Notes on Bayesian Confirmation Theory

Michael Strevens

June 2017

Contents

List of Figures

List of Tech Boxes

There were three ravens sat on a tree,
Downe a downe, hay downe, hay downe
There were three ravens sat on a tree,
With a downe
There were three ravens sat on a tree,
They were as blacke as they might be.
With a downe derrie, derrie, derrie, downe, downe.

Anonymous, 16th century

1. Introduction

Bayesian confirmation theory—abbreviated to BCT in these notes—is the predominant approach to confirmation in late twentieth century philosophy of science. It has many critics, but no rival theory can claim anything like the same following. The popularity of the Bayesian approach is due to its flexibility, its apparently effortless handling of various technical problems, the existence of various a priori arguments for its validity, and its injection of subjective and contextual elements into the process of confirmation in just the places where critics of earlier approaches had come to think that subjectivity and sensitivity to context were necessary.

There are three basic elements to BCT. First, it is assumed that the scientist assigns what we will call *credences* or *subjective probabilities* to different competing hypotheses. These credences are numbers between zero and one reflecting something like the scientist's level of expectation that a particular hypothesis will turn out to be true, with a credence of one corresponding to absolute certainty.

Second, the credences are assumed to behave mathematically like probabilities. Thus they can be legitimately called *subjective probabilities* (subjective because they reflect one particular person's views, however rational).

Third, scientists are assumed to learn from the evidence by what is called the Bayesian conditionalization rule. Under suitable assumptions the conditionalization rule directs you to update your credences in the light of new evidence in a quantitatively exact way—that is, it provides precise new credences to replace the old credences that existed before the evidence came in—provided only that you had precise credences for the competing hypotheses before the evidence arrived. That is, as long as you have some particular opinion about how plausible each of a set of competing hypotheses is before you observe any evidence, the conditionalization rule will tell you *exactly* how to update your opinions as more and more evidence arrives.

My approach to BCT is more pragmatic than a priori, and more in the

mode of the philosophy of science than that of epistemology or inductive logic. There is not much emphasis, then, on the considerations, such as the Dutch book argument (see section 3.4), that purport to show that we must all become Bayesians. Bayesianism is offered to the reader as a superior (though far from perfect) choice, rather than as the only alternative to gross stupidity.

This is, I think, the way that most philosophers of science see things—you will find the same tone in Horwich (1982) and Earman (1992)—but you should be warned that it is not the approach of the most prominent Bayesian proselytizers. These latter tend to be strict apriorists, concerned to prove above all that there is no rational alternative to Bayesianism. They would not, on the whole, approve of my methods.

A note to aficionados: Perhaps the most distinctive feature of my approach overall is an emphasis on the need to set subjective likelihoods according to the physical likelihoods, using what is often called Miller's Principle. While Miller's Principle is not itself especially controversial, I depart from the usual Bayesian strategy in assuming that, wherever inductive scientific inference is to proceed, a physical likelihood *must be found*, using auxiliary hypotheses if necessary, to constrain the subjective likelihood.

A note to all readers: some more technical or incidental material is separated from the main text in lovely little boxes. I refer to these as *tech boxes*. Other advanced material, occurring at the end of sections, is separated from what precedes it by a horizontal line, like so.

On a first reading, you should skip this material. Unlike the material in tech boxes, however, it will eventually become relevant.

2. Credence or Subjective Probability

Bayesianism is built on the notion of credence or subjective probability. We will use the term *credence* until we are able to conclude that credences have the mathematical properties of probability; thereafter, we will call credences *subjective probabilities*.

A credence is something like a person's level of expectation for a hypothesis or event: your credence that it will rain tomorrow, for example, is a measure of the degree to which you expect rain. If your credence for rain is very low, you will be surprised if it rains; if it is very high, you will be surprised if it does not rain. Credence, then, is psychological property. Everyone has their own credences for various events.

The Bayesian's first major assumption is that scientists, and other rational creatures, have credences not only for mundane occurrences like rain, but concerning the truth of various scientific hypotheses. If I am very confident about a hypothesis, my credence for that hypothesis is very high. If I am not at all confident, it is low.

The Bayesian's model of a scientist's mind is much richer, then, than the model typically assumed in classical confirmation theory. In the classical model, the scientist can have one of three attitudes towards a theory:[1] they accept the theory, they reject the theory, or they neither accept nor reject it. A theory is accepted once the evidence in its favor is sufficiently strong, and it is rejected once the evidence against it is sufficiently strong; if the evidence is strong in neither way, it is neither accepted nor rejected.

On the Bayesian model, by contrast, a scientist's attitude to a hypothesis is encapsulated in a level of confidence, or credence, that may take any of a range of different values from total disbelief to total belief. Rather than laying down, as does classical confirmation theory, a set of rules dictating when

1. The classical theorist does not necessarily deny the existence of a richer psychology in individual scientists; what it denies is the relevance of this psychology to questions concerning confirmation.

the evidence is sufficient to accept or reject a theory, BCT lays down a set of rules dictating how an individual's credences should change in response to the evidence.

> *2.1 Bayesian Theories of Acceptance*
> Some Bayesian fellow travellers (for example, Levi 1967) add to the Bayesian infrastructure a set of rules for accepting or rejecting hypotheses. The idea is that, once you have decided on your credences over the range of available hypotheses, you then have another decision to make, namely, which of those hypotheses, if any, to accept or reject based on your credences. The conventional Bayesian reaction to this sort of theory is that the second decision is unnecessary: your credences express all the relevant facts about your epistemic commitments.

In order to establish credence as a solid foundation on which to build a theory of confirmation, the Bayesian must, first, provide a formal mathematical apparatus for manipulating credences, and second, provide a material basis for the notion of credence, that is, an argument that credences are psychologically real.

The formal apparatus comes very easily. Credences are asserted to be, as the term *subjective probability* suggests, a kind of probability. That is, they are real numbers between zero and one, with a credence of one for a theory meaning that the scientist is practically certain that the theory is true, and a credence of zero meaning that the scientist is practically certain that the theory is false. (The difference between practical certainty and absolute certainty is explained in section 6.3.) Declaring credences to be probabilities gives BCT much of its power: the mathematical properties of probabilities turn out to be very apt for representing the relation between theory and evidence.

The psychological reality of credences presents more serious problems for the Bayesian. While no one denies the existence of levels of expectation

2.2 What Do Credences Range Over?

In probability mathematics, probabilities may be attached either to events, such as the event of its raining tomorrow, or to propositions, such as the proposition "It will rain tomorrow". It is more natural to think of a theory as a set of propositions than as an "event", for which reason BCT is usually presented as a theory in which probabilities range over propositions. My formal presentation respects this custom, but the commentary uses the notion of event wherever it seems natural.

for events such as tomorrow's rain, what can reasonably be denied is the existence of a complete set of precisely specified numbers characterizing a level of expectation for all the various events and theories that play a role in the scientific process.

The original response to this skeptical attitude was developed by Frank Ramsey (1931), who suggested that credences are closely connected to dispositions to make or accept certain bets. For example, if my credence for rain tomorrow is 0.5, I will accept anything up to an even money bet on rain tomorrow. Suppose we decide that, if it rains, you pay me $10, while if it does not rain, I pay you $5. I will eagerly accept this bet. If I have to pay you $10, so that we are both putting up the same amount of money, I will be indifferent to the bet; I may accept it or I may not. If I have to pay you $15, I will certainly not make the bet. (Some of the formal principles behind Ramsey's definition will be laid out more precisely in section 3.4.)

Ramsey's argued that betting patterns are sufficiently widespread—since humans can bet on anything—and sufficiently consistent, to underwrite the existence of credences for all important propositions; one of his major contributions to the topic was to show that only very weak assumptions need be made to achieve the desired level of consistency.

What, exactly, is the nature of the connection between credences and betting behavior? The simplest and cleanest answer is to define credences

in terms of betting behavior, so that, for example, your having a credence of one half for a proposition is no more or less than your being prepared to accept anything up to even odds on the proposition's turning out to be true.

Many philosophers, however, resist such a definition. They worry, for example, about the possibility that an aversion to gambling may distort the relation between a person's credences and their betting behavior. The idea underlying this and other such concerns is that credences are not dispositions to bet, but are rather psychological properties in their own right that are intimately, but not indefeasibly, connected to betting behavior (and, one might add, to felt levels of expectation). Ramsey himself held such a view.

This picture strikes me as being a satisfactory basis for modern Bayesian confirmation theory (though some Bayesian apriorists—see below—would likely disagree). Psychologists may one day tell us that there are no credences, or at least not enough for the Bayesian; for the sake of these notes on BCT, though, let me assume that we have all the credences that BCT requires.

3. Axioms of Probability

3.1 The Axioms

The branch of mathematics that deals with the properties of probabilities is called the *probability calculus*. The calculus posits certain *axioms* that state properties asserted to be both necessary and sufficient for a set of quantities to count as probabilities. (Mathematicians normally think of the axioms as constituting a kind of definition of the notion of probability.)

It is very important to the workings of BCT that credences count as probabilities in this mathematical sense, that is, that they satisfy all the axioms of the probability calculus. This section will spell out the content of the axioms; section 3.4 asks why it is reasonable to think that the psychological entities we are calling credences have the necessary mathematical properties.

Begin with an example, a typical statement about a probability, the claim

that the probability of obtaining heads on a tossed coin is one half. You may think of this as a credence if you like; for the purposes of this section it does not matter what sort of probability it is.

The claim about the coin involves two elements: an outcome, heads, and a corresponding number, 0.5. It is natural to think of the probabilistic facts about the coin toss as mapping possible outcomes of the toss to probabilities. These facts, then, would be expressed by a simple function $P(\cdot)$ defined for just two outcomes, *heads* and *tails*:

$$P(\textit{heads}) = 0.5;$$
$$P(\textit{tails}) = 0.5.$$

This is indeed just how mathematicians think of probabilistic information: they see it as encoded in a function mapping outcomes to numbers that are the probabilities of those outcomes. The mathematics of probability takes as basic, then, two entities: a set of outcomes, and a function mapping the elements of that set to probabilities. The set of outcomes is sometimes called the *outcome space*; the whole thing the *probability space*.

Given such a structure, the axioms of the probability calculus can then be expressed as constraints on the probability function. There are just three axioms, which I will first state rather informally.

1. The probability function must map every outcome to a real number between zero and one.

2. The probability function must map an inevitable outcome (e.g., getting a number less than seven on a toss of a single die) to one, and an impossible outcome (e.g., getting a seven on a toss of single die) to zero.

3. If two outcomes are mutually exclusive, meaning that they cannot both occur at the same time, then the probability of obtaining either

one or the other is equal to the sum of the probabilities for the two outcomes occurring separately. For example, since heads and tails are mutually exclusive—you cannot get both heads *and* tails on a single coin toss—the probability of getting either heads or tails is the sum of the probability of heads and the probability of tails, that is, $0.5 + 0.5 = 1$, as you would expect.

On the most conservative versions of the probability calculus, these are the only constraints placed on the probability function. As we will see, a surprising number of properties can be derived from these three simple axioms.

Note that axiom 3 assumes that the probability function ranges over combinations of outcomes as well as individual outcomes. For example, it is assumed that there is a probability not just for heads and for tails, but for the outcome *heads or tails*. A formal statement of the axioms makes explicit just what combinations of outcomes must have probabilities assigned to them. For our purposes, it is enough to know that any simple combination of outcomes is allowed. For example, if the basic outcomes for a die throw are the first six integers, then a probability must be assigned to outcomes such as *either an even number other than six, or a five* (an outcome that occurs if the die shows two, four, or five). Note that we are allowed to refer to general properties of the outcomes (e.g., being even), and to use the usual logical connectives.

It is useful to have a shorthand for these complex outcomes. Let e and d be two possible outcomes of a die throw. Say that e is the event of getting an odd number, and d is the event of getting a number less than three. Then by ed, I mean the event of both e and d occurring (i.e., getting a one), by $e \vee d$ I mean the event of either e or d occurring (i.e., getting one of 1, 2, 3, or 5), and by $\neg e$ I mean the event of e's not occurring (i.e., getting an even number).

Using this new formalism, let me write out the axioms of the probability calculus more formally. Note that in this new version, the axioms appear to

3.1 Sigma Algebra

The main text's characterization of the domain of outcomes over which the probability function must be defined is far too vague for the formal needs of mathematics. Given a set of basic outcomes (which are themselves subsets of an even more basic set, though in simple cases such as a die throw, you may think of them as atomic elements), mathematicians require that the probability function be defined over what they call a *sigma algebra* formed from the basic set. The sigma algebra is composed by taking the closure of the basic set under the set operations of union (infinitely many, though not uncountably many, are allowed), intersection, and complement. The outcome corresponding to the union of two other outcomes is deemed to occur if either outcome occurs; the outcome corresponding to the intersection is deemed to occur if both outcomes occur; and the outcome corresponding to the complement of another outcome is deemed to occur if the latter outcome fails to occur.

contain less information than in the version above. For example, the axioms do not require explicitly that probabilities are less than one. It is easy to use the new axioms to derive the old ones, however; in other words, the extra information is there, but it is implicit. Here are the axioms.

1. For every outcome e, $P(e) \geq 0$.

2. For any inevitable outcome e, $P(e) = 1$.

3. For mutually exclusive outcomes e and d, $P(e \vee d) = P(e) + P(d)$.

The notions of inevitability and mutual exclusivity are typically given a formal interpretation: an outcome is inevitable if it is logically necessary that it occur, and two outcomes are mutually exclusive if it is logically impossible that they both occur.

Now you should use the axioms to prove the following simple theorems of the probability calculus:

3.2 The Axiom of Countable Additivity

Most mathematicians stipulate that axiom 3 should apply to combinations of denumerably many mutually exclusive outcomes. (A set is denumerable if it is infinite but countable.) This additional stipulation is called the *axiom of countable additivity*. Some other mathematicians, and philosophers, concerned to pare the axioms of the probability calculus to the weakest possible set, do their best to argue that the axiom of countable additivity is not necessary for proving any important results.

1. For every outcome e, $P(e) + P(\neg e) = 1$.
2. For every outcome e, $P(e) \leq 1$.
3. For any two logically equivalent propositions e and d, $P(e) = P(d)$. (You might skip this proof the first time through; you will, however, need to use the theorem in the remaining proofs.)
4. For any two outcomes e and d, $P(e) = P(ed) + P(e\neg d)$.
5. For any two outcomes e and d such that e entails d, $P(e) \leq P(d)$.
6. For any two outcomes e and d such that $P(e \supset d) = 1$ (where $\supset$ is material implication), $P(e) \leq P(d)$. (Remember that $e \supset d \equiv d \vee \neg e \equiv \neg(e\neg d)$.)

Having trouble? The main step in all of these proofs is the invocation of axiom 3, the only axiom that relates the probabilities for two different outcomes. In order to invoke axiom 3 in the more complicated proofs, you will need to break down the possible outcomes into mutually exclusive parts. For example, when you are dealing with two events e and d, take a look at the probabilities of the four mutually exclusive events ed, $e\neg d$, $d\neg e$, and $\neg e\neg d$, one of which must occur. When you are done, compare your proofs with those at the end of these notes.

3.2 Conditional Probability

We now make two important additions to the probability calculus. These additions are conceptual rather than substantive: it is not new axioms that are introduced, but new definitions.

The first definition is an attempt to capture the notion of a *conditional probability*, that is, a probability of some outcome conditional on some other outcome's occurrence. For example, I may ask: what is the probability of obtaining a two on a die roll, given that the number shown on the die is even? What is the probability of *Mariner* winning tomorrow's race, given that it rains tonight? What is the probability that the third of three coin tosses landed heads, given that two of the three were tails?

The conditional probability of an outcome e given another outcome d is written $P(e|d)$. Conditional probabilities are introduced into the probability calculus by way of the following definition:

$$P(e|d) = \frac{P(ed)}{P(d)}.$$

(If $P(d)$ is zero, then $P(e|d)$ is undefined.) In the case of the die throw above, for example, the probability of a two given that the outcome is even is, according to the definition, the probability of obtaining a two *and* an even number (i.e., the probability of obtaining a two) divided by the probability of obtaining an even number, that is, 1/6 divided by 1/2, or 1/3, as you might expect.

The definition can be given the following informal justification. To determine $P(e|d)$, you ought, intuitively, to reason as follows. Restrict your view to the possible worlds in which the outcome d occurs. Imagine that these are the only possibilities. Then the probability of e conditional on d is the probability of e in this imaginary, restricted universe. What you are calculating, if you think about it, is the proportion, probabilistically weighted, of the probability space corresponding to d that also corresponds to e.

Conditional probabilities play an essential role in BCT, due to their ap-

3.3 Conditional Probability Introduced Axiomatically
There are some good arguments for introducing the notion of conditional probability as a primitive of the probability calculus rather than by way of a definition. On this view, the erstwhile definition, or something like it, is to be interpreted as a fourth axiom of the calculus that acts as a constraint on conditional probabilities $P(e|d)$ in those cases where $P(d)$ is non-zero. When $P(d)$ is zero, the constraint does not apply. One advantage of the axiomatic approach is that it allows the mathematical treatment of probabilities conditional on events that have either zero probability or an undefined probability.

pearance in two important theorems of which BCT makes extensive use. The first of these theorems is Bayes' theorem:

$$P(e|d) = \frac{P(d|e)}{P(d)}P(e).$$

You do not need to understand the philosophical significance of the theorem yet, but you should be able to prove it. Notice that it follows from the definition of conditional probability alone; you do not need *any* of the axioms to prove it. In this sense, it is hardly correct to call it a theorem at all. All the more reason to marvel at the magic it will work...

The second important theorem states that, for an outcome e and a set of mutually exclusive, exhaustive outcomes $d_1, d_2, \ldots$ that

$$P(e) = P(e|d_1)P(d_1) + P(e|d_2)P(d_2) + \cdots$$

This is a version of what is called the *total probability theorem*. A set of outcomes is exhaustive if at least one of the outcomes must occur. It is mutually exclusive if at most one can occur. Thus, if a set of outcomes is mutually exclusive and exhaustive, it is guaranteed that exactly one outcome in the set will occur.

To prove the total probability theorem, you will need the axioms, and also the theorem that, if $P(k) = 1$, then $P(ek) = P(e)$. First show that

$P(e) = P(ed_1) + P(ed_2) + \cdots$ (this result is itself sometimes called the theorem of total probability). Then use the definition of conditional probability to obtain the theorem. You can make life a bit easier for yourself if you first notice that that axiom 3, which on the surface applies to disjunctions of just two propositions, in fact entails an analogous result for any finite number of propositions. That is, if propositions $e_1, \ldots, e_n$ are mutually exclusive, then axiom 3 implies that $P(e_1 \vee \cdots \vee e_n) = P(e_1) + \cdots + P(e_n)$.

3.3 Probabilistic Independence

Two outcomes e and d are said to be *probabilistically independent* if

$$P(ed) = P(e)P(d).$$

The outcomes of distinct coin tosses are independent, for example, because the probability of getting, say, two heads in a row, is equal to the probability for heads squared.

Independence may also be characterized using the notion of conditional probability: outcomes e and d are independent if $P(e|d) = P(e)$. This characterization, while useful, has two defects. First, it does not apply when the probability of d is zero. Thus it is strictly speaking only a sufficient condition for independence; however, it is necessary and sufficient in all the interesting cases, that is, the cases in which neither probability is zero or one, which is why it is useful all the same. Second, it does not make transparent the symmetry of the independence relation: e is probabilistically independent of d just in case d is probabilistically independent of e. (Of course, if you happen to notice that, for non-zero $P(e)$ and $P(d)$, $P(e|d) = P(e)$ just in case $P(d|e) = P(d)$, then the symmetry can be divined just below the surface.)

In probability mathematics, independence normally appears as an assumption. It is assumed that some set of outcomes is independent, and some other result is shown to follow. For example, you might show (go

ahead) that, if e and d are independent, then

$$P(e) + P(d) = P(e \vee d) + P(e)P(d).$$

(Hint: start out by showing, without invoking the independence assumption, that $P(e) + P(d) = P(e \vee d) + P(ed)$.)

In applying these results to real world problems, it becomes very important to know when a pair of outcomes can be safely assumed to be independent. An often used rule of thumb assures us that outcomes produced by causally independent processes are probabilistically independent. (Note that the word *independent* appears twice in the statement of the rule, meaning two rather different things: probabilistic independence is a mathematical relation, relative to a probability function, whereas causal independence is a physical or metaphysical relation.) The rule is very useful; however, in many sciences, for example, kinetic theory and population genetics, outcomes are assumed to be independent even though they are produced by processes that are not causally independent. For an explanation of why these outcomes nevertheless tend to be probabilistically independent, run, don't walk, to the nearest bookstore to get yourself a copy of Strevens (2003).

In problems concerning confirmation, the probabilistic independence relation almost never holds between outcomes of interest, for reasons that I will explain later. Thus, the notion of independence is not so important to BCT, though we have certainly not heard the last of it.

3.4 Justifying the Axioms

We have seen that calling credence a species of mathematical probability is not just a matter of naming: it imputes to credences certain mathematical properties that are crucial to the functioning of the Bayesian machinery. We have, so far, identified credences as psychological properties. We have not shown that they have any particular mathematical properties. Or rather—since the aim of confirmation theory is more to prescribe than to describe—

we have not shown that credences *ought to have* any particular mathematical properties, that is, that people ought to ensure that their credences conform to the axioms of the probability calculus.

To put things more formally, what we want to do is to show that the credence function—the function $C(\cdot)$ giving a person's credence for any particular hypothesis or event—has all the properties specified for a generic probability function $P(\cdot)$ above. If we succeed, we have shown that $C(\cdot)$ is, or rather ought to be, a probability function; it will follow that everything we have proved for $P(\cdot)$ will be true for $C(\cdot)$.

This issue is especially important to those Bayesians who wish to establish a priori the validity of the Bayesian method. They would like to *prove* that credences should obey the axioms of the probability calculus. For this reason, a prominent strand in the Bayesian literature revolves around attempts to argue that it is irrational to allow your credences to violate the axioms.

The best known argument for this conclusion is known as the *Dutch book* argument. (The relevant mathematical results were motivated and proved independently by Ramsey and de Finetti, neither a Dutchman.) Recall that there is a strong relation, on the Bayesian view, between your credence for an event and your willingness to bet for or against the occurrence of the event in various circumstances. A Dutch book argument establishes that, if your credences do not conform to the axioms, it is possible to concoct a series of gambles that you will accept, yet which is sure to lead to a net loss, however things turn out. (Such a series is called a Dutch book.) To put yourself into a state of mind in which you are disposed to make a series of bets that must lose money is irrational; therefore, to fail to follow the axioms of probability is irrational.

The Dutch book argument assumes the strong connection between credence and betting behavior mentioned in section 2. Let me now specify exactly what the connection is supposed to be.

If your credence for an outcome e is p, then you should accept odds of up to $p : (1 - p)$ to bet on e, and odds of up to $(1 - p) : p$ to bet against e. To accept odds of $a : b$ on e is to accept a bet in which you put an amount proportional to a into the pot, and your opponent puts an amount proportional to b into the pot, on the understanding that, if e occurs, you take the entire pot, while if e does not occur, your opponent takes the pot. The important fact about the odds, note, is the ratio of a to b: the odds $1 : 1.5$, the odds $2 : 3$ and the odds $4 : 6$ are exactly the same odds. Consider some examples of the credence/betting relation.

1. Suppose that your credence for an event e is 0.5, as it might be if, say, e is the event of a tossed coin's landing heads. Then you will accept odds of up to $1 : 1$ (the same as $0.5 : 0.5$) to bet on e. If your opponent puts, say, \$10 into the pot, you will accept a bet that involves your putting any amount of money up to \$10 in the pot yourself, but not more than \$10. (This is the example I used in section 2.)

2. Suppose that your credence for e is 0.8. Then you will accept odds of up to $4 : 1$ (the same as $0.8 : 0.2$) to bet on e. If your opponent puts \$10 into the pot, you will accept a bet that involves your putting any amount of money up to \$40 in the pot yourself.

3. Suppose that your credence for e is 1. Then you will accept any odds on e. Even if you have to put a million dollars in the pot and your opponent puts in only one dollar, you will take the bet. Why not? You are sure that you will win a dollar. If your credence for e is 0, by contrast, you will never bet on e, no matter how favorable the odds.

I will not present the complete Dutch book argument here, but to give you the flavor of the thing, here is the portion of the argument that shows how to make a Dutch book against someone whose credences violate axiom 2. Such a person has a credence for an inevitable event e that is less than 1, say, 0.9. They are therefore prepared to bet against e at odds of $1 : 9$ or

better. But they are sure to lose such a bet. Moral: assign probability one to inevitable events at all times.

It is worth noting that the Dutch book argument says nothing about conditional probabilities. This is because conditional probabilities do not appear in the axioms; they were introduced by definition. Consequently, any step in mathematical reasoning about credences that involves only the definition of conditional probabilities need not be justified; not to take the step would be to reject the definition. Interestingly enough, the mathematical result about probability that has the greatest significance for BCT—Bayes' theorem—invokes only the definition. Thus the Dutch book argument is not needed to justify Bayes' theorem!

The Dutch book argument has been subjected to a number of criticisms, of which I will mention two. The first objection questions the very strong connection between credence and betting behavior required by the argument. As I noted in section 2, the tactic of defining credence so as to establish the connection as a matter of definition has fallen out of favor, but a connection that is any weaker seems to result in a conclusion, not that the violator of the axioms is guaranteed to accept a Dutch book, but that they have a tendency, all other things being equal, in the right circumstances, to accept a Dutch book. That is good enough for me, but it is not good enough for many aprioristic Bayesians.

The second objection to the Dutch book argument is that it seeks to establish too much. No one can be blamed for failing in some ways to arrange their credences in accordance with the axioms. Consider, for example, the second axiom. In order to follow the axiom, you would have to know which outcomes are inevitable. The axiom is normally interpreted fairly narrowly, so that an outcome is regarded as inevitable only if its non-occurrence is a *conceptual* impossibility (as opposed to, say, a physical impossibility). But even so, conforming to the axiom would involve your being aware of all the conceptual possibilities, which means, among other things, being aware of

all the theorems of logic. If only we could have such knowledge! The implausibility of Bayesianism's assumption that we are aware of all the conceptual possibilities, or as is sometimes said, that we are *logically omniscient*, will be a theme of the discussion of the problem of old evidence in section 11.

Bayesians have offered a number of modifications of and alternatives to the Dutch book argument. All are attempts to establish the irrationality of violating the probability axioms. All, then, are affected by the second, logical omniscience objection; but each hopes in its own way to accommodate a weaker link between credence and subjective probability, and so to avoid at least the first objection.

Enough. Let us from this point on assume that a scientist's credences tend to, or ought to, behave in accordance with the axioms of probability. For this reason, I will now call credences, as promised, *subjective probabilities*.

4. Bayesian Conditionalization

4.1 Bayes' Rule

We have now gone as far in the direction of BCT as the axioms of probability can take us. The final step is to introduce the Bayesian conditionalization rule, a rule that, however intuitive, does not follow from any purely mathematical precepts about the nature of probability.

Suppose that your probability for rain tomorrow, conditional on a sudden drop in temperature tonight, is 0.8, whereas your probability for rain given no temperature drop is 0.3. The temperature drops. What should be your new subjective probability for rain? It seems intuitively obvious that it ought to be 0.8.

The Bayesian conditionalization rule simply formalizes this intuition. It dictates that, if your subjective probability for some outcome d conditional on another outcome e is p, and if you learn that e has in fact occurred (and

you do not learn anything else), you should set your unconditional subjective probability for d, that is, $C(d)$, equal to p.

Bayes' rule, then, relates subjective probabilities at two different times, an earlier time when either e has not occurred or you do not know that e has occurred, and a later time when you learn that e has indeed occurred. To write down the rule formally, we need a notation that distinguishes a person's subjective probability distribution at two different times. I write a subjective probability at the earlier time as $C(\cdot)$, and a subjective probability at the later time as $C^+(\cdot)$. Then Bayes' rule for conditionalization can be written:

$$C^+(d) = C(d|e),$$

on the understanding that the sole piece of information learned in the interval between the two times is that e has occurred. More generally, if $e_1 \ldots e_n$ are all the pieces of information learned between the two times, then Bayes' rule takes the form

$$C^+(d) = C(d|e_1 \ldots e_n).$$

If you think of what is learned, that is $e_1 \ldots e_n$, as the *evidence*, then Bayes' rule tells you how to update your beliefs in the light of the evidence, and thus constitutes a theory of confirmation. Before I move on to the application of Bayes' rule to confirmation theory in section 5, however, I have a number of important observations to make about conditionalization in itself.

4.2 Observation

Let me begin by saying some more about what a Bayesian considers to be the kind of event that prompts the application of Bayes' rule. I have said that a Bayesian conditionalizes on e—that is, applies Bayes' rule to e—just when they "learn that e has occurred". In classical Bayesianism, to learn e is to have one's subjective probability for e go to one as the result of some kind of observation. This observation-driven change of e's probability is

not, note, due to an application of Bayes' rule. It is, as it were, prompted by a perception, not an inference.

The Bayesian, then, postulates two mechanisms by means of which subjective probabilities may justifiably change:

1. An observation process, which has the effect of sending the subjective probability for some observable state of affairs (or if you prefer, of some observation sentence) to one. The process is not itself a reasoning process, and affects only individual subjective probabilities.

2. A reasoning process, governed by Bayes' rule. The reasoning process is reactive, in that it must be triggered by a probability change due to some other process; normally, the only such process envisaged by the Bayesian is observation.

That the Bayesian relies on observation to provide the impetus to Bayesian conditionalization prompts two questions. First, what if observation raises the probability of some e, but not all the way to one? Second, what kind of justification can be given for our relying on the probability changes induced by observation?

To the first question, there is a standard answer. If observation changes the credence for some e to a value x not equal to one, use the following rule instead of Bayes' rule:

$$C^+(d) = C(d|e)x + C(d|\neg e)(1 - x).$$

You will see that this rule is equivalent to Bayes' rule in the case where x is one. The more general rule is called *Jeffrey conditionalization*, after Jeffrey (1983).

The second question, concerning the justification of our reliance on observation, is not provided with any special answer by BCT. Indeed, philosophers of science typically leave this question to the epistemologists, and take the epistemic status of observation as given.

When I observe e, I am, according to Bayes' rule, to set my new probability for d equal to $C(d|e)$. But $C(d|e)$, it seems, only expresses the relation between e and d in isolation. What if e is, on its own, irrelevant to d, but is highly relevant when other information is taken into account? It seems that I ought to set $C^+(d)$ equal not to $C(d|e)$, but to $C(d|ek)$, where k is all my background knowledge.

Let me give an example. Suppose that d is the proposition "The room contains at least two philosophy professors" and e is the proposition "Professor Wittgenstein is in the room". Then $C(d|e)$ should be, it seems, moderately large, or at least, greater than $C(d)$. But suppose that I know independently that Professor Wittgenstein despises other philosophers and will leave the room immediately if another philosopher enters. The conditional probability that takes into account this background knowledge, $C(d|ek)$, will then be close to zero. Clearly, upon seeing Professor Wittgenstein in the room, I should take my background knowledge into account, setting $C(d)$ equal to this latter probability. Thus Bayes' rule must incorporate k.

In fact, although there is no harm in incorporating background knowledge explicitly into Bayes' rule, it is not necessary. The reason is that any relevant background knowledge is already figured into the subjective probability $C(d|e)$; in other words, at all times, $C(d|e) = C(d|ek)$.[2] This follows from the assumption that we assign our background knowledge subjective probability one and the following theorem of the probability calculus:

If $P(k) = 1$, then $P(d|ek) = P(d|e)$.

which follows in turn from another theorem: if $P(k) = 1$, then $P(ek) = P(e)$.

I misled you in the example above by suggesting that $C(d|e)$ is moderately large. In fact, it is equal to $C(d|ek)$ and therefore close to zero. Precisely

2. This is only true of subjective probability, not of other varieties of probability you may come across, such as physical probability and logical probability.

because it is easy to be misled in this way, however, it is in some circumstances worth putting the background knowledge explicitly into Bayes' rule, just to remind yourself and others that it is always there regardless.

4.4 Justifying Bayes' Rule

Bayes' rule does not follow from the axioms of the probability calculus. You can see this at a glance by noting that the rule relates two different probability functions, $C(\cdot)$ and $C^{+}(\cdot)$, whereas the axioms concern only a single function. Less formally, the axioms put a constraint on the form of the assignment of subjective probabilities at a particular time, whereas Bayes' rule dictates a relation between subjective probability assignments at two different times.

To get a better feel for this claim, imagine that we have a number of cognitively diverse people whose subjective probabilities obey, at all times, the axioms of the probability calculus, and who conditionalize according to Bayes' rule. Construct a kind of mental Frankenstein's monster, by cutting each persons stream of thoughts into pieces, and stitching them together haphazardly. At any particular moment in the hybrid stream, the subjective probability assignments will obey the calculus, because they belong to the mind of some individual, who by assumption obeys the calculus at all times. But the stream will not obey Bayes' rule, since the value of $C(d|e)$ at one time may belong to a different person than the value of $C(d)$ at a later time. Indeed, there is no coherence at all to the hybrid stream; the moral is that a stream of thoughts that at all times satisfies the probability calculus can be as messed up as you like when examined for consistency through time.

To justify Bayes' rule, then, you need to posit some kind of connection between a person's thoughts at different times. A number of Bayesians have tried to find a connection secure enough to participate as a premise in an a priori argument for Bayes' rule. One suggestion, due to David Lewis, is to postulate a connection between subjective probabilities at one time and

4.1 What the Apriorist Must Do

Let me summarize the things that the Bayesian apriorist must prove in order to establish the apparatus of BCT as compulsory for any rational being. It must be shown that:

1. Everyone has, or ought to have, subjective probabilities for all the elements that play a part in scientific confirmation, in particular, hypotheses and outcomes that would count as evidence.

2. The subjective probabilities ought to conform to the axioms of the probability calculus.

3. The subjective probabilities ought to change in accordance with Bayes' rule. (The exceptions are probability changes due to observation; see section 4.2.)

The "old-fashioned" apriorist tries to get (1) and (2) together by defining subjective probability in terms of betting behavior. In note 3 I suggested trying to get (2) and (3) with a stricter definition of subjective probability; the drawback to this is that the definition, being much more stringent, would no longer be satisfied by a person's betting behavior under the rather weak conditions shown by Ramsey to be sufficient for the existence of subjective probabilities on the "old-fashioned" definition. Thus a definition in terms of extended betting behavior would weaken the argument for the existence of subjective probabilities, to some extent undermining (1).

Two other conclusions that we have not encountered yet may also be the object of the apriorist's desire:

4. Subjective probabilities ought to conform to the probability coordination principle (see section 5.2). To show this is not compulsory, but it is highly desirable.

5. Initial probabilities over hypotheses ought to conform to some kind of symmetry principle (see section 9.4). To show this is not compulsory; many would regard it as completely unnecessary.

betting behavior at a later time, and to run a kind of Dutch book argument.

The connection in question is just what you are thinking: if a person has subjective probability p for d conditional on e, and if e (and nothing else) is observed at some later time, then at that later time, the person should accept odds of up to $p : 1 - p$ on d. Note that this is not the connection between subjective probability and betting behavior used in section 3.4 to run the conventional Dutch book argument. That connection relates subjective probabilities at a time and betting behavior at the same point in time; the new principle relates subjective probabilities at a time and betting behavior at a strictly later time, once new evidence has come in.

This raises a problem for an old-fashioned apriorist. The old-fashioned apriorist justifies the old connection between subjective probability and betting behavior by defining one in terms of the other. But the definitional maneuver cannot be used twice. Once necessary and sufficient conditions are given for having a subjective probability at time t in terms of betting behavior at time t, additional necessary conditions cannot be given in terms of betting behavior at time t'.[3]

There are other approaches to justifying Bayes' rule, but let me move on.

5. The Machinery of Modern Bayesianism

5.1 From Conditionalization to Confirmation

In two short steps we will turn Bayesianism into a theory of confirmation. The first step—really just a change of attitude—is to focus on the use of Bayes' rule to update subjective probabilities concerning scientific hypotheses. To mark this newfound focus I will from this point on write Bayes' rule

3. Exercise to the reader: what are the pitfalls of defining subjective probability so that having a certain subjective probability entails both present *and* future betting behavior? The answer is in tech box 4.1.

as follows:

$$C^+(h) = C(h|e).$$

The rule tells you what your new subjective probability for a hypothesis h should be, upon observation of the evidence e.

In this context, it is no longer natural to call the argument h of the probability function an "outcome" (though if you insist, you can think of the outcome as *the hypothesis's being true*). It is far more natural to think of it as a proposition. At the same time, it is more natural to think of most kinds of evidence as events; consider, for example, the event of the litmus paper's turning blue. For this reason, most expositors of BCT talk apparently inconsistently as if subjective probability functions range over both propositions and events. There is no harm in this custom (see tech box 2.2), and I will follow it with relish.

The second step in the transformation of Bayesianism into a theory of confirmation is, I think, the maximally revelatory moment in all of BCT. So far, Bayes' rule does not appear to be especially useful. It says that you should, upon observing e, set your new probability for h equal to your old probability for h conditional on e. But what ought that old probability to be? It seems that there are very few constraints on a probability such as $C(h|e)$, and so that Bayes' rule is not giving you terribly helpful advice. A skeptic might even suggest reading the rule backwards: $C(h|e)$ is just the probability that you would assign to h, if you were to learn that e. (But this would be a mistake: see tech box 5.1.)

The power of BCT consists in its ability to tell you, these appearances to the contrary, what your value for $C(h|e)$ should be, given only information about your probabilities for h and its competitors, that is, given only values of the form $C(h)$. It will take some time (the remainder of this section) to explain exactly how this is done.

I have promised you a revelatory moment, and now it is time to deliver. Bayes' rule sets $C^+(h)$ equal to $C(h|e)$. We encountered, in section 3.2, a

5.1 Conditional Probability Characterized Dispositionally

There are some interesting counterexamples to the view that $C(h|e)$ is the probability that I would assign to h, if I were to learn that e (due to Richmond Thomason). Suppose that h is the proposition that I am philosophically without talent, and e is some piece of (purely hypothetical!) evidence that incontrovertibly shows that I am philosophically untalented. Then my $C(h|e)$ is very high. But I may be vain enough that, were I to learn that e, I would resist the conclusion that h. Of course, I would be violating Bayes' rule, and (Bayesians would say "therefore") I would be reasoning irrationally. But the scenario is quite possible—plausible, even—and shows that human psychology is such that a dispositional interpretation of conditional probability in these terms is not realistic. The example does not, of course, rule out the possibility, mentioned above, of defining $C(h|e)$ as the probability I *ought to have* for h on learning that e.

result that I called Bayes' theorem:

$$C(d|e) = \frac{C(e|d)}{C(e)}C(d).$$

Writing h instead of d, substitute this into Bayes' rule, obtaining

$$C^+(h) = \frac{C(e|h)}{C(e)}C(h).$$

In this formulation, Bayes' rule is suddenly full of possibilities. I had no idea what value to give to $C(h|e)$, but I know exactly what value to give to $C(e|h)$: it is just the probability that h itself ascribes to a phenomenon such as e. The value of $C(e)$ is perhaps less certain, but for an observable event e, it seems that I am likely to have some opinion or other. (We will see shortly that I have a more definite opinion than I may suppose.) Now, given values for $C(e|h)$ and $C(e)$, I have determined a value for what you might call the *Bayesian multiplier*, the value by which I multiply my old probability for h to arrive at my new probability for h after observing e. What more could you ask from a theory of confirmation?

To better appreciate the virtues of BCT, before I tax you with the details of the same, let me show how BCT deals with a special case, namely, the case in which a hypothesis h entails an observed phenomenon e. I assume that $C(e)$ is less than one, that is, that it was an open question, in the circumstances, whether e would be observed. Because h entails e, $C(e|h)$ is equal to one. (You proved this back in section 3.1.) On the observation of e, then, your old subjective probability for h is multiplied by the Bayesian multiplier

$$\frac{C(e|h)}{C(e)}$$

which is, because $C(e)$ is less than one, greater than one. Thus the observation of e will increase your subjective probability for h, confirming the hypothesis, as you would expect.

This simple calculation, then, has reproduced the central principle of hypothetico-deductivism, and—what HD itself never does—*given an argument* for the principle. (We will later see how Bayesianism avoids some of the pitfalls of HD.) What's more, the argument does not turn on any of the subjective aspects of your assignment of subjective probabilities. All that matters is that $C(e|h)$ is equal to one, an assignment which is forced upon all reasoners by the axioms of probability. You should now be starting to appreciate the miracle of Bayesianism.

5.2 Constraining the Likelihood

You have begun to see, and will later see far more clearly, that the power of BCT lies to a great extent in the fact that it can appeal to certain powerful constraints on the way that we set the subjective probabilities of the form $C(e|h)$.

I will call these probabilities the *subjective likelihoods*. (A likelihood is any probability of the form $P(e|h)$, where h is some hypothesis and e a piece of potential evidence.)

5.2 Logical Probability

Properly constrained subjective probability, or credence, is now the almost universal choice for providing a probabilistic formulation of inductive reasoning. But it was not always so. The first half of the twentieth century saw the ascendancy of what is often called *logical probability* (Keynes 1921; Carnap 1950).

A logical probability relates two propositions, which might be called the hypothesis and the evidence. Like any probability, it is a number between zero and one. Logical probabilists hold that their probabilities quantify the evidential relation between the evidence and the hypothesis, meaning roughly that a logical probability quantifies the degree to which the evidence supports or undermines the hypothesis. When the probability of the hypothesis on the evidence is equal to one, the evidence establishes the truth of the hypothesis for sure. When it is equal to zero, the evidence establishes the falsehood of the hypothesis for sure. When it is between zero and one, the evidence has some lesser effect, positive or negative. (For one particular value, the evidence is irrelevant to the hypothesis. This value is of necessity equal to the probability of the hypothesis on the empty set of evidence. It will differ for different hypotheses.) The degree of support quantified by a logical probability is supposed to be an objective matter—the objective logic in question being, naturally, inductive logic. (The sense of *objective* varies: Carnap's logical probabilities are relative to a "linguistic framework", and so may differ from framework to framework.)

Logical probability has fallen out of favor for the most part because its assumption that there is always a fully determined, objectively correct degree of support between a given body of evidence and a given hypothesis has come to be seen as unrealistic. Yet it should be noted that when BCT is combined with PCP and an objectivist approach to the priors (section 9.4), we are back in a world that is not too different from that of the logical probabilists.

At the end of the last section, I appealed to a very simple constraint on the subjective likelihood $C(e|h)$: if h entails e, then the subjective likelihood must be equal to one. This is a theorem of the probability calculus, and so all who are playing the Bayesian game must conform to the constraint. (Apriorist Bayesians would threaten to inflict Dutch books on non-conformists.)

If the hypothesis in question does not entail the evidence e (and does not entail $\neg e$), however, this constraint does not apply. There are two reasons that h might not make a definitive pronouncement on the occurrence of e. Either

1. The hypothesis concerns physical probabilities of events such as e; it assigns a probability greater than zero but less than one to e, or

2. The hypothesis has nothing definite to say about e at all.

I will discuss the first possibility in this section, and the second in section 6.4.

If h assigns a definite probability to e, then it seems obviously correct to set the subjective likelihood equal to the probability assigned by h, which I will call the *physical likelihood*. For example, if e is the event of a tossed coin's landing heads and h is the hypothesis that tossed coins land heads with a probability of one half, then $C(e|h)$ should also be set to one half. Writing $P_h(e)$ for the probability that h ascribes to e, the rule that we seem tacitly to be following is:

$$C(e|h) = P_h(e).$$

That is, your subjective probability for e conditional on h's turning out to be correct should be the physical probability that h assigns to e. Call this rule the *probability coordination principle*, or PCP.

The principle has an important implication for BCT: if PCP is true, then BCT always favors, relatively speaking, hypotheses that ascribe higher physical probabilities to the observed evidence.

5.3 The Probability Coordination Principle

The rule that subjective likelihoods ought to be set equal to the corresponding physical probabilities is sometimes called Miller's Principle, after David Miller. The name is not particularly apt, since Miller thought that the rule was incoherent (his naming rights are due to his having emphasized its importance in BCT, which he wished to refute). Later, David Lewis (1980) proposed a very influential version of the principle that he called the *Principal Principle*. Lewis later decided (due to issues involving admissibility; see tech box 5.5) that the Principal Principle was false, and ought to be replaced with a new rule of probability coordination called the *New Principle* (Lewis 1994). It seems useful to have a way to talk about all of these principles without favoring any particular one; for this reason, I have introduced the term *probability coordination principle*. In these notes, what I mean by PCP is *whatever probability coordination principle turns out to be correct*.

I cannot emphasize strongly enough that $C(e|h)$ and $P_h(e)$ are two quite different kinds of things. The first is a subjective probability, a psychological fact about a scientist. The second is a physical probability of the sort that might be prescribed by the laws of nature; intuitively, it is a feature of the world that might be present even if there were no sentient beings in the world and so no psychological facts at all. We have beliefs about the values of different physical probabilities; the strength of these beliefs is given by our subjective probabilities.

Because subjective probability and physical probability are such different things, it is an open question why we ought to constrain our subjective probabilities in accordance with PCP. Though Bayesians have offered many proofs that one ought to conform to the constraints imposed by the axioms of probability and Bayes' rule, there has been much less work on PCP.

Before I continue, I ought to tell you two important things about PCP. First, not all Bayesians insist that subjective probabilities conform to PCP:

5.4 Subjectivism about Physical Probability

An important view about the nature of physical probability holds that the probabilities apparently imputed to the physical world by scientific theories are nothing but projections of scientists' own subjective probabilities. This thesis is usually called *subjectivism*, though confusingly, sometimes when people say *subjectivism* they mean *Bayesianism*. The idea that drives subjectivism, due originally to Bruno de Finetti, is that certain subjective probabilities are especially robust, in the sense that conditionalizing on most information does not affect the value of the probabilities. An example might be our subjective probability that a tossed coin lands heads: conditionalizing on almost anything we know about the world (except for some fairly specific information about the initial conditions of the coin toss) will not alter the probability of one half. (Actually, the robustness is more subtle than this; Strevens (2006) provides a more complete picture.) According to the subjectivists, this robustness gives the probabilities an objective aspect that is usually interpreted literally, but mistakenly so. Almost all subjectivists are Bayesians, but Bayesians certainly need not be subjectivists.

Subjectivists do not have to worry about justifying PCP; since subjectivism more or less identifies physical probabilities with certain subjective likelihoods, PCP is trivially true.

the *radical personalists* are quite happy for the subjective likelihoods to be, well, subjective. Radical personalism, which has strong ties to the subjectivist interpretation of probability (see tech box 5.4), will be discussed in section 9.3.

Second, the formulation of PCP stated above, though adequate for the everyday workings of Bayesian confirmation theory, cannot be entirely right. In some circumstances, it is irrational for me to set my subjective likelihood equal to the corresponding physical probability. The physical probability of obtaining heads on a coin toss is one half. But suppose I know the exact velocity and trajectory of a certain tossed coin. Then I can use this information to calculate whether or not it will land heads. Let's say that I figure it will land heads. Then I should set my subjective probability for heads to one, not one half. (Remember that this subjective probability takes into account tacitly all my background knowledge, including my knowledge of the toss's initial conditions and their consequences, as explained in section 4.2.) David Lewis's now classic paper on the probability coordination principle is, in part, an attempt to frame the principle so that it recuses itself when we have information of the sort just described, which Lewis calls *inadmissible information*. You will find more on this problem in tech box 5.5.

The probability coordination principle is a powerful constraint on the subjective likelihoods, but, as I noted above, it seems that it can only brought to bear when the hypotheses under consideration assign specific physical probabilities to the evidence. Some ways around this limitation will be explored in section 6.4.

5.3 Constraining the Probability of the Evidence

How should you set a value for $C(e)$, the probability of observing a particular piece of evidence? A very helpful answer is to be found in a theorem of the probability calculus presented in section 3.2, the theorem of total prob-

5.5 Inadmissible Information

David Lewis's version of PCP says that you should set $C(e|h)$ equal to $P_h(e)$ provided that your background knowledge includes no inadmissible information about e. Lewis does not provide a definition of inadmissibility, but suggests the following heuristic: normally, all information about the world up until the time that the process producing e begins is admissible. (My talk of the process producing e is a fairly crude paraphrase of Lewis's actual view.)

The heuristic is supposed to be foolproof: Lewis mentions, as an exception, the fact that a reading from a crystal ball predicting whether or not e occurs is inadmissible. There are less outré examples of inadmissible information about the past than this, however: the example in the main text, of information about the initial conditions of a coin toss, is a case in point. Cases such as these make it very difficult to provide a good definition of inadmissibility, except to say: evidence is inadmissible relative to some $P_h(e)$ if it contains information relevant to e that is not contained in the physical probability h ascribes to e. But then how to decide what information is relevant to e? Perhaps you look to confirmation theory . . . oh.

By the way, Lewis would not sanction the application of PCP to the probability of heads, because he would not count it as a "real" probability, due to the fact that the process that determines the outcome of a coin toss is, at root, deterministic, or nearly so. (This means that he does not need to worry about the inadmissibility of information about initial conditions.) There may be some metaphysical justification for this view, but it is not very helpful to students of confirmation theory. Many scientific theories concern probabilities that Lewis would regard as unreal, statistical mechanics and population genetics (more generally, modern evolutionary biology) being perhaps the two most notable examples. If we want to use BCT to make judgments about theories of this sort, we will want to constrain the subjective likelihoods using the physical probabilities and we will need a probability coordination principle to do so.

ability, which states that for mutually exclusive and exhaustive $d_1, d_2, \ldots$,

$$C(e) = C(e|d_1)C(d_1) + C(e|d_2)C(d_2) + \cdots .$$

Suppose that the d_is are a set of competing hypotheses. Let us change their name to reflect this supposition: henceforth, they shall be the h_i. This small change in notation gives the theorem the following suggestive formulation:

$$C(e) = C(e|h_1)C(h_1) + C(e|h_2)C(h_2) + \cdots .$$

If you use PCP to set the likelihoods $C(e|h_i)$, then the total probability theorem prescribes a technique for setting $C(e)$ that depends only on your probabilities over the hypotheses h_i, that is, only on the $C(h_i)$.

In order to use the total probability theorem in this way, your set of competing hypotheses must satisfy three requirements:

1. The hypotheses each assign an explicit physical probability to events of e's kind,

2. The hypotheses are mutually exclusive, and

3. The hypotheses are exhaustive.

Of these, assumption (1) has already been made (section 5.2) in order to set a value for the subjective likelihood that constitutes the numerator of the Bayesian multiplier. Assumption (2), that no two of the competing hypotheses could both be true, may seem obviously satisfied in virtue of the meaning of "competing". A little thought, however, shows that often hypotheses compete in science although they are strictly speaking consistent, as in the case of competing Darwinian explanations of a trait that pick out different properties of the trait as selectively advantageous. (Perhaps the competitors can be reformulated so that they are mutually exclusive, however.)

Assumption (3) is perhaps the shakiest of the three. The hypotheses are exhaustive only if there is not some further theory that is incompatible with

each of the h_is. To be sure that there is no such theory, it seems, I would have to have at least some passing acquaintance with all the possible theories that claim to predict events like *e*. But this is unrealistic. The most striking examples of scientific progress are those where an entirely new explanation of the evidence comes to light.

This problem, it turns out, not only undermines a certain application of the theorem of total probability, but goes to the very root of Bayesianism, for the reason that the whole apparatus of subjective probability assumes that the individual puts a probability distribution over all possibilities. The issues arising will be examined in section 11.4. Until then, put these worries aside.

5.4 Modern Bayesianism: A Summary

We have come a long way. Back in section 5.1, Bayes' rule seemed to be telling you, rather unhelpfully, to set your probabilities after the observation of *e* equal to your earlier probabilities conditional on *e*. With the help of Bayes' theorem, this was seen to amount to multiplying your earlier probability for *h* by the Bayesian multiplier:

$$\frac{C(e|h)}{C(e)}$$

Given the constraints imposed by PCP and the theorem of total probability, the Bayesian multiplier turns out to depend entirely on your earlier probabilities for *h* and its rivals. It is equal to

$$\frac{P_h(e)}{P_{h_1}(e)C(h_1) + P_{h_2}(e)C(h_2) + \cdots}$$

(where *h* is, by assumption, one of the h_is).

Once you have set subjective probabilities for all of a set of competing hypotheses $h_1, h_2, \ldots$, then, the Bayesian apparatus tells you, given the assumptions stated in the previous two sections, how to update these proba-

bilities upon observation of any piece of evidence *e*. In other words, provided that you have some initial view about the relative plausibility of a set of competing hypotheses, BCT prescribes the *exact, quantitative* changes you must make to these on the observation of any set of evidence. The Bayesian apparatus, it seems, is a complete guide to how your beliefs must change in response to the evidence.

The probabilities that are assigned to various hypotheses before any evidence comes in are referred to as *prior probabilities* (see tech box 5.6 for an important terminological note). If you manage to set your prior probabilities to some values before any evidence comes in, then, BCT will take care of the rest. This should be enough to make you acquire a set of prior probabilities even if you do not already have them!

> *5.6 Prior Probabilities*
> The term *prior probability* is used in three distinct ways in the Bayesian literature. First, it can mean, as in the main text, your subjective probability for a hypothesis before you have received any evidence. Second, it can mean your subjective probability for a hypothesis before some particular piece of evidence *e* comes in. The probability after receipt of *e* is then called your *posterior probability*. In this sense, the term is entirely relative: your posterior probability for a hypothesis relative to one piece of evidence will be your prior probability for the hypothesis relative to the next. In the third sense, a prior probability is any subjective probability that is unconditioned by evidence. Your prior probabilities for the hypotheses are priors in this sense, but they are not the only priors. Your probabilities that you will see various kinds of evidence, for example, also count as priors, as do subjective likelihoods (probabilities of the form $C(e|h)$). The achievement of modern Bayesianism might then be stated as follows: of all the prior probabilities in the third sense, only the priors in the first sense really matter.

It is as well to enumerate the more important assumptions that were

made in order to reach the dramatic conclusion that the prior probabilities for a set of hypotheses determine the pattern of all subsequent inductive reasoning concerning those hypotheses:

1. You ought to set your subjective probabilities in accordance with PCP.
2. Each of the hypotheses ascribes an exact physical probability to any relevant evidence.
3. The hypotheses are mutually exclusive and exhaustive.

As noted in the previous two sections, ways to relax assumptions (2) and (3) will be explored later in these notes.

6. Modern Bayesianism in Action

6.1 A Worked Example

Although there is much to say about the abstract properties of BCT, let us first look at an example of BCT in action.

Suppose that we have on the table just three hypotheses, concerning the physical probability that any given raven is black.

h_1 The probability that any given raven is black is one (thus, all ravens are black).

h_2 The probability that any given raven is black is one half (thus, it is overwhelmingly likely that about one half of all ravens are black).

h_3 The probability that any given raven is black is zero (thus, no ravens are black).

We have yet to observe any ravens. Let us therefore assign equal prior probabilities to these three hypotheses of 1/3 each. (There is nothing in

BCT, as presented so far, to constrain this choice, but it does seem rather reasonable; more on this in section 9.4.)

Now we observe a black raven. We apply the Bayesian conditionalization formula to see how our subjective probabilities change upon receiving this evidence. The formula is:

$$C^{+}(h) = \frac{P_h(e)}{C(e)} C(h)$$

where $P_{h_1}(e)$ is 1, $P_{h_2}(e)$ is 1/2, $P_{h_3}(e)$ is 0, and

$$\begin{aligned} C(e) &= P_{h_1}(e)C(h_1) + P_{h_2}(e)C(h_2) + P_{h_3}(e)C(h_3) \\ &= 1/3 + 1/6 + 0 \\ &= 1/2 \end{aligned}$$

(Why can we apply the theorem of total evidence, when the three hypotheses clearly fail to exhaust the possibilities? It is enough that the probabilities of the hypotheses sum to one, that is, that I, the amateur ornithologist, *think* that the probabilities exhaust the possibilities. This ornithological belief is rather artificial, but the simplicity is welcome.)

Applying the conditionalization rule, we find that the probability of h_1 is doubled, going to 2/3, that of h_2 remains the same, at 1/3, and that of h_3 is multiplied by zero, going to zero. There are now only two hypotheses in the running, h_1 and h_2. Of the two, h_1 is ahead, due to its having assigned a higher physical probability to the observed evidence, the fact of a certain raven's being black.

Now suppose that another black raven is observed. Is the probability of h_1 doubled again? One would hope not, or it would be greater than one. It is not, because $C(e)$ has taken on a new value closer to one, reflecting our

n	$C(h_1)$	$C(h_2)$	$C(h_3)$	$C(e)$
0	1/3	1/3	1/3	1/2
1	2/3	1/3	0	5/6
2	4/5	1/5	0	9/10
3	8/9	1/9	0	?

Table 1: Change in the probabilities for e, the event of the next raven's being black, and the three raven hypotheses h_1, h_2 and h_3 (see main text for meanings) as more and more black ravens are observed, where n is the number of ravens observed so far

increased confidence in h_1. Let us do the calculation:

$$\begin{aligned} C(e) &= P_{h_1}(e)C(h_1) + P_{h_2}(e)C(h_2) \\ &= 2/3 + 1/6 \\ &= 5/6 \end{aligned}$$

where $C(\cdot)$, note, now represents our subjective probability distribution *after* observing the first black raven but before observing the second.

On observing the second raven, then, the probability of h_1 is multiplied by 6/5, and that of h_2 by 3/5, yielding probabilities of 4/5 and 1/5 respectively. Our subjective probability of a third raven's being black is now 9/10 (since we think that h_1 is very likely correct). If the third raven is indeed black, our probabilities for h_1 and h_2 will go to 8/9 and 1/9. To be sure you understand what is going on, calculate the new value for $C(e)$, the probability that a fourth raven will be black, and the probabilities for h_1 and h_2 if the fourth raven is as black as we expect. The evolution of the subjective probabilities as ravens are observed is shown in table 1.

This simple example illustrates a number of facts about BCT:

1. If a hypothesis is logically inconsistent with the evidence, upon conditionalization its probability goes to zero.

2. Once a hypothesis's probability goes to zero, it can never come back. The hypothesis is eliminated.

3. The hypothesis that assigns the highest probability to the observed evidence (h_1 in the example) receives the biggest probability boost from the observation of the evidence. A hypothesis that assigns probability one to the evidence will receive the largest boost possible in the circumstances.

4. If a hypothesis is consistent with the evidence, its probability can never go to zero, though it can go as near zero as you like (as would h_2's probability if nothing but black ravens were observed).

5. After conditionalization, your subjective probabilities for a set of mutually exclusive, exhaustive hypotheses (such as h_1, h_2, and h_3) will always sum to one.

6. As a certain hypothesis becomes dominant, in the sense that its probability approaches one, its probability boost from further successful predictions declines (though there is always a boost).

6.2 General Properties of Bayesian Confirmation

Now to generalize from the last section's example. First, a little terminology. When the observation of e causes the probability of h to increase, say that h is *confirmed*. When the observation of e causes the probability of h to decrease, say that h is *disconfirmed*. When the observation of e causes the probability of h_1 to increase by a greater proportion than does the probability of h_2, say that h_1 is *more strongly confirmed* than h_2. This is not the only measure of strength of confirmation that may be used within a Bayesian context—some other possibilities are mentioned in tech box 6.1—but it is the most useful for my present expository purposes.

6.1 Weight of Evidence

The degree to which a piece of evidence confirms a hypothesis can be quantified in various ways using the Bayesian framework of subjective probabilities. The *difference measure* of relevance equates the degree of confirmation with the difference between the relevant posterior and prior probabilities, so that the degree to which e confirms h is equal to $C(h|e) - C(h)$. The *ratio measure* equates it with the ratio of the same two probabilities, that is, with the Bayesian multiplier; this is the measure used in the main text. The *likelihood ratio measure* sets the degree of relevance to $C(e|h)/C(e|\neg h)$, and the *log likelihood ratio measure* to the logarithm of this quantity (which has the effect of making degree of confirmation additive). Each of these measures has its uses.

Some writers argue that there is a single, correct way to measure degree of confirmation that gives sense to our everyday talk about the weight of evidence. Most Bayesians take a more pluralistic and pragmatic line, noting that unlike some systems of inductive logic, BCT does not give a notion of weight of evidence any crucial role to play.

Let me define the same terms more compactly. As above, let the Bayesian multiplier be $P_h(e)/C(e)$, the amount by which the probability of h is multiplied when conditionalizing on e. Then, first, e confirms h if the Bayesian multiplier is greater than one and disconfirms h if the multiplier is less than one. Second, e confirms h_1 more strongly than it confirms h_2 if the Bayesian multiplier for h_1 is greater than the multiplier for h_2.

I will state some generalizations about BCT in the form of five "principles", followed by two "remarks".

The Hypothetico-Deductive Principle
If h entails e, then the observation of e confirms h.

We have already seen that this is true. If h entails e, then $P_h(e)$ is equal to one, so the Bayesian multiplier must be greater than one, on the assumption that $C(e)$ is less than one.

The smaller the value of $C(e)$, the greater will be the value of the multiplier. Thus surprising predictions confirm a hypothesis more strongly than unsurprising predictions. The exception to the hypothetico-deductive principle that occurs when $C(e) = 1$ is an extreme case of this observation: evidence that is expected with certainty has no confirmatory (or disconfirmatory) weight whatsoever. You might think that in practice, this could never happen; but in a certain sense it often does, as we will see in section 11.

The Likelihood Lover's Principle
The higher the physical probability that h assigns to e, the more strongly h is confirmed by the observation of e.[4]

The denominator of the Bayesian multiplier is the same, at any given time, for all competing hypotheses, equal to $C(e)$. The numerator is $P_h(e)$.

4. The principle requires that the hypotheses have non-zero priors, and that the background knowledge includes no inadmissible evidence (see tech box 5.5 for inadmissibility).

Therefore the Bayesian multiplier varies among competing hypotheses in proportion to $P_h(e)$, the physical probability that they assign to the evidence.

Some more specific claims along the lines of the likelihood lover's principle:

1. If a hypothesis assigns a physical probability to e that is higher than $C(e)$, it is confirmed by the observation of e (obviously!). To put it more qualitatively, if h predicts e with a higher probability than (the probabilistically weighted) average, it is confirmed by the observation of e.

2. If there are only two hypotheses in the running, and one ascribes e a higher probability than the other, then e confirms the one and disconfirms the other. This is because, as a consequence of the theorem of total probability, $C(e)$ is always somewhere between the physical probability assigned by one hypothesis and that assigned by the other.

(The likelihood lover's principle should not be confused with what some philosophers call the likelihood principle, according to which subjective likelihoods (and only subjective likelihoods) are responsible for a piece of evidence's differential impact on two or more different hypotheses. The likelihood principle differs from the likelihood lover's principle in two ways: it concerns subjective, not physical, likelihoods (and thus does not depend on PCP), and it says not only that strength of confirmation varies with the likelihoods, but that it is entirely determined, relatively speaking, by the likelihoods.)

The Equal Probability Principle

Hypotheses that assign equal physical probabilities to a body of evidence are equally strongly confirmed by that evidence.

It is easy to see that such hypotheses have equal Bayesian multipliers. Though it is trivially true, the equal probability principle is essential for an

understanding of what BCT can and cannot do; see, in particular, section 7.

The Good Housekeeping Principle

No matter what evidence is observed, the probabilities of a set of mutually exclusive, exhaustive hypotheses will always sum to one.

Bayesianism's ability to keep all of your subjective probabilities in good order as you conditionalize your way through the evidence can seem quite miraculous, not least when you are in the middle of a complex Bayesian calculation. The reason that the good housekeeping principle holds true, though, is easy to see. Consider the case in which there are just two competing hypotheses h_1 and h_2. The probabilities of h_1 and h_2 after conditionalization on some e are

$$C^+(h_1) = \frac{P_{h_1}(e)C(h_1)}{P_{h_1}(e)C(h_1) + P_{h_2}(e)C(h_2)} \quad \text{and}$$

$$C^+(h_2) = \frac{P_{h_2}(e)C(h_2)}{P_{h_1}(e)C(h_1) + P_{h_2}(e)C(h_2)}.$$

These probabilities have the form

$$\frac{a}{a+b} \quad \text{and} \quad \frac{b}{a+b}$$

and so they must sum to one.[5] You should have no trouble seeing, first, that the result generalizes to the case where there are any number of competing hypotheses, and second, that the fact that the subjective likelihoods are set in accordance with PCP makes no difference to this result.

The Commutativity Principle

The order in which the evidence is observed does not alter the cumulative effect on the probability of a hypothesis.

5. It is not necessary, note, to assume that $C(h_1)$ and $C(h_2)$ sum to one. Conditionalization actually repairs subjective probabilities that are out of whack! (But of course, you would have no reason to apply the theorem of total probability in a case such as this.)

For example, conditionalizing on e and then on d will leave you with the same probability for h as conditionalizing on d and then on e, or as conditionalizing on e and d at the same time.

This result is a straightforward consequence of the definition of conditional probability, but it can be difficult to represent multiple conditionalizations in just the right way to see how. Here is one approach. Define $C_d(\cdot)$ to mean $C(\cdot \mid d)$. Then upon conditionalizing on d, one's new probability for h ought to be equal to the old $C_d(h)$. Further conditionalizing on e will result in a probability for h equal to the original $C_d(h \mid e)$. Now

$$\begin{aligned} C_d(h \mid e) &= \frac{C_d(he)}{C_d(e)} \\ &= \frac{C(hed)/C(d)}{C(ed)/C(d)} \\ &= \frac{C(hed)}{C(ed)} \\ &= C(h \mid ed) \end{aligned}$$

as desired. By symmetrical reasoning, $C_e(h \mid d) = C(h \mid ed)$.[6]

Remark: The Contrastive Aspect of Bayesian Confirmation Theory

Observe that a hypothesis's degree of confirmation depends not only its own predictions, but on the predictions of its competitors. One notable consequence of this property of BCT is that a hypothesis that is quite inaccurate can be assigned a subjective probability as near one as you like if its competitors are even more inaccurate. To take a very simple example, if the only hypotheses about raven blackness that have non-zero subjective probability

6. It is an open question whether commutativity holds for Jeffrey conditionalization (Field 1978). (Jeffrey conditionalization is described in section 4.2.) The difficulty is more philosophical than mathematical: it is unclear what would count as making the same observations in two different orders. On one interpretation, Jeffrey conditionalization is commutative, but on others, not.

are are *One half of all ravens are black* and *No ravens are black*, then if all observed ravens are black, the first hypothesis will come to have subjective probability one. Similarly, a hypothesis that is very accurate can have its subjective probability greatly decrease (though slowly) if its competitors are even more accurate. There is only a limited amount of subjective probability; it flows towards those hypotheses with the physical likelihoods that are relatively the highest, regardless of their absolute size.

Remark: The Contextual Aspect of Bayesian Confirmation Theory

In the Bayesian scheme, any piece of knowledge can, potentially, affect the impact of a piece of evidence on a hypothesis. Relative to one set of background knowledge, a piece of evidence may confirm a hypothesis a great deal, while relative to another, it may confirm it not at all. This is in part because the background can make a difference to whether or not the hypothesis entails the evidence, or to what physical probability it assigns the hypothesis (some examples are given in section 8.1). But only in part: in principle, almost any kind of information might affect your subjective likelihood $C(e|h)$.

This contextual sensitivity of BCT has been generally regarded as a good thing, since it has long been accepted that inductive reasoning has a certain holistic aspect. But it also makes it more difficult to establish that BCT will be well behaved and that it will lead to some kind of scientific consensus (see section 9).

6.3 Working with Infinitely Many Hypotheses

In the example of the ravens in section 6.1, I assumed that there were just three hypotheses under serious consideration, hypotheses assigning probabilities of zero, one half, and one, to the event of the next observed raven's being black. In reality, I would want to consider all the possible probabilities

for blackness, which is to say that I would seriously entertain all hypotheses of the form *The probability of the next raven's being black is p*.

There are infinitely many such hypotheses. This creates some rather tricky problems for the Bayesian. Most worrying, the only way to assign probabilities to an infinite set of mutually exclusive hypotheses so that the probabilities sum to one (as the axioms insist they must) is to assign almost every hypothesis a probability of zero. But a hypothesis that is assigned a prior probability of zero can never have its probability increase. So it seems that I must rule almost every hypothesis out of the running before I even begin to collect evidence. Nevermore indeed!

There is a purely technical solution to these problems that will be familiar to anyone who has done a little work on probability or statistics. It is to introduce the notion of a *probability density*. Even those of you who know no statistics have very likely seen at least one example of a probability density: the normal curve.

The fundamental idea behind a probability density is that, in a case where there are infinitely many possibilities, I should assign subjective probabilities not to individual possibilities, but to sets of possibilities. For example, in the case of the ravens, I may not have an interesting (that is, non-zero) subjective probability for the hypothesis that the probability of black raven is exactly 0.875, but I may have a subjective probability for the hypothesis that the probability of a black raven is somewhere between 0.8 and 0.9. That is, I may have a subjective probability for the set of hypotheses of the form

$$\{\text{ The probability of the next raven's being black is } x : 0.8 < x < 0.9 \}$$

Provided that I have enough subjective probabilities of this sort (in a well-defined sense of *enough*), I can take advantage of all that the Bayesian has to offer by applying conditionalization to the probabilities over sets.

It turns out that there is a very simple way for me to assign subjective probabilities to all the relevant sets. What I can do is to adopt a *probability*

density over the competing hypotheses. The properties of a probability density are easiest to appreciate in graphical form. Consider a two-dimensional graph on which the points along the x-axis between zero and one each represent the corresponding raven hypothesis, that is, on which the point $x = p$ represents the hypothesis *The probability of the next raven's being black is p*.

A probability density can then be represented by a well-behaved function $f(x)$ defined for arguments between zero and one, such as the function shown in figure 1. The assignment of subjective probabilities inherent in

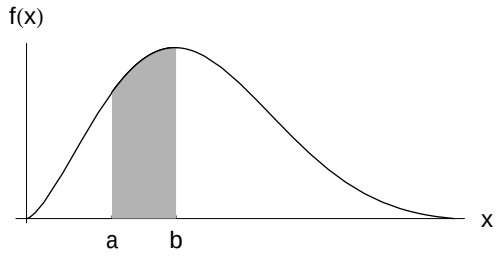

Figure 1: A probability density over the raven hypotheses. The area of the shaded region is by definition equal to the subjective probability that the physical probability of raven blackness is somewhere between a and b.

$f(x)$ is to be interpreted as follows: the subjective probability that the physical probability of raven blackness is between a and b is equal to the area under $f(x)$ between a and b. (The area under $f(x)$ between zero and one, then, had better be equal to one.)

Before I observe any ravens, then, I can set my prior probabilities over the infinitely many hypotheses about the physical probability of blackness by adopting a subjective probability density. If I think that all hypotheses are equally likely, I will adopt a flat density like that shown in figure 2a. If I think that middling probabilities for blackness are more likely, I will adopt a humped density like that shown in figure 2b. If I think that higher probabilities are more likely, my density might look like the density in figure 2c. And so on.

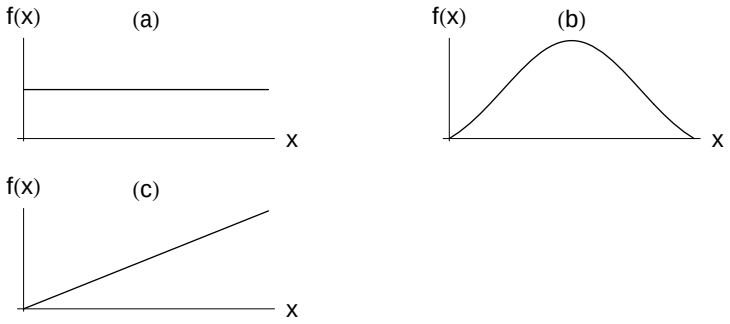

Figure 2: Some choices of prior probability density for the raven blackness hypotheses

How to conditionalize? Very conveniently, I can conditionalize by pretending that $f(x)$ for any x is a probability. That is, when I observe e, my new probability density over the hypotheses $f^+(x)$ should be related to my old density $f(x)$ by the following familiar-looking relation:

$$f^+(x) = \frac{P_{h_x}(e)}{C(e)} f(x).$$

The probability $P_{h_x}(e)$ is the probability assigned to the evidence by the hypothesis that the probability of a black raven is x. If e is the observation of a black raven, then $P_{h_x}(e) = x$.

The theorem of total probability can be applied by using the integral calculus. You do not need to understand this. For those of you who do,

$$C(e) = \int_0^1 P_{h_x}(e) f(x)\, dx.$$

In the case where e is the observation of a black raven, then,

$$C(e) = \int_0^1 x f(x)\, dx.$$

Now we can crunch some numbers. Suppose that I begin with a flat prior over the raven probability hypotheses (figure 3a). After the observation of

a black raven, my new density is that shown in figure 3b. After two black ravens, my density is as in figure 3c. After three and four black ravens, it is as in figures 3d and 3e. After ten black ravens, it is as in figure 3f.

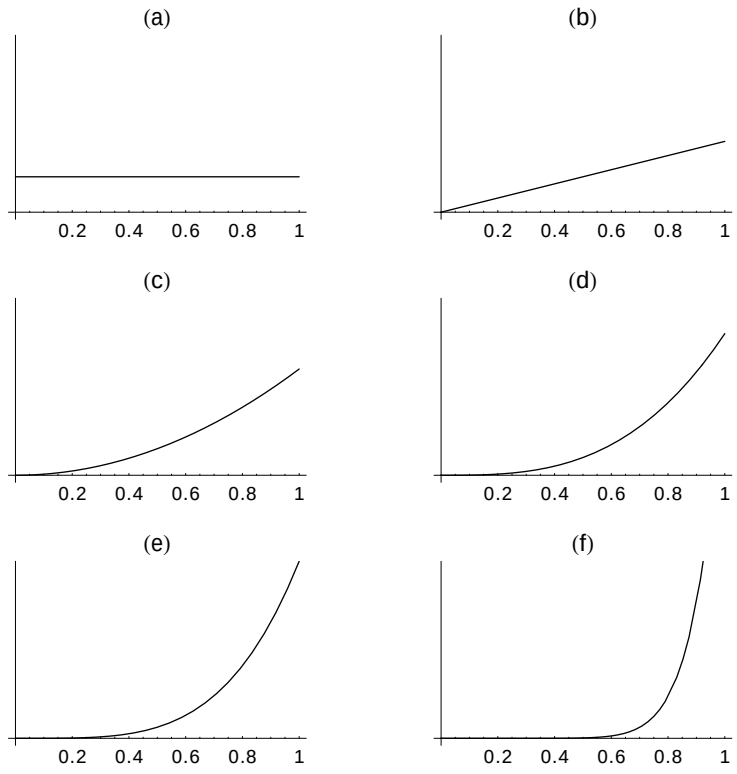

Figure 3: The change in my subjective probability density as more and more black ravens are observed

By way of contrast, figure 4 shows the change in my prior if only some of the observed ravens are black. The prior is shown in (a), then (b)–(f) show the density after observing 2, 6, 10, 20, and 36 ravens respectively, in each case assuming that only one half of the observed ravens are black.

As you can see, my density becomes heaped fairly quickly around those hypotheses that ascribe a physical probability to blackness that is close to the observed frequency of blackness. As more and more evidence comes in,

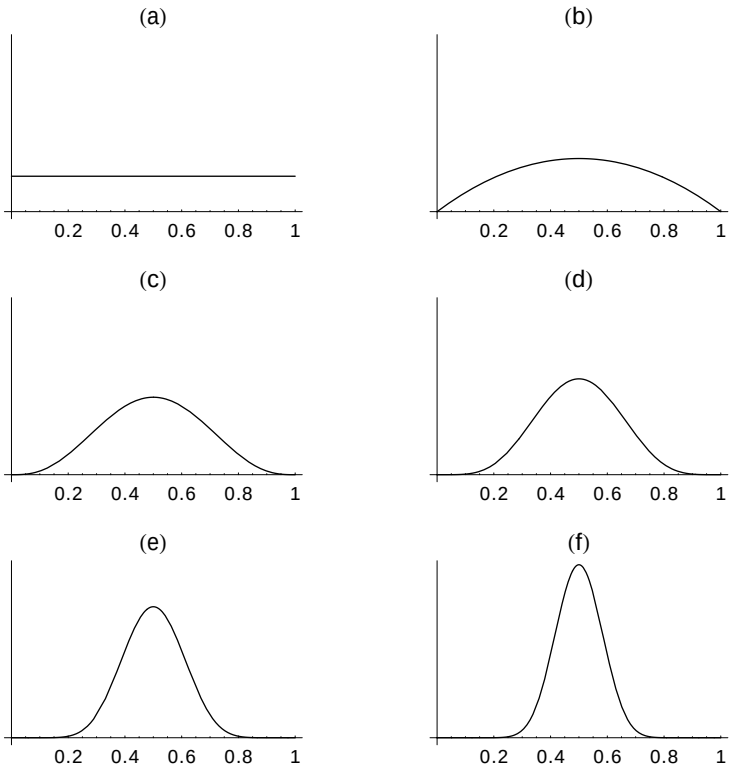

Figure 4: The change in my subjective probability density as more and more ravens are observed, if half of all observed ravens at any stage are black

my subjective probability density will (with very high physical probability) become more and more heaped around the true hypothesis. (For more on the mathematics of convergence, see tech box 6.2. For more on the use of convergence results to undergird BCT, see section 9.2.) After observing 100 ravens, exactly 50 of which are black, for example, my density will be close to zero for any hypothesis that assigns a probability to blackness that differs from one half by more than about 0.1.

6.2 The Law of Large Numbers

The heaping, over time, of subjective probability around the correct hypothesis concerning raven blackness, depends on a mathematical result called the law of large numbers. The law concerns a series of experiments each of which produces an outcome e with a probability x. It states that, as more and more experiments are performed, the proportion of experiments that produce e will, with high probability, converge on x. For example, if the probability of observing a black raven is 0.7, then, as more and more ravens are observed, the proportion of observed ravens that are black will converge to 70% with very high probability. Note that there is always a small probability that the proportion of black ravens is quite different from 70%; this probability, though, becomes smaller and smaller as more and more ravens are observed, tending to zero in the long run.

There is more than one version of the law of large numbers. The result that I stated informally in the last paragraph is called the *weak law of large numbers*. It may be stated more formally as follows. Let b be the proportion of observed ravens that are black, and x the probability that any given raven is black. Assume that the observations are probabilistically independent (section 3.3). Then, as the number of ravens observed tends to infinity, for any small quantity ε, the probability that b differs from x by more than ε tends to zero.

What, in this case, is my subjective probability for the hypothesis that the probability of blackness is exactly one half? It is zero! It is zero even after

I observe two million ravens, exactly one million of which are black. This is because, although almost all my subjective probability is piled in the area very, very close to $x = 0.5$, there are infinitely many hypotheses even in this tiny area, and the large but finite quantity of probability piled in the area must be shared between these infinitely many hypotheses.

As a consequence, no matter how much evidence I accumulate, I am never in a position to say: I am very confident that the probability of blackness is one half. The best I can do is to say: I am very confident that the probability of blackness is very close to one half. Quite reasonably, most Bayesians would say.

You might think that if I have a subjective probability of zero for a hypothesis, I am certain that it is false. But as the raven case shows, this is not correct. When I have the sort of probability density over the rival raven hypotheses shown in figures 2, 3, and 4, I assign each of them a zero probability. But at the same time, I am sure that one of them is correct. Unless I contradict myself, my assignment of zero probability cannot entail certainty of falsehood. It rather represents something that you might call practical certainty of falsehood. Given one of the raven hypotheses in particular, I will accept any odds you like that it is false. Yet I acknowledge that it might turn out to be true.

Assigning probability zero to a hypothesis, then, can mean one of two things. It can mean that I am absolutely sure that the hypothesis is false, as when I assign zero probability to an antitautology (a proposition that is false of logical necessity). Or it can mean mere practical certainty of falsehood, that is, a willingness to bet against the hypothesis at any odds, coupled with the belief that it might nevertheless be true. The same is true for probability one.

6.4 When Explicit Physical Probabilities Are Not Available

We have proceeded so far on the assumption that our competing hypotheses each ascribe a definite physical probability to the evidence e. But in many cases, this assumption is false.

Consider a very simple case. Suppose that our competing hypotheses concerning raven color specify not a physical probability of blackness, but a frequency of black ravens. For example, $h_{0.5}$ says not *The next observed raven will be black with probability 0.5*, but rather *Half of all ravens are black*. This hypothesis does not entail that the next observed raven will be black with probability 0.5. It is consistent with the possibility that, for example, all of the ravens around here are black, but enough ravens elsewhere are white to compensate for the local predominance of blackness. In such a scenario, the probability of the next raven's being black is more or less one. Then again, it might be that all the local ravens are white, in which case the probability of the next raven's being black is about zero. How to apply BCT to hypotheses such as these?[7]

The usual solution to the problem is to introduce one or more *auxiliary hypotheses*. Traditionally, an auxiliary hypothesis is a posit used to extract from a theory a definite verdict as to whether some piece of evidence e will be observed or not—to get the theory to entail either e or $\neg e$—but in the present context, it is a hypothesis that, in conjunction with a hypothesis h, assigns a definite physical probability to e.

A simple example of an auxiliary hypothesis suitable for the ravens case is the hypothesis that raven color is uncorrelated with raven habitat, so that roughly the same percentage of black ravens live in any locale. Conjoining this hypothesis with a hypothesis that the frequency of black ravens is x yields the conclusion that the probability of the next observed raven's being

7. The Bayesian apparatus does not strictly speaking require a physical probability, only a subjective likelihood $C(e|h)$. By insisting on physical probabilities now, I make possible a much stronger response to later worries about the subjectivity of BCT (section 9).

black is x.[8]

Call the auxiliary hypothesis a. Let h be the thesis that the proportion of black ravens is x. Then h and a together ascribe a physical probability x to e, or in symbols, $P_{ha}(e) = x$. The subjective likelihood $C(e|ha)$ can therefore be set, in accordance with PCP, to x.

This is not quite what we wanted. It allows us to use PCP to calculate a shift in our subjective probability for ha, but not in our subjective probability for h. It turns out that, in certain circumstances, a piece of evidence can have a qualitatively different effect on h and ha, raising the subjective probability of one while lowering the subjective probability of the other. Although e's impact on the subjective probability for ha is regimented by PCP, then, its impact on h is, in general, not. The problem, and some solutions, are investigated further in section 10.

It is worth noting that, in some circumstances, you can make your auxiliary hypotheses true. This is, in effect, what is done in statistical sampling. By choosing ravens from a population at random, you ensure that the probability of the next observed raven's being black is equal to the proportion of black ravens in the population as a whole. In this way, you can learn about the population even if unplanned encounters with population members are far from random.

More generally, when scientists conduct experiments in carefully controlled circumstances, they are ensuring, as best they can, the truth of auxiliary hypotheses that, in combination with the hypotheses being tested, assign definite physical probabilities to the various possible experimental outcomes.

The prospect for a Bayesian philosophy of science that is mediated entirely by physical likelihoods is for these reasons far better than some writers (e.g., Earman 1992, §6.3) would have you believe.

8. Strictly speaking, of course, you need in addition the assumption that you are sampling randomly from the local habitat.

7. Does Bayesianism Solve the Problem of Induction?

7.1 Subjective and Objective In Bayesian Confirmation Theory

The old-fashioned problem of induction is the problem of finding a priori, objective grounds for preferring some hypotheses to others on the basis of what we observe (where the hypotheses have a scope that goes beyond our observations). Bayesian confirmation theory tells us how to change our degrees of belief for the hypotheses given the evidence. Does BCT therefore solve the problem of induction?

Although it is true that BCT tells us how to update our subjective probabilities in the light of the evidence, that is not enough in itself to solve the problem of induction. The recommended changes must be founded in a priori, objective constraints. Modern Bayesian confirmation theory claims three such constraints: subjective probabilities should be set in accordance with

1. The axioms of probability,
2. Bayes' conditionalization rule, and
3. The probability coordination principle.

Let us grant, for the sake of the argument, that these constraints are indeed a priori and objective. Now does BCT solve the problem of induction?

Not quite. The reason is that BCT's recommendations are determined not only by the three objective, a priori constraints, but also by your prior probabilities, that is, the subjective probabilities you assign before any evidence comes in. Assuming, as we have been so far, that the priors are subjectively determined, it follows from the fact that the priors make an essential contribution to BCT's recommendations that these recommendations are partly subjectively determined. Thus, the fact that BCT makes such recommendations does not qualify it as a solution to the old-fashioned problem of induction.

In order to use BCT to solve the problem of induction we must do one of two things. First, we might find some further objective, a priori rule that constrains the prior probabilities so effectively that BCT's recommendations are completely determined by the constraints. The subjective element of Bayesianism would be eliminated entirely. This possibility is investigated further in section 9.4.

The second option is to find some aspect of BCT's recommendations that is entirely determined by the constraints. Such an aspect is enunciated by LLP, the likelihood lover's principle (section 6.2):

> The higher the physical probability that a hypothesis *h* assigns to the evidence *e*, the more strongly *h* is confirmed by the observation of *e*.[9]

Although the quantitative details of BCT's recommendations depend on the subjectively determined priors, no matter how I set my priors, I will respect LLP. That I should follow LLP, then, is determined by the axioms of probability, Bayes' rule, and PCP. On our working assumption that these constraints are all objective and a priori, LLP is also, for the Bayesian, objective and a priori.

It would seem, then, we have made some progress on induction. Bayesian confirmation theory puts a real constraint on inductive reasoning. But how much, exactly, follows from LLP? And how much depends, by contrast, on the priors?

In what follows, I ask, concerning three features of our inductive reasoning, whether they follow from LLP or some other objective aspect of the Bayesian machinery, or whether they depend on the way that one's prior probabilities are assigned. The three features are:

1. Our expectation that the future will resemble the past, that is, our belief in the uniformity of nature,

9. The principle requires that the hypotheses have non-zero priors, and that the background knowledge includes no inadmissible evidence (see tech box 5.5 for inadmissibility).

2. Our preference for hypotheses that are not framed using "grue-like" predicates, and

3. Our preference for simpler over more complex hypotheses, when the competing hypotheses are otherwise equally successful in accounting for the data.

In each case, sad to say, it is the assignment of the priors, and not anything more objective, that accounts, on the Bayesian story, for these elements of our inductive practice.

7.2 The Uniformity of Nature

Can BCT justify a belief—or a high degree of belief—in the uniformity of nature? At first blush, it seems that if I am committed to the likelihood lover's principle, I ought to expect the future to be like the past, as illustrated by the raven case. Suppose that every raven I see is black. Then, if I follow LLP, my subjective probabilities will converge on the hypotheses that assign the highest probabilities to these observations, that is, the hypotheses that assign the highest physical probabilities to blackness in ravens. As the convergence occurs, the probability I assign to any future raven's being black grows higher and higher, thanks to the theorem of total probability—as shown by the examples in section 6.1 (see especially table 1).

More generally, if all my hypotheses have the form *The probability of an r being b is x*, and I observe a large number of *r*s of which a proportion *p* are *b*, then my subjective probability for the next *r*'s being *b* will be roughly equal to *p*, the more so as more *r*s are observed. Thus, I will expect the future proportion of *r*s that are *b* to be equal, at least roughly, the past proportion of the same—which is to say, I will set my expectations in accordance with the principle of the uniformity of nature.

Did I make any additional assumptions in reaching this conclusion? Unfortunately, yes. I made the least acceptable assumption possible: I assumed

that I started out with a very high subjective probability for the future's resembling the past.

The assumption came by way of the assumption that all my hypotheses have the form *The probability of an r being b is x*, which implies the existence of a single, unchanging physical probability for raven blackness. To see how powerful this assumption is, consider the following alternative form for a hypothesis covering the same phenomena:

> The probability of an r first observed before the year 3000's being b is x; the probability of an r first observed after the year 3000's being b is $1 - x$.

For the value $x = 1$, a hypothesis of this form will be strongly confirmed by all black ravens observed before 3000. Indeed, it will receive precisely the same Bayesian multiplier as the raven hypothesis, *All ravens are black*, thanks to the equal probability principle (section 6.2, and explained further in section 7.3). If it is assigned the same prior as the raven hypothesis, then it will have the same posterior probability in the year 3000, and so will count as heavily *against* the next observed raven's being black as the raven hypothesis counts for it.

Now you see that the argument from LLP to an expectation that the future will resemble the past depends on my discounting—that is, assigning very low or zero probability to—hypotheses that explicitly state that the future will (very likely) not resemble the past.

The point, that inductive expectations were assumed, not justified, can be made even more dramatically. Suppose that my hypotheses concerning the blackness of ravens are $h_1, \ldots, h_n$. Let u be the proposition that the future, with high probability, resembles the past, at least with respect to the color of ravens. Then

$$C(u) = C(u|h_1)C(h_1) + \cdots + C(u|h_n)C(h_n).$$

But any individual h_i having the form assumed above—*The probability of an r being b is x*—entails u. Thus

$$C(u) = C(h_1) + \cdots + C(h_n) = 1.$$

So by assuming that you consider only hypotheses of the given form to be viable candidates for laws about raven blackness, I assume that you already assign subjective probability one to the uniformity of nature. It is not BCT that compels your allegiance to uniformity; the passion for order has all along been burning in your heart.

7.3 Goodman's New Riddle

If you followed the argument in the last section, then you already strongly suspect that BCT will fail to solve Goodman's new riddle of induction, where a solution to the riddle is an a priori argument for preferring hypotheses couched using standard predicates such as *black* to hypotheses couched using non-standard predicates such as *blite*. (An object is *blite* if it is black and first observed before the year 3000, or white and first observed after 3000.)

In order to solve the new riddle, we need a reason currently to prefer, for example, the hypothesis *All ravens are black* to the hypothesis *All ravens are blite*, such that the preference does *not* depend in any way on the prior probabilities for the hypotheses. But because both hypotheses assign exactly the same physical probability to the evidence observed so far (some large number of ravens, all black), BCT cannot discriminate between them.

Let me explain the point in greater detail. Call the observed evidence e. Your current subjective probability for the first hypothesis, call it h, is

$$C^+(h) = \frac{C(e|h)}{C(e)}C(h);$$

that for the second hypothesis, call it h', is

$$C^+(h') = \frac{C(e|h')}{C(e)}C(h').$$

Since both hypotheses predict that all ravens observed until now will be black, their likelihoods are both equal to one, that is, $C(e|h) = C(e|h') = 1$. So we have

$$C^+(h) = \frac{1}{C(e)}C(h); \quad C^+(h') = \frac{1}{C(e)}C(h').$$

The only difference between these two expressions is in the prior probabilities assigned to each, thus, any preference a Bayesian has for one over the other must be expressed in the prior probabilities, that is, the probabilities assigned *before* any evidence comes in. If you are predisposed against hypotheses involving the predicate *blite*, and you thus assigned a low prior to h', then BCT will preserve your bias. Equally, if you are predisposed against hypotheses involving *black* and in favor of *blite*, then BCT will preserve that bias, too. If Bayesianism is correct, then, no considerations from confirmation theory alone militate for or against Goodman predicates.

You will see that the source of BCT's even-handedness is the equal probability principle. This principle is just the flip side of the LLP: whereas LLP directs us to change our probabilities in accordance with the likelihoods, the equal probability principle directs us *only* to change our probabilities in accordance with the likelihoods.

7.4 Simplicity

All things being equal, we prefer the simpler to the more complex of two competing theories. What is the meaning of the ceteris paribus clause, *all things being equal*? Something like: *if the two theories, in conjunction with the same auxiliary hypotheses, assign the same physical probabilities to all the phenomena in their domain*. Say that two such hypotheses are *empirically equivalent*. Of course, our preference for simplicity is not confined to cases of empirical equivalence; we consider theoretical simplicity a virtue wherever we find it. But confining our attention to the case of empirically equivalent

theories is like conducting an experiment in a clean room: the influence of confounding variables is minimized.

You should see straight away that BCT does not impose a preference for simplicity in these circumstances. The argument is, of course, identical to the argument of the last section. When two theories are empirically equivalent, their likelihoods relative to any given body of evidence are equal. Thus the difference in anyone's subjective probabilities for the theories must be due entirely to the difference in the prior probabilities that were assigned to the theories before any evidence came in. Bayesian confirmation theory preserves a prior bias towards simplicity, but it implements no additional bias of its own.

You will find an additional comment about the confirmatory virtues of simplicity in tech box 11.1.

7.5 Conclusion

Let us take stock. Bayesian confirmation theory does impose an objective constraint on inductive inference, in the form of the likelihood lover's principle, but this is not sufficient to commit the Bayesian to assuming the uniformity of nature, or the superiority of "non-grueish" vocabulary or simple theories. The first of these failures, in particular, implies that BCT does not solve the problem of induction in its old-fashioned sense.

If the old-fashioned problem of induction cannot be solved, what can we nevertheless say about BCT's contribution to the justification of induction? There are two kinds of comments that can be made. First, we can identify unconditional, though relatively weak, constraints that BCT puts on induction, most notably the likelihood lover's principle. Second, we can identify conditional constraints on induction, that is, constraints that hold given other, reasonable, or at least psychologically compelling, assumptions. We can say, for example, that *if* we assign low priors to grueish hypotheses, BCT directs us to expect a future that resembles the past. This is, remember,

considerably more than we had before we began.

8. Bayesian Confirmation Theory and the Problems of Confirmation

One of Bayesian confirmation theory's many advantages is its ability to solve, in a principled way, some of the most famous problems of confirmation. It does not solve every problem, and some of its solutions are open to question, but it is far ahead of most of its competitors, which, as every good Bayesian knows, is what really matters (see the remark on the contrastive aspect of BCT in section 6.2).

8.1 The Paradox of the Ravens

Can a non-black non-raven, such as a white shoe, count as evidence for the hypothesis *All ravens are black*? Hempel showed that the answer must be yes, if we accept a few simple and innocent-looking assumptions about confirmation (see tech box 8.1). This is what he called the *paradox of the ravens* (Hempel 1945a).

Bayesianism does not accept all of Hempel's premises, but it nevertheless has its own ravens paradox: according to BCT, the white shoe would seem, given a few entirely reasonable assumptions, to confirm the raven hypothesis, with or without Hempel's equivalence assumption. The reason is as follows. Suppose I come across a white shoe. The probability of my doing so, given that all ravens are black, is higher, it can reasonably be argued, than the probability of my doing so, given that not all ravens are black.

Can you see why? Think of the shoe as a non-black non-raven. If for all I know, some ravens are white, the next object I came across could have fallen into any of four classes relevant to the raven hypothesis: it could have been a black raven, a non-black raven, a black non-raven, or a non-black non-raven. Suppose that I have subjective probabilities for each of these

8.1 Hempel's Ravens Paradox

Hempel's assumptions are:

1. Hypotheses are confirmed by their positive instances; for example, the hypothesis *All Fs are G* is confirmed by the observation of an *F* that is also a *G*.

2. The hypothesis *All Fs are G* is logically equivalent to the hypothesis *All non-G things are non-Fs*.

3. A piece of evidence that confirms a hypothesis confirms anything logically equivalent to the hypothesis.

From (1) it follows that *All non-black objects are non-ravens* is confirmed by the observation of any non-black non-raven, for example, by a white shoe. From (2) and (3) it follows that anything that confirms this hypothesis also confirms the raven hypothesis, *All ravens are black*. Thus the observation of a white shoe confirms the raven hypothesis.

Note that BCT rejects Hempel's first premise (see Good's counterexample explained at the end of this section), is agnostic about his second premise, and accepts his third premise.

possibilities. I am then told that, in fact, all ravens are black. My subjective probability for the second possibility must now go to zero. In order that my subjective probabilities still sum to one, that subjective probability must then be distributed among the other three possibilities. Assuming that I do not arbitrarily withhold it from the last possibility, my probability that the next object I encounter will be a non-black non-raven will go up. Thus, if e is the event of the next encountered object's being a non-black non-raven, and h is the hypothesis that all ravens are black (and I am a Bayesian), then $C(e|h) > C(e)$. But as you well know, this means that the observation of e—and my coming across a white shoe constitutes such an observation—will, according to BCT, confirm h.

Goodman famously remarked that the ravens paradox seems to open up the possibility of indoor ornithology. By rummaging through the contents of my closet, which is full of non-black non-ravens, it seems that I can confirm the raven hypothesis, that is, that I can make some progress on learning the characteristic color of ravens. This cannot be right—can it?

A standard response to the ravens paradox—I should perhaps say a standard accommodation of the paradox—was proposed early on by the Polish logician and philosopher Janina Hosiasson (later Hosiasson-Lindenbaum, 1899–1942). Hosiasson's view was that, once confirmation is put in a probabilistic framework (applause from the Bayesians), and a few plausible probabilistic assumptions are made, it follows that the white shoe *does* confirm the raven hypothesis, but by a negligibly small amount. The Bayesian multiplier is so close to one that it is simply not worth opening the closet door.

Some philosophers have responded, rightly I think, that this cannot be the whole story. Rummaging in my closet cannot confirm the raven hypothesis, not even by a negligible amount. What I find in the closet is simply irrelevant to the hypothesis; equivalently, looking in the closet does not constitute a test of the hypothesis.

A number of Bayesians and their friends (see especially Horwich (1982))

have pointed out that BCT provides the resources to understand this point. I know for certain that there are no ravens in my closet. Given that assumption, there are only two possibilities as regards the next object in the closet: it will be a black non-raven or a non-black non-raven. Under these circumstances, learning that all ravens are black will not close off any live possibilities, as it did above. There is no probability to redistribute, and so $C(e|h) = C(e)$. In effect, my background knowledge renders e irrelevant to h; thus, whatever I find, it will not change my subjective probability for *All ravens are black*.

On this way of dealing with the problem, indoor ornithology is impossible, but there are still some white shoes that will confirm the raven hypothesis. There are (those few) white shoes that occur in the wild, where I might also encounter a raven. In the wild, the probabilistic reasoning that led to the ravens paradox above goes through in many circumstances, and the raven hypothesis is confirmed. The best that the Bayesian can say about this is to repeat Hosiasson's insight: the amount of confirmation is negligible.

Is there any more systematic way of assessing the significance of white shoes, black ravens, and so on, for the ravens hypothesis? A useful tool is the notion of an auxiliary hypothesis introduced in section 6.4. Perhaps even a necessary tool: it is very difficult to interpret the significance of a found object for the raven hypothesis without an auxiliary hypothesis that, in conjunction with the raven hypothesis, assigns a definite physical probability to the observation of the object in question. (Exception: a non-black raven will always send your subjective probability for the raven hypothesis to zero.)

To see how auxiliary hypotheses may clarify questions about confirmation, suppose that a friend sends you a black raven in the mail. How should the receipt of the bird affect your subjective probabilities for hypotheses involving raven color? That very much depends on how your friend came to decide to send you the raven. Suppose, for example, you know that your

friend decided to send you a raven selected at random. As we saw in section 6.1, this auxiliary hypothesis, in conjunction with any hypothesis of the form *The proportion of ravens that are black is x*, assigns a definite physical probability x to your receiving a black raven. Receiving a black raven, then, will confirm the raven hypothesis.

By contrast, if you know that your friend set out from the beginning to send you a black raven, and that she had the means to do so, then your subjective probability for receiving a black raven is one regardless of the hypothesis. Conjoining this auxiliary hypothesis with any hypothesis h about the proportion of black ravens yields a probability of one for receiving a black raven. The auxiliary hypothesis in effect renders the evidence irrelevant to h. (Note that the auxiliary hypothesis is inconsistent with a raven hypothesis on which there are no black ravens, for in such a world, your friend would not have the means to send you a black raven.)

Finally, suppose that your friend has decided to send you a bird randomly selected from all the black birds in the world. Does your receiving a black raven confirm the raven hypothesis? There is a temptation to surrender to a kind of Popperianism here, and to say that there can be no confirmation because the object you received in the mail could not possibly have falsified the raven hypothesis. Thus, opening the package does not constitute a real test of the hypothesis. But this is wrong. Finding a black raven in the package should, given some plausible background assumptions, increase your confidence in the raven hypothesis. As an exercise, you should explain why (and what assumptions figure in the process).[10]

The same lessons are to be applied if you receive a white shoe in the mail. Suppose that you know that your friend has sent you a randomly selected non-black object. This auxiliary hypothesis, in conjunction with the raven hypothesis, assigns a probability to your receiving a non-raven of one.

10. Begin with the question: given that you were going to be sent a black bird, what facts are relevant to the probability that it would turn out to be a *raven*?

Alternatives to the raven hypothesis that allow that some ravens are non-black assign a probability of less than one. (Only very slightly less than one, if most objects are not ravens.) Thus the raven hypothesis will receive a boost, though a very small one, from the observation of the shoe, in particular, from the observation of the fact that it is not a raven. If, by contrast, you know that your friend has sent you a random item from her closet, the probability assigned to receiving a non-black non-raven will depend only on your beliefs about the ratio of black to non-black items in your friend's (ravenless) closet, and not at all on whatever ravens hypothesis it might be joined with. All ravens hypotheses will therefore assign the same probability to the receipt of a non-black non-raven; thus, the object will not change your subjective probability distribution over the ravens hypotheses.

Given some auxiliaries, a black raven may even disconfirm the raven hypothesis. Suppose, for example, that you believe that your friend, a famous ornithologist, has a special code in which a black raven in the mail means that she has discovered that not all ravens are black. Or consider a case (due to I. J. Good) in which you believe that you are in one of two universes. In the first universe, there are a million birds, one hundred of which are ravens, and all ravens are black. In the second universe, there are a million birds, one hundred thousand of which are ravens, and 90% of ravens are black. A bird is chosen at random; it is a black raven. Because it is much more likely that a randomly chosen bird is a black raven in the second universe than it is in the first, this ought to raise your subjective probability significantly that you are in the second universe, and so lower your probability that all ravens are black accordingly.

In summary, we have provided a method for assessing the significance of any particular found object for the raven hypothesis: find auxiliary hypotheses that, in conjunction with the hypothesis, assign a definite physical probability to the evidence. Conditionalization can then proceed with the help of PCP.

This advice is simple to follow when there is one, clearly correct such auxiliary, for example, if you know exactly what sampling methods your friend is using. But you can follow it even if you are not so sure what your friend is up to, by looking to your subjective probability distribution over the different possible auxiliaries $a_1, a_2, \ldots$ conditional on h, and using the theorem of total probability to derive a likelihood based on physical probabilities:

$$C(e|h) = P_{ha_1}(e)C(a_1|h) + P_{ha_2}(e)C(a_2|h) + \cdots$$ [11]

As noted in section 6.4, the use of uncertain auxiliary hypotheses brings its own problems; this issue is discussed further in section 10.

8.2 Variety of Evidence

A theory is better confirmed, or so we think, when the evidence in its favor is varied, as opposed to all of a very similar kind. If, for example, I want to know whether all ravens are black, a sampling of ravens from Europe, Asia, and North America will provide me with much better evidence than a sampling of the same number of ravens from a small town in France. Various theorists of confirmation, Bayesians especially, have tried to give a general explanation for this aspect of confirmation.

Before I discuss the Bayesian strategies for explaining the significance of evidential diversity, let me explore the question at a more intuitive level. A variety of evidence seems useful because it puts to the test all the different aspects or parts of a hypothesis. A very narrow evidential base, by contrast,

11. It would be nice if we could assume the probabilistic indepndence of the a_i and h, so that $C(a_i|h) = C(a_i)$, but realistically, hypotheses are going to be correlated with the auxiliaries in virtue of which they deliver predictions about the data, at least once the evidence comes in. Imagine, for example, that you have received a black raven in the mail. Then your subjective probability for the raven hypothesis, conditional on your friend's having a policy of sending you a black raven just in case the raven hypothesis is false, will certainly not be the same as your unconditional subjective probability for the raven hypothesis. This point is also made in section 10.5.

seems mainly to test the truth of the hypothesis in one particular locale. Having seen many black ravens in a French town, I become very confident that the ravens in the region are all black, but my confidence that the rule of blackness is equally exceptionless in other parts of the world receives a much smaller boost. The French ravens are not irrelevant to the broader hypothesis: if all French ravens are black, that gives me some reason to think that ravens everywhere are black, but no matter how confident I become about the French ravens, I will never become quite so confident (perhaps not anywhere near so confident) about the color of ravens elsewhere.

The guiding idea, then, is that many hypotheses can be broken into parts, and that a given piece of evidence bears more strongly on the part to which it belongs than on the other parts. Or perhaps this is not quite right, since there need not be clear boundaries separating one part from another. The hypothesis *All ravens in Siberia are black* is better confirmed by a selection of black ravens from all parts of Siberia, but in saying this, I do not assume that there is a division of Siberia into particular parts separated by definite boundaries. The parts shade into one another. Nevertheless, it is convenient to talk as though they are discrete.

The question about evidential diversity, then, can be divided into the following two questions:

1. What distinguishes one part of a hypothesis from another?

2. Why do particular pieces of evidence confirm some parts of a hypothesis more than they confirm others?

You might think that the answers to these questions depend on our high level theories about the domains in question. For example, the extent to which I am inclined to generalize the blackness of French ravens to ravens in other parts of the world will depend on my high level biological beliefs. Bayesian writers have, however, tried to avoid this issue, either by incorporating the effects of high level beliefs implicitly, or by focusing on the ques-

tion as it arises in domains where there are no relevant high level beliefs, or least, where our high level beliefs are too tentative to rely on.

An example of the first strategy may be found in Paul Horwich's approach to the diversity problem. In Horwich's view, the value of diverse evidence in supporting a hypothesis lies in its ability to eliminate, or at least to strongly disconfirm, the hypothesis's competitors. Consider, for example, the geographically salient rivals of the hypothesis that all ravens are black. These might include:

1. All ravens in France are black, but those elsewhere are white,

2. All ravens in Europe are black, but those in Asia and North America are white,

3. All ravens in France are black, but outside France, the proportion of black ravens declines as the distance from Paris increases,

and so on.

All of these hypotheses predict that the ravens I observe in France are black. If I observe many black French ravens, then, all of these hypotheses will be confirmed, including the raven hypothesis itself, the claim that all ravens, everywhere, are black. Note two things. First, the total probability boost due to the French ravens must be shared out among all the aforementioned hypotheses; since Bayesian confirmation is a zero sum game, the boost received by each individual hypothesis is commensurably small. Second, since the above hypotheses all make the same predictions about French ravens, by the equal probability principle no one of them in particular can draw ahead of the others in the confirmation stakes. Thus, no matter how many French ravens I observe, the above hypotheses will always be in the running, and so I will have to entertain a certain probability that ravens somewhere else in the world are not black. In short, no number of black French ravens will come close to convincing me that all ravens, everywhere, are black.

A selection of black ravens from around the world has, by contrast, much more resolving power. It will work against hypotheses (1), (2), and (3) above as well as all the other plausible rivals to the raven hypothesis. In particular, first, the raven hypothesis will share its probability boosts with fewer rivals (and because more rivals are being disconfirmed, there is more to share in the first place), so that the raven hypothesis receives larger boosts in probability from each observed raven, and second, when sufficiently many ravens from sufficiently many places have been observed, all plausible rivals will be strongly disconfirmed, so that the probability for the raven hypothesis approaches one. A varied collection of ravens, then, will confirm the raven hypothesis much faster, and push its probability higher in absolute terms, than a less varied collection.

Horwich's view is that a similar explanation is to be given of the confirmatory virtue of diversity wherever it is to be found. The value of diversity, then, is its resolving power: diverse evidence eliminates rivals to the truth more quickly and more completely than non-diverse evidence.

This approach can be seen as an account of how a hypothesis is to be divided into parts. The parts correspond to the regions over which plausible rivals disagree. Because there is a hypothesis that all ravens are black, and another that European ravens are black but Asian ravens are not, there is a natural division of the domain of raven theories into the continents of Europe and Asia. There is no such division into, say, countries whose names begin with a vowel and those whose names do not begin with a vowel, because the disagreements of plausible rivals are not captured by domains individuated in this way.

What determines the different domains, and thus the different parts of a theory that ought to be tested, on this view, are the prior probabilities that determine what theories are plausible and what theories are not. Suppose that I assign substantive prior probabilities only to the following hypotheses about raven color:

1. All ravens are black,

2. All ravens observed on Wednesdays are black; the remainder are white, and

3. All ravens observed on Sundays are black; the remainder are white.

Then, on the Horwich approach, what would count as a diverse set of evidence is a set of ravens observed on all days of the week (or at least, a set some of which are observed on Wednesday, some on Sunday, and some on other days). This set would count as being equally diverse whether all the ravens were French or whether they were a truly multinational collection. Even a truly multinational collection would count as non-diverse, by contrast, if all specimens were observed on, say, the weekend.

You can see that the Horwich view does not lay down absolute standards for the diversity of evidence. What counts as diverse depends on the assignments of probability you make before you begin observing ravens. These need not be guesses, or a priori suppositions. You may already know quite a bit of biology, and this knowledge may be what is primarily responsible for determining which rivals to the raven hypothesis you consider plausible and which not.

There is an important lesson here, which is repeated again and again in the Bayesian literature. Bayesianism is able to implement many features we would like in a complete theory of confirmation. But the nature of the implementation is not built into BCT itself. Rather, it is determined by the assignment of prior probabilities to competing hypotheses. This raises the question as to what restrictions, if any, apply to the assignment of priors, to be taken up in section 9.

An alternative to Horwich's approach is what might be called the *correlation approach*: a set of evidence is more varied the smaller the correlation (in terms of subjective probability) between the individual pieces of evidence, or in other words, the less one of the pieces of evidence confirms

the others (Earman 1992, §3.5). The idea is as follows: if I observe a black raven in a town in France, that gives me greater reason to expect another raven from the same town to be black than it gives me reason to expect an Australian raven to be black. So the correlation between two black French ravens is greater than the correlation between a black French raven and a black Australian raven. This has immediate implications for the power of the two raven pairs to confirm the raven hypothesis. Suppose I observe a black French raven. Then I observe another black raven, either French or Australian. The confirming power of this second observation is proportional to the Bayesian multiplier, given by the formula $C(e|h)/C(e)$. The likelihood $C(e|h)$ is the same for either a French or an Australian raven (equal to one, of course). The prior probability of the evidence $C(e)$, however, is—all other things being equal—higher for the French raven, since it has been bumped up more by the observation of the first French raven. So, subsequent to the observation of one black French raven, the observation of a second French raven offers less confirming power than the observation of an Australian raven.

The correlation approach does not, I think, offer an especially helpful gloss of the notion of variety of evidence, since it does not say anything about the reasons that some kinds of data are more correlated than others. But technically, the correlation and Horwich approaches have much in common. You might like to try to show that they are formally equivalent, that is, that the question which pieces of evidence are correlated and question which pieces of evidence have greater "resolving power" depend in the same way on the same facts about the prior probability distribution. (Familiarity with the themes of section 7 will help.)

8.3 The Problem of Irrelevant Conjuncts

The hypothetico-deductive (HD) theory of confirmation suffers from the following well-known problem, which I will call the problem of *irrelevant*

conjuncts. According to the HD theory, a hypothesis is confirmed by the observation of any of its logical consequences, or more usefully, by the observation of any logical consequences of the conjunction of the hypothesis itself with the known initial conditions, or as some philosophers say, by the observation of any of the hypothesis's *predictions*. Thus on the HD theory, what I will call the *conjunction principle* is true:

If *e* confirms *h* then *e* confirms *hj*, for any hypothesis *j*.

(The reason: if, according to HD, *e* confirms *h*, then *h*, together with whatever initial conditions, must entail *e*. But then *hj* and the same initial conditions also entail *e*, and so *e* confirms *hj*.)

This looks odd. We have put no constraint on *j* at all; it might be something completely irrelevant to *h* and *e*. For example, where *h* is (what else?) the raven hypothesis, and *e* is the observation of a particular black raven, we might choose *j* to be the thesis that the Pope is infallible. Then we have the result that the observation of a black raven confirms the hypothesis that

All ravens are black and the Pope is infallible.

This is not quite disastrous; what would be disastrous is if the black raven confirmed *j* itself, that is, confirmed that the Pope is infallible.

Some writers (Hempel 1945a; Glymour 1980) have attempted to nail HD by arguing that it is a truism of confirmation that

If *e* confirms *h*, then *e* confirms any logical consequence of *h*.

Hempel called this the *special consequence* principle (it is special in that it is a restricted version of Hempel's more general *consequence principle*). If the special consequence principle is, as Hempel claimed, obviously true, then, because *j* is a consequence of *hj*, anything that confirms *hj* confirms *j*. The black raven therefore confirms the Pope's infallibility. Indeed, it confirms everything, since we have put no restriction on *j* whatsoever.

Naturally, proponents of HD reject the special consequence principle. They find themselves, then, occupying the following uneasy position: necessarily, if *e* confirms *h*, it confirms *hj*, but it does not necessarily confirm *j*. Not necessarily, but not necessarily not. There are an infinitude of confirmed conjunctions, on the HD theory, and no guidance as to which of the conjuncts are separately confirmed. What a relief it would be if there were some criterion that said, of a certain large class of choices for *j*, including hypotheses about papal infallibility, that, although a black raven confirms *hj*, it does not confirm *j*.

Enter BCT. Bayesianism shares with HD a commitment to the conjunction principle, more or less, for the reason that a piece of evidence *e* that raises the probability of a hypothesis *h* will also raise the probability of *hj*, for almost any *hj*. (The exceptions are cases where *e* disconfirms *j* as well as confirming *h*, or where *h* itself bears negatively on *j*.)

As with HD, this poses the question whether *e*, in confirming *hj*, also confirms *j*. Unlike HD, Bayesian confirmation theory provides an apparently straightforward way to answer the question: perform the calculations, and see if the probability of *j* goes up along with that of *hj*.

It is straightforward, indeed, to do the calculations for any particular choice of *h*, *j*, *e*, and background knowledge. But what we would like is a more general result, to the effect that, for such-and-such a kind of conjunct, the probability increase will depend on such-and-such factors.

In the case at hand, the general result would concern irrelevant conjuncts. Let us proceed. One way of capturing the notion of irrelevance in the probabilistic language of BCT is as follows:

> A hypothesis *j* is irrelevant to *h* and *e* if *j* is probabilistically independent of *h* and *e* (and, you should probably add, of *he*).

From the irrelevance of *j* to *e* it follows immediately that $C(j|e)$ is equal to $C(j)$, and so that the observation of *e* does not confirm *j*. But I think that this provides much insight into the question why irrelevant conjuncts are

not confirmed. To assume that *j* is irrelevant to *e* just is to assume that *e* does not confirm *j*. This was the relation that was supposed to be explained.

A more helpful result would show that, if *j* is irrelevant to *h*, then an *e* that confirms *h* does not also confirm *j*. This is not a theorem of the probability calculus, but I conjecture that it is almost true, in the sense that, for almost any correlation between *h* and *e*, it is possible for *j* to be irrelevant to *h* only if *e* is irrelevant to *j*. If this is correct, we can conclude that most irrelevant conjuncts are not confirmed (and in the remainder of cases, we can at least conclude, with the HD theorists, that the irrelevant conjuncts are not *necessarily* confirmed).

9. The Subjectivity of Bayesian Confirmation Theory

9.1 The Problem of Subjectivity

The discussion of induction in section 7 posed a question that goes far beyond the search for an a priori justification for inductive reasoning. It is the question as to which aspects of a Bayesian reasoner's inferences are fully determined by the Bayesian machinery, and so are the same for all Bayesian reasoners, and which aspects are determined in part by the allocation of the prior probabilities. As we saw, although BCT is capable of accommodating much (even all) of our inductive behavior, it *merely* accommodates it: there is nothing about being a Bayesian that makes this behavior any more natural than various alternative behaviors, such as expecting the future to break radically from the past, or preferring more complex to simpler hypotheses. This worries even philosophers who long ago abandoned the search for an a priori justification for inductive practices. Let me explain the source of the worry.

The scientific enterprise requires a certain amount of agreement among scientists as to how the evidence bears on various competing hypotheses. The agreement need not be absolute, but without some common ground

as to what data supports what theories, it seems, there can be no scientific progress.

Many philosophers worry that BCT does not provide enough common ground, that is, that scientists who learn by Bayesian conditionalization may disagree on sufficiently many important questions that the consensus required for scientific progress is undermined.

We saw in section 7 that BCT enforces the likelihood lover's principle, but that LLP alone is not a sufficiently strong constraint to force typical inductive behaviors—not even an expectation that the future will resemble the past. In this section, I will focus not on broad inductive behaviors, but on more specific questions about the impact of evidence on individual hypotheses.

Let me begin with the bad news. Scientists using Bayesian conditionalization, and who are therefore committed to the likelihood lover's principle, may disagree about any of the following matters.

1. Which of several competing theories is most likely to be true, given a certain body of evidence.

2. Which of several competing theories received the greatest boost in its probability from a given piece of evidence, where the boost is the difference between the relevant prior and the posterior probabilities. (The LLP, though, completely fixes the relative size of the Bayesian multipliers for competing theories; scientists will agree as to which theory had the highest Bayesian multiplier: it is the theory with the highest likelihood on the evidence.)

3. Whether a theory's probability ought to increase or decrease given—whether it is confirmed or disconfirmed by—a particular piece of evidence.

This last possibility for disagreement is particularly dismaying. To see how the disagreement may arise, recall that a hypothesis h's probability increases on the observation of e just in case the relevant Bayesian multiplier

is greater than one, which is to say, just in case the physical probability that h assigns to e is greater than the unconditional subjective probability for e; in symbols, just in case $P_h(e) > C(e)$. Whether this is so depends on a scientist's probability for e, that is, on $C(e)$, which depends in turn—by way of the theorem of total probability—on the scientist's prior probability distribution over the hypotheses. Scientists with different priors will have different values for $C(e)$. Even though they agree on the likelihood of a hypothesis h, then, they may disagree as to whether the likelihood is greater or less than $C(e)$, and so they may disagree as to whether h is confirmed or disconfirmed by e. (There is not complete anarchy, however: if two scientists agree that h is confirmed by e, they will agree that every hypothesis with a greater physical likelihood than h—every hypothesis that assigns a higher physical probability to e than h—is confirmed.) How can there be scientific progress if scientists fail to agree even on whether a piece of evidence confirms or disconfirms a given hypothesis? We will explore various answers to this question in what follows.

Before I continue, let me say something about BCT's rival, the classical view of statistical inference. The classical view is as solicitous of scientific consensus as the Bayesian view is indifferent. The aim of the classical view is to set up a system for assessing the significance of the data in which there is absolute epistemic unanimity. Everyone who agrees on the data, and assents to the system, will agree on the impact of the data on the relevant hypotheses.

The price of consensus is that the conclusions that are licensed by the system are often weak, as when a null hypothesis is rejected, or even nonexistent, as when a result is not "statistically significant". (For a Bayesian, there is no such thing as "statistically insignificant" data.) It is a commonplace that scientists supposedly working under the aegis of classical statistics frequently reach conclusions that are stronger than anything officially licensed by the system, and that their daring seems to advance, rather than

to hold back, the progress of science. This (along with many other considerations) suggests that classical statistics is too conservative in its judgments of evidential impact to capture the full range of inductive behavior that makes science so successful.

9.2 Washing Out and Convergence

The most common and in many ways the most effective Bayesian response to the subjectivity objection is the convergence response: although in the short term scientists may disagree on the significance of the evidence, in the longer term their subjective probabilities will converge on the same hypotheses, and so a consensus will emerge. The differences in scientists' priors will be *washed out* over time, that is, they will become less and less relevant to scientists' opinions as more and more evidence comes in. Scientific progress may not be easy to see on a day by day basis, but with the perspective of decades or centuries, everyone, the argument goes, can agree on what the evidence has shown and what it has not shown.

The foundations of the convergence response are certain mathematical results showing that, with very high probability, opinions will indeed converge. These convergence results can be divided, on philosophical as well as mathematical grounds, into two classes. The first class assumes that the subjective likelihoods used in conditionalization are set according to the probability coordination principle, and so correspond to physical probabilities, and derive a high physical probability of convergence. The second class does not make this assumption; the likelihoods are not objectively constrained, and therefore may differ from scientist to scientist and from the corresponding physical probabilities (if there are such). The conclusion is weaker: each person will (or should) have a high subjective probability their beliefs, and the beliefs of other scientists, will converge on the truth. Because I have focused exclusively on a version of BCT in which PCP is used at all times, by invoking auxiliary hypotheses wherever possible, I will focus in what follows

on the first kind of result; a brief discussion of the second kind of result may be found at the end of this section.

Suppose we have a set of competing hypotheses and a group of scientists with different prior probability distributions over the hypotheses. We assume that no hypothesis is ruled out of contention in the priors; in the simple case where there are a finite number of competing hypotheses, this means that no hypothesis is assigned probability zero by any scientist. (You should be able to see immediately why this assumption is necessary for a convergence result.) Then we can show the following: as more and more data comes in, the scientists' subjective probability distributions will, with very high physical probability, converge, and this convergence will be (again with very high physical probability) on the true hypothesis.

There are a number of caveats to this result, but let us first try to understand how it works. The three most important facts bearing on convergence are the following:

1. Bayesian conditionalizers favor the hypotheses with the highest subjective likelihoods,

2. Conditionalizers set the subjective likelihoods equal to the physical likelihoods, thus, they agree on which hypotheses should be favored and on the relative magnitude with which one hypothesis should be favored over another (i.e., the ratio of the Bayesian multipliers, equal to the ratio of the physical likelihoods), and

3. The more evidence that comes in, the lower the physical likelihood of the false hypotheses on all the evidence relative to that of the true hypothesis (with very high probability).

To help you to appreciate this last fact, take a look at the graph of the physical likelihoods in the ravens case. As in section 6.3, I assume a set of competing hypotheses each of which posits a different real number between zero and one as the physical probability of a given raven's being black; I

also assume that, at any time, two-thirds of the observed ravens have in fact turned out to be black. The physical likelihoods for different numbers n of observed ravens—the physical probabilities ascribed by the hypotheses to the event that two-thirds of a set of n ravens are black—are shown in figure 5. As you will see, as n increases, the likelihood for any hypothesis that is not very close to the hypothesis that attributes a two-thirds probability for raven blackness approaches zero. If this is indeed the correct hypothesis, then the likelihood of any hypothesis not very close to the truth approaches zero.

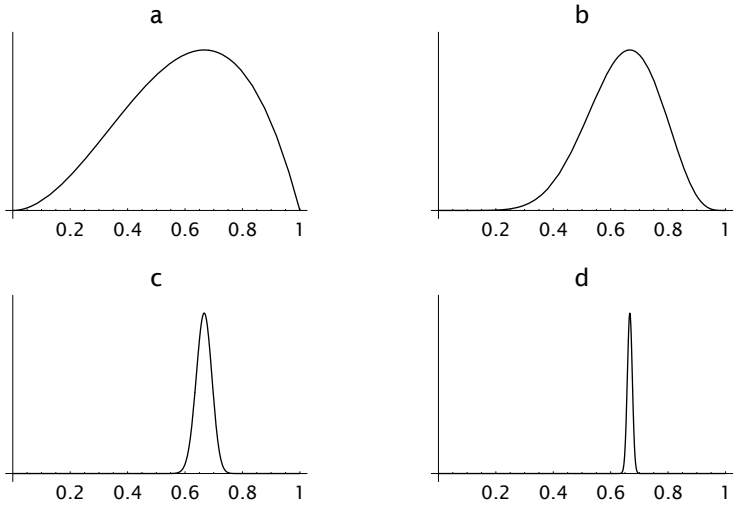

Figure 5: Physical likelihoods for four data sets, consisting respectively of 3, 12, 300 and 3000 ravens, in each of which two-thirds of ravens are black. Heights are normalized. (Drawn to scale, the spike in (d) would much lower than the hump in (a)—it is highly unlikely on any of the hypotheses that *exactly* 2000 out of 3000 ravens will be black.)

The cumulative effect of many individual pieces of evidence is, because of the commutativity principle (section 6.2), equivalent to the effect that the evidence would have if it were all revealed at once. We can therefore understand the convergence result by thinking about the effect of observing, in one single ornithological swoop, the color of vast numbers of ravens.

Let e be the observation of n ravens, then, for some large n. As I have noted, when n is large, the true hypothesis and a small group of other hypotheses near the true hypothesis assign a probability to e that is much higher than that assigned by the rest. The Bayesian multiplier for hypotheses near the truth, then, will be much, much greater than the Bayesian multiplier for the remainder of the hypotheses. Since the probabilities of all the hypotheses must sum to one, this will result in a posterior probability distribution that is heaped almost entirely around the truth.

For your edification, the evolution of three rather different prior probability distributions over the ravens hypotheses is shown in figure 6, under the assumption that, at each stage, exactly two thirds of all observed ravens have been black.

How did we get past the fact that scientists with different prior probabilities may disagree about the confirmatory power of evidence, and in particular, about whether a given piece of evidence confirms or disconfirms a hypothesis? The problem, recall, was that whether a piece of evidence e confirms or disconfirms a hypothesis h depends not only on the physical likelihood of h—that is, the physical probability that h ascribes to e—but also on the prior probability for e, which depends in turn on the prior probability distribution over h and its rivals. If the likelihood is greater than the probability of e, then h is confirmed, if it is less than the probability of e, then h is disconfirmed.

All of this continues to be true when e comprises a vast number of individual data points. But in this case, the likelihoods are either relatively high, for hypotheses that are near the truth, or relatively negligible, for every other hypothesis. The prior probability of e will be somewhere in between (because it is a weighted average of the likelihoods). Thus, although the prior probability for e will vary from scientist to scientist, all scientists will agree that e confirms a small group of hypotheses, those that are in fact near the truth, and that it disconfirms the rest.

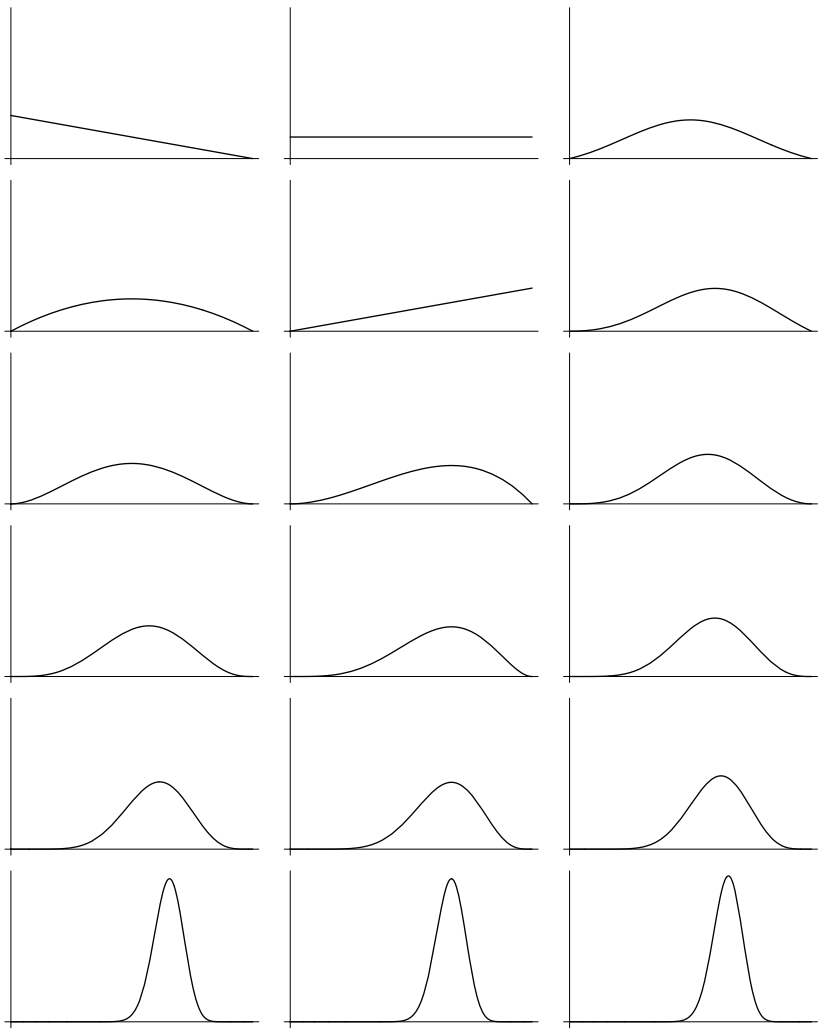

Figure 6: Priors are washed out as the data set becomes larger and larger. From left to right are three different prior probability distributions; from top to bottom the distributions change as more evidence comes in. At any given time, about two thirds of the ravens are black. The six data sets, from top to bottom contain zero ravens (i.e., the unconditioned priors are shown), one raven, three ravens, six ravens, twelve ravens, and sixty ravens.

Now let me say something about the limitations of this result. First, as you already know, no matter how much evidence comes in, Bayesian conditionalization cannot distinguish between two hypotheses each of which assigns exactly the same physical probability to the evidence. No convergence result can guarantee convergence on the truth, then; at best, what you get is convergence on the true hypothesis and any other hypotheses empirically equivalent to the true hypothesis.

Second, convergence on the truth is with high probability only. There will always be some chance that scientists converge on a false hypothesis, or that they fail to converge at all. You can see this quite easily by thinking of the case of the ravens. Suppose that the truth is that a given raven has a two thirds probability of being black. It is possible that, after observing a million ravens, exactly one third have turned out to be black—extremely improbable, but possible (see tech box 6.2). Obviously, if this did happen, scientific opinion would not converge on the true hypothesis. As the example suggests, however, the probability of convergence on the truth is very high, certainly high enough to satisfy most commentators.

Third, the key premise required for convergence, that as more evidence arrives, the contrast between the likelihood for the true hypothesis and the likelihoods for the false hypotheses increases, depends on the physical probability distributions posited by both the true and false hypotheses taking a certain form. Roughly speaking, they must be the kind of distributions to which the law of large numbers (tech box 6.2), or certain of its analogs or extensions, applies.[12]

In theory, this is a great limitation. In practice, just about every scientific hypothesis about physical probabilities, past and present, has had the right

12. More specifically, you need the probability distribution to have two properties relative to the evidence. First, there should be some kind of "long run" result that says that, given a certain distribution, the probability of the evidence having a certain statistical profile tends, as more evidence comes in, to 1. Second, the probabilified statistical profile should be different for different distributions.

sort of form. In Bayesian terms, your average scientist's prior probability for the vast set of "wrong" hypotheses tends to be just about zero.

Fourth, the explanation of convergence turned on the fact that, on the observation of a vast body of evidence, the Bayesian multiplier for hypotheses near the truth is much, much higher than the Bayesian multiplier for the remainder of the hypotheses. You may well have noticed that this in itself will not necessarily get us convergence, if some scientists have a prior probability for the true hypothesis and its neighbors that is much, much lower than their prior probability for the remainder of the hypotheses. These scientists have a very strong prior bias, then, against the hypothesis that is in fact true, and others like it.

The more evidence that comes in, the more biased against the true hypothesis a scientist will have to be not to be swayed by the evidence in the same way as all the other scientists.[13] For any given level of bias against the actual truth, there is some quantity of evidence that is sufficient to heap the biased scientist's subjective probability distribution around the truth. But equally, for any given quantity of evidence, there is some level of bias sufficient to resist the pull of the evidence towards the truth.

Fifth, I have been using the notion of hypotheses that are "near the truth" rather loosely. As you will see from the treatment above, in the context of a convergence argument, a hypothesis "near" the truth is one that assigns about the same physical probability to any given set of evidence as the truth does. In a simple case like that of the competing ravens hypotheses, hypotheses that are near the true hypothesis in this sense are also near the truth in another sense: they are neighbors in parameter space, meaning that

13. There is an additional factor working to push these scientists towards the truth: they will have a lower prior probability for the evidence than less biased scientists, and so their Bayesian multipliers for the near true hypotheses—the physical likelihoods of those hypotheses divided by the probability of the evidence—will be even higher than their fellows'. You can see from figure 6, however, that this additional factor is not enough in itself to erase entirely the effect of the bias.

they are, for example, physically adjacent in a probability density graph of the sort pictured in figure 6. This has the very convenient consequence that a probability density is concentrated on those hypotheses that are near the truth just in case it is heaped around the truth in the geometric sense. You should bear in mind that things may not always turn out this nicely.

Sixth, I have so far ignored the question of auxiliary hypotheses, that is, hypotheses that are needed in order for the main hypotheses to assign definite physical probabilities to the evidence. How will the introduction of auxiliary hypotheses affect the convergence result?

Because it is the combination of a hypothesis h and an auxiliary hypothesis a that ascribes a physical probability to a piece of evidence, the convergence result will apply not to the probability distribution over competing hypotheses, but to the distribution over what might be called competing models, where each model is a conjunction of a hypothesis and an auxiliary hypothesis. What we can conclude, then, is that, as the evidence comes in, scientists' subjective probabilities will converge on the true model and other models that assign similar physical probabilities to the evidence as the true model.

Suppose that we are in the fortunate situation in which no plausible rival to the true model—the true hypothesis h conjoined with whatever true auxiliary hypothesis a is need to extract physical probabilities from h—is empirically equivalent to the true model. Then the convergence result tells us that scientists' opinion will converge on the true model ha. If there are only finitely many competing models, the probability of ha will be driven to near one (scientists will continue to disagree just how near one). Since the probability of each conjunct must be at least as great as the probability of the conjunction, it follows that each scientist ascribes a probability near one to both h and a. We have a conclusion, then, that is just as strong as in the case where auxiliary hypotheses were not required: scientists will eventually agree on the true h. What if there are infinitely many competitors? Then the

probability converges on the true model *ha* and on nearby models. We may hope that whenever two models h_1a_1 and h_2a_2 are nearby in this sense, h_1 is near to h_2 in some useful sense.

It is, however, rather unlikely that there no rivals to the true model. Typically, there will be one or more theories that paint a very different picture of the world than the true theory, but which, when supplemented with the right auxiliaries, deliver the same predictions as the true theory about the outcomes of some set of experiments. The most you can expect from the experiments in a case such as this is convergence of opinion on the empirically equivalent models. (The techniques for dealing with auxiliary hypotheses to be discussed in section 10 are of some use in distinguishing rival models; however, because they draw on the investigators' prior probabilities, they cannot themselves secure any further convergence of opinion.)

Seventh and finally, the convergence result assumes that the true hypothesis has already been formulated and so is among the known alternatives at the beginning of the investigation. No mere convergence result can guarantee that scientists will construct the true theory in response to the evidence if they have not already done so. Thus the convergence result should be seen, not as a guarantee of scientific progress *tout court*, but only as a guarantee (with the appropriate qualifications) that initial differences in subjective probability distributions will not significantly impede progress.

Let me conclude with a positive comment about convergence. You will find in the literature two kinds of convergence results, which might be called eventual convergence results and constant convergence results. An eventual convergence result says that there exists some point in the future at which a certain amount of convergence will have (with high probability) occurred, without saying how far in the future that point lies. A constant convergence result says that convergence is occurring (with high probability) at all times. In other words, pick a time span, and there is a high probability that over that span, the subjective probability distribution has converged towards, rather

than diverged from or moved no nearer to, the truth.

On the question of consensus and progress, then, an eventual convergence result says that there is some point in the future, perhaps very far away, at which there will be a certain amount of consensus, and therefore, if the convergence is on the truth, a certain amount of progress. *Nothing whatsoever* is guaranteed until that point is reached. A constant convergence result says (with the usual probabilistic qualification) that consensus is increasing all the time, and therefore, if the convergence is on the truth, that science is progressing all the time.

This difference is especially important when it comes to progress. If the best we have is an eventual convergence result, we can never know, no matter how much consensus there is, that we have arrived at or near the truth. I do not mean merely that we cannot be absolutely sure; convergence results never give us absolute certainty, because of the probabilistic rider. What I mean is that we never know at all: we know that there exists a point at which there will have been progress, but never that we have reached that point. For all we know, conditionalization has carried us further away from the truth than ever before (albeit temporarily); we can not even say that this alarming possibility is unlikely. A constant convergence result, by contrast, tells us that we are always, with high probability, progressing, *a fortiori*, we are always at a point where progress has likely occurred. The good news: the above convergence result is a constant convergence result.

A few words about the subjective convergence results that do not assume PCP, thus do not assume that different scientists' subjective likelihoods are dragooned into agreement by (hypothesized) physical probabilities.

You might think that when subjective likelihoods are allowed to wander freely (or as freely as the axioms of probability allow), convergence is a pipe dream. Suppose I believe that if the benevolent, omniscient Christian god existed, it is very likely that our life on earth would be paradisiacal, a string of long summer days and nothing to do but to feast on the abundant low-

hanging fruit. You may think that if such a god existed, our life on earth would likely be a grim test of faith, overflowing with hunger, disease, warfare, and suburban angst. Looking around, I see endless evidence against this god's existence; you see the same evidence, but it confirms your belief in the same god. My subjective probability of the evidence conditional on the hypothesis of existence is low; yours is high. How can we ever agree? The observation of further catastrophes can only widen the divide. As more evidence comes in, our theological beliefs must diverge.

It is rather surprising, then, to learn of the existence of theorems that guarantee the convergence of opinion on the truth. These results guarantee (with high probability, that is) that whatever our priors, and whatever our subjective probabilities, we will eventually agree on everything, and the everything we agree on will be exactly right. The influx of evidence, then, washes out not only differences in unconditional priors, but differences in subjective likelihoods.

What kind of magic is this? There is a whiff of Oz, I am afraid: the subjective convergence theorems turn out to require some version of what Earman calls an assumption of "observational distinguishability", according to which there is, for any pair of competing hypotheses, a piece of evidence that will decide between them on deductive grounds. That is, for any two competing hypotheses, there must be some piece of evidence that is implied to be true by one, and implied to be false by the other (Earman 1992, 167).

Now you can see, very roughly, how the mathematics works. It does an end-run around the many significant differences in subjective likelihoods altogether, relying on the small measure of agreement concerning subjective likelihoods that exists, because of the probability calculus, in the special case where a hypothesis entails the evidence: if h entails e, then all scientists must set their subjective likelihood $C(e|h)$ equal to one. As the evidence comes in, the "observational distinguishers" do their work by way of these likelihoods, ruling out ever more alternatives to the truth. (This is not to deny that it

is a mathematical feat to have shown that there is *convergence* on the truth as the evidence comes in; the rough sketch in this paragraph does not come anywhere near to establishing this conclusion. Convergence is, however, not *constant* in the sense defined just above the divider.)

The subjective convergence results are, then, less exciting than they first appear. Another complaint may be lodged against them: what they prove is that each scientist's *subjective probability* for convergence on the truth should be very high (and should approach one as the quantity of evidence increases). This gives them what Earman calls a "narcissistic character" (Earman 1992, 147). Is there anything to worry about here? Nothing terribly important, I think. The probability may be subjective, but it is derived from the fact (well, the assumption) that there is a complete set of "observational distinguishers" out there capable of pointing to the truth in the most objective possible way, by ruling out deductively the untrue hypotheses. It is, then, one of those objectively-based subjective probabilities that give modern Bayesianism its satisfyingly normative edge.

9.3 Radical Personalism

Under the heading of *radical personalism* I class any attempt to embrace, rather than to wriggle out of, the subjective aspect of BCT. A radical personalist admits that many properties of BCT are determined by the way in which the priors are set, but argues that this is acceptable, or even good.

Let me briefly describe two lines of thought in the radical personalist vein. The first addresses the fear that the prior probabilities that shape a Bayesian's inductive inferences are without ground. Meditation on the fact of this groundlessness may give rise to a Hamlet-like indecision, in which the Bayesian reasoner is unable to summon up the will to conditionalize on the evidence, because he knows that the conclusions drawn will owe so much to his subjective, therefore groundless, priors. Radical personalists urge the reasoner to have "the courage of his convictions", that is, to reason on the

9.1 Why Radical?

Is there such a thing as moderate personalism? Yes; all Bayesians are called personalists: think of *personal probability* as being a synonym for *subjective probability*. In the same way, Bayesians are often called *subjectivists*.

There is, however, a radical air to the writings of many radical personalists (Howson being a notable exception). These writers see their work as in part revisionary: rather than capturing our pre-existing intuitions about confirmation, BCT is seen as a corrective to those intuitions. The a priori arguments for BCT are supposed to lead one to embrace Bayesianism; the subjective elements of Bayesianism are then to be understood as discoveries about the true nature of confirmation. If BCT turns out to be more subjective than we had expected a theory of confirmation to be, then so much the worse for our expectations.

basis of the priors just because they are *his* priors and they therefore reflect his complete present judgment as to how things stand.

Of course, if subjective probabilities are defined as dispositions to act in certain ways in certain circumstances, then this exhortation comes too late: the waverer, by wavering, no longer determinately has the subjective probabilities that he once did. (Maybe the right sort of therapy, say a thespian event, could restore the disposition to act and so the probabilities, but it is hard to know how to persuade a reasoner to *want* to restore the dispositions/probabilities, unless by way of, perhaps, a supernatural experience.)

If you do succeed in finding the courage of your convictions, a certain kind of intellectual paralysis is avoided. But whatever solace lies in action is not enough, I think, to quell the worry about scientific consensus with which I introduced the problem of subjectivity. That scientists conditionalize swiftly and unequivocally is no doubt, on the whole, a good thing, but it is insufficiently good in itself to establish a consensus, or even the probability of a consensus. As we saw in sections 7.2 and 7.3, Professors Rosencrantz

and Guildenstern may agree on all the evidence—say, on all the outcomes of all the coin tosses they have seen so far—yet, if their priors differ in a certain way, they may disagree as much as you like on the most probable outcome of the next coin toss.

The second line of thought proposes that BCT is a less ambitious undertaking than these notes have so far supposed. Howson (2001) argues that we ought to regard BCT not as a system of inductive reasoning, but as a framework for implementing a system of inductive reasoning. Bayesianism is, on this view, not itself a theory of confirmation, but rather a meta-theory of confirmation, that is, a theory of what theories of confirmation should look like. Under Bayesianism's tolerant rule, many different approaches to confirmation may flourish together in friendly competition. These theories will differ in what Howson calls their *inductive commitments*.

An example of an inductive commitment is the principle of the uniformity of nature. As we saw in section 7.2, you can, while conforming entirely to all the strictures of BCT, either reason in accordance with or in violation of the uniformity principle. Howson's thesis is that Bayesianism shows you how to do either—it is your choice as to which—in a consistent manner. Bayesianism does not supply inductive commitments, then; it tells you how to apply whatever inductive commitments you happen to fancy, *consistently*.[14] Other writers have talked of Bayesianism as being purely a theory of *coherence*, meaning by this more or less the same thing as Howson.

Howson compares the rules imposed by Bayesianism to the rules of deductive logic. Both sets of rules, Howson claims, are intended to keep the reasoner from a certain kind of inconsistency: deductive logic protects you from contradiction, while Bayesian reasoning protects you from the kind of inconsistency that is exposed by Dutch book arguments. There is at least

14. Readers familiar with Carnap's system of inductive logic will note a parallel: Carnap provides a framework with a single parameter λ which, depending on its value, results in a wide range of inductive behaviors, or if you like, in a wide range of inductive logics (Carnap 1950).

one disanalogy, however: whereas all disciples of deductive logic will reach the same deductive conclusions given the same evidence (insofar as any deductive conclusion can be drawn at all), Bayesians will not reach the same inductive conclusions given the same evidence. It seems reasonable to say that all deductive reasoners share the same deductive commitments; by contrast, as we have seen, even by Howson's own lights, not all Bayesians share the same inductive commitments.

Howson's reply is that your priors, and thus all the inductive commitments implicit in your priors, should be considered on a par with the premises in a deductive argument. Then all Bayesians will reach the same conclusions given the same premises, because the premises will include the inductive commitments. (The "premises", on this view of things, extend far beyond what we have been calling the evidence.)

This is, I suggest, tendentious. Certainly, it seems reasonable to equate some aspects of the priors with the premises of a deductive argument—perhaps, for example, the ratio of my probability for *All ravens are black* to that for *50% of ravens are black*. But other aspects of the priors—I am thinking, of course, of those aspects that bear directly on a reasoner's inductive commitments—seem to have no parallel in deductive logic's premises. An example, I think, would be the ratio of my probability for *All ravens are black* to that for *All ravens observed before the year 3000 are black, the rest white*.

Howson, no doubt, would point to the similarity between these two examples as a reason to think my distinction invidious. But in making use of the notion of an inductive commitment, he himself acknowledges that the ratio in the second example gets at something inductively deeper than that in the first—something that belongs to the subject matter of inductive logic itself, and that is therefore more than a mere premise.

In summary, the arguments of the radical personalists, though they do convey some insights into the nature of BCT, do not—and really, could not—

assuage the worry that BCT provides no ground for scientific consensus, let alone scientific progress. At best, the radical personalist can persuade us that no such grounds are needed, or at least, that none are available. As you might expect, this has not satisfied those who see Bayesianism as one among several viable approaches to confirmation, rather than as the one true religion.

9.4 Constraining the Priors

Varieties of Constraint

So far we have assumed that Bayesian reasoners are free to set their priors in any way that they like, subject only to the axioms of probability. But surely some assignments of priors are more reasonable than others? Or at the very least, surely a few are downright pathological? If so, then we can add to the Bayesian cocktail a new ingredient, some sort of objective constraint on the way in which the priors may be set. There are many ways to concoct this particular mixture, some involving just a dash and some a healthy dollop of constraint; there are at least as many justifications for the constraint, whatever it may be. We will survey some of the more popular suggestions.

Begin with the kinds of justification you might advance for your preferred constraint. Here are three flavors, arranged in order of decreasing Bayesian purity.

1. The argument for the constraint may be a putatively rigorous a priori argument, that is, it may have the same virtues as the arguments for the other parts of BCT.

2. The constraints may be constructed so as to enhance the power of the convergence argument, that is, they may be justified on the grounds that they make convergence to the truth faster and more likely.

3. In real life, by contrast with the Bayesian fairy tale, priors are not set

in a vacuum. The third approach to constraining the priors takes into account the milieu in which the priors are set in such a way as to influence their setting.

Of these three approaches, constraints derived in the first way tend to be very strong, those derived in the second way rather weak, and those derived in the third way range from weak to strong.

The A Priori Purist Approach

The classic a priori constraint on prior probabilities is the *principle of indifference* (sometimes called the *principle of insufficient reason*), which enjoins you to set equal probabilities for mutually exclusive alternatives concerning which you have no distinguishing information. In the case of the ravens, for example, you should assign the same probability to each hypothesis of the form *x% of ravens are black*, which is to say, you should assign a uniform (i.e., flat) probability density over the different values of x, resulting in the density shown in figure 2a.

> *9.2 Origins of the Principle of Indifference*
> The principle of indifference is closely associated with Leibniz (1646–1716) and Laplace (1749–1847), and through them with what is called the classical interpretation of probability. On the classical interpretation, indifference plays a part in the very definition of probability. The definition of subjective or physical probability, you ask? Neither, exactly: the classical interpretation fails to clearly distinguish between these two kinds of probability (Hacking 1975). The standard view among modern proponents of the principle is that, while it plays no role in determining physical probabilities, it is an important constraint on subjective probabilities. (In the earlier twentieth century, it was seen as a constraint on what is called logical probability; see tech box 5.2.)

What does it mean to say that you have *no distinguishing information*

about two alternatives? It means that the alternatives are relevantly similar, or that they differ only in a respect which is epistemically irrelevant. (If you see problems looming, you see clearly.) The justification for the principle is that, if there is no relevant difference between two outcomes or hypotheses, then you have no reason ("insufficient reason") to expect one over the other, and so you should assign them equal probabilities.

If the principle of indifference functioned as it is supposed to, then everyone would have the same prior probability distribution for, say, the ravens hypotheses. Thus, provided that everyone saw the same evidence and conditionalized on it in the Bayesian way, everyone would have the same probability distribution over the ravens hypotheses at any later time. This provides all the scientific consensus one might want, and perhaps more than one might reasonably want. The convergence results are no longer needed to promise consensus (though they are needed to promise that the consensus will likely move towards the truth).

In the later part of the nineteenth century,[15] decisive arguments against the principle of indifference in its initial form were formulated, notably by Joseph Bertrand. The problem is that the principle does not dictate, as it claims to, a unique prior probability distribution over a set of hypotheses. To see how *Bertrand's paradox* undercuts the principle, consider the following example of van Fraassen's. There is a factory that produces cubes with sides of a length between one and three centimeters. You must choose a prior probability distribution over the competing hypotheses of the form *The next cube produced by the factory will have a side of length x cm*. It seems that the indifference principle commands you, in the absence of any information distinguishing the different hypotheses, to put a uniform probability over the side length x. As a result, your subjective probability that, say, the next cube has a side of length less than 2 cm will be 0.5.

15. By which time, perhaps relevantly, the classical interpretation of probability (see tech box 9.2) was already dead.

But now consider the competing hypotheses of the form *The next cube produced by the factory will have a volume of v cm*^{3}. Each of these is logically equivalent to a hypothesis about side length; for example, the hypothesis that the volume of the next cube will be 8 cm^3 is equivalent to the hypothesis that the side length of the next cube will be 2 cm. The principle of indifference, as applied to this set of hypotheses, commands us, it would seem, to put a uniform distribution over v. But a uniform probability distribution over v is a different distribution than a uniform probability distribution over x. As a result of a uniform distribution over v, for example, the probability that a cube has a side of length less than 2 cm is the probability that it has a volume of less that 8 cm^3, which, since the volumes range from 1 cm^3 (i.e., 1^3) to 27 cm^3 (i.e., 3^3), is 7/26, or about 0.27. Depending on how we formulate the competing hypotheses about the next cube, then, the principle of indifference seems to offer conflicting advice: on the one hand, to set the probability of a cube of side length less than 2 cm to 0.5, but on the other hand, using the volume formulation, to set the probability for the same outcome to 0.27.

The principle of indifference also offers no help with grue; in fact, it makes the grue problem as bad as it could possibly be. Consider the rival hypotheses of the form *All ravens observed before time t are black, and all others are non-black* for the range of times t extending from now to infinity. (When $t = \infty$, the hypothesis becomes the raven hypothesis *All ravens are black*.) In a state of total ignorance, it seems that we should put a uniform probability distribution over t. But then, even if all ravens observed until now have been black, we will assign a negligible probability to the raven hypothesis, and a probability of effectively one that, at some time in the future, we will begin to observe non-black ravens.[16] The principle of indifference, then, appears to force us to take seriously grueish hypotheses that, if our priors were unconstrained, we would have the freedom (though not an obli-

16. In this simple example we will, however, always expect the next raven to be black.

gation) to ignore.

Proponents of the principle of indifference have reacted to Bertrand's paradox and to other complaints by refining the principle so that it delivers univocal judgments about the proper probability distribution in as many cases as possible. They have had less to say about the grue problem. Most Bayesians have given up on the indifference approach to constraining priors as both technically infeasible and overweening, the latter in the sense that it seeks to constrain the priors far too much, eliminating one of BCT's most attractive features, its ability to accommodate diversity of opinion.

The Convergence Approach

Suppose that the best defense against the subjectivity inherent in BCT is the convergence argument. Then you have an ambivalent attitude to your priors. On the one hand, they reflect your feeling about the plausibility of various different hypotheses, and a commitment to act on that feeling. On the other hand, your reason for adopting the Bayesian machinery, and so taking the priors seriously in the first place, is that they will, ultimately, have no influence on your views. Given such a mindset, it might be reasonable to adopt a set of priors not because they reflect your judgments about various hypotheses' plausibility, but because they are especially conducive to convergence.

Consider, for example, the following very modest proposal. A hypothesis that is assigned a prior probability of zero is doomed always to have a probability of zero. Likewise, in the case of infinitely many hypotheses, a region of hypotheses that is assigned a prior probability of zero—a region where the prior probability density is zero—will always have zero probability. Consequently, if you are so unfortunate as to assign a prior of zero to the true hypothesis, or in the infinite case, to the region containing the true hypothesis, your subjective probabilities can never converge on the truth.

If convergence on the truth is the paramount Bayesian imperative, then,

it seems that you should do your best to avoid assigning a prior of zero to the truth. The only way to be sure of doing this, it seems, is to assign non-zero priors to every hypothesis, or in the infinite case, to assign a prior probability density that never touches zero. This policy of not ruling out any hypothesis altogether from the start is sometimes called *open-mindedness*. You need not care about convergence to be open-minded, of course, but if you care about convergence, it seems that you ought to be open-minded.

How open-minded ought you to be about hypotheses involving grue and its brethren? You ought to assign them a probability greater than zero—in the infinite case, a probability density that does not touch zero—but you may otherwise discriminate against them as much as you like. For example, you might assign your priors over the various raven hypotheses, grueish and non-grueish, so that 99% goes to the non-grueish hypotheses of form *x% of ravens are black* and only 1% to all the grueish hypotheses (and there are a lot of them!) combined. This gives the grueish hypotheses enough of a toehold that, in the event that ravens observed after, say, 3000 are all white, your subjective probabilities can converge on the appropriate grueish options.

But why stop there? The more you discriminate against the hypothesis that turns out to be true, the longer it will take for your probabilities to converge on the truth (though you will have to be highly discriminatory to hold out against the evidence for long). So why not, as well as being open-minded, be somewhat *fair-minded*, that is, why not assign a prior probability distribution that does not discriminate too much against any particular group of hypotheses?

Fair-mindedness will have to have its limits. We most likely want to allow continued discrimination against grueish hypotheses. But this can be done while still allowing grueish hypotheses enough of a chance that convergence to such hypotheses, if we start to observe white ravens, is adequately fast. (*Adequately* here means about as fast as you would want to converge on the grueish options, in the circumstances.) It takes only a little fairness to con-

verge on the truth in a tolerably short time. Even the bigoted can be saved, if they have an eye for the evidence.

A fair-minded approach to the priors is similar to an indifference-driven approach, insofar as it results in reasonably smooth or flat priors over the usual suspects. There are, however, two important differences between the fair-minded and indifference approaches. First, fair-mindedness allows for quite a bit of diversity in the priors, while indifference, as it is usually understood, requires complete conformity. Second, the motivations for the two approaches are entirely different. The apriorist proponent of indifference regards a flat distribution as an end in itself, a uniquely rational set of opinions, given the paucity of evidence, whereas the proponent of fair-mindedness regards a flat distribution only as a means to an end, namely, convergence on the truth. The appeal of the fair-minded approach, then, is that it allows for a range of opinions when knowledge is hard to come by, while at the same time upholding the ideal of consensus when the evidence is decisive.

The Contextual Approach

The contextual approach encompasses a number of different ways of thinking about the priors, though all have a pragmatic feel. Let me consider one particular contextual line of thought.

Certain kinds of hypotheses have better track records than others. Grueish hypotheses, for example, have tended not to be as useful to science, so far, as their non-grueish counterparts. Why not use this, then, as reason to prefer non-grueish to grueish hypotheses when setting priors? The proposal, note, is not to prefer a non-grueish hypothesis h to a grueish hypothesis g on the grounds that h itself has been more useful than g. Since we are setting priors for h and g, we have, by assumption, no evidence as yet bearing on either hypothesis. Rather, it is *other* grueish hypotheses that have disappointed in the past (not least in the philosophy of induction's watershed

year 2000); this previous disappointment is taken to reflect unfavorably on grueishness in general, and therefore on *g* in particular.

To take another example, we may use the past success of simple over complex hypotheses as a reason to set priors in a new realm of inquiry that favor simple hypotheses in that realm over their more complex competitors. More generally, the idea goes, we can use the information we have about what kinds of hypotheses have been right in the past to set our priors for new hypotheses that have, as yet, no track record themselves.

At first, this suggestion might appear to be utterly confused. Clearly, using a past track record to set expectations for future success is a kind of inductive reasoning. But the prior probabilities are supposed to precede any inductive reasoning; inductive reasoning is, if you are a Bayesian, to be carried out by way of the conditionalization rule, which requires preexisting priors.

Contextualists are well aware of this. What they are suggesting is, in effect, a second inductive method to supplement BCT. This second method is less sophisticated than BCT—it is, more or less, what is sometimes called *straight induction*, in which expectations are based purely on frequencies—but by the same token it is independent of BCT. The contextualist prescription is: set your priors using straight induction, then learn from the evidence using BCT. Straight induction acts, then, as a kind of boot-strapping mechanism, to get BCT up and running as efficiently as possible.

The failure of contextualism to exert a hold on the Bayesian imagination is, surely, its construction of a second inductive edifice to stand alongside BCT. Why two inductive methods, if you can get by with just one?

The contextualist reply is as follows. When we set the priors for a set of hypotheses in a new realm of inquiry, we surely do have information that is inductively relevant to the task, for example, information that suggests that the laws of nature tend to be on the whole vastly more simple than the phenomena they describe. To ignore such information would be perverse. Yet

BCT provides no easy way to incorporate it: Bayesians learn by conditionalization, and conditionalization requires that priors are already set.[17] Thus, like it or not, there is a need for a non-Bayesian inductive procedure to take the information into account.

Note that this reply depends on the observation that we formulate new theories as we go along and in the light of the existing evidence, rather than all at once before inquiry begins. Some notorious consequences that this commonplace has for BCT are examined in section 11.

10. Bayesianism, Holism, and Auxiliary Hypotheses

10.1 Auxiliary Hypotheses

Modern Bayesianism deals in hypotheses that assign precise physical probabilities to the evidence. Most scientific theories assign such probabilities only in conjunction with auxiliary hypotheses. By an *auxiliary hypothesis* for a theory h, I mean any hypothesis that helps h to fix a particular physical probability for the evidence. No hypothesis is intrinsically auxiliary, then; rather, to be auxiliary is play, relative to what I will call the *main hypothesis*, a particular inferential role.

Typically, auxiliary hypotheses will do one of the following:

1. State that certain background conditions obtain, or state that there is a certain probability distribution over background or initial conditions,

2. Assert a certain relationship between the readings on a piece of instrumentation and the unobserved properties that the instrumentation is supposed to measure,

17. Exercise for the reader: could the Bayesian introduce "prior priors", and then conditionalize on the information about simplicity and so on, to reach the priors that will be used once the evidence begins to come in? What difficulties will have to be surmounted to implement such a scheme?

3. Help to derive, from the instantiation of some theoretical properties (presumably the subject of the main hypothesis) the instantiation of certain other, easier to observe properties. An example is the use of general relativity to determine the magnitude of the *gravitational lensing* effects that would be present if the universe's dark matter were to consist mainly of large, dark, star-like entities (MACHOs). Here general relativity plays an auxiliary role, helping a certain theory of dark matter to make predictions about observable phenomena.[18]

As these examples show, an auxiliary hypothesis may be anything from a low level empirical generalization to a high level theory.

10.2 The Bayesian's Quine-Duhem Problem

The ubiquity of auxiliary hypotheses creates a problem in confirmation theory called the *Quine-Duhem* problem.[19] Given that a main hypothesis makes predictions about observable phenomena only in conjunction with its auxiliary hypotheses, how to distribute the epistemic bounty among the individual hypotheses when the predictions are borne out, and (this side of the question has always attracted more attention) how to distribute blame when the predictions fail? The old workhorse of theoretical confirmation, hypothetico-deductivism, provides no mechanism for sharing out either the credit or the blame. The *Quine-Duhem* thesis holds that there is no such method: what is confirmed or disconfirmed will always be a conglomerate of main theory and auxiliary hypotheses. It has as a consequence what is often called epistemological holism, the doctrine that it is (almost) never correct to say of a bare theory, such as the general theory of relativity or the theory of evolution, that it is confirmed or disconfirmed.

18. Of course, gravitational lensing itself is not directly observable; still other auxiliaries must be used to deduce, from images of the stars, that it is happening.

19. The hyphenated eponymy presumably reflects a reluctance to distribute praise or blame among the individuals named.

The analog of the Quine-Duhem problem for BCT (a difficulty first encountered in section 6.4) is as follows. The heart of BCT is Bayes' rule, which tells us how to adjust our subjective probabilities for a hypothesis in the light of the evidence. But as is evident from the rule, the hypothesis whose probability is in the first instance adjusted is the hypothesis that bestows the particular physical probability on the evidence—the likelihood—that is used for conditionalization. This is not the main hypothesis itself, but the main hypothesis conjoined with one more more auxiliary hypotheses. Thus, Bayes' rule tells you how your probability for this conjunction should change, but what about the probability for the hypothesis itself?

There is some good news and some bad news. The good news is that, provided all the appropriate priors are well defined, Bayesian conditionalization on a piece of evidence has definite implications for the subjective probability of *every hypothesis* (indeed, every proposition) in the conditionalizer's head space. Even though my main hypothesis h does not assign a definite physical probability to the evidence e, then, conditionalizing on e will result in some definite posterior probability for h. (This is because, although there is no well-defined physical likelihood $P_h(e)$, there is a well-defined subjective likelihood $C(e|h)$; or at least, there is if all the necessary priors are well defined.)

The bad news is that, although the posterior for h is well defined, the dynamics of this probability—the way that it changes over time as the evidence comes in—does not necessarily have any of the features that make modern Bayesianism so attractive. Suppose, for example, that the main hypothesis h when conjoined with an auxiliary hypothesis a predicts a piece of observed evidence e. The conjunction ha is duly confirmed by e: your probability for ha increases. But, depending on your other priors, your probability for h may change in a number of ways. It may even decrease. (The reasons for this will be laid out in section 10.5.) The likelihood lover's principle does not apply to the individual constituents of a hypothesis-auxiliary conjunc-

tion, then, so neither do any of the consequences of the principle, not least the convergence results.[20]

Perhaps above all else, it is no longer true that fixing the prior probabilities for the individual hypotheses is sufficient, by way of the probability coordination principle and the total probability theorem, to determine the way your probabilities should change. Other subjective probabilities will make a difference to probability updates (see section 10.5).

Is there any way to get back the clockwork behavior of modern Bayesianism? We cannot have the transcendent simplicity promised back in sections 5 and 6, but if certain limits are placed on the theories in play—if certain assumptions are made about the prior probabilities over various combinations of main and auxiliary hypotheses—a nice, approximately clockwork aspect will emerge. At the very least, there are conditions under which the subjective probability of *h* will behave in the same decent and forthright way as the subjective probability of *ha*, increasing when *ha* predicts well and decreasing when *ha* predicts badly. I will sketch some of this work, presented in Strevens (2001), later in this section.

10.3 The Problem of Ad Hoc Reasoning

Let me introduce a particular problem in confirmation theory for the solution of which it seems especially important to have a solution to the Quine-Duhem problem—a method for distributing praise and blame differentially across main/auxiliary conglomerates—and where it seems that considerable scientific progress has in fact been made by knowing which allocation of praise and blame is correct. We will view Quine-Duhem through the lens of this problem, the problem of how to judge *ad hoc reasoning*.

20. There is one important convergence result that is unaffected (see §9.2): as the probability of the true main/auxiliary conglomerate *ha* approaches one, it drives the probabilities of *h* and *a*, which must each be greater than the probability of *ha*, before it.

Suppose that a hypothesis-auxiliary conglomerate—what I earlier called a scientific *model*—is falsified. The model makes a prediction that is at odds with the evidence. At least one hypothesis in the model must therefore be false.

A proponent of the model's main hypothesis may be tempted to tweak its auxiliaries so that they no longer yield the false prediction, or even better, so that they yield a new version of the prediction that, unlike the old, accords with the data. Tweaking the auxiliaries in this way is what I call ad hoc reasoning. (There may be other kinds of ad hoc reasoning, but I will not discuss them here.)

Some writers have come close to condemning ad hoc reasoning outright. But there are some famous historical examples of ad hoc reasoning that not only turned out to be very good science, but were recognized as such at the time. The most famous of all, at least in philosophical circles, is the episode that may justly be called the discovery of Neptune.

In 1800 only seven planets were known to exist, the outermost of which was Uranus. Astronomers observed certain irregularities in the orbit of Uranus, which were at odds with the model consisting of, first, the Newtonian theory of gravitation, and second, the prevailing view of the structure of the solar system, in which Uranus was the outermost planet. Adams and Le Verrier independently postulated the existence of a planet orbiting the sun beyond Uranus, which would account for the apparent perturbations in Uranus's orbit. When they showed that a particular choice of orbit for such a planet would exactly account for Uranus's apparently unNewtonian excursions, the eminent astronomer John Herschel declared that Neptune's existence had been demonstrated "with a certainty hardly inferior to that of ocular demonstration". Not only did the ad hoc tweaking of the solar system model save Newtonian theory, then, the tweaked auxiliary hypothesis was itself elevated to a position of supreme epistemic standing.

Other episodes of ad hoc reasoning have served to damn rather than

to save the theories involved. Howson and Urbach (1993) cite an example from Velikovsky's *Worlds in Collision*, a book that sought to explain various ancient myths and dramatic biblical events such as the parting of the Red Sea as real events caused by giant comets passing close to the earth. In answer to the question why we have only reports of isolated "miracles" at such times, when the comets were supposed to have caused planet-wide upheavals, Velikovsky appeals to the hypothesis of collective amnesia: the events were so terrible that almost all memory of them has been repressed. (The repressed memories, by the way, emerge in the form of stories about the gods, which turn out to be simple allegories of astronomical events—the myth of Athena's birth from the head of Zeus, for example, refers to a "comet" being ejected from the planet Jupiter, almost colliding with the earth, and later settling into a stable orbit as the planet Venus. Velikovsky described himself as psychoanalyzing the entire human race.) Why does this auxiliary maneuver elicit sniggers, rather than Herschelian paeans?

One answer: in the Velikovsky case, the auxiliary hypothesis is extremely self-serving. But is the positing of an additional planet to rescue your hypotheses about the laws of celestial mechanics any less self-serving? Another answer: the hypothesis of collective amnesia is highly implausible, or as a Bayesian would say, would have had a very low subjective probability. But so would the Neptune hypothesis: bear in mind that to save Newtonian theory, it was necessary to assume not only the existence of an extra planet, but that the planet had a very particular orbit. No one would have bet on finding a planet in that orbit before the work of Adams and Le Verrier.[21] Could there be social forces at work? The central characters involved in the discovery of Neptune were highly respected mathematicians and scientists, occupy-

21. This is not to say that the probability of the tweaked auxiliary is always irrelevant: in the case where it is high even before the falsifying data was observed, there will be no problem endorsing the new auxiliary—indeed, in Bayesian terms, it has been endorsed all along.

ing central roles in their fields.[22] Velikovsky was a psychiatrist who had no standing in the world of astronomy. In a sense, social standing did play an important role in distinguishing these two cases, on the Bayesian approach; not the social standing of an individual, however, but the hard-won social standing of a *theory*.

Before I turn to the Bayesian treatment of ad hoc reasoning, let me provide one more example, and then reestablish the link between the problem of ad hoc reasoning and the more general Quine-Duhem problem. First, the example. Observations of galactic dynamics—the way that galaxies move and change—are at odds with current gravitational theory. Or rather, they are at odds unless large amounts of "dark matter" exist in the universe. This dark matter, though unobservable, would have to constitute about 90% of the mass of a typical galaxy. The dark matter hypothesis (or rather hypotheses, since as mentioned above, there are competing views about the nature of dark matter) are recognizably auxiliary tweaks very similar to the tweak that initiated the discovery of Neptune. Interestingly, there is an alternative, though less well known, account of the galactic aberrations that revises the theory of gravity itself, so that it is no longer valid for bodies undergoing acceleration very near zero. Which tweak is on the right track? Is it the main hypothesis, gravitational theory, or the auxiliary, concerning the constitution of galaxies, that needs to be revised?

This is a matter of controversy. Dark matter theory is more popular, but the physical revisionists are by no means regarded as cranks.[23] In this third case, then, tweaking the auxiliary is neither clearly the right thing to do nor clearly a sign of impending theoretical implosion.

22. Actually, Adams was still an undergraduate when he did his calculations (and what have *you* calculated recently?). But his contribution was at first ignored; it was only when Le Verrier's work became known a few months later that he received the attention he deserved.

23. The principal author of the theory, Mordehai Milgrom, wrote it up recently for *Scientific American* (August 2002). At the time of these notes' most recent revision (August 2006), evidence has emerged that casts doubt on Milgrom's theory and seems to point to the existence of dark matter. The work is not yet published.

What does this have to do with Quine-Duhem? If there were a solution to the Quine-Duhem problem, we would know, when a main hypothesis and an auxiliary hypothesis jointly issued a false prediction, which of the two to blame (or more subtly, which to blame more). If the auxiliary is to be blamed, then it is right and proper to tweak it, that is to discard it in favor of another, revised version, to save the irreproachable main hypothesis from the stigma of empirical inadequacy. If it is the main hypothesis, tweaking the auxiliary will seem to be the last resort of the scientific charlatan—or at least, of the scientist who loved too much.

When the auxiliary deserves the blame, and so is rightly replaced by a new, tweaked version, say that the main hypothesis has been subject to a *glorious rescue*. When the main hypothesis deserves the blame, and so the replacement of the auxiliary by a tweaked version is the wrong move, say that the main hypothesis has been subject to a *desperate rescue*. The discovery of Neptune, then, constitutes a glorious rescue of Newton's laws, the glory accruing not so much to the laws themselves as to the rescuing auxiliary, the posited existence of a eighth planet in a certain orbit, which sees its epistemic status raised from outside possibility to near certitude. Velikovsky's use of the hypothesis of collective amnesia is, by contrast, a desperate rescue of his theory of planetary Ping-Pong. The rest of this section will explore a Bayesian approach to distinguishing glorious from desperate rescues.

10.4 The Old Bayesian Approach to the Quine-Duhem Problem

Begin with what might be called the standard Bayesian approach to the Quine-Duhem Problem. The standard Bayesian I have in mind was introduced in section 10.2. They solve the problem simply by noting that conditionalization on the critical piece of evidence will result in some well-defined posterior probability or other for each of the main hypothesis, the auxiliary hypothesis, and the tweaked auxiliary.

Since we are about to start pushing some probabilities around, let me

put this into the Bayesian formalism. As above, call the main hypothesis h and the auxiliary hypothesis a. Call the tweaked version of the auxiliary hypothesis a'. The piece of evidence that is inconsistent with the predictions of ha I call e. In the scenario under consideration, ha entails $\neg e$, whereas ha' is consistent with e, and perhaps even predicts it. I will assume that ha' indeed entails e, as it does in the examples described above. Finally, I assume that a and a' are mutually exclusive; thus, ha and ha' are mutually exclusive.

Then the Bayesian approach is as follows. Upon conditionalizing on e, the probability of ha will go to zero. The probabilities of the rival models that predict the evidence, including ha', will increase. That is all we can say for sure. The probabilities of h, a, and a' will all change in some way. Their new values will be well defined, but it is impossible to say in general whether they will increase or decrease. That will depend on various priors. What, then, is the difference between a glorious and a desperate rescue? A glorious rescue is one in which the probability of h does not decrease much or at all, and the probability of a' becomes quite high itself. The epistemic credibility of the main hypothesis is preserved, then, and the tweaked hypothesis a' gains credibility that it did not have any more—it is "discovered to be true". A desperate rescue is one in which the probability of h falls and the probability of a' does not increase to any great degree.

Is this a solution to the problem of ad hoc reasoning? It sounds more like a redescription of the problem. We have asked: in what circumstances is a rescue glorious, rather than desperate? The standard Bayesian has answered: when it is glorious, rather than desperate. The reason that the standard Bayesian can say no more is, of course, that BCT in itself puts no constraints on the way that the probabilities of h and a change when ha is refuted by a piece of evidence e—the same reason, as explained in section 10.2, that BCT provides a rather unhelpful solution to the Quine-Duhem problem.

To do better than the standard Bayesian, we must make some additional assumptions about the epistemic situation. To see the form that these assumptions will take, consider why, when e refutes ha, the probability of h may nevertheless go up.

There are two features of the Bayesian scenario that leave open this possibility. First, there may be a negative correlation between h and a, in which case (ceteris paribus), as the probability of one increases, the other will decrease; they cannot, then, both increase. Second, and more importantly, the evidence e may bear on h in other ways than its refutation of ha. (Perhaps h is a part of other models that make predictions about e, possibly playing the role of the auxiliary rather than the main hypothesis.) These other ways may have much more impact on the posterior for h than does the refutation of ha, and whether the impact is positive or negative is left entirely undetermined by the description of the situation.

Put the first possibility aside for now. The second possibility can be avoided by limiting the scope of the investigation to cases where it does not eventuate, that is, to cases where the impact of e on h and a goes entirely by way of the falsification of ha. In what follows I will suppose, then, that

$$C(h|e) = C(h|\neg(ha))$$

In the cases of ad hoc reasoning presented above, this seems to be a reasonable assumption. Insofar as the observed perturbations in the orbit of Uranus impacted Newtonian gravitational theory, they did so by showing that Newtonian gravitation made the wrong predictions when conjoined with the then-current model of the solar system, and so on for the other ad hoc scenarios.[24]

24. You may wonder if this sort of informal reasoning is not question-begging: in determining that all the evidential impact of e on h is due to the impact of the falsification of ha on h, am I not assuming that I can follow the flow of confirmation from e to the in-

If the assumption is allowed, real constraints on the way that e impacts the probabilities of h and a exist. I will leave the details of the derivation to you (or you can find them in Strevens (2001)), but the following result about the posterior probability $C^+(h)$ for h after the observation of e can be derived:

$$\begin{aligned} C^+(h) &= C(h|\neg(ha)) \\ &= \frac{1 - C(a|h)}{1 - C(a|h)C(h)} C(h) \end{aligned}$$

The change in the probability of the main hypothesis, then, will be determined by the properties of the multiplier—the fractional expression by which $C(h)$ is multiplied to obtain $C^+(h)$.

A careful study of the multiplier shows that it has the following features (depicted in figure 7).

1. It is always less than one (unless $C(h)$ is already one). Thus, when ha is falsified, the probability of h must always decrease.

2. The greater the probability of h prior to the observation of e—the greater the value of $C(h)$—the smaller the decrease in the probability of h. Thus main hypotheses that are more credible suffer less from the falsification of ha.

3. The lower the probability $C(a|h)$ prior to the observation of e, the smaller the decrease in the probability of h. Roughly speaking, the less credible the auxiliary hypothesis, the less the main hypothesis will suffer from the falsification of ha. But only roughly, because $C(a|h)$ is not the same as $C(a)$. This issue will be addressed shortly.

---

dividual parts of ha? That is, that I already have a solution to the Quine-Duhem problem in hand? Touché. For a more careful, but therefore more roundabout, presentation of the assumption required to secure the above equality, see Strevens (2001). The roundabout presentation has the advantage that it isolates precisely those features of the examples above in virtue of which it is reasonable to assume that $C(h|e)$ is equal to, or at least approximately equal to, $C(h|\neg(ha))$.

4. Because we could just as well have called h the auxiliary hypothesis and a the main hypothesis, without affecting the derivation, all of the above holds for a as well: its probability decreases when ha is falsified, and the size of the probability drop is smaller as a is more credible and h is less credible.

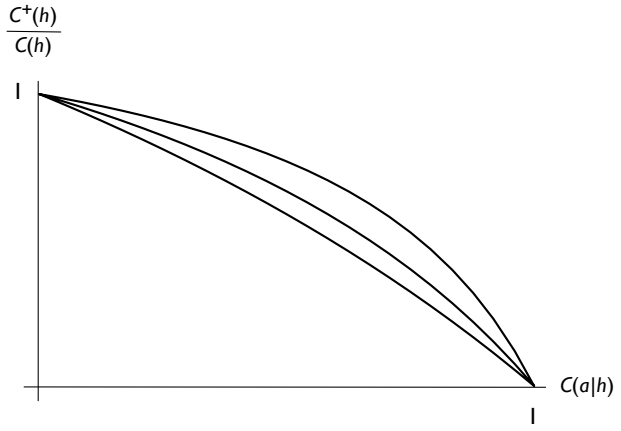

Figure 7: The ratio of the posterior probability $C^+(h)$ to the prior $C(h)$ (i.e., the Bayesian multiplier), as a function of $C(a|h)$, for three values of $C(h)$. From top to bottom, $C(h) = 0.7, 0.5, 0.3$. In each case, as $C(a|h)$ increases, the multiplier for $C(h)$ decreases, but the higher $C(h)$, the slower the initial decrease.

In summary, then, when ha is falsified, there is a certain amount of blame that must be shared between h and a in the form of probability decreases for both. The blame is distributed, roughly, by comparing the probabilities of h and a before the falsification: as you might expect, the hypothesis with the lower probability gets more of the blame. There is a Matthew effect at work here: hypotheses that already have low probability will suffer disproportionately, while hypotheses that have high probability will be relatively well protected.[25]

25. "For to every one who has will more be given, and he will have abundance; but from

There is nothing surprising about this; rather, it is a case of Bayesianism's reproducing a piece of conventional wisdom about confirmation. Platitudinous though it may be, however, the observation that the less certain theory is the likelier culprit in an episode of falsification had become so problematized by Quine and others by the middle of the last century that a little triangulation ought not to be scorned.

However, we have gone too fast. I have been assuming that what can be said about $C(a)$ can also be said about $C(a|h)$, the expression that actually occurs in the all-important multiplier, and in particular, that when $C(a)$ is high, $C(a|h)$ is relatively high and vice-versa. What is the justification for this?

The mathematics itself puts few constraints on $C(a|h)$. It can be high when $C(a)$ is low, and low when $C(a)$ is high. In order to impose some discipline on $C(a|h)$, then, we are going to have to make a substantive assumption about probabilistic situations of interest, an assumption that both justifies the assumption that the two probabilities move more or less in tandem, and that holds in the kinds of cases of ad hoc reasoning that are under investigation.

It is sometimes said, in connection with this sort of problem, that since a and h typically concern rather different kinds of things—auxiliary hypotheses are often about particular matters of fact, whereas main hypotheses are often abstract theories of the behaviors of things—it is reasonable to assume that a and h are probabilistically independent, so that $C(a|h)$ is equal to $C(a)$. It would be very convenient if this reasoning were sound, but it is not. The two hypotheses' "being about rather different kinds of things" might justify independence in the prior probability distribution before any evidence comes in,[26] but once evidence arrives, a main hypothesis and its

him who has not, even what he has will be taken away", Gospel of Matthew, 25:29. The original Matthew effect, concerning the assignment of credit for scientific discoveries, was described by Merton (1968).

26. Though even then, it seems a bit much to assume that the prior distribution over,

auxiliaries will invariably become probabilistically correlated.

To see this, observe that if the main and auxiliary hypotheses really were independent, then on learning that the auxiliary was false, your subjective probability for the main hypothesis would be, by Bayes' rule, unaffected. (Independence entails that $C(h|a)$ is equal to $C(h)$, provided that $C(a)$ is non-zero.) But this is the same auxiliary that has been used in many past tests of the hypothesis. If it is false, these experiments have little or no credibility, and your probability for the main hypothesis should be adjusted (downward, if the tests were confirmatory).

This same argument shows, however, that in the usual case—for typical pairs of main and auxiliary hypotheses—the probabilities of the pair will be tethered, and so movements in the probability of a will be broadly reflected in the movements of $C(a|h)$. The exact relationship will depend on the content of h and a, but the more plausible a is, the higher $C(a|h)$ will be. This is what we need for behaviors (1) to (4) above to be typical of cases in which a main/auxiliary pair is falsified.

So what can we say about ad hoc reasoning? The problem, recall, is to lay down conditions under which an auxiliary tweak will constitute a glorious rescue, and conditions under which it will constitute a desperate rescue. In a glorious rescue, the drop in the probability of h due to the falsification of ha is small or non-existent; in a desperate rescue, it is substantial. In the light of the above results, we can then say:

1. A tweak constitutes a glorious rescue when the probability of the main hypothesis is high relative to the probability of the auxiliary. Most glorious of all are rescues where the probability of the main hypothesis is not only relatively high but absolutely high, that is, near one. The Neptune case fits this specification very well: the probability for

say, the different possible arrangements of the planets will be independent of the prior distribution over the different physical theories characterizing the dynamics of those same planets.

Newtonian theory was, at the time, very close to one in the scientific consciousness.[27] The probability of the auxiliary, the prevailing model of the solar system, was high, but still well short of one: the possibility of another planet was considered a real one.

2. A tweak constitutes a desperate rescue when the probability of the main hypothesis is low relative to the probability of the auxiliary. Most desperate of all are rescues where the probability of the main hypothesis is not only relatively low but absolutely low. It is in these cases that you have permission to laugh and point.[28] The unfortunate Velikovsky illustrates this precept very well: the probability of the main hypothesis of *Worlds in Collision* seemed rather low to most people compared to the probability of the auxiliary, which in this case is the absence of collective amnesia. Was Velikovsky irrational, then? Perhaps not: his own subjective probability for the hypothesis of colliding worlds was perhaps much higher than for the auxiliary. There are Bayesian sirens for every reef...

One feature of glorious rescues remains to be explained. Why does the probability of the tweaked auxiliary jump so high? Why, for example, was Couch and Le Verrier's achievement not just to preserve Newtonian theory, but to discover the existence of Neptune?

In a glorious rescue, the brunt of the falsification of *ha* is borne by *a*. The probability of *a*, then, will plunge dramatically. Assuming that it was reasonably high before the falsifying evidence was observed, this leaves a lot of probability to redistribute among *a*'s rivals. In many cases, most of the rival auxiliaries are no better able to predict the falsifying evidence than *a* itself. The redistributed probability will be concentrated, then, on a small

27. Real Bayesians do not, of course, talk this way.

28. Bearing in the mind that if the main hypothesis turns out, against all odds, to be correct, you will go down in history as a reactionary blockhead who stood in the way of scientific progress—only to become roadkill on the superhighway to truth.

subset of the rivals, perhaps even on a single possibility, the tweaked auxiliary. Hence its sudden ascendancy.

Exercise to the reader: explain why a tweaked auxiliary that, in conjunction with the main hypothesis, predicts the falsifying evidence, will tend to do better out of a glorious rescue than a tweaked auxiliary that merely insulates the main hypothesis from the evidence, that is, that in conjunction with the main hypothesis, has nothing special to say about the falsifying evidence. (You should look beyond the admittedly relevant fact that the former auxiliary, but not the latter, will be confirmed by the falsifying evidence. Think of some examples.)

Enough about ad hoc reasoning. Consider the Quine-Duhem problem more generally. You have seen how the blame should be distributed when a main/auxiliary pair *ha* is falsified. How should praise be distributed when *ha* is confirmed? You know that in the falsification case, the higher the probability of *a*, the bigger the drop in the probability of *h*. How do you think it goes in the confirmation case? When *ha* is confirmed, the probability of *h* will go up; how will the size of this increase depend on the probability of *a*? As the probability of *a* goes higher, the increase in the probability of *h* due to confirmation . . . gets higher or lower?

Answer: it gets higher. That is, the more sure we are about *a*, the more strongly *h* is confirmed when *ha* is confirmed. This also goes for disconfirmation of *ha* that stops short of falsification: the more sure we are about *a*, the more strongly *h* is disconfirmed when *ha* is disconfirmed. (These claims are all justified in Strevens (2001).) In short, the more sure we are of *a*, the more we learn about *h* by either the confirmation or the disconfirmation of *ha*. Bayesian plaudits for another platitude: the more secure your auxiliaries, the more the evidence tells you about your theories.

11. The Problem of Old Evidence

11.1 The Problem

You are a scientist, and your life's goal is to explain a mysterious phenomenon, the Kumquat effect. No known theory is very successful in explaining the effect. All of them assign (in conjunction with the relevant auxiliary hypotheses) a very low physical probability to the effect's occurring. Yet occur it does.

Then one happy day, you hit upon a new theory that predicts the effect. Your theory, let us say, entails the effect; it therefore assigns the effect a physical probability of one, in contrast to the very low physical probabilities assigned by previous theories. Since BCT favors theories that assign relatively high physical probabilities to the evidence, and disfavors those that assign relatively low probabilities, your theory should immediately become everyone's favorite—correct? Surprisingly, it is not at all clear that BCT delivers this verdict.

Indeed, on the Bayesian picture, it looks rather like your theory's prediction of the Kumquat effect will fail to count in its favor at all. This terrible embarrassment for Bayesianism, pressed enthusiastically by Clark Glymour, is called the *problem of old evidence*.

Let us see how the problem arises. When you first conceive of your new theory, you must assign it a prior probability. This in itself poses a problem for the Bayesian, which will be discussed in section 11.4, but let us suppose, for now, that a prior is successfully determined. Now you wish to update the prior in the light of the fact that your theory h predicts the Kumquat effect, the observation of which I will call e. The updating should proceed in accordance with the conditionalization rule:

$$C^{+}(h) = \frac{C(e|h)}{C(e)} C(h).$$

Because h entails e, the relevant likelihood $C(e|h)$ is equal to one. The probability for h, then, will increase by a Bayes multiplier that is the reciprocal of

$C(e)$. But what is $C(e)$? We already know that the Kumquat effect occurs. Thus our subjective probability for e will be one. It follows that the Bayes multiplier is also one, which is to say that conditionalizing on the Kumquat effect does not affect our confidence in h at all, despite the fact that h predicts the effect whereas no rival can.

More generally, whenever we update our subjective probability for a new theory by conditionalizing on old evidence, we will encounter the same problem. Because the evidence is already known for certain, our subjective probability for the evidence is equal to one. Thus conditionalizing on the evidence cannot increase our subjective probability for the new theory.[29] A dire difficulty for BCT.

The usual example of the old evidence problem is the confirmation of Einstein's theory of general relativity. One of the strongest pieces of evidence for the theory, in the view of physicists then and now, was its prediction of the precession of the perihelion of Mercury. The precession was discovered in the nineteenth century, but Einstein formulated his theory only in the years between 1905 and 1915. Scientists' probability for the fact of the precession was one, or close to one, then, before Einstein began work; consequently, the precession ought not, on the Bayesian approach, to have confirmed the theory.

You should be able to see that evidence is often old in the sense that causes trouble for BCT: Darwin knew about his finches before he became an evolutionist; physicists knew some of the puzzling properties of the specific heats of gases for decades before quantum mechanics was able to account for them; and theories of, say, the business cycle in economics have all been for-

29. You might think that if a hypothesis h assigns a probability to a piece of old evidence e that is less than one, the probability of h would actually *decrease* when conditionalizing on e. But this is not so: remember that the subjective likelihood $C(e|h)$ implicitly accommodates background knowledge (section 4.3); for this reason, $C(e|h)$ is equal to one, so that there is no change in the probability of h. Why does PCP not apply? Because the observation of e is inadmissible evidence.

mulated after the behavior of the cycle was already well known. Bayesianism promises to give you a complete theory of how your evidence should impact your beliefs. But in the case of old evidence, it fails to do so.

Two remarks. First, it is not quite fair to say that BCT gives the wrong answer the question of how old evidence should affect your subjective probabilities. It is rather that, in its idealizing way, BCT assumes that the problem will never crop up in the first place, because it assumes that you are aware of all the logical possibilities—in particular, all the possible scientific theories—in advance of your receiving any evidence at all. The problem, then, is that there seems to be no straightforward way to relax this assumption. Or more exactly, if you make BCT more realistic by supposing that some theoretical possibilities will be discovered in the course of investigation, you seem to be committed to making it at the same time more *unrealistic*, in that it declares the impact of old evidence on these new theories to be nonexistent.

Second, a number of writers have tried to make Bayesians feel better about the problem of old evidence by defining a sense in which it is possible to say that old evidence "supports" a theory. One way to do this is to say that the evidence supports a theory if the probability of the theory given the evidence is higher than the probability of the theory given the negation of the evidence, according to some probability distribution that (if you think about it) cannot be your current subjective probability distribution in which the probability of the evidence is one. Such approaches to the problem do not actually help you to adjust the probability of a new theory in the light of the evidence already accumulated, which is to say, they do not do for you what BCT, or at least "clockwork Bayesianism", is supposed to do for you. As such, they should be considered palliatives, rather than solutions, to the problem of old evidence—symptomatic relief, perhaps, but not cures.[30]

30. They may also be considered as solutions to a distinct problem, that of providing an ahistorical notion of evidential support—a problem that arises whether the evidence is"old"

11.2 Replaying History

I will consider two popular attempts to solve the problem of old evidence. The first, favored by, for example, Howson and Urbach (1993), is motivated by a wish: if only the evidence had been uncovered after the theory was formulated, or the theory had been formulated before the evidence came to light, there would have been no problem. Would that it were possible to go back in time and replay the history of science, but with events reordered so that the problem of old evidence did not arise! If only, for example, humanity had been so clever as to see all the theoretical possibilities *before* the evidence began to accumulate . . .

Howson and Urbach (following a strategy panned in advance by Glymour) suggest that we do just this. More exactly, when we formulate a new theory for which pertinent evidence e already exists, we should conditionalize on e as though we had not yet observed it, using the value for $C(e)$, then, that we would have had before e came to light. The numerical effect is just as if the theory really had been formulated before the observation of e.

Most Bayesians regard this as a pretty hairy maneuver. First off, it involves flagrantly violating the Bayesian conditionalization rule, which enjoins us to use our present subjective probability for e when conditionalizing, not some other historical or counterfactual probability. Howson and Urbach defend themselves against this worry by arguing that it can never be right to conditionalize using your present subjective probability for e since, by the time you get around to conditionalizing, your probability for e has already gone to one. You should always conditionalize, then, using your probability for e before e was observed.[31] This argument seems to me to

---

or not. The distinction between the historical problem of old evidence (the problem tackled in these notes) and the ahistorical problem of evidential support is introduced by Garber (1983).

31. What Howson and Urbach require, more precisely, is that you use the subjective probability for e that is determined by your background knowledge at the time of conditionalization *not counting e*. This is a peculiar suggestion for a Bayesian to make; the idea

conflate the merely practical problem raised by having a number of things to do at one time with a serious theoretical problem.

Second, it is hard to say what your subjective probability for *e* would be if you did not already know it. Perhaps *e* was observed long before you were born, and has always been a part of your mental firmament. Perhaps the best you can do is to estimate the subjective probability for *e* commonly entertained before its discovery—but are such mental exercises really a legitimate part of the story about confirmation?

Third, even if these other difficulties can be assuaged, we are a long way from the clockwork Bayesianism of sections 5 and 6. In those sections, recall, we envisaged a Bayesian theory of confirmation in which the scientist only had to fix a prior probability distribution over the competing hypotheses, and everything else was taken care of. As the evidence rolled in, subjective probabilities were updated to as many places after the decimal point as you liked. The "replay" solution to problem of old evidence threatens to make BCT just as fuzzy, subjective, and open to human psychological vagaries as many of its opponents fear.

11.3 Learning about Entailment

An alternative solution to the problem of old evidence, also first suggested and rejected by Glymour, focuses on the fact that, in many historical examples of the confirmatory power of old evidence, the moment at which the

that there is some probability for *e* determined by a certain set of propositions, such as your background knowledge less *e*, which has at no point constituted your actual background knowledge, belongs to the world of logical probability (see section 5.2), not subjective probability. There is also, of course, the more practical worry that removing the effect of an ancient observation *e* from your subjective probabilities would be a technical nightmare. Howson and Urbach invoke the work of Gärdenfors and others on belief revision, but this work does not address the case where background knowledge is represented, not only by a set of propositions, but by a probability distribution over those propositions. Later, Howson and Urbach invoke the "counterfactual supposition that one does not yet know *e*" (p. 405).

new theory is seen to be supported by the old evidence is the moment at which it is shown that the new theory predicts the evidence. To keep things simple, let us say that the theory entails the evidence. Then the critical moment is the moment at which it becomes apparent that this entailment holds. This suggests that perhaps, though conditionalizing on the long known old evidence itself may not have any effect on a new theory's probability, conditionalizing on the newly learned fact of the entailment may have such an effect. The proposal, then, is that the probability of a new theory h is raised not by conditionalizing on e, but by conditionalizing, as it were, on $h \vdash e$ (using $\vdash$ to represent logical entailment).

This approach has two major advantages over the replay approach. First, it involves the use of the conditionalization rule as always envisaged by BCT, that is, using a real, current subjective probability—the subjective probability $C(h \vdash e)$—rather than historical or counterfactual subjective probabilities. Second, it holds out the promise of a precise technical recipe for assessing the effect of old evidence on new theories, provided that a good theory of conditionalizing on facts about entailment can be constructed.

There is a serious obstacle, however, to the construction of such a theory. According to the axioms of subjective probability, your subjective probability for a logical fact such as $h \vdash e$ should always be one. If your subjective probabilities conform to the axioms, then, your probability for the entailment will be, like your probability for the old evidence itself, equal to one, and thus your probability for the new theory will be unaffected by conditionalization.

Yet although it is true that your subjective probabilities ought to conform to the axioms, and so that, ideally, you should assign a probability of one to all logical facts, it is not psychologically realistic to expect Bayesian reasoners to follow this policy at all times. We are not *logically omniscient*; we do not perceive of all logical truths that they are in fact logical truths. The Bayesian apparatus ought to allow for this. In a new Bayesianism with a human face

(more exactly, with a human mind), a logical truth such as $h \vdash e$ could have a subjective probability of less than one, corresponding to the believer's (in most cases, no doubt, very low) level of expectation that h will in fact turn out to entail e.

Suppose that all of this is quite feasible, and that we are equipped with a new Bayesianism; how much can we then show? A little it turns out, but less than you might expect—or at least, less than you might hope. There exist various attempts to demonstrate that conditionalizing on $h \vdash e$ will, even if $C(e)$ is one, raise the probability of h. But there is no consensus on which demonstration is best, nor is there a stronger result showing that the probability lift is large in cases where the rivals to h have much less to say about e (Earman 1992, chap. 5). This is something of a disappointment, but perhaps progress is around the corner.

In any case, there is a deeper worry about the approach described in this section, voiced by Earman and Howson and Urbach, among others. It seems quite possible that a theory might be formulated with the express aim that it predict a piece of old evidence e. In a case such as this, by the time the theory is complete and we are ready to assess the evidence for the theory, we already know that the theory entails e. The fact of the entailment itself is "old evidence". Then we have the same old problem on our hands. Yet it seems that the theory's predicting the evidence ought to count in its favor all the same (tech box 11.1).

This objection suggests that the entailment-learning approach, however successfully it may deal with some cases, is not getting at the gist of the problem. If BCT is truly to reflect our practices of confirmation, it must find another way to handle the impact of old evidence.

11.4 The Problem of Novel Theories

On our way to the problem of old evidence, we encountered a problem that might be called the *problem of novel theories*. Suppose that you for-

11.1 Prediction versus Accommodation

Some writers have doubted that, in the case where a theory h is explicitly designed to entail e, it ought to receive much of a boost from e. In such cases, they say, e is merely accommodated, rather than being predicted.

This view is at odds with scientific practice. Einstein's special theory of relativity was designed to provide a symmetrical treatment of moving coils and moving magnets, Darwin's theory to account for speciation, and so on. Surely these theories received considerable credit for achieving, or partially achieving, their aims.

What the belittlers of accommodation have in mind, I think, is a special case in which evidence obtained from an experiment is accommodated by adjusting the parameters of a theory already known to have sufficient flexibility to accommodate any kind of evidence likely to result from the experiment. In such a case the chosen values of the parameters receive a probability boost, but the theory itself, considered in abstraction from the parameters, does not.

This is, by the way, one way in which a theory that is simple in a certain sense—having few adjustable parameters, and thus little flexibility in accommodating the evidence—may receive a greater probability boost than a more complex theory from a given set of evidence, even though both theories have instantiations (i.e., versions with the parameters set to certain specific values) that entail the evidence.

I leave it as a (fairly involved) exercise to the reader to give a formal treatment of this phenomenon. An even more involved, yet worthwhile, exercise, is to generalize to the case where even the more complex theory could not have accommodated just any evidence.

mulate a new theory. If your new theory is to get anywhere in the Bayesian world, it must receive a prior probability greater than zero. But the theory's pre-existing competitors have probabilities that, we have been assuming throughout, already sum to one. This leaves nothing over for the new theory. Where is its prior probability to come from?

There are several possibilities. First, foreseeing the problem, you may have reserved some prior probability for cases such as this. That is, you may have included among your rival hypotheses from the very start a hypothesis that says, in effect, "Some other hypothesis that I haven't yet thought of is the true one". As new competitors are formulated, they may take a share of this reserve probability. Call this equivalent of "(e) None of the above" in science's great multiple choice exam the *reserve hypothesis*. (It is sometimes called the catch-all hypothesis.)

Here are two problems with the reserve hypothesis approach. First, our clockwork Bayesianism is again thrown out of whack. No matter how creative you are with auxiliary hypotheses, you will not find one that is both reasonable and that, in combination with the reserve hypothesis, assigns a definite physical probability to the evidence. Assigning prior probabilities to all the hypotheses and auxiliary hypotheses is not enough, then, to determine how your subjective probabilities will change on observation of the evidence.

There are ways around this problem. You might, for example, take as the physical likelihood of the evidence on the reserve hypothesis the weighted average of the other physical likelihoods, so that the reserve hypothesis has no net effect on the probability of the evidence *e*. This strategy is technically clean; can it be justified in the Bayesian framework? Do you really suppose that your unformulated theories will say, concerning the evidence, much the same thing as the existing theories? If so, why bother to formulate them?

The second problem is simply that, if you have been overly optimistic in your estimate of your grasp of the theoretical possibilities, then you may

run out of reserve probability. It is hard to see a conventional Bayesian way out of this difficulty, except to encourage undue pessimism among scientists. Perhaps that is what philosophy is for.

There is an unconventional solution to the problem, however: just as learning a new fact about logical entailment relations might impact your subjective probability distribution (section 11.3), so learning a new fact about the space of logical possibilities might rearrange your probabilities. For example, learning that a new theory of gravity is possible, quite unlike anything you had previously considered possible, might cause you to reduce your subjective probability for other theories of gravity. In this way, an expansion of your conceptual horizons might itself liberate some probability that was previously tied to other, known theories.

Another approach to the problem of novel theories is to rejuggle your subjective probabilities every time a new theory is formulated. Preexisting theories are "taxed" to provide the subjective probability you want to assign as a prior to the new theory.[32] If there is a problem with this approach, it is that, because your priors for the competing hypotheses at any time sum to one, you are committed to taking a bet at *very* unfavorable odds that no plausible new theory will turn up. Surely no scientist would take this bet, or at least, not in most circumstances. Thus, the relation between subjective probabilities and betting is in this respect severed, a rather big deal for the Bayesian.

One way out of this difficulty is to regard the Bayesian setup as a model of only part of a scientist's total epistemic state; their opinions about unknown theories, in particular, are on this approach simply not represented within the Bayesian framework. For an example of this sort of use of the apparatus of BCT, see Strevens (forthcoming).

32. Exercise to the reader: suppose that the probability for a new theory h assumes the prior x. What happens if you Jeffrey conditionalize on this probability change? (Jeffrey conditionalization was discussed briefly in section 4.2.) You should be pleasantly surprised at your answer. Assume, as always, that the competing theories are mutually exclusive.

12. Further Reading

This is very much a work in progress (and is now also a bit out of date). Any suggestions for making the list less idiosyncratic and more complete (while bearing in mind the introductory function of these notes) is most welcome!

General Some helpful guides to BCT, already referenced many times in the main text, are Horwich (1982), Earman (1992), and Howson and Urbach (1993). All three have something useful to say about almost every issue discussed in the main text.

Mathematical Foundations On the elements of the mathematics of probability that are most relevant to BCT, Howson and Urbach (1993) is perhaps the best place to start.

On introducing conditional probability as a primitive and treating the "definition" of conditional probability as an additional axiom of the probability calculus, see Hájek (2003).

If you are at all interested in the history of the mathematics of probability, there are a number of excellent books in the area: Hacking (1975, 1990), Porter (1986), Stigler (1986), and Daston (1988) are all strongly recommended.

Epistemological Foundations A recent presentation of the idea that credences are psychologically real and related to, but distinct from, betting behavior, see Osherson et al. (1994).

There is a huge and ever-increasing literature on Dutch book and other approaches to arguing for the irrationality of credences that violate the axioms of the probability calculus. Howson and Urbach (1993) will point you to some of this work. It all begins with Ramsey (1931) and de Finetti (1964). For a recent, revealing overview, try Hájek (2005).

For the extension of Dutch book arguments to Bayes' rule, see Teller (1973). Howson and Urbach is also handy here. For a justification of Bayes' rule that does not make use of Dutch books, see van Fraassen (1989), chapter 13.

On the debate between a priori and pragmatic approaches to justifying BCT, take a look at some of the papers in Maxwell and Anderson (1975). Most of these writers are philosophers of science who favor the more pragmatic approach. Reading this volume in the 1970s, you might have predicted that apriorism was on the wane. Wrong; in the matter of ideology, never bet against the fanatics.

The question of the justification of the probability coordination principle was raised by Miller (1966), who argued that it was inconsistent. Lewis (1980) is a classic formulation of the principle and an equally classic discussion of its role in Bayesian thinking. For an attempt to justify PCP, see Howson and Urbach (1993, chap. 13). Strevens (1999) crtiticizes Howson and Urbach's strategy and argues that the problem of justifying the principle is on a par with the problem of induction. Lewis's recantation of his original formulation of the principle may be found in Lewis (1994); Strevens (1995) argues that there is no need to recant.

The subjectivist theory of physical probability was originally presented by de Finetti (1964), and has been expanded by Skyrms (1979).

On defining a quantitative measure of evidential relevance or support using subjective probabilities, see Fitelson (1999).

Bayesianism and Induction For the view that BCT does not, and ought not to, solve the problem of induction, see Howson (2001).

Goodman's new riddle: Earman (1992), §4.7 presents the modern consensus on BCT's handling of the problem. If you cannot get enough grue, read Stalker (1994).

The question whether simpler hypotheses should be regarded as more

plausible has a life of its own outside of BCT. There is a nice discussion in the Stanford Encyclopedia of Philosophy by Alan Baker; http://plato.stanford.edu/entries/simplicity/.

Problems of Confirmation Ravens: Hempel introduced the problem and gave an unpopular solution in Hempel (1945b); he also mentions Hosiasson's probabilistic solution, which went on to become very popular. Vranas (2004) argues that the probabilistic solution has some difficulties, as well as providing a useful bibliographical resource. The importance of the sampling procedure is emphasized by Horwich (1982) You will find the famous case of the black raven that disconfirms the raven hypothesis in Good (1967).

For an argument that variety in evidence is not always a good thing, see Bovens and Hartmann (2003), §4.4.

Hawthorne and Fitelson (2004) is a recent discussion of irrelevant conjuncts.

Subjectivity A physical probability version of the convergence result is presented by Savage (1954) and critiqued by Hesse (1975) (who I think underestimates the scope of such results).

For a good (though at times abstruse) overview of the subjective probability versions, see Earman (1992), chapter 6 (see also the later comment on p. 167 on the importance of "observational distinguishability" for such results).

A recent defense of a fairly radical personalism is Howson and Urbach (1993), §15i.

On the principle of indifference, van Fraassen (1989), chapter 12 is very helpful. For more advanced (and more favorable) treatments of the principle, see Rosenkrantz (1981) and Jaynes (1983). Strevens (1998) argues that even today, there is confusion between the principle of indifference and quite different methods used to infer, on a posteriori grounds, the values of physical probabilities.

More pragmatic "convergentist" and "contextualist" motivations of mild constraints on the priors have long been popular; see for example Shimony (1970) and, for the view on which frequencies in particular constrain priors, Salmon (1990).

Auxiliary Hypotheses The Quine-Duhem problem is raised by Duhem (1954). Quine's sentiment that the problem is insoluble is supposedly conveyed by his famous formulation "our statements about the external world face the tribunal of sense experience not individually but only as a corporate body" (Quine 1951); he claims as inspiration Carnap's *Aufbau*. A clear and uncompromising exponent of insolubility is Feyerabend (1975).

Dorling (1979) gave what I call the standard Bayesian solution to the problem. Bovens and Hartmann (2003), §4.5 investigate the behavior of auxiliary hypotheses given some very specific assumptions about the structure of the relevant probabilities.

For other answers to the question of how to evaluate ad hoc hypotheses, look at Howson and Urbach (1993), §7j.

Old Evidence Glymour (1980), chapter 3 is the classic exposition of the old evidence problem. Howson and Urbach (1993), §15g suggest a replay solution to the problem, arguing that all Bayesian conditionalization involves replay of some sort. Earman (1992) provides good coverage of the "learning entailment" view, with references to the original work of Garber, Jeffrey and Niiniluoto and others. A very helpful attempt to distinguish various different issues raised by old evidence can be found in Joyce (1999).

Many Bayesians and others have written about the difference, if any, between prediction and accommodation. See in particular Horwich (1982), chapter 5, Earman, §4.8, Howson and Urbach, §15h, and White (2003).

You will find the reserve or catch-all hypothesis characterized and put to use in Shimony (1970).

Other Objections to Bayesian Confirmation Theory The Bayesian system makes a number of epistemic idealizations, such as the assumption of logical omniscience that became salient in the discussion of the old evidence problem, or the assumption that all thinkers are sufficiently subtle in their epistemic musings that they assign a single, well-defined subjective probability to every possibility. Various writers have investigated the question of how to relax these assumptions.

Some writers have argued that BCT gives no weight to certain relevant aspects of the methods used to produce the evidence (Mayo 1996).

Proofs

These theorems should be interpreted as tacitly requiring that all conditional probabilities mentioned are well defined, that is, that the probability of what is conditionalized upon is non-zero.

Exercise 1. For every outcome e, $P(e) + P(\neg e) = 1$.

Proof. Because the outcomes e and $\neg e$ are mutually exclusive, by axiom 3,

$$P(e) + P(\neg e) = P(e \vee \neg e).$$

Since $e \vee \neg e$ is a tautology, it is inevitable, and so by axiom 2, $P(e \vee \neg e) = 1$, as desired. □

Exercise 2. For every outcome e, $P(e) \leq 1$.

Proof. By exercise 1, $P(e) + P(\neg e) = 1$, thus

$$P(e) = 1 - P(\neg e).$$

By axiom 1, $P(\neg e)$ is greater than or equal to zero. Thus $P(e)$ is less than or equal to one. □

Exercise 3. If e is logically equivalent to d, then $P(e) = P(d)$.

Proof. Since e and d are logically equivalent, e and $\neg d$ are mutually exclusive. Also, $e \vee \neg d$ is a tautology. By axioms 2 and 3, then,

$$P(e \vee \neg d) = P(e) + P(\neg d) = 1.$$

Thus $P(e) = 1 - P(\neg d)$. But also, by exercise 1,

$$P(d) = 1 - P(\neg d),$$

so $P(e) = P(d)$. □

Exercise 4. For any two outcomes e and d, $P(e) = P(ed) + P(e\neg d)$.

Proof. The outcomes ed and $e\neg d$ are mutually exclusive, so by axiom 3,

$$P(ed) + P(e\neg d) = P(ed \vee e\neg d).$$

The outcome $ed \vee e\neg d$ is logically equivalent to e, and so by exercise 3, it has the same probability as e. □

EXERCISE 5. For any two outcomes e and d such that e entails d, $P(e) \leq P(d)$.

Proof. By exercise 4,

$$P(d) = P(de) + P(d\neg e).$$

Because e entails d, de is equivalent to e. Thus by exercise 3,

$$P(d) = P(e) + P(d\neg e).$$

By axiom 1, $P(d\neg e)$ is non-negative, thus $P(e)$ is less than or equal to $P(d)$, as desired. □

EXERCISE 6. For any two outcomes e and d such that $P(e \supset d) = 1$ (where $\supset$ is material implication), $P(e) \leq P(d)$.

Proof. The proof is as for exercise 5, with the following amendment. It is no longer true that e is logically equivalent to ed, so we cannot use exercise 3 to deduce that $P(e) = P(ed)$. Instead we deduce the equality as follows. By exercise 4,

$$P(e) = P(ed) + P(e\neg d).$$

But if $P(e \supset d) = 1$, then, because $e \supset d$ is logically equivalent to $\neg(e\neg d)$, by exercise 3 and then exercise 1, $P(e\neg d) = 0$. Thus $P(e) = P(ed)$. □

EXERCISE 7. Bayes' theorem:

$$P(e|d) = \frac{P(d|e)}{P(d)}P(e).$$

Proof. As mentioned in the main text, all we need for the proof of Bayes' theorem is the definition of conditional probability.

$$\begin{aligned} P(e|d) &= \frac{P(ed)}{P(d)} \\ &= \frac{P(ed)}{P(d)}\frac{P(e)}{P(e)} \\ &= \frac{P(ed)}{P(e)}\frac{P(e)}{P(d)} \\ &= P(d|e)\frac{P(e)}{P(d)} \end{aligned}$$

as desired. □

EXERCISE 8. If $P(d) = 1$, then $P(e) = P(ed)$.

Proof. By exercise 4,

$$P(e) = P(ed) + P(e\neg d).$$

We would like to show that $P(e\neg d) = 0$. Since $P(d)$ is one, from exercise 1, we know that $P(\neg d)$ is zero. Because $e\neg d$ entails $\neg d$, we have by exercise 5

$$P(e\neg d) \leq P(\neg d) \quad \text{thus} \quad P(e\neg d) \leq 0,$$

and so by axiom 1, we can conclude that $P(e\neg d)$ equals zero, as desired. □

EXERCISE 9. The Bayesian's favorite version of the theorem of total probability: for mutually exclusive, exhaustive outcomes d_i,

$$P(e) = P(e|d_1)P(d_1) + P(e|d_2)P(d_2) + \cdots$$

where a set of outcomes d_i is exhaustive if $P(d_1 \vee d_2 \vee \dots) = 1$.

Proof. Because the outcomes ed_i are mutually exclusive,

$$P(e(d_1 \vee d_2 \vee \dots)) = P(ed_1) + P(ed_2) + \cdots$$

(compare the reasoning used for exercise 4). Then, because $P(d_1 \vee d_2 \vee \dots) = 1$, by exercise 8,

$$P(e) = P(ed_1) + P(ed_2) + \cdots .$$

From the definition of conditional probability, $P(ed_i) = P(e \,|\, d_i)P(d_i)$, we finally obtain our result. □

Exercise 10. $P(e) + P(d) = P(e \vee d) + P(ed)$.

Proof. The outcome $e \vee d$ is logically equivalent to $ed \vee e\neg d \vee d\neg e$. The three disjuncts are mutually exclusive, so

$$P(e \vee d) = P(ed) \vee P(e\neg d) \vee P(d\neg e).$$

Meanwhile, by exercise 4,

$$P(e) = P(ed) + P(e\neg d) \quad \text{and} \quad P(d) = P(de) + P(d\neg e).$$

Thus

$$\begin{aligned} P(e) + P(d) &= P(ed) + P(e\neg d) + P(de) + P(d\neg e) \\ &= P(ed) + P(e \vee d). \end{aligned}$$

as desired. □

Exercise 11. If e and d are independent, then

$$P(e) + P(d) = P(e \vee d) + P(e)P(d).$$

Proof. By exercise 10 and the definition of independence. □

Exercise 12. If $P(k) = 1$, then $P(d \,|\, ek) = P(d \,|\, e)$.

Proof. First note that by exercise 8, $P(ek) = P(e)$ and $P(dek) = P(de)$. Then

$$\begin{aligned} P(d \,|\, ek) &= \frac{P(dek)}{P(ek)} \\ &= \frac{P(de)}{P(e)} \\ &= P(d \,|\, e) \end{aligned}$$

as desired. □

EXERCISE 13. $P(e|h) > P(e)$ just in case $P(\neg e|h) < P(\neg e)$.

Proof. Assume that the probabilities of h and e are neither zero nor one (if not, the theorem is trivially true). Then $P(e|h) > P(e)$ just in case $P(he) > P(h)P(e)$ (in words, h and e are positively correlated). For the same reason, $P(\neg e|h) < P(\neg e)$ just in case $P(h\neg e) < P(h)P(\neg e)$. We will show that $P(he) > P(h)P(e)$ just in case $P(h\neg e) < P(h)P(\neg e)$. Observe that

$$\begin{aligned} P(he) + P(h\neg e) &= P(h) && \text{(by exercise 4), and} \\ P(h)P(e) + P(h)P(\neg e) &= P(h)\big(P(e) + P(\neg e)\big) \\ &= P(h) && \text{(by exercise 1)} \end{aligned}$$

Thus

$$P(he) + P(h\neg e) = P(h)P(e) + P(h)P(\neg e)$$

If the first term on the left hand side of the equals sign is greater than the first term on the right hand side, it follows that the second term on the left hand side must be less than the second term on the right hand side, as desired. □

Glossary

Bayes' Rule When evidence e (and nothing else) is observed, change your old subjective probability $C(h)$ for a hypothesis h to a new probability $C^+(h)$ equal to $C(h|e)$. Bayes' rule tells you how your subjective probability distributions at two different times ought to be related; by contrast, Bayes' theorem tells you about the relationship between different elements of your subjective probability distribution at a single time. See sections 4.1 and 5.1.

Bayes' Theorem $C(h|e) = C(e|h)C(h)/C(e)$. Bayes' theorem tells you about the relationship between different elements of your subjective probability distribution at a single time; by contrast, Bayes' rule tells you how your subjective probability distributions at two different times ought to be related. See sections 3.2 and 5.1.

Bayesian Multiplier When a piece of evidence comes in, the old probability for each hypothesis is multiplied by the Bayesian multiplier to determine the new posterior probability. The Bayesian multiplier is the likelihood divided by the probability of the evidence, or $C(e|h)/C(e)$. When the likelihood is set using PCP, it is equal to $P_h(e)/C(e)$. See section 5.1.

Empirical Equivalence From a Bayesian perspective, two theories are empirically equivalent if they assign the same physical probabilities to any piece of evidence, or better, if conjoined with the same auxiliary hypotheses, they assign the same physical probabilities to any piece of evidence.

Inadmissible Evidence Evidence is inadmissible if it invalidates the application of the probability coordination principle. If you possess evidence that is inadmissible relative to e and h, then, you may not, on the basis of PCP, set the subjective likelihood $C(e|h)$ equal to the physical likelihood $P_h(e)$. For example, if e is the proposition that a tossed coin lands heads, then in-

formation about initial conditions that allows you to deduce the probability for heads is inadmissible: you should set your subjective probability for e to one, not to the physical probability for e of one half. See tech box 5.5.

Likelihood The likelihood of some hypothesis on the evidence is the probability of the evidence given the hypothesis. (Note that although it is normal to talk of the likelihood of the hypothesis, a likelihood is in fact the probability of the evidence, not the hypothesis.) A *subjective likelihood* is a subjective conditional probability: the subjective likelihood of h on e is $C(e|h)$. A *physical likelihood* (some writers might say *objective likelihood*) is the physical probability ascribed to the evidence by the hypothesis $P_h(e)$, if any. See section 5.2.

Likelihood Lover's Principle The principle, entailed by BCT, that the degree to which a piece of evidence confirms a hypothesis increases with the physical likelihood of the hypothesis on the evidence (or the degree to which the evidence disconfirms the hypothesis increases as the physical likelihood decreases). The principle asssumes that subjective likelihoods are set equal to physical likelihoods, thus that there is no inadmissible evidence. See section 6.2.

Model The conjunction of a hypothesis and one or more auxiliary hypotheses. Of interest when the hypothesis alone does not assign a physical probability to the evidence, but the hypothesis plus auxiliaries does.

Prior Probability Either (a) the probability that you assign to a hypothesis before any evidence at all comes in, or (b) the probability you assign to a hypothesis right before some particular, salient piece of evidence comes in. See tech box 5.6.

Probability Coordination Principle The principle that enjoins you to set your subjective likelihoods equal to the corresponding physical likelihoods, that is, to set your subjective probability for e conditional on a hypothesis h, equal to the physical probability that h assigns to e (if any). Example: you should assign your subjective probability that a coin lands heads, conditional on the hypothesis that the physical probability of heads is one half, to one half. The principle does not apply if you have inadmissible evidence. See section 5.2.

To-morrow the rediscovery of romantic love,
The photographing of ravens . . .

W. H. Auden, Spain (April 1937)

References

Bovens, L. and S. Hartmann. (2003). *Bayesian Epistemology*. Oxford University Press, Oxford.

Carnap, R. (1950). *Logical Foundations of Probability*. University of Chicago Press, Chicago.

Daston, L. (1988). *Classical Probability in the Enlightenment*. Princeton University Press, Princeton, NJ.

Dorling, J. (1979). Bayesian personalism, the methodology of scientific research programmes, and Duhem's problem. *Studies in History and Philosophy of Science* 10:177–187.

Duhem, P. (1954). *The Aim and Structure of Physical Theory*. Translated by P. P. Wiener. Princeton University Press, Princeton, NJ.

Earman, J. (1992). *Bayes or Bust?* MIT Press, Cambridge, MA.

Feyerabend, P. K. (1975). *Against Method*. Verso, London.

Field, H. (1978). A note on Jeffrey conditionalization. *Philosophy of Science* 45:361–367.

de Finetti, B. (1964). Foresight: its logical laws, its subjective sources. In H. E. Kyburg and H. Smokler (eds.), *Studies in Subjective Probability*. Wiley, New York.

Fitelson, B. (1999). The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science* 66:S362–S378.

van Fraassen, B. C. (1989). *Laws and Symmetry*. Oxford University Press, Oxford.

Garber, D. (1983). Old evidence and logical omniscience in Bayesian confirmation theory. In J. Earman (ed.), *Testing Scientific Theories*, volume 10 of *Minnesota Studies in the Philosophy of Science*. University of Minnesota Press, Minneapolis.

Glymour, C. (1980). *Theory and Evidence*. Princeton University Press, Princeton, NJ.

Good, I. J. (1967). The white shoe is a red herring. *British Journal for the Philosophy of Science* 17:322.

Hacking, I. (1975). *The Emergence of Probability*. Cambridge University Press, Cambridge.

———. (1990). *The Taming of Chance*. Cambridge University Press, Cambridge.

Hájek, A. (2003). What conditional probability could not be. *Synthese* 137:273–323.

Hájek, A. (2005). Scotching dutch books? *Philosophical Perspectives* 17.

Hawthorne, J. and B. Fitelson. (2004). Re-solving irrelevant conjunction with probabilistic independence. Forthcoming, *Philosophy of Science*. Available online at http://fitelson.org/research.htm.

Hempel, C. G. (1945a). Studies in the logic of confirmation. *Mind* 54:1–26, 97–121. Reprinted in *Aspects of Scientific Explanation*, chap. 1, pp. 3–51. Free Press, New York.

———. (1945b). Studies in the logic of confirmation. *Mind* 54:1–26, 97–121.

Hesse, M. (1975). Bayesian methods and the initial probabilities of theories. In Maxwell and Anderson (1975).

Horwich, P. (1982). *Probability and Evidence*. Cambridge University Press, Cambridge.

Howson, C. (2001). *Hume's Problem: Induction and the Justification of Belief*. Oxford University Press, Oxford.

Howson, C. and P. Urbach. (1993). *Scientific Reasoning: The Bayesian Approach*. Second edition. Open Court, Chicago.

Jaynes, E. T. (1983). *Papers on Probability, Statistics, and Statistical Physics*. Edited by R. Rosenkrantz. D. Reidel, Dordrecht.

Jeffrey, R. C. (1983). *The Logic of Decision*. Second edition. University of Chicago Press, Chicago.

Joyce, J. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press, Cambridge.

Keynes, J. M. (1921). *A Treatise on Probability*. Macmillan, London.

Levi, I. (1967). *Gambling with the Truth*. MIT Press, Cambridge, MA.

Lewis, D. (1980). A subjectivist's guide to objective chance. In R. C. Jeffrey (ed.), *Studies in Inductive Logic and Probability*, volume 2. University of California Press, Berkeley, CA.

———. (1994). Humean supervenience debugged. *Mind* 103:473–490.

Maxwell, G. and R. M. Anderson, Jr. (eds.). (1975). *Induction, Probability, and Confirmation*, volume 6 of *Minnesota Studies in the Philosophy of Science*. University of Minnesota Press, Minneapolis.

Mayo, D. G. (1996). *Error and the Growth of Experimental Knowledge*. University of Chicago Press, Chicago.

Merton, R. K. (1968). The Matthew effect in science. *Science* 159:56–63.

Miller, D. (1966). A paradox of information. *British Journal for the Philosophy of Science* 17:59–61.

Osherson, D., E. Shafir, and E. Smith. (1994). Extracting the coherent core of human probability judgment: A research program for cognitive psychology. *Cognition* 50:299–313.

Porter, T. M. (1986). *The Rise of Statistical Thinking 1820–1900*. Princeton University Press, Princeton, NJ.

Quine, W. V. O. (1951). Two dogmas of empiricism. *Philosophical Review* 60:20–43.

Ramsey, F. (1931). Truth and probability. In *The Foundations of Mathematics and Other Logical Essays*. Edited by R. B. Braithwaite. Routledge and Kegan Paul, London.

Rosenkrantz, R. D. (1981). *Foundations and Applications of Inductive Probability*. Ridgeview, Atascadero, CA.

Salmon, W. C. (1990). Rationality and objectivity in science, or Tom Kuhn meets Tom Bayes. In C. W. Savage (ed.), *Scientific Theories*, volume 14 of *Minnesota Studies in the Philosophy of Science*. University of Minnesota Press, Minneapolis.

Savage, L. J. (1954). *The Foundations of Statistics*. Wiley, New York.

Shimony, A. (1970). Scientific inference. In R. G. Colodny (ed.), *Pittsburgh Studies in the Philosophy of Science*, volume 4. University of Pittsburgh Press, Pittsburgh.

Skyrms, B. (1979). Resilience, propensities and causal necessity. *Journal of Philosophy* 74:704–713.

Stalker, D. (ed.). (1994). *Grue!: The New Riddle of Induction*. Open Court, Chicago.

Stigler, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*. Harvard University Press, Cambridge, MA.

Strevens, M. (1995). A closer look at the 'New' Principle. *British Journal for the Philosophy of Science* 46:545–561.

———. (1998). Inferring probabilities from symmetries. *Noûs* 32:231–246.

———. (1999). Objective probabilities as a guide to the world. *Philosophical Studies* 95:243–275.

———. (2001). The Bayesian treatment of auxiliary hypotheses. *British Journal for the Philosophy of Science* 52:515–538.

———. (2003). *Bigger than Chaos: Understanding Complexity through Probability*. Harvard University Press, Cambridge, MA.

———. (2006). Probability and chance. In D. M. Borchert (ed.), *Encyclopedia of Philosophy*, second edition. Macmillan Reference USA, Detroit.

———. (Forthcoming). Reconsidering authority: Scientific expertise, bounded rationality, and epistemic backtracking. *Oxford Studies in Epistemology* 3.

Teller, P. (1973). Conditionalization and observation. *Synthese* 26:218–258.

Vranas, P. B. M. (2004). Hempel's raven paradox: A lacuna in the standard Bayesian solution. *British Journal for the Philosophy of Science* 55:545–560.

White, R. (2003). The epistemic advantage of prediction over accommodation. *Mind* 112:653–683.