

# Probability and Chance

Michael Strevens

For the *Macmillan Encyclopedia of Philosophy*, second edition

The weather report says that the chance of a hurricane arriving later today is 90%. Forewarned is forearmed: expecting a hurricane, before leaving home you pack your hurricane lantern.

Probability enters into this scenario twice, first in the form of a physical probability, sometimes called a *chance*, quantifying certain aspects of the local weather that make a hurricane very likely, and second in the form of an epistemic probability capturing a certain attitude to the proposition that a hurricane will strike, in this case one of considerable confidence.

It is not immediately obvious that these two probabilities are two different kinds of thing, but a *prima facie* case can be made for their distinctness by observing that they can vary independently of one another: for example, if the meteorologists are mistaken, the chance of a hurricane may be very low though both they and I am confident that one is on its way.

Most philosophers now believe that the apparent distinctness is real. They are therefore also inclined to say that my belief that the physical probability of a hurricane is very high is distinct from my high epistemic probability for a hurricane. There must be some principle of inference that takes me from one to the other, a principle that dictates the epistemic impact of the physical probabilities—or at least, of my beliefs about the physical probabilities—telling me, in the usual cases, to expect what is physically probable and not what is physically improbable. I call such a principle, me-

diating as it does between two different kinds of probability, a *probability coordination principle*.

The three principal topics of this entry will be, in the order considered, epistemic probability, physical probability, and probability coordination. Two preliminary sections will discuss the common mathematical basis of epistemic and physical probability and the classical notion of probability.

### The Mathematical Basis

What all probabilities, epistemic and physical, have in common is a certain mathematical structure. The most important elements of this structure are contained in the axioms of probability, which may be paraphrased as follows:

1. All probabilities are real numbers between zero and one inclusive (for any proposition  $a$ ,  $0 \leq P(a) \leq 1$ ),
2. The probability of an inconsistent proposition is zero; the probability of a logical truth, or tautology, is one.
3. The probability that either one or the other of two mutually exclusive propositions is true is equal to the sum of the probabilities of the individual propositions. (Two propositions are mutually exclusive if they cannot both be true; the *cannot* is interpreted as a matter of logical consistency, so that the axiom says that for any two propositions  $a$  and  $b$  such that  $a \vdash \neg b$ ,  $P(a \vee b) = P(a) + P(b)$ .)

The axioms as stated here assume that probabilities are attached to propositions, such as the proposition that *A hurricane will strike New York at some time on the afternoon of 1/20/05*. The axioms may also be stated in a way that assumes that probabilities attach to events. It is more natural to attach epistemic probabilities to propositions and physical probabilities to

events, but when the two kinds of probability are discussed side by side it is less confusing, and quite tolerable, to take propositions as the primary bearers of both kinds of probability. Nothing important is thought to turn on the choice.

The three axioms of probability, though simple, may be used to prove a wide range of interesting and strong mathematical theorems. Because all probabilities conform to the axioms, all probabilities conform to the theorems. It is possible, then, to do significant work on probability without presupposing either epistemic or physical probability as your subject matter, let alone some particular construal of either variety. Such work is for the most part the province of mathematicians.

Philosophical work on probability may also be mathematical, but is most often directed to one or the other variety of probability, usually attempting a philosophical analysis of probability statements made in a certain vein, for example, of probability claims made in quantum mechanics or evolutionary biology (both apparently claims about physical probability) or of probability claims made in statistical testing or decision theory (both apparently claims about epistemic probability).

Two important notions encountered in statements of the mathematical behavior of probability are *conditional probability* and *probabilistic independence*. Both are introduced into the mathematics of probability by way of definitions, not additional axioms, so neither adds anything to the content of the mathematics.

*Conditional Probability* The probability of a proposition  $a$  conditional on another proposition  $b$ , written  $P(a|b)$ , is defined to be  $P(ab)/P(b)$ , where  $ab$  is the conjunction of  $a$  and  $b$ . (The conditional probability is undefined when the probability of  $b$  is zero.) For example, the probability of obtaining three heads on three successive tosses of a coin, conditional on the first toss yielding heads, is the probability of obtaining three heads in a row, namely one eighth, divided by the probability of obtaining heads on the first coin,

namely one half—in other words, one quarter.

Some writers suggest taking conditional probability as the basis for all of probability mathematics, a move that allows, among other things, the possibility of conditional probabilities that are well defined even when the probabilities of the propositions conditionalized on are zero (Hájek 2003). On this view, the mathematical posit stated above linking conditional and unconditional probabilities is reinterpreted as an additional axiom.

The act of conditionalization may be used to create an entirely new probability distribution. Given an old probability distribution  $P(\cdot)$  and a proposition  $b$ , the function  $P(\cdot|b)$  is provably also, mathematically speaking, a probability distribution. If  $k$  is a proposition stating all of your background knowledge, for example, then a new probability distribution  $P(\cdot|k)$  can be formed by conditionalizing on this background knowledge, a distribution that gives, intuitively, the probabilities for everything once your background knowledge is taken into account. This fact is especially important in the context of epistemic probability.

*Probabilistic Independence* Two propositions  $a$  and  $b$  are probabilistically independent just in case  $P(ab) = P(a)P(b)$ . When the probability of  $b$  is non-zero, this is equivalent to  $P(a|b) = P(a)$ , or in intuitive terms, the claim that the truth or otherwise of  $b$  has no impact on the probability of  $a$ .

Several of the most important and powerful theorems in probability mathematics make independence assumptions. The theorem of most use to philosophers is the law of large numbers. There are in fact several theorems of this name; the most familiar say, very roughly, that when a large, finite set of propositions are independent, but have the same probability  $p$ , then the proportion of propositions that turn out to be true will, with high probability, be approximately equal to  $p$ . (The generalization to countably infinite sets of propositions is easy if the propositions are ordered; substitute *limiting frequency* for *proportion*.)

For example, the propositions might all be of the form *Coin toss  $x$  will*

*produce heads*, where the  $x$  stands for any one of a number of different tosses of the same coin. If the probability of each of the propositions is one half, then the law of large numbers says, in effect, that provided the tosses are independent, it is very likely that about one half will yield heads.

It is natural to interpret the probabilities in this example as physical probabilities, but the law of large numbers applies equally to any kind of probability, provided that independence holds. There are, in fact, many variants of the law of large numbers, but the details are beyond the scope of this entry.

### Classical Probability

The development of the mathematics, and then the philosophy, of probability was spurred to a perhaps surprising degree by an interest, both practical and theoretical, in the properties of simple gambling devices such as rolled dice, tossed coins, and shuffled cards. Though there was from the beginning a great enthusiasm for extending the dominion of the “empire of chance” to the ends of the earth, gambling devices were—and to some extent are still—the paradigmatic chance setups.

A striking feature of gambling devices is their probabilistic transparency: the discerning eye can “read off” their outcomes’ physical probabilities from various physical symmetries of the device itself, seeing in the bilateral symmetry of the tossed coin a probability of one half each for heads and tails, or in the six-way symmetry of the die a probability of one sixth that any particular face is uppermost at the end of a roll (Strevens 1998).

The classical definition of probability, paramount from the time of Leibniz to the time of Laplace (the late seventeenth century to the early nineteenth century) takes its inspiration from the alignment of probability with symmetry. The best-known formulation of the classical account is due to Laplace:

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. The ratio of this number to that of all the cases possible is the measure of this probability, which is thus simply a fraction whose numerator is the number of favorable cases and whose denominator is the number of all the cases possible (Laplace 1902, 6–7).

As many commentators have noted, this formulation, typical of the classical probabilists, appears to involve two parallel definitions, the first based on the notion of *equal possibility* and the second on the notion of *equal undecidedness*. Laplace's relation of *equal possibility* between two cases probably ought to be understood as picking out a certain physical symmetry in virtue of which the cases have equal physical probabilities. All classical probabilities, on the *equal possibility* definition, have their basis in such physical symmetries, and so would seem to be physical probabilities. The relation of *equal undecidedness* between two cases refers to some sort of epistemic symmetry, though perhaps one founded in the physical facts. A probability that has its basis in undecidedness would seem to be, by its very nature, an epistemic probability. Classical probability, then, is at the same time a kind of physical probability and a kind of epistemic probability.

This dual nature, historians have argued, is quite intentional (Hacking 1975; Daston 1988). In its epistemic guise, classical probability can be called on to do work not normally thought to lie within the province of an objective notion of probability, such as measuring the reliability of testimony, the strength of evidence for a scientific hypothesis, or participating in decision-theoretic arguments such as Pascal's famous wager on the existence of God. In its physical guise, classical probability is able to cloak itself in the aura of unrevisability and reality that attaches to the gambling probabilities such as

the one half probability of heads.

The classical definition could not last. Gradually, it came to be acknowledged that although the epistemic probabilities may, or at least ought to, shadow the physical probabilities wherever the latter are found, they play a number of roles in which there is no physical probability, nor anything with the same objective status as a physical probability, to mimic. The classical definition was split into its two natural parts, and distinct notions of physical and epistemic probability allowed to find their separate ways in the world.

At first, in the middle and later nineteenth century, physical probability commanded attention almost to the exclusion of its epistemic counterpart. Developments in social science, due to Quetelet, in statistical physics, due to Maxwell and Boltzmann, and eventually (around 1930) in the synthesis of evolutionary biology and genetics, due to Fisher and many others, turned on the successful deployment of physical probability distributions.

Beginning in the early twentieth century, however, epistemic probability came into its own, freeing itself over the decades from what came to be seen as the classical probabilists' futile attempt to provide strict guidelines dictating unique rational epistemic probabilities in every situation.

Modern philosophy remade itself in the twentieth century, imposing a historical horizon near the year 1900. The story of the interpretation of probability is often told beginning around that year, with the result that the development of epistemic probability, and "logical probability" in particular, comes first—a convention that will be followed here.

### **Epistemic Probability**

Epistemic probability takes two compatible forms. In its first form, it is a measure of a person's degree of confidence in a proposition, increasing from zero to one as their attitude goes from almost total disbelief to near cer-

tainty. This kind of epistemic probability is called *credence*, *degree of belief*, or *subjective probability*. The propositional attitude you get when you attach a subjective probability to a proposition is sometimes called a *partial belief*.

In its second form, associated most often with the term *logical probability*, epistemic probability measures the impact of a piece or pieces of evidence on a proposition. As such, it may not have the structure of a probability distribution, but it is related to a probability distribution in some straightforward way, and as you will see shortly, is quite capable of providing a basis for a complete system of epistemic probability.

There is a foundational dispute between the proponents of the two forms of epistemic probability. It is not a fight for existence but for primacy: the question is which of the two kinds of epistemic probability is the more epistemologically basic.

*Logical Probability* The second form of epistemic probability has, since 1900, most often taken the guise of *logical probability*. A logical probability is attached not to a proposition but to a complete inductive inference. It is a measure of the degree to which the evidence contained in the premises of an inductive inference, considered in isolation, probabilifies the conclusion. The idea of probabilistic inference was an important part of classical probability theory, but from the post-1900 perspective it is associated first with Keynes (1921)—who was more famous, of course, as an economist.

In explaining the nature of logical probability, and in particular the tag *logical* itself, Keynes draws a close analogy with deductive inference: whereas in a deductive inference, the premises entail the conclusion, in an inductive inference, they *partially entail* the conclusion, the degree of entailment being represented by a number between zero and one, namely, a logical probability. (Note that a degree zero entailment of a proposition is equivalent to full entailment of the proposition's negation; see below.) Just as the first form of epistemic probability generalizes from belief to partial belief, then, the second form generalizes, in Keynes's hands, from entailment to partial

entailment.

An example: take as your conclusion the proposition that the next observed raven will be black. A proposition stating that a single raven has been observed to be black might entail this conclusion only to a relatively small degree, this logical probability representing the slightness of a single raven's color as evidence for the color of any other raven. A proposition stating that many hundreds of ravens have been observed to be black will entail the conclusion to some much greater degree. And so on.

It is an objective matter of fact whether one proposition deductively entails another; so, Keynes conjectured, it is in many cases a matter of objective fact to what degree one proposition partially entails another. These facts themselves comprise inductive logic; the logical probabilities are at base, then, logical entities, just as the name suggests.

Although exact logical probabilities are for Keynes the ideal, he allows that in many cases logic will fix only an approximate degree of entailment for an inductive inference. The presentation in this entry will for simplicity's sake focus on the ideal case.

Keynes's logical probability is not only compatible with subjective probability, the other form of epistemic probability; it mandates certain values for a person's subjective probabilities. If the premises in an inductive inference are known for certain, and they exhaust the available evidence, then their inductive impact on the conclusion—the degree of entailment, or logical probability attached to the inference, from the premises to the conclusion—is itself the degree of belief, that is, the subjective probability, that a rational person ought to attach to the conclusion, reflecting as it does all and only the evidence for the conclusion.

Keynes uses this argument as a basis for taking as a formal representation of logical probabilities the probability calculus itself: the degree to which proposition  $b$  entails proposition  $a$  is written as a conditional probability  $P(a|b)$ . Note that these probabilities do not change as the evidence

comes in, any more than facts about deductive entailment can change as the evidence comes in. The logical probability  $P(a|b)$  must be interpreted as a quantification of the inductive bearing of  $b$  alone on  $a$ , not of  $b$  together with some body of accepted knowledge.

The unconditional probability  $P(a)$ , then, is the inductive bearing on  $a$  of an empty set of evidence—the degree to which  $a$  is entailed, if you like, by the set of logical truths, or tautologies, alone. You might think that the degree of entailment is zero. But this cannot be right: if you have no evidence at all, you must set your subjective probabilities for both  $a$  and its negation equal to their respective degrees of entailment by the tautologies. But you cannot set both subjective probabilities to zero—it cannot be that you are certain that neither  $a$  nor its negation is true, since one of the two must be true. Your complete lack of evidence would be better represented by setting both subjective probabilities to intermediate values, say one half. The logical probabilist, in endorsing this assignment, implicitly asserts that the empty set of evidence, or the set of tautologies, entails both  $a$  and its negation to degree one half.

Although its subject matter is the bearing of evidence on hypotheses, then, logical probability theory finds itself having to take a position on what you should believe when you have no evidence (under the guise of the question of the tautologies' partial entailments). To answer this question, it has turned to the *principle of indifference*, which recommends that when there is no evidence favoring one of several mutually exclusive possibilities over the others, the available probability be equally distributed among them. This is, of course, the very same principle that comprises one strand of the classical definition of probability: Laplace suggested (see above) assigning equal probabilities to cases “such as we may be equally undecided about in regard to their existence”. It has also played an important role in the development of the theory of subjective probability, and so is discussed in a separate section below.

As the role of indifference shows, logical probability is very close in spirit to the epistemic strand of classical probability. It posits, at least as an ideal, a single system of right reasoning, allowing no inductive latitude whatsoever, to which all rational beings ought to conform. Insofar as rational beings ever disagree on questions of evidential impact, it must be because they differ on the nature of the evidence itself.

Many philosophers find this ideal of inductive logic hard to swallow; even those sympathetic to the idea of strong objective constraints on inductive reasoning are often skeptical that the constraints take the form of logical truths, or something analogous to logical truths. This skepticism has two sources.

First is the perception that inductive practices vary widely. Whereas there exists a widespread consensus as to which propositions deductively entail which other propositions, there is no such consensus on degrees of evidential support. That is not to say, of course, that there is disagreement about every aspect of inductive reasoning, but there is far less agreement than would be necessary to build, in the same way that deductive logic was constructed, a useful inductive logic.

Second, there are compelling (though not irresistible) reasons to believe that it is impossible to formulate a principle of indifference that is both consistent and strong enough to do the work asked of it by logical probabilists. These reasons are sketched in the discussion of the principle below.

Carnap (1950) attempted to revive the idea of a system of induction founded on logic alone in the mid-century. His innovation—drawing on his general philosophy of logic—was to allow that there are many systems of inductive logic that are, from a purely logical viewpoint, on a par. We may freely choose from these a logic, that is, a set of logical probabilities, that suits our particular non-logical ends.

Carnap relativized induction in two ways. First, his version of the principle of indifference was indexed to a choice of language; how you distribute

probability among rival possibilities concerning which you know nothing depends on your canonical system for representing the possibilities. Second, even when a canonical language is chosen, Carnap's rule for determining inductive support—that this, degrees of entailment or logical probabilities—contains a parameter whose value may be chosen freely. The parameter determines, roughly, how quickly you learn from the evidence. Choose one extreme, and from the observation of a single black raven you will infer with certainty that the next raven will also be black (“straight induction”). Choose the other extreme, and no number of black ravens is great enough to count as any evidence at all for the blackness of the next raven. A sensible choice would seem to lie somewhere in the middle, but on Carnap's view, logic alone determined no preference ranking whatsoever among the different choices, rating all values apart from the extremes as equally good.

Carnap did give extra-logical arguments for preferring a particular value for the parameter, arriving at an inductive rule equivalent to Laplace's own “rule of succession”. Given that, say,  $i$  out of  $n$  observed ravens have been black, both Carnap and Laplace assign a probability of  $(i + 1)/(n + 2)$  to the proposition that the next raven will be black.

One awkward feature of Carnap's system is that, no matter what value is chosen for the inductive parameter, universal generalizations cannot be learned: the inductive bearing of any number of black ravens on the hypothesis *All ravens are black* is zero.

Carnap's system is of great intrinsic interest, but from the time of its presentation, its principal constituency—philosophers of science—was beginning to move in an entirely different direction. Such considerations as Goodman's new riddle of induction and arguments by Bayesians and others that background knowledge played a part in determining degree of inductive support, though not beyond the reach of Carnap's approach, strongly suggested that the nature of inductive support could not be purely logical.

Today, the logical approach to inductive inference has been supplanted

to a great extent by (though not only by) the Bayesian approach (see the entry on Bayesianism). Yet in Bayesianism itself some have seen the seeds of a new inductive logic.

*Subjective Probability* Whereas logical probability is a logical entity—a quantification of the supposed logical facts about “partial entailment”—the other kind of epistemic probability, subjective probability, is a psychological entity, reflecting an actual cognitive fact about a particular person or (if they are sufficiently agreed) a group of people. The rationality of a person’s subjective probabilities may be a matter of logic, then, but the probabilities themselves are a matter of psychology.

That for a number of propositions, we tend to have a degree of confidence intermediate between the extremes associated with total disbelief and total belief, no one will deny. The advocates of subjective probability as a key epistemological notion—who call themselves *Bayesians* or simply *subjectivists*—go much further than this. They characteristically hold that humans have, or ought to have, well-defined subjective probabilities for every proposition, and that these subjective probabilities play a central role in epistemology, both in inductive inference, by way of Bayes’ conditionalization rule (see the entry on Bayesianism), and in practical deliberation, by way of the usual mechanisms of decision theory.

The subjectivist’s first challenge is to give a substantial characterization of subjective probability and to argue that subjective probabilities are instrumental in human cognition, while at the same time finding a foothold in the descriptive, psychological scheme for the normative concerns of epistemology. Much of this groundwork was laid in Frank Ramsey’s influential paper “Truth and Probability” (Ramsey 1931).

Ramsey does not define subjective probability as such, and even goes so far as to acknowledge that the ideal of a definite subjective probability for every proposition is just that—an ideal that goes a long way towards capturing actual human epistemology without being accurate in every respect.

What he posits instead is a connection—whether conceptual or empirical he does not say—between the value of a person’s subjective probability for a proposition and their betting behavior.

If you have a subjective probability  $p$  for a proposition  $a$ , Ramsey claims, you will be prepared to accept odds of up to  $p : (1 - p)$  on the truth of  $a$ . That is, given a game in which you stand to win  $\$n$  if  $a$  is true, you will pay up to  $\$pn$  to play the game; equivalently, if you will pay up to  $\$m$  to play a game in which you stand to win  $\$n$  if  $a$  is true, your subjective probability for  $a$  must be  $m/n$ . (Decision theorists, note, talk about utility, not dollars.)

Importantly, all human choice under uncertainty is interpreted as a kind of betting. For example, suppose you must decide whether to wear a seat belt on a long drive. You are in effect betting on whether you will be involved in an auto accident along the way. If the cost of wearing a belt, in discomfort, inconvenience, and forsaken cool, is equivalent to losing  $\$m$ , and the cost of being beltless in an accident, in pain, suffering and higher insurance premiums, is  $\$n$ , then you will accept the risk of going beltless just in case your subjective probability for there being an accident is less than or equal to  $m/n$ . (Here the “prize” is negative. The cost of playing is also negative, so just by agreeing to play the game, you gain something, the increase in comfort, cool and so on. Your aim is to play while avoiding a win.) The central doctrine of decision theory is, then, built into the characterization of subjective probability.

Ramsey uses this fact to argue that, provided a person’s behavior is coherent enough to be described, at least approximately, by the machinery of decision theory, their subjective probabilities for any proposition you like may be inferred from their choices. In effect, their subjective probabilities are inferred from the nature of the bets, in the broadest sense, they are prepared to accept. Because our overt behavior can be systematized, approximately, using a decision-theoretic framework, then, we must have subjective probabilities for every proposition, and these probabilities must play a cen-

tral role in our decision theory.

What is the force of the *must* in the preceding sentence? That depends on the nature of the posit that your having a certain subjective probability for a proposition means that you are prepared to accept certain odds on the proposition's being true. Some writers, especially in the mid-century heyday of conceptual analysis and psychological behaviorism, interpret the posit as a definition of subjective probability; on this view, your having certain subjective probabilities just is your having a certain betting behavior. Others, like Ramsey, opt for a looser connection. On any approach, there is a certain amount of latitude in the phrase *prepared to accept*. If you are prepared to accept certain odds, *must* you play a game in which you are offered those odds? Or only if you are in a betting mood? The former answer vastly simplifies the subjectivist enterprise, but at a cost in psychological plausibility: it is surely true that we frequently gamble in the broad sense that we take measured risks, but it is not nearly so obvious that we are compulsive gamblers intent on taking on every favorable risk we can find. Recent work on the psychology of decision-making also suggests that it is a mistake to found the subjectivist enterprise on too strong a conception of the connection between subjective probability and betting behavior.

Subjective probabilities are supposed to conform, as the name suggests, to the axioms of probability theory. In a theory such as Ramsey's, a certain amount of probability mathematics is built into the technique for extracting the subjective probabilities; that humans not only have subjective probabilities, but arrange them in accord with the axioms, is therefore a condition for the success of Ramsey's project.

Insofar as subjective probability is not simply defined as whatever comes out of the Ramsey project, however, there is a question as to whether subjective probabilities obey the axioms. If they do not, there is really very little that they are good for, so the question is an important one for subjectivists, who tend to follow Ramsey in giving a normative rather than a descriptive

answer: it is rational to arrange your subjective probabilities in accordance with the axioms. (It is not unreasonable, of course, to see this normative claim, if true, as evidence for the corresponding descriptive claim, since humans are in certain respects reliably rational.)

The vehicle of Ramsey's argument is what is called the Dutch book theorem: it can be shown that, if your subjective probabilities violate the axioms, then you will be prepared to accept certain sets of bets (which bets depends on the nature of the violation) that will cause you a sure loss, in the sense that you will lose whether the propositions that are the subjects of the bets turn out to be true or false.

The details of the argument are beyond the scope of this entry (for a more advanced introduction, see Howson and Urbach 1993), but an example will illustrate the strategy. The axioms of the probability calculus require that the probability of a proposition and that of its negation sum to one. Suppose you violate this axiom by assigning a probability of 0.8 both to a certain proposition  $a$  and to its negation. Then you are prepared to accept odds of 4 : 1 on both  $a$  and  $\neg a$ , which means a commitment to playing, at the same time, two games, in one of which you pay \$8 and win \$10 (i.e., your original \$8 plus a \$2 "profit") if  $a$  is true, and in one of which you pay \$8 and win \$10 if  $a$  is false. Whether  $a$  is true or false, you pay \$16 but win only \$10—a certain loss. To play such a game is irrational; thus, you should conform your subjective probabilities to the probability calculus. Needless to say, the Dutch book argument works best on the dubious interpretation of *prepared to accept* as equivalent to *compelled to accept*; there have been many attempts to reform or replace the argument with something that makes weaker, or even no, assumptions about betting behavior.

Subjectivism has been developed in several important directions. First are various weakenings or generalizations of the subjectivist machinery. The question of the connection between subjective probability and betting behavior is, as noted, one locus of activity. Another attempts to generalize

the notion of a subjective probability to a subjective probability interval, the idea being that where you do not have an exact subjective probability for a proposition, you may have an approximate level of confidence that can be captured by a mathematical interval, the equivalent of saying that your subjective probability is indeterminately somewhere between two determinate values.

Second, and closely related, is all the work that has been put into developing decision theory over the last 100 years (see for example Jeffrey 1983).

Finally, subjectivism provides the foundation for the Bayesian theory of inference. At the root of the Bayesian system is a thought much like the logical probabilist's doctrine that, if  $k$  is your background knowledge, then your subjective probability for a hypothesis  $a$  ought to be  $P(a|k)$ . Whereas for a logical probabilist a conditional probability  $P(a|b)$  is a timeless logical constant, for a subjectivist it is something that constantly changes as further evidence comes in (even holding  $a$  and  $b$  fixed). For this reason, the subjectivist theory of inference must be an inherently dynamic theory; what is perhaps its best-known weakness, the "problem of old evidence", arises from this very fact. See the entry on Bayesianism for further discussion.

Subjectivism had almost entirely eclipsed logical probabilism by the late twentieth century; as the celestial metaphor unwittingly implies, however, there is a cyclic aspect to philosophical history: an interest in the central notion of logical probability theory, evidential weight, is on the rise.

There are three strands to this new movement. First is the perception among philosophers of science that scientific discourse about evidence is almost never about the subjective probability scientists should have for a hypothesis, and almost always about the degree of support that the evidence lends to the hypothesis. Second is the development of new and safer (though limited) versions of the principle of indifference. Third is technical progress on the project of extracting from the principles of Bayesian inductive inference a measure of weight. Note that this third project conceives of inductive

weight as something derived from the more basic Bayesian principles governing the dynamics of subjective probability, a view quite opposed to the logical probabilists' derivation of rational subjective probabilities from the (by their lights) more basic logical principles governing the nature of inductive support. See the entry on Bayesianism for further discussion.

*Indifference* The principle of indifference distributes probability among various alternatives—in the usual case, mutually exclusive and exhaustive propositions—concerning which little or nothing is known. The principle's rationale is that certain probability distributions reflect ignorance better than others. If I know nothing that distinguishes two mutually exclusive possibilities, picked out by propositions  $a$  and  $b$ , then I have no reason to expect one more than the other: I should assign the propositions equal probabilities. Any asymmetric assignment, say assigning twice the probability to  $a$  that I assign to  $b$ , would reflect some access on my part to facts supporting  $a$  at the expense of  $b$ . Thus ignorance and probabilistic symmetry ought to go hand in hand—or so the principle of indifference would have it.

The principle is an essential part of logical probability theory, for the reasons given above, but there have always been subjectivists who appeal to the principle as well. It is most useful within the Bayesian approach to inductive inference; see the entry on Bayesianism for further discussion.

The epistemic strand of classical probability theory also invokes the principle, of course, running it together with the discernment of “equally possible cases” in the paradigmatic gambling setups. This conflation has confused the discussion of the principle ever since, with proponents of the principle continuing to take aid and comfort in the principle's apparent virtuoso handling of cases such as the one half probability of heads. Our reasoning about the gambling probabilities, however, as the classical probabilists for the most part themselves dimly saw, is a matter of inferring physical probabilities from physical symmetries, not of setting epistemic probabilities to reflect symmetric degrees of ignorance (Strevens 1998).

The most famous arguments against the principle of indifference were developed in the nineteenth century, as already noted a time of hegemony for physical over epistemic probability. They take their name from Joseph Bertrand, who pointed to the difficulty of finding a unique symmetry in certain indifference-style problems.

Consider, for example, two leading theories of dark matter in the universe: the MACHO and the WIMP theories. Each posits a certain generic form for dark matter objects, respectively very large and very small. If you have no evidence to distinguish them, it seems that the principle of indifference directs you to assign each a probability of one half (assuming for the sake of the argument that there are no other possibilities). But suppose that there are four distinct schools of thought among the MACHO theorists, corresponding to four distinct ways that MACHOS might be physically realized, and eight such schools of thought among WIMP theorists. Now there are twelve possibilities, and once probability is distributed equally among them, the generic MACHO theory will have a probability of one third and the WIMP theory a probability of two thirds. Cases such as this make the principle seem capricious, if not simply inconsistent (as it would be if it failed to pick out a privileged symmetry).

Matters become far worse, as Bertrand noted, when there are uncountably many alternatives to choose among, as is the case in science when the value of a physical parameter, such as the cosmological constant, is unknown. Even in the simplest such cases, the principle equivocates (van Fraassen 1989, chap. 12). As noted above, some progress has been made in solving these problems, Jaynes (1983) being a ringleader. Most philosophers, though, doubt that there will ever be a workable principle of indifference suited to the needs of general inductive inference.

## Physical Probability

The paradigms of physical probability are the probabilities attached to gambling setups; there are, however, many more interesting examples: the probabilities of quantum mechanics and kinetic theory in physics, the probabilities of population genetics in evolutionary theory, actuarial probabilities such as the chance of dying before reaching a certain age, and the probabilities in many social science models.

It is by no means clear that there is a single phenomenon to be explained here; the physical probabilities ascribed to phenomena by our best scientific theories may differ in their makeup from theory to theory. There is a commonality in the phenomena themselves, however: whenever the notion of physical probability is put to scientific work, it is to predict or explain what might be called probabilistic patterns of outcomes. These patterns are characterized by a certain kind of long run order, discernible only over a number of different outcomes, and a certain kind of short term disorder, the details of the order and disorder depending on the variety of probability distribution.

The simplest and best-known of the patterns is the Bernoulli pattern, which takes its name from the corresponding probability distribution. This is the pattern typical of the outcomes produced by gambling devices, such as the pattern of heads and tails obtained by tossing a coin. The long term order takes the form of a stable frequency equal to the corresponding probability; in the case of the tossed coin, this is of course the one half frequency with which heads and tails occur (almost always) in the long run. The short term disorder, though an objective property of the pattern itself, is perhaps best gotten at epistemically: once you know that the long run frequency of heads is one half, the outcome of one toss gives you no useful information about the outcome of the next. The law of large numbers implies that a chance setup will produce its characteristic probabilistic patterns in the long run with a very high (physical) probability.

When discussing physical probability, it is more natural to talk of probabilities attaching to events than to propositions; what follows will be formulated accordingly.

*The Frequency Theory* The frequentist theory of physical probability has its roots in the empiricist interpretation of law statements according to which they assert only the existence of certain regularities in nature (on the regularity theory, see Armstrong 1983). What is usually called the *actual frequency* theory of probability understands physical probability statements, such as the claim that the probability of a coin toss's yielding heads is one half, as asserting in like spirit the existence of the appropriate probabilistic patterns—in the case of the coin toss, for example, a pattern of heads and tails in the actual outcomes of coin tosses exemplifying both the order and the disorder characteristic of the Bernoulli patterns.

The characteristic order in a Bernoulli pattern is a long run frequency approximately equal to the relevant probability; in the case of the coin, then, it is a long run frequency for heads of one half. It is from this aspect of the pattern that frequentism takes its name. (One complication: a distinction must be made between the case in which the set of events exemplifying the pattern is finite and the case in which it is countably infinite. In a finite case, what matters is the proportion or relative frequency, whereas in the infinite case, it is instead the *limiting frequency*, that is, the value of the relative frequency in the limit, if it exists, as it must for the Bernoulli pattern to exist.)

Although their account is named for frequencies, most frequentists insist also on the presence of appropriate short term disorder in the patterns. It is less easy to characterize this disorder in the purely extensional terms implicit in a commitment to regularity metaphysics. Suffice it to say that there is a broad range of characterizations, some very strict, some rather lax. Among frequentists, von Mises tends to a strict, Reichenbach to a lax, requirement (though Reichenbach holds, characteristically, that there is no uniquely correct level of strictness). For a discussion of the technical prob-

lems in constructing such a requirement, see Fine (1973).

The probability that a particular coin toss lands heads is one half, according to frequentism, because the outcome of the toss belongs to a series that exemplifies the Bernoulli pattern with frequency one half. The “truth-maker” for the probability claim is a fact, then, about a class of outcomes, not just about the particular outcome to which the probability is nominally attached. But which class? If I am tossing an American quarter, does the class include all American quarters? All American and Canadian quarters? All fair coins? Or—ominously—all coin tosses producing heads? To give an answer to this question is to solve what has become known as the problem of the reference class.

The standard frequentist solution to the problem is to understand probability claims as including a (perhaps implicit) specification of the class. All physical probability claims are, in other words, made relative to a reference class. This doctrine reveals that the frequency theory is best seen as an account, in the first instance, of statements of statistical laws. A claim about the one half probability of heads, for example, is on the frequency interpretation in essence a statement of a probabilistic law concerning a class of coin tosses, not a claim about a property of a particular toss.

The kinship between the regularity account of deterministic laws and the frequency account of probability is, then, even closer than it first appears. Note that the regularity account has its own analog of singular probability claims, namely, singular claims about deterministic “tendencies”, such as a particular brick’s tendency to fall to earth when released. Regularity theorists interpret a tendency claim not as picking out an intrinsic property of the object possessing the tendency, but as a veiled law statement.

The case of probability introduces a complication, however, that is not present in the case of exceptionless regularities: a particular coin toss will belong to many reference classes, some with different frequencies for heads. There may be, then, no determinate fact of the matter about an individual

coin toss's "probabilistic tendency" to produce heads, or equivalently, about what are often called *single case probabilities*. Frequentists have made their peace with this consequence of their view.

Opponents of the frequency view argue that single case probabilities are metaphysically, inductively, and explanatorily indispensable. Are they right? The case for metaphysical indispensability: some writers, especially propensity theorists (see below), hold that there is clearly a fact of the matter about the value of the probability that some particular coin toss lands heads, independent of any choice of reference class. Frequentists may simply deny the intuition, or may try explain away the appearance of a single case fact (for related versions of the explanation, see Reichenbach 1949, §68 and Strevens 2003, pp. 61–62).

The case for predictive indispensability: in order to settle, for predictive and decision-theoretic purposes, on a rational subjective probability for an event using the probability coordination principle, a corresponding physical probability must be found (see the discussion of probability coordination below). The corresponding probability is often understood to be the physical probability of that very event, hence, a single case probability. Frequentists must find an alternative understanding. Reichenbach (1949, §71–72) proposes using the frequentist probability relative to the narrowest reference class "for which reliable statistics can be compiled" (p. 374).

The case for explanatory indispensability rests principally on the intuition that the probabilistic explanation of a single outcome requires a single case probability. The philosophy of scientific explanation, much of it developed by regularity theorists and other metaphysical empiricists, offers a number of alternative ways of thinking about explanation, for example, as a matter of showing that the outcome to be explained was to be expected, or as a matter of subsuming the outcome to be explained under a general pattern of outcomes (both ideas proposed by Hempel; see the entry on *Explanation*). The fate of frequentism, and more generally of the regularity

approach to laws of nature, depends to some extent, then, on the adequacy of these conceptions of explanation.

Why be a frequentist? The view has two principal advantages. First is its light metaphysical touch, shared with the regularity account of laws. Second is the basis it gives for the mathematics of probability: frequencies, as mathematical objects, conform to almost all the axioms of probability. Only almost all, because they violate the axiom of countable additivity, an extension to the countably infinite case of the third axiom described earlier. Countable additivity plays an important role in the derivation of some of probability mathematics' more striking results, but whether it is necessary to provide a foundation for the scientific role of physical probability claims is unclear.

There is more than one way to be a frequentist. A naive actual frequentist holds that there is a probability wherever there is a frequency, so that, in a universe where only three coin tosses have ever occurred, two coming up heads, there is a probability for heads of two-thirds. This view has been widely criticized, though never held. Compare with the naive regularity theory of laws (Armstrong 1983, §2.1).

What might be called *ideal actual frequentism* is the theory developed by Reichenbach (1949) and von Mises (1957). On this view, probability statements are construed as ideally concerning only infinite classes of events. In practice, however, they may be applied to large finite classes that in some sense come close to having the properties of infinite classes. Thus Reichenbach distinguishes the *logical meaning* of a probability statement, which asserts the probabilistic patterning of an infinite class of outcomes, and the *finitist meaning* that is given to probability claims in “physical applications”, that is, in the scientific attribution of a physical probability (p. 349). On the finitist interpretation, then, a physical probability claim concerns the probabilistic patterning of some actual, finite class of events—albeit a class large enough to have what Reichenbach calls a “practical limiting frequency”. (Re-

Reichenbach's wariness about logical meaning owes as much, incidentally, to his desire to have his theory of probability conform to the verifiability theory of meaning as to a concern with, say, the validity of probability claims in a finite universe.)

David Lewis, reviving Frank Ramsey's account of laws of nature, proposes that the fundamental laws are nothing but the axioms of the theory that best systematizes, or unifies, the phenomena. A systematization is good to the degree that it is simple, to the degree that it makes claims about a large proportion of the phenomena (ideally all the phenomena, of course), and to the degree that its claims are accurate. (Lewis 1994) extends the definition of accuracy, or as he calls it, *fit*, to accommodate axioms attributing physical probabilities: a set of phenomena are a good fit to a physical probability statement if the phenomena exemplify the probabilistic patterns appropriate to the probability ascribed. A system of probabilistic axioms will be a good systematization, then, only if the physical probabilities it assigns to the phenomena are reflected, for the most part, in corresponding probabilistic patterns.

In this respect, Lewis's view is a form of frequentism: although there is not some particular set of outcomes whose probabilistic patterning is necessary and sufficient for the truth of a given probabilistic law statement, it is nevertheless the world's probabilistic patterns, taken as a whole, that provide the basis for all true statements of probabilistic law.

Some writers suggest that a claim such as *The probability of obtaining heads on a toss of this coin is one half* is equivalent to the claim that, if the coin were tossed infinitely many times, it would yield heads with a limiting frequency of one half. The truth-makers for physical probability claims, then, are modal facts (except in the case where there actually are an infinite number of tosses). This view is known as *hypothetical frequentism*.

Though much discussed in the literature, hypothetical frequentism is seldom advocated. Reichenbach and von Mises are sometimes labeled hy-

pothetical frequentists, but the textual evidence is thin, perhaps even non-existent. Howson and Urbach (1993, chap. 13) advocate a hypothetical frequency view. Bas van Fraassen's frequencies are also hypothetical, but because he holds that the literal meaning of theoretical claims is irrelevant to the scientific enterprise, the spirit of his account of probability is, in its empiricism, closer to Reichenbach's ideal actual frequentism (van Fraassen 1980, chap. 6).

The weaknesses of frequentism are in large part the weaknesses of the regularity theory of laws. An interesting objection with no parallel in the regularity account is as follows: in the case of reference classes containing countably infinite numbers of events, the value (indeed, the existence) of the limiting frequency will vary depending on how the outcomes are ordered. There appear to be no objective facts, then, about limiting frequencies. Or rather, if there are to be objective facts, there must be some canonical ordering of outcomes, either specified along with the reference class or fixed as a part of the scientific background. How serious an impediment this is to the frequentist is unclear.

*The Propensity Theory* If frequentism is the regularity theorist's natural interpretation of physical probability claims, then the propensity account is the interpretation for "realists" about laws, that is, for philosophers who believe that law statements assert the existence of relations of nomic necessity, causal tendencies, and so on (Armstrong 1983). For the propensity theorist, probabilities are propensities, and propensities are a certain kind of distinctly probabilistic causal tendency or disposition.

The propensity theorist's home territory is single case probability, the kind of probability attached to a particular physical process or outcome independently of the specification of a reference class or ordering of outcomes. Because propensities are supposed to be intrinsic properties of token processes, on the propensity view every probability is a single case probability. Given some particular outcome that you wish to predict or explain, then,

there is an absolute fact of the matter as to the physical probability of the outcome, which you may—and presumably, must—use in your prediction or explanation.

Of course, knowledge of this fact, if it is to be obtained by observing the statistics of repeated experiments, will require the choice of a reference class, the aim being to find a class containing processes that are sufficiently similar that their statistics reveal the nature of each of the underlying propensities in the class. Further, by analogy with the case of deterministic causal tendencies, propensities may owe their existence to probabilistic laws governing classes of processes. Thus, something not unlike the frequentist's reference classes may turn up in both the epistemology and the metaphysics of propensities, but this does not detract from the fact that on the propensity view, there are real, observer-independent single case probabilities.

To identify probabilities with propensities is revealing because we think that we have a good intuitive sense of the nature of propensities in the deterministic case; we are reasonably clear on what it is to be fragile, aggressive, or paramagnetic. Though the metaphysics of dispositions is still a matter of dispute, it seems that we come to deterministic propensities, at least at first, by grasping what they are propensities for: breaking, violent behavior, magnetic attraction, and so on. To adopt a propensity theory of probability, then, with the sense of familiarity the word *propensity* brings, is to make an implicit commitment to elucidating what probabilistic propensities are propensities for.

A straightforward answer to this question was given by Popper (1959) in one of the earliest modern presentations of the propensity theory: a probabilistic propensity is a disposition to produce probabilistically patterned outcomes. A particular coin's probability for heads of one half, then, is a disposition to produce a sequence of heads and tails that is disordered in the short term, but in the long term contains heads with a frequency of one half. (Popper in fact omits the disorder requirement, and allows that the sequence

may be long and finite or infinite.) On Popper's view, then, a probabilistic propensity differs from a deterministic propensity not in the means of production, but only in what is produced: a probabilistic pattern over a long series of trials, rather than a single discrete episode of, say, shattering or magnetic attraction.

Popperian propensity theory is committed to the claim that, if the probability of a tossed coin's landing heads is one half (and remains so), then continued tossing of the coin will eventually yield a set of outcomes of which about one half are heads. But this sits badly with our intuitive conception of the workings of probability: if the probability of heads is one half, then it is possible, though unlikely, that it will produce all heads for as long as you like, even forever.

This intuition has an analog in probability mathematics. The law of large numbers prescribes a very high probability that the long run frequency with which an outcome occurs will match its probability; by the same token, however, there is a non-zero probability that any (finite) long run will fail to produce a probability-matching frequency. There is some physical probability, then, that a probabilistic propensity will fail to produce what, according to the Popperian propensity view, it must produce. If this physical probability is itself a Popperian propensity—and surely it is just another manifestation of the original one half propensity for heads—then it must produce, by Popper's definition, a matching frequency, which is to say that it must occasionally produce the supposedly impossible series of heads. If it is to be consistent, Popper's definition must be carefully circumscribed. (There is a lesson here for frequentists, too.)

Most propensity theorists accept that probabilistic setups will occasionally fail to produce probability-matching frequencies. Thus they repudiate Popper's version of the propensity theory. What, then, can they say about the nature of the propensity? Typically, they hold that the probability of, say, heads is a propensity to produce the appropriate probabilistic patterns

with a very high physical probability (Fetzer 1971; Giere 1973)—thus such a probabilistic propensity is probabilistic not only in its characteristic effect, which is, as on Popper’s definition, a probabilistic pattern, but also in its relation to the effect. (Mellor (1971) offers an interesting variant on this view.)

Whereas the Popperian definition comes close to inconsistency, this new definition is manifestly circular. Its proponents accept the circularity, so committing themselves to the ineffability of probabilistic propensities.

The ineffability of propensities, it is asserted, is not a problem provided that their values can be inferred; the usual apparatus of statistical inference is tendered for this purpose. Critics of the post-Popperian propensity interpretation naturally fasten on the question of whether it succeeds in saying anything substantive about probability at all—anything, for example, that illuminates the question of why physical probabilities conform to the axioms of the probability calculus or explain the outcomes that they produce. It does seem that modern propensity theorists are not so far from what is sometimes called the “semantic interpretation” of probability, on which probabilities are considered to be model-theoretic constructs that ought not to be interpreted at all, but simply accepted as formal waypoints between evidence and prediction in probabilistic reasoning (Braithwaite 1953). Compare Carnap’s notion of partial interpretation and Suppes (1973).

A characteristic doctrine of the propensity theory is that probabilistic propensities, hence probabilities, are metaphysically irreducible: they are in some sense fundamental building blocks of the universe. The corollary to this doctrine is that the physical probabilities science assigns to outcomes that are deterministically produced—including, according to many philosophers, the probabilities of statistical mechanics, evolutionary biology, and so on—are, because not irreducible, not propensities, hence not “real” probabilities. Giere writes that they must be given an “as if” interpretation, but propensity theorists offer no account of “as if” probability’s scientific role.

On a broader understanding of the nature of a propensity, however, at least some of the physical probabilities assigned by science to the outcomes of deterministic processes might count as probabilistic propensities. As explained in the entry on chaos, certain sub-classes of chaotic systems have dynamic properties in virtue of which they tend to generate probabilistic patterns of outcomes (Strevens 2003). These dynamic properties may be understood, then, as endowing the systems with a propensity to produce probabilistic patterns, and the propensity itself may be identified with the physical probabilities that science ascribes to the outcomes.

There is one, not inconsiderable, complication: the systems in question will generate the probabilistic patterns only given appropriate initial conditions. Almost all, but not all, initial conditions will do. This raises two important questions that need to be answered if chaos is to provide a part of the foundation for the metaphysics of physical probability. First, ought the necessary properties of the initial conditions to be considered a part of the propensity? If so, the propensity seems not to be an intrinsic causal property of the process. Second, the initial conditions are, in this context, most naturally described using a probability distribution. Thus the basis of the probabilistic propensity is a further probabilistic element itself in need of analysis.

*The Subjectivist Theory* It is something of a mystery why the mathematics of the probability calculus should be useful both for capturing elements of belief and inductive inference and for describing the processes that give rise to probabilistic patterns, or in other words, why two such different things as epistemic and physical probability should share the same formal structure.

According to the subjectivist theory of physical probability, there is no mystery at all: physical probabilities are nothing but a certain kind of subjective probability. The intuition that, say, the probability of heads is a quantification of some physical property of the tossed coin is, on the subjectivist approach, an illusion: there are frequencies and mechanical properties out

in the world, but “physical probabilities” exist entirely in the descriptive apparatus of our theories, or in our minds.

For the principal architect of subjectivism, Bruno de Finetti, the appeal of the theory is not only its neoclassical reunification of epistemic and physical probability, but also its empiricism: subjectivism is most at home in what is now called a Humean world. Of course, frequentism is also a theory of physical probability that the metaphysical empiricist can embrace; the main advantage of subjectivism over frequentism is its provision—if such is truly necessary—of single case probabilities (de Finetti 1964).

Subjectivism asserts the identity of the subjective probability for heads and the physical probability for heads. But it does not claim that, say, my subjective probability for the MACHO theory of dark matter is also a physical probability for the theory. Rightly so, because we do not acknowledge the existence of physical probabilities wherever there are subjective probabilities. A plausible subjectivism must have the consequence that we project only a small subset of our subjective probabilities onto the world as physical probabilities.

At the heart of the subjectivist theory, then, must be a criterion that picks out just those subjective probabilities that are experienced as physical, and that accounts for their particular, peculiar phenomenology. The key notion in the criterion is one of *resilience*: unlike most subjective probabilities, which change as more evidence comes in, the subjective probabilities we call *physical* have attained a certain kind of stability under the impact of additional information. This stability gives them the appearance of objectivity, hence of reality, hence of physicality, or so the subjectivist story goes. Skyrms (1980) employs this same notion of resilience to give a projectivist account of causal tendencies and lawhood in the deterministic as well as the probabilistic case; subjectivism, then, like frequentism and the propensity theory, can be seen as a part of a larger project embracing all causal and nomological metaphysics.

There is an obvious difficulty with the subjectivist position as elaborated so far: my subjective probability for an outcome such as a coin's landing heads may very well change as the evidence comes in. I may begin by believing that a certain coin is fair, and so that the physical probability of its yielding heads when tossed is one half. As I continue to toss it, however, I may come to the realization that it is biased, settling eventually on the hypothesis that the physical probability of heads is three quarters. Throughout the process of experimentation, I project (according to the subjectivist) a physical probability distribution onto the coin, yet throughout the process, because the projected physical probability for heads is changing, increasing from one half to three quarters, my subjective probability for heads is also changing. Where is the resilience?

De Finetti's achievement is to find a kind of resilience, or constancy, in my subjective probabilities even as my subjective probability for heads is changing. This resilience is captured by the property de Finetti calls *exchangeability*. Consider my subjective probability distribution over, say, the outcomes of the next four tosses of my coin. Every possible sequence of four outcomes will be assigned some subjective probability. The probability assignment—the subjective probability distribution—is said to be exchangeable if any two sequences having the same number of heads and tails are assigned equal probabilities. For example, exchangeability implies that *HTHT* and *HHTT*, each having two heads and two tails, are assigned the same probability, but allows this probability to differ from that assigned to, say, *HHHT*. In an exchangeable distribution, then, the probability assigned to a sequence of heads and tails depends only on the relative frequency with which heads and tails occur in the sequence (in the case of infinite sequences, which de Finetti uses in his mathematical construction, substitute *limiting frequency*).

If my subjective probability distribution over heads and tails is exchangeable, then the order in which the heads and tails come in as I experiment

with my coin will not in itself affect my subjective probability for heads. The frequency with which heads and tails come in will, by contrast, most definitely affect my subjective probability. Thus exchangeability is a kind of partial resilience; it is resilience to information about order, but not frequency.

De Finetti claims, uncontroversially, that our subjective probability distributions over future sequences of heads and tails (and the outcomes of other Bernoulli setups) are exchangeable. He goes on to prove a theorem—his celebrated representation theorem—that shows that the following two reasoners will be outwardly indistinguishable: first, a reasoner who has various hypotheses about the physical probability of heads and updates the subjective probabilities for these hypotheses in the usual way as evidence comes in, and second, a reasoner who has no beliefs about physical probabilities, but simply has an exchangeable subjective probability distribution over future sequences of outcomes. The only difference between the two reasoners, then, will be that the first will claim, presumably as a result of introspection, to be learning about the values of physical probabilities in the world.

The subjectivist's sly suggestion is that we are all in fact reasoners of the second kind, falsely believing that we are reasoners of the first kind. Or, in a more revisionist mood the subjectivist may argue that, though we are reasoners of the first kind, we will give up nothing but dubious metaphysical commitments by becoming reasoners of the second kind.

Critics of subjectivism question the aptness of exchangeability as a psychological foundation for probabilistic reasoning. The sole reason that we assign exchangeable subjective probability distributions to certain classes of sequences, according to these writers, is that we believe the sequences to be produced by physical probabilities (Bernoulli distributions, to be exact), and we know that an exchangeable subjective probability distribution is appropriate for outcomes so produced. This argument, note, has both a descriptive and a normative dimension: against a descriptive subjectivist, who holds that beliefs about physical probability play no role in our probabilistic

reasoning, the critic proposes that such beliefs cause us to assign exchangeable distributions. Against a normative subjectivist, who holds that beliefs about physical probability *should not* play a role in our probabilistic reasoning, the critic proposes that such beliefs are required to *justify* our assigning exchangeable distributions.

A different line of criticism targets subjectivism's metaphysics: why not identify physical probability with whatever produces the probabilistic patterns? Why not say that the probability of heads is a quantification of, at least in part, the physical symmetry of the coin? Such a position has its problems, of course (see the discussion of the propensity theory above), but they are not obviously insurmountable. More generally, given the rich array of options available for understanding the nature of physical probability, the subjectivist's flight from any attempt to give a metaphysics seems to many, as yet, insufficiently motivated.

### Probability Coordination

It is generally accepted that it is rational, in normal circumstances, to set your subjective probability for an event equal to the physical probability ascribed by science to that event, or to that type of event. Returning to the very first paragraph of this entry, if the physical probability of a hurricane is high, you should expect—you should assign a high subjective probability to—a hurricane strike. This is the principle of *probability coordination*.

Because the equation of physical and epistemic probability is made explicit in the classical definition of probability, classicists are probability coordinators *par excellence*. Leibniz, for example, articulates what appears to be an early formulation of the probability coordination principle when he writes *quod facile est in re, id probabile est in mente*; Hacking (1975, 128) glosses this as “our judgment of probability ‘in the mind’ is proportional to (what we believe to be) the facility or propensity of things” (the parenthe-

sized phrase is not in the Latin). But strictly speaking, of course, classicists cannot conceive of this as a coordination of different kinds of probability, since they allow only one kind of probability.

In the twentieth century, probability coordination was introduced as a topic in its own right by David Miller, who argued, as a part of a Popperian case against inductive inference, that a probability coordination principle would have to be inconsistent. Commentators soon pointed out that there are consistent versions of the principle, and some years later David Lewis wrote what is still the most influential paper about the proper form of a principle of coordination and its role in scientific inference, conjecturing that such a principle “capture[s] all we know about [physical probability]” (Lewis 1980).

Modern attempts at a formulation of a probability coordination principle contain two elements not present in Leibniz’s maxim. First is the modification interpolated by Hacking above: the principle commands that you set your subjective probabilities equal not to the corresponding physical probabilities, but to what you believe the values of those probabilities to be, or more generally, to the mean of the different possible values, weighted by your subjective probability that each value is the correct one. Such a principle might be loosely interpreted as saying that you should do your best to set your subjective probabilities equal to the physical probabilities.

Second is a restriction of the range of the principle: when you possess certain kinds of information, probability coordination is not necessarily rational. Suppose, for example, that you know for some science-fictional reason that the coin you are about to toss will land heads. Then you should set your subjective probability for heads equal to one, not equal to the physical probability of one half. The information that the coin will land heads is what Lewis calls *inadmissible information*; in the presence of inadmissible information, the principle of probability coordination does not apply. Note that what is admissible is relative to the outcome in question; knowing how

the coin lands is admissible when I am setting my subjective probability for the outcome of a different toss.

An attempt at a probability coordination principle might, then, have the following form: your subjective probability for an event  $e$ , conditional both on the proposition that the physical probability of  $e$  is  $p$  and on any admissible information  $k$ , should be set equal to  $p$ . (Your unconditional subjective probability for  $e$ , then, will be the weighted sum of the physical probabilities, as mentioned above.) In symbols: if your background knowledge is admissible, then set

$$C(e|tk) = P_t(e),$$

where  $C(\cdot)$  is your subjective probability distribution,  $t$  is the proposition that the correct physical probability distribution for  $e$  is  $P_t(\cdot)$ , and  $k$  is any other admissible information.

Note that propositions such as  $t$  are normally consequences of two kinds of fact: probabilistic laws of nature, and some properties of  $e$  in virtue of which it falls under the laws. For example, if  $e$  is the event of a particular coin toss's landing heads, then the law might be *All tosses of a fair coin land heads with physical probability one half* and the additional fact the fairness of the coin in question. In what follows I assume that the latter facts are part of the background knowledge, and that  $t$  simply asserts some probabilistic law of nature, as suggested by the above notation.

The most puzzling aspect of the probability coordination principle is the nature of admissibility. Lewis proposes what might be called a working definition of admissibility (he says that it is a "sufficient or almost sufficient" condition for admissibility) on which information is admissible either if it is historical—if it concerns only facts about the past up to the point where the principle is invoked—or if it is purely probabilistic, that is, if it is information about physical probabilities themselves.

The characterization is problematic for two reasons. One difficulty is explicitly identified by Lewis and for many years prevented him from ad-

vancing the frequency-based theory of physical probability (described under *Frequentism* earlier in this entry) that he wished to give. As noted above, when you are coordinating probabilities for a given outcome, your ought to count information about the future occurrence or otherwise of that outcome as inadmissible. It turns out that frequency-based probabilities provide information of this sort. Lewis, then, has three choices. The first is to revise the working definition of admissibility so as to rule out such information, in which case information about physical probabilities will be inadmissible and the resulting probability coordination principle useless. The second is to stay with the working definition of admissibility, allowing the information provided by frequency-based probabilities to count as admissible by fiat. It can be shown, however, that the resulting principle—that is, Lewis’s original principle—clearly sets the wrong subjective probabilities in certain circumstances: there are certain complex facts about the future that a frequency-based probability distribution entails cannot obtain, yet assigns a non-zero probability. If such a probability distribution is known to be the correct one, then the right subjective probability for the facts is zero, but probability coordination results in a non-zero subjective probability. The third option is to abandon probability coordination as such. Lewis takes the third way out, proposing a new kind of probability coordination principle that has the form (using the notation from above)  $C(e|tk) = P_t(e|t)$ . Strevens (1995) points out that both Lewis’s new principle and his original principle are consequences of a more general probability coordination principle according to which conditional subjective probabilities should be set equal to conditional physical probabilities. This principle yields Lewis’s original principle when information about physical probability distributions is admissible and Lewis’s new principle when it is not.

A different problem with Lewis’s working definition of admissibility is that it makes no sense of probability coordination in deterministic systems. If you conditionalize on the exact initial conditions of a coin toss, you ought

not to set your subjective probability for heads to the physical probability of heads, one half, but either to zero or to one depending on whether those particular initial conditions cause the coin to land heads or tails. If a probability coordination principle is to be applied to the probability of heads, exact information about initial conditions must therefore be ruled inadmissible. Lewis's working definition of admissibility counts initial conditions, like all historical facts, as admissible.

Lewis does not regard this as a problem, since he agrees with the propensity theorists that in deterministic systems there could be only ersatz physical probabilities. Even if this is correct as a metaphysical doctrine, however, it remains a matter of fact that we coordinate our subjective probabilities with such ersatz probabilities all the time, as when we form expectations about the outcomes of a tossed coin. Whatever you call it, then, there is a coordination principle for systems such as gambling devices that apparently has the same form as the genuine probability coordination principle. (See Loewer (2001) for a reconciliation of Lewis's account of physical probability and probability coordination in deterministic systems.)

There is clearly more work to be done elucidating the form of the probability coordination process, and in understanding admissibility in particular. A different project attempts to justify the practice of probability coordination, by giving an a priori argument that subjective probabilities should track physical probabilities, or beliefs about such. Lewis himself says no more than that he can "see dimly" why probability coordination is rational. Howson and Urbach (1993) attempt a full-blown justification. Strevens (1999) argues that Howson and Urbach's argument appeals implicitly to a principle of indifference, and goes on to make a case that there is a strong parallel between providing an a priori justification for probability coordination and providing an a priori justification for inductive inference, that is, solving the problem of induction.

A final question about the relation between epistemic and physical prob-

ability was adumbrated above (see *Subjectivism*): why should the same formal structure be central to our understanding of two such different things as the production of the probabilistic patterns and the nature of inductive reasoning?

## References

- Armstrong, D. (1983). *What Is a Law of Nature?* Cambridge University Press, Cambridge.
- Braithwaite, R. B. (1953). *Scientific Explanation*. Cambridge University Press, Cambridge.
- Carnap, R. (1950). *Logical Foundations of Probability*. Chicago University Press, Chicago.
- Daston, L. (1988). *Classical Probability in the Enlightenment*. Princeton University Press, Princeton, NJ.
- Fetzer, J. (1971). Dispositional probabilities. *Boston Studies in the Philosophy of Science* 8:473–482.
- Fine, T. (1973). *Theories of Probability*. Academic Press, New York.
- de Finetti, B. (1964). Foresight: its logical laws, its subjective sources. In H. Kyburg and H. Smokler (eds.), *Studies in Subjective Probability*. Wiley, New York.
- van Fraassen, B. (1980). *The Scientific Image*. Oxford University Press, Oxford.
- . (1989). *Laws and Symmetries*. Oxford University Press, Oxford.
- Giere, R. (1973). Objective single case probabilities and the foundation of statistics. In Suppes et al. (1973).
- Gillies, D. A. (2000). *Philosophical Theories of Probability*. Routledge, London.
- Hacking, I. (1975). *The Emergence of Probability*. Cambridge University Press, Cambridge.

- Hájek, A. (2003). What conditional probability could not be. *Synthese* 137:273–323.
- Howson, C. and P. Urbach. (1993). *Scientific Reasoning: The Bayesian Approach*. Second edition. Open Court, Chicago.
- Jaynes, E. T. (1983). *Papers on Probability, Statistics, and Statistical Physics*. Edited by R. Rosenkrantz. D. Reidel, Dordrecht.
- Jeffrey, R. C. (1983). *The Logic of Decision*. Second edition. Chicago University Press, Chicago.
- Keynes, J. M. (1921). *A Treatise on Probability*. Macmillan, London.
- Laplace, P. S. (1902). *A Philosophical Essay on Probabilities*. Translated by F. W. Truscott and F. L. Emory. Wiley, New York.
- Lewis, D. (1980). A subjectivist's guide to objective chance. In R. C. Jeffrey (ed.), *Studies in Inductive Logic and Probability*, volume 2. University of California Press, Berkeley, CA.
- . (1994). Humean supervenience debugged. *Mind* 103:473–490.
- Loewer, B. (2001). Determinism and chance. *Studies in History and Philosophy of Modern Physics* 32:609–620.
- Mellor, D. H. (1971). *The Matter of Chance*. Cambridge University Press, Cambridge.
- von Mises, R. (1957). *Probability, Statistics and Truth*. Macmillan, New York. Second revised English edition, prepared by Hilda Geiringer.
- Popper, K. (1959). The propensity interpretation of probability. *British Journal for the Philosophy of Science* 10:25–42.

- Ramsey, F. (1931). Truth and probability. Reprinted in *Philosophical Papers*. Edited by D. H. Mellor. Cambridge University Press, Cambridge.
- Reichenbach, H. (1949). *The Theory of Probability*. University of California Press, Berkeley, CA.
- Skyrms, B. (1980). *Causal Necessity*. Yale University Press, New Haven, CT.
- Strevens, M. (1995). A closer look at the 'New' Principle. *British Journal for the Philosophy of Science* 46:545–561.
- . (1998). Inferring probabilities from symmetries. *Noûs* 32:231–246.
- . (1999). Objective probabilities as a guide to the world. *Philosophical Studies* 95:243–275.
- . (2003). *Bigger than Chaos: Understanding Complexity through Probability*. Harvard University Press, Cambridge, MA.
- Suppes, P. (1973). New foundations of objective probability: Axioms for propensities. In Suppes et al. (1973).
- Suppes, P., L. Henkin, G. C. Moisil, and A. Joja (eds.). (1973). *Logic, Methodology and Philosophy of Science IV: Proceedings of the Fourth International Congress for Logic, Methodology and Philosophy of Science, Bucharest, 1971*. North Holland, Amsterdam.