Rosenberg, Alexander (1983), *The Structure of Biological Science*. Cambridge: Cambridge University Press.
——— (1993), "*Scientific Innovation and the Limits of Social Scientific Prediction*," *Synthese* 97: 161–182.
——— (1994), *Instrumental Biology or the Disunity of Science*. Chicago: University of Chicago Press.
Scerri, Eric R., and John Worrall (2001), "*Prediction and the Periodic Table*," *Studies in History and Philosophy of Science* 32A: 407–452.
Schlesinger, George N. (1987), "Accommodation and Prediction," *Australasian Journal of Philosophy* 65: 33–42.
van Fraassen, Bas (1980), *The Scientific Image*. Oxford: Clarendon Press.
Williams, Mary B. (1982), "The Importance of Prediction Testing in Evolutionary Biology," *Erkenntnis* 17: 291–306.
Worrall, John (1978), "*The Ways in Which the Methodology of Scientific Research Programmes Improve on Popper's Methodology*," in G. Radnitzky and G. Andersson (eds.), *Boston Studies in the Philosophy of Science*, vol. 58: *Progress and Rationality in Science*. Dordrecht, Netherlands: Reidel, 321–338.
——— (1989), "*Fresnel, Poisson and the White Spot: The Role of Successful Predictions in the Acceptance of Scientific Theories*," in D. Gooding, T. Pinch and S. Shaffer (eds.), *The Uses of Experiment: Studies in the Natural Sciences*. Cambridge: Cambridge University Press, 135–157.
Zahar, Elie (1973), "Why Did Einstein's Programme Supersede Lorentz's?" parts I and II, *British Journal for the Philosophy of Science* 24: 95–125, 223–262.

*See also* **Bayesianism; Causality; Confirmation Theory; Determinism; Empiricism; Explanation; Induction, Problem of; Laws of Nature; Logical Empiricism; Phenomenalism; Popper, Karl Raimund; Probability; Reichenbach, Hans; Theories; Verificationism**

# PROBABILITY

There are two central questions concerning probability. First, what are its formal features? That is a *mathematical* question, to which there is a standard, widely, though not universally, agreed upon answer (reviewed in the next section). Second, what sorts of things *are* probabilities—what, that is, is the *subject matter* of probability theory? This is a *philosophical* question, and while the mathematical theory of probability certainly bears on it, the answer must come from elsewhere. To see why, observe that there are many things in the world that have the *mathematical structure* of probabilities (e.g., the set of measurable regions on the surface of a table) but that would never be mistaken for *being* probabilities. So probability is distinguished by more than just its formal characteristics. The bulk of this essay will be taken up with the central question of what this "more" might be.

## Kolmogorov's Axiomatization

Probability theory was inspired by games of chance in seventeenth-century France and inaugurated by the Fermat–Pascal correspondence, which culminated in the *Port Royal Logic* (Arnauld [1662] 1964). Its axiomatization had to wait nearly another three centuries. The locus classicus of the mathematical theory of probability is Kolmogorov's ([1933] 1950) *Foundations of Probability*. Inspired by measure theory, Kolmogorov's axiomatization has become orthodoxy. Let $\Omega$ be a nonempty set. A *field* (*algebra*) on $\Omega$ is a set $F$ of subsets of $\Omega$ that has $\Omega$ as a member and that is closed under complementation (with respect to $\Omega$) and union. Assume for now that $F$ is finite. Let $P$ be a function from $F$ to the real numbers, obeying the following axioms:

$$P(a) \geq 0 \text{ for all } a \in F, \qquad \text{A1}$$

$$P(\Omega) = 1, \text{and} \qquad \text{A2}$$

$$P(a\, \Omega\, b) = P(a) + P(b)$$
$$\text{for all } a, b \in F \text{ such that } a \cap b = \phi. \qquad \text{A3}$$

Call $P$ a probability function, and ($\Omega$, $F$, $P$) a probability space.

One could instead attach probabilities to members of a collection of *sentences* of a formal language, closed under truth-functional combinations. Either way, a kind of reflective equilibrium is achieved between these axioms, which are thought to be intuitively plausible, and various important interpretations of probability (to be discussed in the subsequent sections), which obey them and bring them to life in applications.

It is often thought that the only nonconventional part of the axiomatization is A3. That is too quick, for it is substantive that probabilities are

1. Defined by *functions* (rather than by one-many or many-many mappings);
2. Functions of *one variable* (unlike primitive conditional probability functions, which are functions of two variables);
3. Defined on a *field* (rather than a set with weaker closure conditions);
4. Represented *numerically* (rather than qualitatively, as is "possibility"; or comparatively, as is "similarity to a given world" in the Stalnaker/Lewis style of semantics for counterfactuals [Lewis 1986a]);
5. *Real* numbers (rather than those of some other number system);
6. *Bounded* (unlike other quantities that are treated measure-theoretically, such as lengths);
7. Bounded by *Maximal* and *minimal* values (thus prohibiting open or half-open ranges).

For a discussion of rival theories that relax or replace points 2, 3, 4, and 6 above, see Fine 1973. Complex-valued probabilities are proposed by Feynman and Cox (Mückenheim et al. 1986); infinitesimal probabilities (of nonstandard analysis) by Skyrms (1980) and Lewis (1986b), among others; and unbounded probabilities by Renyi (1970). Primitive conditional probability functions will be briefly discussed at the end of this section.

Kolmogorov extends his axiomatization to cover infinite probability spaces. Probabilities are now defined on a *σ-field* (*σ-algebra*)—a field that is further closed under *countable* unions—and A3 is correspondingly strengthened:

A3′ (Countable additivity) If $a_1, a_2, a_3, \ldots$ is a countable sequence of (pairwise) disjoint sets, each belonging to $F$, then $P(\bigcup_{n=1}^{\infty} a_n) = \sum_{n=1}^{\infty} P(a_n)$.

De Finetti (1990) is a notable opponent of countable additivity.

Kolmogorov then defines the conditional probability of $a$ given $b$ by the ratio of unconditional probabilities:

$$P(a|b) = \frac{P(a \cap b)}{P(b)}, \text{provided } P(b) > 0.$$

Note that this ratio is undefined if either or both of the unconditional probabilities are undefined, or if $P(b) = 0$. Yet in uncountable spaces there can be genuine, nontrivial events whose probabilities are undefined (so-called "nonmeasurable" sets), and others whose probabilities are 0 ("probability 0 does not imply impossible," as textbooks and Kolmogorov himself caution us). So Kolmogorov's definition does not guarantee that certain intuitive constraints on conditional probability are met—for example, that the probability of an event, *given itself*, is 1.

Kolmogorov addresses the probability-0 problem with a more sophisticated account of conditional probability as a random variable conditional on a sigma algebra, appealing to the Radon-Nikodym theorem to guarantee the existence of such a random variable (see, e.g., Billingsley 1995). A rival approach takes conditional probability $P(\_,\_)$ as primitive and defines the unconditional probability of $a$ as $P(a, \mathbf{T})$, where $\mathbf{T}$ is a necessary (e.g., tautological) proposition. Various axiomatizations of primitive conditional probability have been defended in the literature, typically differing only in the handling of conditional probabilities with zero unconditional probability antecedents. In many ways, the most general and elegant of the proposed axiomatizations is Popper's (1959). (See Roeper and Leblanc 1999 for an encyclopedic discussion of competing theories of conditional probability, and Keynes 1921, Carnap 1950, Popper 1959, and Hájek 2003b for arguments that probability is inherently a two-place function.)

Versions of Bayes's theorem can now be proven (see Bayesianism):

$$P(a \mid b) = \frac{P(b \mid a)P(a)}{P(b)} = \frac{P(b \mid a)P(a)}{P(b \mid a)P(a) + P(b \mid \neg a)P(\neg a)}.$$

More generally, suppose there is a partition of hypotheses $\{h_1, h_2, \ldots, h_n\}$ and evidence $e$. Then for each $i$,

$$P(h_i \mid e) = \frac{P(e \mid h_i)P(h_i)}{\sum_{j=1}^{n} P(e \mid h_j)P(h_j)}.$$

The $P(e|h_i)$ terms are called likelihoods, and the $P(h_i)$ terms are called priors. Finally, Kolmogorov defines $a$ and $b$ to be *independent* iff (if and only if) $P(a|b) = P(a)$; equivalently, iff $P(b|a) = P(b)$; equivalently, iff $P(a \cap b) = P(a)P(b)$ (for $P(a) \neq 0 \neq P(b)$). The terminology suppresses the fact that such independence is really a *three*-place relation between an event, another event, and a *probability function.* This distinguishes probabilistic independence from such two-place relations as logical, causal, and counterfactual independence.

The next section turns to the so-called *interpretations* of probability—attempts to answer the central philosophical question: What is probability?

## Frequentism

Ask a scientist what probability is, and one will typically get a *frequentist* answer: The probability of an event is the relative frequency of trials of a repeatable experiment on which that event occurs; sometimes the words "in the long run" are added. This leaves open important questions: Which are the trials to be counted? How long does the run have to be? One may confine one's attention to *actual* trials, realized in this world, or countenance *hypothetical* trials. And one may have merely finitely many trials to contend with or infinitely many, in which case probability will be identified with the *limit* of the relative frequency in a *sequence* of trials. One may thus immediately distinguish $2 \times 2 = 4$ variants of frequentism. However, the actual world typically delivers only finitely many trials of any given experiment. And it is often thought that if one is going to allow the trials to be hypothetical anyway, there is no obstacle to letting the sequence of trials be infinite, thus *guaranteeing* a "long run." So one may confine one's attention, as frequentists typically do, to just two of the possible positions: finite actual frequentism and infinite hypothetical frequentism.

In his discussion of the proportion of births of males and females, Venn (1866) contends that "probability is nothing but that proportion" (84)—a version of finite actual frequentism. Von Mises (1957), by contrast, insists that probabilities exist only relative to virtual infinite sequences of "attributes" called *collectives*. In a collective, the limiting relative frequency of any attribute exists and is the same on any recursively specified subsequence. (Von Mises' original definition, in terms of "place selections," is finessed by Church.) The probability of a given attribute, relative to a collective, is then identified with its limiting relative frequency in that collective. Von Mises' position is thus a version of infinite hypothetical frequentism, as are those of Reichenbach and van Fraassen.

Any version of frequentism faces the notorious *reference class problem.* Any event, in all its detail, occurs exactly once, so if nontrivial frequencies are to be associated with it, it must be regarded as a token of a more general event type, whose instances constitute its reference class. However, there are indefinitely many ways of typing a given event. This would not be a problem if its relative frequency were the same in each reference class, or if one such class stood out as natural or privileged. The problem gains teeth to the extent that various competing reference classes have equal claim to determining the probability and that they yield different relative frequencies for the event.

In some cases, the reference class problem may be solved for the actual, finite frequentist, but at the price of creating the equally notorious *problem of the single case:* Intuitively, the objective probability of a one-off event may be less than 1, but finite frequentism cannot respect this intuition. Many events occur only once by any reasonable standard of typing: the 2000 presidential election, the invasion of Iraq, the last Lakers–Bulls game, and so on. The only natural reference class for such an event is the singleton set consisting of itself, and thus it has relative frequency 1 (and its nonoccurrence has relative frequency 0). Nonetheless, it seems natural to think of nonextreme probabilities attaching to at least some of these "single-case" events.

The problem of the single case is particularly striking, but there is really a sequence of related "granularity" problems: the problem of the double case, the problem of the triple case, and so on. A finite reference class of size $n$ can produce relative frequencies at only a certain level of "grain," namely $\frac{1}{n}$. Among other things, this rules out irrational probabilities; yet, the best physical theories say otherwise (for example, various decay probabilities delivered by quantum mechanics are irrational). Furthermore, there is a sense in which any of these problems can be transformed into the problem of the single case. Suppose that a coin is tossed a thousand times. This can be regarded as a *single* trial of a thousand-tosses-of-the-coin experiment. Yet one does not want to be committed to saying that *that* experiment yields its actual result with probability 1.

The move to infinite hypothetical frequentism makes the reference class problem only worse, for not only must a set of events be chosen in which to place a given event, but since the set is now infinite, an *ordering* among the events must be chosen. After all, in nontrivial cases a limiting relative frequency can be made whatever value one likes simply by reordering the results of a given sequence. Consider the limiting relative frequency of even numbers among positive integers. In the "natural" ordering $<1, 2, 3, \ldots>$ it is $\frac{1}{2}$; however, one can make it $\frac{1}{4}$ by reordering the integers so that the even numbers occur at every fourth place in the sequence: $<1, 3, 5, 2, 7, 9, 11, 4, 13, \ldots>$; and so on. Thus, limiting relative frequencies are sensitive to apparently arbitrary choices of ordering, while it appears that probabilities need not be. One might call this the *reference sequence problem.*

A sequence of events is said to be *exchangeable* with respect to a given probability function if all the joint probabilities of the events are invariant under finitely many permutations of the sequence: Every event has the same probability, every conjunction of two events has the same probability, every conjunction of three events has the same probability, and so on. A sequence of events is automatically exchangeable with respect to the relative frequency function: The frequency of an event is insensitive to *which* trials the event occurs at. Yet various events intuitively are not exchangeable with respect to the relevant probability function. Consider someone learning to throw a dart at a bull's-eye: The sequence <MISS, MISS, HIT> is presumably more probable than <HIT, MISS, MISS>, because the dart thrower's accuracy improves with practice. Yet the (finite) relative frequency of HIT is $\frac{1}{3}$ either way. Since relative frequencies *force* a kind of symmetry that probabilities need not obey, they cannot be the same thing. (Ironically, it was the failure of a more thoroughgoing "infinite exchangeability" that proved to be the undoing of hypothetical infinite frequentism in the previous paragraph.)

## The Classical Interpretation

The brainchild of such founding fathers of probability as Pascal, Fermat, Huygens, and Leibniz, and clearly articulated in Laplace ([1814] 1951), the classical interpretation is the oldest interpretation of probability—indeed, it dates back to a time when the axiomatization and interpretation of probability were not clearly distinguished. It seeks to characterize the probability assignment of a rational agent in a state of epistemic neutrality with respect to a finite set of "equipossibilities": The agent has either *no evidence* or *symmetrically balanced evidence* regarding the possibilities. It appeals to the so-called principle of indifference: Whenever there is no evidence favoring one possibility over another, each should be assigned the same probability as the others. So

$$P(e) = \frac{\text{number of equipossibilities in which } e \text{ occurs}}{\text{total number of equipossibilities.}}$$

But the notion of "equipossibilities" seems to presuppose some prior notion of probability. After all, the most obvious characterization of "symmetrically balanced evidence" is in terms of equality of conditional probabilities: Given evidence $e$ and possible outcomes $o_1, o_2, \ldots, o_n$, the evidence is symmetrically balanced with respect to the outcomes iff $P(o_1|e) = P(o_2|e) = \ldots = P(o_n|e)$. Perhaps, then, one should regard the classical interpretation as an attempt to reduce *quantitative* probability to *comparative* probability: All *numerical* probabilities are ultimately based on facts about *equalities* among probabilities.

Note the structural resemblance of the classical theory to finite frequentism. Both theories see probability as a matter of evenhanded counting and ratio taking:

$$P(e) = \frac{\text{number of cases favorable to } e}{\text{total number of cases}}.$$

It is just that for frequentism, the cases are *actual* outcomes of a *repeated* experiment, whereas for the classical theory they are *possible* outcomes of a *single* experiment. And indeed the classical theory faces many of the same problems as frequentism. There is the granularity problem: Clearly, every classical probability is some fraction of the form $\frac{m}{n}$, where $n$ is the number of possibilities. There is the exchangeability problem: Classical probabilities are invariant under permutation of the labeling of the possibilities (for example, relabeling the faces of a die makes no difference to their probabilities of coming up). Thus, the classical interpretation cannot readily provide *asymmetric* probability distributions (e.g., for biased dice or coins), and it cannot handle distributions that evolve over time (e.g., for the dart thrower's hitting the bull's-eye).

Moreover, the reference class problem reappears. If one is truly ignorant about the results of some experiment, then presumably there is nothing to favor various competing choices of sample space. One should then be indifferent between, for example, {heads, tails} and {heads, tails, edge}. And one should be indifferent between various refinements of the original space: for example, between spaces that refine in different ways the heads outcome according to its final orientation relative to due north. Thus, probabilities will be determined by an apparently arbitrary choice of sample space. To adapt an example from physics: Bose-Einstein statistics, Fermi-Dirac statistics, and Maxwell-Boltzmann statistics each arise by considering the ways in which particles can be assigned to states and then partitioning the set of alternatives in different ways (see, e.g., Fine 1973). Someone ignorant of which statistics apply to a given type of particle can make only an arbitrary choice and hope for the best.

In typical applications of the classical theory (gambling, for example), one is not wholly ignorant, but the evidence that one has is symmetrically balanced regarding the possibilities. There are two problems here: in the evidence and in the symmetry. Classical probabilities are acutely sensitive to the evidence. If the evidence becomes *unbalanced*, favoring some outcomes over others, then classical probabilities are not merely revised; they are *destroyed.* And there may be competing respects of symmetry, each equally compelling. This problem arises especially when there are infinitely many possible outcomes. Then, the equipossibilities must be a finite partition of the outcomes. But *which* partition?

A tempting answer may be: the most "natural" partition. However, "Bertrand's paradoxes" show that there need not be any. The trick is to give competing parameterizations of a given problem that are nonlinearly related to one another but equally natural. Suppose one is told only that a car traveled 100 miles at an average speed between 50 and 100 mph. What is the probability that its average speed was between 75 and 100 mph? Perhaps 0.5, since (50, 75) and (75, 100) are equipossible intervals for the average speed. But the question could be equivalently formulated: A car took between 1 and 2 hours to travel 100 miles. What is the probability that it took between 1 hour and $1\frac{1}{3}$ hours? Now it seems that there are three equipossible intervals for the time taken: $(1, 1\frac{1}{3})$, $(1\frac{1}{3}, 1\frac{2}{3})$, and $(1\frac{2}{3}, 2)$; whence the answer should be $\frac{1}{3}$.

## Logical Probability

Many philosophers—Leibniz, von Kries, Keynes, Wittgenstein, Waismann, Carnap, and others—have tried to explicate the following "logical" concept of conditional probability:

$$P(p \mid q) = \frac{\text{the proportion of logically possible worlds in which both } p \text{ and } q \text{ are true}}{\text{the proportion of logically possible worlds in which } q \text{ is true}}.$$

An obvious problem has been to justify a *measure* of the "proportion of logically possible worlds in which a proposition is true." Early attempts (including those by Carnap that will be the focus here) tried to apply the controversial principle of indifference (see Carnap, Rudolf). Carnap's (1950) early constructions are very similar to systems developed earlier by W. E. Johnson (1921) (see Inductive Logic for further references on Carnapian inductive logic and logical probability).

Begin with a first-order language $L$ containing a finite number of monadic predicates: $F, G, H, \ldots$, and a finite or denumerable number of individual constants $a, b, c, \ldots$ . Then define an (a priori) unconditional probability function $P(\bullet)$ over the sentences of $L$, in a way that appeals to only their *syntactic structure* (whence the name "logical" probability). Finally, use the standard ratio definition to construct a conditional probability function $P(\bullet \mid \bullet)$ over pairs of sentences of $L$.

The results of this procedure will be *language relative:* If one describes the same phenomena by means of a different language $L^*$—equipped with a different stock of monadic predicates—one will typically not recover the same probabilities. Consider two languages used to represent the outcomes of random draws from an urn filled with colored balls. Let $L$ contain the color predicates "blue" and "green" and let $L^*$ contain the predicates "grue" and "bleen." The intended interpretation is that a draw is grue just in case it is one of the first million and green or a later one and blue; a draw is bleen just in case it is one of the first million and blue or a later one and green. Starting with $L$, use whatever is the appropriate procedure to calculate

$P$(draw 1,000,001 is green | the first 1,000,000 draws are green).

Starting with $L^*$, use this procedure to calculate

$P$(draw 1,000,001 is grue | the first 1,000,000 draws are grue).

If syntax is all that matters, then these conditional probability values will be identical—and surely greater than $\frac{1}{2}$, at least if logical probability is to have a hope of modeling actual inductive reasoning (see Inductive Logic). The trouble is that the second conditional probability, translated into $L$, is just

$P$(draw 1,000,001 is blue | the first 1,000,000 draws are green).

One can avoid contradiction, but only by explicitly insisting that probability is language-relative. And that raises a serious problem—really, the reference class problem in a new guise: If one wishes to employ logical probability as a foundation for inductive inference, which is the "right" language to use? The remainder of this discussion will presuppose that an answer to this question has been found (for Carnap [1980], this question was "external" to inductive logic anyway, and his later systems did not have this blatant form of language relativity).

Returning now to Carnap's early systems, consider a simple language with only two monadic predicates $F$ and $G$ and only two individual constants $a$ and $b$. This language yields exactly sixteen maximally specific descriptions of the world—the *state descriptions* of $L$: $(Fa \wedge Ga \wedge Fb \wedge Gb)$, $(Fa \wedge Ga \wedge Fb \wedge \neg Gb)$, etc. Two state descriptions $S_1$ and $S_2$ are *permutations* of each other if $S_1$ can be obtained from $S_2$ by some permutation of the individual constants. For example, $Fa \wedge \neg Ga \wedge \neg Fb \wedge Gb$ and $\neg Fa \wedge Ga \wedge Fb \wedge \neg Gb$ are permutations of each other. A *structure description* in $L$ is a disjunction of state descriptions, closed under permutation. The $L$ language provides these ten structure descriptions:

| | |
|---|---|
| $Fa \wedge Ga \wedge Fb \wedge Gb$ | $(Fa \wedge \neg Ga \wedge \neg Fb \wedge Gb) \vee (\neg Fa \wedge Ga \wedge Fb \wedge \neg Gb)$ |
| $(Fa \wedge Ga \wedge Fb \wedge \neg Gb) \vee (Fa \wedge \neg Ga \wedge Fb \wedge Gb)$ | $(Fa \wedge \neg Ga \wedge \neg Fb \wedge \neg Gb) \vee (\neg Fa \wedge \neg Ga \wedge Fb \wedge \neg Gb)$ |
| $(Fa \wedge Ga \wedge \neg Fb \wedge Gb) \vee (\neg Fa \wedge Ga \wedge Fb \wedge Gb)$ | $\neg Fa \wedge Ga \wedge \neg Fb \wedge Gb$ |
| $Fa \wedge Ga \wedge \neg Fb \wedge \neg Gb) \vee (\neg Fa \wedge \neg Ga \wedge Fb \wedge Gb)$ | $(\neg Fa \wedge Ga \wedge \neg Fb \wedge \neg Gb) \vee (\neg Fa \wedge \neg Ga \wedge \neg Fb \wedge Gb)$ |
| $Fa \wedge \neg Ga \wedge Fb \wedge \neg Gb$ | $\neg Fa \wedge \neg Ga \wedge \neg Fb \wedge \neg Gb$ |

Now, assign nonnegative real numbers to the state descriptions, so that these sixteen numbers sum to 1. Any such assignment will constitute an (a priori) unconditional probability function $P(\bullet)$ over the state descriptions of $L$. To extend $P(\bullet)$ to the entire language $L$, note that the probability of a disjunction of mutually exclusive sentences is the sum of the probabilities of its disjuncts. Since every sentence in $L$ is equivalent to some disjunction of state descriptions, and all the state descriptions are mutually exclusive, this gives a complete unconditional probability function $P(\bullet)$ over $L$—typically called a *measure function.* The standard ratio definition then yields a conditional probability function $P(\bullet|\bullet)$ over pairs of sentences in $L$. Carnap (1950) discusses two natural measure functions. The first, $m\dagger$, treats each state description as equiprobable a priori: If there are $N$ state descriptions in $L$, then $m^{\dagger}$ assigns $\frac{1}{N}$ to each. However natural this measure function may seem, it has the consequence that the resulting probabilities cannot undergird *learning from experience.* To see why, observe that

$$P(Fb \mid Fa) = \frac{m^{\dagger}(Fb \wedge Fa)}{m^{\dagger}(Fa)} = \frac{1}{2} = m^{\dagger}(Fb) = P(Fb).$$

So "learning" that one object has property $F$ cannot affect the probability that any other object will also have property $F$. Indeed, it can be shown that *no matter how many objects are assumed to be F*, this will (according to probability functions based on $m^{\dagger}$) *always be irrelevant* to the hypothesis that a distinct object will also be $F$—a feature widely viewed as a serious shortcoming of $m^{\dagger}$.

As a result, Carnap formulated an alternative measure function $m^*$: First, assign equal probabilities to each *structure* description. Then, each state description entailing a given structure description is assigned an equal portion of the probability assigned to that structure description. So, in the present toy language, the state description $Fa \wedge Ga \wedge \neg Fb \wedge Gb$ gets assigned an a priori probability of $\frac{1}{20}$ ($\frac{1}{2}$ of $\frac{1}{10}$), but the state description $Fa \wedge Ga \wedge Fb \wedge Gb$ receives an a priori probability of $\frac{1}{10}$ ($\frac{1}{1}$ of $\frac{1}{10}$). Unlike $m^{\dagger}$, $m^*$ *does* allow for learning from experience; for example, $P(Fa \mid Fb) = \frac{3}{5} > \frac{1}{2} = P(Fa)$. Still, even $m^*$ can give unintuitive results in more complex languages (see Carnap 1952 for discussion). Also, note that the state descriptions are exchangeable with respect to $m^*$, an omen that logical probabilities will face some of the problems that plagued the frequentist and the classical probabilist.

Carnap (1952) presents a more complicated "continuum" of conditional probability functions. This continuum depends on a parameter $\lambda$ intended to reflect the "speed" with which learning from experience is possible. $\lambda = 0$ corresponds to the "straight rule," which says that the probability that the next object observed will be *F,* conditional upon a sequence of past observations, is simply the frequency of $F$ objects in that sequence; $\lambda = +\infty$ yields a conditional probability function much like that derived from the measure function $m^{\dagger}$ (i.e., $\lambda = +\infty$ implies that there is no learning from experience); $\lambda = \kappa$ (which is the number of independent families of predicates in Carnap's more elaborate [1952] linguistic framework) yields a conditional probability function equivalent to that generated by the measure function $m^*$.

Problems remain. None of the Carnapian systems allow universal generalizations to have nonzero probability. Carnap's early systems also failed to allow for *analogical effects*, since in these systems the fact that two objects have several properties in common is (in many cases) *irrelevant* to whether they have any *other* properties in common. Carnap's most recent (and most complex) theories of logical probability (1980) include two additional parameters designed to provide the theory with enough flexibility to overcome these (and other) limitations. Unfortunately, no Carnapian logical theory of probability to date has dealt successfully with the problem of analogical effects (see Maher 2001 for further discussion). The consensus now seems to be that the Carnapian project of constructing an adequate logical theory of probability is all but hopeless: The syntactical constraints implicit in any such theory will inevitably prevent the theory from being able to model certain essential features of statistical inference and/or inductive logic (see Inductive Logic).

## Subjectivism

In slogan form, subjectivism regards probabilities as *degrees of belief,* or *credences.* But what are credences? Subjectivists since Ramsey ([1926] 1980) have insisted that they must be intimately tied to the behavioral dispositions of suitable agents. In one influential account, advocated by de Finetti ([1937] 1980):

> An agent's credence in $e$ is $p$ iff $p$ units of utility is the price at which the agent would buy or sell a bet that pays 1 unit of utility if $e$, 0 if $\neg e$.

This is at best a first approximation to an analysis of credence. One surely should allow the buying and selling prices of at least some bets to come apart. And even when they agree, there are problems. How does one separate the agent's epistemic attitude to $e$ from his or her attitude (favorable, unfavorable, or neutral) to gambling? Indeed, one may insist on separating epistemic attitudes from desire-based attitudes altogether; one can imagine, for example, a chronic apathetic who has opinions but lacks corresponding desires (for bets or for anything). Moreover, the very placement of the bet may change the world in ways that affect the agent's credences.

Be that as it may, there are famous arguments that credences must conform to the probability calculus, at least if one demands that the agent be in some sense *ideally rational.* For example, if one's credences do *not* so conform, one is susceptible to a **Dutch Book**, a sequence of bets that one regards as acceptable taken individually but that collectively guarantee one's loss, however the world turns out. Conversely, if one's credences *do* so conform, one is immune to a Dutch Book. Rationality, it is concluded, requires obedience to the probability calculus (see Dutch Book Argument).

Utilities (desirabilities) of outcomes, their probabilities, and rational preferences are all intimately linked. The *Port Royal Logic* (Arnauld [1662] 1964) showed how utilities and probabilities together determine rational preferences; de Finetti's betting interpretation derives probabilities from utilities and rational preferences; von Neumann and Morgenstern (1944) derive utilities from probabilities and rational preferences. And most remarkably, Ramsey ([1926] 1980) (and later, Savage 1954 and Jeffrey 1983) derives *both* probabilities *and* utilities from rational preferences alone (see Ramsey, Frank Plumpton).

First, Ramsey defines a proposition to be *ethically neutral*—relative to an agent and an outcome—if the agent is indifferent between having that outcome when the proposition is true and when it is false. Suppose that the agent prefers $a$ to $b$. Then an ethically neutral proposition $n$ has probability $\frac{1}{2}$ iff the agent is indifferent between the gambles

> $a$ if $n$, $b$ if not.
> $b$ if $n$, $a$ if not.

One may assign arbitrarily to $a$ and $b$ any two real numbers $u(a)$ and $u(b)$ such that $u(a) > u(b)$, thought of as their respective desirabilities. Having done this for the one arbitrarily chosen pair $a$ and $b$, the utilities of all other propositions are determined. Given various assumptions about the richness of the preference space, and certain "consistency assumptions," Ramsey can define a real-valued utility function of the outcomes $a$, $b$, etc.—in fact, various such functions will represent the agent's preferences. He is then able to define equality of differences in utility for any outcomes over which the agent has preferences. It turns out that ratios of utility differences are invariant—the same whichever representative utility function one chooses. This fact allows Ramsey to define degrees of belief as ratios of such differences. For example, suppose the agent is indifferent between $a$ and the gamble "$b$ if $x$, $c$ otherwise." Then his or her degree of belief in $x$, $P(x)$, is given by:

$$P(x) = \frac{u(a) - u(c)}{u(b) - u(c)}.$$

Ramsey shows that degrees of belief so derived obey the probability calculus (with finite additivity). He calls what results "the logic of partial belief."

Ramsey avoids some of the objections to the betting interpretation, but not all of them. Notably, the essential appeal to gambles again raises the concern that the wrong quantities are being measured. And his account has new difficulties. It is unclear what facts about agents fix their preference rankings. It is also dubious that *consistency* alone requires one to have a set of preferences as rich as Ramsey requires, or that one can find ethically neutral propositions of probability $\frac{1}{2}$ This in turn casts some doubt on Ramsey's claim to assimilate probability theory to logic.

Savage (1954) likewise derives probabilities and utilities from preferences among options that are constrained by certain putative "rationality" principles. For a given set of such preferences, he generates a class of utility functions, each a positive linear transformation of the other (i.e., of the form $u_1 = au_2 + b$, where $a > 0$) and a unique probability function. Together these are said to "represent" the agent's preferences. Jeffrey (1983) refines the method further. The result is theory of decision according to which rational choice maximizes "expected utility," a certain probability-weighted average of utilities.

So far, this is a static picture of a rational agent. How should one update one's degrees of belief in the light of new evidence? The favored rule among subjectivists is *conditionalization:* Where $e$ is the strongest proposition of which one becomes certain, one's new credence function is related to the old by:

$$\text{(Conditionalization)} \quad C_{new}(\bullet) = C_{old}(\bullet \mid e),$$

using $C(\bullet)$ here and in what follows to distinguish credence from other kinds of probability.

So-called *subjective Bayesianism* holds that an agent's epistemic trajectory is rational iff the agent's credences are representable at any moment by a probability function and the agent always updates by conditionalization. This is at once a highly demanding and highly permissive epistemology. It is demanding because conformity to probability theory is demanding. It is permissive because there is no requirement that degrees of belief in any way correspond to the way the world is. So someone who assigns probability 1 to the universe being ruled by a rubber chicken can meet the Bayesian standards for rationality—as long as the agent obeys the probability calculus in all other assignments and always updates by conditionalizing. Bayesians reply that various convergence theorems show roughly that in the long run, agents who do not give probability 0 to genuine possibilities, and whose stream of evidence is sufficiently rich, will eventually be arbitrarily close to being certain regarding the truth about the world in which they live. For skepticism about the value of these theorems, see Earman (1992).

In any case, there are numerous proposals for further constraints on priors. Some (e.g., Jeffreys and Jaynes) appeal to a version of the principle of indifference. Some can be regarded as instances of a certain schema, proposed by Gaifman (1988). He coins the term "expert probability" for a probability assignment that a given agent strives to track, codifying this idea as follows:

$$\text{(Expert)} \quad C(a \mid pr(a) = x) = x, \text{ for all } x \text{ such that } C(pr(a) = x) > 0.$$

Here $pr(a)$ is the assignment that the agent regards as expert. For example, if one regards the local weather forecaster as an expert, and he or she assigns probability 0.1 to it raining tomorrow, then one may well follow suit:

$$C(\text{rain} \mid pr(\text{rain}) = 0.1) = 0.1.$$

More generally, one might speak of an entire probability function as being such a guide for an agent, over a specified set of propositions—so that (Expert) holds for any choice of $A$ from that set. A *universal expert function* would guide *all* of the agent's probability assignments in this way. Van Fraassen (1995) argues that an agent's *future* probability functions are universal expert functions for that agent—his *reflection principle* is:

$$C_t(a \mid C_{t'}(a) = x) = x, \text{ for all } a \text{ and for all } x \text{ such that } C_t(C_{t'}(a) = x) > 0,$$

where $C_t$ is the agent's probability function at time $t$, and $C_{t'}$ his or her function at a later time $t'$. The principle encapsulates a certain demand for "diachronic coherence" imposed by rationality. Van Fraassen defends it with a diachronic Dutch Book argument (one that considers bets placed at different times) and by analogizing violations of it to the sort of pragmatic inconsistency that one finds in Moore's paradox. For example, suppose an agent is certain that he will tomorrow assign probability $\frac{1}{2}$ to it raining the day after but that he nonetheless assigns it probability $\frac{1}{3}$ now. While this is not logically inconsistent, it is surely puzzling.

One may go still further. There may be universal expert functions for all rational agents. The *principle of direct probability* regards the relative frequency function as a universal expert function. Let $a$ be an event type, and let $relfreq(a)$ be the relative frequency of $a$ (in some suitable reference class). Then for any rational agent:

$$C(a \mid relfreq(a) = x) = x, \text{ for all } a \text{ and for all } x \text{ such that } C(relfreq(a) = x) > 0.$$

The next section takes up what many consider the most important such universal expert function.

## Objective Chance

De Finetti (1990, x), the great probabilist, quipped that "probability does not exist." What he meant was that all probability is subjective. Yet there is a strong *prima facie* case for recognizing the existence of *objective chances:* probabilities that attach to physical systems and their behavior independently of anyone's mental state and that capture *contingent* facts about those systems, not merely quasi-logical relations among propositions concerning them. One wonders, for example, whether a certain coin is biased, and if so, to what degree and in what direction. Translation: One wonders what the *chance of heads* would be were the coin tossed fairly.

This example would not faze a committed subjectivist such as de Finetti or a frequentist like von Mises (1957), who denies that probability ever applies in the single case. But more serious examples from physics suggest that ultimately, resistance is futile. For starters, *statistical physics* says that the entire universe *could* evolve toward a state of lower entropy but that the chance of its doing so is vanishingly small. Such chances are seemingly compatible with determinism at the level of the fundamental dynamical laws (this is controversial; see Irreversibility; Statistical Mechanics), and following Bernoulli, various authors have for that reason doubted their credentials (however, see Levi 1990 for a valuable discussion of authors since Cournot and Venn who countenance such compatibilist chances). The compatibility is putatively secured by relativizing chances to a kind of trial. For example, a coin may have chance $\frac{1}{2}$ of landing heads relative to the specification of being tossed high above a flat surface, while it has a chance 1 of landing heads relative to a precise specification of the initial conditions of a particular toss in a deterministic world. Levi argues that the applicability of chance hypotheses and statistical techniques does not presuppose an underlying indeterminism, and so a theory of chance should remain neutral vis-à-vis determinism.

In any case, commitment to chances has a second source: collapse theories of quantum mechanics. These theories explicitly introduce indeterministic dynamical laws that not only specify what courses of evolution are possible for a given physical system with a given initial state, but also specify exact *probabilities* for each such trajectory (see Quantum Measurement Problem; Quantum Mechanics). The subjectivist or von Mises–style frequentist seems left only with the option of denying—from the armchair!—that the physical theories that postulate them are true or coherent. It is better simply to acknowledge that objective chances are or at least could be real, and then to go on to consider what sort of account one could give of them.

Reductionist accounts attempt to reduce facts about objective chances to the totality of nonmodal facts about a world. Nonreductionist accounts deny that chance even supervenes on the nonmodal facts. Actual frequentism is clearly a reductionist view. More sophisticated is Lewis's (1994) "best systems" approach, which sees the laws for a world $W$ as, roughly, being theorems of that axiomatic system for describing nonmodal facts about $W$ that achieves an optimal balance of simplicity and informativeness. In the case of probabilistic laws, Lewis invokes a third criterion: A system for $W$ is "better" to the extent that it assigns a higher probability to the total history of $W$. The simplest form of nonreductionism is *primitivism*, which takes chances to be unanalyzable features of the world. Alternatively, one might try to explain one's nonreductionist chances by appeal to some other bit of metaphysical gadgetry, such as Armstrong's interpretation of chances as consisting in higher-order relations of "probabilification" that obtain between universals.

No mention has been made so far of "propensity" interpretations of probability. Everyone, reductionist and nonreductionist alike, can agree that in a chancy world, some physical systems will have propensities to exhibit certain behaviors under certain conditions, for all one need *mean* by that is that counterfactuals of the following form are true of these systems: "Were the system in conditions $C$, there would be a chance of $x$ that it would manifest behavior $B$." So propensities—understood as tendencies, or variable-strength dispositions—can be analyzed straightforwardly in terms of subjunctive conditionals whose consequents make reference to objective chances.

Propensity interpretations of probability aim to reverse this order of analysis, explaining objective chances directly in terms of propensities. For some authors (Popper, Gillies), chances are dispositions for a chance setup to produce long-run relative frequencies; for others (Giere, Fetzer, Miller), they are dispositions for a chance setup to produce outcomes on single trials. Subtle variations can be found in the work of Hacking, Mellor, and Levi (see Gillies 2000 and Hájek 2003a for surveys.)

But there are some general problems that any propensity account faces. Suppose that system $S$ has a certain tendency to manifest behavior $B$ under conditions $C$. One must be able to attach numbers to such a tendency as a measure of its strength *without* appealing to the concept of chance; it is not clear how this is to be done or why the results should obey the probability calculus. Moreover, how do propensities for distinct systems yield propensities for the composite systems they make up? Here are two coins, each with a propensity of 0.5 of landing heads if tossed. Suppose both are tossed at once. If there is a chance that both will land heads, then there must be a propensity possessed by the *combined* two-coin system. If so, what guarantees that the marginal probabilities (for each coin considered separately) will be recovered correctly from this composite propensity? And one cannot stop here, but had better say that *the world as a whole* exhibits, at each moment, propensities to evolve in various different ways. Having gone thus far, one might as well simply say that instead of exhibiting propensities, the world exhibits *chances,* thus avoiding (by stipulation) the original problem of their conformity to the probability calculus—and thus arriving at primitivism about chance. If that is right, then it is not clear that propensity accounts offer a genuinely new option for understanding probability.

Although distinct, objective and subjective probability display an extremely important connection. Lewis (1980) formulates it in his principal principle:

$$\text{(PP)} \quad C_0(a \mid e \wedge ch_t(a) = x) = x.$$

Here $C_0$ is some reasonable "initial" (a priori) credence function; $a$ an arbitrary proposition; $ch_t(a) = x$ the claim that the chance, at time $t$, of $a$ is $x$; and $e$ an "admissible" proposition—one that does not contain information relevant to $a$ beyond that given by its chance at $t$ (thus, e.g., $a$ itself is inadmissible).

One can apply (PP) to a non-initial agent by modeling credence $C$ as the result of conditionalizing some reasonable initial credence $C_0$ on some suitable evidence. Let $h$ describe a complete possible course of history until time $t$. Let $l$ describe some possible fundamental laws compatible with $h$, and assume that the way in which chances depend on history is underwritten by these laws. Then the conjunction $h \wedge l$ picks out a unique chance distribution $P(\bullet)$ for time $t$. Thus, if a proposition of the form $(h \wedge l \wedge ch_t(a) = x)$ is consistent, then the third conjunct is entailed by the first two. Assuming, as seems reasonable, that the conjunction $h \wedge l$ is admissible, it follows that

$$\text{(PP}^*) \quad C_0(a \mid h \wedge l) = P(a).$$

Much of the debate between reductionists and nonreductionists consists in a war of intuitions. For instance, the reductionist claims to find the nonreductionist's extra, irreducibly modal feature of metaphysical reality unintelligible, while the nonreductionist claims to "show" that distinct chances can give rise to exactly the same total histories of nonmodal fact—a draw, perhaps. But (PP) and (PP*) appear to open up new lines of argument.

The nonreductionist alleges that reductionism is *inconsistent* with (PP*). Typical reductionist views will allow that the chance laws can have some nonzero chance of failing to obtain, for the reductionist says that these laws are determined by the total history of nonmodal fact. But these laws issue in chance distributions over possible total histories of nonmodal fact. Thus, it may turn out that positive chances are assigned to total histories that would specify *different* laws—the "undermining" of the chance laws by themselves (see Lewis 1994). Example: A coin is about to be tossed exactly $10^{10}$ times. As it happens, exactly half the tosses will land heads. A reductionist might say that it *follows* that the chance of heads on each toss is 0.5, adding that the correct chance laws will treat the tosses as independent. So there is now a large chance that the frequency of heads will be *different* from what it actually is—and if so, the *laws* will be different as well.

The inconsistency with (PP*) is now manifest. Consider those consistent history–law conjunctions $h \wedge l$ that entail that $P(l) < 1$. Pick such a conjunction; by (PP*),

$$C_0(l \mid h \wedge l) = P(l) < 1.$$

But by the probability calculus,

$$C_0(l \mid h \wedge l) = 1.$$

Lewis responds by amending (PP*) to what he calls the "new principle":

$$\text{(NP)} \quad C_0(a \mid h \wedge l) = P(a \mid l),$$

thus avoiding the inconsistency. The consensus in the literature seems to be that unlike the original principal principle, (NP) is unintuitive and, in application, unwieldy.

The reductionist (e.g., Lewis 1994) retorts that nonreductionists are hard-pressed to show how chances, understood their way, constrain rational credences according to (PP). But can the *reductionist* meet this challenge? Presumably, in addition to his reductionist analysis of chance, he ought to provide a *derivation* of (PP) from constraints on rational credence to which he is *already* committed. The literature provides no such derivation. And while a nonreductionist may also be unable to supply such a derivation, it is not clear why it would be needed. Arguably, both the reductionist and the nonreductionist are committed to the existence of substantive constraints on rational credence; why can't the nonreductionist simply include (PP) as one of them? (See Hall 2003 for further discussion.) Perhaps, then, the debate between reductionism and nonreductionism remains a stalemate.

Finally, a "deflationary" account of chance, associated with de Finetti and his followers, has proved to be very influential. Consider an infinite exchangeable sequence of events with respect to a probability function $P$. De Finetti's representation theorem states that the probability according to $P$ of exactly $k$ of the events occurring in $n$ trials is given by

$$\int_0^1 \binom{n}{k} p^k (1-p)^{n-k} f(p)\,dp,$$

for all $n$ and $k$ and for some density function $f$. The upshot is that any such probability distribution is representable as a "weighted average" of distributions. Each distribution corresponds to a hypothesis about the value of the probability $p$ of an event occurring on a single trial; it gives the probability of $k$ such events occurring in $n$ independent, identically distributed trials, given that fixed value of $p$. One can then average these distributions using the probabilities of their corresponding hypotheses about the value of $p$ as weights. The result is significant because it enables a subjectivist to "simulate" being an objectivist about chance when the exchangeability assumption holds, and for many situations this seems reasonable. If $P$ is one's subjective probability function, then it is *as if* one spread probability over various hypotheses about the single-case objective chance of the event, which remains fixed across infinitely many independent trials of the experiment in question. (See Skyrms 1994 for an excellent discussion of generalizations of exchangeability and their use in formulating various Goodmanian theses about projectability.) Indeed, common sense often (but not invariably) seems to require that probabilities be exchangeable over "green"-like hypotheses but not "grue"-like hypotheses.

## Conclusion

Feller (1957, 19) writes: "All possible definitions of probability fall short of the actual practice." Certainly, a lot is asked of the concept of probability. It is supposed at once to capture a quasi-logical notion, a subjective notion, and an objective notion instantiated in the mind-independent world. Perhaps one would do better to think of these as distinct *concepts* of probability. Each of the leading interpretations, then, attempts to illuminate one of these concepts, while leaving the others in the dark. In that sense, the interpretations might be regarded as complementary, although to be sure each may need some further refinement. Clearly, much work remains to be done on the philosophical foundations of probability. Equally clearly, the field has come a long way since the *Port Royal Logic*.

BRANDEN FITELSON
ALAN HÁJEK
NED HALL

## References

Arnauld, Antoine ([1662] 1964), *Logic, or, The Art of Thinking ("The Port Royal Logic")*. Translated by J. Dickoff and P. James. Indianapolis, IN: Bobbs-Merrill.

Billingsley, Patrick (1995), *Probability and Measure*, 3rd ed. New York: John Wiley & Sons.

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

——— (1952), *The Continuum of Inductive Methods*. Chicago: University of Chicago Press.

——— (1962), *Logical Foundations of Probability*, 2nd ed. Chicago: University of Chicago Press.

——— (1980), "*A Basic System of Inductive Logic II*," in Richard Jeffrey (ed.), *Studies in Inductive Logic and Probability, Volume II*. Berkeley and Los Angeles: University of California Press, 7–155.

de Finetti, Bruno ([1937] 1980), "*Foresight: Its Logical Laws, Its Subjective Sources*," in H. E. Kyburg Jr. and H. E. Smokler (eds.), *Studies in Subjective Probability*, 2nd ed. Huntington, NY: Robert E. Krieger, 53–118. Originally published as "*La prévision: Ses lois logiques, ses sources subjectives*," *Annales de l'Institut Henri Poincaré*, 7: 1–68.

——— (1990), *Theory of Probability*, vol. 1. Chichester, UK: Wiley Classics Library/Wiley & Sons.

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

Feller, William (1957), *An Introduction to Probability Theory and Its Applications*. New York: John Wiley & Sons, Inc.

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

Gaifman, Haim (1988), "*A Theory of Higher Order Probabilities*," in Brian Skyrms and William L. Harper (eds.), *Causation, Chance, and Credence*. Dordrecht, Holland: Kluwer Academic Publishers.

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

Hájek, Alan (2003a), "*Interpretations of Probability*," in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/sum2003/entries/probability-interpret

——— (2003b), "*What Conditional Probability Could Not Be*," *Synthese* 137: 273–323.

Hall, Ned (2003), "*Two Mistakes about Credence and Chance*," *Australasian Journal of Philosophy*: 93–111.

Jeffrey, Richard (1983), *The Logic of Decision*, 2nd ed. Chicago: University of Chicago Press.

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

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

Kolmogorov, Andrei. N. ([1933] 1950), *Foundations of Probability* [*Grundbegriffe der Wahrscheinlichkeitrechnung, Ergebnisse der Mathematik*]. New York: Chelsea Publishing.

Laplace, Pierre Simon ([1814] 1951), *A Philosophical Essay on Probabilities*. New York: Dover Publications.

Levi, Isaac (1990), "*Chance*," *Philosophical Topics* 18: 117–148.

Lewis, David (1986a), *Philosophical Papers: Volume II*. New York: Oxford University Press.

——— (1986b), "*A Subjectivist's Guide to Objective Chance*," in *Philosophical Papers: Volume II*. New York: Oxford University Press, 83–132.

——— (1994), "*Humean Supervenience Debugged*," *Mind* 103: 473–490.

Maher, Patrick (2001), "*Probabilities for Multiple Properties: The Models of Hesse and Carnap and Kemeny*," *Erkenntnis* 55: 183–216.

Mückenheim, W., G. Ludwig, C. Dewdney, P. Holland, A. Kyprianidis, J. Vigier, N. Petroni, M. Bartlett, and

E. Jaynes (1986), "*A Review of Extended Probability*," *Physics Reports* 133: 337–401.
Popper, Karl (1959), *The Logic of Scientific Discovery*. London: Hutchinson & Co.
Ramsey, Frank P. ([1926] 1980), "*Truth and Probability*," in H. E. Kyburg Jr. and H. E. Smokler (eds.), *Studies in Subjective Probability*, 2nd ed. Huntington, NY: R. E. Krieger, 23–52.
Renyi, Alfred (1970), *Foundations of Probability*. San Francisco: Holden-Day.
Roeper, Peter, and Hughes Leblanc (1999), *Probability Theory and Probability Logic*. Toronto: University of Toronto Press.
Savage, Leonard J. (1954), *The Foundations of Statistics*. New York: John Wiley.
Skyrms, Brian (1980), *Causal Necessity*. New Haven, CT: Yale University Press.
——— (1994), "*Bayesian Projectibility*," in D. Stalker (ed.), *Grue: Essays on the New Riddle of Induction*. Chicago: Open Court, 241–262.
van Fraassen, Bas (1995), "*Belief and the Problem of Ulysses and the Sirens*," *Philosophical Studies* 77: 7–37.
Venn, John (1866), *The Logic of Chance*. London and Cambridge: Macmillan.
von Mises, Richard (1957), *Probability, Statistics and Truth*. New York: Macmillan.
von Neumann, John, and Oskar Morgenstern (1944), *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

# PROGRESS

*See* **Scientific Progress**

# PROTOCOL SENTENCES

Protocol sentences are reports of an individual's experiences. The simplest and paradigmatic example is a report of "red here now." The term 'protocol sentence' was introduced by Rudolf Carnap (1932a, b) (and the example here is his). It reflects two chief aspects of logical empiricism (or positivism): (i) the importance of linguistic form and (ii) the role of experience as the source of acceptability and cognitive significance of scientific beliefs (see Logical Empiricism). The form, role, and status of protocol sentences became the topic of an important philosophical debate in the early 1930s involving logical empiricists such as Carnap, Moritz Schlick, Edgar Zilsel, Otto Neurath, Karl Popper, and a few others.

In *Der logische Afbau der Welt*, Carnap ([1928] 1967) investigated the logical "construction" of objects of intersubjective knowledge from several possible bases. These included a physicalist basis, but Carnap epistemically privileged a basis consisting of the autopsychological objects of private sense experience, termed "elementary experiences," which were supposed to provide the simplest and natural starting point for epistemological constructions. The main problem then became how objective knowledge is possible. To solve that problem required the existence of connecting statements or rules linking those other statements to these epistemically privileged experiential statements.

The *Aufbau*'s focus on an empiricist or phenomenalist model of knowledge in terms of the immediate experiential basis was interpreted by other members of the Vienna Circle as manifesting three philosophical positions: reductionism, atomism, and foundationalism. Reductionism took one set of terms to be fundamental or primitive; the rest