

Varieties of Bayesianism*

Jonathan Weisberg
University of Toronto

For the Handbook of the History of Logic, vol. 10
Eds. Dov Gabbay, Stephan Hartmann, and John Woods

Contents

1	Introduction	3
1.1	The Standard Bayesian Machinery	3
1.2	Alternatives to the Standard Machinery	7
1.2.1	Primitive Conditional Probabilities	7
1.2.2	Infinitesimal Probabilities	8
1.2.3	Indeterminate Probabilities	9
2	Interpretations of Probability	12
2.1	The Logical Interpretation	12
2.2	The Degree of Belief Interpretation	14
2.2.1	Different Notions of Degree of Belief	15
2.3	Primitivism	17
2.4	Non-Epistemic Interpretations: Frequency and Chance	18
2.5	Summary So Far	20
3	The Subjective-Objective Continuum	21
3.1	Countable Additivity	22
3.2	Regularity	25
3.3	Conditionalization	27
3.4	Jeffrey Conditionalization	30
3.5	Infomin	33
3.6	Principles of Indifference	34
3.6.1	Discrete Indifference	35
3.6.2	Continuous Indifference	38
3.6.3	Summary of Principles of Indifference	42
3.7	Principles of Reflection	42

* Thanks to Frank Arntzenius, Kenny Easwaran, Branden Fitelson, James Hawthorne, Franz Huber, James Joyce, Phil Kremer, and Chris Meacham for helpful discussion and feedback.

3.8	Chance Principles	45
3.9	Summary So Far	46
4	Justifications	47
4.1	Dutch Book Arguments	47
4.1.1	The DBA for Probabilism	47
4.1.2	The DBA for Countable Additivity	49
4.1.3	The DBA for Conditionalization	49
4.1.4	The DBA for Jeffrey Conditionalization	51
4.1.5	The DBA for Reflection	51
4.1.6	The DBA for Regularity	52
4.1.7	Critical Discussion of DBAs	52
4.2	Representation Theorem Arguments	54
4.3	Axiomatic Arguments	56
4.4	Calibration and Accuracy Arguments	57
4.5	Expected Cognitive Utility and Conditionalization	58
4.6	Summary	59
5	Decision Theory	59
5.1	Savage's Decision Theory	60
5.2	Evidential Decision Theory	61
5.3	Causal Decision Theory	63
6	Confirmation Theory	68
6.1	The Naive Account and Old Evidence	68
6.2	Alternate Measures of Confirmation	69
6.3	The Popper-Rényi Solution	71
6.4	Objective, Ternary Solutions	72
6.5	Ravens and the Tacking Problem	74
7	Theories of Belief (a.k.a. Acceptance)	74
8	Summary	79

1 Introduction

Loosely speaking, a Bayesian theory is any theory of non-deductive reasoning that uses the mathematical theory of probability to formulate its rules. Within this broad class of theories there is room for disagreement along several dimensions. There is much disagreement about exactly what the subject matter of such theories should be, i.e. about what the probabilities in these theories should be taken to represent. There is also the question which probabilistic rules are the right ones. These two dimensions mark out the primary divides within the Bayesian viewpoint, and we will begin our Bayesian taxonomy by examining the most popular points along them. We'll then turn to another important point of disagreement: the kinds of justification that can and should be given for the aforementioned rules. Applications of the theory — to account for rational decision making, scientific confirmation, and qualitative belief — provide other significant points of intra-Bayesian dispute, and we will consider these questions later on. We begin, however, with an outline of the mathematical machinery that lies at the heart of Bayesianism.

1.1 The Standard Bayesian Machinery

Probabilities are numbers assigned to possibilities according to a few simple rules. We may choose either sets or sentences as our mathematical representations of the possibilities to which probabilities are assigned. If we use sets, then we start with a set of objects called the *outcome set*, denoted Ω . The singleton subsets of Ω represent fully-specified possible outcomes, and the other subsets represent less-than-fully-specified possible outcomes. For example, if we're considering a roll of a six-sided die, we might use the set

$$\Omega = \{1, 2, 3, 4, 5, 6\},$$

with $\{1\}$ representing the possible outcome where the die comes up 1, $\{2, 4, 6\}$ representing the possible outcome where an even-numbered face comes up, and so on. We then use a real-valued function on these subsets, p , to represent the assigned probabilities. A typical probability assignment might have

$$\begin{aligned} p(\{1\}) &= 1/6, \\ p(\{2, 4, 6\}) &= 1/2, \end{aligned}$$

and so on.

In simple cases like the die example, every subset of Ω is assigned a probability. But this won't always be the case, for reasons explained below (fn. 21). What

is always the case is that there is a σ -algebra on Ω that has probabilities assigned throughout:

σ -Algebra Given a set Ω , a σ -algebra on Ω is a set of subsets of Ω that contains Ω and is closed under complementation and countable union:

$$(\sigma_1) \sigma \subseteq \mathcal{P}(\Omega)$$

$$(\sigma_2) \Omega \in \sigma$$

$$(\sigma_3) \text{ If } A \in \sigma, \text{ then } \bar{A} \in \sigma$$

$$(\sigma_4) \text{ If } \{A_i\} \text{ is a countable collection of subsets of } \sigma, \text{ then } \bigcup \{A_i\} \in \sigma.$$

Clearly, $\mathcal{P}(\Omega)$ is a σ -algebra on Ω , and we typically assume that it is the σ -algebra under discussion unless specified otherwise.

Given an outcome set, Ω , and a σ -algebra on it, σ , we call the ordered pair (Ω, σ) an *outcome space*, and we define a probability function relative to it.

Probability Function Given the outcome space (Ω, σ) , a *probability function* on (Ω, σ) is a total, real-valued function on σ , p , satisfying the following three axioms:

$$(P_1) p(A) \in [0, 1] \text{ for each } A \in \sigma.$$

$$(P_2) p(\Omega) = 1.$$

$$(P_3) p(A \cup B) = p(A) + p(B) \text{ whenever } A \cap B = \emptyset.$$

The ordered triple (Ω, σ, p) is called a *probability space*, and jointly represents the bearers of probability as well as the probabilities they bear.

What if we use sentences rather than sets to represent the possibilities to which probabilities are assigned? In that case, we start with a standard propositional language built out of a countable set of atomic propositions, $\{A_i\}$, and the connectives \neg and \vee . We define the relation of logical entailment on \mathcal{L} , written \models , in the usual way, and the other standard logical connectives, \wedge , \supset , and \equiv , are defined out of \neg and \vee in the usual way as well. For some applications the language needs to be enriched into first-order logic, but for most purposes, and for us here, \mathcal{L} need only be standard propositional logic.

We call the ordered pair (\mathcal{L}, \models) a *logic*, and we define a probability function relative to it:

Probability Function A *probability function on the logic* (\mathcal{L}, \models) is a real-valued, total function on \mathcal{L} satisfying:

$$(P_1') p(A) \in [0, 1] \text{ for each } A \in \mathcal{L}.$$

$$(P_2') \quad p(A) = 1 \text{ if } \models A.$$

$$(P_3') \quad p(A \vee B) = p(A) + p(B) \text{ when } \models \neg(A \wedge B).$$

Call the triple $(\mathcal{L}, \models, p)$, a *logical probability space*.

The three axioms governing probability functions on logics, (P_1') – (P_3') , deliberately mirror those governing probability functions on outcome spaces, (P_1) – (P_3) . In fact, the set-based and sentence-based formulations are pretty much interchangeable in most contexts. For example, we have the following elementary consequences for probability spaces:

Proposition 1.1 $p(\emptyset) = 0$.

Proposition 1.2 $p(\bar{A}) = 1 - p(A)$ for any A .

Proposition 1.3 If $\{A_i\}$ is a finite set of mutually disjoint sets, then

$$p(\bigcup \{A_i\}) = \sum_i p(A_i).$$

And analogously, for logical probability spaces, we have:

Proposition 1.4 If $\models \neg A$ then $p(A) = 0$.

Proposition 1.5 $p(\neg A) = 1 - p(A)$ for any A .

Proposition 1.6 If $\{A_i\}$ is a finite set of mutually logically incompatible sentences, then

$$p(\bigvee \{A_i\}) = \sum_i p(A_i).$$

If the set-based and sentence-based approaches are so similar, why have both? In some ways the set-based approach is tidier, and hence preferable for certain applications. For example, in the set-based approach each way things could turn out is represented by exactly one thing, a set. On the sentence-based approach, however, there may be infinitely many logically equivalent sentences that represent the same eventuality. We can prove that they will all have the same probability, so that our choice of representation does not matter, but the fact that we have to verify this adds a minor complication to our machinery.

On the other hand, philosophical considerations can make the sentence-based approach more perspicuous. Consider, for example, the sentences “Superman will save the world today” and “Clark Kent will save the world today”. To Lois Lane, these two sentences represent very distinct eventualities, though in fact

they correspond to the same possible outcome. On the set-based approach, these two eventualities are represented by the same set, and hence must have the same probability. It seems reasonable, however, for Lois to think that they have different probabilities, which is allowed if we use different atomic propositions, A and B , to represent them. Then it is only once Lois realizes that A and B are equivalent that she must assign them the same probability:

Proposition 1.7 *For any A, B in a logical probability space, if $p(A \equiv B) = 1$, then $p(A) = p(B)$.*

Of course, we can get something similar out of the set-based approach, if we just interpret the elements of Ω so that they correspond more closely to what Lois regards as possible. But the sentence-based approach seems (to many, at least) to provide a more natural description of the situation.

Also, for many philosophical applications where the objects of probability are evidence and hypotheses, it is often more natural to talk about sentences or propositions, in which case the sentence-based approach is again more natural. Furthermore, some theorists seek to derive probabilistic facts from logical features of the hypotheses in question, so that a logical foundation is needed (see the discussion of Carnap below, p. 12).

For the most part our discussion will be ambiguous between the two approaches, slipping occasionally into one or the other approach for convenience. Our notation will often reflect this ambiguity. I will often use \bar{A} to mean $\neg A$, and AB as ambiguous between $A \wedge B$ and $A \cap B$. So, for example, I will write

$$p(A) = p(AB) + p(A\bar{B})$$

instead of the more cumbersome

$$p(A) = p(A \wedge B) + p(A \wedge \neg B).$$

These conventions are common and help to reduce clutter.

We need one final definition to complete our basic mathematical arsenal:

Conditional Probability The *conditional probability of B given A* is written $p(B|A)$, and is defined by

$$p(B|A) = \frac{p(BA)}{p(A)}$$

when $p(A) \neq 0$.

As we will see below (p. 8), some Bayesians prefer to take conditional probability as the basic notion, and treat unconditional probability as the defined one. Treating conditional probabilities as the derived notion is the more traditional approach, however, so we will use unconditional probability as our basic notion until we have reason to do otherwise.

1.2 Alternatives to the Standard Machinery

The mathematical system of probability that we just outlined is often thought to be inappropriately restrictive. For one thing, our definition of conditional probability leaves $p(B|A)$ undefined when $p(A) = 0$, though there are cases where it seems it should still be defined. For example, Hájek (2003) asks us what the probability is that a randomly chosen point on the Earth's surface will be in the Western hemisphere, given that it is on the equator. The intuitive answer is $1/2$. And yet the probability that the point will be on the equator is 0. So our framework does not allow us to have conditional probabilities where it seems we should have them.

Our framework also demands a level of precision that may not always be appropriate. Every probability that is defined is perfectly determinate, picking out a precise point in $[0, 1]$. But couldn't probabilities be less precise in some cases? Consider the probability that it will snow tomorrow. Where I am right now, Toronto in February, I can only say that this probability is greater than .05 but no more than .5. Maybe that lack of precision just reflects my ignorance about the true probability. But if we want to use probabilities to represent degrees of belief, as many do (see section 2.2), then those estimations of mine may just reflect indeterminacy in my opinions. Being inexpert about the weather as I am, I simply may not be able to cite any reasons that motivate a more precise opinion than, "greater than .05 but less than .5."

We'll now consider modifications to the traditional probabilistic framework that are designed to fix these, and other problems. Our first two variations on the traditional machinery address the problem of defining conditional probabilities when the condition has probability 0. The third variation addresses indeterminacy in probabilities.

1.2.1 Primitive Conditional Probabilities

The classic way to get around the zero-denominator problem for our definition of conditional probability is to reverse the order of definition, taking conditional probabilities as primitive and defining unconditional probabilities in terms of the conditional ones. This allows us to have the conditional probabilities defined in

all cases, but still be able to talk about probabilities unconditionally. There are several competing axiomatizations of conditional probability in the literature; ours is a minimal characterization in the spirit of the approach devised by Popper (1959) and refined by Rényi (1970):

Popper-Rényi Function Given an outcome space (Ω, σ) , a *Popper-Rényi function* is a real-valued, two-place function on $\sigma \times (\sigma - \{\emptyset\})$ satisfying the following axioms for any $A, B, C \in \sigma$:

(P₁'') $p(\cdot|A)$ is a probability function.

(P₂'') $p(A|A) = 1$.

(P₃'') $p(BC|A) = p(B|AC)p(C|A)$.

We then define the unconditional probability of A to be $p(A|\Omega)$. Then, from these axioms, we can recover the ratio formula for conditional probabilities:

Proposition 1.8 *If p is a Popper-Rényi function and $p(A) > 0$, then*

$$p(B|A) = \frac{p(BA)}{p(A)}.$$

So what was a definition of conditional probability is now a theorem about a basic concept. Since this ratio formula is only a constraint now, and not an analysis, Popper-Rényi probabilities allow us to have conditional probabilities even when the condition has probability 0.

1.2.2 Infinitesimal Probabilities

Another way around the zero-denominator problem is to ensure that the denominator is never 0. This looks impossible in cases like the Equator example. Consider the longitudes: there are uncountably many and it seems they should all be equally probable. But if we assign one of them a positive real number, then we must assign it to them all, in which case our probabilities will add up to more than 1, violating (P₁). So if we cannot assign 0, and we cannot assign a positive real, we must assign something in between: an infinitesimal.

Robinson (1966) showed that there must be such things as infinitesimals. Start with a first-order language containing a name for every real number, and state the theory of real analysis. Now state the theory containing all the sentences of the form $a < \omega$, where a is the name of a real number and ω is a new name of our choosing. Take the union of these two theories. Each finite subset of the unified theory has a model and hence, by compactness, the whole theory does. ω denotes

an infinite element of the model, one that is greater than all the reals. Better yet, $\epsilon = 1/\omega$ denotes an infinitesimal element of the model, one that is smaller than all the reals but greater than 0. Best of all, if we allow our probability functions to take on such infinitesimal values we can avoid assigning 0 to any of the latitudes, longitudes, or any set of points on the Earth’s surface. We can even arrange things so that the conditional probability of a randomly chosen point on the Earth’s surface lying in the Western hemisphere, given that it is on the equator, is $1/2$ (Skyrms 1995). To say much more than this, however, we have to say a lot more about the nature of infinitesimals, for which Skyrms refers us to (Parikh & Parnes 1974) and (Hoover 1980).

Infinitesimals take us into tricky and unfamiliar territory, so most Bayesians who want to solve the zero-denominator problem prefer to go with a Popper-Rényi-style solution. A result of McGee’s (1994) shows that the two approaches are to some extent equivalent, and (Halpern 2001) explores the equivalence further. But infinitesimals may have another use, as we’ll see in section 3.2.

1.2.3 Indeterminate Probabilities

Suppose that, for whatever reason, we want to allow probabilities to be indeterminate. There are two obvious ways to go about this. The first obvious way is to allow our probability functions to take on sets of values instead of single values, and the second is to talk about sets of probability functions rather than single functions.

Let’s take the first one first. If we have an indeterminate probability function returning sets of points instead of points, what constraints are reasonable? The most natural one is that our indeterminate probability function be, in some sense, resolvable into a determinate one. Resolvable how?

Precisification Given a function $\tilde{p} : \sigma \rightarrow \mathcal{P}([0, 1])$, a probability function p is a *precisification* of \tilde{p} if and only if $p(A) \in \tilde{p}(A)$ for each $A \in \sigma$.

We then define indeterminate probability functions as those that are ambiguous between their various precisifications:

Indeterminate Probability Function An *indeterminate probability function* on (Ω, σ) is a function $\tilde{p} : \sigma \rightarrow \mathcal{P}([0, 1])$ such that, whenever $x \in \tilde{p}(A)$, there is some precisification of \tilde{p} , p , for which $p(A) = x$.

Notice, we don’t just require that \tilde{p} be precisifiable, but also that every value it encompasses be part of a potential precisification. If we didn’t require this then \tilde{p} could include “junk” values in its outputs, ones that could never be the “true” probability, or a probability into which \tilde{p} gets resolved.

We can define conditional indeterminate probabilities in the obvious way:

Conditional Indeterminate Probability If \tilde{p} is an indeterminate probability function then the *indeterminate probability of B given A*, $\tilde{p}(B|A)$, is defined

$$\tilde{p}(B|A) = \{x : p(B|A) = x \text{ for some precisification } p \text{ of } \tilde{p}\}.$$

Using the same idea, we can talk about other operations on indeterminate probability functions, like sums and mixtures. Whatever the operation normally is, just apply it to each precisification and gather up the results.

The other obvious way to think about indeterminate probabilities was to work with classes of probability functions, rather than functions that returned classes of values. How do the two approaches compare? Clearly any class of probability functions picks out a unique indeterminate probability function, namely the one whose outputs include exactly those values assigned by some function in the class.

Ambiguation If P is a set of probability functions, the *ambiguation* of P is the indeterminate probability function that assigns for each A ,

$$\tilde{p}(A) = \{x : p(A) = x \text{ for some } p \in P\}.$$

The map that takes an indeterminate probability function to the class of its precisifications is clearly 1-1. However, the ambiguation of a set of probability functions can have precisifications not in the ambiguated set.

A natural extra constraint on indeterminate probability functions is to restrict the outputs to intervals. This is especially natural when we regard the indeterminacy as resulting from our ignorance about some objective but unknown probability function, where we have established lower and upper bounds but can't pin down the exact values.

When thinking in terms of classes of probability functions, the constraint most often added is convexity:

Convexity A class of probability functions, P , is *convex* if and only if whenever $p, q \in P$, every mixture of p and q is in P as well. That is, if and only if whenever $p, q \in P$ we also have $\alpha p + (1 - \alpha)q \in P$, $\alpha \in (0, 1)$.

Convexity is a bit harder to motivate than the interval requirement. Intuitively what it says is that the class of functions contains all the functions on the straight line between any two of its members. Why should we demand that? Levi (1980) endorses convexity on the grounds that a mixture of p and q can be seen as a sort of resolution of the conflict between the two states of opinion that p and q represent. Thus if we are torn between the state of opinion p and the state of

opinion q such that we suspend judgment, we should not rule out any compromise or resolution of that conflict.

Convexity is closely related to the interval requirement:

Proposition 1.9 *If P is convex with \tilde{p} its ambiguation, then $\tilde{p}(A)$ is an interval for each A .*

However, there are also non-convex sets of probability functions whose ambiguations are interval-valued.

If our indeterminate probabilities obey the interval constraint, it becomes natural to think about them in terms of the upper and lower bounds they place on the probability of each proposition.

Upper and Lower Probabilities If \tilde{p} is an indeterminate probability function, define

$$\begin{aligned}\tilde{p}_*(A) &= \inf\{x : x \in \tilde{p}(A)\} \\ \tilde{p}^*(A) &= \sup\{x : x \in \tilde{p}(A)\}.\end{aligned}$$

\tilde{p}_* and \tilde{p}^* are called the *lower* and *upper* probabilities of \tilde{p} .

Upper and lower probabilities are a quite general and well-studied way of representing uncertainty. In fact, they lead fairly naturally to one of Bayesianism's closest competitors, the Dempster-Shafer theory of belief functions (Shafer 1976). If we impose the requirement of n -monotonicity on lower probabilities:

n -monotonicity A lower probability function \tilde{p}_* is *n -monotone* if and only if for any A_1, \dots, A_n in σ ,

$$\tilde{p}_*(A_1 \cup \dots \cup A_n) \geq \sum_{I \subseteq \{1, \dots, n\}} (-1)^{|I|+1} \tilde{p}_*\left(\bigcap_{i \in I} A_i\right).$$

then we have the class of *belief functions* (Kyburg 1987), which are the topic of Dempster-Shafer theory (see the entry on Dempster-Shafer theory in this volume).

These are elementary remarks on indeterminate probabilities, and they serve only to make the reader aware of the possibility and to give a sense of the idea. For the remainder of our discussion, we will ignore indeterminate probabilities almost entirely. For further discussion, the reader is referred to (Levi 1974), (Jeffrey 1983), (Kyburg 1987, 1992), (van Fraassen 1990), (Walley 1991), and (Halpern 2003). Halpern and Kyburg are especially useful as surveys of, and inroads to, the literature on indeterminate probabilities.

2 Interpretations of Probability

Now that we have the mathematical framework of probability in hand, let's survey the sorts of epistemic subject matter it might be used to represent. (For a more detailed and historically-oriented survey of our first two interpretations, the logical and degree of belief interpretations, see the entry "Logicism and Subjectivism" in this volume.)

2.1 The Logical Interpretation

One way to think about probability is as a logical property, one that generalizes the deductive notion of logical validity.¹ Logical truths have probability 1 and logical falsehoods have probability 0, so it is natural to think of intermediate probabilities as "degrees" of logical truth. Similarly, the relationship of conditional probability generalizes the notion of logical entailment, with $p(B|A)$ representing the degree to which A entails B . Since we have

Proposition 2.1 *If $A \models B$ then $p(B|A) = 1$; if $A \models \neg B$ then $p(B|A) = 0$,*

it seems sensible to think of $p(B|A)$ as representing the extent to which A entails B , with $p(B|A) = 1$ representing the extreme case given by deductive validity. On this interpretation, the probability axioms (P1)–(P3) are to be understood as constraints on this generalized concept of validity, extending the usual rules of logical consistency.

But it's hard to understand the idea of partial entailment, since no analysis of the concept is given, and an analogy with the usual analyses of deductive entailment is not forthcoming. Usually deductive entailment is explained semantically, modally, or syntactically:

$A \models B \iff$ In every world/model where A is true, B is true.

$A \models B \iff$ Necessarily, if A then B .

$A \models B \iff B$ is deducible from A by the rules of a given logic.

The obvious way to extend the semantic definition is to say that A entails B to degree x iff $100x\%$ of the worlds/models where A is true are worlds where B is true. But then we need some way of counting or measuring sets of worlds, a notoriously problematic endeavor (see section 3.6.2). As for the modal and syntactic analyses, they don't seem to admit of any obvious extension to degrees.

¹ Keynes (1921) and Carnap (1950) were leading proponents of this interpretation in the 20th Century; Keynes attributes the origins of his views to Leibniz.

Etchemendy (1990) argues that none of the usual analyses of logical validity offer a proper analysis anyway. Instead they should be understood as providing converging approximations to a basic concept. Exploring semantic and syntactic validity and the relations between them helps us to refine our grip on our basic concept of logical validity, but does not tell us what logical validity is. But it wouldn't do for the proponent of the logical interpretation to say that, likewise, no analysis of *partial* logical entailment should be expected. Even if analyses of core logical notions cannot be demanded, we should expect at least some help in getting a grip on them — say, for example, by their approximate relationships to semantics, syntax, and modality. If similar approximating paths of convergence to partial entailment were given, then it would be on a par with deductive entailment, but there don't seem to be any such paths to help us converge on the postulated concept of partial entailment.

That's not to say that this sort of thing hasn't been tried. Carnap, one of the chief proponents of the logical interpretation, made a famous and heroic effort to characterize the concept of partial entailment syntactically. In his (1950), and later (1952), Carnap outlined basic principles from which a particular partial entailment relation on a simple language could be derived. This relation had some nice, intuitive properties that we might expect from a relation of partial entailment. For example, a larger sample of green emeralds entails "All emeralds are green" more strongly than a small sample does.

But because Carnap's method for specifying the partial entailment relation is bound to the syntax of the underlying language, it falls victim to Goodman's (1954) new riddle of induction. Define *grue* to mean "green and observed or blue and not observed". If we formulate the underlying language in terms of 'grue' rather than 'green', Carnap's partial entailment relation will tell us that our observation of a large sample of green emeralds highly entails "All emeralds are grue," since the green emeralds we have observed are all grue. But "All emeralds are grue" says that all emeralds not yet observed are blue, and it seems that this should not be highly entailed by our observation of many green emeralds. Insofar as we understand partial entailment, it does not seem it should be a purely syntactic matter, as Carnap's approach made it.² Carnap's approach is also regarded as problematic because it uncovered many candidates for the relationship of partial entailment (the famous " λ -continuum"), and none of them seems to be especially self-recommending, not to mention logically required.

² For more on Goodman's riddle, see the entry "Goodman on the Demise of the Syntactic Approach" in this volume.

2.2 The Degree of Belief Interpretation

Carnap's difficulties led many Bayesians to embrace the (already existing) degree-of-belief interpretation of probability, according to which probabilities represent a subject's levels of certainty. On this view, the force of the probability axioms (P₁)–(P₃) is that of constraints on rational belief. Someone whose degrees of belief violated those axioms, say by thinking rain tomorrow 70% likely and no rain 50% likely, would violate a canon of rationality.

Many notable thinkers in this tradition have thought that this rational force is on a par with the rational force of the rules of deductive logic — that the probability axioms provide the rules for degrees of belief just as the laws of deductive logic provide the rules of full belief. Some even claim that one whose degrees of belief violate the probability axioms is logically inconsistent in the deductive sense. But their view is still very different from the logical interpretation. Even if general rules of probability like (P₁)–(P₃) express logical facts, statements of particular probabilities represent psychological facts. For them, to say that $p(B|A) = 1/3$ is to say that the ratio of my degree of belief in AB to my degree of belief in B is $1/3$, not that there is any logical relationship between A and B that achieves a degree of $1/3$.

In fact, many in the degree-of-belief tradition have thought that there are few or no rules of rationality for degrees of belief beyond (P₁)–(P₃). According to them, there are many degrees of belief one can have for $p(B|A)$, any of which would be reasonable to hold (except in the special case where A deductively entails either B or its negation). Thus there is no such thing as the degree to which A entails B , merely the degrees of belief one can hold without violating the canons of logic.

The degree-of-belief interpretation is sometimes called the “subjective” interpretation, because probabilities are used to describe a subjective state, degree of belief. But ‘subjectivism’ has other connotations: specifically, the view just mentioned, according to which the probability axioms capture (nearly) all the rules of rationality for degrees of belief, thus leaving a great deal of room for reasonable inter-subjective variation. Because the degree-of-belief interpretation is characterized by the *particular* subjective state it concerns, degree of belief, I prefer the more explicit title ‘the degree of belief interpretation’, and I save ‘subjectivism’ for the view that the rules of rationality leave much room for inter-subjective variation. We will discuss subjectivism in this sense later (section 3).

2.2.1 Different Notions of Degree of Belief

Bayesians differ on exactly what degrees of belief are. The first precise accounts were given by Ramsey ([1926] 1990) and de Finetti (1937). Their definitions were heavily operationalist, defining degrees of belief in terms of what a person prefers, or would choose if given the option. This operationalist tendency may reflect the heavy influence of logical positivism at the time, motivating those using a psychological interpretation to offer empirically respectable analyses of the psychological state they were postulating.³

Let's start with de Finetti's definition, since it is simpler. According to de Finetti, your degree of belief in a proposition A is the odds at which you would regard a bet on A that pays \$1 as fair. For example, if you are willing to pay \$.50 but no more to play a game that pays \$1 if A comes true and nothing otherwise, that must be because you think it 50% likely that A is true. For then you would stand a 50% chance of gaining \$.50, but a 50% chance at losing \$.50. Were your degree of belief in A lower, you would think it more likely that you would lose the \$.50 than gain it, and would not regard the bet as fair. And if your degree of belief were higher, you would think yourself more likely to gain it than lose it, making the bet unfair (if advantageous).

This sort of definition has been criticized for failing to take into account factors other than the subject's degrees of belief that might affect her betting behavior (see, for example, (Earman 1992; Weatherson 1999; Christensen 2001)). For example, she might be risk averse, in which case her betting behavior would underrate her degrees of belief. She might also have other motives or interests besides money, like entertainment or showing off. She could even be misled by the format in which the deal is offered, failing to appreciate its consequences or falling prey to a framing effect.

The standard response to such worries is to specify an idealized setup in which all such factors have been ruled out, and to say that your degrees of belief are the odds you would regard as fair in that idealized scenario (Goldstick 2000). There is, of course, the worry that such conditions cannot be spelled out non-circularly. There is also the worry that comes with any counterfactual analysis, namely that moving to such a remote, hypothetical situation may alter the subject's epistemic state so that the degrees of belief that guide her there are not the actual ones that we were trying to mine.

³ Ramsey actually prefaces his definition saying, "It is a common view that belief and other psychological variables are not measurable, and if this is true our inquiry will be vain [...] for if the phrase 'a belief two-thirds of certainty' is meaningless, a calculus whose sole object is to enjoy such beliefs will be meaningless also." (Ramsey [1926] 1990: p. 166) The worry that we can move from 'not measurable' to 'meaningless' suggests a logical positivist perspective.

Ramsey's approach is similar in spirit but more sophisticated and technical. Rather than assume that our subject is interested in money, Ramsey tries to extract her degrees of beliefs from her preferences whatever they may be. To do this, Ramsey proved his famous representation theorem, which says roughly this:

Representation Theorem Suppose our subject's preferences are rational in that they obey a set of constraints C (not specified here). Then there is exactly one probability function-utility function pair, (p, u) , that represents the agent's preferences in this sense: she prefers A to B if and only if the expected utility of A relative to p and u is greater than that of B .⁴

The idea is to use the theorem to reverse-engineer the agent's beliefs and desires from her preferences. Her beliefs are given by p and her desires by u , since these are the only functions that coincide with her preferences, and hence must be the beliefs and desires that generated her preferences. Since Ramsey proved his theorem, variations to the same effect have been proved by Savage (1954), Jeffrey (1965), and others.⁵

A classic criticism of the Ramsey-Savage-Jeffrey approach is that actual people do not obey the constraints on preferences, C , assumed by the theorem. For example, the famous Allais paradox (Allais 1979) has been shown by Kahneman and Tversky (1979) to habitually lead people into violations of Savage's key Axiom of Independence (section 4.2). Even the elementary constraint of transitivity — that you should prefer A to C if you prefer A to B and B to C — has been claimed to be habitually violated by actual people (Lichtenstein & Slovic 1971, 1973).

Savage and others typically regard the constraints in C as normative, not descriptive. Hence they see no problem if people do not obey the constraints assumed by the theorem, so long as it is still true that they *should* obey them. But even if the constraints in C are normatively correct, this does nothing to ameliorate the problem for Ramsey and others who want to use representation theorems to *define* 'degree of belief'. For if nobody satisfies the constraints needed by the theorem, then nobody's degrees of belief can be defined via the theorem. One might respond that the theorem still offers a definition of degree of belief for ideal agents who do satisfy the constraints in C . But it is questionable whether one can (without begging the question) show that ideal agents really would have the degrees of belief attributed to them by the theorem. And it is also unclear how that would help us understand what degrees of belief are for actual people. For further

⁴ See section 5 for a definition of 'expected utility'.

⁵ There are also non-Bayesian representation theorems, which allow for preferences that violate the usual constraint-set C in certain ways, and which deliver non-Bayesian representations of those preferences. See, for example, (Kahneman & Tversky 1979) and (Wakker & Tversky 1993).

development and discussion of these worries, see Zynda (2000), Christensen (2001), and Meacham and Weisberg (unpublished).

If the standard, operationalist definitions fail, what can be put in their place? We might take degrees of belief as primitive on the grounds that they are theoretically fruitful (Eriksson & Hájek 2007). We might also hope that a precise characterization will ultimately be provided implicitly in our psychological theorizing. Psychological theorizing about uncertain reasoning has boomed in the last 25 years, and we might take the naturalist attitude that degrees of belief are roughly characterized by our folk-psychological theory of confidence/certainty, and that the concept will be refined as that theory is refined by the empirical work in psychology. Of course, there is the possibility that, at the end of the day, nothing remotely like our folk concept will show up in the finished psychological theory, in which case the degree of belief interpretation will be out of luck. But the meaningfulness of our normative theorizing about a state always hangs on that state actually existing, and we should always be prepared for the eventuality that our empirical theorizing will eliminate it, rendering our normative theorizing otiose.⁶

2.3 Primitivism

Another take on the interpretation of probability is one we might call *primitivism*, according to which Bayesianism can be understood as a theory about an important epistemological notion of probability that need not be given any analysis (and maybe can't be given one). Timothy Williamson (2000) frankly adopts this stance. Williamson doesn't necessarily reject the degree of belief interpretation of probability, but he thinks that there is an important concept of probability that is not captured by that interpretation, and cannot be captured in any analysis. If someone asks, "how likely is the theory of evolution given our evidence?", Williamson thinks that there is a definite, and objective answer, one that needn't correspond to any actual (or ideal) person's degrees of belief.⁷ Such questions are about a concept Williamson dubs *evidential probability*. Evidential probability is something like the confirmational notion of probability pursued by Carnap, except that Williamson does not endorse the analogy with deductive entailment, and rejects any attempt to specify probabilities syntactically. Any attempt to

⁶ Compare, for example, the argument that virtue ethics is undermined by empirical research showing that the character traits it presupposes do not exist (Doris 2002).

⁷ Maher (1996) takes something of an intermediate stance between the degree of belief interpretation and Williamson's primitivism. On Maher's view there is a (fairly) definite and objective answer to such questions, but the answer is the degree of belief one ought to have (not the degree of belief one actually does have).

analyze evidential probability would, on Williamson's view, be a mistake on a par with trying to analyze modality or sets. Instead, Williamson thinks our approach should be to go ahead and use the concept in our epistemological theorizing. Our grip on the concept will strengthen as our theory develops and evidential probability's role is outlined, as happens with possibility in modal logic and sets in set theory.

2.4 Non-Epistemic Interpretations: Frequency and Chance

There are other classic interpretations of probability that may be legitimate but which do not yield an epistemological theory, and so do not enter into our taxonomy of Bayesian theories. Still, they deserve mention since they are widely discussed and are important to Bayesian theory in other ways.

The first is the *frequency interpretation*, according to which the probability of an event is the frequency with which such things occur. For example, the probability that the coin I am about to flip will come up heads is $1/2$ because this coin and coins like it come up heads half the times they are flipped. As stated, this interpretation is ambiguous in two respects. First, we need to know which coins count as "coins like this one". If trick coins that are asymmetrically weighted or have heads on both sides count as "like this one", then the frequency of heads may be something other than $1/2$. This is the famous problem of the reference class: with reference to which class of coin-flips should we calculate the frequency of heads to determine *this* coin's probability of coming up heads? There is an accompanying ambiguity that is less often stressed but is also crucial: what is it we are supposed to count the frequency of? If we want to know the probability of the sentence "the coin will come up heads", we might count the frequency with which such coins come up heads, the frequency with which they come up at all (instead of disappearing or landing on a side), or even the frequency with which such coins are coins. Which feature of the sentence is it that we must count the frequency of to determine the sentence's probability?

These two ambiguities correspond to the two variables F and G in the question, "how many F s are G s?". Given a sentence whose probability we want to know, we must specify what F and G are before we can say what the frequency is that determines the probability. For a sentence like, "the coin-flip will come up heads," what F and G are supposed to be may be obvious. But it isn't always obvious, as is shown by sentences like "Louis XIV wore brown trousers on Jan. 14, 1713." So a recipe for specifying F and G must be given for this interpretation to have much content.

Setting these problems aside, why isn't the frequency interpretation part of our taxonomy of Bayesian theories? Because frequency is not an epistemic subject,

strictly speaking. There are no facts about what the frequencies ought to be, nor do frequencies represent some epistemic relationship like confirmation, partial entailment, or evidential probability. That is not to say that frequencies are epistemically irrelevant. Far from it. Frequencies are one of the most useful kinds of data we can have for determining a theory's degree of logical truth/degree of belief/evidential probability. But every Bayesian theory acknowledges that probabilities can be used to represent frequencies⁸ and that frequencies are a crucial kind of data. The interesting question is what we should do with our knowledge of frequencies. Should they determine our degrees of belief, our judgments about partial entailment, or our judgments about evidential probability? And what are the rules by which they determine these judgments? Why is the frequency with which we observe green emeralds relevant in a way that the frequency of grue emeralds is not? Frequencies are just one kind of data out of many, and the fact that these data are important is not controversial. What is controversial is what kind of epistemic facts or states these data determine, and what the exact determination relation is.

The same is true for another common interpretation known as the *physical chance* interpretation. On this interpretation, probabilities represent objective physical properties of objects or events. The classic exemplars of physical probability are the probabilities in quantum mechanics. For example, the probability that a given atom of U_{238} will decay in the next 10 minutes is supposed to be an objective, physical fact that is entirely independent of the epistemic status of any agent. It is a property of that atom, or of the events it is a part of, and does not merely reflect our ignorance about the state of some hidden variables since, ostensibly, there are none.

Another, more contentious example of physical chance comes from statistical mechanics. Because classical statistical mechanics is compatible with a deterministic underlying physics, some think that its probabilities cannot be physical chances. Lewis (1980), for example, holds that the chances in a deterministic universe are always 0 or 1. Others, however, hold that statistical mechanical chances must be objective and physical since they would otherwise be incapable of doing the explanatory work demanded of them (Albert 2001; Loewer 2001). Statistical mechanical probabilities are supposed to explain why ice cubes always melt when we put them in water, but if the chances were not something physical — and especially if they are merely something epistemic — then they could not *explain* why this occurs. They would, at best, explain why we expect it to occur, which is

⁸ It is an elementary truth of mathematics that frequencies obey the probability axioms. Given a set of F s, the ratio that are G s in the set is always a (rational) real number in $[0, 1]$, the ratio that are F s is always 1, and the ratio that are either G s or G' s is the ratio that are G s plus the ratio that are G' s, if G and G' are incompatible.

not the same as explaining why our expectations are repeatedly met.

The probabilities in special sciences like biology and sociology are also held by some to be physical chances, but these cases are even more contentious. At any rate though, most agree that quantum mechanical probabilities are examples of physical chances.

The physical chance interpretation is sometimes called the “propensity” interpretation, but this label is not entirely neutral, conveying a more specific view about what physical chances are like. On the propensity interpretation, physical chances are likened to dispositions like fragility and solubility. Just as a thing can be more or less fragile, a thing can have a higher or lower propensity to do a certain thing. On the propensity view, physical chances are just the measures of these propensities. Not all authors who believe in physical chances endorse the analogy with propensities, however.

With respect to our taxonomy of Bayesian theories, physical chances are in the same boat as frequencies. Some think there are such chances and some think there are not, but this debate is a matter of physics or metaphysics, not epistemology. But, like frequencies, chances are very epistemically important if they exist, since they would guide our degrees of belief, evidential probabilities, or whatever. Chances will enter into our taxonomy when we consider the particular epistemic norms endorsed by various Bayesian theories, since the exact way in which chances should guide degrees of belief is a point of disagreement amongst Bayesians. But as far as the subject matter of our epistemic theorizing goes — the question what epistemic properties, states, or relations are being subjected to normative rules by our theory — chances are not a point of decision.

2.5 Summary So Far

Let’s summarize our taxonomy so far. We began with the question: what epistemic states or relations can probability be used to model and formulate norms about? We considered three candidate answers. First there was the logical interpretation, according to which probability generalizes logical validity and entailment to allow for intermediate degrees. Problems with understanding how logical entailment could be a matter of degree led us instead to the degree of belief interpretation, according to which probabilities represent levels of confidence and the rules of probability theory are normative rules governing them. We also considered various ways of explicating the notion of degree of belief and their drawbacks. Finally, we considered the primitivist stance, which treats probability as its own, basic epistemic notion, for which an analysis is neither demanded nor given.

These interpretations are not necessarily incompatible. We might think that all these epistemic phenomena exist, and that they are all appropriately modeled

using probabilities. For example, a pluralist might think that we can have a degree of belief about the evidential probability of the level of partial entailment between *A* and *B*.

Whichever interpretations we accept as appropriate, the next question we face is what precise rules they ought to follow. When does *A* entail *B* to degree *x*? Given that I am *x* confident in *A*, how confident should I be in *B*? What is the evidential probability of *B* given my evidence? These kinds of questions bring us to the second dimension of our taxonomy. In addition to the three basic rules that define Bayesianism, the probability axioms (P₁)–(P₃), what other rules should we include in our theory?

3 The Subjective-Objective Continuum

Probabilism is the view that degrees of belief (or degrees of entailment or evidential probabilities) should obey the probability axioms, (P₁)–(P₃). But Probabilism leaves a tremendous amount unsettled, since many wildly different probability assignments can satisfy (P₁)–(P₃). Take a coin-flip for example. An assignment of 1 and 0 to heads and tails satisfies (P₁)–(P₃), but so would an assignment of 0 and 1, 1/2 and 1/2, 1/3 and 2/3, etc. This means that our Bayesian theory so far allows for a wide range of inter-subjective variation, since you may choose one assignment and I another without either of us violating any rule. Thus our theory so far is heavily “subjective”.⁹ If we were to add more rules beyond just (P₁)–(P₃), we would move more and more towards objectivity as we ruled out more and more possible assignments. At the objective end of this spectrum is a theory that specifies a single probability distribution for each individual or situation.

On the degree-of-belief interpretation, this spectrum of subjectivity is just a matter of how much room there is for people’s beliefs to reasonably vary. On the other interpretations it’s less clear what a more subjective theory amounts to. If our theory of partial logical entailment demands only (P₁)–(P₃), does that mean

⁹ The “subjective vs. objective” terminology is deeply entrenched, but can be very misleading. Presumably, any correct theory of non-deductive reasoning must allow for massive inter-subjective disagreement, since different people have different evidence, conceptual frameworks, interests, etc. The idea typically had in mind, I think, is that more subjective brands of Bayesianism allow for inter-subjective disagreement even between agents with the same evidence. But one might want to acknowledge other epistemological factors besides evidence that determine the probabilities an agent should assign, e.g. attention and epistemic interests.

A scale from “subjective” to “objective” thus obscures a central question: which variables should determine a subject’s probabilities? Just evidence, or are there other properly epistemic factors to consider? Once that question is settled, we can place views on a subjective-continuum according to how much they think these variables do to determine a subject’s probabilities. Without having settled which variables are relevant though, there is a danger of cross-talk.

that degrees of entailment are generally indeterminate? What about for primitive or evidential probability? On these interpretations a theory at the subjective end of the spectrum looks dangerously at odds with the very notion of probability offered. But even on the degree of belief interpretation a heavily subjective theory looks problematic. Someone who is convinced that heads will come up but can offer no epistemological reason for their conviction seems completely unreasonable. The mere fact that he satisfies (P₁)–(P₃) because he simultaneously has no confidence in tails does little to raise our estimation of his rationality.

So all parties have an interest in filling out their theory to push it more towards the objective end of the spectrum. This is usually done by formulating specific rules, over and above (P₁)–(P₃), to eliminate possible probability distributions. Which such rules should Bayesians endorse? Here we consider the most common proposals.

3.1 Countable Additivity

According to (P₃), the probability of two disjoint possibilities is the sum of their individual probabilities. We can derive from there that the same sort of additivity applies to any finite number of disjoint possibilities (Proposition 1.3). But what about a countable infinity of disjoint possibilities, does additivity apply there too? (P₁)–(P₃) do not entail that it does. There are probability distributions satisfying (P₁)–(P₃) but violating

Countable Additivity If $\{A_i\}$ is a countably infinite set of mutually disjoint sets, then

$$p\left(\bigcup_i A_i\right) = \sum_i p(A_i).$$

The existence of probability distributions satisfying (P₃) but not Countable Additivity can be proved using the ultrafilter lemma.¹⁰ So we have the option of adding countable additivity to our theory or not. Should we add it?¹¹

De Finetti (1970; 1972) famously argued that we should reject Countable Additivity, since it rules out fair, countable lotteries. Couldn't God pick out a natural number at random? Not if Countable Additivity holds. Any probability function that assigned the same value to each natural number would have to assign

¹⁰ The ultrafilter lemma is a consequence of ZFC set theory, though not of ZF, so this way of proving the existence of such distributions is non-constructive. I do not know whether a construction is possible.

¹¹ Many authors contemplate adding a condition called 'continuity' which, in the context of (P₁)–(P₃), is equivalent to Countable Additivity. But the content and point of Countable Additivity is much more readily apparent, so our discussion is cast in terms of it.

them all 0. If we assigned some $a > 0$ to each number, we would violate (P₃) since some finite collection of numbers would have individual probabilities adding up to more than 1. But if we assign 0 everywhere, we violate Countable Additivity:

$$p\left(\bigcup_{i \in \mathbb{N}} \{i\}\right) = 1 \neq \sum_{i \in \mathbb{N}} p_i = 0.$$

But, de Finetti thought, the possibility of a fair, countable lottery is a real one, so Countable Additivity cannot be true.

On the other hand, Countable Additivity is crucial to the proof of some allegedly important theorems. There is a family of theorems known as *Bayesian convergence theorems* which say, in various more precise forms, that a Bayesian agent’s conditional probabilities on true data are certain to converge to certainty in the truth. For example, consider the following gloss of a theorem from Gaifman and Snir (1982):

Gaifman-Snir Theorem Let \mathcal{L} be a first-order language with finitely many “empirical terms”, and $\{E_i\}$ a sequence of sentences that “separates” the models of \mathcal{L} . For each model ω of \mathcal{L} , let $\omega(A) = 1$ if A is true on ω , 0 otherwise. Finally, let E_i^ω denote E_i if $\omega(E_i) = 1$, $\neg E_i$ otherwise. Then for any H in \mathcal{L} ,

$$\lim_{n \rightarrow \infty} p(H \mid \bigwedge_{0 \leq i < n} E_i^\omega) = \omega(H)$$

almost everywhere.¹²

This says, in effect, that given sufficiently fine-grained evidence, the probability of H will almost certainly converge to 1 if H is true and 0 if it is false. Such theorems are supposed to provide a vindication of Bayesianism, showing that it is guaranteed to find the truth eventually, whatever the truth turns out to be. But they require Countable Additivity to be proved.

Many do not think these theorems provide any real vindication, and would not miss them if they were lost. For one thing, the theorems only show that the convergence will happen “almost everywhere”, i.e. on a set of models with probability 1, where that certainty is judged by the probability function whose success is in question. From an impartial perspective, one that does not assume that the agent’s initial probabilities have any bearing on the truth, this guarantee is no guarantee at all. Moreover, the convergence is only guaranteed to happen at infinity, which is of no interest to actual people, who will surely be dead in the long run.

¹² See (Gaifman & Snir 1982: p. 507) for the full and precise statement of the theorem. A clear explanation can also be found in (Huber 2005).

Countable Additivity's role in the derivation of these theorems has perhaps been most forcefully criticized by Kelly (1996). Kelly points out that the fairness de Finetti complained was ruled out by Countable Additivity is actually crucial to the convergence theorems. Countable Additivity serves to rule out as having probability 0 those possible scenarios in which the agent will not converge on the truth. Thus the agent assures herself of success by, in effect, starting with a 100% bias against the eventuality that she might fail. For an excellent discussion of convergence theorems, the role of Countable Additivity in their proofs, and the accompanying criticisms, see chapter 13 of Kelly's (1996).¹³

There are convergence theorems that do not depend on Countable Additivity, so even those who value convergence results might not miss Countable Additivity if it were abandoned. For example, James Hawthorne has pointed out to me that his Likelihood Ratio Convergence Theorem (Hawthorne 2008) does not depend on Countable Additivity. It should be noted, however, that Hawthorne's theorem does not guarantee convergence to the truth in the same sense that the Gaifman-Snir theorem does. What Hawthorne's theorem guarantees is that the likelihood ratio of a true hypothesis to a false hypothesis will converge, which is different from the conditional (or "posterior") probability of a true hypothesis converging to 1. On the plus side, the precise kind of convergence that Hawthorne guarantees has a finitistic character that may immunize it against the "dead in the long run" objection (see (Hawthorne 2008) for the details). On the minus side, Hawthorne's theorem only guarantees convergence with probability 1, where that certainty is again judged by the probability function whose convergence is in question. Thus Hawthorne's result is, like the Gaifman-Snir theorem, not perspective-neutral in an important way. It expresses a kind of self-affirmation, whereby each probability distribution sees itself as sure to converge to the truth. Such a result might be seen as speaking more to internal consistency than to truth-orientedness.

There is a standard argument for axioms (P1)-(P3), the Dutch book argument, which can be extended to Countable Additivity. The Dutch book argument purports to show that violating the axiom in question puts one in the dubious position of regarding a deal as fair even though it is a simple matter of deductive logic to show that it will lead you into a sure loss. Since such a state of mind is supposed to be irrational, so is violating the axiom (Adams 1964; Williamson 1999). Dutch book arguments are questionable, as we will see, and the Dutch book argument for Countable Additivity is especially questionable thanks to certain anomalies of infinitary decision problems. We will discuss the matter more fully in section 4.1.7, but for now it suffices to say that the Dutch book argument

¹³ Interestingly, Kelly's own favored methodology of formal learning theory is a prime candidate for the "dead in the long run" objection, since its results primarily concern convergence at infinity. Kelly responds to this concern in his (1996) and (2000).

for Countable Additivity is controversial.

How decisive are these considerations? The intuitive possibility of a fair, countable lottery seems a weak consideration on which to reject such a powerful principle. On the other hand, Dutch book arguments are not terribly well regarded in general, and especially not when it comes to a countable number of possibilities. Given that Kelly's analysis puts a point on the fact that Countable Additivity forces us into biased probability distributions, we might think Countable Additivity too suspect to embrace. And yet, it may be the very moral of Hume and Goodman's puzzles about induction that good inductive reasoning is of necessity biased. Apparently, the available arguments do not settle the issue.

3.2 Regularity

Another commonly cited constraint is *regularity*, which says that we should not assign probability 0 to any possible state of affairs. We can distinguish two versions of the regularity principle:

Initial Regularity At the initial stage of inquiry, before any evidence is acquired, nothing except a contradiction should be assigned 0 probability.

Continuing Regularity At every stage of inquiry, nothing except a contradiction should be assigned 0 probability.

The motivating thought behind Initial Regularity is that we should begin our inquiry with an open mind, only ruling things out once our evidence warrants it. Continuing Regularity is motivated by the further dictum that we can never rule anything out for sure. Because Continuing Regularity will figure significantly in our discussion of Jeffrey Conditionalization (section 3.4), I will focus on Initial Regularity here.

Initial Regularity seems a perfectly sensible bit of advice, and complaints about it usually derive from technical problems with its implementation. We saw a typical problem case back in section 1.2.3, when we tried to assign probabilities to a randomly chosen latitudes on the Earth's surface. Each latitude had to get the same probability, but any positive real would make the collective probabilities add up to more than 1, contra (P1). The only solution, it seems, is to assign a probability greater than 0 but less than every positive real, i.e. an infinitesimal.

Turning to infinitesimals to solve this technical problem has a few philosophical sticking points. First, on the degree of belief interpretation, it may not be plausible to talk of infinitesimal probabilities. Folk psychology certainly doesn't deal in infinitesimal levels of confidence, and it's hard to see how a more refined theory could. There's always the temptation to try for an operational definition,

maybe in terms of conditional bets or infinitesimal stakes, but we've seen that operational definitions are passé for a reason (section 2.2.1). To make things worse, Hájek points out that infinitesimals are "ineffable". We know they exist because we can prove it, but the proofs are non-constructive. The Robinson proof we sketched earlier (section 1.2.2) appeals to compactness to tell us that there is a model containing infinitesimal elements, but does not provide a particular model by construction.¹⁴ So we don't seem able to actually use infinitesimals in our epistemology, since we have no way of glomming onto a particular one to say, "this one, this is the one that is the probability in question!" This is a problem for the infinitesimal treatment in general, and makes the notion of an infinitesimal degree of belief look especially spooky.¹⁵ The final sticking point with infinitesimals is that they violate countable additivity... sort of. As Hájek notes, a countable sum of infinitesimals cannot be expressed because the definition of the appropriate limit cannot be expressed in the language of non-standard analysis, wherein infinitesimals live (Hájek 2003: p. 280).

If we can't turn to infinitesimals to satisfy our demand for Initial Regularity, maybe we can use a bit of philosophical therapy instead. Recall that the alleged problem with assigning 0 to each latitude is that it represents a sort of closed-mindedness. But am I really being closed-minded by assigning the minimal probability to this eventuality? After all, it's one of an uncountable infinity of equiprobable possibilities, and thus is about as unlikely as anything could be. It's not on a par with a logical contradiction or metaphysical impossibility, of course, but my attitude towards the point ending up on the equator isn't necessarily the same as my attitude towards the possibility of there being a round square. The kinds of reasons I have for these two beliefs are different, and so they will have different levels of resilience. The equator proposition could be easily raised in my estimation by all kinds of evidence. But pretty much nothing could convince me that there's a round square outside. Furthermore, I am aware of the different metaphysical and epistemic statuses of these two possibilities, and so my reasoning about them will be of different kinds. In short, there is more than just the probability I assign that distinguishes the equator proposition from the possibility of a round square, and this might help us to come to terms with their both having 0 probability.

¹⁴ There are constructions of models containing infinitesimals, e.g. that of Kanovei and Shelah (2004), which make use of ultra-filters. As Philip Kremer and Kenny Easwaran have explained to me, the existence of ultra-filters is proved non-constructively, using the axiom of choice. So the ineffability of ultra-filters infects models so constructed.

¹⁵ Though whether the unavailability of a construction really does make infinitesimals less accessible than other mathematical objects will depend on one's views in the epistemology and semantics of mathematics.

Of course, this bit of therapy will have to be coupled with a regimen of mathematical role-reversal, since we will have to appeal to Popper-Rényi functions (section 1.2.1) or something similar to have the appropriate conditional probabilities be defined. And there may be a problem understanding Popper-Rényi functions on the degree of belief interpretation of probability. It seems kosher to talk of my degree of belief in AB and in A , and thus of their ratio. We understand belief as an attitude towards a proposition, and we just need to apply the notion of gradation to that concept, the result of which we may already have in the concepts of confidence and certainty. But can we make sense of degree of belief as an attitude towards a pair of propositions? Or have we moved so far from folk-psychology that we don't really have a grip on what we're talking about? There may be precedent for such "binary" propositional attitudes in the concept of desire, since many things we desire we do not desire simpliciter, but only given this or that. (Would you rather have fish or the meat? Well, it depends which is fresher or which restaurant you're at.) But conditional desire may itself be parasitic on desire simpliciter. Supposing you have an absolute ranking of all the possibilities in mind, then you might prefer A given C but B given D because the best C -possibilities are A -possibilities, but the best D -possibilities are B -possibilities. To my knowledge, this is an unexplored concern about the Popper-Rényi approach, but it may well not be serious. And, at any rate, Popper-Rényi degrees of belief are not nearly as spooky as infinitesimal ones.

As a final remark, Dutch book type considerations have been called on to support Regularity (Shimony 1955). Assuming that someone whose probability for A is 0 must bet at odds of 1:0 on \bar{A} , such a probability assignment is worrisome because it means that you should be willing to bet any amount of money on \bar{A} for the chance to win absolutely nothing. While this won't lead you to a certain loss, it *may* lead you to a loss and *cannot* lead to any gain, a position sometimes described as "not strictly coherent".

3.3 Conditionalization

So far we have been concerned solely with synchronic rules, rules that constrain what probabilities must be like at a time, as opposed to how they must be related across time. But for a Bayesian, probabilities are central to inductive inference, and so we must know how to change them over time as data comes in and circumstances change. How do we do this? The classic Bayesian answer is

Conditionalization When you acquire new evidence E , the new probability of any proposition H should be the previous conditional probability of H given E . That is, $q(H) = p(H|E)$.

Conditionalization has some elementary, desirable properties.

Proposition 3.1 *If p is a probability function and $q(H) = p(H|E)$ for each H , then q is a probability function too.*

Thus the deliverance of conditionalization is always a probability function.

Proposition 3.2 *If q comes from p by conditionalizing on E_1 , and r comes from q by conditionalizing on E_2 , the result of conditionalizing on E_2 first and then E_1 would have been the same, namely $r(\cdot) = p(\cdot|E_1E_2)$.*

So the order in which we gain information doesn't affect the final probabilities, so long as the same total information is gleaned in the end, as seems appropriate.

Conditionalization not only gives intuitively plausible answers in many applications, but also unifies some classic truisms about the confirmation of theories by evidence. Philosophers of science have long observed that theories are confirmed when their predictions are borne out, refuted when their predictions are not borne out, especially strongly supported when otherwise implausible predictions are borne out, and so on. These truisms find a natural place when we look at conditional probabilities in one of their most classic formulations:

Proposition 3.3 (*Bayes's Theorem*) *For any H and E where $p(E) > 0$,*

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

If H predicts E , then $p(E|H)$ is high, typically making the ratio on the right high, and so $p(H)$ gets multiplied by a number greater than 1, making $p(H|E) > p(H)$. Thus, when we conditionalize on E , the probability of H goes up. Similarly, if H predicts \bar{E} but we find E , H 's probability will typically go down since $p(E|H)$ will be low, and thus $p(H)$ will typically be multiplied by a number near 0. And if H predicts E where E would be surprising otherwise, then the fraction on the right is especially large, raising the probability of H significantly when we conditionalize on E .

It is important to note that discussions of Conditionalization are often ambiguous between two importantly different interpretations of the same formal idea. Suppose we are understanding probabilities as degrees of belief. Then what Conditionalization says is that, when you learn new information E , you should adopt as your new degree of belief in H your previous conditional degree of belief in H given E , *whatever your degrees of belief may happen to have been just before you learned E* . On this understanding of the rule, the probabilities you plug into the rule are just whatever degrees of belief you happen to have when you learn E . But

what if your degrees of belief right before you learned E were badly misguided, treating E and H as mutually irrelevant when they ought to be regarded as importantly related? This thought prompts a different reading of Conditionalization, one where the probabilities plugged into the rule are not the degrees of belief you happen to have at the moment but, rather, something more objective. An obvious candidate is the probability function that you *ought* to have had — the probabilities that would have been reasonable given your available evidence up to that time. This reading of Conditionalization, however, requires constraints strongly objective enough to specify what degrees of belief would be “reasonable” given your evidence up to that time. Thus which reading of Conditionalization is used usually depends on the author’s leanings on the subjective-objective continuum.

These remarks, it should be noted, apply quite generally to pretty much any update rule, including the ones about to be discussed in the next few sections. Such rules always take in “prior” probabilities, modifying them in light of the new evidence. Which probabilities should be regarded as the appropriate prior ones — e.g. the degrees of belief you happened to have vs. the objectively correct evidential probabilities — is a major point of disagreement between objectivists and subjectivists. That said, this ambiguity in the interpretation of diachronic rules will be suppressed for the remainder of the section.

The downside to Conditionalization is that it flouts Continuing Regularity (section 3.2) by giving probability 1 to all evidence:

Proposition 3.4 *If q comes from p by conditionalizing on E , then $q(E) = 1$.*

But we are rarely if ever entitled to be absolutely certain that our evidence is true. Notice that, using Conditionalization, we could not distinguish the different grades of plausibility amongst various pieces of our evidence.

Another common complaint about Conditionalization is that it guarantees that evidence not only becomes certain, but stays certain forever. If all changes in our probabilities happen by Conditionalization, then certainties stay certainties. Conditionalizing on anything consistent with a certainty leaves it as a certainty:

Proposition 3.5 *If $p(E_1) = 1$ then $p(E_1|E_2) = 1$ for any E_2 consistent with E_1 .*

And we can’t conditionalize on anything inconsistent with past evidence, since we can’t conditionalize on anything inconsistent with a certainty:

Proposition 3.6 *If $p(E_1) = 1$ then $p(A|E_2)$ is undefined whenever E_2 is inconsistent with E_1 , since $p(E_2) = 0$.*

So evidence stays certain once certain, and evidence to the contrary cannot even be assimilated (unless we turn to Popper-Rényi functions (section 1.2.1)). These prob-

lems led Jeffrey (1965) to offer a less demanding extension of Conditionalization, and many have followed him.

3.4 Jeffrey Conditionalization

Jeffrey motivated his rule by considering cases of unclear observation, like the observation of a cloth in dim candlelight, where the color appears red but might also be green or blue. Suppose the probabilities the experience licenses in the red, green, and blue hypotheses are $q(R)$, $q(G)$, and $q(B)$. What, then, is the probability of H , that this is your sock and not your roommate's? Well the appearance of the cloth tells you something about its color, but nothing about how its color bears on the probability of it being your sock as opposed to your roommate's. So we can set $q(H|R) = p(H|R)$, and similarly for G and B . We are then in a position to calculate $q(H)$:

$$\begin{aligned} q(H) &= q(H|R)q(R) + q(H|G)q(G) + q(H|B)q(B) \\ &= p(H|R)q(R) + p(H|G)q(G) + p(H|B)q(B). \end{aligned}$$

The example suggests the following as a general rule:

Jeffrey Conditionalization When an observation bears directly on the probabilities over a partition $\{E_i\}$, changing them from $p(E_i)$ to $q(E_i)$, the new probability for any proposition H should be

$$q(H) = \sum_i p(H|E_i)q(E_i).$$

Notice that Jeffrey Conditionalization gets Conditionalization as a special case when the partition is $\{E, \bar{E}\}$ and $q(E) = 1$.

Proposition 3.7 *If q is obtained from p by Jeffrey Conditionalization on the partition $\{E, \bar{E}\}$ with $q(E) = 1$, then $q(\cdot) = p(\cdot|E)$.*

The driving assumption behind Jeffrey Conditionalization is that the bearing of the observation is captured entirely by the new probabilities on the partition, so that the conditional probabilities on the elements of the partition need not be changed. It is the probabilities of the E_i that we have been informed about, not their evidential bearing on other questions. This key assumption, that the conditional probabilities on $\{E_i\}$ should remain unchanged, is called *rigidity*.

Historically, the most common complaint about Jeffrey Conditionalization has been that it is not indifferent to the order of evidence in the way that Conditionalization is (Proposition 3.2). (See, for example, (Levi 1967b), (Domotor 1980)

and (van Fraassen 1989) for this complaint.) For example, suppose we use Jeffrey Conditionalization to get q from p and then r from q , using the partition $\{E, \bar{E}\}$ both times with $q(E) = x$ and $r(E) = y$, $x \neq y$. Now reverse the order, using y first and then x . In the first case the final probability of E is x , in the second case it will be y . So order matters.

Lange (2000) counters that this kind of order-dependence is not problematic, since it does not correspond to an order-reversal of anything whose order should not matter. For example, the quality of the observations yielding the x and y values in the first scenario will not be the same as in the second. To illustrate, let E be the statement that the raven outside is black, let $x = 1/10$, and let $y = 9/10$. Supposing I start with $p(E) = 1/2$, then the first scenario must be one in which the raven appears pretty clearly non-black at first glance (a shift from $1/2$ to $1/10$), but then looks *very* clearly black at a second glance (a shift from $1/10$ to $9/10$). If we reverse the order of the numbers, however, then the raven must have seemed pretty clearly black on the first glance ($1/2$ to $9/10$), but *very* clearly non-black on the second ($9/10$ to $1/10$). So reversing the order of the input values does not amount to having the same qualitative experiences in reverse order. Why, then, should we expect order-invariance when we reverse the order of the input probabilities?

Another concern about Jeffrey Conditionalization is that it is incomplete in a very important way. Without some supplementary rule telling us which partition an experience bears on, and what probabilities on that partition the observation warrants, we cannot apply the rule. Jeffrey Conditionalization needs a partition and a distribution over it as inputs, and we haven't been told how to select these inputs. It's worth noting that a similar problem afflicts Conditionalization, since it does not specify when a proposition counts as new evidence, and should thus be conditionalized on. The problem is especially acute for Jeffrey Conditionalization, however, since it is nearly vacuous without any constraints on what the inputs should be. If we may select any inputs we like, then we can get any q from a given p . We just use the set of singletons in Ω as the partition and use as input probabilities whatever probabilities we want to end up with for q .¹⁶

Field (1978) approached this problem, sketching a rule for assigning input probabilities to experiences. Garber (1980) showed that Field's proposal had the unwelcome consequence that repetition of the same experience could boost a hypothesis's probability without bound, even though it seems no real information should be gained from redundant observations. Wagner (2002) shows that Field's proposal is actually the only one that will make Jeffrey Conditionaliza-

¹⁶ For Ω 's larger than \mathbb{R} , we may not be able to get just any q from a given p , but we will still have a tremendous amount of freedom.

tion order-invariant on experiences (though Wagner actually takes his result to vindicate Jeffrey Conditionalization in the face of the order-invariance objection). Christensen (1992) worried that Jeffrey Conditionalization might be in tension with the epistemological doctrine of holism, according to which a belief's empirical justification is always sensitive to background assumptions. He points out that the tension becomes an outright clash when Jeffrey Conditionalization is supplemented by Field's rule. Weisberg (2009) argues that, in light of Wagner's result, Christensen's worry about holism becomes a dilemma for Jeffrey Conditionalization: Wagner shows that Field's rule is the only way to make Jeffrey Conditionalization commutative, but Field's rule is anti-holistic, so Jeffrey Conditionalization cannot satisfy both commutativity and holism.

An entirely different sort of concern about Jeffrey Conditionalization is that it may not be general enough. Does the rule apply in every case where we need to change our probabilities? Let's set aside cases where we *lose* information rather than gain new information, say through memory degradation, cognitive mishap, or loss of self-locating information as in the Sleeping Beauty puzzle (Elga 2000). Even setting aside all these kinds of probability changes, there are alleged cases where evidence is gained but cannot be treated by Jeffrey Conditionalization. Perhaps the most famous is van Fraassen's Private Judy Benjamin Problem (van Fraassen 1981). Judy Benjamin is dropped from a plane in the middle of a field, parachuting to the ground. The field is divided up into four sectors of equal area — NW, NE, SW, and SE — and Judy thinks the probability that she is in a given quadrant is $1/4$ in each case. She radios base to describe her surroundings and they tell her that they cannot say whether she is in the North or the South, but they can say that, if she is in the North, the odds are 3:1 that she is in the East. What should Judy's new probabilities be?

The trouble is supposed to be that what Judy has learned is a conditional probability, $p(NE|N) = 3/4$, rather than a distribution over a partition. So Jeffrey Conditionalization does not apply. To handle this sort of problem we might look for a more general rule (see the next section). But that might be too hasty. Arguably, the problem is only apparent, arising from an overly simplistic representation of Judy's probabilities. What Judy has really learned, after all, is that home base reported a conditional probability $p(NE|N) = 3/4$. However Judy conceives of those probabilities — as home base's degrees of belief, as physical chances, or whatever — she is only disposed to accept them as her own in so far as she trusts home base's opinions. But then her acceptance of a conditional probability of $p(NE|N) = 3/4$ is based on the evidence that home base reported that they thought $3/4$ to be the correct conditional probability. So Judy should just conditionalize her prior probabilities on the fact that home base made the report they did. There is no need for a new rule that takes conditional probabilities

as inputs (Grove & Halpern 1997). Richard Bradley (2005) offers other examples where the evidence is, allegedly, properly represented as a conditional probability, thereby stumping Jeffrey Conditionalization. Bradley's examples may be amenable to similar treatment, but they may not.

3.5 Infomin

If we are not satisfied with the Conditionalization-based treatment of the Judy Benjamin problem, or we think that other examples show a need to be able to take conditional probabilities as inputs, then we should look for a more general rule. The best-known proposal is that of information minimization, or *Infomin*:

Infomin If your initial probability function is p and your evidence mandates a probability function in the set S , adopt the probability function $q \in S$ that minimizes the quantity

$$H(p, q) = \sum_{i \in \Omega} p_i \log \left(\frac{p_i}{q_i} \right),$$

where $p_i = p(\{i\})$ and similarly for q_i .

Infomin has the virtue of being extremely general. Not only could we constrain S by insisting on a particular conditional probability, we could also constrain it by insisting on unconditional probabilities over a non-exclusive set of propositions, or even by insisting on a set of expectation values.¹⁷ Notice, though, that Infomin is incomplete in the same way that (Jeffrey) Conditionalization is, since it does not say how a particular experience, observation, or bit of evidence fixes a constraint on S . Just as those rules need to be supplemented with a rule determining the inputs, so does Infomin.

Why adopt Infomin? The usual advertisement is that Infomin makes the minimal changes in p necessary to meet the constraints imposed by the evidence. $H(p, q)$ is commonly regarded as a measure of the difference in information between p and q , and p being your existing opinions, you should seek to preserve them as much as possible while still respecting your new evidence. Also to its credit, Infomin has the nice property of agreeing with Conditionalization and Jeffrey Conditionalization when they apply (Williams 1980).

Proposition 3.8 *If S contains just those probability distributions on (Ω, σ) that assign the values x_i over the partition $\{E_i\}$, then $H(p, q)$ is minimized when $q(\cdot) = \sum_i x_i p(\cdot | E_i)$.*

¹⁷ For a nice discussion of exactly how general Infomin is, see (Williams 1980).

But Infomin is not the only generalization that agrees with (Jeffrey) Conditionalization, nor is H the only way to measure the distance between probability functions. As a number of authors have pointed out, H isn't even symmetric — $H(p, q) \neq H(q, p)$ — so it's not a proper metric. Variational distance (a.k.a. Kolmogorov distance), defined

$$\delta(p, q) = \frac{1}{2} \sum_{i \in \Omega} |p_i - q_i|,$$

is a more standard way of measuring the distance between functions, one that does satisfy the definition of a metric.

Proposition 3.9 *For any probability functions p , q , and r ,*

$$\delta(p, p) = 0,$$

$$\delta(p, q) = \delta(q, p), \text{ and}$$

$$\delta(p, q) + \delta(q, r) \geq \delta(p, r).$$

If we minimize $\delta(p, q)$ instead of $H(p, q)$, we have a rule that still agrees with (Jeffrey) Conditionalization but sometimes disagrees with Infomin. Interestingly though, the posterior probability delivered by Jeffrey Conditionalization is not always the unique one that minimizes δ , whereas Jeffrey Conditionalization does uniquely minimize H . For discussion of these concerns and the options in this area, see (Diaconis & Zabell 1982) and (Howson & Franklin 1994).

3.6 Principles of Indifference

The rules suggested so far still leave tremendous room for inter-subjective variation. When faced with a coin about to be flipped, we can assign any probability to heads we want (except 1 and 0, thanks to Regularity), provided we just assign one minus that probability to tails. While Conditionalization and its successors determine your new probabilities given the ones you start with, what probabilities you may start with is still pretty much a free-for-all. Countable Additivity and Initial Regularity just don't rule out very much. But our next proposal aims to rectify this problem, specifying a unique starting distribution.

3.6.1 Discrete Indifference

The classic principle for assigning initial probabilities is the *Principle of Indifference*,¹⁸ which says that you should assign the same probability to each possibility in the absence of any (relevant) evidence.

Principle of Indifference Given a finite outcome set Ω with cardinality N , if you have no (relevant) evidence then assign $p(\omega_i) = 1/N$ to each singleton $\omega_i \subseteq \Omega$.

When Ω is countably infinite, we can extend the principle to mandate $p(\omega_i) = 0$, assuming we are not endorsing Countable Additivity. But then the Principle of Indifference does not determine all of p , since we cannot determine the probabilities of infinite sets by additivity. And if we do accept Countable Additivity, then the principle has no application for a countable Ω .

The qualifier “in the absence of any (relevant) evidence” is deliberately vague and ambiguous.¹⁹ There are at least two interpretations to be considered. First, the principle might be viewed as a way of assigning *ur-probabilities*, the probabilities one ought to assign absent any evidence whatsoever, and which ought to determine one’s probabilities when one does have evidence by updating on the total evidence. This interpretation is troubled by the fact that the full space of possibilities we are aware of may be too large for the Principle of Indifference to apply. There is a formulation for outcomes spaces of cardinality $|\mathbb{R}|$, which we will consider momentarily (section 3.6.2). Yet we know that there are more than $|\mathbb{R}|$ possible ways our universe could be. For every element in $\mathcal{P}(\mathbb{R})$, there corresponds the possibility that precisely its members describe the physical constants of the universe. Most of these possibilities are highly improbable, but they are certainly possible. To my knowledge, the Principle of Indifference has not been extended to larger-than- $|\mathbb{R}|$ outcome spaces.

A second interpretation makes crucial use of the parenthetical “relevant” in “(relevant) evidence”. On this interpretation, the Principle of Indifference is not to be applied to the grand-world problem of assigning ur-probabilities to the total space of possibilities. Rather, it applies to agents who have evidence, but whose evidence is not relevant to a certain partition; that is, their evidence does not favor any one element of the partition over any other (or maybe doesn’t favor any of

¹⁸ Originally it was called ‘The Principle of Insufficient Reason’, but Keynes renamed it in an attempt to make it sound less arbitrary.

¹⁹ It is also something of a modernism: early statements of the principle, most notoriously those of Laplace and Bernoulli, spoke of “equipossible” outcomes, rather than outcomes for which relevant evidence is lacking. See (Hacking 1971) for an excellent historical discussion of the notion of equipossibility and the classic formulation’s connections to the modern one.

them at all). In that case, Ω represents the partition in question, and the Principle of Indifference tells us to assign the same probability to each ω_i . A concern for this interpretation is that the principle becomes vacuous. It tells us to assign the same probability to ω_1 and ω_2 when our evidence doesn't favor either one over the other, but what does it mean for our evidence not to favor either one, if not that they should have the same probability?²⁰

Use of the Principle of Indifference in actual practice tends to be more in the spirit of the second interpretation. Statements of the principle commonly include the “relevant” qualifier or something similar, and the outcome spaces to which the principle is applied are typically finite or continuous, and do not represent the full space of possibilities in maximal detail. Such applications might still be consistent with the first interpretation: evidence that is not “relevant” to a finite or continuous partition might be understood as evidence which, once the ur-probabilities are conditionalized on it, yields uniform probabilities over the partition in question. Nevertheless, we will proceed to look at applications of the principle with the second interpretation in mind.

For toy cases like the flip of a coin or the roll of a die, the Principle of Indifference looks quite sensible. If you have no information about the coin, the principle will tell you to assign a probability of $1/2$ to both heads and tails. And for the die it will tell you to assign $1/6$ to each face. But some cases yield terrible results. To illustrate, suppose that the coin is to be flipped ten times. The possible outcomes here are the 2^{10} possible 10-bit sequences of the sort

HTTTHHTTTH.

The Principle of Indifference instructs you to regard each such sequence as equally likely. If you do, what will be the probability that the last toss will come up heads given that the first nine do? Let H_i mean that the i -th toss was a heads. Then

$$\begin{aligned} p(H_{10}|H_{1-9}) &= \frac{p(H_{1-10})}{p(H_{1-9})} \\ &= \frac{1/2^{10}}{1/2^9} \\ &= 1/2. \end{aligned}$$

In fact, that probability will be the same no matter how many times the coin is to be flipped and how many times it comes up heads. So the Principle of Indifference enforces a kind of stubborn anti-inductivism.

²⁰ This concern goes back at least as far as (von Mises [1928] 1981) and (Reichenbach 1949).

This result led Carnap (1950) to prefer a different application of the principle, where we assign the same probability to each hypothesis of the form “ m heads and n tails”, and then divide that probability equally amongst the possible sequences instantiating that hypothesis. This yields a much more sensible probability distribution, one that starts out assigning $1/2$ to each H_i but increases that estimate as more and more heads are observed.

Another approach to the same problem, proposed by Jon Williamson (2007), appeals to implicit background information encoded in our linguistic formulation of the problem. The Principle of Indifference is only supposed to apply when we have no relevant evidence, and Williamson’s idea is that we do have this kind of evidence in these cases, since we know that H_{10} would be another instance of the same property, heads, as is instantiated 9 times in H_{1-9} . This background information rules out the probability distribution that gives the same probability distribution to each possible sequence of heads and tails. (Exactly what this knowledge rules out and how it does so takes us into details of Williamson’s proposal that I won’t go into here.)

If our implicit, linguistic knowledge rules out the uniform distribution mandated by the Principle of Indifference, what probabilities should we assign? There is a generalization of the principle that we can apply in its stead, called the *Principle of Maximum Entropy*, or

MaxEnt If your evidence is consistent with all and only those probability distributions in the set S , then use the probability distribution p that maximize the quantity

$$-H(p) = -\sum_{i \in \Omega} p_i \log(p_i).$$

MaxEnt is strongly reminiscent of InfoMin (section 3.5), and in fact employs the same basic idea. H was supposed to measure information, and entropy defined as $-H$ is its opposite. Minimizing information is the same as maximizing entropy. The difference is that, whereas before we sought to minimize the change in information relative to our initial probabilities, now we are setting our initial probabilities by minimizing the amount of absolute information. MaxEnt just applies Infomin to find the distribution that is informationally closest to assigning 1 to each $\{i\}$. (We have to take it on faith that minimizing information relative to this assignment amounts to minimizing information absolutely.)

MaxEnt is a generalization of the Principle of Indifference, selecting the uniform distribution recommended by indifference when S includes it.

Proposition 3.10 Suppose p is the uniform distribution on (Ω, σ) and $p \in S$. Then p maximizes $-H$ on S .

But we can also apply MaxEnt when the uniform distribution isn't in S . To solve the problem we are facing, that the uniform distribution can be stubbornly anti-inductive, Williamson uses Bayes-nets and the linguistic structure of the problem to rule the uniform distribution out of S , and then applies MaxEnt. The end result is not the uniform distribution but, rather, the "straight-rule". This distribution says that the conditional probability of H_i is the frequency with which heads has been observed so far.

The straight-rule is actually not such a great result to end up with. It assigns $p(H_2|H_1) = 1$, so after observing just a single heads we conclude with certainty that the next flip will be heads too. But we might be able to give a formula for narrowing S down that yields a more sensible distribution. There is a deeper problem that deserves our attention instead.

Williamson appeals to knowledge implicit in the syntactic features of our language in order to shape S . It's supposed to be the match between the predicates we use to formulate H_{10} and H_{1-9} that tells us that these propositions are inductively related. But grounding induction and probability in syntax is notoriously problematic, as Goodman showed and we saw in section 2.1. If the agent uses a gruesome language to describe her hypotheses, she will end up with a different S and hence a different distribution. Williamson might not be troubled by this, since he might think that an agent who speaks a gruesome language has different implicit knowledge and so *should* assign different probabilities. But, thanks to Goodman, we have *both* 'green' and 'grue' in our language, so we can't use our language to shape S according to Williamson's recipe. Clearly Williamson will want us to focus on the 'green' fragment of our language when we apply his recipe, but then it's plain that it isn't really our language that encodes the implicit knowledge that leads us to project 'green' instead of 'grue'. Indeed, it's hard to see what implicit knowledge Williamson could be appealing to, except maybe our knowledge that heads is very likely given lots of heads initially, that green emeralds are more likely than grue ones, and so on.

3.6.2 Continuous Indifference

What if Ω is uncountable? We can extend the Principle of Indifference to cover this case if we have a continuous parameterization of Ω , i.e. if we have a 1-1 map from Ω onto an interval of reals so that each element of the interval represents an element of Ω . Supposing we have such a parameterization, indifference takes the following form:

Principle of Indifference (Continuous Version) Suppose you have no evidence relevant to Ω , and Ω is parameterized on the interval from a to b (open

or closed). Then the probability of a subset is the area under the curve $f(x) = 1$ on that subset, divided by the total area on the interval, $b - a$. Formally,

$$p(S) = \frac{1}{b-a} \int_S 1 dx.$$

For sub-intervals, their probability will be their length divided by $(b - a)$. For single points the probability will be 0, and likewise for any countable set of points.²¹ In general, when the probabilities over a real parameter are encoded by the areas under a curve like $f(x) = 1$, the encoding function f is known as a *probability density*. The Principle of Indifference tells us to use the uniform density function, $f(x) = 1$, on the interval in question, since it treats each possible value the same.

Now we know how to apply the Principle of Indifference given a parameterization. But how do we pick a parameterization? Does it matter? A famous puzzle of Bertrand's ([1888] 2007) shows that the choice of parameterization does matter to the probabilities we end up with. Here's a nice illustration from van Fraassen (1989). Suppose a factory makes cubes, always between 1 and 3 inches on a side. What is the probability that the next cube to come off the line will be between 2 and 3 inches on a side? If we apply the Principle of Indifference to the parameter of side-length, the answer is

$$\frac{1}{2} \int_2^3 1 dx = 1/2.$$

But notice that the volume of the cube is another possible parameterization. If we use that parameterization instead, the Principle of Indifference gives us

$$\frac{1}{26} \int_8^{27} 1 dx = 19/26.$$

So we get different probabilities if we use different parameterizations.

To resolve this problem, Jaynes (1968; 1973), developing an idea due to Jeffreys ([1939] 2004), suggested a variation on the Principle of Indifference. Rather than using a uniform density on the grounds that it treats each possible value the same,

²¹ For some sets, known as Vitali sets, the probability is not defined at all, and so they can't be included in the σ -algebra implicitly under discussion. To be explicit, the σ -algebra for which p is defined here is the family of *Borel sets* in the interval from a to b , i.e. those sets that can be obtained from subintervals by countable union and intersection. It is a basic consequence of measure theory that there must be Vitali sets, assuming the axiom of choice. For an explanation why, see any standard textbook in measure theory, such as (Halmos 1974).

we should look for a density that treats each parameterization the same. That is, we want a density that is invariant across all parameterizations of the problem. In the cube factory example, the only density that is invariant between the length and volume parameterizations is $f(x) = 1/x$.²² Recall that

$$\int_a^b 1/x \, dx = \log(b) - \log(a).$$

So in terms of side-length, the probability of an interval $[a, b]$ will be

$$\frac{\log(b) - \log(a)}{\log(3) - \log(1)}.$$

Recall also that $\log(x^n) = n \log(x)$. So when we convert from side-length to volume, the probability of that same range of possibilities will be

$$\frac{\log(b^3) - \log(a^3)}{\log(3^3) - \log(1^3)}$$

or

$$\frac{3[\log(b) - \log(a)]}{3[\log(3) - \log(1)]}.$$

So changing the parameter from length to volume doesn't change the probabilities encoded by $f(x) = 1/x$. In general, because areas under $f(x) = 1/x$ are given by $\log(b) - \log(a)$, and $\log(x^n) = n \log(x)$, any transformation of x to x^n will leave the probabilities invariant, since the n 's will divide out. Something similar happens if we transform x into nx , except that the n 's will subtract out, since $\log(nx) = \log(x) + \log(n)$.

But for Jaynes's proposal to work in general, there must always be a unique density that leaves the probabilities invariant across all parameterizations. Unfortunately, it isn't so. A famous Bertrand-style problem due to von Mises ([1928] 1981) gives an example where there is no density that will yield invariant probabilities across all plausible parameterizations. Suppose we have a container with 10cc of a mixture of water and wine; at least 1cc is water and at least 1cc is wine. What is the probability that at least 5cc is wine? Consider two parameterizations: the ratios of water-to-wine and of wine-to-water. As before, $f(x) = 1/x$ is the only density that is invariant between these two parameterizations, and the probability

²² Actually, any constant multiple of $f(x) = 1/x$ will be invariant, but they will all yield the same probability distribution, since the constant will divide out.

it gives is

$$\begin{aligned} p([1,9]) &= \frac{\log(9) - \log(1)}{\log(9) - \log(1/9)} \\ &= 1/2. \end{aligned}$$

But what if we move to a third parameterization, wine-to-total-liquid? In this case we get a different answer:

$$\begin{aligned} p([1/2,9/10]) &= \frac{\log(9/10) - \log(1/2)}{\log(9/10) - \log(1/10)} \\ &\approx .268 \end{aligned}$$

Since $f(x) = 1/x$ is the only density invariant across the first two parameterizations, there cannot be another density that would be invariant across all three.

Jaynes was well aware of the water-wine puzzle, and even mentions it towards the end of his seminal 1973 essay on the invariantist approach, *The Well-Posed Problem*:

There remains the interesting, and still unanswered, question of how to define precisely the class of problems which can be solved by the method illustrated here. There are many problems in which we do not see how to apply it unambiguously; von Mises' water-and-wine problem is a good example [...]. On the usual viewpoint this problem is underdetermined; nothing tells us which quantity should be regarded as uniformly distributed. However, from the standpoint of the invariance [approach], it may be more useful to regard such problems as *overdetermined*; so many things are left unspecified that the invariance group is too large, and no solution can conform to it (Jaynes 1973: p. 10).

Jaynes's view seems to be that there is no answer to the problem *as posed*, but there might be a solution if the problem were posed in such a way as to specify a class of parameterizations for which there is an invariant distribution.

But what the correct probabilities are cannot depend on the way the problem is posed. Whether we're talking about the degrees of belief we ought to have, or what the logical or evidential probabilities are, how could they be relative to how the problem is phrased? Maybe what Jaynes has in mind is not the way the problem is phrased, but what assumptions we bring to the table in an actual case, and thus which parameterizations we are entitled to regard as on a par. But the whole point of the von Mises problem is that we don't bring any assumptions

to the table except the ones stated. So the three parameterizations discussed *are* all legitimate. It simply won't do to say that there is no fact of the matter what the probabilities are, or that the probabilities are assumption-relative. The whole point of the Principle of Indifference is to tell us what the correct probabilities are given no (relevant) assumptions, and either of these responses is simply a non-answer.

3.6.3 Summary of Principles of Indifference

In the discrete case, we saw that the probabilities assigned by the Principle of Indifference were badly anti-inductive unless we fiddled with the representation, first partitioning the space of possibilities coarsely according to frequencies. We also saw that the probabilities we got from Williamson's recipe depended on the language in which the space of possibilities was characterized, making the principle susceptible to grueification. And a similar representation-sensitivity troubled us in the continuous case. The variable by which the possibilities are parameterized affects what probabilities the Principle of Indifference assigns.

The bottom line looks to be that the Principle of Indifference is inappropriately sensitive to the way the space of possibilities is represented. To get unique answers from the principle, we need a recipe for choosing a partition or parameterization. But there is no obvious way of doing this, except to pick the one that respects the symmetry of our evidence, e.g. the finite partition such that our evidence favors no one element over any other. And such recipes threaten to rob the Principle of Indifference of its content (but see (White 2009) for a contemporary defense).

3.7 Principles of Reflection

The Principle of Indifference is ambitious but checkered. Let's turn our attention to a more modest principle, one with a much narrower focus: probabilities of future probabilities. Suppose you are at the track just before a race. Assuming you will be sure that your horse will lose the race as it comes around the final bend, what should you think about your horse's chances now, before the race starts? It would be odd to be confident that your horse will win, given that you will have the opposite attitude as the race draws to a close. So we might adopt the following general principle:

Reflection Principle Let p represent your current probabilities, and p_t your probabilities at some future time, t . Then for each A you should have

$$p(A|p_t(A)=x)=x.$$

An immediate and important consequence of Reflection is

Proposition 3.11 *If p satisfies the Reflection Principle then, for each A ,*

$$p(A) = \sum_i x_i p(p_t(A) = x_i).$$

Thus, if you obey Reflection, your current probability for A must be the expected value of the probabilities you may come to have at future time t . So to speak, your current probability must match the probability you expect to have in the future.

I've cast Reflection's motivation and formulation in terms very close to the degree of belief interpretation, since that's the context in which it was first advanced by van Fraassen (1984), and in which it is usually discussed. It is also the context in which it is most controversial. As many have pointed out, Reflection forces you to follow your future degrees of belief now, even if you are well aware that they will be unreasonable or misinformed. Even if you know that you will be drunk (Christensen 1991), or that you will have forgotten things you know now (Talbot 1991), Reflection still insists that you adopt those unreasonable/uninformed opinions. Reflection demands tomorrow's failures today.

On other interpretations of probability, Reflection may be less troubled. If the future probability in question is logical or evidential, then it may not be subject to the kind of cognitive imperfections that your future degrees of belief are subject to, and so it might be appropriate for today's probabilities to follow them. Still, future evidential or logical probabilities might depend on what evidence you will have in the future, and evidence can be lost. Williamson points out that his evidential probabilities can in fact lead to violations of Reflection (Williamson 2000: pp. 230-237).

Assuming the degree of belief interpretation though, why would anyone endorse such an implausible principle? Van Fraassen pointed out that the same Dutch book argument used to support Conditionalization can be used in support of Reflection (see section 4.1). Some take this as a sign that there is something amiss with the argument, but van Fraassen takes the opposite view, that Reflection is a requirement of rationality.

He later defended Reflection against the Talbot and Christensen objections in his (1995) and, perhaps more importantly, offered a more general version of Reflection with a new justification:

General Reflection Your current probability for A must be a mixture of your foreseeable probabilities at any future time t .

General Reflection is essentially a weakening of the idea captured in Proposition 3.11. Proposition 3.11 says that your current probability must be a specific mixture of your future probabilities, namely the expected value. General Reflection says that it need only be some mixture — any mixture. That is, $p(A)$ must equal $\sum_i x_i y_i$, where the x_i are your foreseeable future probabilities, and the y_i are any sequence of positive reals adding up to 1. In fact, this requirement is equivalent to $p(A)$ lying between the greatest and smallest foreseeable values for $p_t(A)$. So why does van Fraassen phrase General Reflection in terms of mixtures? Because he allows for indeterminate probabilities (section 1.2.3), in which case mixtures do not have this trivial formulation. For simplicity, our discussion will continue in terms of point-probabilities. Because point-probabilities are a special case of indeterminate probabilities, the critical points raised will apply *mutatis mutandis* to indeterminate probabilities.

In support of General Reflection, van Fraassen argues that it is entailed by Conditionalization (van Fraassen 1995: p. 17). He claims that any conditionalizer will satisfy General Reflection automatically, so we cannot violate General Reflection without violating Conditionalization. Weisberg (2007) argues that this is in fact not so. Rather, it is that someone who *thinks* she will Conditionalize (and also thinks that her future evidence forms a partition, and also knows her current conditional probabilities) will satisfy Reflection. But we are not rationally required to satisfy these conditions — indeed, it seems we should not satisfy them — so violating General Reflection does not entail a violation of anything required of us. Van Fraassen also argues that General Reflection, under certain assumptions, entails the original Reflection Principle, which he redubs *Special Reflection* (1995: pp. 18-19). And van Fraassen (1999) argues that General Reflection also entails Conditionalization when the foreseeable evidence forms a partition. But Weisberg (2007) argues that this latter claim makes a mistake similar to the one in the alleged entailment from Conditionalization to General Reflection.

More tempered views on the connection between current and future degrees of belief can be found in (Elga 2007) and (Briggs forthcoming). Elga defends the view that your probability for A should match the probability your future self or a third party would have if their judgment were absolutely trustworthy (a “guru”) and they had all the relevant information you have. (This is a rough characterization; see Elga’s paper for an exact statement of the view.) Elga goes on to ask how you should respond to the judgment of those whose judgment you do not trust completely, and addresses the special case where you disagree with an epistemic peer — someone whose judgment you regard as equally as good as your own, and who has the same information you do. In this case, Elga argues that you should average your opinion with theirs. Briggs proposes a generalization of the Reflection principle designed to handle cases where your future degrees of belief

are of varying trustworthiness and informedness. For other views on these and similar questions, see also (Plantinga 2000), (Kelly 2005), and (Christensen 2007).

3.8 Chance Principles

In section 2.4 we briefly discussed the physical chance interpretation of probability. There I said that, because physical chance is a (meta)physical notion, this interpretation did not yield its own brand of Bayesianism. But Bayesians who do think that there is such a (meta)physical thing as physical chance tend to think that it has important epistemic connections, since what the chances are should influence what the logical probabilities, correct degrees of belief, or evidential probabilities are. What is the exact nature of this connection?

The obvious, naive proposal is that your probability for A , given that the physical chance of A is x , should be x :

Miller's Principle Let p represent a reasonable probability function, and let X be the proposition that the chance of A is x . Then

$$p(A|X) = x$$

The principle is jokingly called 'Miller's Principle', since Miller (1966) argued that it is inconsistent. Jeffrey (1970b) argued that Miller's argument is invalid, and this is now the accepted view. The principle is still problematic though, since it forbids omniscience and foreknowledge. Assuming Conditionalization, someone who learns X must assign probability x to A , even if they have already consulted a perfectly reliable crystal ball that told them whether A will turn out true.

A more canonical view, designed to take account of such problems, comes from Lewis (1980). Lewis endorses both the degree of belief and physical chance interpretations of 'probability', and proposes the following rule connecting them:

Principal Principle Let p be any reasonable initial credence function.²³ Let t be any time and let x be any real number in the unit interval. Let X be the proposition that the chance at time t of A 's holding equals x . Let E be any proposition compatible with X that is admissible at time t . Then

$$p(A|XE) = x.$$

²³ What do we mean by "reasonable, initial credence function"? Lewis is thinking of p as a credence function which, "if you started out with it as your initial credence function, and if you always learned from experience by conditionalizing on your total evidence, then no matter what course of experience you might undergo your beliefs would be reasonable for one who had undergone that course of experience." (Lewis 1980: p. 268). Thus the Principal Principle should be thought of as a constraint on reasonable "ur-credences", (p. 35) rather than as a constraint on reasonable credences at any given time.

The idea is that your degree of belief in A , on the assumption that the chance at t of A is x , should be x . Unless, that is, you have “inadmissible” information, information that “circumvents” the chances, so to speak. The kind of thing Lewis has in mind as inadmissible is the consultation of a reliable crystal ball, time-traveller, oracle, or any other source of information from the future after t .

Of course, admissibility is everything here. As Lewis says, “if nothing is admissible, [the principle] is vacuous. If everything is admissible, it is inconsistent.” (1980: p. 272) Lewis did not give a full account of admissibility, but did identify two very broad kinds of admissible information: (i) information about the history of the world before t , and (ii) information about how chances depend on the history. These views famously ran into trouble when paired with his metaphysical view that what the true chances are depends only on how the history of the world turns out (Thau 1994; Lewis 1994; Hall 1994; Arntzenius & Hall 2003), though Vranaas (1998) argues that the problem is solvable. Of more general concern are worries raised by Meacham (2005) about the compatibility of Lewis’s view with the chance theories actually used in physics. Meacham argues that Lewis’s view is incompatible with the the Aharonov-Bergman-Lebowitz theory of quantum mechanics, as well as classical statistical mechanics. The trouble, in short, is that Lewis’s view makes chances time-relative, and assigns only chances of 0 and 1 if the true physics is deterministic. The Aharonov-Bergman-Lebowitz theory falsifies the first view, and classical statistical mechanics falsifies both. Meacham also argues that the concept of admissibility is actually dispensable if we formulate the principle correctly, which Meacham claims to do. This is good news if true, since no account of admissibility has received general acceptance.

3.9 Summary So Far

We have now surveyed the best-known proposals for rules constraining the range of permissible probability functions. Where do things stand today? Most Bayesians accept that something close to (Jeffrey) Conditionalization is correct, as well as something close to the Principal Principle, and maybe, in certain limited cases, a version of the Principal of Indifference. Because of worries arising from infinitary scenarios, Initial Regularity is less widely accepted, though Countable Additivity seems to do better. Reflection, on the other hand, is almost universally rejected.

A disclaimer: these are my own, informal estimations, and may reflect my own biases, and selective exposure to a certain kind of Bayesian.

Assuming that these estimations are not completely off though, the picture that emerges is of a fairly subjective trend. Diachronic rules like Conditionalization do a fair bit to fix what your new probabilities should be given what they were before, but there is still lots of room for disagreement about what probabilities

one may start with. The Principle of Indifference would eliminate much or all of that latitude, but it is only accepted by a few Bayesians and often in only a limited way. Countable additivity and Initial Regularity, as mentioned earlier (p. 34), do very little to constrain the range of permissible probability functions, and the Principal Principle, though more substantial, is still quite permissive.

Some Bayesians probably think that all that leeway is actually not allowed, and we just haven't yet found the correct formulation of the principles that disallow it. Others, however, think that such disagreement is just an inevitable fact of life, and does not demonstrate the incompleteness of the Bayesian theory so-far developed. The latter view is certainly counterintuitive at first blush, and whether it can be made more palatable depends on questions that go well beyond the scope of our discussion here. What one thinks about the nature of rationality, and about the aims of logic and epistemology, will heavily inform one's conceptions about objectivity. So we must leave off the subject of the subjective-objective continuum here, and move on to survey justificatory arguments for the foregoing rules.

4 Justifications

We will look at four kinds of justificatory argument for the various rules we have considered. The first kind, Dutch book arguments (DBA), are supposed to justify Probabilism, (Jeffrey) Conditionalization, Reflection, and Regularity. The second and third kinds, representation theorem arguments and axiomatic arguments, apply only to Probabilism. The fourth and final kind, accuracy/calibration/cognitive utility arguments, have been used to support both Probabilism and Conditionalization.

4.1 Dutch Book Arguments

Dutch book arguments (DBAs) have been given for Probabilism, Countable Additivity, Conditionalization, Jeffrey Conditionalization, Reflection, and Regularity. The crux of these arguments is that if the probabilities violate the rule in question, they will regard as fair a collection of deals that, as a matter of deductive logic, will lead to a sure loss. Since it is supposed to be unreasonable to regard as fair an obviously necessarily losing deal, it would be irrational to violate the rule in question. Let's look at the individual arguments.

4.1.1 The DBA for Probabilism

To get the argument going we need to establish a connection between probabilities and fairness. The standard connection assumed is

Probabilities Are Fair Odds If the probability of A is x , then a bet that pays $\$S$ if A is true, pays $\$0$ otherwise, and costs $\$xS$, is fair.

Usually the interpretation of probability assumed for DBAs is the degree of belief interpretation. Indeed, as we saw above (section 2.2.1), de Finetti took Probabilities Are Fair Odds to be a definitional truth about degrees of belief, since your willingness to pay at most $\$x$ for a $\$1$ bet on A is what makes it the case that your degree of belief in A is x . For others, Probabilities Are Fair Odds may be a law of psychology, or a conceptual truth about the commitments you implicitly make when adopting a probability. As a matter of convenience and tradition, our discussion of DBAs will occasionally use the lingo of the degree of belief interpretation, though this is not essential to the formal properties of the arguments.

To represent a bet that pays $\$S$ if A is true, pays nothing otherwise, and costs $\$x$, we'll use the notation $[A : S, x]$. Define the *possible worlds* of an outcome space (Ω, σ) as the functions on σ assigning

$$\omega_i(A) = \begin{cases} 1 & \text{if } \omega_i \in A \\ 0 & \text{if } \omega_i \notin A \end{cases}$$

for each ω_i belonging to some member of σ . Then we can say that your net in possible world ω_i for the bet $B_j = [A_j : S_j, x_j]$ is $S_j \omega_i(A_j) - x_j$. Your total net on the set of bets $\{B_j\}$ in world ω_i is

$$\sum_j S_j \omega_i(A_j) - \sum_j x_j.$$

Finally, we define a *finite Dutch book* for a real-valued function p on σ as a set of bets $\{B_j\}$ such that p regards each B_j as fair and yet

$$\forall_i : \sum_j S_j \omega_i(A_j) - \sum_j x_j < 0.$$

That is, a Dutch book is a set of seemingly fair bets that leads to a net loss no matter what the truth turns out to be.

The DBA for Probabilism then turns on the following theorem:

The Dutch Book Theorem Let p be a real-valued function on the outcome space (Ω, σ) . If p is not a probability function, then there is a finite Dutch book for p .

To illustrate, suppose p violates (P₃) because $A \cap B = \emptyset$ and $p(A) = p(B) = .4$, but $p(A \cup B) = .7$. Then these bets make a dutch book:

$$\begin{aligned} B_1 &= [A : 1, .4] \\ B_2 &= [B : 1, .4] \\ B_3 &= [A \cup B : -1, -.7] \end{aligned}$$

Notice the negative values in B_3 . These correspond to playing the role of the bookie on that bet instead of the bettor, since you now “pay” $-.70$ for the opportunity to “win” -1 . I.e. you accept $.70$ as payment and agree to pay out $\$1$ if $A \cup B$ comes true. Now, when the bets are made, you will pay out $\$.80$ and collect $\$.70$, for a net loss of $\$.10$. But you cannot win your money back. If either of A or B comes true, you will collect a dollar and pay out a dollar. If neither does, no more money changes hands. And because $A \cap B = \emptyset$, they cannot both come true.

It would be unfortunate to get suckered like that, but is obeying (P₁)–(P₃) any guarantee against it? In fact, it is:

The Converse Dutch Book Theorem If (Ω, σ, p) is a probability space then there is no finite Dutch book for p .

The reason you are immune is that each B_j must have 0 expected value to be fair and so, p being a probability function, the expected value of the whole set must be 0 as well. But that couldn’t be if the net-payoff were always negative — then the expected value would have to be negative. So obeying (P₁)–(P₃) is not only necessary, but is also sufficient, for avoiding a finite Dutch book.

4.1.2 The DBA for Countable Additivity

The DBA for Countable Additivity is much the same, except that now there is an infinite set of bets that leads to a sure loss if you violate the principle, but no such set if you obey. We can even construct the bets in such a way that someone with a finite bankroll could place them all. So there is no worry that you would have to have an infinite amount of money to be suckered, in which case you wouldn’t much care since you could arrange to have an infinite balance left after suffering an infinite loss (Adams 1964; Williamson 1999).

4.1.3 The DBA for Conditionalization

The DBA for Conditionalization is a bit different, in that there are two stages to the betting. To ensure that you incur a net loss, the bookie has to wait and see

what happens, possibly offering you a second round of bets depending on how things unfold. To model the situation, fix an outcome space (Ω, σ) and let P be the set of all probability distributions on it. Define an *update rule* as a function $r : P \times \sigma \rightarrow P$. Conditionalization is one such rule. Now define a *strategy* for the distribution p and the rule r on the partition $\{E_i\} \subseteq \sigma$ as a function from $\{E_i\}$ to ordered pairs of books (i.e. ordered pairs of sets of bets) where the first book is always the same. A strategy is *fair* for p and r if, whenever it assigns to E_i the books $\{B_j\}$ and $\{C_j\}$, $\{B_j\}$ is fair according to p and $\{C_j\}$ is fair according to $r(p, E_i)$. A *Dutch strategy* for r and p is a fair strategy for r and p that yields a net loss for each E_i .

It is important that strategies are fixed, in that they commit to their second-round bets for each eventuality beforehand. Thus the strategy is as much in the dark about which of its commitments it will end up having to follow through on as the update rule is. If the strategy is Dutch, then it bilks the update rule without taking advantage of any information that is not available to both parties. Thus it's especially interesting that we have

The Conditionalization Dutch Book Theorem Suppose p is a probability distribution and r an update rule such that for some E , $r(p, E) \neq p(\cdot|E)$. Then there is a Dutch strategy for p and r .

The Dutch strategy that vindicates the theorem was concocted by David Lewis and reported in (Teller 1973). Let p be your probability function, E some element of σ for which $r(p, E) \neq p(\cdot|E)$, and $q = r(p, E)$. Define $y = p(H|E) - q(H)$ for some H for which $p(H|E) > q(H)$. We then make $\{B_j\}$ the following trio of bets:

$$\begin{aligned} B_1 &= [HE : 1, p(HE)] \\ B_2 &= [\bar{E} : p(H|E), p(H|E)p(\bar{E})] \\ B_3 &= [E : yp(E), y] \end{aligned}$$

If E turns out to be false, you lose $yp(E)$, and the game is over, $\{C_j\} = \emptyset$. If it turns out true, we sell you a fourth bet:

$$C_1 = [H : -q(H), -1]$$

Whether or not H turns out true, you'll find you've still lost $yp(E)$. What Lewis's strategy does, in effect, is get you to make the same conditional bet on H given E twice, but at different odds.

Lewis's Dutch strategy is important, but the all-important question whether Conditionalization also immunizes you against Dutch strategies is settled by the following companion theorem (Skyrms 1987):

The Converse Conditionalization Dutch Book Theorem Let p be a probability distribution and suppose $r(p, E) = p(\cdot|E)$ for every E . Then there is no Dutch strategy for p and r .

The reason you are immunized by Conditionalization is that any fair strategy for a conditionalizer amounts to a set of bets that is fair relative to p . Whatever bets are in $\{C_j\}$ can be made into conditional bets on E that will be fair at the outset. But, since p is immune to Dutch books in virtue of being a probability function, no fair strategy can be Dutch. So, as with Probabilism and Countable Additivity, Conditionalization is both necessary and sufficient to immunize you against a certain kind of bilking.

4.1.4 The DBA for Jeffrey Conditionalization

When it comes to Jeffrey Conditionalization things get a bit trickier, since it is no longer clear just how to model an update rule or how a strategy for playing against an update rule should be conceived. Before, a rule just mapped old distributions and new evidence to new distributions. But now no member of σ can be identified as new evidence to which the rule can respond and in response to which a strategy can offer second-round bets. An analysis of rules and strategies for bilking them is not intractable, but it requires a level of detail that would be inappropriate here. Suffice to say that we can prove, for Jeffrey Conditionalization, theorems respectably analogous to those for Conditionalization. For the details, the reader is referred to Skyrms's (1987) excellent and careful discussion.

4.1.5 The DBA for Reflection

The Dutch book argument for Reflection is very closely related to the one for Conditionalization. We have

The Reflection Dutch Book Theorem If $p(H|p_t(H) = x) \neq x$ for some H and x , there is a Dutch strategy for p (no matter what r).

At first this result appears to contradict the Converse Conditionalization Dutch Book Theorem. How can you be Dutch booked no matter what r if, supposedly, you can use Conditionalization to immunize yourself against any Dutch strategy? To resolve the apparent conflict, notice that you cannot conditionalize in this scenario if E is $p_t(H) = x$ and E turns out true. If you did conditionalize on E when it turned out to be true, you would have $p_t(H) \neq x$, since $p(H|E) = p(H|p_t(H) = x) \neq x$, which contradicts E . So if you violate Reflection then you

cannot Conditionalize when E turns out true, making you vulnerable to Lewis's Dutch strategy. This is why the Reflection Dutch Book Theorem holds.

Is there a converse theorem for Reflection? There cannot be. We know from the Conditionalization Dutch Book Theorem that, even if you satisfy Reflection, you can still fall victim to a Dutch strategy by failing to Conditionalize. So obeying Reflection cannot protect you from the possibility of a Dutch strategy. The absence of a converse theorem may not undermine the DBA for Reflection, however. Obeying Reflection is a necessary part of any overall plan to avoid Dutch strategies by Conditionalizing. If you violate Reflection, you put yourself in a position where you may not be able to execute this strategy. So while obeying Reflection may not be sufficient to immunize yourself, it is a necessary part of a sufficient plan. But for those who, like van Fraassen (1995), want Reflection to *supplant* Conditionalization, the force of the DBA is undermined. To avoid the Dutch strategy that bilks non-Reflective distributions in general, you would have to adopt Conditionalization as a general policy.

4.1.6 The DBA for Regularity

The DBA for Initial and Continuing Regularity is of a slightly different character from the preceding ones. If you violate either kind of Regularity, there is no Dutch book strictly speaking. But there is a *weak* Dutch book, one that does not lead you to a sure loss but may lead you to a loss and *cannot* lead you to a gain. If $p(A) = 0$, then a bet on \bar{A} that pays $-\$1,000,000$ if A and costs $\$0$ is fair, since there is no chance that you will have to pay out the $\$1,000,000$. But unless A is impossible, it is possible that you will end up losing $\$1,000,000$, and yet you stand to gain nothing, so how can the deal be fair? (Shimony 1955)

4.1.7 Critical Discussion of DBAs

There are numerous points on which to criticize DBAs, and everyone has their favorite criticism. In the interest of brevity I will briefly mention some of the more popular criticisms with references to fuller discussions, and briefly discuss those criticisms that I think are especially instructive.

First, we may contest Probabilities Are Fair Odds for presupposing a false connection between money and value. If I regard losing $\$.50$ as more bad than gaining it would be good, the bet $[A : 1, .5]$ will not be fair for me even if $p(A) = 1/2$. Probabilities Are Fair Odds appears to be false because it assumes that the value of money must be linear, which it needn't be. But we may just be able to regard money as a stand-in for whatever goods you do value.

Another complaint is that the threat of being Dutch booked is avoidable by

simply not accepting the bets, or by betting in a way not in accordance with the probabilities. So probabilities don't have to obey the relevant axiom for you to avoid being suckered, it's just the odds at which you bet that have to obey those rules. The usual response to this point is that it's not the actual threat of being suckered that's the point. The point is that you *would* be suckered if you took the bets even though they're supposed to be fair. How can they be fair if they lead to a sure loss?

But perhaps the argument shows the wrong thing, since it makes an appeal to pragmatic rather than epistemic concerns. While it might be interesting that, as a matter of pragmatic rationality, probabilities must obey the axiom in question, this says nothing about the rules of probability qua rules of epistemic rationality. One response would be to insist that there is no difference between epistemic and pragmatic rationality; or, at least, epistemic rationality is subservient to pragmatic rationality in such a way that any epistemic rule which, if violated, leads to a violation pragmatic rationality, is a rule of epistemic rationality. Another response is that being in the position of being Dutch bookable actually demonstrates an *epistemic* defect, not a pragmatic one. Specifically, the probabilities that lead to the Dutch book regard the bets as fair even though they lead to a certain loss, which is *logically inconsistent*. A deal cannot be fair if one party has an advantage, which the bookie surely does since he necessarily walks away with a profit. So probabilities violating the axiom in question are actually logically inconsistent (Howson & Urbach 1993).

Finally, many of the DBAs make the implicit assumption that, because each bet in a collection is fair individually, the collection of them all together is fair too (Schick 1986). This is known as “the packaging assumption”. The packaging assumption is, arguably, false in some cases, so its use here may be illicit. Perhaps more importantly, its use here may beg the question. Consider the example above (p. 49) where we gave a Dutch book for a violation of (P₃). We implicitly assumed that, because the bets

$$\begin{aligned} B_1 &= [A : 1, .4] \\ B_2 &= [B : 1, .4] \\ B_3 &= [A \cup B : -1, -.7] \end{aligned}$$

were fair individually, they were fair as a package. But treating B_1 and B_2 as fair together because they are fair individually is very close to insisting that the bet $[A \cup B : 1, .8]$ is fair because $p(A) = .4$ and $p(B) = .4$, i.e. because (P₃) is true. In fact, it seems so close as to rob the argument of any dialectical weight.

In the case of Dutch strategies, where the package of bets is distributed between two times, it seems even less reasonable to insist that they are all fair together

because they are fair individually. After all, some of the bets are evaluated as fair according to one set of probabilities, and the others as fair relative to a different set of probabilities. There is no single epistemic state or point of view relative to which all the bets are fair, so why think that any epistemic defect has been exposed? (Christensen 1991) For a fuller discussion of this and other concerns about Dutch strategies, see (Earman 1992).

Moreover, paradoxical cases involving countable sets of deals seem to tell against the packaging assumption, at least for infinite sets of decisions. Consider the following example from Arntzenius, Elga, and Hawthorne (2004). God has divided an apple into a countable number of pieces, and tells Eve that she may eat any of the pieces she likes, though she will be expelled from the Garden of Eden if she eats an infinite number of them. Above all else Eve wants to stay in the garden, though she would also like to eat as much of the apple as she can without getting expelled. Now, for any given piece individually, the prospect of taking it is advantageous; no matter which of the other pieces she takes, taking this one too can't worsen her situation and surely improves it. Of course, Eve cannot thereby conclude that taking all the pieces is advantageous. See (Arntzenius et al. 2004) for further discussion of packaging-type reasoning in infinitary scenarios.

For a fuller survey of concerns about DBAs, see (Hájek 2008).

4.2 Representation Theorem Arguments

Earlier (p. 16) we discussed Ramsey's use of The Representation Theorem to analyze 'degree of belief'. The same theorem has also been put into service in support of Probabilism. Recall the theorem:

Representation Theorem Suppose a subject's preferences obey a set of constraints C (not specified here). Then there is exactly one probability function-utility function pair, (p, u) , that represents the agent's preferences, in the sense that she prefers A to B if and only if $EU(A) > EU(B)$.

To turn this into an argument for Probabilism, assume that the probabilities and utilities whose existence is guaranteed by the theorem describe the agent's actual beliefs and desires. Assume also that satisfying the constraints in C is rationally required. Then anyone who satisfies those constraints, as is required, has probabilistic degrees of belief. Thus we have an argument for Probabilism, at least under the degree of belief interpretation of probability.

There are essentially two main threads of criticism here. Traditionally, the target has been the claim that the constraints in C are rationally required. Allais (1979) famously devised a decision problem where it seems intuitive to violate Savage's Independence Axiom:

Independence Axiom Suppose that acts A_1 and A_2 yield the same outcome, O , in the event that E is false. Then A_1 is preferable to A_2 if and only if an act that has the same outcomes as A_1 , except that some other outcome O' happens if E is false, is preferable to an act that has the same outcomes as A_2 except that it too leads to O' if E is false.

Other, more technical axioms needed to complete the theorem are also a common point of attack, such as Jeffrey's (1965) assumption that the objects of one's preferences form an atomless boolean algebra.

In response to these complaints, proponents of the RTA tend to insist that axioms like Independence are correct if unintuitive in some cases, and that the more technical axioms are in-principle dispensable. Joyce (1999), for example, suggests that we should only expect a rational agent's preferences to be *embedable* in the fine-grained sort of preference structure outlined by the axioms. Then the theorem shows that one's state of opinion must agree with some probabilistic state of opinion, though it may not be as thorough and fine-grained as an actual probability function.

Perhaps a more damning problem is the one raised in our earlier discussion of The Representation Theorem (p. 16), namely that we cannot say that an agent's actual degrees of belief are the ones the theorem provides. Actual people do not form their preferences in accordance with expected utility maximization (Kahneman & Tversky 1979), so why think that the p assigned by the theorem, which is only unique in the sense that it alone is part of an expected utility representation of those preferences, represent anyone's actual probabilities? Christensen (2001) has responded that, while we cannot assume that the p guaranteed by the theorem is your actual degree of belief function, we can say that p *should* be your actual degree of belief function. If it weren't, your degrees of belief would violate an intuitively plausible normative constraint connecting degrees of belief and preferences:

Informed Preference An ideally rational agent prefers the option of getting a desirable prize if B obtains to the option of getting the same prize if A obtains, just in case B is more probable for that agent than A .

Meacham and Weisberg (unpublished) show, however, that this is not true. One can satisfy the constraints in C , together with Informed Preference, and not have the degrees of belief assigned by the Representation Theorem. Christensen's reason for thinking that this is not possible turns on an ambiguity in standard statements of the theorem, but counterexamples to his claim are easy to generate. If p represents the degrees of belief assigned by the theorem, then the function p^2 will also satisfy Informed Preference, since it preserves the ordinal properties of p .

A third criticism often leveled at the Representation Theorem approach is that this argument, like DBAs, conflates pragmatic and epistemic normativity. The constraints on preferences given in C appear to be pragmatic norms, so the argument's conclusion, if it worked, would seem to be that Probabilism is a requirement of pragmatic rationality, whereas epistemic rationality is the topic of interest. As with DBAs, one might respond that there is no difference, or that epistemic rationality is subservient to pragmatic rationality.

4.3 Axiomatic Arguments

Another approach to justifying Probabilism is to try to derive it from unassailable axioms. The classic result here is due to Cox (1946), who started with a few elementary assumptions that any measure of plausibility should obey, and then proved that any such measure would be isomorphic to a probability function. Here is the full statement of the theorem:

Cox's Theorem Let $b(\cdot|\cdot)$ be a two-place, real-valued function on $\sigma \times (\sigma - \{\emptyset\})$ satisfying the following two conditions for any A, B , and C :

- i. $b(A|B)$ is a function f of $b(\bar{A}|B)$, and
- ii. $b(AB|C)$ is a function g of $b(A|BC)$ and $b(B|C)$.

If f'' and g'' exist and f'' is continuous, then there is a continuous 1-1 function $i: \mathbb{R} \rightarrow \mathbb{R}$ such that $p = i \circ b$ is a probability function in the sense that $p(\cdot|\Omega)$ satisfies (P1)–(P3), and $p(A|B)p(B|\Omega) = p(AB|\Omega)$ for any A, B .

The idea is that any reasonable numerical representation of conditional plausibility will be a notational variant of a probabilistic representation.

It turns out that there are technical difficulties with Cox's theorem. Halpern (1999), for example, argues that there are counterexamples to the theorem, and that the hole can be plugged only at the price of appealing to unmotivated assumptions. There are other theorems to similar effects, e.g. those of Aczél (1966) and Fine (1973). A common concern with such results, and with the axiomatic approach in general, is that the assumptions needed to nail down the theorems tend to be less plausible or intuitive than the rules of probability are on their own terms. (P1) simply establishes a scale, and (P2) says that logical truths are maximally probable, which seems almost self-evident. The only really substantial and controversial axiom is thus additivity, (P3). And it's hard to imagine that any formal argument deriving it could be based on intuitively obvious assumptions about the way probability should work that are more obvious than additivity itself. It is also hard to see why the probability function that b is shown to be isomorphic to in

these theorems should be regarded as a notational variant. Why, in Cox’s theorem for example, should $p = i \circ b$ be regarded as just a re-scaled statement of b ?

4.4 Calibration and Accuracy Arguments

Another approach to justifying Probabilism is to try to show that probability functions are specially situated, veridically speaking — that they do better than any alternative at getting at the truth. This approach has been taken by Rosenkrantz (1981), van Fraassen (1983), Shimony (1988), and Joyce (1998), all of whom aim to show that probabilistic beliefs are in some sense more accurate than non-probabilistic beliefs. Suppose, for example, that we adopt as a measure of inaccuracy the quadratic loss function,

$$Q(b, \omega) = \sum_{A \in \sigma} (\omega(A) - b(A))^2,$$

where b is some real-valued function on σ and ω is, as earlier, a singleton of Ω , with $\omega(A) = 1$ if $\omega \subseteq A$ and 0 otherwise. De Finetti then showed the following:

Proposition 4.1 *If b is a non-probabilistic function on σ then there is a probability function, p , for which $Q(p, \omega) < Q(b, \omega)$ for every $\omega \in \Omega$.*

Thus non-probabilistic beliefs can always be brought closer to the truth by going probabilistic *no matter what the truth may turn out to be*.

An argument for Probabilism based on De Finetti’s result must, however, motivate the use of Q as a measure of inaccuracy. The result can be generalized to any quadratic loss function, where the terms in the sum are given varying weights, but we still want some argument that there are no acceptable measures of inaccuracy that deviate from this general form. Alternatively, we could try to generalize the result even further, which is what Joyce (1998) does. He offers six axioms constraining any reasonable measure of inaccuracy, and then shows that a result analogous to De Finetti’s holds for any such measure. Joyce’s axioms not only entail that non-probabilistic functions are dominated by probabilistic functions in this way, but also that probabilistic functions are not dominated by non-probabilistic functions.²⁴

Maher (2002) criticizes Joyce’s axioms on the grounds that (i) his arguments motivating them are unsound, and (ii) they are too strong, ruling out reasonable measures of inaccuracy. In particular, the simple measure

$$\rho(b, \omega) = \sum_{A \in \sigma} |b(A) - \omega(A)|$$

²⁴ This is not explicitly shown in Joyce’s (1998), but Joyce has verified this additional result in personal correspondence.

is incompatible with two of Joyce’s axioms, and without them the theorem does not follow. Joyce (2009) shows that analogous dominance results can be proved from alternative constraints on inaccuracy measures, and defends these alternatives as reasonable constraints.

4.5 Expected Cognitive Utility and Conditionalization

Greaves and Wallace (2006) have given an argument for Conditionalization that is similar in spirit to the preceding one, joining those ideas with Levi’s idea of maximizing expected epistemic utility.²⁵ The rule of Conditionalization, they argue, maximizes expected epistemic utility. To make such an argument, of course, we need to say something about the structure of epistemic utility; which actions are likely to lead to epistemically good outcomes depends in large part on what counts as an epistemically good outcome. A particularly interesting class of utility functions are characterized by the following definition:

Strong Stability A utility function is *everywhere strongly stable* if and only if the expected utility of holding any probability function, calculated relative to that same probability function, is higher than the expected utility function of holding any other probability function. I.e., u is such that for any p, q ,

$$\sum_i p(O_i | \text{Hold } p) u(O_i \wedge \text{Hold } p) > \sum_i p(O_i | \text{Hold } q) u(O_i \wedge \text{Hold } q).$$

The argument then proceeds from the following theorem

Greaves-Wallace Theorem If u is everywhere strongly stable, then updating by Conditionalization uniquely maximizes expected utility.

If we replace the $>$ in the definition of Strong Stability by \geq , we get the set of *everywhere stable* utility functions, in which case the corresponding theorem is that Conditionalization maximizes expected utility, though not necessarily uniquely — some other update rule might be equally promising.

Greaves and Wallace’s argument turns on the assumption that epistemic utility is, or should be, everywhere (strongly) stable. On this subject they remark, “we find this rationality constraint plausible, but we offer no argument for it here.” (Greaves & Wallace 2006: p. 626) Admittedly, the constraint has strong pull. An unstable utility function would be one that was empirically biased, in a sense, valuing one state of opinion over another even assuming that the latter one was correct. But I do not know how to turn this fuzzy thought into a proper argument.

²⁵ See section 5 for a definition of ‘expected utility’ and the role of expected utility in decision making; see section 7 for Levi’s use of expected epistemic utility

4.6 Summary

Our discussion has touched on at least one argument for each of Probabilism, Conditionalization, Jeffrey Conditionalization, Reflection, and Regularity, with Probabilism and Conditionalization receiving most of the attention. We did not discuss any arguments for Infomin, the Principle of Indifference, and the Principal Principle. This treatment does represent the literature roughly, but not exactly. We mentioned some arguments for Infomin back in section 3.5; and the Principle of Indifference and the Principal Principle can be motivated intuitively. The Principle of Indifference says that, when your evidence is symmetric over a partition, your probabilities should be too. And the Principal Principle says that your credences should match the physical chances, absent any reason to think you “know better”. Both are intuitively compelling suggestions that yield plausible results in many cases (indeed, Lewis (1980) motivates the Principal Principle by appealing to a list of its applications).

5 Decision Theory

So far we have been looking just at the logical and epistemological faces of Bayesianism. But one of Bayesianism’s great virtues is that it lends itself quite readily to theorizing about pragmatic norms — norms of decision making — as well. Assuming that we have a numerical notion of value to work with, Bayesianism gives us an algorithm for deciding what to do when faced with a choice. Let u be a *utility* function, i.e. a function that assigns to each possible outcome of an act a number reflecting how good or bad that outcome is. Then act A_1 is better than act A_2 just in case

$$\sum_i p_i u_i > \sum_j p_j u_j,$$

where the p_i are the probabilities of the possible results of A_1 and the u_i their utilities, and similarly for the p_j and u_j with respect to the results of act A_2 . The quantity

$$\sum_i p_i u_i$$

is called the *expected utility (EU)* of act A , and it serves to weigh the value of each possible outcome of the act against the likelihood that it will occur, summing the results to capture the overall expected “goodness” promised by that act. The decision rule to choose the act with the highest expected utility is called *expected utility maximization*, or *EU-max*.

The rule to maximize expected utility has near-universal acceptance amongst Bayesians. (Indeed, departures from the *EU-max* rule tend to be regarded as

non-Bayesian almost as a matter of definition.) Disagreements usually arise over the details. There are questions about how to understand and represent acts and outcomes, and there is disagreement about how we should interpret the p_i , the probabilities of various eventualities of the act in question. The best way to appreciate the subtleties here is just to follow the historical dialectic that brought them to light. We start with Savage's classic formulation of Bayesian decision theory, move on to the motivations and amendments of Jeffrey's theory, and finish with a look at causal decision theories.

5.1 Savage's Decision Theory

Savage (1954) represented a decision problem in terms of three components: states, acts, and outcomes. To illustrate, consider a simple gambling problem, such as deciding whether to bet on red or black in a game of roulette. There are two possible states of the world: the wheel will land on red, or the wheel will land on black. The outcome of each act depends on the true state: if you bet red, landing on red leads to a win and landing on black leads to a loss. Vice versa for betting on black. Thus we may think of acts as functions from states to outcomes. Since the outcomes are what we care about, utilities attach to them. But our ignorance is about which state will actually obtain, so states are the bearers of probability. Thus a decision problem is represented, in Savage's framework, as a set of states, acts, and outcomes, with probabilities distributed over the states and utilities distributed over outcomes.

How do we calculate expected utility in this framework? Given act A , we want to multiply the probability of each state by the utility of the outcome it will result in given act A , $p(S_i)u(A(S_i))$, where $A(S_i)$ is the outcome that act A maps state S_i to. Thus expected utility in Savage's framework is

$$EU(A) = \sum_i p(S_i)u(A(S_i)).$$

In the roulette example, supposing that utility equals money and that our two options are to bet \$1 on black or bet \$1 on red, the expected utility of betting on red is

$$p(R)u(+\$1) + p(B)u(-\$1) = (1/2) \times 1 + (1/2) \times -1 = 0,$$

as expected, and similarly for betting on black. So neither act is superior, as seems right. (In the case of a tie for best, EU -max says that it does not matter which act we choose.)

Savage's framework is simple, elegant, and in many cases perfectly adequate. But there are decision problems for which it seems to be inappropriate, as Jeffrey

(1965) pointed out with examples like this one.²⁶ Suppose that I park my car in a sketchy neighborhood and am approached by a suspicious character who offers to “protect” it for me while I am about my business; his price is \$10. Now suppose I reason as follows. There are two possible states of the world, one in which this suspicious character will smash my windshield and one in which he will not. If I pay him the \$10 then the possible outcomes here are one in which my windshield is not smashed and I am out \$10, and one in which my windshield is smashed and I am still out \$10. If I don’t pay him, the possible outcomes are smash vs. not-smashed, though I am not out \$10 in either case. Whatever the probabilities of the states are here, the expected utility of not paying him will clearly be higher on Savage’s account, since the outcomes determined by that act are better in each possible state. But this is surely mistaken. Clearly I will end up with a smashed windshield if I don’t pay, and may well not if I do. Given the low price of \$10 to avoid the extreme badness of having my windshield smashed, I should pay.

5.2 Evidential Decision Theory

What Savage’s framework overlooks, and this problem exploits, is that my action is relevant to which state of the world will obtain, smashed or not-smashed. To remedy this problem, Jeffrey proposed that we allow the probabilities of the possible states to vary depending on which act we are considering. In particular, Jeffrey thought that the relevant probabilities were the probabilities *conditional* on the assumption that we take that act. Once we notice that there is an important probabilistic relationship between actions and states, we also notice that the connection between action-state pairs and outcomes can be probabilistic as well. Even once we fix my act (bet on red) and the state of the world (land on black), there is some chance that I will win \$1 anyway, through some oversight on the part of the casino, for example. Thus it looks as though acts, states, and outcomes should all be propositions under the domain of the probability function.

In Savage’s framework, states are basic elements in the domain of the probability function, outcomes are basic elements in the domain of the utility function, and acts are deterministic functions connecting these two otherwise distinct domains. In contrast to this, Jeffrey’s system treats acts, states, and outcomes as all the same kind of thing, propositions, which can be arguments for both the probability and utility functions. We can talk about the probability that we will take a given act, the probability that a certain state obtains, and the probability that a certain outcome will result. Similarly for utilities. In fact, Jeffrey’s system doesn’t officially distinguish between acts, states, and outcomes at all. On Jeffrey’s

²⁶ This particular example is from Joyce (1999).

system, a decision problem is just composed of a propositional language, with a probability function and a utility function over it.

How do we figure the expected utility of an act in Jeffrey's system? We consider all the ways things could turn out, all the singletons of Ω , the ω_i ,²⁷ and multiply the probability of each one conditional on doing the act by the utility of that act and that outcome obtaining, then sum up. That is, we calculate the *evidential expected utility* of A ,

$$\begin{aligned} EEU(A) &= \sum_i p(\omega_i|A)u(\omega_i \wedge A) \\ &= \sum_i p(\omega_i|A)u(\omega_i) \end{aligned}$$

Applying this new system to the car-park example, we find that a plausible assignment of utilities and probabilities leads to a higher *EEU* for paying the \$10, since this significantly increases the probability of the vastly superior outcome where my windshield is not smashed. Unlike in Savage's system, the $p(\omega_i|A)$ vary depending on what A is.

Savage was aware of cases like the car-park example, where the probability of the outcomes depends on the act being considered. He thought they should be handled by individuating the possible states in such a way that the states would be independent of the acts. So, for example, we might distinguish:

- S_1 : The windshield will be broken whatever I do.
- S_2 : The windshield will be broken if you don't pay, but not if you do.
- S_3 : The windshield won't be broken if you don't pay, but will be if you do.
- S_4 : The windshield will not be broken regardless.

Then the expected utility calculation comes out as desired on Savage's formula. So if we partition the space of possible states appropriately, we can find a formulation of the problem that gets the right answer. The problem, of course, is that we then need a general theory of how to partition the states. The obvious answer is to use a partition where the states are probabilistically independent of the acts, but acts do not have probabilities on Savage's framework; there is no way to express the thought that $p(AS) = p(A)p(S)$, since the first two terms are undefined. This is why Jeffrey's theory is preferable.

²⁷ Actually, for technical reasons, Jeffrey's system works with an atomless algebra, so there is no Ω , just an ever more finely divisible set of propositions. To keep things simple and more in harmony with our formalism, we'll ignore this feature of Jeffrey's system. We also assume that Ω is countable, so that we can do everything in terms of sums.

Not only does evidential decision theory allow us to express probabilistic relationships between acts and outcomes, but it also gives us the same answer no matter how we set up the problem. The expected utility of a proposition is equal to the expected value of the expected utility over any partition:

Proposition 5.1 *If $\{O_j\}$ is a partition of Ω , then*

$$\begin{aligned} EEU(A) &= \sum_i p(\omega_i|A)u(\omega_i) \\ &= \sum_j p(O_j|A)EEU(O_j \wedge A). \end{aligned}$$

Thus it doesn't matter how we partition a problem, contra Savage, provided we plug the right values in for the utilities, namely the evidential expected utilities.

This partition-invariance is important because it shows that Jeffrey's theory is just as applicable to "small-world" problems as it is to the "grand-world" problem of managing your decisions over the most fine-grained possible representation of the problem. Whether we look at a decision problem as we normally do, only considering very coarse outcomes like "windshield gets smashed" and "windshield doesn't get smashed", or we look at the possible outcomes in maximal detail, everything coordinates. The expected utility of an act will come out the same either way. So, while Savage's framework depended on finding the "right" partition, Jeffrey's evidential decision theory does not.

5.3 Causal Decision Theory

Still, there are problem cases for evidential decision theory. Suppose, for example, that new research shows us that smoking does not actually cause lung cancer. As it turns out, the correlation between smoking and lung cancer is due to a common cause — there is a gene that disposes one to smoke, and also disposes one to develop lung cancer. In this case, would you have reason to stop smoking (assuming you do smoke)? On the one hand, the probability of you not having the gene, and thus not developing lung cancer, conditional on you quitting, is quite high. So the *EEU* of quitting is high. On the other hand, either you have the gene or you don't, and there's nothing you can do about it. So you might as well go ahead and enjoy your favorite habit.²⁸

Most find the latter argument the persuasive one: while it might be good news if you can quit, since it decreases your chance of getting lung cancer, quitting is not a way of preventing lung cancer, and thus has no *practical* value, despite its

²⁸ This is a version of the famous Newcomb Paradox.

evidential value. If you find this a plausible diagnosis of the problem, then you should be on the market for a new kind of decision theory, one that weights acts based on their *efficacy*, i.e. their ability to bring about good outcomes, rather than based on their value qua good news, which is what evidential decision theory was evaluating.

How to formulate such a theory? The classic formulation is due to Skyrms (1980). Skyrms’s idea is to look for a special partition of Ω , one that separates Ω into what David Lewis (1981) calls *dependency hypotheses*, theories about how the various outcomes you care about depend causally on what you do. Each element of such a partition would fully specify how your potential actions sway the possible outcomes. Call such a partition a *K-partition*. We can then calculate the *causal expected utility* of an act as

$$CEU(A) = \sum_i p(K_i) \sum_j p(\omega_j | A \wedge K_i) u(\omega_j).$$

The idea is that we calculate the expected evidential value of the act on each possible assumption about what the causal connections might be,

$$\sum_j p(\omega_j | A \wedge K_i) u(\omega_j),$$

and then sum up, weighting by the probability of each assumption about what the causal connections may be, $p(K_i)$. Causal decision theory agrees that evidential expected utility is what matters when the causal structure of the world is taken as a given. The thought is that, when the causal structure is given, the tendency of an action to bring about an outcome coincides with the conditional probability of the outcome given the action. But, since the causal relationships between your actions and possible outcomes are not known, we have to calculate the evidential expected utility for each possible dependency hypothesis, and then weight each one by its probability.

There are other ways of trying to capture the same basic idea, that it is the causal efficacy of your actions, not their evidential value, that matters to decision making. The original proposal, due to Gibbard and Harper (1978), used subjunctive conditionals to capture the causal impact of your actions. Subjunctive conditionals have the form, “were A the case then B would be the case”, and are standardly abbreviated $A \square \rightarrow B$. Subjunctive conditionals are very closely related to causation; generally speaking, if B wouldn’t have happened had A not happened, then A is a cause of B . Motivated by this thought, and by subjunctive analyses of causation like Lewis’s (1973a), Gibbard and Harper proposed that we calculate the

expected utility of act A by

$$\sum_i p(A \square \rightarrow \omega_i) u(\omega_i).$$

The idea is that we weight the utilities of the ways things could turn out by the probabilities that they would turn out that way if we did A . Thus we use A 's disposition to bring about good outcomes as our guide. Lewis (1981) argues that the Gibbard-Harper approach and others can be seen as alternate formulations of the same idea, being equivalent to CEU as we defined it in terms of K -partitions. For fuller discussion of various formulations of causal decision theory and their relationships, see (Lewis 1981) and (Joyce 1999: pp. 161-176).

Applying the K -partition approach gets the intuitively correct result in cases like the smoking example: that you need not quit. But the cost is that we have to go back to the Savage approach, where applying the theory requires fixing on the right kind of partition, a K -partition. Is this essential to causal decision theory?

Joyce argues that it is not. He suggests that we formulate causal expected utility as

$$CEU(A) = \sum_i p^A(\omega_i) u(\omega_i),$$

where p^A is the probability function we get from p by a process called *imaging*. Imaging p on A is supposed to capture the idea that we suppose A causally, rather than evidentially. How does imaging work? We start with a notion of “closeness” or “similarity” between the ω_i , so that we can pick out the closest singleton to a given ω_i within a certain set, A . Let ω_i^A be the singleton closest to ω_i that is in A . Then the image of p on A is given by

$$p^A(O) = \sum_i p(\omega_i) p(O|\omega_i^A),$$

which amounts to moving the probability from worlds outside of A to their nearest respective worlds inside of A . This is supposed to capture the kind of supposing we do when we ask how probable O would be if A were true. This definition of p^A implicitly assumes that there is always a “closest” world within A , an implausible assumption, as Lewis (1973b) argues. But Joyce adopts a more general definition, due to Gärdenfors (1988), that does not require this assumption. See (Joyce 1999: pp. 198–9) for the details.

The notion of imaging in hand, Joyce then defines a notion of conditional, causal expected value,

$$V(X|A) = \sum_i p^A(\omega_i|X) u(\omega_i),$$

which expresses A 's causal tendency to bring about good outcomes within the range of possibilities specified by X . We then have partition-invariance in the form,

Proposition 5.2 *If $\{X_j\}$ is a partition of X , then*

$$V(X \setminus A) = \sum_j p^A(X_j | X) V(X_j \setminus A).$$

As a result, when we calculate the causal expected utility of an act, we can do it using any partition:

Proposition 5.3 *If $\{X_j\}$ is a partition of A , then*

$$\begin{aligned} CEU(A) &= \sum_i p^A(\omega_i) u(\omega_i) \\ &= \sum_j p^A(X_j | A) V(X_j \setminus A) \\ &= V(\Omega \setminus A). \end{aligned}$$

This lends Joyce's formulation a virtue similar to Jeffrey's theory. The imaging-based formulation also has the nice effect of isolating the difference between causal and evidential expected utility, placing it squarely in the epistemic half of the formula. The difference between evidential and causal decision theory can then be seen as a difference in the way in which we suppose A : causally versus evidentially, by imaging as opposed to by conditioning. That this is the crucial difference between EEU and CEU is not apparent on a Skyrms-style, K -partition account.

And yet, causal decision theory may itself be troubled. Here is a problem case recently raised by Egan (2007).²⁹ Johnny has devised a button which, if pressed, will kill all psychopaths. Johnny believes himself not to be a psychopath and places a high value on eliminating psychopaths from the world. And yet, Johnny believes that only a psychopath would push the button, and he values his own preservation much more than he values eliminating all psychopaths from existence. Intuitively, it seems to many, Johnny should not push the button, since that would tell him that his action is very likely to cause his own death (note the mixture of evidential and causal considerations here).

But causal decision theory seems to say that he should press the button. Consider the question first in terms of the K -partition formulation. There are two relevant dependency hypotheses here:

²⁹ Egan offers two illustrations of the problematic phenomenon; he attributes this one to David Braddon-Mitchell.

K_1 : If Johnny presses the button he and all psychos will die, but if he doesn't press it nobody will.

K_2 : If Johnny presses the button all psychos will die yet he will survive. If he doesn't, nobody dies.

Given K_1 , pressing is a terrible idea, whereas it is a good idea given K_2 . As for not pressing, well it's pretty much neutral, maintaining the status quo either way. Since Johnny thinks it very unlikely that he is a psycho, however, K_2 seems much more likely, and so pressing will come out on top, since the expected good on K_2 will outweigh the very improbable badness that would result on K_1 . The trouble seems to stem from the fact that Johnny's action is *evidentially* relevant to which K_i obtains, but this factor is not accounted for by the causal account, since we use just $p(K_i)$ to calculate *CEU*.

How do things work out on the Joycean formulation? Consider how likely Johnny would be to die if he were to push the button. Of the button-pushing ω 's, the ones where Johnny is a psycho and hence dies bear more of the probability initially. But those improbable ones in which Johnny is not a psycho, and hence survives, will acquire a good deal of probability once we image; all the ω 's where Johnny is not a psycho and doesn't push the button will have their probability transferred to the worlds where Johnny is still not a psycho but pushes the button, since they are more similar, on any plausible reading of "similar". Viewing the probabilities over the ω 's schematically may help here. Initially they might look something like this:

	Press	\neg Press
Psycho	.1	.09
\neg Psycho	.01	.8

After imaging, they would look something like this:

	Press	\neg Press
Psycho	.19	0
\neg Psycho	.81	0

So it's pretty likely that he would survive if he pressed the button and, assuming he wants to kill all psychopaths badly enough, he should press it. What seems to be missing is, again, that the act of pressing is not just causally relevant to what happens, but also evidentially relevant to what causal connections obtain.

6 Confirmation Theory

Not only in science, but in everyday reasoning too, we say that some bit of evidence confirms this or that theory, that this is evidence for or against that hypothesis, or that one theory is better confirmed than another. Explicating such talk about the confirmation and evidential support of theories is a classic problem in the philosophy of science (see the entry on confirmation in this volume). Bayesianism seems especially well equipped to solve it, and indeed Bayesians have had a great deal to say about the concept of confirmation.

Traditionally, three different questions about confirmation are distinguished. First, when does a given bit of evidence confirm or disconfirm a given hypothesis? Second, by how much does the evidence confirm the hypothesis? And third, how well confirmed is a given hypothesis, absolutely speaking? Corresponding to these three questions, we have three separate concepts of confirmation:

Qualitative Evidential Confirmation A qualitative relation between evidence and hypothesis, specifying whether the evidence tells for or against the hypothesis (or is neutral).

Quantitative Evidential Confirmation A quantitative relation between evidence and hypothesis specifying how much the evidence tells for or against the hypothesis (or is neutral).

Absolute Confirmation A quantitative property of a hypothesis, specifying its overall evidential standing to date.

Bayesians tend to equate Absolute Confirmation with probability, saying that a hypothesis is confirmed to degree x just in case it has probability x . The plausibility of this proposal immediately suggests that we try to give probabilistic analyses of the other two concepts of confirmation. This is where things get tricky, and where Bayesians diverge from one another.

6.1 The Naive Account and Old Evidence

The simplest and most obvious Bayesian account of both Qualitative and Quantitative Evidential Confirmation is this:

Naive Bayesian Confirmation E confirms H when $p(H|E) > p(H)$, and the degree to which E confirms H is $p(H|E) - p(H)$.

The idea is intuitively appealing: evidence supports a hypothesis when it increases its probability, and the extent of the support is just the degree of the increase. How could something so obvious and simple possibly go wrong?

Glymour (1980) posed a notorious problem for the naive account. One of the most crucial bits of evidence supporting Einstein’s General Theory of Relativity comes from an anomalous advance in Mercury’s perihelion. The anomaly was unexplained in classical physics but Einstein showed that it was a consequence of his theory, thus lending his theory crucial support. Let E be the fact of the anomaly and H Einstein’s theory. Because E is evidence that we’ve had for some time, $p(E) = 1$. But then, Glymour points out, it follows that $p(H|E) = p(H)$, so E does not support H on the naive account. Yet it surely does intuitively.

The naive account has some accordingly flat-footed responses to this problem of old evidence. The first is what we might call the historical response, which says that, while $p(E)$ may equal 1 now, it was not always 1. It may well be that, before E became old evidence, $p(H|E)$ was significantly higher than $p(H)$, and that is why E confirms H . But this response is troubled on the degree of belief interpretation. What if we knew about the anomaly before ever encountering Einstein’s theory? Then there never was a time when $p(H|E) > p(H)$. This seems to be the case for most of us living now, and was surely the case for Einstein. But even if we did formulate H before learning E , it’s hard to accept that this historical accident is what makes it correct when we say *now* that E confirms H . A different approach is to abandon historical probabilities in favor of counterfactual probabilities. Maybe the reason we say that E confirms H is that, if we didn’t already know E , then it would be true that $p(H|E) > p(H)$. But this strategy runs into a similar problem, since there is no guarantee that the probabilities in the counterfactual scenario will be as desired. Maybe if we didn’t already know E our degrees of belief would be different in such a way that $p(H|E) > p(H)$ actually wouldn’t hold. See (Maher 1996: p. 156) for a nice example of this problem.

6.2 Alternate Measures of Confirmation

A more promising line of response to Glymour’s problem begins with an appeal to Continuing Regularity (section 3.2), the thesis that we should never assign degree of belief 1 to anything but a tautology. Then it will not be true that $p(E) = 1$ for the anomaly evidence, and it may be that $p(H|E) > p(H)$. The problem still remains in its quantitative form, however, since $p(H|E)$ and $p(H)$ are still approximately equal, so that the *degree* to which E confirms H is very little on the naive account, which is almost as bad as it not confirming H at all.

This brings us to the second step in the strategy, where we tweak the quantitative aspect of the naive account. The quantity $p(H|E) - p(H)$ is only one possible way of measuring E ’s support for H . Another candidate is the quantity $\log[p(E|H)/p(E|\bar{H})]$. The idea behind this measure is that we are using E to test between H and its alternatives. If E is more likely on the supposition that H than

on the supposition that \bar{H} , then this quantity is positive. And if H makes E much more likely than \bar{H} does, then the quantity is large. This latter measure is called the *log-likelihood ratio* measure, and our original measure is the *difference measure*. There are actually quite a few measures that have been considered in the literature, the most prominent ones being:

$$\begin{aligned} d(H,E) &= p(H|E) - p(H) \\ r(H,E) &= \log\left(\frac{p(H|E)}{p(H)}\right) \\ l(H,E) &= \log\left(\frac{p(E|H)}{p(E|\bar{H})}\right) \\ s(H,E) &= p(H|E) - p(H|\bar{E}) \end{aligned}$$

And corresponding to each candidate measure, we have a variant of the naive account :

The c_x Account of Confirmation E confirms H if and only if $p(H|E) - p(H)$, and the degree to which E confirms H is $c_x(H,E)$.

$c_x(H,E)$ can be whatever measure we favor, usually one of the four listed. Outside of the context of Glymour's problem, there are various desiderata that have been used to argue that one or another measure provides a superior account. For historical surveys and thorough coverage of this debate, see (Fitelson 1999) and (Fitelson 2001).

As far as Glymour's challenge goes, we've already seen that d does poorly, and a similar problem afflicts r . The virtues of l and s are explored by Christensen (1999), who raises the following problem case for l . Suppose we roll a fair die and have yet to observe how it turns up. Let L be the proposition that the roll was a low number — 1, 2, or 3 — and O the proposition that it was an odd number. Intuitively, L should confirm O . Suppose that L becomes old evidence, however; our reliable friend tells us that the toss was low, raising the probability of L to .99, and we update our other probabilities by Jeffrey Conditionalization (section 3.4). A little arithmetic shows that $p(L|O) = .99$ and $p(L|\bar{O}) \approx .98$ so that $l(O,L) \approx .01$, saying that L is effectively neutral with respect to O . So l is subject to the problem of old evidence too. But s does better: it says that L confirms O both before and after L becomes old evidence, and even by the same amount at each time.

And yet, as Christensen shows, there are problems even for s . Suppose H is the hypothesis that deer live in a given wood, D is the possibility that there are

deer droppings in location x , and A is the possibility that there is a shed deer antler at location y . Initially, $p(H) = 1/2$, $p(D) = .001$, and $p(A) = .0001$. Now suppose that we find what look like deer droppings in location x , so that D becomes old evidence, making $p(D)$ and $p(H)$ very high. As before, s will still say that D confirms H , which is good. The trouble is that s will now say that A provides much less confirmation for H , despite the fact that A is just as good evidence for H as D is. But, because H now has a high probability whether or not A is true, s will miss out on this fact. The moral Christensen draws is that intuitive judgments of confirmation are relative to a sort of base-line body of evidence, wherein D is not already assumed. So, once the probabilities reflect our knowledge of D , the probabilistic correlations will no longer reflect our intuitive judgments of confirmation.

Christensen draws the pessimistic conclusion that no Bayesian measure will be able to account for intuitive judgments of confirmation. Eells and Fitelson (2000) disagree, arguing that Christensen gives up on Bayesian confirmation too quickly. Their reasoning, however, appears to me to be incompatible with any Bayesianism that interprets probability as degree of belief and lives towards the subjective end of the subjective-objective continuum (section 3), as many Bayesians do.

6.3 The Popper-Rényi Solution

The strategy we just considered, of appealing to an alternate measure of confirmation, depended crucially on the appeal to Continuing Regularity. If $p(E) = 1$, then d and r become neutral, l becomes a measure of absolute confirmation, and s becomes undefined. But even if Continuing Regularity is true, it may be that it would not help with Glymour's problem. Continuing Regularity says that you shouldn't become certain of your evidence, but if you do, it seems you should still be able to say whether E supports H . Why should your evaluation of the evidential connection between E and H depend crucially on whether you are certain or just nearly certain of E ? Arguably, even if E did attain probability 1, it would still be true that it confirms H . To see this, notice that a being with perfect evidence-gathering faculties, for whom assigning $p(E) = 1$ to evidence is reasonable, should still say that E confirms H , even after E is old evidence.

We can eliminate this limitation of the alternate-measure strategy if we appeal to Popper-Rényi functions, which is what Joyce (1999) recommends. Using Popper-Rényi functions, s can still be defined when $p(E) = 1$, and can even be arbitrarily close to its maximum value of 1. The fact that given evidence still confirms a theory is, according to Joyce, captured by the fact that s takes a positive value when calculated according to the Popper-Rényi function that represents the agent's beliefs. Joyce does not reject d as a measure of confirmation, however.

Instead he adopts a pluralist stance on which both d and s measure quantities important to confirmation.

Fitelson (2003) faults Joyce for introducing problematic ambiguity into the Bayesian analysis of confirmation. Fitelson points out that, when s is positive despite $p(E) = 1$, it is still the case that d takes a 0 value. So the move to Popper-Rényi probabilities allows these two measures to yield different *qualitative* judgments when $p(E) = 1$: s can say that the evidence supports the hypothesis even though d says that the evidence is neutral. Fitelson's concern is that both d and s are supposed to be measures of confirmation, and yet they have different qualitative properties, which would seem to make confirmation ambiguous, not just quantitatively, but qualitatively too. The move to Popper-Rényi functions also leaves us with the same old problem of old evidence for the confirmational quantity $d(H, E)$, which is still going to be 0.

To avoid the ambiguity concern we might adopt Joyce's Popper-Rényi approach without his pluralism, embracing just s as our measure of confirmation. But Fitelson also finds fault with the measure s on its own terms, since it violates what he takes to be desiderata on any adequate measure of confirmation (Eells & Fitelson 2000; Fitelson 2006). For example, it is possible to have $p(H|E_1) > p(H|E_2)$ but $s(H, E_1) < s(H, E_2)$, which Fitelson finds unacceptable. And, as Christensen showed with his deer-in-the-woods example, even s seems to face a version of the old evidence problem.

6.4 Objective, Ternary Solutions

Above I said that the historical and counterfactual solutions don't work for a degree of belief interpretation of probability that lives at the subjective end of the subjective-objective continuum, because the historical or counterfactual degrees of belief can easily be such that $p(H|E) \neq p(H)$. But a more objectivist view may be able to more successfully exploit the idea behind the historical and counterfactual responses. The driving idea is that, if we delete E from our stock of knowledge, then we do have $p(H|E) > p(H)$. We've seen that the probabilities here can't be historical or counterfactual, but they could be *normative*. That is, we might be able to say that E confirms H , even after we know E , because if we didn't know E then the *right* probabilistic evaluation to make would be $p(H|E) > p(H)$. And we might be able to say something similar if the probabilities are logical or primitive. So it looks like the problem of old evidence really only affects subjective, degree of belief brands of Bayesianism.

It's not that simple, of course. We can't just use the counterfactual supposition "if we didn't know E ," because who knows what other things we would not know if we didn't know E , but which are still relevant to our judgment that E

confirms H . So we need a more stable way of “deleting” E from our evidential corpus. One route to go on this is to appeal to the belief revision theory pioneered by Alchourrón, Gärdenfors, and Makinson (1985). This theory is designed to give answers to questions like “how should I revise my corpus of knowledge if I must add/remove proposition E ?”. Belief revision lays down a set of axioms that the new corpus of beliefs must satisfy after E is deleted but, unfortunately, these axioms tend to leave the end result extremely under-determined. The under-determination can be controlled by regarding some beliefs as more “entrenched” than others, but introducing the notion of entrenchment takes us well out of Bayesian territory, and raises the worry that we might be circling back on a probabilistic notion.

A more popular approach has been to make the corpus of assumptions explicit in the concept of confirmation, introducing a third variable to represent the background assumptions relative to which E 's bearing on H is being evaluated. On this view, the proper question to ask is whether E confirms H relative to background assumptions B . To answer this question, we start with the objectively correct probabilities before any evidence is gathered, $p(\cdot)$, and then consult $p(\cdot|B)$. This yields

The Objective, Ternary Account of Confirmation E confirms H relative to B if and only if $p(H|EB) > p(H|B)$, and the extent of the confirmation is $c_x(H,E)$ computed using the probability function $p(\cdot|B)$.

This account depends crucially on being able to make sense of $p(\cdot)$, the correct probability function sans any evidence, which a strongly objectivist view can do.³⁰ We could have $p(\cdot)$ be somewhat under-determined, but it must be fairly tightly constrained to agree with the judgments of confirmation that we actually make. Maher (1996) endorses this kind of objectivism, and gives a version of the objective, ternary account.³¹

The objective, ternary account might be accused of cheating by introducing the variable B , since we often simply ask whether E confirms H , making no mention of relativity to a body of assumptions. Maher addresses this issue, suggesting that the relative body of assumptions is implicit in the context of discourse when we talk this way. An account of how the context fixes B would help secure this view's respectability, and Maher does give a sketch (Maher 1996: pp. 166-

³⁰ Though the “correct” prior probability function may yet be interpretable in a subjectivist-friendly way, as the probabilities the agent thinks someone with no evidence ought to have, for example.

³¹ Eells and Fitelson adopt a formally similar account of confirmation (they call it “evidence for”, to distinguish it from a different notion of confirmation they discuss) in their aforementioned criticism of Christensen (Eells & Fitelson 2000), though they seem to want to interpret $p(\cdot)$ historically, as the agent's past degrees of belief.

168). Arguably, insisting on a complete account would be overly demanding, since understanding the mechanics of context is a major project in linguistics and philosophy of language. A better approach might be to just see whether ‘confirms’ has the properties that typify other, paradigmatic context-sensitive terms. This would vindicate the view that B is contextually specified, and then we could leave the details of how for the linguists and philosophers of language. Alternatively, we might treat the relativity to B by analogy with the relativity of simultaneity. Just as relativity theory teaches us that simultaneity is frame-relative, so too confirmation theory teaches us that confirmation is B -relative. We used to talk about confirmation as if it were a two-place relation but our theorizing has shown us the error of our ways. It was just that, in many cases, this relativity was insignificant and hence easy to overlook.

6.5 Ravens and the Tacking Problem

Outside of debates about which analysis of confirmation is correct and how Glymour’s problem ought to be solved, Bayesian discussions of confirmation tend to focus on solving classic puzzles of confirmation, especially Hempel’s Raven Paradox (Hempel 1937, 1945) and the tacking problem (a.k.a. the problem of irrelevant conjunction). For summaries of this literature and the current state of affairs, see (Vranas 2004; Fitelson 2006) on the Raven Paradox, and (Fitelson 2002; Fitelson & Hawthorne 2004) on the tacking problem. We will not cover these topics here, as they do not expose deep divides in the Bayesian camp.

7 Theories of Belief (a.k.a. Acceptance)

A good deal of Bayesian theorizing is concerned with the degree to which we ought to believe something. But what about the question whether you should believe something *tout court*? Over and above questions about probabilities and levels of confidence, there seems to be an additional, qualitative question about what one ought to believe. Should you believe that there is a God, or that your ticket won’t win the lottery? Probabilistic theorizing may answer the quantitative question of how confident you should be, but it leaves a major epistemological question unanswered. When should/shouldn’t you believe a given proposition, qualitatively speaking? Or, as some prefer to phrase it, which propositions should you *accept*?³²

³² The terms ‘belief’ and ‘acceptance’ get used in a confusing variety of ways, to mark a variety of distinctions. I will adopt the following convention here. I will use ‘belief’ and ‘acceptance’ interchangeably to refer to the same, qualitative propositional attitude, which is to be distinguished from the gradable attitude denoted by ‘degree of belief’ and ‘credence’. I will sometimes use

The obvious thing to conjecture is that you should believe those propositions that have attained a certain minimum threshold of probability, say .99. But Kyburg's (1961) lottery paradox shows that this conjecture leads to inconsistent belief states. Suppose that there is a fair lottery with 100 tickets, one of which will be the winner. Each ticket has a .99 probability of losing, and so the threshold conjecture says that you should believe of each ticket that it will lose. But the resulting set of beliefs is inconsistent, since you believe of each ticket that it will lose, and you also believe that one will win. Things get even worse if we endorse the principle that, if you believe *A* and you believe *B*, then you should believe *AB*.³³ Then you will be led to believe the explicit contradiction that all the tickets will lose and yet one will win.

The lottery paradox (and also the related preface paradox (Makinson 1965)) puts a point on the problem of elaborating the connection between probability and belief, and this might push us in either of two directions. One would be to eliminate belief talk in favor of degree-of-belief talk. Jeffrey (1968; 1970a), for example, seems to have felt that the folk notion of belief should be replaced by the more refined notion of degree of belief, since talk of belief is just a sloppy approximation of degree of belief. If we go this route, then we avoid the lottery paradox and get to skip out on the job of elaborating the probability-belief connection.³⁴

The other direction we might go is to look for a more sophisticated account of that connection. For example, we might say that high probability warrants belief, except when certain logical or probabilistic facts obtain. Pollock (1995), Ryan (1991; 1996), Nelkin (2000), and Douven (2002) represent some recent proposals in this vein. One such proposal is that high probability is sufficient for belief, except when the belief in question is a member of a set of similarly probable propositions whose conjunction is improbable.³⁵ Douven and Williamson (2006) argue that such proposals fail because every probable proposition is a member of such a set. They go on to argue that any proposal that restricts itself to probabilistic and

'qualitative belief' to stress the contrast with degree of belief. The distinction between 'belief' and 'acceptance' often introduced in discussions of scientific realism (e.g., (van Fraassen 1980)) will be ignored.

³³ Foley (2009) responds to the lottery paradox by rejecting this principle.

³⁴ Christensen (2004) endorses a less extreme option in this neighborhood. He holds that qualitative belief talk is a way of sorting more fine-grained epistemic attitudes, similar to our sorting of dogs into the the categories 'large', 'medium', and 'small'. Just as sorting dogs in this way has great utility but does not carve nature at the joints, so too sorting our finer-grained attitudes into beliefs and non-beliefs is useful but does not expose any "real" kind. According to Christensen, whatever our scheme for sorting finer-grained attitudes into beliefs and non-beliefs, the states so sorted do not need to obey a norm of deductive cogency. Deductive logic acts as an epistemic norm only insofar as it governs degrees of belief, via the probability axioms.

³⁵ This is roughly Ryan's (1996) proposal.

logical criteria will fail on similar grounds.

To avoid this problem we could look for other variables that, together with probability, tell us what to believe. This approach was pioneered by Levi (1967a), who argues that probability determines what we should believe when conjoined with a notion of epistemic value. Roughly, Levi's view is that you should believe something if doing so maximizes expected *epistemic* utility, also known as *cognitive* utility. Thus belief is a sort of decision problem, where we use a special notion of utility, one that captures cognitive values like true belief. Levi originally took epistemic utility to weigh two competing epistemic concerns, amalgamating them into a single scale of value. On the one hand, our inquiry is aimed at eliminating agnosticism in favor of giving informative answers to questions posed. On the other hand, we want those answers to be true as often as possible. Epistemic utility takes account of both aims: informativeness and truth. We should then believe exactly those propositions that maximize expected epistemic utility. In later work, Levi (1980) incorporates additional epistemic values into the scale of epistemic utility, such as simplicity and explanatory power.

There is a troublesome tension for views that, like Levi's, endorse both degrees of belief and belief simpliciter. On such views, belief threatens to be epiphenomenal, in the sense that it is determined by degree of belief but does nothing to determine degree of belief. Just like with epiphenomenalism in the philosophy of mind, this has the further consequence that there is no room for the epiphenomenon to determine action, since that job is taken. On the standard Bayesian view, rational action is determined by your degrees of belief via expected utility maximization. So qualitative belief is doubly idle; it neither influences your levels of confidence, nor what decisions you make. On this epiphenomenalist picture, it becomes hard to see the point of keeping belief in our theory, tempting us to follow Jeffrey's lead and just forget about qualitative belief altogether.

To avoid the epiphenomenalist conundrum, we might give qualitative belief a more substantial role, as Levi does in later work. Levi (1980) takes a view of belief where the believed propositions are treated more like evidence, with all accepted propositions receiving probability 1. Going in this direction is tricky though. According to expected utility maximization, attributing probability 1 to a proposition obliges you to bet on it at any odds, yet we believe many things we would not stake our lives on. A related worry is that everything we believe becomes equally certain on this view, and yet we are able to distinguish grades of certainty in the things we believe. In short, if belief is not epiphenomenal because accepted beliefs acquire probability 1, then belief threatens to become too dogmatic.

To avoid these worries we might take a more tentative view of belief. Levi's own view is that beliefs, while "infallible" in the sense of having credence 1, are

not “incorrigible”, since they can, and often should, be retracted. In particular, Levi thinks there are two cases where beliefs should be retracted: (i) when an inconsistency is discovered, or (ii) when we want to contemplate accepting a proposition that has been rejected, but might improve our explanatory picture if added. Such retractions are not enough to solve our problems though. My belief that Albany is the capital of New York does not contradict any of my other beliefs, nor is it inconsistent with any hypothesis that promises to improve my explanatory picture, so I have no reason of kind (i) or (ii) to retract it. Nevertheless, I would not bet on it at any odds, and I can appreciate that it is less certain than my belief that Albany is in New York.

Because it is difficult to give belief a substantial cognitive role without inappropriately interfering with the role of degrees of belief, some Bayesians prefer to stick to the epiphenomenalist conception of belief, arguing that it is an important epistemic attitude even if it is idle in the ways mentioned. Maher (1993) takes this approach, arguing that the notion of acceptance is crucial to understanding the history of science. Without it, he argues, we cannot explain classic episodes of scientific confirmation, the Kuhnian observation that paradigms are only rejected once an alternative is available, nor the truism that it is never unreasonable to gather cost-free evidence.

Would it be an adequate defense of the epiphenomenal view if we showed that belief is an interesting epiphenomenon? As Maher acknowledges, “it is standardly assumed that you believe *H* just in case you are willing to act as if *H* were true,” (Maher 1993: p. 152) and an epiphenomenal view is hard to reconcile with this truism, no matter how good a job we do at showing that belief is crucial to our understanding of science. For how could belief be reliably tied to action if it is epiphenomenal? Maher’s reaction is to conclude that the folk concept of belief is simply flawed. It presupposes that two distinct states are really one; or at least that they are necessarily correlated. On the one hand there is the willingness to act as if *H* were true, and on the other there is the state expressed by sincere assertions that *H*. Maher’s conception of acceptance is directed at the latter state, whereas the truism about willingness to act is directed at the former. The right approach to take, according to Maher, is to acknowledge the difference between these two states by theorizing about them separately, instead of looking for a notion of belief that covers both.

Alternatively, one might try to reconcile full belief with degrees of belief by telling a story on which both states play a significant role in determining assertions, inferences, and actions. Frankish (2009), for example, suggests that full beliefs are realized in certain degrees of belief and utilities. When you believe that *H*, this is because you are highly confident that you have adopted a policy of using *H* as a premise in certain deliberative contexts (both epistemic and pragmatic),

and you attach a high utility to adhering to that policy. On Frankish's view, those degrees of belief and utilities are what make it the case that you believe H . Thus, when you assert H , draw conclusions from H , and act based on H , there are two, compatible explanations: that you believe that H , and that you have certain degrees of belief and utilities. Because you believe H in virtue of having those degrees of belief and utilities, both explanations are correct.

One might worry that this story does not reconcile your belief that H with the right degrees of belief. On Frankish's view, when you use H as a premise in a piece of practical reasoning, the degrees of belief and utilities that explain why you did so are your degrees of belief about what policies you've adopted and the utilities you attach to sticking by those policies. But, on the standard Bayesian story, the explanation proceeds in terms of your degrees of belief about the possible outcomes of the act in question. What if you believe that you have adopted a policy that mandates assuming H in this context, but your degrees of belief about the consequences of acting on H make such action sub-optimal? According to the Bayesian, you should not so act, but according to Frankish, it seems that you would. But evaluating the seriousness of this worry would require us to go into more detail on the nature of premising policies and their adoption. See (Frankish 2009: pp. 83–90) for more detail on premising policies, and for responses to some worries in this neighborhood.

Another approach is to allow the distinct existence of beliefs and degrees of belief, while trying to give both states a significant cognitive role. Consider the way we normally operate, unreflectively, when going about our business. In most contexts it feels as if we don't distinguish between what is more or less probable; we simply work off of a body of assumptions that we take for granted. When I set my alarm at night, wake up to its ring, step in and out of the shower, and head down the street to catch my bus, I am not thinking reflectively and I seem to treat the assumptions that drive my planning and behavior as all on a par. None are more or less likely, I simply accept that my alarm will ring at the right time, that it is time to get up when it rings, that my shower will turn on when I twist the knob, that my bus will be there in the morning just like it always is, and that it will take me to campus on time like it always does. If you asked me which of the assumptions that go into my daily routine are more or less likely, I could surely give some differentiating answers. My bus is more likely to be late than my shower is to not turn on when I twist the knob. But these differences are ones that I have decided not to worry about, because the assumptions in play are all reliable enough that it would not be worth thinking about my morning routine at a more fine-grained level. If I were more careful about how much to rely on each assumption, I might be able to decrease the chances that, one day, I will be late to class. But as it is, I am late rarely enough that it is worth my while to simplify by

just taking them all for granted, and to act accordingly.

These considerations suggest two levels of cognitive operation, one at which cognition and planning happen in a qualitative mode, and one where they happen in a more fine-grained, quantitative mode. The obvious thing to conjecture is that the purely qualitative mode is useful because it is simpler and more efficient, making daily tasks and life in general manageable. But the qualitative mode is rough, and may need to be reset, rearranged, or overridden altogether when we encounter an eventuality that we assumed would not happen, when we face a new situation, when greater accuracy is desired, etc. If my shower doesn't turn on I'm going to have to think more carefully about how to rearrange my morning routine, and if I move to a new location I'm going to have to create an entirely new morning routine, based on a whole new set of qualitative assumptions and intermediate goals. On such a model, quantitative and qualitative belief might be able to coexist, and even be complementary. But developing this thought into a substantive view poses a serious challenge, since we would want a precise story about when each mode gets engaged, whether/how the two modes interact, share information, and so on.

8 Summary

We distinguished the varieties of Bayesianism by considering six questions Bayesians disagree about. First we asked what the proper subject matter of Bayesianism is, and we considered three (not necessarily exclusive) answers: degrees of logical truth, degrees of belief, and epistemic probabilities taken as primitive. Second, we asked what rules over and above the probability axioms are correct, exploring eight proposals: five synchronic rules, and three diachronic rules of increasing generality. Third, we considered what justificatory arguments were appropriate for these various rules, noting the many rebuttals and criticisms that have been offered over the years.

Our discussion of those three questions laid out a variety of Bayesian views on core issues. The remaining three questions exposed a variety of views on applications of Bayesianism to related subject matter.

Our fourth question was how Bayesianism applies to decision making. Presupposing a numerical notion of value ("utility"), we considered three versions of the expected utility rule, each resulting from a different view about the salient probabilistic relationship between an act and its possible outcomes. Fifth, we considered Bayesian attempts to analyze discourse about confirmation. In our attempts to deal with Glymour's challenge for the naive account, we considered four ways of measuring probabilistic support, the merits of Popper-Rényi probabilities, and the possible advantage of treating confirmation as a three-place relation (as opposed

to two-place). Sixth and finally, we considered how qualitative belief fits with the Bayesian framework. We saw that the lottery paradox motivates a dispute between those who prefer to eliminate belief in favor of degrees of belief and those who see an important role for belief in cognition. We also saw how degrees of belief threaten to take over the role of qualitative belief in cognition, making qualitative belief epiphenomenal.

In the course of our discussion we encountered many connections between possible answers to these six questions. For example, some proposed rules were not subject to the same sorts of justificatory arguments as others; there are Dutch book arguments for the probability axioms but not for the Principle of Indifference. Thus those who only feel comfortable with principles that can be defended by a Dutch book argument may prefer a more subjective brand of Bayesianism. Another connection we encountered was between Popper-Rényi probabilities and confirmation; Popper-Rényi probabilities allowed us to measure confirmation in ways we could not on the standard axiomatization. Thus considerations in confirmation theory may motivate a preferred axiomatization of probability.

Still, the list of possible views to be generated by mixing and matching answers to our six questions is too large to be considered explicitly. And our list of contentious questions is incomplete anyway. Hence the famous quip that “there must be at least as many Bayesian positions as there are Bayesians.” (Edwards et al. 1963)

References

- Aczél, J. 1966. *Lectures on Functional Equations and Their Applications*. Academic Press.
- Adams, Ernest W. 1964. On rational betting systems. *Archiv für Mathematische Logik und Grundlagenforschung* 6: 7–29.
- Albert, David. 2001. *Time and Chance*. Harvard University Press.
- Alchourrón, Carlos E., Peter Gärdenfors & David Makinson. 1985. On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic* 50(2): 510–530.
- Allais, Maurice. 1979. The so-called allais paradox and rational decisions under uncertainty. In Maurice Allais & Ole Hagen (eds.) *Expected Utility Hypotheses and the Allais Paradox*. D. Reidel.
- Arntzenius, Frank, Adam Elga & John Hawthorne. 2004. Bayesianism, infinite decisions, and binding. *Mind* 113: 251–283.
- Arntzenius, Frank & Ned Hall. 2003. On what we know about chance. *British Journal for the Philosophy of Science* 54: 171–179.

- Bertrand, Joseph L. F. [1888] 2007. *Calcul Des Probabilités*. Oxford University Press.
- Bradley, Richard. 2005. Probability kinematics and bayesian conditioning. *Philosophy of Science* 72.
- Briggs, Rachael Amy. forthcoming. Distorted reflection. *The Philosophical Review*
- Carnap, Rudolph. 1950. *Logical Foundations of Probability*. Chicago: University of Chicago Press.
- Carnap, Rudolph. 1952. *The Continuum of Inductive Methods*. Chicago: University of Chicago Press.
- Christensen, David. 1991. Clever bookies and coherent beliefs. *The Philosophical Review* 100(2): 229–247.
- Christensen, David. 1992. Confirmational holism and bayesian epistemology. *Philosophy of Science* 59.
- Christensen, David. 1999. Measuring confirmation. *Journal of Philosophy* 96: 437–461.
- Christensen, David. 2001. Preference-based arguments for probabilism. *Philosophy of Science* 68.
- Christensen, David. 2004. *Putting Logic in its Place*. Oxford University Press.
- Christensen, David. 2007. Epistemology of disagreement: The good news. *The Philosophical Review* 116(2): 187–217.
- Cox, Richard T. 1946. Probability, frequency, and reasonable expectation. *American Journal of Physics* 14: 1–13.
- De Finetti, Bruno. 1937. La prévision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré* 17.
- De Finetti, Bruno. 1970. *Theory of Probability*. New York: John Wiley.
- De Finetti, Bruno. 1972. *Probability, Induction, and Statistics*. New York: John Wiley.
- Diaconis, Persi & Sandy L. Zabell. 1982. Updating subjective probability. *Journal of the American Statistical Association* 77(380): 822–830.
- Domotor, Zoltan. 1980. Probability kinematics and representation of belief change. *Philosophy of Science* 47.
- Doris, John M. 2002. *Lack of Character: Personality and Moral Behavior*. Cambridge University Press.
- Douven, Igor. 2002. A new solution to the paradoxes of rational acceptability. *British Journal for the Philosophy of Science* 53(3): 391–410.
- Douven, Igor & Timothy Williamson. 2006. Generalizing the lottery paradox. *British Journal for the Philosophy of Science* 57(4): 755–779.
- Earman, John. 1992. *Bayes or Bust: A Critical Examination of Bayesian Confirmation Theory*. The MIT Press.

- Edwards, Ward, Harold Lindman & Leonard J. Savage. 1963. Bayesian statistical inference for psychological research. *Psychological Review* 7(3): 193–242.
- Eells, Ellery & Branden Fitelson. 2000. Comments and criticism: Measuring confirmation and evidence. *Journal of Philosophy* 97(12): 663–672.
- Egan, Andy. 2007. Some counterexamples to causal decision theory. *The Philosophical Review* 116(1): 93–114.
- Elga, Adam. 2000. Self-locating beliefs and the sleeping beauty problem. *Analysis* 60(2): 143–147.
- Elga, Adam. 2007. Reflection and disagreement. *Noûs* 41(3): 478–502.
- Eriksson, Lina & Alan Hájek. 2007. What are degrees of belief? *Studia Logica* 86: 185–215.
- Etchemendy, John. 1990. *The Concept of Logical Consequence*. Harvard University Press.
- Field, Hartry. 1978. A note on jeffrey conditionalization. *Philosophy of Science* 45.
- Fine, Terence L. 1973. *Theories of Probability*. Academic Press.
- Fitelson, Branden. 1999. The plurality of bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science* 66 (Proceedings): S362–S378.
- Fitelson, Branden. 2001. *Studies in Bayesian Confirmation Theory*. Ph.D. thesis, University of Wisconsin, Madison. URL <http://fitelson.org/thesis.pdf>.
- Fitelson, Branden. 2002. Putting the irrelevance back into the problem of irrelevant conjunction. *Philosophy of Science* 69: 611–622.
- Fitelson, Branden. 2003. Review of the foundations of causal decision theory. *Mind* 112: 545–551.
- Fitelson, Branden. 2006. The paradox of confirmation. *Philosophy Compass* 1: 95.
- Fitelson, Branden & James Hawthorne. 2004. Discussion: Re-solving irrelevant conjunction with probabilistic independence. *Philosophy of Science* 71: 505–514.
- Foley, Richard. 2009. Beliefs, degrees of belief, and the lockean thesis. In Franz Huber & Christoph Schmidt-Petri (eds.) *Degrees of Belief, Synthese Library*, vol. 342, 37–47. Springer.
- Frankish, Keith. 2009. Partial belief and flat-out belief. In Franz Huber & Christoph Schmidt-Petri (eds.) *Degrees of Belief, Synthese Library*, vol. 342. Springer.
- Gaifman, Haim & Marc Snir. 1982. Probabilities over rich languages, testing, and randomness. *Journal of Symbolic Logic* 47: 495–548.
- Garber, Daniel. 1980. Field and jeffrey conditionalization. *Philosophy of Science* 47.
- Gärdenfors, Peter. 1988. *Knowledge in Flux: Modelling the Dynamics of Epistemic States*. The MIT Press.
- Gibbard, Allan & William Harper. 1978. Counterfactuals and two kinds of ex-

- pected utility. In A. Hooker, J. J. Leach & E. F. McClennen (eds.) *Foundations and Applications of Decision Theory*. D. Reidel.
- Glymour, Clark. 1980. *Theory and Evidence*. Princeton University Press.
- Goldstick, Daniel. 2000. Three epistemic senses of probability. *Philosophical Studies* 101: 59–76.
- Goodman, Nelson. 1954. *Fact, Fiction, and Forecast*. Cambridge: Harvard University Press.
- Greaves, Hilary & David Wallace. 2006. Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind* 115: 607–632.
- Grove, Adam & Joseph Halpern. 1997. Probability update: Conditioning vs. cross-entropy. In *Proceedings of the Thirteenth Annual Conference on Uncertainty in Artificial Intelligence*.
- Hacking, Ian. 1971. Equipossibility theories of probability. *British Journal for the Philosophy of Science* 22(4): 339–355.
- Hájek, Alan. 2003. What conditional probability could not be. *Synthese* 137: 273–323.
- Hájek, Alan. 2008. Dutch book arguments. In Paul Anand, Prasanta Pattanaik & Clemens Puppe (eds.) *The Oxford Handbook of Rational and Social Choice*. Oxford University Press.
- Hall, Ned. 1994. Correcting the guide to objective chance. *Mind* 103: 505–517.
- Halmos, Paul R. 1974. *Measure Theory (Graduate Texts in Mathematics)*. Springer-Verlag.
- Halpern, Joseph Y. 1999. A counterexample to theorems of Cox and Fine. *Journal of Artificial Intelligence* 10: 67–85.
- Halpern, Joseph Y. 2001. Lexicographic probability, conditional probability, and nonstandard probability. In *Proceedings of the Eighth Conference on Theoretical Aspects of Rationality and Knowledge*, 17–30. Morgan Kaufmann Publishers Inc.
- Halpern, Joseph Y. 2003. *Reasoning About Uncertainty*. The MIT Press.
- Hawthorne, James. 2008. Inductive logic. URL <http://plato.stanford.edu/entries/logic-inductive/>.
- Hempel, Carl G. 1937. Le problème de la vérité. *Theoria* 3: 206–246.
- Hempel, Carl G. 1945. Studies in the logic of confirmation i. *Mind* 54: 1–26.
- Hoover, Douglas N. 1980. A note on regularity. In Richard C. Jeffrey (ed.) *Studies in Inductive Logic and Probability*. Berkeley: University of California Press.
- Howson, Colin & Allan Franklin. 1994. Bayesian conditionalization and probability kinematics. *British Journal for the Philosophy of Science* 45: 451–466.
- Howson, Colin & Peter Urbach. 1993. *Scientific Reasoning: The Bayesian Approach*. Open Court.
- Huber, Franz. 2005. What is the point of confirmation? *Philosophy of Science*

- 72(5): 1146–1159.
- Jaynes, Edwin T. 1968. Prior probabilities. *IEEE Transactions On Systems and Cybernetics* SSC-4(3): 227–241.
- Jaynes, Edwin T. 1973. The well-posed problem. *Foundations of Physics* 3: 477–493.
- Jeffrey, Richard C. 1965. *The Logic of Decision*. University of Chicago Press.
- Jeffrey, Richard C. 1968. Review of gambling with truth: An essay on induction and the aims of science. *The Journal of Philosophy* 65(10): 313–322.
- Jeffrey, Richard C. 1970a. Dracula meets wolfman: Acceptance vs. partial belief. In Marshall Swain (ed.) *Induction, Acceptance, and Rational Belief*. D. Reidel.
- Jeffrey, Richard C. 1970b. Untitled review. *The Journal of Symbolic Logic* 35(1): 124–127.
- Jeffrey, Richard C. 1983. Bayesianism with a human face. In John Earman (ed.) *Testing Scientific Theories*. University of Minnesota Press.
- Jeffreys, Harold. [1939] 2004. *Theory of Probability*. Oxford University Press.
- Joyce, James. 1998. A nonpragmatic vindication of probabilism. *Philosophy of Science* 65(4): 575–603.
- Joyce, James. 1999. *The Foundations of Causal Decision Theory*. Cambridge University Press.
- Joyce, James. 2009. Accuracy and coherence: Prospects for an alethic epistemology of partial belief. In Franz Huber & Christoph Schmidt-Petri (eds.) *Degrees of Belief, Synthese Library*, vol. 342, 263–297. Synthese.
- Kahneman, Daniel & Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47.
- Kanovei, Vladimir & Saharon Shelah. 2004. A definable non-standard model of the reals. *Journal of Symbolic Logic* 69(1): 159–164.
- Kelly, Kevin T. 1996. *The Logic of Reliable Inquiry*. Oxford University Press.
- Kelly, Kevin T. 2000. The logic of success. *British Journal for the Philosophy of Science* 51: 639–666.
- Kelly, Thomas. 2005. The epistemic significance of disagreement. In *Oxford Studies in Epistemology*, vol. 1. Oxford University Press.
- Keynes, John Maynard. 1921. *A Treatise on Probability*. New York: MacMillan.
- Kyburg, Henry E. 1961. *Probability and the Logic of Rational Belief*. Wesleyan University Press.
- Kyburg, Henry E. 1987. Bayesian and non-bayesian evidence and updating. *Artificial Intelligence* 31: 271–293.
- Kyburg, Henry E. 1992. Getting fancy with probability. *Synthese* 90: 189–203.
- Lange, Marc. 2000. Is jeffrey conditionalization defective by virtue of being non-commutative? remarks on the sameness of sensory experience. *Synthese* 123.
- Levi, Isaac. 1967a. *Gambling With Truth*. A. Knopf.

- Levi, Isaac. 1967b. Probability kinematics. *British Journal for the Philosophy of Science* 18: 197–209.
- Levi, Isaac. 1974. On indeterminate probabilities. *Journal of Philosophy* 71: 391–418.
- Levi, Isaac. 1980. *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*. The MIT Press.
- Lewis, David. 1973a. Causation. *Journal of Philosophy* 70: 556–567.
- Lewis, David. 1973b. *Counterfactuals*. Oxford: Blackwell.
- Lewis, David. 1980. A subjectivist's guide to objective chance. In Richard C. Jeffrey (ed.) *Studies in Inductive Logic and Probability*, vol. II. University of California Press.
- Lewis, David. 1981. Causal decision theory. *Australasian Journal of Philosophy* 59(1): 5–30.
- Lewis, David. 1994. Humean supervenience debugged. *Mind* 103: 473–490.
- Lichtenstein, Sarah & Paul Slovic. 1971. Reversals of preferences between bids and choices in gambling decisions. *Experimental Psychology* 89.
- Lichtenstein, Sarah & Paul Slovic. 1973. Response-induced reversals of preference in gambling decisions: An extended replication in las vegas. *Journal of Experimental Psychology* 101.
- Loewer, Barry. 2001. Determinism and chance. *Studies in the History of Modern Physics* 609–620.
- Maher, Patrick. 1993. *Betting on Theories*. Cambridge University Press.
- Maher, Patrick. 1996. Subjective and objective confirmation. *Philosophy of Science* 63.
- Maher, Patrick. 2002. Joyce's argument for probabilism. *Philosophy of Science* 69: 73–81.
- Makinson, David C. 1965. The paradox of the preface. *Analysis* 25: 205–207.
- McGee, Vann. 1994. Learning the impossible. In Ellery Eells & Brian Skyrms (eds.) *Probability and Conditionals: Belief Revision and Rational Decision*, 179–199. Cambridge University Press.
- Meacham, Christopher J. G. 2005. Three proposals regarding a theory of chance. *Philosophical Perspectives* 19: 281–307.
- Meacham, Christopher J. G. & Jonathan Weisberg. unpublished. Debunking representation theorem arguments. *manuscript*.
- Miller, David W. 1966. A paradox of information. *British Journal for the Philosophy of Science* 17: 59–61.
- Nelkin, Dana K. 2000. The lottery paradox, knowledge, and rationality. *The Philosophical Review* 109(3): 373–408.
- Parikh, Rohit & Milton Parnes. 1974. Conditional probabilities and uniform sets. In A. Hurd & P. Loeb (eds.) *Victoria Symposium on Non-Standard Analysis*.

- New York: Springer Verlag.
- Plantinga, Alvin. 2000. Pluralism: A defense of religious exclusivism. In Philip L. Quinn & Kevin Meeker (eds.) *The Philosophical Challenge of Religious Diversity*. Oxford University Press.
- Pollock, John L. 1995. *Cognitive Carpentry*. Cambridge: MIT Press.
- Popper, Karl. 1959. *The Logic of Scientific Discovery*. London: Hutchinson & Co.
- Ramsey, Frank Plumpton. [1926] 1990. Truth and probability. In D. H. Mellor (ed.) *Philosophical Papers*. Cambridge: Cambridge University Press.
- Reichenbach, Hans. 1949. *The Theory of Probability: An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability*. English translation by E. H. Hutton and Maria Reichenbach. Berkeley and Los Angeles: University of California Press.
- Rényi, Alfred. 1970. *Foundations of Probability*. San Francisco: Holden-Day Inc.
- Robinson, Abraham. 1966. *Non-Standard Analysis: Studies in Logic and the Foundations of Mathematics*. Amsterdam: North-Holland.
- Rosenkrantz, Roger D. 1981. *Foundations and Applications of Inductive Probability*. Ridgeview Press.
- Ryan, Sharon. 1991. The preface paradox. *Philosophical Studies* 64(3): 293–307.
- Ryan, Sharon. 1996. The epistemic virtues of consistency. *Synthese* 109(22): 121–141.
- Savage, Leonard J. 1954. *The Foundations of Statistics*. Wiley Publications in Statistics.
- Schick, Frederic. 1986. Dutch books and money pumps. *Journal of Philosophy* 83: 112–119.
- Shafer, Glenn. 1976. *A Mathematical Theory of Evidence*. Princeton University Press.
- Shimony, Abner. 1955. Coherence and the axioms of confirmation. *Journal of Symbolic Logic* 20: 1–28.
- Shimony, Abner. 1988. An adamic derivation of the calculus of probability. In J. H. Fetzer (ed.) *Probability and Causality*. D. Reidel.
- Skyrms, Brian. 1980. *The Role of Causal Factors in Rational Decision*, chap. 2. Yale University Press.
- Skyrms, Brian. 1987. Dynamic coherence and probability kinematics. *Philosophy of Science* 54: 1–20.
- Skyrms, Brian. 1995. Strict coherence, sigma coherence and the metaphysics of quantity. *Philosophical Studies* 77: 39–55.
- Talbott, William J. 1991. Two principles of bayesian epistemology. *Philosophical Studies* 62: 135–150.
- Teller, Paul. 1973. Conditionalisation and observation. *Synthese* 26: 218–258.
- Thau, Michael. 1994. Undermining and admissibility. *Mind* 103: 491–503.

- van Fraassen, Bas. 1980. *The Scientific Image*. Oxford University Press.
- van Fraassen, Bas. 1981. A problem for relative information minimizers. *British Journal for the Philosophy of Science* 32.
- van Fraassen, Bas. 1983. Calibration: A frequency justification for personal probability. In R. Cohen & L. Laudan (eds.) *Physics, Philosophy, and Psychoanalysis*. D. Reidel.
- van Fraassen, Bas. 1984. Belief and the will. *The Journal of Philosophy* 81(5): 235–256.
- van Fraassen, Bas. 1989. *Laws and Symmetry*. Oxford University Press.
- van Fraassen, Bas. 1990. Figures in a probability landscape. In J. Dunn & A. Gupta (eds.) *Truth or Consequences*. Kluwer.
- van Fraassen, Bas. 1995. Belief and the problem of ulysses and the sirens. *Philosophical Studies* 77: 7–37.
- van Fraassen, Bas. 1999. Conditionalization, a new argument for. *Topoi* 18: 93–96.
- von Mises, Richard E. [1928] 1981. *Probability, Statistics, and Truth*. New York: Dover.
- Vranas, Peter B. M. 1998. Who's afraid of undermining. *Erkenntnis* 57(2): 151–174.
- Vranas, Peter B. M. 2004. Hempel's raven paradox: A lacuna in the standard bayesian account. *British Journal for the Philosophy of Science* 55: 545–560.
- Wagner, Carl. 2002. Probability kinematics and commutativity. *Philosophy of Science* 69.
- Wakker, Peter & Amos Tversky. 1993. An axiomatization of cumulative prospect theory. *Journal of Risk and Uncertainty* 7(7): 147–176.
- Walley, Peter. 1991. *Statistical Reasoning With Imprecise Probabilities*. Chapman & Hall.
- Weatherson, Brian. 1999. Begging the question and bayesianism. *Studies in History and Philosophy of Science* 30: 687–697.
- Weisberg, Jonathan. 2007. Conditionalization, reflection, and self-knowledge. *Philosophical Studies* 135(2): 179–197.
- Weisberg, Jonathan. 2009. Commutativity or holism? a dilemma for conditionalizers. *British Journal for the Philosophy of Science* forthcoming.
- White, Roger. 2009. Evidential symmetry and mushy credence. In *Oxford Studies in Epistemology*. Oxford University Press.
- Williams, P. M. 1980. Bayesian conditionalisation and the principle of minimum information. *British Journal for the Philosophy of Science* 32(2): 131–144.
- Williamson, Jon. 1999. Countable additivity and subjective probability. *British Journal for the Philosophy of Science* 50: 401–416.
- Williamson, Jon. 2007. Inductive influence. *British Journal for the Philosophy of Science* 58(4): 689–708.
- Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford University Press.

Zynda, Lyle. 2000. Representation theorems and realism about degrees of belief.
Philosophy of Science 67.