Philosophy of Probability

Aidan Lyon

Abstract

In the philosophy of probability there are two central questions we are concerned with. The first is: what is the correct formal theory of probability? Orthodoxy has it that Kolmogorov's axioms are the correct axioms of probability. However, we shall see that there are good reasons to consider alternative axiom systems. The second central question is: what do probability statements mean? Are probabilities "out there", in the world as frequencies, propensities, or some other objective feature of reality, or are probabilities "in the head", as subjective degrees of belief? We will survey some of the answers that philosophers, mathematicians and physicists have given to these questions.

1. Introduction

The famous mathematician Henri Poincaré once wrote of the probability calculus: "if this calculus be condemned, then the whole of the sciences must also be condemned" (Poincaré [1902], p. 186).

Indeed, every branch of science makes extensive use of probability in some form or other. Quantum mechanics is well–known for making heavy use of probability. The second law of thermodynamics is a statistical law and, formulated one way, states that the entropy of a closed system is most *likely* to increase. In statistical mechanics, a probability distribution known as the micro–canonical distribution is used to make predictions concerning the macro-properties of gases. In evolutionary theory, the concept of fitness is often defined in terms of a probability function (one such definition says that fitness is expected number of offspring). Probability also plays central roles in natural selection, drift, and macro-evolutionary models. The theory of three dimensional random walks plays an important role in the biomechanical theory of a diverse range of rubbery materials: from the resilin in tendons that help flap insect wings, to arteries near the heart that regulate blood flow. In ecology and conservation biology, we see concepts like the expected time to extinction of a population. In economics, stochastic differential equations are used extensively to model all sorts of quantities: from inflation to investment flows, interest rates, and unemployment figures. And all of the sciences make extensive use of probability in the form of statistical inference: from hypothesis testing, to model selection, to parameter estimation, to confirmation, to confidence intervals. In science, probability is truly everywhere.

But the sciences do not have exclusive rights to probability theory. Probability also plays an important role in our everyday reasoning. It figures prominently in our formal theories of decision making and game playing. In fact, probability is so pervasive in our lives that we may even be tempted to say that "the most important questions of life, are indeed for the most part only problems of probability", as Pierre–Simon Laplace once did (Laplace [1814], p. 1).

In philosophy of probability, there are two main questions that we are concerned with. The first question is: what is the correct mathematical theory of probability? Orthodoxy has it that this question was laid to rest by Andrei Kolmogorov in 1933 (Kolmogorov [1933]). But as we shall see in §2, this is far from true; there are many competing formal theories of probability, and it is not clear that we can single one of these out as *the* correct formal theory of probability.

These formal theories of probability tell us how probabilities behave, how to calculate probabilities from other probabilities, but they do not tell us what probabilities *are*. This leads us to the second central question in philosophy of probability: just what are probabilities? Or put another way: what do probability statements *mean*. Do probability claims merely reflect facts about human ignorance? Or do they represent facts about the world? If so, *which* facts? In §3, we will see some of the various ways in which philosophers have tried to answer this question. Such answers are typically called *interpretations of probability*, or philosophical theories of probability.

These two central questions are by no means independent of each other. What the correct formal theory of probability is clearly constrains the space of philosophical interpretations. But it is not a one way street. As we shall see, the philosophical theory of probability has a significant impact on the formal theory of probability too.

2. The Mathematical Theory of Probability

In probability theory, we see two types of probabilities: absolute probabilities and conditional probabilities. Absolute probabilities (also known as unconditional probabilities) are probabilities of the form "$P(A)$", while conditional probabilities are probabilities of the form "$P(A, B)$"—read as "the probability of A, given B".[1] These two types of probability can be defined in terms of each other. So when formalizing the notion of probability we have a choice: do we define conditional probability in terms of unconditional probability, or *vice versa*? The next section, §2.1, will focus on some of the various formal theories of probability that take absolute probability as primitive and define conditional probability in terms of the former. Then in §2.2, we will look at some of the various formal theories that take conditional probability as the primitive type of probability.

2.1. Absolute Probability as Primitive

Kolmogorov's theory of probability (Kolmogorov [1933]) is the best known formal theory of probability and it is what you will learn if you take a course on probability. First, we start with a set of elementary events, which we will refer to by Ω. For example, if we are considering the roll of a die where the possible

[1] Another common notation for conditional probability is $P(A|B)$, though some authors take this notation to have a very particular definition that does not correspond to our concept of conditional probability (see e.g., Hájek [2003]). Some authors also reverse A and B, so that $P(B, A)$ stands for the conditional probability of A given B, though this notation is not often used.

outcomes are the die landing with "one" face up, or with "two" face up, etc., then Ω would be the set $\{1, 2, 3, 4, 5, 6\}$. From this set of elementary events, we can construct other, less fine-grained events. For example, there is the event that an odd number comes up. We represent this event with the set $\{1, 3, 5\}$. Or, there is the event that some number greater than two comes up. We represent this event with set $\{3, 4, 5, 6\}$. In general, any event constructed from the elementary events will be a subset of Ω. The least fine-grained event is the event that *something* happens—this event is represented by Ω itself. There is also the event that cannot happen, which is represented by the empty set, $\varnothing$.[2]

In probability theory, we often want to work with the set of all events that can be constructed from Ω. In our die example, this is because we may want to speak of the probability of any particular number coming up, or of an even or odd number coming up, of a multiple of three, a prime number, etc. It is typical to refer to this set by $\mathscr{F}$. In our example, if $\mathscr{F}$ contains every event that can be constructed from Ω, then it would be rather large. A partial listing of its elements would be: $\mathscr{F} = \{\varnothing, \Omega, \{1\}, \{2\}, \ldots, \{6\}, \{1, 2, 3\}, \{4, 5, 6\}, \{1, 2\}, \{1, 2, 3, 4, 5\}, \ldots\}$. In fact, there are a total of $2^6 = 64$ elements in $\mathscr{F}$.

In general, if an event, A, is in $\mathscr{F}$, then so is its complement, which we write as $\Omega \setminus A$. For example, if $\{3, 5, 6\}$, is in $\mathscr{F}$, then its complement, $\Omega \setminus \{3, 5, 6\} = \{1, 2, 4\}$ is in $\mathscr{F}$. Also, if any two events are in $\mathscr{F}$, then so is their union. For example, if $\{1, 2\}$ and $\{4, 6\}$ are in $\mathscr{F}$, then their union $\{1, 2\} \cup \{4, 6\} = \{1, 2, 4, 6\}$ is in $\mathscr{F}$. If a set, S, has the first property (i.e., if A is in S, then $\Omega \setminus A$ is in S), then we say S is *closed under* Ω*-complementation*. If S has the second property (i.e., if A and B are in S, then $A \cup B$ is in S), then we say S is *closed under union*. And if S has both of these properties, i.e., if it is closed under both Ω-complementation and union, then we say that S is an *algebra* on Ω. If a set, S, is an algebra, then it follows that it is also closed under intersection, i.e., that if A and B are in S, then $A \cap B$ is also in S.

In our die example, it can be seen that $\mathscr{F}$ (the set that contains all the subsets of Ω) is an algebra on Ω. However, there are also algebras on Ω that do not contain every event that can be constructed from Ω. For example, consider the following set: $\mathscr{F} = \{\varnothing, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}$. The elements of this $\mathscr{F}$ would correspond to the events: (i) nothing happens; (ii) an odd number comes up; (iii) an even number comes up; and (iv) some number comes up. This is an important example of an algebra because it illustrates how algebras work. For example, not every "event"—intuitively understood—gets a probability. For instance, the event that the number two comes up gets no probability because $\{2\}$ is not in $\mathscr{F}$. Also note that even though this $\mathscr{F}$ does not contain every subset of Ω, it is still closed under union and Ω-complementation. For example, the union of $\{1, 3, 5\}$ and $\{2, 4, 6\}$ is Ω, which is in $\mathscr{F}$, and the Ω-complement of, say, $\{1, 3, 5\}$ is $\{2, 4, 6\}$, which is also clearly in $\mathscr{F}$.

Once we have specified an algebra, $\mathscr{F}$, we can then define a probability function that attaches probabilities to every element of $\mathscr{F}$. Let P be a function from $\mathscr{F}$ to the real numbers, $\mathbb{R}$, that obeys the following axioms:

(KP1) $P(A) \geq 0$

(KP2) $P(\Omega) = 1$

(KP3) $P(A \cup B) = P(A) + P(B)$, if $A \cap B = \varnothing$

[2] It is a theorem of standard set theory that the empty set is a subset of every subset.

for every A and B in $\mathscr{F}$. We call any function, P, that satisfies the above constraints a *probability function*.[3]

So far we have assumed that $\mathscr{F}$ only contains a finite number of events. But sometimes we want to work with infinitely many events (e.g., consider choosing a random point on a line; there are infnitely many points than can be chosen). When $\mathscr{F}$ is countably infinite[4], we replace KP3 with:

$$\text{(KP4)} \quad P\left(\bigcup_{i=1}^{i=\infty} A_i\right) = \sum_{i=1}^{i=\infty} P(A_i)$$

on the condition that the intersection of any two of the A_i is the empty set. A simple example will help illustrate how these last two axioms work. Suppose we randomly choose a positive integer in a way such that the number 1 has a probability of $1/2$ being chosen, the number 2 has a probability of $1/4$ being chosen, and so on. In general, the number n has a probability of $1/2^n$ of being chosen. Let us write the event that the number n gets chosen as A_n (this will be the set $\{n\}$). So for example, the event that some integer below 4 is chosen is: $A_1 \cup A_2 \cup A_3$ which is the set $\{1, 2, 3\}$. The probability of this event is then:

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2) + P(A_3) \\ &= 1/2 + 1/4 + 1/8 \end{aligned}$$

And the probability that *some* number is chosen is:

$$P\left(\bigcup_{i=1}^{i=\infty} A_i\right) = \sum_{i=1}^{i=\infty} P(A_i)$$

which can be expanded as:

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3 \cup ...) &= P(A_1) + P(A_2) + P(A_3) + ... \\ &= 1/2 + 1/4 + 1/8 + ... \\ &= 1 \end{aligned}$$

This fourth axiom—known as countable additivity—is by far the most controversial. Bruno de Finetti (de Finetti [1974]) famously used the following example as an objection to KP4. Suppose you have entered a fair lottery that has a countably infinite number of tickets. Since the lottery is fair, each ticket has an equal probability of being the winning ticket. But there are only two ways in which the tickets have equal probabilities of winning, and on both ways we run into trouble. On the first way, each ticket has some positive probability of winning—call this positive probability ϵ. But then, by KP4, the probability that *some* ticket wins is ϵ added to itself infinitely many times, which equals infinity, and so violates KP2. The only other way that the tickets can have equal prob-

[3]Also, for any Ω, $\mathscr{F}$, and P that satisfy the above constraints, we call the triple $(\Omega, \mathscr{F}, P)$ a *probability space*.

[4]A set is countable if there is a one to one correspondence between it and some subset of the Natural Numbers, and uncountable if there is not. Some examples: any finite set is countable, the set of even integers is countable (and thus countably infinite), the set of rational numbers is countable (and countably infinite), the set of real numbers is *not* countable, and the set of real numbers between 0 and 1 is also not countable.

ability of winning is if they all have zero probability. But then KP4 entails that the probability that some ticket wins is 0 added to itself infinitely many times, which is equal to zero and so again KP2 is violated. It is a matter of either too much or too little!

Axioms KP1–4 define absolute probability functions on *sets*. However, many philosophers and logicians prefer to define probability functions on *statements*, or even other abstract objects instead. One reason for this is because Kolmogorov's axioms are incompatible with many philosophical interpretations of probability. For example, Karl Popper points out that the formal theory of probability should be sensitive to the needs of the philosophical theory of probability:

> In Kolmogorov's approach it is assumed that the objects a and b in $p(a, b)$ are sets (or aggregates). But this assumption is not shared by all interpretations: some interpret a and b as states of affairs, or as properties, or as events, or as statements, or as sentences. In view of this fact, I felt that in a formal development, no assumption concerning the nature of the "objects" or "elements" a and b should be made [...]. Popper [1959b], p. 40

A typical alternative to Kolmogorov's set theoretic approach to probability is an axiom system where the bearers of probability are sentences (in §2.2, we shall see an axiom system inspired by Popper's work that makes less assumptions concerning what the bearers of probability are). For cases where there are only finitely many sentences, it is fairly easy to "translate" axioms KP1–3 into axioms that define probability functions on sentences. Instead of an algebra, $\mathscr{F}$, we have a language, $\mathcal{L}$. $\mathcal{L}$ is a set of atomic sentences and their Boolean combinations. So, if A and B are in $\mathcal{L}$, then so is: their conjunction, $A \wedge B$, read as "A and B"; their disjunction $A \vee B$, read as "A or B"; their negations, e.g., $\neg A$, read as "not A"; their equivalence, $A \equiv B$, read as "A is equivalent to B"; and their material implications, e.g., $A \supset B$, read as "if A, then B". We also define a consequence relation $\vdash$ over the language $\mathcal{L}$. So for example, $A \vdash B$ is read as "A entails B", or "B is a consequence of A". Because tautologies are always true, if A is a tautology, we write $\vdash A$, and since logical falsehoods are always false, if A is a logical falsehood, we write $\vdash \neg A$.[5] We then let P be a function from $\mathcal{L}$ to $\mathbb{R}$ that satisfies:

$$(P1) \quad P(A) \geq 0$$
$$(P2) \quad P(A) = 1, \text{ if } \vdash A$$
$$(P3) \quad P(A \vee B) = P(A) + P(B), \text{ if } \vdash \neg(A \wedge B)$$

for every A and B in $\mathcal{L}$. However, if we are interested in languages with infinitely many sentences, we cannot simply add on a fourth axiom for countable additivity as we did with axiom KP4. This is because we don't have infinite disjunction our logic. A new logic has to be used if we wish to have a countable additivity axiom for probabilities attached to sentences.[6] Axioms P1–3 are what most philosophers use, so I will refer to them as the standard theory of probability (however, the reader should not take this to mean they are the "correct" axioms of probability).

[5] This overview of classical logic is necessarily very brief. The reader is who is not familiar with these ideas is encouraged to read CHAPTER ON LOGIC.

[6] See e.g., Roeper and Leblanc [1999], p. 26 for details on how this can be done.

We also have the notion of *conditional* probability to formalize. On the accounts just mentioned it is typical to define conditional probability in terms of unconditional probability. For example, in the standard theory of probability, conditional probability is defined in the following way:

$$\text{(CP)} \quad P(A,B) \overset{\text{df}}{=} \frac{P(A \wedge B)}{P(B)}, \text{ where } P(B) \neq 0$$

Many have pointed out that this definition leaves an important class of conditional probabilities undefined when they should be defined. This class is comprised of those conditional probabilities of the form, $P(A,B)$ where $P(B) = 0$. Consider the following example, due to Émile Borel. Suppose a point on the Earth is chosen randomly—assume the Earth is a perfect sphere. What is the probability that the point chosen is in the western hemisphere, given that it lies on the equator? The answer intuitively ought to be $1/2$. However, CP does not deliver this result, because the denominator—the probability that the point lies on the equator—is zero.[7]

There are many responses one can give to such a problem. For instance, some insist that any event that has a probability of zero cannot happen. So the probability that the point is on the equator must be greater than zero, and so CP is not undefined. The problem though is that it can be proven that for any probability space with uncountably many events, uncountably many of these events *must* be assigned zero probability, as otherwise we would have a violation of the probability axioms.[8] This proof relies on particular properties of the real number system, $\mathbb{R}$. So some philosophers have said so much the worse for the real number system, opting to use a probability theory where the values of probabilities are not real numbers, but rather something more mathematically rich, like the hyper-reals, $\mathbb{HR}$ (see e.g., Lewis [1980] and Skyrms [1980]).[9]

Another response that philosophers have made to Borel's problem is to opt for a formal theory that takes conditional probability as the fundamental notion of probability (we will see some of these theories in §2.2). The idea is that by defining absolute probability in terms of conditional probability while taking the latter as the primitive probability notion to be axiomatized, conditional probabilities of the form $P(A,B)$ where $P(B) = 0$ can be defined.[10]

There are, also, other reasons to take conditional probability as the fundamental notion of probability. One such reason is that sometimes the unconditional probabilities $P(A \wedge B)$ and $P(B)$ are undefined while the conditional probability $P(A,B)$ *is* defined, so it is impossible to define the latter in terms of the former. The following is an example due to Alan Hájek (Hájek [2003]). Consider the conditional probability that heads comes up, given that I toss a coin

[7]It is zero because in this example probability corresponds to area. Since the equator has an area of zero, the probability that the chosen point lies on the equator is zero, even though this event is *possible*.

[8]See e.g., Hájek [2003] for the proof.

[9]The hyper-reals are all the real numbers (1, $3/4$, $\sqrt{2}$, $-\pi$, etc.), plus some more. For example, in the hyper-reals there is a number greater than zero but that is smaller than every positive real number (such a number is known as an *infinitesimal*).

[10]Kolmogorov was also well aware of Borel's problem and in response gave a more sophisticated treatment of conditional probability using the Radon–Nikodym theorem. The details of this approach are beyond the scope of this survey. Suffice to say though, this approach is not completely satisfactory either—see Hájek [2003] and Seidenfeld et al. [2001] for reasons why.

fairly. Surely, this should be 1/2, but CP defines this conditional probability as:

$$P(\text{heads}, \text{I toss the coin fairly}) = \frac{P(\text{heads} \wedge \text{I toss the coin fairly})}{P(\text{I toss the coin fairly})}$$

But you have no information about how likely it is that I will toss the coin fairly. For all you know, I never toss coins fairly, or perhaps I always toss them fairly. Without this information, the terms on the right hand side of the above equality may be undefined, yet the conditional probability on the left *is* defined.

There are other problems with taking absolute probability as the fundamental notion of probability, see Hájek [2003] for a discussion. These problems have led many authors to take conditional probability as primitive. However, this requires a new approach to the formal theory of probability. And so we now turn to theories of probability that take conditional probability as the primitive notion of probability.

2.2. Conditional Probability as Primitive

The following axiom system—based on the work of Alfred Rényi (Rényi [1955])—is a formal theory of probability where conditional probability is the fundamental concept. Let Ω be a non-empty set, $\mathscr{A}$ be an algebra on Ω, and $\mathscr{B}$ be a non-empty subset of $\mathscr{A}$. We then define a function, P, from $\mathscr{A} \times \mathscr{B}$ to $\mathbb{R}$ such that:[11]

(RCP1) $P(A, B) \geq 0,$

(RCP2) $P(B, B) = 1,$

(RCP3) $P(A_1 \cup A_2, B) = P(A_1, B) + P(A_2, B), \text{ if } A_1 \cap A_2 = \varnothing$

(RCP4) $P(A_1 \cap A_2, B) = P(A_1, A_2 \cap B) \cdot P(A_2, B)$

where the As are in $\mathscr{A}$ and the Bs are in $\mathscr{B}$. Any function that satisfies these axioms is called a *Rényi conditional probability* function.[12] RCP3 is the conditional analogue of the KP3 finite additivity axiom and it also has a countable version:

$$\text{(RCP5)} \quad P\left(\bigcup_{i=1}^{i=\infty} A_i, B\right) = \sum_{i=1}^{i=\infty} P(A_i, B)$$

on the condition that the A_i are mutually exclusive. RCP4 is the conditional analogue of CP, and absolute probability can then be defined in the following way:

$$\text{(AP)} \quad P(A) \overset{\text{df}}{=} P(A, \Omega)$$

Popper, in many places, gives alternative axiomatizations of probability where conditional probability is the primitive notion (see, e.g., Popper [1938], Popper [1955], Popper [1959a], and Popper [1959b]). The following set of ax-

[11] $X \times Y$ is the *Cartesian product* of the sets X and Y. So for example, if $X = \{x_1, x_2, x_3\}$ and $Y = \{y_1, y_2\}$, then $X \times Y = \{(x_1, y_1), (x_1, y_2), (x_2, y_1), (x_2, y_2), (x_3, y_1), (x_3, y_2)\}$.

[12] And any quadruple, $(\Omega, \mathscr{A}, \mathscr{B}, P)$, that satisfies the above conditions is called a *R'enyi conditional probability space*.

ioms are a user-friendly version of Popper's axioms (these are adapted from Roeper and Leblanc [1999], p. 12):

(PCP1) $P(A, B) \geq 0$

(PCP2) $P(A, A) = 1$

(PCP3) $P(A, B) + P(\neg A, B) = 1$ if B is P–normal

(PCP4) $P(A \wedge B, C) = P(A, B \wedge C) \cdot P(B, C)$

(PCP5) $P(A \wedge B, C) \leq P(B \wedge A, C)$

(PCP6) $P(A, B \wedge C) \leq P(A, C \wedge B)$

(PCP7) There is a D in $\mathcal{O}$ such that D is P–normal

for every A, B, and C in $\mathcal{O}$. $\mathcal{O}$ is the set of "objects" of probability—they could be sentences, events, states of affairs, propositions, etc. An "object", B, is P–abnormal if and only if $P(A, B) = 1$ for every A in $\mathcal{O}$, and it is P–normal if and only if it is not P–abnormal. P–abnormality plays the same role as logical falsehood. Any function, P, that satisfies the above axioms is known as a *Popper conditional probability function*, or often just as a *Popper function*, for short. This axiom system differs from Rényi's in that: (i) it is *symmetric* (i.e., if $P(A, B)$ exists, then $P(B, A)$ exists) and (ii) it is *autonomous*. An axiom system is autonomous if, in that system, probability conclusions can be derived only from probability premises. For example, the axiom system P1–3 is not autonomous, because, for instance, we can derive that $P(A) = 1$, from the premise that A is a logical truth.

2.3. Other Formal Theories of Probability

We have just seen what may be the four most prominent formal theories of probability. But there are many other theories also on the market—too many to go into their full details here, so I will merely give a brief overview of the range of possibilities.

Typically it is assumed that the logic of the language that probabilities are defined over is classical logic. However, there are probability theories that are based on other logics. Brian Weatherson, for instance, introduces an intuitionistic theory of probability (Weatherson [2003]). He argues that this probability theory, used as a theory of rational credences, is the best way to meet certain objections to Bayesianism (see §3.5.2). The defining feature of this formal account of probability is that it allows an agent to have credences in A and $\neg A$ that do not sum to 1, but are still additive. This can be done because it is not a theorem in this formal theory of probability that $P(A \vee \neg A) = 1$.

Another example of a "non-classical" probability theory is quantum probability. Quantum probability is based on a non-distributive logic, so it is not a theorem that $P((A \wedge B) \vee C)) = P((A \vee C) \wedge (B \vee C))$. Hilary Putnam uses this fact to argue that such a logic and probability makes quantum mechanics less mysterious than it is when classical logic and probability theory are used (Putnam [1968]). One of his examples is how the incorrect classical probability result for the famous two–slit experiment does not go through in quantum probability (see Putnam [1968] for more details and examples). See Dickson [2001]—who argues that quantum logic (and probability) is still a live option for making sense of quantum mechanics—for more details and references.

As we will see in §3.5, the probability calculus is often taken to be set of rationality constraints on the credences (degrees of belief) of an individual. A consequence of this—that many philosophers find unappealing—is that an individual, to be rational, should be logically omniscient. Ian Hacking introduces a set of probability axioms that relax the demand that an agent be logically omniscient (Hacking [1967]).

Other formal theories of probability vary from probability values being *negative* numbers (see, e.g., Feynman [1987]), to *imaginary* numbers (see, e.g., Cox [1955]), to *unbounded* real numbers (see, e.g., Rényi [1970]), to real numbered *intervals* (see, e.g, Levi [1980]). Dempster-Shafer Theory is also often said to be a competing formal theory of probability (see, e.g., Shafer [1976]). For more discussion of other formal theories of probability see Fine [1973].

That concludes our survey of the various formal theories of probability. So far though, we have only half of the picture. We do not yet have any account of what probabilities *are*, only how they *behave*. This is important because there are many things in the world that behave like probabilities, but are not probabilities. Take for example, areas of various regions of a tabletop where one unit of area is the entire area of the tabletop. The areas of such regions satisfy, for example, Kolmogorov's axioms, but are clearly not probabilities.

3. The Philosophical Theory of Probability

Interpretations of probability are typically categorized into two kinds: subjective interpretations and objective interpretations. Roughly, the difference is that subjective interpretations identify probabilities with the credences, or "degrees of belief" of a particular individual, while objective interpretations identify probability with something that is independent of any individual—the most common somethings being relative frequencies and propensities. The following is a brief survey of some of the interpretations of probability that philosophers have proposed. It is impossible to give a full and just discussion of each interpretation in the space available, so only a small selection of issues surrounding each will be discussed.

3.1. The Classical Interpretation

The central idea behind the classical interpretation of probability—historically the first of all the interpretations—is that the probability of an event is the ratio between the number of equally possible outcomes in which the event occurs and the total number of equally possible outcomes. This conception of probability is particularly well suited for probability statements concerning games of chance. Take, for example, a fair roll of a fair die. We quite naturally say that the probability of an even number coming up is 3/6 (which, of course, is equal to 1/2). The "3" is for the three ways in which an even number comes up (2, 4, and 6) and the "6" is for all of the possible numbers that could come up (1, 2, 3, 4, 5, and 6).

The idea of relating probabilities to equally possible outcomes can be found in the works of many great authors—e.g., Cardano [1663], Laplace [1814] and Keynes [1921]. However, among these authors there is a considerable degree of variation in how this idea is fleshed out. In particular, they vary on how we are to understand what it means for events to be “equally possible”. In the hands of some, the equally possible outcomes are those outcomes that are *symmetric* in some physical way. For example, the possible outcomes of a fair roll of a fair die might be said to be all equally possible because of the physical symmetries of the die and in the way the die is rolled. If we understand “equally possible” this way, then the classical interpretation is an *objective* interpretation. However, the most canonical understanding of the term “equally possible” is in terms of our knowledge (or lack thereof). Laplace is a famous proponent of this understanding of “equally possible”:

> The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as *we* may be equally *undecided* about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought. (my emphasis) Laplace [1814], p. 6

Understood this way, the classical interpretation is a *subjective* interpretation of probability. From now on, I will assume that the classical interpretation is a subjective interpretation, as this is the most popular understanding of the interpretation—see Hacking [1971] for a historical study of the notion of equal possibilities and the ambiguities in the classical interpretation.

If we follow Laplace, then the classical interpretation puts constraints on how we ought to assign probabilities to events. More specifically, it says we ought to assign equal probability to events that we are “equally undecided about”. This norm was formulated as a principle now known as the Principle of Indifference by John Maynard Keynes:

> If there is no known reason for predicating of our subject one rather than another of several alternatives, then relative to such knowledge the assertions of each of these alternatives have an equal probability. Keynes [1921], p. 42

It is well known that the Principle of Indifference is fraught with paradoxes—many of which originate with Joseph Bertrand (Bertrand [1888]). Some of these paradoxes are rather mathematically complicated, but the following is a simple one due to Bas van Fraassen (van Fraassen [1989]). Consider a factory that produces cubic boxes with edge lengths anywhere between (but not including) 0 and 1 meter, and consider two possible events: (a) the next box has an edge length between 0 and 1/2 meters or (b) it has an edge length between 1/2 and 1 meters. Given these considerations, there is no reason to think either (a) or (b) is more likely than the other, so by the Principle of Indifference we ought to assign them equal probability: 1/2 each. Now consider the following four events: (i) the next box has a face area between 0 and 1/4 square meters; (ii) it has a face area between 1/4 and 1/2 square meters; (iii) it has a face area between 1/2 and 3/4 square meters; or (iv) it has a face area between 3/4 and 1 square meters. It seems we have no reason to suppose any of these four events to be more probable than any other, so by the Principle of Indifference we ought to assign them all equal probability: 1/4 each. But this is in conflict with our earlier assignment, for (a) and (i) are different descriptions of the same event (a length of 1/2 meters corresponds to an area of 1/4 square meters). So the probability assignment that the Principle of Indifference tells us to assign depends on how we describe the box factory: we get one assignment for the “side length” description, and another for the “face area” description.

There have been several attempts to save the classical interpretation and the Principle of Indifference from paradoxes like the one above, but many authors consider the paradoxes to be decisive. See Keynes [1921] and van Fraassen [1989] for a detailed discussion of the various paradoxes, and see Jaynes [1973], Marinoff [1994], and Mikkelson [2004] for a defense of the principle. Also see Shackel [2007] for a contemporary overview of the debate. The existence of paradoxes like the one above were one source of motivation for many authors to abandon the classical interpretation and adopt the frequency interpretation of probability.

3.2. The Frequency Interpretation

3.2.1. Actual Frequencies

Ask any random scientist or mathematician what the definition of probability is and they will probably respond to you with an incredulous stare or, after they have regained their composure, with some version of the frequency interpretation. The frequency interpretation says that the probability of an outcome is the number of experiments in which the outcome occurs divided by the number of experiments performed (where the notion of an "experiment" is understood very broadly). This interpretation has the advantage that it makes probability empirically respectable, for it is very easy to measure probabilities: we just go out into the world and measure frequencies. For example, to say that the probability of an even number coming up on a fair roll of a fair die is 1/2 just means that out of all the fair rolls of that die, 50% of them were rolls in which an even number came up. Or to say that there is a 1/100 chance that John Smith, a consumptive Englishman aged fifty, will live to sixty-one is to say that out of all the people like John, 1% of them live to the age of sixty-one.

But which people are like John? If we consider all those Englishman aged fifty, then we will include consumptive Englishman aged fifty and all the healthy ones too. Intuitively, the fact that John is sickly should mean we only consider consumptive Englishman aged fifty, but where do we draw the line? Should we restrict the class of those people we consider to those who are also named John? Surely not, but is there a principled way to draw the line? If there is, it is hard to say exactly what that principle is. This is important because where we draw the line affects the value of the probability. This problem is known as the *reference class problem*. John Venn was one of the first to notice it:

> It is obvious that every individual thing or event has an indefinite number of properties or attributes observable in it, and might therefore be considered as belonging to an indefinite number of different classes of things [...]. Venn [1876], p. 194

This can have quite serious consequences when we use probability in our decision making (see, e.g., Colyvan et al. [2001]). Many have taken the reference class problem to be a difficulty for the frequency interpretation, though Mark Colyvan et al. (Colyvan et al. [2001]) and Hájek (Hájek [2007c]) point out that it is also a difficulty for many other interpretations of probability.

The frequency interpretation is like the classical interpretation in that it identifies the probability of an event with the ratio of favorable cases to cases. However, it is unlike the classical interpretation in that the cases have to be *actual* cases. Unfortunately, this means that the interpretation is shackled too tightly to how the world turns out to be. If it just happens that I never flip this coin, then the probability of "tails" is undefined. Or if it is flipped only once and it lands "tails", then the probability of "heads" is 1 (this is known as a single case probability). To get around these difficulties many move from defining probability in terms of actual frequencies to defining it in terms of *hypothetical* frequencies. There are many other problems with defining probability in terms of actual frequencies (see Hájek [1996] for fifteen objections to the idea), but we now move on to hypothetical frequencies.

3.2.2. Hypothetical Frequencies

The hypothetical frequency interpretation tries to put some of the modality back into probability. It says that the probability of an event is the number of trials in which the event occurs divided by the number of trials, *if the trials were to occur*. On this frequency interpretation, the trials do not have to actually happen for the probability to be defined. So for the coin that I never flipped, the hypothetical frequentist can say that the probability of "tails" is 1/2 because this is the frequency we would observe, *if* the coin were tossed.

Maybe. But we definitely would not observe this frequency if the coin were flipped an odd number of times, for then it would be impossible to observe an even number of "heads" and "tails" events. To get around this sort of problem, it is typically assumed that the number of trials is countably infinite, so the frequency is a *limiting* frequency. Defenders of this type of view include Richard von Mises (von Mises [1957]) and Hans Reichenbach (Reichenbach [1949]). Consider the following sequence of outcomes of a series of fair coin flips:

$$THTTHTHHT\ldots$$

where T is for "tails" and H is for "heads". We calculate the limiting frequency by calculating the frequencies of successively increasing finite subsequences. So for example, the first subsequence is just T, so the frequency of "tails" is 1. The next larger subsequence is TH, which gives a frequency of 1/2. Then the next subsequence is THT, so the frequency becomes 2/3. Continuing on in this fashion:

$$\begin{array}{cc} THTT & 3/4 \\ THTTH & 3/5 \\ THTTHT & 4/6 \\ THTTHTH & 4/7 \\ THTTHTHH & 4/8 \\ THTTHTHHT & 5/9 \\ \vdots & \vdots \end{array}$$

These frequencies appear to be settling down to the value of 1/2. If this is the case, we say that the limiting frequency is 1/2. However, the value of the limiting frequency depends on how we order the trials. If we change the order of the trials, then we change the limiting frequency. To take a simple example, consider the following sequence of natural numbers: $(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, \ldots)$. The limiting frequency of even numbers is 1/2. Now consider a different sequence that also has all of the natural numbers as elements, but in a different order: $(1, 3, 5, 2, 7, 9, 11, 4, \ldots)$. Now the limiting frequency of even numbers is 1/4. This means that the value of a limiting frequency is sensitive to how we order the trials, and so if probabilities are limiting frequencies, then probabilities depend on the order of the trials too. This is problematic because it seems probabilities should be independent of how we order the trials to calculate limiting frequencies.[13]

Another worry with the hypothetical frequency interpretation is that it does not allow limiting frequencies to come apart from probabilities. Suppose a coin, whenever flipped, has a chance of 1/2 that "tails" comes up on any particular flip. Although highly improbable, it is entirely possible that "tails" never comes up. Yet the hypothetical frequency interpretation says that this statement of fifty percent chance of "tails" *means* that the limiting frequency of "tails" *will* be 1/2. So a chance of 1/2 just means that "tails" has to come up at least once (in fact, half of the time). Many philosophers find this unappealing, for it seems that it is part of the concept of probability that frequencies (both finite and limiting frequencies) can come apart from probabilities.

One of the motivations for the move from the actual frequency interpretation to the hypothetical frequency interpretation was the problem of single–case probabilities. This was the problem that the actual frequency interpretation cannot sensibly assign probabilities to one–time–only events. This problem was also a main motivation for another interpretation of probability, the propensity interpretation.

3.3. The Propensity Interpretation

The propensity interpretation of probability originates with Popper in Popper [1957], and was developed in more detail in Popper [1959b]. His motivation for introducing this new interpretation was the need, that he saw, for a theory of probability that was objective, but that could also make sense of single case probabilities—particularly the single case probabilities which he thought were indispensable to quantum mechanics. His idea was (roughly) that a probability is not a frequency, but rather it is the tendency, the disposition, or the *propensity* of an outcome to occur.

Popper, who was originally a hypothetical frequentist, developed the propensity theory of probability as a slight modification of the frequency theory. The modification was that instead of probabilities being properties of sequences (*viz.*, frequencies), they are rather properties of the conditions that generate those sequences, when the conditions are repeated:

> This modification of the frequency interpretation leads almost inevitably to the conjecture that probabilities are dispositional properties of these conditions—that is to say, propensities. Popper [1959b], p. 37

And earlier:

> Now these propensities turn out to be *propensities to realize singular events.* (emphasis in original) Popper [1959b], p. 28

Perhaps the best known and most influential objection to Popper's original propensity interpretation is due to Paul Humpreys, and is known as Humphreys'

[13] This problem is similar to the problem in decision theory where the expected utility of an action can depend on how we order the terms in the expected utility calculation. See Nover and Hájek [2004] for further discussion.

paradox—though Humphreys himself did not intend the objection to be one against the propensity interpretation (Humphreys [1985]). The objection, in a nutshell, is that propensities are not symmetric, but according to the standard formal theory of probability, probabilities are.[14] For example, it is often possible to work out the probability of a fire having been started by a cigarette given the smoking remains of a building, but it seems strange to say that the smoking remains has a *propensity*, or *disposition* for a cigarette to have stated the fire. The standard reaction to this fact has been "if probabilities are symmetric and propensities are not, then too bad for the propensity interpretation". Humphreys, however, intended his point to be an objection to the standard formal theory of probability (Humphreys [1985], p. 557) and to the whole enterprise of interpreting probability in a way that takes the formal theory of probability as sacrosanct:

> It is time, I believe, to give up the criterion of admissibility [the criterion that a philosophical theory of probability should satisfy "the" probability calculus]. We have seen that it places an unreasonable demand upon one plausible construal of propensities. Add to this the facts that limiting relative frequencies violate the axiom of countable additivity and that their probability spaces are not sigma-fields unless further constraints are added; that rational degrees of belief, according to some accounts, are not and cannot sensibly be required to be countably additive; and that there is serious doubt as to whether the traditional theory of probability is the correct account for use in quantum theory. Then the project of constraining semantics by syntax begins to look quite implausible in this area. Humphreys [1985], pp. 569–70

In response to Humphreys' paradox, some authors have offered new formal accounts of propensities. For example, James Fetzer and Donald Nute developed a probabilistic causal calculus as a formal theory of propensities (see Fetzer [1981]). A premise of the argument that leads to the paradox is that probabilities are symmetric. But as we saw in §2.2, there are formal theories of probability that are *asymmetric*—Rényi's axioms for conditional probability, for instance. A proponent of Popper's propensity interpretation could thus avoid the paradox by adopting an asymmetric formal theory of probability. Unfortunately for Popper though, his own formal theory of probability is symmetric.

There are now many so–called propensity interpretations of probability that differ from Popper's original account. Following Donald Gillies, we can divide these accounts into two kinds: long–run propensity interpretations and single–case propensity interpretations (Gillies [2000b]). Long–run propensity interpretations treat propensities as tendencies for certain conditions to produce frequencies identical (at least approximately) to the probabilities in a sequence of repetitions of those conditions. Single–case propensity interpretations treat propensities as dispositions to produce a certain result on a specific occasion. The propensity interpretation initially developed by Popper (Popper [1957, 1959b]) is both a long–run and single case propensity interpretation. This is because Popper associates propensities with repeatable "generating conditions" to generate singular events. The propensity interpretations developed later by Popper (Popper [1990]), and David Miller (Miller [1994, 1996]) can be seen as only single–case propensity interpretations. These propensity interpretations attribute propensities not to repeatable conditions, but to entire states of the universe. One problem with this kind of propensity interpretation is

[14]Remember from §2.2, a formal theory of probability is symmetric if whenever $P(A,B)$ is defined, $P(B,A)$ is also defined.

that probability claims are no longer testable (a cost noted by Popper himself, Popper [1990], p. 17). This is because probabilities are now properties of entire states of the universe—events that are not repeatable—and Popper believed that to test a probability claim, the event needs to be repeatable so that a frequency can be measured.[15]

For a general survey and classification of the various propensity theories see Gillies [2000b], and see Eagle [2004] for twenty–one objections to them.

3.4. Logical Probability

In classical logic, if $A \vdash B$, then we say A entails B. In model theoretic terms, this corresponds to every model in which A is true, B is true. The logical interpretation of probability is an attempt to generalize the notion of entailment to *partial* entailment. Keynes was one of the earliest to hit upon this idea:

> Inasmuch as it is always assumed that we can sometimes judge directly that a conclusion *follows from* a premiss, it is no great extension of this assumption to suppose that we can sometimes recognize that a conclusion *partially follows from*, or stands in a relation of probability to a premiss. Keynes [1921], p. 52

On this interpretation "$P(B, A) = x$" means A entails B to degree x. This idea has been pursued by many philosophers—e.g., William Johnson (Johnson [1921]), Keynes (Keynes [1921]), though Rudolf Carnap gives by the far the most developed account of logical probability (e.g., Carnap [1950]).

By generalizing the notion of entailment to partial entailment, some of these philosophers hoped that the logic of *deduction* could be generalized to a logic of *induction*. If we let c be a two–place function that represents the confirmation relation, then the hope was that:

$$c(B, A) = P(B, A)$$

For example, the observation of ten black ravens deductively entails that there are ten black ravens in the world, while the observation of five black ravens only partially entails, or *confirms* that there are ten black ravens, and the observation of two black ravens confirms this hypothesis to a lesser degree.

One seemingly natural way to formalize the notion of partial entailment is by generalizing the model theory of full entailment. Instead of B being true in *every* model in which A is true, we relax this to there being some percentage of the models in which A is true. So "$P(B, A) = x$", which is to say, "A partially entails B to degree x" is true, if the number of models where B and A are true, divided by the number of models where A is true is equal to x.[16] If we think of models as like "possible worlds", or possible outcomes then this definition is the same as the classical definition of probability. We might suspect then that the logical interpretation shares some of the same difficulties (in particular, the language relativity of probability) that the classical interpretation has. Indeed, this is so (see e.g., Gillies [2000a], pp. 29–49).

[15] This is, perhaps, not the only way in which a probability claim can be tested. For example, it may be possible to test the claim "this coin has a chance 0.5 to land heads when flipped" by investigating whether or not the coin is physically symmetrical.

[16] Unconditional, or absolute probability, $P(A)$, is understood as the probability of A given a tautology T, so $P(A) = P(A, T)$ in which case $P(A)$ is just the number of models where A is true divided by the total number of models (since a tautology is true in every model).

Carnap maintains that $c(B,A) = P(B,A)$, but investigates other ways to define the probability function, P. In contrast to the approach above, Carnap's way of defining P is purely syntactic. He starts with a language with predicates and constants, and from this language defines what are called *state descriptions*. A state description can be thought of as a maximally specific description of the world. For example, in a language with predicates F and G, and constants a and b, one state description is $Fa \wedge Fb \wedge Gb \wedge \neg Ga$. Any state description is equivalent to a conjunction of predications where every predicate or its negation is applied to every constant in the language. Carnap then tried to define the probability function, P, in terms of some measure, m, over all of the state descriptions. In Carnap [1950], he thought that such a measure was unique. Later on, in Carnap [1963], he thought there were many such measures. Unfortunately, every way Carnap tried to define P in terms of a measure over state descriptions failed for one reason or another (see e.g., Hájek [2007b]).

Nearly every philosopher now agrees that the logical interpretation of probability is fundamentally flawed. However, if they are correct, this does not entail that a formal account of inductive inference is not possible. Recent attempts at developing an account of inductive logic reject the sole use of conditional probability and instead measure the degree to which evidence E confirms hypothesis H by how much E affects the probability of H (see e.g., Fitelson [2006]). For example, one way to formalize the degree to which E supports or confirms H is by how much E *raises* the probability of H:

$$c(H,E) = P(H,E) - P(H)$$

This is one such measure among many.[17] The function c, or some other function like it, may formally capture the notion of evidential impact that we have, but these functions are defined in terms of *probabilities*. So an important and natural question to ask is: what are these probabilities? Perhaps the most popular response is that these probabilities are subjective probabilities, i.e., the credences of an individual. According to this type of theory of confirmation (known as Bayesian confirmation theory), the degree to which some evidence confirms a hypothesis is relative to the epistemic state of an individual. So E may confirm H for one individual, but disconfirm H for another. This moves us away from the strictly objective relationship between evidence and hypothesis that the logical interpretation postulated, to a more subjective one.

3.5. The Subjective Interpretation

While the frequency and propensity interpretations see the various formal accounts of probability as theories of how frequencies and propensities behave, the subjective interpretation sees them as theories of how people's beliefs *ought* to behave. We can find this idea first published in English by Frank Ramsey (Ramsey [1931]) and de Finetti (de Finetti [1931a,b]). The normativity of the "ought" is meant to be one of ideal epistemic rationality. So subjectivists traditionally claim that for one to be ideally epistemically rational, one's beliefs must conform to the standard probability calculus. Despite the intuitive appeal of this claim (which by the way is typically called *probabilism*), many have felt the need to provide some type of argument for it. Indeed, there is now a

[17] See Eells and Fitelson [2002] for an overview of some of the other possible measures.

formidable literature on such arguments. Perhaps the most famous argument for probabilism is the Dutch Book Argument.

3.5.1. The Dutch Book Argument

A *Dutch Book* is any collection of bets that collectively guarantee a sure monetary loss. An example will help illustrate the idea. Suppose Bob assigns a credence of 0.6 to a statement, A, and a credence of 0.5 to that statement's negation, $\neg A$. Bob's credences thus do not satisfy the probability calculus since his credence in A and his credence in $\neg A$ sum to 1.1. Suppose further that Bob bets in accordance with his credences, that is, if he assigns a credence of x to A, then he will buy a bet that pays \$$y$ if A, for at most \$$xy$. Now consider the following two bets:

- Bet 1: This bet costs \$0.6 and pays \$1 if A is true.
- Bet 2: This bet costs \$0.5 and pays \$1 if $\neg A$ is true.

Bob evaluates both of these bets as fair, since the expected return—by his lights—of each bet is the price of that bet.[18] But suppose Bob bought both of these bets. This would be apparently equivalent to him buying the following bet:

- Bet 3: This bet costs \$0.6 + \$0.5 = \$1.1 and pays \$1 if A is true, and \$1 if $\neg A$ is true (i.e., the bet pays \$1 no matter what).

If Bob were to accept Bet 3, then Bob would be guaranteed to lose \$0.1, no matter what. The problem for Bob is that he evaluates Bet 1 and Bet 2 as both individually fair, but by purchasing both Bet 1 and Bet 2, Bob effectively buys Bet 3, which he does not evaluate as fair (since his expected return of the bet is less than the price of the bet).

There is a theorem called The Dutch Book Theorem which, when read informally, says that if an agent has credences like Bob's—i.e., credences that do not obey axioms P1–3—then there is always a Dutch Book that the agent would be willing to buy. So having credences that do not obey axioms P1–3, results in you being susceptible to a Dutch Book. Conversely, there is a theorem called the The Converse Dutch Book Theorem which, when also read informally, says that if an agent has credences that *do* obey P1–3, then there is no Dutch Book that that agent would be willing to buy. Taken together these two theorems give us:

> (DBT & CDBT) An agent is susceptible to a Dutch Book if and only if the agent has credences that violate axioms P1–3.

Then with the following rationality principle:

> (RP) If an agent is ideally epistemically rational, then that agent is not susceptible to a Dutch Book.

we get the following result:

[18] To work out the expected return of a bet we multiply the probability of each pay-off by the value of that pay-off and sum these numbers together. For example, in Bet 1 there is only one pay-off—\$1, when A is true—so we multiply that by Bob's credence in A, 0.6, so the expected return is \$0.6.

(C) If an agent is ideally epistemically rational, then that agent's credences obey axioms P1–3.

This is known as the Dutch Book Argument. It is important that CDBT is included, because it blocks an obvious challenge to RP. Without CDBT one might claim that it is *impossible* to avoid being susceptible to a Dutch book, but it is still possible to be ideally epistemically rational. CDBT guarantees that it is possible to avoid a Dutch book, and combined with DBT it entails that the only way to do this is to have one's credences satisfy the axioms P1–3.

There are many criticisms of the Dutch Book Argument—too many to list all of them here, but I will mention a few.[19] One criticism is that it is not clear that the notion of rationality at issue is of the right kind. For example, David Christensen writes:

> Suppose, for example, that those who violated the probability calculus were regularly detected and tortured by the Bayesian Thought Police. In such circumstances, it might well be argued that violating the probability calculus was imprudent, or even "irrational" in a practical sense. But I do not think that this would do anything toward showing that probabilistic consistency was a component of rationality in the epistemic sense relevant here." Christensen [1991], p. 238

In response to this worry, some have offered what are called depragmatized dutch book arguments, in support of probabilism (see e.g., Christensen [1996]). Others have stressed that the Dutch Book Argument should not be interpreted literally and rather that it merely dramatizes the inconsistency of a system of beliefs that do not obey the probability calculus (e.g., Skyrms [1984], p. 22 and Armendt [1993], p. 3).

Other criticisms focus on the assumptions of the Dutch Book and Converse Dutch Book Theorems. For instance, the proofs of these theorems assume that if an agent evaluates two bets as both fair when taken individually, then that agent will, and should, also consider them to be fair when taken collectively. This assumption is known as the package principle (see Schick [1986] and Maher [1993] for criticisms of this principle). The standard Dutch Book Argument is meant to establish that our credences ought to satisfy axioms P1–3, but what about a countable additivity axiom? Dutch book arguments that try to establish a countable additivity axiom as a rationality constraint rely on a countably infinite version of the package principle (see Arntzenius et al. [2004] for objections to this principle).

These objections to the Dutch Book Argument—and others—have led some authors to search for other arguments for probabilism. For instance, Patrick Maher argues that if you cannot be *represented* as an expected utility maximizer, relative to a probability and utility function, then you are irrational (Maher [1993]). Some have argued that one's credences ought to obey the probability calculus because for any non-probability function, there is a probability function that better matches the relative frequencies in the world, no matter how the world turns out. This is known as a calibration argument (see e.g., van Fraassen [1984]). James Joyce argues for probabilism by proving that for any non-probability function, there is a probability function that is "closer" to the truth, no matter how the world turns out (Joyce [1998]). This is known as a gradational accuracy argument. For criticisms of all these arguments see Hájek [forthcoming].

[19] For a detailed discussion of these and other objections see Hájek [2007a].

Suppose for the moment that it has been established that one's credences ought to satisfy the probability axioms. Are these the *only* normative constraints on credences? One feature our beliefs have is that they *change* over time, especially when we learn new facts about the world. And it seems that there are rational and irrational ways of changing one's beliefs. In fact, perhaps most probabilists believe that there are rational and irrational ways to respond to evidence, beyond simply remaining in synch with the probability calculus. One particularly large subgroup of these probabilists are known as Bayesians.

3.5.2. Bayesianism

Orthodox Bayesianism is the view that an agent's credences: should at all times obey the probability axioms; should change only when the agent acquires new information; and, in such cases, the agent's credences should be updated by Bayesian Conditionalisation. Suppose that an agent has a prior credence function Cr_{old}. Then, according to this theory of updating, the agent's posterior credence function, Cr_{new}, after acquiring evidence E ought to be:

$$\text{(BC)} \quad Cr_{\text{new}}(H) = Cr_{\text{old}}(H, E)$$

for every H in $\mathcal{L}$. BC is said to be a *diachronic* constraint on credences, whereas for example, P1–3 are said to be *synchronic* constraints on credences. There are Dutch Book Arguments for why credences ought to be diachronically constrained by Bayesian Conditionalization (see e.g., Lewis [1999]). Arguments of this type suppose that BC is a *rationality* constraint, and that violations of it are a type of inconsistency. Christensen argues that since the beliefs are changing across time, violations of BC are not, strictly speaking, inconsistencies (Christensen [1991]).

One important criticism of orthodox Bayesianism, due to Richard Jeffrey, is that it assumes that facts are always acquired (learned) with full certainty (Jeffrey [1983]). Critics argue that it should at least be *possible* for evidence that you are not entirely certain of to impact your credences. For this reason, Jeffrey developed an alternative account for how credences should be updated in the light of new evidence, which generalized BC to account for cases when we acquire evidence without full certainty. Jeffrey called his theory *Probability Kinematics*, but it is now known as *Jeffrey Conditionalization*. According to Jeffrey Conditionalization, an agent's new credence in H after acquiring some information that has affected the agent's credence in E should be:

$$\text{(JC)} \quad Cr_{\text{new}}(H) = Cr_{\text{old}}(H, E)Cr_{\text{new}}(E) + Cr_{\text{old}}(H, \neg E)Cr_{\text{new}}(\neg E)$$

for every H in $\mathcal{L}_{\mathcal{PC}}$.[20] Notice that the right-hand side of JC contains both the old and new credence function, whereas BC only had the old credence function. At first glance this may give the impression that JC is circular. It is not though. Initially you have a probability in E and $\neg E$, $Cr_{\text{old}}(E)$ and $Cr_{\text{old}}(\neg E)$, respectively. Then you acquire some information that causes you to change your credences concerning E and $\neg E$ (and only these statements). These new credences are

[20] Actually, this is a special case of Jeffrey Conditionalisation. The general equation is: $Cr_{\text{new}}(H) = \sum_i Cr_{\text{old}}(H, E_i)Cr_{\text{new}}(E_i)$, where the E_i are mutually exclusive and exhaustive in $\mathcal{L}$.

$Cr_{\text{new}}(E)$ and $Cr_{\text{new}}(\neg E)$, respectively. JC then tells you how the information you acquired should affect your other credences, given how that information affected your credences concerning E and $\neg E$.

So, Bayesianism is a theory of epistemic rationality that says our credences at any given time should obey the probability calculus and should be updated by conditionalization (either BC or JC). However, some insist that there is still more to a full theory of epistemic rationality.

3.5.3. Objective and Subjective Bayesianism

Within the group of those probabilists who call themselves Bayesians, is another division between so–called objective Bayesians and subjective Bayesians. As we saw in the previous sections, Bayesians believe that credences should obey the probability calculus and should be updated according to conditionalisation, when new information is obtained. So far though, nothing has been said about which credence function one should have *before* any information is obtained—apart from the fact that it should obey the probability calculus.

Subjective Bayesians believe there ought to be no further constraint on initial credences. They say: given that it satisfies the probability calculus, no initial credence function is anymore rational than any other. But if subjective Bayesians believe any coherent initial credence function is a rational one, then according to them, a credence function that assigns only 1s and 0s to all statements—including statements that express contingent propositions—is also a rational credence function. Many philosophers (including those that call themselves subjective Bayesians) balk at this idea and so insist that any initial credence function must be *regular*. A regular credence function is any probability function that assigns 1s and 0s only to logical truths and falsehoods; all contingent sentences must be assigned strictly intermediate probability values.[21] The idea roughly is that an initial credence function should not assume the truth of any contingency, since nothing contingent about the world is known by the agent.

However, we may worry that this is still not enough, for a credence function that assigns a credence of, say, 0.9999999 to some contingent sentence (e.g., that the Earth is flat) is still counted as a rational initial credence function. There are two responses that Bayesians make here. The first is to point to so–called Bayesian convergence results. The idea, roughly, is that as more and more evidence comes in, such peculiarities in the initial credence function are in a sense "washed out" through the process of repeated applications of conditionalization. More formally, for any initial credence function, there is an amount of possible evidence that can be conditionalized on to ensure the resulting credence function is arbitrarily close to the truth. See Earman [1992] (pp. 141–149) for a more rigorous and critical discussion of the various Bayesian convergence results.

The second response to the original worry that some Bayesians make is that there are in fact further constraints on rational initial credence functions. Bayesians who make this response are known as objective Bayesians. One worry

[21] There are some technical difficulties with this condition of regularity because it is impossible (in some formal accounts of probability) for there to be a regular probability function over uncountably many sentences (or sets, propositions, or whatever the bearers of probability are). See e.g., Hájek [2003] for discussion.

with the prior that assigned 1s and 0s to contingent statements was that such a prior does not truly reflect our epistemic state—we do not know anything about any contingent proposition before we have learned anything, yet our credence function says we do. A similar worry may be had about the prior that assigns 0.9999999 to a contingent statement. This type of prior reports an overwhelming confidence in contingent statements before anything about the world is known. Surely such blind confidence cannot be rational. Reasoning along these lines, E. T. Jaynes, perhaps the most famous proponent of objective Bayesianism, claims that our initial credence function should be an accurate description of how much information we have:

> [A]n ancient principle of wisdom—that one ought to acknowledge frankly the full extent of his ignorance—tells us that the distribution that maximizes H subject to constraints which represent whatever information we have, provides the most honest description of what we know. The probability is, by this process, "spread out" as widely as possible without contradicting the available information. Jaynes [1967], p. 97

The quantity H is from information theory and is known as the Shannon entropy.[22] Roughly speaking, H measures the information content of a distribution. According to this view, in the case where we have no information at all, the distribution that provides the most honest description of our epistemic state is the uniform distribution. We see then that the principle that Jaynes advocates—which is known as the Principle of Maximum Entropy—is a generalization of the Principle of Indifference. This version of objective Bayesianism thus faces problems similar to those that plague the logical and classical interpretations. Most versions of objective Bayesianism ultimately rely on some version of the Principle of Indifference and so suffer a similar fate. As a result, subjective Bayesianism with the condition that a prior should be regular is perhaps the most popular type of Bayesianism amongst philosophers.

3.5.4. Other Norms

At this stage, we have the following orthodox norms on partial beliefs:

1. One's credence function must always satisfy the standard probability calculus.
2. One's credence function must only change in accordance with conditionalisation.
3. One's intial credence function must be regular.

However, there are still more norms that are often said to apply to beliefs. One important such norm is David Lewis' Principal Principle (LPP) (Lewis [1980]). Roughly, the idea behind this principle is that one's credences should be in line with any of the objective probabilities in the world, if they are known. More formally, if $Ch_t(A)$ is the chance of A at time t (e.g., on a propensity interpretation of probability this would be the propensity at time t of A to obtain), then:

$$\text{(LPP)} \quad Cr(A, Ch_t(A) = x \wedge E) = x$$

[22] To learn more about information theory and Shannon entropy, see Shannon and Weaver [1962].

where E is any proposition, so long as it is not relevant to A.[23] LPP, as originally formulated by Lewis, is a synchronic norm on an agent's *initial* credence function, though LPP is commonly used as synchronic constraint on an agent's credence function at any point in time.

Another synchronic norm is van Fraassen's Reflection Principle (VFRP) (van Fraassen [1995]). Roughly, the idea behind this principle is that if, upon reflection, you realize that you will come to have a certain belief, then you ought to have that belief *now*. More formally, the Reflection Principle is:

$$\text{(VFRP)} \quad Cr_{t_1}(A, Cr_{t_2}(A) = x) = x$$

where $t_2 > t_1$ in time.

Another more controversial norm is Adam Elga's principle of indifference for indexical statements, used to defend a particular solution to the Sleeping Beauty Problem. The problem is that Sleeping Beauty is told by scientists on Sunday that they are going to put her to sleep and flip a fair coin. If the coin lands "tails", they will wake her on Monday, wipe her memory, put her back to sleep, and wake her again on Tuesday. If the coin lands "heads", they will simply wake her on Monday. When Sleeping Beauty finds herself having just woken up, what should her credence be that the coin landed "heads"? According to Lewis, it should be 1/2 since this is the chance of the event and LPP says Sleeping Beauty's credence should be equal to the chance. According to Elga, there are three possibilities: (i) she is being woken for the first time, on Monday; (ii) she is being woken for the second time, on Tuesday; or (iii) she is being woken for the first time, on Monday. All of these situations are indistinguishable from Sleeping Beauty's point of view, and Elga argues that an agent should assign equal credence to indistinguishable situations—this is his indifference principle. So according to Elga, Sleeping Beauty should assign equal probability to each possibility, and so her credence that the coin landed "heads" ought to be 1/3. See Elga [2000] for more on the Sleeping Beauty Problem and Weatherson [2005] for criticism of Elga's version of the Principle of Indifference.

4. Conclusion

In a short amount of space we have covered a lot of territory in the philosophy of probability. In §2, we considered various formal theories of probability. We saw that not only are there rival theories to Kolmogorov's axioms, but these rivals arguably have desirable features that Kolmogorov's axioms lack. In §3, we saw some of the various interpretations of probability and some of the issues connected with each interpretation. The discussion of each interpretation was necessarily brief, but each of these interpretations suffers from one problem or another. In fact, the failures of each interpretation have motivated some to take probability as a primitive, undefined concept (e.g., Sober [forthcoming]). We see then, that despite the ubiquity of probability in our lives, the mathematical and philosophical foundations of this fruitful theory remain in contentious dispute.[24]

[23] Actually, strictly speaking E should be what Lewis calls *admissible*. See Lewis [1980] for details on this issue.

[24] Thanks to Alan Hájek for many discussions and feedback on earlier drafts, and to Jens Christian Bjerring, Mark Colyvan, Fritz Allhoff, John Matthewson, Joe Salerno, Mike Smithson, and

References

Armendt, B. (1993). Dutch Books, Additivity and Utility Theory. *Philosophical Topics*, 21(1):1–20.

Arntzenius, F., Elga, A., and Hawthorne, J. (2004). Bayesianism, Infinite Decisions, and Binding. *Mind*, 113(450):1–34.

Bertrand, J. (1888). *Calcul des Probabilités.* Paris: Gauthier-Villars.

Cardano, G. (1663). *The Book on Games of Chance.* English translation in Ore, 0. (1953). *Cardano: The Gambling Scholar*, (1965) Dover: New York.

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

Carnap, R. (1963). *Replies and Systematic Expositions.* Open Court, Illinois: La Salle.

Christensen, D. (1991). Clever Bookies and Coherent Beliefs. *The Philosophical Review*, 100(2):229–247.

Christensen, D. (1996). Dutch-Book Arguments Depragmatized: Epistemic Consistency for Partial Believers. *The Journal of Philosophy*, 93(9):pp. 450–479.

Colyvan, M., Regan, H. M., and Ferson, S. (2001). Is it a Crime to Belong to a Reference Class? *Journal of Political Philosophy*, 9(2):168–81.

Cox, D. R. (1955). A Use of Complex Probabilities in the Theory of Stochastic Processes. *Proceedings of the Cambdrigde Philosophical Society*, 51:313–319.

de Finetti, B. (1931a). On the Subjective Meaning of Probability. In *Induction and Probability*. Clueb: Bologna, 1993.

de Finetti, B. (1931b). Probabilism. *Erkenntnis*, 31:169–223.

de Finetti, B. (1974). *Theory of Probability*, volume 1. (1990) Wiley Classics Library, John Wiley and Sons.

Dickson, M. (2001). Quantum Logic is Alive $\wedge$ (It Is True $\vee$ It Is False). *Philosophy of Science*, 68(3):S274–S287.

Eagle, A. (2004). Twenty-One Arguments Against Propensity Analyses of Probability. *Erkenntnis*, 60:371–416.

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

Eells, E. and Fitelson, B. (2002). Symmetries and Asymmetries in Evidential Support. *Philosophical Studies*, 107(2):129–142.

Elga, A. (2000). Self-Locating Belief and the Sleeping Beauty Problem. *Analysis*, 60(2):143–147.

Fetzer, H. (1981). *Scientific Knowledge: Causation, Explanation, and Corroboration, Boston Studied in the Philosophy of Science.* Dordrecht: D. Reidel.

Weng Hong Tang for helpful comments on an earlier draft.

Feynman, R. P. (1987). Negative Probability. In *Quantum Implications: Essays in Honour of David Bohm*. Routledge and P. Kegan.

Fine, T. (1973). *Theories of Probability*. New York: Academic Press.

Fitelson, B. (2006). Logical Foundations of Evidential Support. *Philosophy of Science*, 73:500–512.

Gillies, D. (2000a). *Philosophical Theories of Probability*. Routledge: New York.

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

Hacking, I. (1967). Slightly More Realistic Personal Probability. *Philosophy of Science*, 34(4):311–325.

Hacking, I. (1971). Equipossibility Theories of Probability. *The British Journal for the Philosophy of Science*, 22(4):339–355.

Hájek, A. (2003). What Conditional Probability Could Not Be. *Synthese*, 137(3):273–323.

Hájek, A. (2007a). Dutch Book Arguments. In Anand, P., Pattanik, P., and Puppe, C., editors, *The Oxford Handbook of Corporate Social Responsibility*. Oxford: Oxford University Press.

Hájek, A. (2007b). Interpretations of Probability. http://plato.stanford.edu/archives/fall2007/entries/probability-interpret/.

Hájek, A. (2007c). The Reference Class is Your Problem Too. *Synthese*, 156:563–85.

Hájek, A. (forthcoming). Arguments For—Or Against—Probabilism. In Huber, F. and Schmidt-Petri, C., editors, *Degrees of Belief*. Oxford University Press.

Humphreys, P. (1985). Why Propensities Cannot Be Probabilities. *Philosophical Review*, 94:557–570.

Jaynes, E. T. (1967). Foundations of Probability Theory and Statistical Mechanics. In Bunge, M., editor, *Popular Electronics*. Springer-Verlag, Berlin.

Jaynes, E. T. (1973). The Well Posed Problem. *Foundations of Physics*, 4(3):477–492.

Jeffrey, R. C. (1983). *The Logic of Decision*. Chicago: University of Chicago Press.

Johnson, W. E. (1921). *Logic*. Cambridge: Cambridge University Press.

Joyce, J. (1998). A Non-Pragmatic Vindication of Probabilism. *Philosophy of Science*, 65(575–603).

Keynes, J. M. (1921). *A Treatise on Probability*. London: Macmillan.

Kolmogorov, A. N. (1933). *Foundations of Probability*. (1950) Chelsea Publishing Company.

Laplace, P. (1814). *A Philosophical Essay on Probabilities.* (1951) New York: Dover.

Levi, I. (1980). *The Enterprise of Knowledge.* MIT Press: Cambridge, Massachussetts.

Lewis, D. (1980). A Subjectivist's Guide to Objective Chance. In Jeffrey, R. C., editor, *Studies in Inductive Logic and Probability*, volume II. Berkeley: University of California Press.

Lewis, D. (1999). Why Conditionalize? In *Papers in Metaphysics and Epistemology*, volume 2. Cambridge: Cambridge University Press.

Maher, P. (1993). *Betting On Theories.* Cambridge: Cambridge University Press.

Marinoff, L. (1994). A Resolution of Bertrand's Paradox. *Philosophy of Science*, 61(1):1–24.

Mikkelson, J. M. (2004). Dissolving the Wine/Water Paradox. *British Journal for the Philosophy of Science*, 55:137–145.

Miller, D. W. (1994). *Critical Rationalism. A Restatement and Defence.* Open Court.

Miller, D. W. (1996). Propensities and Indeterminism. In *Karl Popper: Philosophy and Problems.* Cambridge: Cambridge University Press.

Nover, H. and Hájek, A. (2004). Vexing Expectations. *Mind*, 113:237–249.

Poincaré, H. (1902). *Science and Hypothesis.* (1952) English Translation. New York: Dover.

Popper, K. R. (1938). A Set of Independent Axioms for Probability. *Mind*, 47(186):275–277.

Popper, K. R. (1955). Two Autonomous Axiom Systems for the Calculus of Probabilities. *The British Journal for the Philosophy of Science*, 6(21):51–57.

Popper, K. R. (1957). The Propensity Interpretation of the Calculus of Probability and the Quantum Theory. In Körner, S., editor, *The Colston Papers*, pages 65–70. 9.

Popper, K. R. (1959a). *The Logic of Scientific Discovery.* Basic Books: New York.

Popper, K. R. (1959b). The Propensity Interpretation of Probability. *The British Journal for the Philosophy of Science*, 10(37):25–42.

Popper, K. R. (1990). *A World of Propensities.* Bristol: Thoemmes.

Putnam, H. (1968). Is Logic Empirical? In Cohen, R. and Wartofsky, M., editors, *Boston Studies in the Philosophy of Science*, volume 5, pages 216–241. Dordrecht: Reidel.

Ramsey, F. (1931). Truth and Probability. In Braithwaite, R. B., editor, *Foundations of Mathematics and other Essays.* Routledge and P. Kegan.

Reichenbach, H. (1949). *The Theory of Probability*. Berkeley: University of California Press.

Rényi, A. (1955). On a New Axiomatic Theory of Probability. *Acta Mathematica Academiae Scientiarum Hungaricae*, 6:286–335.

Rényi, A. (1970). *Foundations of Probability*. Holden-Day, Inc.

Roeper, P. and Leblanc, H. (1999). *Probability Theory and Probability Logic*. Toronto: University of Toronto Press.

Schick (1986). Frederic. *The Journal of Philosophy*, 83(2):pp. 112–119.

Seidenfeld, T., Schervish, M. J., and Kadane, J. B. (2001). Improper Regular Conditional Distributions. *The Annals of Probability*, 29(4):1612–1624.

Shackel, N. (2007). Bertrand's Paradox and the Principle of Indifference. *Philosophy of Science*, 74:150–175.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press.

Shannon, C. E. and Weaver, W. (1962). *The Mathematical Theory of Communication*. University of Illinois Press.

Skyrms, B. (1980). *Causal Necessity*. New Haven: Yale University Press.

Skyrms, B. (1984). *Pragmatics and Empricism*. New Haven: Yale University Press.

Sober, E. (forthcoming). Evolutionary Theory and the Reality of Macro Probabilities. In Eells, E. and Fetzer, J., editors, *Probability in Science*. Open Court.

van Fraassen, B. (1984). Belief and Will. *The Journal of Philosophy*, 81(5):235–256.

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

van Fraassen, B. (1995). Belief and the Problem of Ulysses and the Sirens. *Philosophical Studies*, 77:7–37.

Venn, J. (1876). *The Logic of Chance*. Macmillan and Co., 2nd edition.

von Mises, R. (1957). *Probability, Statistics and Truth*. New York: Macmillan, English edition.

Weatherson, B. (2003). From Classical to Intuitionistic Probability. *Notre Dame Journal of Formal Logic*, 44(2):111–123.

Weatherson, B. (2005). Should We Respond to Evil with Indifference? *Philosophy and Phenomenological Research*, 70:613–635.