RANDOMNESS AND INFINITY

Michiel van Lambalgen*
Faculteit Wiskunde en Informatica
Universiteit van Amsterdam
Plantage Muidergracht 24
1018 TV Amsterdam
michiell@fwi.uva.nl

PART I: RANDOMNESS

Everyone knows examples of phenomena commonly associated with chance: tossing a coin, throwing a die, playing roulette or lotto. Less familiar perhaps are examples from physics such as radioactive decay or Brownian motion. What is chance, and why is it important to us?

If we have no way to predict whether a certain event (throwing a 6 with a die, say) will occur, we commonly ascribe the occurrence of the event to chance. A sequence of such chance phenomena, say a sequence of throws of a die, will look very irregular and chaotic to us; even if we already observed the outcomes of a thousand throws, we are not in a position to predict the outcome of the next throw. The 'gambler's fallacy' which holds that after a long spell of 6's, it becomes more likely to throw one of 1, .. ,5, is really a fallacy.

We might call this phenomenon local irregularity, or disorder. However, some chance phenomena exhibit remarkable global regularities: even though we cannot predict the outcome of the next throw, we feel confident that in the long run, the relative frequency of, say, 6 will stabilize around a certain value (1/6 in the case of un unbiased die). We shall call these global regularities *statistical* regularities. As a first approximation, we shall call a sequence *random* if it exhibits both these characteristics, local irregularity and statistical regularity. A sequence can exhibit statistical regularity without being (locally) irregular; an example is Champernowne's number obtained by writing the positive integers in an infinite sequence:

0.12345678910111213141516171819202122232425262728293031323334...

This sequence has the remarkable property that in the limit, the relative frequency of each digit converges to 1/10; more generally, the relative frequency of each block of digits of

* The author is deeply grateful to the Netherlands Organization for Scientific Research (NWO) for support in the form of the PIONIER-project "Reasoning with Uncertainty", under grant PGS 22-262.

length k converges to $1/10^k$. Clearly, however, this sequence is generated by a law, so it cannot be called disordered.

On the other hand, disorder does not necessarily lead to statistical regularity in the sense introduced above; examples, involving chance fluctations in waiting times, are somewhat too involved to be discussed here, so for details we refer the interested reader to Feller[1]. Roughly speaking, in some disordered sequences, the relative frequencies of events may oscillate erratically between 0 and 1, instead of converging to a fixed value.

Hence randomness is a non-trivial combination of properties. In fact it is so highly non-trivial that some have wondered whether it is amenable to theorizing at all: 'It may be taken for granted that any attempt at defining disorder in a formal way will lead to a contradiction. This does not mean that the notion of disorder is contradictory. It is so, however, as soon as I try to formalize it.'[2]

Here is a sample of questions one would like to ask about randomness. (1) Granted that a random sequence looks unpredictable *to us*, is there something objective about its unpredictability? At first sight, Champernowne's sequence may look as disordered and chaotic as a sequence generated by tossing a coin, but this impression is dispelled as soon as we learn its law of formation. Do there exist random sequences which are in some sense inherently unpredictable? (2) We have a more or less clear idea what the mechanics of, say, coin tossing is. That is, given knowledge of initial conditions such as momentum and angular velocity, we can calculate whether the coin will land heads or tails (see e.g. Keller[3]). What is the relation between this description and the observed sequence of outcomes? How (if at all) can we explain both the disorder and the statistical regularity in the outcome sequence starting from this deterministic mechanism? More generally, if we have a large number of independent trials of the same experiment (say an experiment to test Mendel's laws) we expect stabilization of relative frequencies; why is this expectation justified? (3) A priori, the aspect of randomness most difficult to formalize would seem to be disorderedness, not so much statistical regularity. Is it actually of practical importance to have a formal theory of randomness, incorporating a formal treatment of disorder? Some might argue that all we shall ever be interested in are simple statistical properties.

Below, we shall begin by answering the third question, examining the role of randomness in statistics. A word of caution is necessary here: the statistical inference problems we shall

[1]W. Feller, An Introduction to probability theory and its applications, Vol. I, Wiley (1970); for example Ch. III.

[2]H. Freudenthal, Realistic models in probability, in: I. Lakatos ed., Problems in inductive logic, North-Holland (1969).

[3]J.B. Keller, The probability of heads, *Am. Math. Monthly* 1986, 191 - 197

consider assume the so-called frequency interpretation of probability, which says that probability should be identified with (limiting) relative frequency in a large (infinite) number of trials. This is not the only possible conception of probability; for example, at the other extreme we find the interpretation of probability as degree of belief. On this interpretation, the role of randomness is much less prominent, so we shall not consider it here.

We then proceed gradually to more abstract matters, by considering various attempts to formalize the notion of randomness. In the second part of the paper, INFINITY AND FREE CHOICE, we show that consideration of the foundations of mathematics also forces us to take randomness, in one form or another, seriously. The reader with little mathematical inclination may find this part too abstract for his taste; but even a perusal of the material contained there may convince him that foundational research can be an exciting subject.

I.1 Statistical Inference

Consider the following example from psychological research.[4] A farmer visited the psychological laboratory, carrying with him a carved whalebone. He claimed that in his hands the bone was an extremely powerful instrument capable of detecting even small quantities of water; he told the psychologist that he had been able to locate waterwells for several of his neighbours using the bone. The farmer was somewhat disturbed by his ability and asked the psychologist to investigate him.

The psychologist decided to ascertain first whether the farmer indeed had this remarkable ability, by means of the following experiment.

On the floor was an array of 10 tin cans covered with plywood. The pieces of plywood carried numbers 1, ..., 10. Five of the cans had been filled with water by the psychologist's laboratory assistant, five had been left empty. The psychologist emphasized that the arrangement of the empty and filled cans was purely 'random' and that she herself did not know which of the cans contained water. She now wanted the farmer to take his bone and divide the set of tin cans into two groups, one with the cans containing water and the other consisting of the empty cans.

The question which the psychologist hoped to answer by performing this experiment is: can the farmer do any better in his choices than might be expected on the basis of chance? Presumably, if the farmer does better than chance, he indeed has the ability he claims to have.

[4] Taken from: A.L. Edwards, Experimental design in psychological research, Holt,Rhinehart and Whinston 1962.

It is worth while to spell out this experimental design in more detail, to see where randomness comes in. The cans could have several dinstinguishing characteristics, such as size, colour, shape, or minor damages. The laboratory assistant may have been guided in his choices (perhaps subconsciously) by any of these features. Likewise, if the psychologist would know which can contained water, she might give hints to the farmer without her knowing this. Also, there could be some correlation between the numbers 1, ..., 10 and a can's being empty or full; in an extreme case, the first five cans would be filled, while cans 5 to 10 would be empty. Randomness will prevent such things from happening: for instance, if we shuffle a deck of ten cards (corresponding to the cans) very thoroughly, and then draw five cards to determine which of the cans are to be filled, we may be reasonably sure that the allocation of water to cans is independent of all these characteristics.

Recall that the psychologist wanted to test the hypothesis (the so-called *null hypothesis*) that the farmer cannot do better than chance. The argument leading up to this hypothesis goes as follows. Suppose the farmer has no ability to detect water. The random allocation of water to cans ensures that the farmer has no clues whatsoever about which cans contain water and which do not. It follows that his choices can only be haphazard; and we might say that in this case his choices are no better (or worse) than if he would randomly select a group of five cans filled with water from among all possible allocations. But if that is so, we can compute the probability of a specific result under the hypothesis that the farmer has no ability. For instance, suppose all his guesses are correct; the probability that this happens by chance is 0.00397 and we might consider the hypothesis refuted thereby.[5]

I should add here that the above argument contains a subtle inference which hides a wealth of interesting problems: from assuming that the farmer's choices are haphazard to concluding that we can assign a probability to his choices, and in fact model his behaviour by randomly drawing five cards from a deck of ten. There is ample psychological evidence (see for example Bar-Hillel and Wagenaar[6]) to show that, when people are asked to imitate chance, they fail to do so; in particular, the sequences produced differ from random sequences with regard to fluctuations. Why, therefore, should we assume that people's haphazard choices *can* be modelled by probability? This problem points to an important issue in reasoning under uncertainty, namely: when can uncertainty be described by probability?
In any case, it is clear that randomization is *necessary* (though perhaps not sufficient) to make the probability model (which says that under the null hypothesis, the farmer's choices are modelled by independent tosses of a fair coin) applicable. Without some such device, an

---

[5]Strictly speaking, of course, we should be talking here about tests of significance and hypothesis tests.
[6]M. Bar-Hillel and W.A. Wagenaar, The perception of randomness, *Adv. in Applied Math.* **12** (1991), 428-454.

improbable test result could not be interpreted as evidence for an unusual ability on the farmer's part.

Let us next consider a second example of the use of randomness in statistics: random sampling. The statistical essence of sampling is the possibility to draw conclusions about a population on the basis of a small sample from this population; opinion polls are a familiar example. So suppose we want to test the hypothesis that Labour will get more than half of the votes during the next elections. We draw a sample from the population, determine the percentage (say 49%) of pro-Labour votes in this sample, and then, working under the hypothesis that Labour will receive more than 50% of the votes, compute the probability that in a sample 49% of the votes are pro-Labour. If this probability is lower than some antecendently specified value, we can decide to reject the hypothesis.

This probability model is applicable only under the assumption that the sample has been drawn *randomly*, which means (in the simplest case) that every member of the population has an equal chance of being drawn, and that each choice is independent of all the others. In this case, we can calculate the probability of drawing a sample of any prescribed composition. In other words, the first function of random sampling in statistics is to allow us to *compute* the relevant probabilities. Haphazardly choosing individuals from a population does not ensure computability.

However, there is also a second aspect to the use of random sampling: the justification of the conclusion of the inferential process, for example that the null hypothesis is rejected. For the conclusion to be justified, it is imperative that the sample is somehow representative of the population. For example, it must exhibit roughly the same distribution of frequencies as the population from which it was drawn. Let us provisionally call a sample random with respect to a population if it has this property (we shall see below that this notion still has to be refined and extended).

We expect that the method of random sampling will indeed produce a sample whose distribution is an accurate reflection of the distribution of the parent population, *but there is no mathematical guarantee that this will be so.* If we toss a coin a thousand times, the great majority of sequences which could be obtained this way will have a roughly equal number of zeros and ones, but any specific sequence is as likely as any other, so in particular a sequence which consists only of ones has the same probability (namely 2^{-1000}) as a specific sequence with 500 zeros and 500 ones.

It follows that the process of statistical inference outlined above contains a tacit non-trivial assumption, namely that a randomly drawn sample is indeed random. Of course tests for randomness of a sample exist, but they are not applicable here, since the distribution of the parent population is unknown. Therefore the assumption of randomness of the sample is part of the hypothesis being tested. To summarize, in statistical inference we need *both* the notion of random sampling, i.e. a certain probability model, *and* the notion of a random sample as the outcome of the sampling process.

Having established the necessity of a concept of random sample, let us investigate what kind of properties it should have. One property has already been noted: the frequency distribution should be roughly the same as in the parent population. While this property may be sufficient for the simple situation sketched above, in general we need something more, as is made clear by the following experimental set-up.

To test a certain psychological hypothesis, we draw a sample of 1000 individuals from a population, and we specify in advance that the first 500 individuals are to comprise the experimental group, while the remainder will form the control group.

It is clear that if the sample is such that, say, 490 individuals of the experimental group are women, we are in trouble; even if the remaining 510 individuals are men, so that the original sample of 1000 is more or less an accurate reflection of the parent population, it is very likely that our experimental results will be worthless because of biases in the two groups considered separately. What we require here is that the two subsamples are also representative of the parent population. Furthermore, we require this not only for this particular division of the sample into two subsamples; it should hold for *any* antecendently specified method of subsample selection, and even for a *random* assignment of individuals to experimental group or control group (a random assignment might be necessary to ensure objectivity of the experiment, as in the first example). For instance, if we decide to put every odd individual into the experimental group this group should be representative of the whole population; likewise if we toss a coin to determine if an individual is to belong to the control group or the experimental group (note that in this case the sizes of the groups can be different). It is important here that the method of selection is specified in advance (i.e. before drawing the sample); if we select subsamples based on properties of the sample, we are sure to introduce biases. Moreover, ideally these various forms of representativeness should hold for *any* attribute which could possibly be relevant to the hypothesis under consideration. Otherwise, the decision to accept or reject a hypothesis could be based inadvertently on a peculiarity of the sample. This is quite a lot to ask, of course; in general we do not even know which attributes are relevant to a hypothesis. For example, psychological dispositions could be influenced by unknown biochemical factors, and our method of drawing a sample might, unbeknownst to us, introduce a bias with respect to those factors.

Before we go on to consider attempts to make the notion of random sample precise, let us pause to see where we stand. We would like to have a method of randomly drawing a sample which would guarantee that the sample is indeed random in the strong sense outlined above. However, no probabilistic method can achieve this with complete certainty; it is always possible that we draw a sample which is very non-representative in some respects. In this sense the conclusions of a statistical analysis (acceptance or rejection of an hypothesis) always contain a proviso. One might argue that, since we can compute the probability of a non-representative sample being drawn (if we consider known attributes only), the uncertainty surrounding the conclusion can always be quantified. However, this probability *also* has to be interpreted as a relative frequency, and to do so we need a long sequence of drawings of a sample from the parent population; in this long sequence the relative frequency of a non-representative sample would have to approach the probability, or in other words, this sequence has to be random. This means that we cannot avoid making the assumption of randomness at some point, so we might as well do it at the earliest available opportunity, which in this case amounts to assuming the sample is random.[7] It follows, therefore, that we cannot do without the informal notion of randomness outlined above.[8]

I.2 Defining randomness

As we have seen above, randomness is characterized by local irregularity or disorder, and global or statistical regularity. Most definitions proposed can be classified into one of two types: those that take statistical regularity as primary, and those that take their starting point from disorder. We shall consider an example from each of these approaches.

The first more or less formal definition of a random sample was given by von Mises in 1919[9]; with reference to the foregoing examples we can state it as follows. Consider a population P, from which we are going to draw a random sample. For simplicity, we shall consider only sampling with replacement, so that at every trial, each individual has the same probability of being drawn. Having drawn an individual, we test whether or not it has attribute A, and write a 1 or a 0 for the result of the test. Accordingly, our sample consists of a sequence of 0's and 1's, which for simplicity we shall suppose to be infinite at first. Let x

---

[7] The point made just now is related to a common misconception concerning the so-called laws of large numbers. These theorems say that under certain circumstances (e.g. independence assumptions) the relative frequencies in 'almost all' sequences converge to the probabilities (where the meaning of 'almost all' depends on whether the sequences considered are finite or infinite). Many scientists have taken these results to be a justification of the frequency interpretation of probability. They are nothing of the sort, however, since the phrase 'almost all' still has to be given a frequentistic interpretation. This point is customarily glossed over by saying that if the outcome of an experiment has very large probability, then it is practically certain that this outcome will occur on the first trial (see for example A.N. Kolmogorov's famous 'Grundbegriffe der Wahrscheinlichkeitsrechnung', J. Springer (1933), p. 4). However, this is nothing but the frequency interpretation in disguise.

[8] For a thoughtful discussion of this point, one might compare M.G. Kendall's 'A theory of randomness', *Biometrika*, XXXII, Part I (1941), 1-15.

[9] Grundlagen der Wahrscheinlichkeitsrechnung, *Math.Z.* **5**(1919), 52-99.

be this infinite binary sequence. We write x_k for the k^{th} element of the sample. The relative frequency of the occurrence of 1 in the first n elements of the sample (that is, the number of occurrences of 1, divided by n) is denoted as rel(x,n). Let P(A) = p be the frequency of A in P.

Then von Mises' two conditions on random samples (which are the same as the conditions introduced informally above) read as follows:

(i) $\lim_{n\to\infty} \mathrm{rel}(x,n) = p$

(ii) let x' be a subsequence of x which is determined in the following way: the decision to include x_k in x' does not depend on the value of x_k (such a procedure is called an *admissible place selection*). Then also $\lim_{n\to\infty} \mathrm{rel}(x',n) = p$.

The second condition is meant to be an abstract formulation of the requirements for subsample selection which were discussed above. Examples of admissible place selections are: choose x_k if k is even, choose x_k if $x_{k-10},\ldots,x_{k-1}$ are all equal to 0, choose x_k if tossing a coin yields heads.

The guiding intuition behind von Mises' definition is *independence* : because the successive x_k are independently chosen, an admissible place selection cannot yield a subsequence with a different distribution.

An infinite binary sequence x satisfying conditions (i) and (ii) is called a *random sequence*. (Strictly speaking, a sequence is random *with respect to a given distribution* P, but it will usually be clear from the context which distribution is meant.)

By definition, a random sequence is statistically regular. It is also in a sense disordered, because a pattern could be used to define an admissible place selection which would violate condition (ii). For example, Champernowne's sequence is not random (here we consider the obvious extension of von Mises' definition to sequences consisting of digits 0, .., 9); although it satisfies the first condition, we can easily falsify condition (ii) by means of the following admissible place selection, which picks out only 1's: consider Champernowne's number as the sequence of positive integers and for every k = 0,1,2,..., consider the integer n in the sequence which occupies position k.10 + 2 (i.e., n = 1,11,21,...), and choose the last digit of n. Generalizing this example, we can see that a random sequence could never be given by a mathematical law; if it were so given, we could use the law to define an admissible place selection violating (ii). In particular, sequences generated by so-called random number generators on pocket calculators or personal computers can never be random in this sense; although their statistical behaviour is sometimes surprisingly good, their lawlike origin (often a version of the linear congruence algorithm) disqualifies them. This raises serious questions about the mathematical status of random sequences: do they exist at all, in the sense that the concept of randomness is non-contradictory? We shall take up these questions in the second part of this paper.

The main empirical assumption underlying statistical inference can now be stated as follows: random sampling produces a random sequence. However, this formulation is an improvement on the informal discussion given above only if von Mises' definition is sufficiently precise, and here doubts set in. Granted that the examples of admissible place selections are clear, exactly what does it mean that 'the decision to select x_k does not depend on the value of x_k'?

We shall not review all the attempts to turn this appealing informal idea into an exact mathematical concept, but one approach, that of Kolmogorov[10], must be explained at some length.

Kolmogorov is the founding father of modern axiomatic probability theory (cf. his book cited in footnote 7). He strongly believed that a formal theory of randomness is necessary to explain the applicability of probability theory and statistics to the real world, but he was dissatisfied with von Mises' definition, both for its impreciseness and its infinitary character. Kolmogorov's starting point was disorder, or, as he called it, *complexity*.
Imagine a computer C which prints a binary sequence when given some binary sequence as input. We shall take C to be *universal*, which means that C is able to simulate any other computer C'. In other words, given any computer C' and any input p for C', if we concatenate the description of C' (written in binary code) and p into a new binary sequence <C',p> and if we give <C',p> as input to C, then C applied to <C',p> gives the same output as C' applied to p. (Theoretically, universal computers exist; it is clear, however, that such a computer would require unbounded memory, so actual computers are only an approximation to this ideal.)
The intuition behind the notion of complexity is the following. A regularity or a pattern can be seen to be a configuration which is easy to describe, in the sense that the length of a description of the configuration can be much smaller than the length of the configuration itself. For example, a program very much shorter than 1000 bits can cause C to print the first 1000 digits of Champernowne's number. Here, we consider the program to be a description of the sequence printed by C when given this program as input.
On the other hand it seems difficult to construct efficient descriptions of sequences generated by for instance coin tossing. *To put this in other terms, the information contained in a regular*

[10]On tables of random numbers, *Sankhya* **25** (1963), 369-376; Combinatorial basis of information theory and probability theory, *Russ. Math. Surveys* **38** (1983), 29-40; both articles now have been reprinted in his Collected Works , Kluwer 1994.

sequence is compressible, whereas this is not so for random sequences. (For an interesting experiment which suggest that ants use data compression for information transmission, see Li and Vitanyi[11], which contains a wealth of information about Kolmogorov complexity.)
If p is a binary sequence, we write |p| for the length of p, that is, the number of bits of p. We may now define the Kolmogorov complexity K(w) of a binary sequence w as follows:

K(w) is the length of the smallest program p such that C(p) = w.

An upper bound to K(w) is given by the length of w plus some fixed constant, because we can always take our program p to be "print w". We may call a sequence complex or *Kolmogorov random* if its Kolmogorov complexity roughly equals its length. A simple counting argument shows that there will be very many sequences w such that K(w) > |w| - 10, so for large w, K(w) is roughly equal to |w|.
Note that we defined Kolmogorov randomness without reference to statistical properties; but it can be shown that in a Kolmogorov random sequence, the relative frequency of 1 roughly equals 1/2, and that the same holds for subsequences which are selected by sufficiently simple algorithms, i.e. algorithms whose description is very much shorter than the complexity of the sequence. (It is also possible to give bounds on the complexity which ensure that the sequence is random with respect to some other distribution.) In other words, starting from the notion of disorder, we have arrived at a version of von Mises randomness, where 'admissible place selection' is now taken as 'subsequence selection described by a *simple* algorithm'. For example, if we assume that a sample is Kolmogorov random, then the subsample obtained by taking the odd-numbered individuals will also be random, because this algorithm is particularly simple. (But note that we still do not have a way to treat random selection of a subsample, since by definition a random selection is not simple.)

It might now seem that we have made great headway toward a satisfactory definition of randomness: we have a definition that applies to finite samples, i.e. the samples that we meet in practice, so in principle we could use the definition to generate random sequences and to test a given sample for randomness. Surprisingly, this is still not possible, because of the following result:
PROPOSITION (a) There does not exist an algorithm which, when given an integer k as input, produces a Kolmogorov random sequence of length k as output;
(b) There does not exist an algorithm which, when given a binary sequence as input, determines whether it is Kolmogorov random.
For the interested reader, we sketch proofs of (a) and (b); others may proceed directly to QED.

[11] M. Li and P.M.B. Vitanyi, An introdution to Kolmogorov complexity and its applications, Springer (1993); the experiment is described in chapter 8.

First, (b) follows from (a). If there would exist an algorithm which determines whether a sequence is Kolmogorov random, we can transform it into an algorithm which generates Kolmogorov random sequences as follows: given an integer k, generate all sequences of length k, determine for each of them whether it is Kolmogorov random and output the first such sequence (one can prove that there will always be at least one). Hence it suffices to prove (a).
Suppose there were a computer C' such that C' applied to an integer k gives as output a sequence w with K(w) roughly equal to the length of w, that is, k. Clearly the length of the input k of C' is of the order $\log_2 k$. Simulate C' on C, then C applied to <C',k> yields a sequence w with the above property. The length of the input <C',k> is equal to $\log_2 k$ plus some constant. For large enough k, this will be very much smaller than k. Hence for large enough k, if C applied to <C',k> yields w, K(w) (which is less than or equal to the length of <C',k>) is very much smaller than k, that is, the length of w. It follows that those w cannot be Kolmogorov random after all, so there cannot be such a C'. QED

As a consequence, although it can be proved that Kolmogorov random sequences exist, one can never effectively construct or recognize such a sequence. The proof makes it clear that if we would be able to generate a random sequence, this fact itself would count as a regularity. Observe that we encountered the very same phenomenon in the case of von Mises random sequences, when we showed that they could not be given by a mathematical law! Apparently there is an inherent tension in the notion of randomness: the more disordered a sequence is, the more difficult it is to prove its existence. This gives a partial answer to our first question: Do there exist random sequences which are in some sense inherently unpredictable? A fuller answer will be given in the section on infinity.

I.3 Can we explain why randomness occurs?

Above we remarked that the process of statistical inference contains a tacit assumption, namely that a randomly drawn sample is indeed random; and we also said that there is no mathematical guarantee that this will be so. Since this point is disputed, it is worthwhile to discuss this in greater detail. Consider the simple case of tossing a fair coin very often. We expect that the resulting sequence will be random in any of the senses discussed above. Why is this expectation justified? A fairly sophisticated answer would run as follows (cf. the reference cited in footnote 3).

Coin tossing is a deterministic process, and if we knew the initial conditions exactly, we could predict the outcome (heads or tails). The relevant initial conditions here are (vertical) velocity and angular momentum.

However, it is impossible to determine these initial conditions with any precision, hence it is customary to assume a probability distribution over the initial conditions. This distribution is taken to reflect our uncertainty concerning the true values of the initial conditions. The following statement is somewhat inexact, but will do for our present purposes: whatever the particular form of this probability distribution, in the region of large values for the velocity and the angular momentum the set of initial conditions leading to heads has about the same probability as the set of initial conditions leading to tails, hence the probability of heads is (approximately) equal to 1/2. The argument is concluded by an appeal to the law of large numbers: since the successive initial conditions are independent, so are the outcomes, and it follows that with overwhelming probability we observe sequences in which the relative frequency of heads is roughly equal to 1/2.

The virtue (or vice) of this argument is that it pushes the explanandum one step back: 'the outcome is random because the initial conditions are random'. What this explanation achieves is that it shows why frequencies are 1/2; but we have to assume random choice, in particular independence, of initial conditions. In other words, we have to assume that the sequence of initial conditions, corresponding to the sequence of tosses, is a random sequence. The mere fact that we can derive that the probability of observing a sequence in which the relative frequencies approach 1/2 is very high does not suffice, since we still have to interpret *this* as a relative frequency (cf. footnote 7). Clearly, then, in this instance, randomness cannot be reduced to something more primitive. We seem to be forced to accept as an empirical statement that under certain experimental conditions a sequence of outcomes is a random sequence. Probabilistic models of the sampling process can make this statement plausible, but they can never prove it conclusively.

A final remark: we have seen that the probabilistic model of coin tossing, a sequence of independent two-valued variables with the same distribution, can be obtained from a more detailed probabilistic model in which the mechanics of the process has been built in. On the other hand, there also appear to be phenomena, especially in quantum mechanics, which have the same 'surface-level' probabilistic model as coin tossing, but for which there is no underlying 'mechanistic' probabilistic model. Such phenomena may be called *inherently stochastic*; these phenomena also obey probabilistic laws, and thus are amenable to scientific treatment, but there is no underlying explanation using deterministic laws.

PART II: INFINITY AND FREE CHOICE

We shall now use randomness to shed light on one of the deepest problems in the foundations of mathematics: infinity. Let us say at the outset that if the reader, after finishing this part, thinks that the considerations presented here have nothing to do with real mathematics, he is mistaken; although there is a core of accepted mathematical principles (embodied in Zermelo-Fraenkel set theory), this core by itself is not sufficient to settle all mathematical questions. Fundamental concepts of mathematics, such as infinity, therefore have to be analysed time and again to yield additional axioms. Since the questions treated here are difficult, we cannot do more than give the reader a glimpse of what is involved.

The basic structures of mathematics, the natural numbers 0,1,2,3, ... or the real numbers (examples of which are π, $\sqrt{2}$, e, ...) are infinite. On the other hand, human minds are finite; it does not seem possible for us to survey the structure of natural numbers in its entirety. Therefore it is impossible to verify simple statements about *all* natural numbers directly, for example Goldbach's conjecture which says that every even number is the sum of two primes. This statement can be verified for arbitrarily large numbers, but to establish the truth of the conjecture we have to resort not to verification but to proof.

A proof is a logical chain of arguments starting from premises, so called axioms, which are taken as evident for the structure under consideration. These axioms are often derived from a law defining or determining the infinite structure. For example, the natural numbers are generated by the operation '+1', so an important axiom is the principle of induction, which says that for a property P, if we can prove that (i) P holds for 0, and (ii) if P holds for n, then also of n+1, then P holds for all the natural numbers.

Hence in this case infinity has in a sense been tamed because we can conceive of it as being generated by a (particularly simple) law. A fundamental issue in the ongoing debate about the foundations of mathematics is: should all infinite structures be given by a law, and if so, exactly what is a law? The question was first raised in this form around the turn of the century, by a group of French mathematicians including E. Borel and H. Lebesgue. The immediate cause for this question was the development of Cantor's set theory, which threw an unexpected light on the real number system.

For the present exposition, we can take real numbers to be characterized by an infinite decimal expansion (cf. Champernowne's number above); in the case of the rationals, this expansion is periodic, but for irrationals such as $\sqrt{2}$, π, or indeed Champernowne's number, the expansion is aperiodic. It had been taken for granted for some time that real numbers were also given by a law in the sense that there is a lawlike connection between natural numbers n and the n^{th} decimal of the real number. For instance, there exist algorithms which compute arbitrary decimals of $\sqrt{2}$, π or e. This instinctive feeling was at the basis of many criticisms of von Mises' definition of randomness: since all mathematical objects should be given by a law, and since it follows from the definition of a von Mises random sequence that it *cannot* be given by a law, von Mises random sequences do not exist, and perhaps even the notion of randomness itself is inconsistent. However, the idea that all real numbers must be given by a law, even though intuitively appealing, is not unproblematic. For suppose that indeed all real numbers are given by a law, then in a sense there cannot be more real numbers than there are laws; and if we take laws to be given by a definition in some language, then it follows that the real numbers can be 'listed' by making a list of all possible definitions in that language. However, Cantor's famous diagonal argument seemed to shatter this conclusion. Suppose we had a list of all real numbers between 0 and 1, say as follows:

1	0.67459987.........
2	0.45234426.........
3	0.89640287.........
4	0.74915359.........
..	....

We may now define another real number $a = 0.a_1a_2a_3a_4$....... by putting: a_n equals k+1, where k is the n^{th} decimal of the n^{th} number in the list. In this particular case, the decimal expansion of a would start off with 0.7672...... . By construction, a differs from each number in the list, so it cannot occur in the list.[12] But we assumed all real numbers between 0 and 1 to be listed. Contradiction.

One interpretation of this result is that there are many more real numbers than there are definitions, and then immediately the question arises: where do the 'extra' real numbers come from? One possible answer is: these real numbers are generated by 'free choice' or even by 'chance'; and this answer forces us to reconsider randomness for infinite sequences, with all its attendant problems of consistency.

The Dutch mathematician L.E.J. Brouwer (1881 - 1966), working on a revision of the classical foundations of mathematics, was the first to take this possibility seriously. He introduced various notions of 'free choice sequences'[13], one type of which, the so called lawless sequences, can be explained as follows. For Brouwer, mathematics is a free creation of the human mind; incidentally, a point of view diametrically opposed to the more usual conception of mathematics as a science. However that may be, if we assume that the human

[12]We ignore the slight inconvenience caused by reals having two decimal expansions.
[13]Begründung der Mengenlehre unabhängig vom logischen Satz vom ausgeschlossenen Dritten I: Allgemeine Mengenlehre, *Nederl. Acad. Wetensch. Verh. Tweede Afd Nat.* 12/5 (1918).

mind indeed freely creates mathematical objects, then we can imagine a mathematician generating a real number by freely choosing decimals, so that his choices are not subject to any restriction. In particular, this mathematician would not be obliged to make his choices conform to a law, so we would expect that, in a sense, there will be more real numbers than those given by a law. (Note the oblique connection with the psychologist's experiment with which we began this article: it was assumed there that the farmer's choices would be random when not guided by an unusual ability. Similarly, can we say that the mathematician's free choices are actually random? Interestingly, Brouwer's source of inspiration, the French mathematician Borel, argued that these choices are indeed random; this is in fact the historical root of his celebrated Strong Law of Large Numbers[14].)

A mathematician creating a real number not guided by a law, but by means of free choices, faces infinity as it were unarmed; if infinity is tamed by admitting lawlike infinite objects only, it reemerges under a wild aspect when free or arbitrary choice is allowed. Given the finitude of our human minds, what can one still say about sequences so generated? Brouwer's answer was, in a sense: nothing! But this emphatic 'nothing!' leads to a mathematical axiom (called the Principle of Open Data): if α is a sequence freely chosen by a mathematician, and if the statement A is true of α, then, since we can know only a finite piece of α (call it $\alpha(n)$), we must have for every freely chosen β which has $\alpha(n)$ in common with α, A is true of β. Axioms of this kind have fundamental consequences and in fact lead to a revision of logic. For example, the principle of excluded middle, which says that either A or not-A is true, now fails. This can easily be seen as follows: let $A(\alpha)$ be the property "there is a block of ten 7's in α", and suppose that $A(\alpha)$ is false, then not-$A(\alpha)$ would be true. Hence there would be a finite piece $\alpha(n)$ of α such that for all freely chosen β extending $\alpha(n)$, not-$A(\beta)$. But nothing can prevent a mathematician from choosing a β which *does* contain a block of ten 7's. Contradiction.
In other words, if all we know about a freely generated sequence is a finite piece, then the usual logical laws fail, and most mathematicians are not prepared to swallow this.

Nevertheless, concepts of free choice inevitably play a role in modern mathematics. An axiom that is often added to the basic core of accepted principles is the axiom of choice (first proposed by Zermelo in 1904), which in a simple case says the following. Suppose we have a (possibly infinite) collection of pairs of objects; then we may form a new collection

---

[14] The strong law of large numbers says that in 'almost all' reals (written as infinite decimal expansions) the relative frequency of a specific block of integers (say, ten 7's) of length k approaches 10^{-k}. This result, which now belongs to the standard tools of the probabilist, was considered very surprising at the time (1909); even in 1923, the mathematician Steinhaus called it 'le paradoxe de M. Borel'. This shows again that intuitions about randomness are by no means uniform or stable.

which contains exactly one object from each pair; in other words, there exists a function which to each pair assigns an element of that pair. This seems innocuous enough, but a homely example taken from the philosopher Bertrand Russell will make its full force clear.
First, consider an infinite collection of pairs of shoes. We know how to choose a collection which contains exactly one object from each pair of shoes: take all left shoes, for example. Now suppose that we have an infinite collection of pairs of socks; in this case there is no obvious way to distinguish the elements of the pair from one another. It follows that the choices we must make to determine the collection which contains one sock from each pair, are of necessity free, since there is no law to guide us. In this connection the following example from medieval philosophy may be instructive. The philosopher Buridan (*c.* 1295 - *c.* 1358) maintained that the will can only choose what the reason presents as the greater good. This leads to the predicament of 'Buridan's ass', which is faced by two equally appetizing bundles of hay and, unable to choose between them, dies of starvation.[15]

For the mathematically oriented reader we give one more example to make the difference between the two situations clear.
First, consider an infinite collection of pairs of real numbers. It is easy to define the required collection: take the smallest number from each pair. Let us now complicate this situation slightly. For each real number x, let [x] denote the set $\{x+q \mid q \text{ rational}\}$. Now suppose that we have an infinite collection of pairs of the form $\{[x],[y]\}$; in this case there is no natural linear order on the [x]. Can we still perform the choices to determine the required collection?[16] Zermelo-Fraenkel set theory does not give any information about this question.
Here, randomness comes into play again. Surprisingly, it turns out that randomness embodies a type of free choice which is incompatible with the freedom of choice necessary to choose from infinite sets of pairs, the elements of which are indistinguishable. The first ingredient of the argument is Brouwer's axiom of open data, alluded to above, which can be given a form which does not lead to a revision of logical laws. The starting point is the same: if we randomly or freely generate an infinite sequence α, then at any stage we know only a finite initial segment of that sequence. Now suppose A is a property which is insensitive to initial segments in the following sense: if $A(\beta)$ is true and γ differs from β only in a finite initial segment, then $A(\gamma)$ is also true. Call such a property A *asymptotic*. Then we may paraphrase Brouwer's axiom as follows: if A is an asymptotic property and A holds for *some* randomly generated α, then A holds for *all* randomly generated α. (If A were false for some randomly generated β, then A would make distinctions among randomly generated

[15] F.C. Coppleston, A history of medieval philosophy, Methuen (1972), p. 267.
[16] This is not an idle question; for instance if all bounded real functions are Lebesgue integrable, then we cannot.

sequences; but the only way to make distinctions is on the basis of finite initial segments, which A cannot do because it is asymptotic.)

The second ingredient of the argument borrows some ideas from von Mises. First some notation. If z is an infinite binary sequence, let [z] denote the set {x | x differs from z on at most a finite initial segment}. Let z^0 denote the sequence of even coordinates of z, and let z^1 denote the sequence of odd coordinates of z. Using this notation, we may now extrapolate some of von Mises' ideas on independence as follows:

(a) if z is random sequence, so are z^0 and z^1

(b) since we think of the successive choices (or coin tosses) generating z as independent, it follows that z^0 and z^1 are independent of each other

(c) conversely, if x and y are independent randomly generated sequences, and if z is such that $z^0 = x$ and $z^1 = y$, then z is itself random.

'Independent' is used as an intuitive concept here, but (a-c) have some simple consequences which can be stated without using the concept of independence. For example, suppose z is randomly generated; if x is such that $x^0 = z^1$, $x^1 = z^0$, then by (c), x is also random.

Intuitively, one may now argue as follows[17]. Let $\mathcal{C}$ be the collection of pairs $\{\{[z^0],[z^1]\} \mid z$ randomly chosen$\}$ ($\mathcal{C}$ is analogous to Russell's collection of pairs of socks). We show that it is impossible to choose an element from each pair of $\mathcal{C}$. Working toward a contradiction, suppose that g is a function which picks one element from each pair in $\mathcal{C}$.

For definiteness, we may suppose that there is a random z such that $g\{[z^0],[z^1]\} = [z^0]$. Now the property "$g\{[z^0],[z^1]\} = [z^0]$" is asymptotic in z, so by the analogue of Brouwer's axiom it follows that for *all* randomly chosen x, $g\{[x^0],[x^1]\} = [x^0]$. Consider a very special x, namely a sequence which is defined by $x_{2n} = z_{2n+1}$, $x_{2n+1} = z_{2n}$. The sequence x is just as random as z (by (c) above), and we have $[z^0] = [x^1]$ and $[z^1] = [x^0]$, hence $g\{[z^0],[z^1]\}= g\{[x^0],[x^1]\}$ also equals $[z^1]$, a contradiction.

So we have come to a surprising conclusion. When it deals with infinity, mathematics cannot do without some concept of free choice or randomness; but there are several kinds of free choice, which may be incompatible. It follows that question (1), "Do there exist random sequences which are in a sense inherently unpredictable?" does not admit an unambiguous answer; ultimately the answer depends on the mathematical axioms one adopts as basic. In particular, free choice of the Brouwer-von Mises type is not compatible with the type of free

---

[17] Formal proofs can be found in M. van Lambalgen, Independence structures in set theory, Proceedings Logic Colloquium '93, Oxford University Press (to appear). It should be added that the axioms stated informally above are actually true under such hypotheses as 'all sets of reals are Lebesgue measurable'.

choice postulated by Zermelo's axiom of choice. We now leave the reader to ponder the differences between them.