)S( A &amp; B )</td>
</tr>
<tr>
<td>Conditional bet on ( A ) given ( B )</td>
<td>(+1 - P(A</td>
<td>B))S( A</td>
<td>B )</td>
<td>(-P(A</td>
</tr>
</tbody>
</table>

Note what happens if we set \( S_B = -P(A \& B)/P(B) \) and \( S_{A \& B} = 1 \). Then the second and third lines of that table become the first two lines of this table, and the sum of these two individually acceptable bets is the third line:

<table>
<thead>
<tr>
<th></th>
<th>( B &amp; A )</th>
<th>( B &amp; \sim A )</th>
<th>( \sim B &amp; A )</th>
<th>( \sim B &amp; \sim A )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bet on ( B )</td>
<td>(-P(A</td>
<td>B)/P(B)) + ( P(A</td>
<td>B) )</td>
<td>(-P(A</td>
</tr>
<tr>
<td>Bet on ( A &amp; B )</td>
<td>(+1 - P(A</td>
<td>B))</td>
<td>(-P(A</td>
<td>B))</td>
</tr>
<tr>
<td>Sum of both bets</td>
<td>(+1 - P(A</td>
<td>B)/P(B))</td>
<td>(-P(A</td>
<td>B)/P(B))</td>
</tr>
</tbody>
</table>

If \( P(A|B)/P(B) > P(A|B) \) then this sum is worse than a fair conditional bet on \( A \) given \( B \) at stakes of 1, and if \( P(A|B)/P(B) < P(A|B) \) then this sum is better than a fair conditional bet on \( A \) given \( B \) at stakes of 1. But the sum of two fair bets shouldn’t be either strictly better or strictly worse than another fair bet, so we should have \( P(A|B)/P(B) = P(A|B) \) as required by the synchronic rule for conditional probability.

For the diachronic rule, we have to consider a slightly different argument involving bets at different times. Imagine that an agent knows that she is about to learn either \( B \) or \( \sim B \). Consider what bets she is willing to accept if she has \( P_2(A) > P_1(A|B) \) in the scenario where she happens to learn \( B \). In that case, she is currently willing to accept \( P_1(A|B) \) for a conditional bet requiring her to pay 1 if \( A \) and \( B \) are both true that will be called off if \( B \) is false. But then when she learns \( B \), she will now be willing to pay \( P_2(A) \) for a bet that will give her 1 if \( A \) is true. Since \( B \) is true, the first bet won’t be called off, so the two payments she has already made net to a loss for her (since she gained the smaller value at the beginning and then paid the larger value at \( t_2 \)). If \( A \) is true, the two sets of winnings will cancel out, and if \( A \) is false there are no more
further winnings. So she has ended up losing money either way. Reversing the
direction of the bets gives a Dutch book if she has \( P_2(A) < P_1(A|B) \). Thus,
she should have \( P_2(A) = P_1(A|B) \), and similar arguments show that she should
have \( P_2(A) = P_1(A|\neg B) \) if \( \neg B \) is what she ends up learning.

A somewhat different argument has been given by Greaves and Wallace
(2006) for these two rules in the accuracy framework. They assume that an
agent wants to minimize her inaccuracy as measured by a “proper scoring rule”.
If \( I(x, 1) \) is the degree of inaccuracy of having degree of belief \( x \) in a true
proposition, and \( I(x, 0) \) is the degree of inaccuracy of having it in a false one,
then \( I \) is said to be a proper scoring rule iff for a given \( y \), the unique value of
\( x \) to minimize \( yI(x, 1) + (1 - y)I(x, 0) \) is \( x = y \). The idea is that if the agent
herself has degree of belief \( y \) in a proposition \( A \), and she considers possible
alternative degrees of belief \( x \) she might have instead, then she should judge any
alternative degree of belief as being expected to be more inaccurate than her
own. Joyce required that no alternative degrees of belief should be guaranteed
to have less inaccuracy than her own; Greaves and Wallace require further that
no alternative degrees of belief should on average have less inaccuracy than her
own (at least, from her current evaluation). One thing to note about a proper
scoring rule is that if we want to choose \( x \) so as to minimize \( zI(x, 1) + wI(x, 0) \),
where \( z \) and \( w \) are any non-negative numbers, then the unique value of \( x \) that
does so is \( x = z/(z + w) \). This is because \( x = z/(z + w) \) minimizes \( (z/(z + w))I(x, 1)
+ (1 - z/(z + w))I(x, 0) \), and \( zI(x, 1) + wI(x, 0) \) is just this quantity
times \( (z + w) \), which doesn’t depend on \( x \).

Because the scoring rule is proper, as long as the agent has degrees of belief
that satisfy the probability axioms, she will have no reason to unilaterally change
her degrees of belief to some other value. However, she might consider a plan
for what to do if she knows she is about to learn either \( B \) or \( \neg B \). Her plan could
be to update her degree of belief in \( A \) to \( x_B \) if she learns \( B \) and to \( x_{\neg B} \) if she
learns \( \neg B \). The expected accuracy of this plan is thus:

\[
P(A \& B)I(x_B, 1) + P(\neg A \& B)I(x_B, 0) + P(A \& \neg B)I(x_{\neg B}, 1) +
P(\neg A \& \neg B)I(x_{\neg B}, 0).
\]

By adjusting \( x_B \) she can modify the first two components of this plan, and by
adjusting \( x_{\neg B} \) she can modify the second two components. Thus, she should
choose \( x_B \) to minimize

\[
P(A \& B)I(x_B, 1) + P(\neg A \& B)I(x_B, 0),
\]

and she should choose \( x_{\neg B} \) to minimize

\[
P(A \& \neg B)I(x_{\neg B}, 1) + P(\neg A \& \neg B)I(x_{\neg B}, 0).
\]

Using the note at the end of the previous paragraph, she should thus choose
\( x_B = \frac{P(A \& B)}{P(A \& B) + P(\neg A \& B)} = \frac{P(A \& B)}{P(B)} \), and she should choose \( x_{\neg B} = \frac{P(A \& \neg B)}{P(A \& \neg B) + P(\neg A \& \neg B)} =
\frac{P(A \& \neg B)}{P(\neg B)} \). If we think of the conditional degree of belief \( P(A|B) \) as her plan for
updating her degree of belief in \( A \) if she is to learn \( B \), then this just is the
synchronic part of conditionalization, and the diachronic part is the rule that when she updates, she ought to do so in accord with her prior plan.

Thus, the two rules for conditional probability can be given justification both on the action-based and the accuracy-based accounts of the role of degree of belief. Many philosophers have suggested that this diachronic rule is in fact the only way that evidence plays a role in guiding belief, so this seems to help unify some of the perspectives discussed in Chapter 4.

Alternatives to conditionalization

As mentioned above, some authors have objected to conditionalization on the basis that it seems to require that learning proceeds with certainty. Whatever evidence is gained gets probability 1, and the most natural interpretation of what to do with the set of epistemic possibilities is to eliminate the possibilities on which one’s evidence was not true. However, as fallible creatures in an uncertain world, it seems we should never be fully certain that the evidence we have gained is true.

Some authors have suggested that there is no problem if we take the evidence not to be something like, “the ith fire tested was hot”, but rather something more like “what appeared to me to be the ith fire tested appeared to be hot”. But this sort of move will only help if there is some privileged language of sense data that are given to us with Cartesian certainty, from which we hope to infer claims about the uncertain external world. If there can be reasonable doubt or uncertainty about the phenomenal character of our experience, then even the claim that there appeared to be a fire and that it appeared to be hot might be uncertain.

Richard Jeffrey (1965) proposed a rule for updating degrees of belief in cases where no certainty is gained. He considers that in such a case, there may still be some set of propositions that play the role that the evidence gained with certainty in conditionalization. In the simplest version, Jeffrey imagines that there is some particular proposition $E$ that drives the update, but that it doesn’t go all the way to certainty. Instead, the agent’s degree of confidence in $E$ goes from $P_1(E)$ up to some value $P_2(E) = x$, which could be quite high, but still short of certainty. In this case, Jeffrey suggests the experience shouldn’t change the way that the evidence relates to any other proposition. That is, we should have $P_2(A|E) = P_1(A|E)$, and $P_2(A|\neg E) = P_1(A|\neg E)$. Putting these together with the synchronic constraints on conditional probability, we see that $P_2(A) = P_2(A\&E) + P_2(A\&\neg E) = P_2(A|E)P_2(E) + P_2(A|\neg E)P_2(\neg E) = xP_1(A|E) + (1 - x)P_1(A|\neg E)$. In the special case where $x = 1$, we get $P_2(A) = P_1(A|E)$, so standard conditionalization is a special case of Jeffrey conditionalization.

Jeffrey further allowed that this could be generalized to a case where there is not just one proposition driving the update. If $E_1, \ldots, E_n$ are the mutually exclusive and exhaustive alternatives that make up the evidential partition (perhaps the update experience is being driven by catching a glimpse of a tablecloth in candlelight, and each $E_i$ specifies some particular color that the cloth might be), then they can update by the rule $P_2(E_i) = x_i$, where the
add up to 1. Then other propositions update by the calculation \( P_2(A) = x_1 P_1(A|E_1) + \cdots + x_n P_1(A|E_n) \). The condition that \( P(A|E_i) \) stays the same for any proposition \( A \) and any piece of evidence \( E_i \) is known as “rigidity”, and it is what characterizes the sort of update described here.

Under Jeffrey’s characterization, it is hard to say how to characterize the effect of the same experience on agents who had different starting degrees of belief. In ordinary conditionalization, we could just specify the evidence, but for Jeffrey, we need to say how the prior degree of belief turns into the posterior degree of belief. If one person starts at \( P_1(E) = .1 \) and goes to \( P_2(E) = .9 \), then it would be strange for someone else who started at \( P_1(E) = .95 \) to decrease her degree of belief in \( E \) in light of the same evidence. Thus, Hartry Field and Carl Wagner and others have argued that it is better to characterize the update not by the number specifying the value of \( P_2(E) \), or even by specifying how the probability \( P_1(E) \) turns into the probability \( P_2(E) \). Instead, they suggest that we should specify how the odds \( P_1(E)/P_1(\neg E) \) turn into the odds \( P_2(E)/P_2(\neg E) \). Thus, they focus on the “Bayes factor” \( \frac{P_2(E)}{P_2(\neg E)} / \frac{P_1(E)}{P_1(\neg E)} \). By characterizing the update this way, it turns out that we can apply multiple updates on different partitions in different orders and get the same end result. (Though see Weisberg (2009) for the worry that this notion of commutativity, or even just rigidity, just isn’t compatible with the original motivating idea that the evidential relevance of any sort of update could be undermined by learning some other fact later on.)

I won’t discuss the details of the argument for or against Jeffrey’s version of conditionalization here, but they have been given by Skyrms, Myrvold and others (both in Dutch book and accuracy forms). But other authors have argued that updates just aren’t always driven by a partition. (Gallow, 2016, Schoenfield, 2016)

An even more generalized rule has been discussed by Jaynes (2003). Jaynes suggests that we can measure the “distance” between two probability functions by computing the “entropy” of one with regards to the other. This can be thought of as telling us the amount of information needed to get from one distribution to the other, where information is thought of as the reduction in entropy. Thus, from the perspective of any one probability function, it is the probability function of maximum entropy. When performing an update, Jaynes imagines that we have some evidence that specifies some features that the new distribution must have. It could be that the new distribution is required to set a particular proposition \( E \) to certainty. Or it could be that the new distribution is required to set particular probability values \( x_1, \ldots, x_n \) to a partition \( E_1, \ldots, E_n \). Or it could be something else. In each case, Jaynes suggests that we should end up with the probability distribution with the highest entropy out of all the ones that satisfy this condition. If entropy is measured by a particular standard formula, then this agrees with standard conditionalization and Jeffrey conditionalization in the cases just described, but also provides guidance for
A proposition is equivalent to and by distributivity of conjunction and disjunction, the disjunction of these average of $E_i$ will be too. Since we’ll be able to show from here that $P_1(A|P_2(A) = x)$. $P_2(A) = x$ is a proposition about one’s credences after the update. Since one is certain that one will update in line with conditionalization while maintaining awareness about the way one’s update progresses, then reflection follows. Consider first a simple case in which one knows at $t_1$ that one will, by $t_2$, learn which one of the mutually exclusive and exhaustive propositions $E_1, \ldots, E_n$ is true. Now consider $P_1(A|P_2(A) = x)$. $P_2(A) = x$ is a proposition about one’s credences after the update. Since one is certain that one will update by conditionalization, one is certain that $P_2(A)$ is just $P_1(A|E_i)$, for whichever $E_i$ is true. If there is just one $E_i$ for which $P_1(A|E_i) = x$, then $P_2(A) = x$ is true iff $E_i$ is true. If the agent has sufficient self-awareness about her degrees of belief, then she recognizes this fact. Thus, for her, $P_2(A) = x$ can be substituted with $E_i$ in any expression of probability. Thus, $P_1(A|P_2(A) = x) = P_1(A|E_i)$. But this just is $x$.

If there are multiple potential pieces of evidence, say $E_i, E_j, E_k$, that would all result in $P_2(A) = x$, then the argument has a few more steps. In this case, $P_2(A) = x$ is equivalent to $E_i \lor E_j \lor E_k$, and the agent is aware of this fact. Thus,

$$P_1(A|P_2(A) = x) = P_1(A|E_i \lor E_j \lor E_k) = \frac{P(A \land (E_i \lor E_j \lor E_k))}{P(E_i \lor E_j \lor E_k)}.$$

We’ll be able to show from here that $P_1(A|E_i \lor E_j \lor E_k)$ is a kind of weighted average of $P_1(A|E_i)$, $P_1(A|E_j)$, and $P_1(A|E_k)$, which are all equal to $x$, so it will be too. Since $E_i, E_j, E_k$ are mutually exclusive, so are $A \land E_i, A \land E_j, A \land E_k$, and by distributivity of conjunction and disjunction, the disjunction of these propositions is equivalent to $A \land (E_i \lor E_j \lor E_k)$. Thus,

$$\frac{P_1(A \land (E_i \lor E_j \lor E_k))}{P_1(E_i \lor E_j \lor E_k)} = \frac{P(A \land E_i) + P(A \land E_j) + P(A \land E_k)}{P_1(E_i \lor E_j \lor E_k)}.$$

By the synchronic rule of conditional probability, $P_1(A \land E_i) = P_1(A|E_i)P_1(E_i)$, and similarly for the others. Furthermore, by definition of $E_i, E_j, E_k$, we have $P_1(A|E_i) = P_1(A|E_j) = P_1(A|E_k) = x$. Thus, the last expression is equal to

$$x \left( \frac{P_1(E_i) + P_1(E_j) + P_1(E_k)}{P_1(E_i \lor E_j \lor E_k)} \right).$$

**Reflection**

Describe reflection.

Derive it from conditionalization

If one is certain that one will update in line with conditionalization while maintaining awareness about the way one’s update progresses, then reflection follows. Consider first a simple case in which one knows at $t_1$ that one will, by $t_2$, learn which one of the mutually exclusive and exhaustive propositions $E_1, \ldots, E_n$ is true. Now consider $P_1(A|P_2(A) = x)$. $P_2(A) = x$ is a proposition about one’s credences after the update. Since one is certain that one will update by conditionalization, one is certain that $P_2(A)$ is just $P_1(A|E_i)$, for whichever $E_i$ is true. If there is just one $E_i$ for which $P_1(A|E_i) = x$, then $P_2(A) = x$ is true iff $E_i$ is true. If the agent has sufficient self-awareness about her degrees of belief, then she recognizes this fact. Thus, for her, $P_2(A) = x$ can be substituted with $E_i$ in any expression of probability. Thus, $P_1(A|P_2(A) = x) = P_1(A|E_i)$. But this just is $x$.

If there are multiple potential pieces of evidence, say $E_i, E_j, E_k$, that would all result in $P_2(A) = x$, then the argument has a few more steps. In this case, $P_2(A) = x$ is equivalent to $E_i \lor E_j \lor E_k$, and the agent is aware of this fact. Thus,

$$P_1(A|P_2(A) = x) = P_1(A|E_i \lor E_j \lor E_k) = \frac{P(A \land (E_i \lor E_j \lor E_k))}{P(E_i \lor E_j \lor E_k)}.$$
But the numerator is equal to the denominator of that fraction, because $E_i, E_j, E_k$ are mutually exclusive. Thus, we end up with the conclusion that again, $P_1(A | P_2(A) = x) = x$.

By similar reasoning we can show that under these conditions of self-awareness and certainty of conditionalization, $x \leq P_1(A | x \leq P_2(A) \leq y) \leq y$. This argument proceeds as above by first finding all the possible pieces of evidence such that updating on them would leave one’s degree of belief in $A$ within the relevant range. At the end, instead of being able to factor out a single value $x$, we can show that the resulting fraction is at least as great as the one where $x$ is factored out from each term, and no greater than the one where $y$ is factored out from each term, and is this itself between $x$ and $y$.\(^5\)

Similar arguments also work if one is certain not that one will update in line with standard conditionalization, but Jeffrey conditionalization. For Jaynes’ rule of maximum entropy, the situation is more complicated, and some violations can arise.

It is also possible to give a more direct argument for Reflection. Dutch book argument for Reflection.

I have also presented an expected accuracy-based argument for Reflection, in my (2013). This argument proceeds much like that of Greaves and Wallace, but it generalizes to cases where there are infinitely many pieces of evidence that one might learn, and applies to cases of probability 0 as well.

Challenges

Proposed counterexamples to reflection

- Spaghetti
- LSQ
- Self-doubt. I am confident that I won’t have sufficient evidence to justify me having degree of belief .999999 in $A$, so currently I have $P_1(A | P_2(A) = .999999) < .99$.

- van Fraassen’s “death and disability” and “integrity” defenses

- Rachael Briggs (2009) has given interesting arguments for a modified version of Reflection. She allows for a moderate amount of self-doubt and seeks to see what sort of principle can be justified in light of that.

- Christensen’s worry about diachronic consistency

- Titelbaum’s worry about sticking with your plans

- Consider what holds us together. Sarah Moss, Brian Hedden, and “time slice epistemology”.

- Cite Titelbaum Chapter 7. Discuss commitment. Discuss uniqueness.

- Christensen compares diachronic consistency to interpersonal consistency. He says there is no such rational requirement.

- I say we should be more careful — there is a rational requirement not to share your finances with someone if you’re not willing to plan and work together

\(^5\)Note that if we condition on $P_2(A)$ not being in some interval, it is quite possible that the result at the end is inside that interval. It is only because the weighted average of probabilities within an interval is also within that interval that we can draw the relevant conclusion.
to protect those finances. I might prefer the bedroom painted blue while my partner prefers the bedroom painted red. That means that in some sense, I would be willing to give up some money in order to get the bedroom painted blue, and my partner would be willing to give up some money in order to get the bedroom painted red. However, this preference doesn’t translate into direct action the way we might naively assume. If we both had this willingness, then of course we could have money extracted from us by a clever painter.

Are we just lucky that no enterprising painter is quick and efficient enough to paint the bedroom cheaply enough to get each of us to pay in sequence? No! Part of sharing finances is sharing a commitment to one another not to make purchases that might specifically displease the other. Perhaps better, we should have a commitment to use the financial resources of the shared account in a way that represents the beliefs and desires of a fictitious joint entity. In most ordinary cases our beliefs are similar enough that we don’t have to worry about that, and our desires don’t conflict, so we can use our individual desires as proxies for the desires of the pair. But in making decisions that affect the operations of the household, or in betting on things that we know we disagree about, we need to be careful. We may in these sorts of cases agree to let the beliefs and desires of one of us win out for particular topics. Or we may have some compromise rule. But when acting with the joint account, we should do something of the sort, unless we have completely lost our commitment to work together.

I claim that something similar is at work diachronically. I have reason now to commit to certain plans for updating my beliefs in the future. This commitment may be as a result of accuracy considerations showing that this plan for update is the one that has the best expected accuracy over all possible futures. Or it may be because I see that I and my shared future will suffer a combined loss if I don’t update in this way. However, if my future self has a radical change of heart, he will not be irrational to violate my current plan for updating, any more than my partner must be irrational if he has a change of heart and spends our joint account to run away to Chile. There is irrationality only if the person involved intends to maintain the commitment while acting in this way. If the commitment is broken, there may be a kind of tragedy in that things won’t likely work out well for the people involved. But this tragedy is not irrationality.

This may be for the next chapter:

Given a set of beliefs, we have rules about what you should prefer (which include things of higher expected value). Thus, given a set of degrees of belief, if you have a proper scoring rule, then the degrees of belief you should prefer are the ones that you have. This is a defense of a sort of epistemic conservatism. However, it isn’t terribly strong. You should also prefer a strategy that changes your degrees of belief in light of possible signals that you can recognize, which we might call a sort of responsiveness to evidence.

And if at some point you actually do change your degrees of belief, then afterwards we can say that you should prefer the result, regardless of whether or not it was what you should have preferred beforehand. You might not cohere if you have ended up with something that you didn’t prefer beforehand but you
prefer it now. But whatever sort of failing this is, there’s no time at which it is an epistemic failing. Perhaps it’s a diachronic failure of consistency.

References


