

BAS C. VAN FRAASSEN

RELATIVE FREQUENCIES*

The probability of an event is the limit of its relative frequency in the long run. This was the concept or interpretation developed and advocated in Reichenbach's *The Theory of Probability*. It cannot be true.

After so many years and so much discussion, it is a bit *vieux jeu* to attack this position. But the arguments tell equally against such related views as those of Kolmogorov and Cramer, which surround the standard mathematical theory. Even so, I would not attempt this, if it were not that I also want to defend Reichenbach. For the heart of his position is that the statistical theories of physics intend to make assertions about actual frequencies, and about nothing else.

In this, I feel, Reichenbach must be right.

I. ABSOLUTE PROBABILITIES

1. *The Inadequacy of Relative Frequency*

What is now the standard theory of probability is due to Kolmogorov, and it is very simple. A *probability measure* is a map of a Borel field of sets into the real number interval $[0, 1]$, with value 1 for the largest set, and this map countably additive.¹ Somewhat less standard is the notion of a *probability function*, which is similar, but defined on a field of sets and finitely additive. Relative frequency lacks all these features except that it ranges from zero to one.

This is not news; I need simply marshal old evidence (though I must say that it was a surprise to me, and I painfully reconstructed much of this along the way as I studied Reichenbach).

Let the actual long run be counted in days: 1 (today), 2 (tomorrow), and so on. Let $A(n)$ be an event that happens only on the n th day. Then the limit of the relative frequency of the occurrence of $A(n)$ in the first $n + q$ days, as q goes to infinity, equals zero. The sum of all these zeroes, for $n = 1, 2, 3, \dots$ equals zero again. But the union of the events $A(n)$ – that

is, $A(1)$ -or- $A(2)$ -or- $A(3)$ -or...; symbolically $\cup\{A(n) : n \in \mathbb{N}\}$ - has relative frequency 1. It is an event that happens every day.

So relative frequency is not countably additive.² Indeed, its domain of definition is not closed under countable unions, and so is not a Borel field. For let B be an event whose relative frequency does not tend to a limit at all. Let the events $B(n)$ be as follows: $B(n) = A(n)$ if B happens on the n th day, while $B(n) = \Lambda$, the 'empty' event, if B does not occur on the n th day. The limit of the relative frequency of $B(n)$ exists, and equals zero, for each number n . But B is the union of the events $B(n)$, and the limit of its relative frequency does not exist.

A somewhat more complicated argument, due to Rubin and reported by Suppes, establishes that the domain of relative frequency in the long run is not a field either.³ Let us divide the long run of days into segments: $X(n)$ is the segment which stretches from segment $X(n-1)$ to day 2^n inclusive ($X(1)$ is just the first and second day). We note that $X(n)$ is as long as the sum of all preceding segments. Call $X(n)$ an odd segment if n is odd; an even segment otherwise. Let A be an event that happens every day in every odd segment; but on no other days. In that case, the relative frequency of A has no limit in the long run. A is not in the domain of definition of relative frequency.

We let B and C be two events that overlap inside A , in a regular way: Let B happen on all the even dates on which A happens, and on all the odd dates when A does not happen. And let C also happen on all the even dates on which A happens, and in addition, on all even dates when A does not happen. About B and C it is true to say that in each segment, each of them happens every other day. So each has relative frequency $\frac{1}{2}$.

But the intersection $B \cap C$ is curious. Up to the end 2^n of segment $X(n)$, there have been exactly half as many $(B \cap C)$ days as there were A days. So if $B \cap C$ had a relative frequency in the long run, so would A . And A does not.

Let me sum up the findings so far. The domain of relative frequency is not closed under countable unions, nor under finite intersections. But still, countable additivity fares worse than finite additivity. For when the relative frequency of a countable union of disjoint events exists, it need not be the sum of the relative frequencies of those components. But if the relative frequencies of B , C , and $B \cup C$ all exist, while B is disjoint from C , then the relative frequency of $B \cup C$ is the sum of those of B and C .

We cannot say therefore that relative frequencies are probabilities. But we have not ruled out yet that all probabilities are relative frequencies (of specially selected families of events). For this question it is necessary to look at 'large' probability spaces; specifically, at geometric probabilities.

It is sometimes said that a finite or countable sample space is generally just a crude partition of reality – reality is continuous. Of course you cannot have countable additivity in a sensible function on a countable space! But the problem infects continuous sample spaces as well. In the case of a mechanical system, we would like to correlate the probability of a state with the proportion of time the system will spend in that state in the limit, that is, in the fullness of time. But take a particle travelling forever in a straight line. There will be a unique region $R(n)$ filled by the trajectory of the particle on day (n). The proportion of time spent in $R(n)$ tends to zero in the long run; but the particle always stays in the union of all the regions $R(n)$.

Geometric probabilities such as these have further problems. Take Reichenbach's favorite machine gun example, shooting at a circular target. The probability that a given part of the target is hit in a given interval, is proportional to its area, we say. But idealize a bit: let the bullets be point-particles. One region with area zero is hit every time: the set of points *actually* hit in the long run. Its complement, though of area equal to the whole target, is hit with relative frequency zero.

This example is very idealized; but it establishes that in certain cases, it is not *consistent* to claim that all probabilities are relative frequencies.

2. *Weaker Criteria of Adequacy*

When an interpretation is offered for a theory such as that of probability, what criteria should be satisfied? In logic, we tend to set rather high standards perhaps. We expect *soundness* (all theorems true in the models offered) and *completeness* (each non-theorem false in some of those models). In addition we hope for something less language-bound and more informative: the relevant features of all models of the theory must be reflected in the models offered by the interpretation. This is imprecisely phrased, but an example may help. Take the truth-functional interpretation of propositional logic. Any model of that logic is a Boolean algebra. Every Boolean algebra can be mapped homomorphically into

the two-element $\{T, F\}$ Boolean algebra pictured in the truth-tables. And if $a \neq 1$ in the original, the map can be such that the image of a is not T .

Translating this to the case of probability theory, we would ask of the relative frequency interpretation first of all that the structures it describes (long runs, with assignments of real numbers to subsets thereof through the concept of limit) be models of the theory. But that would require relative frequency to be countably additive which it is not. Secondly we would ask that if we have any model of probability theory, we could map its measurable sets into the subsets of a long run, in such a way that if the originals have distinct probabilities, then their images have distinct relative frequencies. The machine gun example shows this is not possible either.

In his book, Reichenbach certainly considered many questions concerning the adequacy of the relative frequency concept of probability. His exact formulation I shall examine in Part III below. But his book appeared in 1934, too early to be addressed to Kolmogorov's mathematical theory (which appeared in 1933), and much too early to guess that the latter would become the standard basis for all work in the subject. So Borel fields and countable additivity were not considered by Reichenbach. In addition, he construed probability theory on the paradigm of a logical calculus. This introduces the extra feature that probabilities are assigned to only denumerably many entities, for these are all denotations of terms in the language. So Reichenbach's discussion is concerned with weaker criteria of adequacy than I have listed.

Turning to representability, remember that Reichenbach could or would not consider more than denumerably many distinct probabilities at once. To show that he was right about this, in a way, I will have to give some precise definitions. A *probability space* is a triple $S = \langle K, F, P \rangle$, where K is a non-empty set, F a Borel field on K , P a map of F into $[0, 1]$ such that $P(K) = 1$ and P is countably additive. The infinite product $S^* = \langle K^*, F^*, P^* \rangle$ is formed as follows: K^* is the set of all maps σ of N (the natural numbers) into K ; F^* is the least Borel field on K^* which includes all the sets

$$\{\sigma \in K^* : \sigma(1) \in A_1, \dots, \sigma(n) \in A_n\} = A_1 \times \dots \times A_n \times K^*$$

for sets A_1, \dots, A_n in F ; and P^* is the probability measure on F^* such

that

$$P^*(A_1 \times \cdots \times A_n \times K^*) = P(A_1) \cdots P(A_n)$$

which exists by a theorem of Kolmogorov. A *long run* is simply a member σ of K^* .

The relative frequency of A in σ – where A is a member of F – call it $\text{relf}(A, \sigma)$, is defined by

$$\text{relf}(A, \sigma) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n a(\sigma(i))$$

where a is the characteristic function of A ; that is, $a(x)$ equals 1 if x is in A and equals zero otherwise. The Strong Law of Large Numbers⁴ has the consequence

$$P^*({\sigma \in K^* : \text{relf}(A, \sigma) = P(A)}) = 1$$

from which we can infer, as a minor consequence, that there will be at least one long run σ such that the probability $P(A)$ is the relative frequency of A therein. I shall attach no other importance to the function P^* , using it *only* to establish the existence of certain long runs.

The intersection of countably many sets of probability measure 1, must have measure 1 again. Thus we generalize that if A_1, \dots, A_k, \dots are all members of F , then

$$P^*({\sigma \in K^* : \text{relf}(A_i, \sigma) = P(A_i) \text{ for } i = 1, 2, \dots}) = 1$$

also, and therefore

If X is a denumerable subfamily of F , there exists a ‘long run’ σ in K^* such that, for each set A in X , $P(A) = \text{relf}(A, \sigma)$.

In other words, any denumerable family of probabilities can consistently be held to be reflected in the relative frequencies in the actual long run.

3. Square Bullets

Although relative frequencies are not countably additive, and so are not probabilities, we have just seen something very encouraging. Any countable family of probability assertions can consistently be interpreted as a family of assertions of relative frequencies. This creates the temptation to do a bit of Procrustean surgery. Perhaps large probability spaces can be

approximated by small ones, and a *small* probability space can be identified with a suitably chosen *part* of a relative frequency model. (On the subject of choosing a suitable part of a relative frequency model, Reichenbach explicitly rejected von Mises' restriction to 'random sets'. I feel that Reichenbach had good reasons for this; and also, I see little gain in the partial representation theorems the randomness tradition provides, so I shall say no more about this.)

Let me give a simple example to show the possibilities in this line, and also its limitations. Let the machine gun still fire point-bullets, but let the target be a line segment, conveniently coordinatized by distances as the interval $[0, 1]$. All the bullets fall on it, and the probability that interval $[a, b]$ is hit equals its length $b - a$. More generally, the probability a Borel set $E \subseteq [0, 1]$ is hit equals its Lebesgue measure. We would have the same problems as before trying to interpret this situation exhaustively in terms of relative frequencies. But why try, when perception thresholds are not infinitely fine? Given the essential human myopia, appearances must be a fragmentary part of reality. (Anyway, bullets are not point-sharp; they may be round or square.)

Any real interval $[a, b]$ is of course suitably approximated by the rational intervals $[a', b']$ such that $a' < a < b < b'$ and a' and b' are rational fractions. There are only countably many such rational intervals. Let us take the field F they generate (still countable), and find a long run in which the relative frequency of hits in a set in that field equals the Lebesgue measure of that set. Let that long run be σ , and let us begin our reconstruction.

$$(1) \quad m(A) = \text{relf}(A, \sigma) \text{ if } A \text{ is in } F.$$

Now in addition, if $B_1 \supseteq \dots \supseteq B_k \supseteq \dots$ is a series of sets in F , and B their intersection, let us say

$$(2) \quad m(B) = \lim_{n \rightarrow \infty} \text{relf}(B_n, \sigma)$$

when that exists.

This function m is a pretty good shot at the probability. (Perhaps even identical with it; let us not stop to inquire.) Since it assigns zero to every countable point-set, it is not identifiable with any relative frequency. But it is *approximated* by the function $\text{relf}(X, \sigma)$ restricted to field F .

Let us suppose for a moment that all and only functions thus approximated by relative frequencies are probabilities. This will remind us of very typical mathematical moves. If a structure is not topologically closed – if the limits of converging series in the structure do not always belong to the structure – then they are ‘put in’. That is, the mathematician’s attention switches to a larger structure, which is closed under such operations. The original topic of concern is looked upon as an arbitrarily hacked out fragment of the important structure. (Consider the definitions of Hilbert space and of tensor products of Hilbert spaces, as example.) So did probability theorists simply widen the original topic of concern – relative frequencies – in the traditional, smooth paving, mathematical fashion?

But no, that cannot be. The family of probability measures does not include relative frequencies in long runs, except on finite sample spaces. The limit points, if that is what they are, are not *added* onto the original family, but the originals are thrown out. Moreover, those limit functions are not extensions or extrapolations of relative frequencies. In the above example, $m(X)$ is not an extension of $\text{relf}(X, \sigma)$, for the two disagree on countable sets; $m(X)$ is only an extension of the result of restricting $\text{relf}(X, \sigma)$ to field F .

The construal of probabilities I have just described in a simple case is explained by Reichenbach for more interesting examples in his Chapter Six. Let the target be a region A on the Cartesian plane. A geometric probability will be a measure m on the plane that gives 1 to A , and for which

$$(3) \quad m(B) = \iint_B \phi(x, y) \, dx \, dy$$

for each subregion of A , for a function ϕ – the *probability density* – which is characteristic of m .

We can proceed as follows: Cut the region A into n subregions of equal area, call them E_1, \dots, E_n . The probability space which has m restricted to the Borel field generated by E_1, \dots, E_n is a finite probability space. Therefore the probability assertions about it can be interpreted through relative frequencies. Call this space S_1 .

Construct space S_2 by refining S_1 : each region E_k is subdivided into n regions of equal area, and the probabilities assigned are again the restriction of m to the (Borel) field generated by this (finite) set of small regions. And so forth.

Given now a series of sets $B_1 \supseteq \dots \supseteq B_k \supseteq \dots$ where B_k is a measurable set in S_k , it is clear that m will still be defined for their intersection, and the value there assigned by m is the limit of the values $m(B_1), \dots, m(B_k), \dots$. From this Reichenbach infers that there must be a parallel 'empirical construction'. The probability of B_k , he says, is 'empirically determined' by estimating the long run relative frequency on the basis of finite samples. If the limit of these probabilities exists, we then assign that limit to the intersection $\cap \{B_k : k = 1, 2, 3, \dots\}$. With the probabilities so determined, we form the probability space S which has as measurable sets the least Borel field containing all the measurable sets in each space S_k .

The idea is clearly that S is the mathematical limit of the series of spaces S_1, \dots, S_k, \dots and that the representation of S as such a limit of finite spaces constitutes a representation in terms of relative frequencies. For each space S_k has only finitely many measurable sets, so there will be a long run σ of points such that $m(B) = \text{relf}(B, \sigma)$ for each measurable set B of S_k .

But in fact we have nothing like a representation in terms of relative frequencies here. Let S_∞ be the space that has as measurable sets all the ones from the spaces S_1, \dots, S_k, \dots ; and let the probability of B in S_∞ be exactly what it is in those spaces S_k in which B is measurable. Then S is the least upper bound, in a natural sense, of the series S_1, \dots, S_k, \dots . In addition, the measurable sets of S_∞ form a field – but not a Borel field. However, there are only countably many of them so there is a long run of points σ' such that the probability of B in S_∞ equals $\text{relf}(B, \sigma')$. Surely, from the frequency point of view, S_∞ is the reasonable extrapolation from the finite spaces for which the probabilities are empirically determined.

The space S contains many more measurable sets than S_∞ does, and there is no single long run in which the relative frequencies reflect the probabilities in S . Thus it is not reasonable to present the relation between S and the finite spaces S_k as showing that the probabilities in S are a mere extrapolation from, or representable in terms of, relative frequencies.

4. *Representation of Probabilities*

We already know that we cannot have a representation of probabilities as relative frequencies. But an approximation theorem, if strong enough, is a sort of representation theorem. If all probability measures are suitably approximated by (series of) relative frequency functions, then we can say that probabilities are at least representable *in terms of* (if not *as*) relative frequencies. And the preceding section holds out this hope.

But this would not be a *good* representation in this case. Indeed, I suspect that Reichenbach was misled in this. Let us try to picture the practical context in which we take samples and estimate probabilities. This work is done in what we might call a 'practical language'. In this we state experimental and sampling results – always proportions in finite classes – and also extrapolate these results (by induction, if that exists; by guessing, if that is any different) into hypotheses about probabilities.

In any experiment or observation I can *explicitly check* only finitely many samples, and these of finite size. Moreover, my language has only countably many expressions in it, so I can *explicitly extrapolate* only to countably many sets. This must be what suggested to Reichenbach that he need only concern himself directly with finite sample spaces, extrapolate to countable series of these, and worry only about finite unions and intersections. After all, the language cannot have explicit designations for all the countable unions of classes designated in it.

But if this was the train of thought, then he was misled. For our practical language does certainly contain expressions like '*limit*', \sum^∞ , and \bigcup^∞ . It has numerals and number variables. And even though it *is* countable, it has a systematic way of designating countable unions and limits of series. Our extrapolations, always *in* this language, are *to* countable unions and infinite sums. And the method of extrapolation, however haphazard in its inductive leaps, rigidly follows countable addition. These assertions I make about the practical activity of going from sampling data to the framing of probabilistic hypotheses, models, and theories; and what I have to say further will be baseless if in fact scientists are happily violating countable additivity when they propose their hypotheses – but I do not believe so.

If all this is true, the picture drawn in the preceding section (and following Reichenbach's Chapter Six) is not realistic. For there the

practical language was assumed to contain itself only descriptions of the finite sample spaces S_1, S_2, \dots ; the 'nearest' limit was a frequency space S_∞ which was ignored; and the extrapolation was to be outside the practical language to a function m defined on a large family of sets not designatable in the language itself. The only criterion of adequacy considered is that m agrees with all the *finitary* extrapolations from sample data (by ordinary conjunction, disjunction, and negation), because that is all that is assumed to rear its head in the practical context.

I shall now use the Law of Large Numbers – or rather the corollaries to it in Section 2 above – to give an explicit representation of all probabilities *in terms of* (not *as*) relative frequencies. This will be different from the representation by approximations so far discussed, in that it will assume that what we extrapolate to is relative frequencies on countable families of sets, and that the extrapolation to countable unions is indeed by countable addition.

A *probability space* is a triple $S = \langle K, B, P \rangle$ where K is a non-empty set, B a Borel field on K , P a countably additive map of B into $[0, 1]$ such that $P(K) = 1$. A *frequency space* is a couple $M = \langle \sigma, F \rangle$ for which there is a set K such that

- (a) σ is a countable sequence of members of K
- (b) F is a family of subsets of K
- (c) $\text{relf}(A, \sigma)$ exists for each A in F .

I shall call $M = \langle \sigma, F \rangle$ a *special frequency space* exactly if $\text{relf}(X, \sigma)$ is countably additive on F in so far as F is closed under countable unions; and F is a field:

- (d) F is a field on K
- (e) if $\{A_i\}, i = 1, 2, \dots$ is a countable family of disjoint sets, all in F , and their union A is also in F , then $\text{relf}(A, \sigma) = \sum_{i=1}^{\infty} \text{relf}(A_i, \sigma)$.

Let us shorten 'special frequency space' to 'sfs'. We have already seen in Section 2 that for each probability space $S = \langle K, B, P \rangle$ and countable subfamily G of B there is a sequence σ such that $P(A) = \text{relf}(A, \sigma)$ for all A in G . Without loss of generality we can say: countable subfield G of B ,

for if G is countable, so is the field it generates. Moreover, because P is countably additive on B , so is the function $\text{relf}(-, \sigma)$ on G . So $\langle \sigma, G \rangle$ is an sfs.

A *good family* will be a family Z of sfs, such that (a) the union of all the fields G of members $\langle \sigma, G \rangle$ of Z is a Borel field, and (b) if $\langle \sigma, G \rangle$ and $\langle \sigma', G' \rangle$ are both in Z , and A in both G and G' , then $\text{relf}(A, \sigma) = \text{relf}(A, \sigma')$. The *space associated* with a family Z of frequency spaces is a triple $S(Z) = \langle K, F, P \rangle$ such that F is the union of all the second members of elements of Z , K is the union of F , and P the function: $P(A) = q$ iff $\text{relf}(A, \sigma) = q$ for all $\langle \sigma, G \rangle$ in Z such that A is in G .

It will be clear that the space $S(Z)$ associated with a good family Z of sfs is a probability space. Moreover, the weak representation result of the section before last shows at once that there is for each probability space a good family of sfs with which it is associated.

II. CONDITIONAL RELATIVE FREQUENCIES

1. *The Inadequacy of Defined Conditional Probabilities*

The standard (Kolmogorov) theory defines the conditional probability by

$$(1) \quad P(B/A) = P(B \cap A)/P(A) \text{ provided } P(A) \neq 0.$$

Thus the probability that it will rain on a given day if the sky is overcast, is well-defined if and only if the probability that the sky is overcast is not itself zero.

There are other theories, specifically those of Renyi [8] and Popper [6] in which $P(B/A)$ is taken as well-defined also in (some) cases in which $P(A) = 0$. The relevant features of those theories here are:

$$(2) \quad P(B \cap C/A) = P(B/A)P(C/B \cap A) \text{ if all these terms are defined,}$$

$$(3) \quad \text{If } P(-/A) \text{ and } P(-/B) \text{ are both well-defined functions, they have the same domain as } P(-/K),$$

when the sets B, C, A are measurable sets in a relevant sort of conditional probability space with 'universe' K - i.e., when $P(A/K), P(B/K), P(C/K)$ are defined.

Clearly Reichenbach's theory does not belong to the first sort. The consideration of relative frequencies is a main motive for dissatisfaction with the definition 1. But as I shall also show, conditional relative frequencies do not quite fit Renyi's or Popper's theory.

When a submarine dives, filling its ballast tanks to some extent with water, and then the engine is switched off, it remains stationary.⁵ Its center of gravity is at a certain exact depth; that is, lies in one of the planes cutting the water horizontally. The probability of its coming to lie in that exact plane, is zero. What is the probability that it lies in that plane and north of the forty-first parallel, given that it lies indeed in that plane?

I imagine the practical mind can always find some way around such questions. Recall, however, the enormous distance of idealization we have seen between actual repetitive phenomena, *even* extrapolated to the fullness of time, and probability spaces. I do not think the way around should be introduced with the statement such 'unreal' or 'impractical' problems have no place in the business-like mind of the probability theorist.

But my point is not to ask for improvements in probability theory here. I only want to point out a further difference between probabilities and relative frequencies. Even when A has relative frequency zero, the relative frequency of B 's among the A 's may just as well make sense. Suppose we toss a die forever, and even numbers come up only ten times. Unlikely, but possible. In such a world, the probability that a given toss yields an even number is, on the frequency view, zero. But surely, even there, the probability that a particular toss yields six, if this toss is one of those that yield even, must be well-defined. It is exactly the proportion of sixes in those ten tosses.

More abstractly, let σ be just the series of natural numbers. If A is the set of integral powers of 93, then $\text{relf}(A, \sigma)$ is zero. But those powers themselves form a subseries σ' . If B is the set of *even* powers of 93, then $\text{relf}(B, \sigma')$ is not only well-defined, but is already $\frac{1}{2}$. And should this not exactly be the relative frequency of the even powers among the integral powers?

The subject cannot be simple, however. If A is a set of members of σ , let $\sigma(A)$ be the subseries of σ consisting of members of A . I shall worry about exact definitions later. Let the *conditional (relative) frequency* $\text{relf}(B/A, \sigma)$ be construed as $\text{relf}(B, \sigma(A))$. In that case we would like to know when that conditional frequency is defined.

As I pointed out, some theories of probability (like Renyi's or Popper's) take conditional probability as basic. But these have feature 3 above; so that if $P(B/K)$, $P(C/K)$, and $P(B/A)$ are all defined, then so is $P(C/A)$. We can get nothing so nice for conditional frequencies.

For let A and B be as in the natural number example of the second paragraph before last. In addition, let C be a subset of A whose relative frequency in $\sigma(A)$ does *not* tend to a limit. In that case, $\text{relf}(A, \sigma) = 0$, and since B and C are subsets of A , $\text{relf}(B, \sigma) = 0 = \text{relf}(C, \sigma)$. Moreover, A is a perfectly good 'antecedent condition' because $\text{relf}(B/A, \sigma)$ is defined and equals $\frac{1}{2}$. But $\text{relf}(C/A, \sigma)$ is not defined.

We cannot localize the problem by noting that C has measure zero from the point of view of σ . First, so did B and $\text{relf}(B/A, \sigma)$ is defined. Second, there are no powers of 93 which are even numbers, but if E is the set of even numbers, then $\text{relf}(E \cup C, \sigma) = \frac{1}{2}$ while $\text{relf}(E \cup C/A, \sigma)$ is not defined.

The subject of conditional frequencies is therefore not correctly treated by *either* Kolmogorov, *or* Popper, *or* Renyi. In the next two sections I shall indicate ways in which one might go about providing a general theory. Reichenbach's attempts I shall consider in Part III.

2. The Natural Frequency Space

In Part I, I defined a *frequency space* as a couple $\langle \sigma, F \rangle$ for which there exists a set K such that

- (a) σ is a countable sequence of members of K
- (b) F is a family of subsets of K
- (c) $\text{relf}(A, \sigma)$ exists for each A in F .

We can easily map any such structure into what I shall call the natural frequency space $\langle \omega, F(\omega) \rangle$ where

- (a) ω is the natural number series $(1, 2, 3, \dots)$
- (b) $F(\omega)$ is the family of sets A of natural numbers such that $\text{relf}(A, \omega)$ exists.

What exactly is in the family $F(\omega)$ is very mysterious. The map $A \rightarrow \{i : \sigma(i) \in A\}$ will relate $\langle \sigma, F \rangle$ to $\langle \omega, F(\omega) \rangle$ in a natural way, preserving relative frequencies. It also preserves the set operations on the members of F which are sets of members of σ .

If K has in it members foreign to σ , these will of course play no role in the determination of relative frequencies, and are ignored in this map.

Before turning to conditionalization, I wish to make relative frequency and conditional relative frequency precise. If A is in $F(\omega)$, let a be its characteristic function: $a(i) = 1$ if i is in A , and $a(i) = 0$ otherwise. In that case:

$$+a(n) = \sum \{a(m) : m \leq n\}$$

$$\text{rel}(a, n) = \frac{+a(n)}{n}$$

$$\text{relf}(A) = \lim_{n \rightarrow \infty} \text{rel}(a, n)$$

$$\text{relf}(B/A) = \lim_{n \rightarrow \infty} \frac{\text{rel}(ba, n)}{\text{rel}(a, n)}$$

$$= \lim_{n \rightarrow \infty} \frac{+ba(n)}{+a(n)}$$

where ba is the characteristic function of AB , namely $ba(n) = b(n)a(n)$.

What I would like to show now is that the natural frequency space is *already* closed under conditionalization (insofar as it can be). That is, there is for each pair of sets B and A in $F(\omega)$ another set $A \rightarrow B$ such that

$\text{relf}(A \rightarrow B)$ exists if and only if $\text{relf}(B/A)$ exists; and if they exist, they are equal.

We find this set by constructing its characteristic function $(a \rightarrow b)$. As a first approximation, I propose

$$(a \rightarrow b)(i) = \begin{cases} 1 & \text{if the } i\text{th member of } A \text{ is in } B \\ 0 & \text{otherwise.} \end{cases}$$

The members of A have of course a natural order: if A is $\{1, 3, 5\}$ then 3 is its *second* and 5 its *third* member. If A is infinite, the above will do very

well, for we shall have

$$\begin{aligned} \text{rel}(a \rightarrow b, n) &= \frac{+(a \rightarrow b)(n)}{n} \\ &= (\text{The number of } B\text{'s among the} \\ &\quad \text{first } n \text{ } A\text{'s})/n \\ &= \frac{+ab(m)}{+a(m)} \end{aligned}$$

where m is the number at which the n th A occurs. The variables n and m go to infinity together, and the lefthand limit equals the righthand limit.

If A is finite, we need a slight emendation. For example, the relative frequency of the even numbers among the first 10, should be $\frac{1}{2}$. If we kept the above definition, it would be zero. So we call k the *index* of A if its characteristic function a takes the value 1 exactly k times. Each number i is of course a multiple of k plus a remainder r (ranging from one to k ; this slightly unordinary usage of 'remainder' makes the sums simpler). We let $a \rightarrow b$ take value 1 at i exactly if the (r)th A is a B . In that way,

$$\begin{aligned} \text{rel}(a \rightarrow b)(k) &= \frac{\text{the number of } B\text{'s among the first } k \text{ } A\text{'s}}{k} \\ \text{rel}(a \rightarrow b)(km) &= \frac{m \times \text{the number of } B\text{'s among the first } k \text{ } A\text{'s}}{mk} \\ &= \frac{+(a \rightarrow b)(k)}{k} \end{aligned}$$

which, if there are only k A 's, is just correct. If we call that number x/k , we also see that

$$\frac{+(a \rightarrow b)(km+r)}{km+r} \leq \frac{mx+(k-1)}{km+(k-1)}$$

which series converges to x/k as m goes to infinity. So this is indeed the correct characteristic function.

As a minor exercise, I shall now make this precise. Let 0 and 1 stand also for those characteristic functions which belong to Λ and N respectively, that is, take constant values 0 and 1.

- (i) $\text{index}(a) = k$ iff $\sum \{a(n) : n \in N\} = k$
 [let k here be any natural number, or ∞]
- (ii) $i \div k = r$ iff either $k = \infty$ and $i = r$, or
 $(\exists m)(mk + r = i)$ where $0 < r \leq k$.
- (iii) $\# a(r) = i$ iff $a(i) = 1$ and $+a(i) = r \div \text{index}(a)$
- (iv) $(a \rightarrow b) = \begin{cases} b \# a & \text{if } a \neq 0 \\ 1 & \text{otherwise} \end{cases}$

Here (iii) must be read as defining ‘the r th (modulo the index of A !) A occurs at i ’. Thus if A is $\{1, \dots, 10\}$, then the 5th A occurs at 5; also the 25th (modulo the index 10) occurs at 5. So $\# a(5) = \# a(25) = 5$ in this example. If this “25th” A occurs at 5, then $a \rightarrow b$ should take the value 1 at 25 exactly if b takes the value 1 at 5. Thus $(a \rightarrow b)(25) = b(\# a(25)) = b(5)$. This is what (iv) says. In the case where a never takes the value 1, I have arbitrarily given $a \rightarrow b$ the value 1 everywhere. This is only a trick to keep the object well-defined.

The class $A \rightarrow B$ is the one which has characteristic function $a \rightarrow b$. While the operation \rightarrow is very ‘linear’, it is not really interesting. It is certainly not a conditional in the sense of a logical implication, because even the analogue to *modus ponens* would not hold. It serves its purpose of showing that the natural frequency space already contains all the *conditional relative frequencies*; but in its own right, it is no more than a mathematical *objet trouvé*.

3. A Partial Algebra of Questions

I shall now continue the general theory of relative frequencies, taking the conditionality for granted. That is, we know at this point that $\text{relf}(B/A, \sigma)$ can be reduced to an assertion of ‘absolute’ form $\text{relf}(X, \sigma)$. Henceforth, the conditional concept $\text{relf}(B/A, \sigma)$ will therefore be used without comment.

Let us view suitable relative frequencies in the long run as answers to questions of the form ‘What is the chance that an A is a B ?’ The terms A and B I take to stand for subsets of a large set K ; the set of possible situations or states or events. The question I shall reify as the couple $\langle B, A \rangle$. Since the answer would be exactly the same if we replaced ‘a B ’ in the question by ‘an A which is a B ’, I shall simplify the matter by requiring that $B \subseteq A$. This couple is a *question on K* , and I shall call B the *Yes-set* and A the *Domain* of the question. The *answer* relative to long run σ (a countable sequence of members of K) will be $\text{relf}(B/A, \sigma)$, if indeed $\text{relf}(B/A, \sigma)$ exists; otherwise the question is mistaken relative to σ .

The word ‘chance’ was proposed by Hacking as a neutral term. I use it here, but mean of course exactly what Reichenbach thought we should mean with ‘probability’. The occurrence of *actual* long run σ in the determination of the answer makes the question empirical.

(a) *Questions.* A *question on set K* is a couple $q = \langle qY, qD \rangle$ with $qY \subseteq qD \subseteq K$. Call questions q and q' *comparable* exactly if $qD = q'D$. The set $[q]$ of questions comparable to q is clearly ordered through the relations on its first members (*Yes-sets*), and we could define

$$q \cap q' = \langle qY \cap q'Y, qD \rangle - q = \langle qD - qY, qD \rangle$$

and so on, to show that $[q]$ is a family isomorphic by a natural mapping to the powerset of qD .

The operation of *conditionalization* which I shall now define, takes us outside $[q]$:

$$q \rightarrow q' = \langle qY \cap q'Y, qY \rangle$$

unless of course $qY = qD$. This leads us to the next topic.

(b) *Unit questions.* If q is a question and $qY = qD$, I shall call it a *unit question*. Its answer must always be 1; and in $[q]$ it plays the role of unit, that is, supremum of the natural partial ordering.

Henceforth let Q be the set of questions on K , and U the set of unit questions thereon; let u, v, \dots always stand for unit questions.

The unit questions are not comparable to each other. But they are easily related nevertheless; they form a structure isomorphic to $P(K)$

under the natural map: A to $\langle A, A \rangle$. Let us use the symbols \neg, \wedge, \vee, \leq in this context, to maintain a distinction between the natural ordering of the unit questions and that of the questions comparable to a given one:

$$\begin{aligned}\neg u &= \langle K - uY, K - uY \rangle \\ u \wedge v &= \langle uY \cap vY, uY \cap vY \rangle \\ u \leq v &\text{ iff } uY \subseteq vY\end{aligned}$$

where of course $u = \langle uY, uD \rangle$ and so on. Every question is a conditionalization of unit questions:

$$q = \langle qY, qD \rangle = \langle qD, qD \rangle \rightarrow \langle qY, qY \rangle.$$

Therefore define

$$\begin{aligned}uq &= \langle qD, qD \rangle \\ vq &= \langle qY, qY \rangle \\ q &= (uq \rightarrow vq).\end{aligned}$$

We note that if q and q' are *comparable*, the operations on them are definable in terms of those on unit questions:

$$\begin{aligned}\neg q &= uq \rightarrow \neg vq \\ (q \wedge q') &= uq \rightarrow (vq \wedge vq') \\ (q \cup q') &= uq \rightarrow (vq \vee vq') \\ q \subseteq q' &\text{ iff } vq \leq vq' .\end{aligned}$$

It would not make much sense to generalize these except for \neg which is of course defined for all questions. However, I want to add one operation on all questions

$$q \cdot q' = uq \rightarrow (vq \wedge vq')$$

which reduces to \cap when q and q' are comparable, but is otherwise not commutative. However, it gives the comparable question to q that comes closest to being its conjunction with q' .

(c) *The logic of questions.* Since I have not given a partial ordering of all questions, it may seem difficult to speak of a logic at all. Should we say

that q implies q' if the Yes-set of q is part of the Yes-set of q' ? Or should we require in addition that their Domains are equal? Or that the No-set (Domain minus Yes-set) of q' be part of the No-set of q ? The minimal relation is certainly that the Yes-set of q be part of the Yes-set of q' . We may think of a Yes-No question q as related to a proposition which is *true* at x in K if $x \in qY$, *false* at x if $x \in (qD - qY)$, and *neither true nor false* in the other cases. In that context the relation of semantic entailment is just that minimal relation (corresponding to valid arguments) of 'if true, then true'. So let us define

$$q \Vdash q' \text{ iff } qY \subseteq q'Y \\ \text{iff } q \rightarrow q' \text{ is a unit question .}$$

Then we note that the analogue to *modus ponens* holds, but only because something stronger does

$$q \cdot (q \rightarrow q') \Vdash q \rightarrow q' \Vdash q$$

which should not be surprising, because these conditionals are just like Belnap conditionals (a Belnap conditional says something only if its antecedent is true; in that case it says that its consequent is true – see [2], [11]).

$$(i) \quad q \rightarrow q' = vq \rightarrow vq' \\ (ii) \quad q_0 \rightarrow (q \rightarrow q') = q_0 \rightarrow (vq \rightarrow vq') \\ = q_0 \rightarrow \langle vqY \cap vq'Y, vqD \rangle \\ = q_0 \rightarrow (vq \wedge vq') \\ = q_0 \rightarrow (q \cdot q')$$

$$\text{Corollary: } u \rightarrow (v \rightarrow v') = u \rightarrow (v \wedge v')$$

$$(iii) \quad (q_0 \rightarrow q) \rightarrow q' = \langle q_0Y \cap qY, q_0Y \rangle \rightarrow q' \\ = \langle q_0Y \cap qY \cap q'Y, q_0Y \cap qY \rangle \\ = (vq_0 \wedge vq) \rightarrow vq' \\ = (q_0 \cdot q) \rightarrow q'$$

$$\text{Corollary: } (u \rightarrow v) \rightarrow v' = (u \wedge v) \rightarrow v'$$

$$(iv) \quad \text{if } u_0 \leq u \leq u' \text{ then } u \rightarrow u_0 = (u' \rightarrow u) \rightarrow (u' \rightarrow u_0) .$$

The second and third show that iteration is trivial; the fourth is a trivial corollary which I mention because of the way it recalls the 'multiplication axiom'. Which brings us to the next topic.

(d) *The multiplication axiom.* Reichenbach's fourth axiom was the 'theorem of multiplication'. In our symbolism, it states

(AM) If $\text{relf}(B/A, \sigma) = p$ and $\text{relf}(C/A \cap B, \sigma) = r$ exist, then $\text{relf}(B \cap C/A, \sigma)$ also exists and equals $p \cdot r$.

The answer to the question q , relative to σ is

$$\text{relf}(qY/qD, \sigma) = m(q).$$

Just for the moment, let A, B, C stand equally for the unit questions, $\langle A, A \rangle, \langle B, B \rangle$, etc. Then the axiom clearly says:

If $m(A \rightarrow B) = p$ and $m(A \cap B \rightarrow C) = r$ exist, then $m(A \rightarrow B \cap C)$ also exists and equals $p \cdot r$.

Because $X \rightarrow Y = X \rightarrow X \cap Y$, there are only three sets really operative here: $A \cap B \cap C, A \cap B, A$. So we can phrase this also as follows:

If $u_0 \leq u \leq u'$, and $m(u' \rightarrow u) = p$ and $m(u \rightarrow u_0) = r$ exist, then so does $m(u' \rightarrow u_0)$ and equals $p \cdot r$.

Thus in the favorable case of $p \neq 0$, which here means only that a certain conditional probability is not zero,

$$(AM^*) \quad m(u \rightarrow u_0) = \frac{m(u' \rightarrow u_0)}{m(u' \rightarrow u)} \text{ if } u_0 \leq u \leq u'.$$

But at the end of subsection (c) we just saw that with $u_0 \leq u \leq u'$ given, $u \rightarrow u_0 = (u' \rightarrow u) \rightarrow (u' \rightarrow u_0)$. So we have

$$m[(u' \rightarrow u) \rightarrow (u' \rightarrow u_0)] = \frac{m(u' \rightarrow u_0)}{m(u' \rightarrow u)}.$$

Let us now generalize this to conditionals for which the antecedent and consequent are not specially related:

$$\begin{aligned} m(q \rightarrow q') &= m(vq \rightarrow (vq \wedge vq')) \\ &= \frac{m(uq \rightarrow (vq \wedge uq'))}{m(uq \rightarrow vq)} \\ &= \frac{m(q \cdot q')}{m(q)} \quad \text{provided } m(q) \neq 0 \end{aligned}$$

where I went from the first line to the second by the reflection that $uq \leq vq \leq vq \wedge vq'$ always, and applying (AM*).

We have seen that there is in the general theory of questions "What is the chance that an A is a B ?" there is a logically reasonable conditionalizing operation; relative frequency of the conditional object looks quite familiar in the relevant special cases.

III. REICHENBACH'S THEORY

Reichenbach had two aims in writing *The Theory of Probability*. He wanted to present a good axiomatization of probability theory, and also to defend his frequency interpretation. The two aims interfered somewhat with each other. But the book is a monumental and instructive work, with many fascinating features that a narrower or less committed enterprise would certainly have lacked.

1. *The Formulation*

Logic had a bad influence. The language is essentially that of quantificational logic, with the quantifier ranging over the positive integers. There are special function terms x, y, \dots such that x_i is an individual (event) if i is an integer, and class terms A, B, \dots . The operators are Boolean (on the classes); the single predicate is the binary \in of membership; and the connectives are those of propositional logic plus a special variable connective

$$\underset{p}{\exists}$$

where p is any numerical expression denoting a real number (if you like, between zero and one inclusive). This special connective is eventually

allowed to combine any sentences, but to begin is considered only in the context

$$(1) \quad (i)(x_i \in A \underset{p}{\ni} y_i \in B).$$

To understand this, reason as follows: If this coin be tossed, the probability of its showing heads equals $\frac{1}{2}$. This asserts a correlation between two sequences of events: the event x_n is the n th toss of the coin, and the event y^n the n th landing of the coin.

There is no gain, formally speaking, in this consideration of more than one sequence of events. If $\{x_i, y_i\}, \dots$ are the sequences of events designated, we could construct a single sequence $\{\omega_i\}$ such that for each n , ω_n is the sequence $\langle x_n, y_n, \dots \rangle$. In that case the class terms A, B, \dots would have to be reinterpreted accordingly; to the old formula " $y_i \in B$ " would correspond a new formula " $\omega_i \in B'$ " with the same truth-conditions. (That is, B' would designate the class of all sequences $\langle a, b, \dots \rangle$ such that b is in the class designated by B .) In that way, all basic probability assertions would take the form

$$(2) \quad (i)(\omega_i \in A \underset{p}{\ni} \omega_i \in B).$$

The event ω_n is the complex totality of all the events happening on the n th day, say, and the sequence of elements $\omega = \{\omega_n\}$ plays the role of the long run σ . Thus 2 would in my symbolism so far be

$$(3) \quad P(B/A, \omega) = p$$

for the truth-conditions of 2 and 3 (as explained respectively by Reichenbach and in part II above) are the same.

2. The Multiplication Axiom

In a paper published in 1932, preceding the book, Reichenbach used special 'reversal axioms' to govern inverse calculations. An example would be the inference (using Reichenbach's own abbreviations):

$$(1) \quad \text{If } A \text{ and } B \text{ are disjoint, then } P(A \cup B) = P(A) + P(B). \\ \text{Therefore, in that case, if } P(A \cup B) = p \text{ and } P(A) = q, \text{ then} \\ P(B) = p - q.$$

This inference contains a hidden assumption, namely that all the relevant terms are well-defined (all the probabilities exist). Strictly speaking, the premise should be:

- (2) If $P(A)$ and $P(B)$ exist, and A and B are disjoint, then $P(A \cup B)$ exists and equals $P(A) + P(B)$.

And a *second* premise needed in the inference is then

- (3) If $P(A \cup B)$ and $P(A)$ exist, and A and B are disjoint, then $P(B)$ exists.

The special reversal axioms were apparently too weak (see page 61 of *The Theory of Probability*), and in the book their role was played by a single ‘Rule of Existence’ (page 53). This rule seems to be correct, and Reichenbach is very careful in its use.

But extreme care does seem to be needed. On page 62 Reichenbach gives as his fourth axiom the ‘Theorem of Multiplication’.

$$(4) \quad (A \underset{p}{\Rightarrow} B) \cdot (A \cdot B \underset{p}{\Rightarrow} C) \supset (\exists w)(A \underset{w}{\Rightarrow} B \cdot C) \cdot (w = p \cdot u)$$

where $(A \underset{p}{\Rightarrow} B)$ is short for “ $(i)(x_i \in A \underset{p}{\Rightarrow} y_i \in B)$ ” and so forth. He concludes via the Rule of Existence that

$$(5) \quad P(B \cap C/A) = P(B/A)P(C/A \cdot B)$$

“can be solved according to the rules for mathematical equations for each of the individual probabilities occurring”. This is certainly true, but only a very careful reader will not be tempted to infer if $P(B \cap C/A)$ and $P(B/A)$ both exist and equal zero, then $P(C/A \cap B)$ exists though its numerical value cannot be determined. This inference would be invalid on the frequency interpretation. (This can be seen from the example at the beginning of Part II: replace ‘ A ’ by ‘ K ’ and ‘ B ’ by ‘ A ’ in equation 5, to get $P(C/K) = P(A/K)P(C/A)$ given that $A \cap C = A$ and all are subsets of K .) Reichenbach does not make this inference, but neither does he point out its invalidity, which is a very special feature of his theory, because it is involved with the fact that $P(B/A)$ may exist while $P(C/A)$ does not, although $P(C/K)$ does – a feature not shared by the theories of Kolmogorov, Popper, and Renyi.

I should mention the very nice deduction Reichenbach gives of the multiplication axiom from the weaker axiom that $P(B \cap C/A) = f[P(B/a), P(C/A \cap B)]$ for some function f . His proof assumes that f is differentiable, but he says that this requirement can be dropped.

3. *The Axiom of Interpretation*

Reichenbach's second axiom (Normalization) has the corollary

$$(1) \quad (i)(x_i \notin A) \supset (A \underset{p}{\rightarrow} B) \quad \text{for all } p.$$

In other words, if an event A never occurs, then the probability of A does not exist. Much later on (Section 65) Reichenbach states also the Axiom of Interpretation

$$(2) \quad \text{If an event } C \text{ is to be expected in a sequence with a probability converging toward } 1, \text{ it will occur at least once in the sequence.}$$

It is not clear whether this axiom is meant as part of the formal theory, but Reichenbach does say that "whoever admits [a contrary] possibility must abandon every attempt at a frequency interpretation" (page 345).

Either of these axioms is sufficient to rule out geometric probability altogether. For any sequence $\{x_i\}$ will be countable, and so receive probability zero in a geometric example. The complement of the set of points in the sequence thus receives 1, but this is an event which never occurs in the sequence.

4. *Geometric Probability*

In Chapter Six, Reichenbach tries to deal with geometric probability. He says that this provides an interpretation for his probability calculus. Without qualification, this is not true; but it would be true if we delete the Axiom of Normalization (and of course, never add the Axiom of Interpretation). It would also be true if we allow the class terms A to stand only for finite or empty regions.

On page 207, Reichenbach actually says that the geometric and frequency theories are isomorphic. There is no reasonable qualification

which makes this true. He discusses the special feature of geometric probability that

- (1) if $B_1 \supseteq \dots \supseteq B_k \dots$ is a series of measurable sets converging to measurable set B , then $P(B) = \lim_{k \rightarrow \infty} P(B_k)$.

His comments on this are curiously phrased. They suggest that it is a special feature, which was not mentioned before, because he wished to deal with finite families of events (as well). But of course, 1 does not fail in the finite case; it holds trivially there. It fails only in the mathematically curious case in which the space is a countable sequence, and P the limit of relative frequencies. If 1 were added to the calculus, that would rule out the relative frequency interpretation; again, unless the designation of the class variables A, B, \dots were restricted so as to eliminate violations.

It would of course be contrary to Reichenbach's basic intentions to solve any problem by restricting the range of the class variables. For he wishes to say that the probability of A exists if the limit of the relative frequency of A in the long run exists – as a general assertion explicating probability.

5. Higher-level Probabilities

In Chapter 8, Reichenbach considers probabilities of probabilities, assertions such as we might like to symbolize

- (1) $P(P(C/B) = p/A) = q$.

He writes these in the forms

- (2) $(k)(x_k \in A \xrightarrow{q} [(i)(y_{ki} \in B \xrightarrow{p} z_{ki} \in C)])$.

That would be a *second-level* assertion. It is abbreviated

- (3) $(A^k \xrightarrow{q} (B^{ki} \xrightarrow{p} C^{ki})^i)^k$.

A *third-level* assertion would take the form

- (4) $(A^k \xrightarrow{r} (B^{mk} \xrightarrow{q} (C^{mki} \xrightarrow{p} D^{mki})^i)^m)^k$.

The way to understand it, I think, is to think of long runs of experiments done in the Central Laboratory at the rate of one per day – these are designated by x_i, y_i, \dots and so on. On every day, however, long runs of experiments are done in the Auxiliary Laboratory; the ones done on the k th day are designated by x_{ki}, y_{ki}, \dots .

In that way, we certainly get non-trivial iterations of the probability implication. I doubt that it is the right explication of what we mean. For although $(B \underset{p}{\Rightarrow} C)$ looks like it is a constituent of formula 3, it really is not.

The truth conditions of 3 and those of

$$(5) \quad B \underset{p}{\Rightarrow} C$$

have nothing to do with each other, because in 5 we correlate sequences of experiments in the Central Laboratory, with respect to B and C , while in 3 we correlate sequences in the Auxiliary Laboratory with respect to B and C .

Let me suggest something that might be a slight improvement. First, I want to do everything in terms of a single function symbol t rather than the diverse ones x, y, z, \dots . For any finite sequence $ijk \dots n$ of positive integers, let $t_{ijk\dots n}$ be a countable sequence of events. So one event would be $t_{21}(3)$ for example. Now, in some of these long runs of events, such as t_{21} , we find they are all A . This is a long run selection from class A . And we might ask: in those selections from A , how probable is it that half the B 's are C 's?

$$(6) \quad \text{relf}(\{k : \text{relf}(C/B, t_k) = q\} / \{k : (i)(t_{ki} \in A)\}, \sigma) = q$$

where σ is the natural number series, is then the explication of assertion 1.

A slightly different construction puts the probability implication in the antecedent. Thus

$$(7) \quad P(A/P(C/B) = p) = q$$

would on the Reichenbachian explication indicate how often an event in class A occurs in the Central Laboratory, on days when the Auxiliary Laboratory reports a relative frequency p of C 's among B 's. On my version it would be explicated as

$$(8) \quad \text{relf}(\{k : (i)(t_{ki} \in A)\} / \{k : \text{relf}(C/B, t_k) = p\}, \sigma) = q.$$

That is, it reports how many long runs in which there is a proportion p of C 's among B 's, are selections from the class A .

My version is semantically just as odd as Reichenbach's, though it has the advantage that in both sides of the slash mark, the same class $\{t_k : k \text{ a positive integer}\}$ is being talked about. To finish let me just indicate what a third level sentence would look like.

$$(9) \quad P(P(D/C) = p/B) = q/A) = r$$

is explicated by Reichenbach with formula 4, and here with

$$(1) \quad \text{relf} (\{k : \text{relf} (\{i : \text{relf} (D/C, t_{ki}) = p\} / \{i : (j)(t_{ki}(j) \in B)\}, \sigma) = q\} / \{k : (i)(t_k(i) \in A)\}, \sigma) = r.$$

IV. THE TRANSCENDENCE OF PROBABILITIES

There is something factual about relative frequencies, and something counterfactual about probabilities. What is probable is a gradation of the possible. And what is likely to happen is what would happen most often if we could realize the same circumstances many times over.

This is the logic of the concept. It implies, to my mind, nothing about the ontology we must accept. But if a physical theory contains probability assertions, then it contains modal assertions, for that is what they are.⁶ It is important to see first what such a theory literally says, before we ask what accepting the theory involves.

The view I shall develop here uses a modal interpretation of statistical theories.⁷ But it is also a version of the frequency interpretation. Literally construed, probabilistic theories may posit irreducible modal facts – propensities – in the world. But a philosophical retrenchment is possible: we can accept such a theory without believing more than statements about actual frequencies.

1. *Propensities Construed*

Probabilities cannot (all, simultaneously) be identified with the relative frequencies in the actual long run. If we wish to identify them with something real in the world, we must postulate a physical counterpart. In the case of modalities and counterfactuals, realist philosophers postulate dispositions. Probabilities are graded modalities; graded dispositions are

propensities. And propensities are just right, because they are postulated to be so.

What exactly is the structure of propensities? Some account is needed, if they are not to be merely an *ad hoc* 'posit'. ('Posit' is a verb which Reichenbach used only in reasonable ways, to indicate empirical hypotheses. In realist metaphysics, however, the operation seems to be conceived on the model of laying eggs, and equally productive.) Kyburg attempted to provide such an account, which represents propensities by means of relative frequencies in alternative possible worlds [5]. In this way he tried to show that there is no real difference between 'hypothetical frequency' and 'propensity' views of probability. This attempt did not quite work because the failure of countable additivity in each possible world infected the function that represented the propensity.

However, the representation I gave in Part I, Section 4, shows how this can be done. In each possible world we select a *privileged* family of sets, with respect to which relative frequency is countably additive in so far as that family contains countable unions. In other words, a *model structure* for the language of probability, will be a set of possible worlds, which together form a *good family of special frequency spaces*. The representation theorem then shows how to interpret probability. So Kyburg was right in the main.

Thus propensities do have structure, which can be described in terms of relative frequencies in different possible worlds. It also allows the following to happen; a certain coin has a propensity of $\frac{1}{2}$ to land heads up, but in fact never does. (The Rosencrantz and Guildenstern Problem, or, if you like, the Tom Stoppard Paradox.) It allows this in the sense of implying that it is possible, occurs in some possible world; which is not to say that our theories would be perfectly all right if it happened.

Of course, if we had a real coin never landing heads up, and since we do not have an ontological telescope which discloses propensities, we would more likely conjecture that this coin has a propensity *zero* to land heads up. This is a typical feature of realist metaphysics: the appearances do not uniquely determine the reality behind them. But it is more than that. It is a typical feature of scientific theories that actual occurrences do not uniquely determine one and only one model (even up to isomorphism) of the theory. Just consider the old problem of mass in classical mechanics: an unaccelerated body may have any mass at all (compatible with the

empirical facts), but in any model of mechanics it has a unique mass. There are many such examples. There are some subtle distinctions to be drawn about possibility here, to which I shall return in the last section.

2. *Nominalist Retrenchment*

The bone of contention between medieval realists and nominalists was not so much properties *überhaupt* as causal properties. Fire burns by virtue of its heat, a real property whose presence explains the regularity in fire-involving phenomena. The phenomena do not uniquely determine the world of real properties behind them. As Ockham pointed out, God could have created the world lawless with respect to any connection between fire and burning, but simply have decreed in addition an actual regularity (He directly causing the burning, e.g., the wood turning to charcoal, on all and only those occasions when fire is sufficiently proximate). So we cannot *infer* the real causal properties, dispositions and hidden powers. But we can postulate them, and then reap the benefit of their explanatory power. At this point, the nominalist can reject only the why-questions the realist wants answered (“No, there is no underlying reason for the regularities in nature – they are actual but not necessary”) and maintain that science has the air of postulating hidden powers only because it wishes to systematize the description of actual regularities.

To give a systematic description of the actual regularities is to exhibit them as an arbitrary fragment of a larger unified whole. Since actuality is all there is, there is only one picture we can form of that larger whole: it is the system of all possible worlds. Thus Kant in his *Inaugural Dissertation*:

... the bond constituting the *essential* form of a world is regarded as the principle of *possible interactions* of the substances constituting the world. For actual interactions do not belong to essence but to state. ([9], page 40.)

To a nominalist, this picture of the actual as but one of a family of possibles can be no more than a picture; but it can be granted to be the picture that governs our thinking and reasoning.

A distinction must be drawn between what a theory *says*, and what we believe when we accept that theory. Science is shot through and through with modal locutions. The picture the scientist paints holds irreducible necessities and probabilities, dispositions and propensities. Translation without loss into a ‘nominalist idiom’ is impossible – *but it is also a*

mistaken ideal. The nominalist should focus on the *use* to which the picture is put, and argue that this use does not involve automatic commitment to the reality of all elements of the picture.

3. *Epistemic Attitudes*

To explain how a nominalist retrenchment is possible in the specific case of probabilistic theories, I shall try to answer two questions: what is it to accept a (statistical) theory? and, what special role is being played by the 'privileged' sets in my representation?

Elsewhere I have given a general account of acceptance versus belief.⁸ A scientific theory specifies a family of models for empirical phenomena. Moreover, it specifies for each of these models a division into the observable parts ('empirical substructures') and the rest. The theory is true if at least one of its models is a faithful replica of the world, correct in all details. If not true, the theory can still be *empirically adequate*: this means that the actual phenomena are faithfully represented by the empirical substructures of one of its models. To believe a theory is to believe that it is true; but accepting a theory involves only belief in its empirical adequacy.

This is a distinction between two epistemic attitudes, belief and acceptance. I am not arguing here, only presenting my view, in a quick and summary manner. A similar distinction must be drawn in the attitudes struck to a particular model, when we say that the model 'fits' the world. This can mean that it is a faithful replica, or merely that the actual phenomena are faithfully mirrored by empirical substructures of this model.

When the model is of a family of possible worlds, then the latter belief, which is about the actual phenomena, only requires that at least one of these possible worlds exhibits the requisite fit. This is to the point when the theory is a probabilistic theory, for in that case it will never narrow down the possibilities to one. It will always say merely: "The actual world is among the members of this family of possible worlds, so-and-so constituted and related."

Now it is possible to see how a statistical theory in general must be constructed. Each model presents a family of possible worlds. In each world, certain substructures are identified as observable. Empirical adequacy consists in this: the observable structures in our world correspond

faithfully to observable structures in one of these possible worlds in one of these models. But all this fits in well with my representation of a probability space (which is what a model of a statistical theory must be). For I represent it as a family of possible worlds, in each of which a particular family of events is distinguished from the other events (as 'privileged'). These privileged events can be the candidates for images of observable events.

To see whether this holds, we must look at the testing situation. It is very important to be careful about the modality in 'observable'. Recall that we are distinguishing two epistemic attitudes to theories, in order to see how much ontological commitment is forced upon *us* when *we* accept a theory. Since we are trying to answer an anthropocentric question, our distinctions must be anthropocentric too. It will not do to say that DNA is observable because there could be creatures who have electron microscopes for eyes, nor that absolute simultaneity is an *observable* relationship, because it could be observed if there were signals faster than light. What is observable is determined by the very science we are trying to interpret, when it discusses our place in the universe. In the testing situation for a statistical theory, we use a language in which we report actual proportions in finite sets, and estimate (hypothesize, postulate, extrapolate) relative frequencies in many other sets. The explicitly checkable sets are those which can be described in this language – and these are the sets which we try to match against the privileged sets in some possible world in the model.

I have already indicated in Part I how I see this. We consider the totality of all sets on whose relative frequency we shall explicitly check in the long run – *whatever they be*. We estimate their probabilities on the basis of actual finite proportions in samples. Since in our language we have the resources to denote limits, by means of expressions like $\lim_{n \rightarrow \infty}$, \sum^∞ , \bigcup^∞ , it follows that we shall sometimes estimate relative frequencies of limits of sets we have treated already; and we make this estimation by a limiting construction. But that means that we do so by countable addition. For in the presence of finite additivity, the postulates

$$(1) \quad m(B) = \lim_{n \rightarrow \infty} m(B_n) \text{ if } B_1 \supseteq \dots \supseteq B_k \supseteq \dots$$

converges to B

$$(2) \quad m(B) = \sum_{n=1}^{\infty} m(B_n) \text{ if } \{B_n\}, n = 1, 2, \dots \text{ is a}$$

countable disjoint family

are equivalent.

Now the sets *explicitly* checked and subject to such estimating hypotheses, are all explicitly named in our language. This language is not static, it can be enriched day by day. But even so, it will be a language with only countably many expressions in it, even in the long run. So the family of privileged sets – the ones encountered explicitly in our dialogue with nature – *whichever they may be*, shall be countable.

The probabilistic theory says that our world is a member of a certain good family of special frequency spaces. Let us consider the Rosencrantz and Guildenstern problem again. Suppose that the theory implies that coins of that construction land heads with probability one-half.⁹ Imagine also that this coin is tossed every day from now on, and lands heads every time. In that case the theory is false, and that because it is not empirically adequate. For we have here a class of events described in our language, whose actual relative frequency is not equal to the probability the theory attributes to it. If this particular coin's behaviour is an isolated anomaly, the theory may still be a useful one, in some attenuated way. But of course, what I am describing here is not a situation of epistemic interest. What I say is an extrapolation from this: we would know only that the first x tosses had landed heads, and we would count this as *prima facie* evidence against the theory; but as long as we accept the theory we shall assert that in the long run, this coin too will land heads half the time. It seems mistaken to me to interpret statistical theories in such a way that they can be true if the relative frequencies do not bear out the probabilities. For if that is done (*and various propensity views suggest it*) then it is possible to hold this theory of our present example, and yet not assert that in the long run Rosencrantz' coin will follow suit – or indeed, *any* coin, *all* coins. The empirical content disappears.

What is important about Rosencrantz' coin is that the model should have in it some possible world (which we emphatically deny to be the actual one) in which an event described by us here shows aberrant relative frequencies. What we must explain is the assertion that this is always possible, but never true. (To be logical: as long as we are looking at a

model of *this* theory, then in each possible world therein it is *true* that if X is a described event, the probability of X equals its relative frequency, and it is *also true* in each world that *possibly*, the relative frequency of X is not its probability. The language in which X is described, is here given a semantics in the fashion of ‘two-dimensional modal logic’, so that a sentence can be true in every world where that sentence expresses a proposition, but still it does not express a necessary proposition. Consider ‘I am here’ which is true in every world in which there are contextual factors (a speaker, a place) fixing the referents of the terms, but which expresses in our world the contingent proposition that van Fraassen is in Toronto.)

To summarize, suppose that a physical theory provides me with probabilistic models. Each such model, as I interpret science, is really a model of a family of possible worlds. The empirical content of the theory lies in the assertion that some member of that family, for some model of the theory, fits our world – insofar as all empirically ascertained phenomena in the long run are concerned, whatever they may be.

To accept a theory, in my view, is only to believe it to be empirically adequate – which is, to believe that empirical content which I just indicated. So to accept a probabilistic theory involves belief only in an assertion about relative frequencies in the long run.

In this, I feel, Reichenbach was right.

University of Toronto

NOTES

* The author wishes to thank Professors R. Giere (Indiana University), H. E. Kyburg Jr. (University of Rochester), and T. Seidenfeld (University of Pittsburgh) for their help, and the Canada Council for supporting this research through grant S74-0590.

¹ I use Kolmogorov’s term ‘Borel field’ rather than the more common ‘ σ -field’, because I shall use the symbol sigma for other purposes.

² R. von Mises claimed countable additivity in [12], Chapter I; but the failure of this property for relative frequencies has long been known. See [4], pages 46 and 53 and [3], pages 67–68; I owe these two references to Professors Seidenfeld and Giere respectively.

³ This argument is reported in [10], pages 3–59 and 3–60; I owe this reference to Professor Giere.

⁴ Professor Kyburg pointed out to me that the conclusion does not follow from the Bernoulli theorem. Needed for the deduction is also the Borel–Cantelli lemma. For details see Ash [1], Chapter 7; especially the theorem called Strong Law of Large Numbers for Identically Distributed Variables.

⁵ This subject, and Kolmogorov's discussion of related problems, will be treated in a forthcoming paper 'Representation of Conditional Probabilities'.

⁶ This point of view about modality in scientific theories was argued strongly in a symposium at the Philosophy of Science Association 1972 Biannual Meeting, by Suppes, Bressan, and myself; the emphasis on probability assertions as modal statements was Suppes'.

⁷ In other publications I have developed a modal interpretation of the mixed states of quantum mechanics which can be regarded as a special case of the present view.

⁸ In my 'To Save the Phenomena'; a version was presented at the Canadian Philosophical Association, Western Division, University of Calgary, October 1975; a new version will be presented at the American Philosophical Association, Eastern Division, 1976.

⁹ I am keeping the example simple and schematic. Our actual beliefs, it seems to me, imply only that in the set of all tosses of all coins, which is most likely finite but very large, half show heads. The same reasoning applies. Our belief is false if the proportion of relative frequency is not one-half. This is quite possible – we also believe *that*. But would be a mistake to infer that we believe that perhaps the proportion is actually not one-half. We believe that our beliefs could be false, not that they are false – what we believe are true but contingent propositions. (This problem can also not be handled by saying that a theory which asserts that the probability is heads equals one-half implies only that the actual proportion is most likely one-half; for then the same problem obviously arises again.)

REFERENCES

- [1] Ash, R. B., *Real Analysis and Probability*, Academic Press, New York, 1972.
- [2] Belnap, Jr., N. D., 'Conditional Assertion and Restricted Quantification', *Nous* 4 (1970), pp. 1–12.
- [3] Fine, T., *Theories of Probability*, Academic Press, New York, 1973.
- [4] Kac, M., *Statistical Independence in Probability, Analysis and Number Theory*. American Mathematical Association, Carus Mathematical Monograph # 12, 1959.
- [5] Kyburg, Jr., H. E., 'Propensities and Probabilities', *British Journal for the Philosophy of Science* 25 (1974), pp. 358–375.
- [6] Popper, K. R., *The Logic of Scientific Discovery*, Appendices *iv and *v. revised ed., Hutchinson, London, 1968.
- [7] Reichenbach, H., *The Theory of Probability*, University of California Press, Berkeley, 1944.
- [8] Renyi, A., 'On a New Axiomatic Theory of Probability', *Acta Mathematica Hungarica* 6 (1955), pp. 285–333.
- [9] Smith, N. K. (ed.), *Kant's Inaugural Dissertation and Early Writings on Space*, tr. J. Handiside, Open Court, LaSalle, Ill., 1929.
- [10] Suppes, P., *Set-Theoretical Structures in Science*, Mimeo'd, Stanford University, 1967.
- [11] van Fraassen, B. C., 'Incomplete Assertion and Belnap Connectives', pp. 43–70 in D. Hockney *et al.* (ed.) *Contemporary Research in Philosophical Logic and Linguistic Semantics*, Reidel, Dordrecht, 1975.
- [12] von Mises, R., *Mathematical Theory of Probability and Statistics*, Academic Press, New York, 1964.