RANDOMNESS

William A. Dembski
Department of Philosophy
Northwestern University
Evanston, IL 60208

August 1993

Article for the *Routledge Encyclopedia of Philosophy*
Subject Editor (Philosophy of Science): Arthur Fine
General Editor: Edward Craig

[B]

As a theoretical notion randomness assumes five distinct senses. These comprise

(1) Randomness as the output of a chance process. Thus an event is random if it is the output of a chance process. Moreover, a sequence of events constitutes a random sample if all events in the sequence derive from a single chance process and no event in the sequence is influenced by the others.

(2) Randomness as mimicking chance. Statisticians frequently wish to obtain a random sample (in the sense of (1)) according to some specified probability distribution. Unfortunately, a chance process corresponding to this probability distribution may be hard to come by. In this case a statistician may employ a computer simulation to mimic the desired chance process (e.g., a random number generator). Randomness qua mimicking chance is also known as pseudo-randomness.

(3) Randomness via mixing. Consider the following situation: Particles are concentrated in some corner of a fluid; forces act on the fluid inducing a global dynamics; eventually the particles become thoroughly mixed throughout the fluid, reaching an equilibrium state. Here randomness is identified with the equilibrium state reached via mixing.

(4) Randomness as a measure of computational complexity. Computers are ideally suited for generating bit strings. The length of the shortest program that generates a given bit string as well as the minimum time it takes for a program to generate the string both assign measures of complexity to the strings. The higher the complexity, the more random the string.

(5) Randomness as pattern-breaking. Given a specified collection of patterns, an object is random if it breaks all the patterns in the collection. If, on the other hand, it fits at least one of the patterns in the collection, then it fails to be random.

[A]

1 Chance

2 Simulation

3 Mixing

4 Complexity

5 Pattern-Breaking

---

1 Chance

By far the most common conception of randomness identifies randomness with chance. Indeed, much of probability theory and statistics does not distinguish the two. Thus for a probabilist or statistician a random event and an event due to chance are typically the same thing. Moreover, the processes that give rise to such events are referred to indiscriminately as random, chance, or stochastic processes.

Within statistics the adjective 'random' assumes a technical sense when it occurs in the phrase 'random sample'. Given a chance process, one may wish to consider not just a single random event from this process, but rather an entire sequence of such events. Such a sequence is then said to constitute a random sample if (1) the same chance process is responsible for each event in the sequence and (2) the occurrence of any event in the sequence is unaffected by the occurrence of other events in the sequence. If the first condition is satisfied one says that the events are 'identically distributed'; if the second, that the events are 'independent'.

Identifying randomness with chance now raises the obvious question, to wit, What is chance? While there exists an entire metaphysics of chance related to causation, determinism, and free will, for this discussion it seems best to take an instrumental approach to chance, characterizing chance in terms of those processes like coin tossing and radioactive decay for which our best understanding is irreducibly probabilistic, that is, no finer level of analysis is

available which circumvents our probabilistic understanding (cf. CHAOS THEORY for which the probabilities are artifacts of underlying deterministic systems).

2 Simulation

Scientific research consists increasingly of computer simulations that generate vast amounts of data. Presumably, if scientists had sufficient time and resources to examine nature directly, computer simulations that imitate nature would be unnecessary. Practical limitations on investigating nature, however, seem to render computer simulations indispensable to scientific research.

A dilemma now confronts the scientist. For many purposes the data a scientist wishes to obtain should properly be the output of a chance process characterized by some well-defined probability distribution. Practical limitations, however, often prevent the scientist from actually sampling such a process and obtaining the desired data set (imagine a scientist who desires as data the sequence of heads and tails gotten by flipping a fair coin a billion times--the scientist's life will expire before the billion flips can be accomplished). In this case, the scientist will want to simulate the chance process computationally.

The dilemma then is this. On the one hand computers are fully deterministic devices--specify an algorithm, and the behavior of the machine is fixed. It follows that any probabilistic features of the data generated by a computer simulation are strictly eliminable. Yet on the other hand, such data are to substitute for data generated by a genuine chance process, data which cannot be characterized except in probabilistic terms. As the output of a chance process, truly random data (in the sense of §1)) are supposed to defy all but post hoc characterizations. As the output of an algorithm the (pseudo-) random data generated by a computer simulation are fully characterized in advance. How then can the twain meet?

Strictly speaking they can't. If randomness is identified with chance, then an event is random just in case it has the right sort of causal history, to wit, it was generated by a chance process. A computer is not a chance process. Ergo the data generated by a computer cannot be

random. John von Neumann summed up the matter as follows: 'Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin'.

Nonetheless, the incongruity of using not merely deterministic systems, but systems whose entire behavior can be precisely specified in advance has not dampened the proliferation of random number generators (RNGs). What then justifies using the data generated in a computer simulation in place of data generated by a chance process? In practice what happens is this. Given an RNG statisticians, as it were, set up a gauntlet of statistical tests that serve to vet it. The tests specify properties which the overwhelming majority of numerical sequences should have if they were generated by the chance process that the RNG is attempting to mimic. If the numerical sequences generated by the RNG don't have these properties, the RNG fails to make it through the gauntlet and is rejected. Otherwise it is considered adequate.

Although in practice RNGs do a lot of useful work, there remains a theoretical problem justifying RNGs in this way (and this unfortunately is the only way RNGs can be justified): any RNG is only as good as the last statistical test that it happened to pass. Indeed, the history of RNGs is strewn with RNGs that were for a time considered adequate, and then shown to be deficient. The problem is that we can never be sure that an RNG incorporates biases which the statistical tests we have thrown at it have simply failed to detect. This is bad. Practically speaking this means that the scientific literature may be filled with type I errors which we shall be unable to root out until appropriate statistical tests are found that detect the biases. For instance, cosmologists whose computer simulations of the early universe rely on RNGs may find their models overturned if the RNGs they employ are subsequently found to be badly biased.

3 Mixing

Take a fresh deck of playing cards and begin to riffle shuffle them. How many riffle shuffles are required before the deck is thoroughly mixed? Persi Diaconis has shown that seven riffle shuffles are needed. What it means for the deck to be thoroughly mixed is that any configuration of the deck is as likely as any other. The deck starts in a specified configuration.

A single shuffle mixes the deck, but not enough to break all connection with the previous configuration. Only after multiple shuffles does the configuration of the deck lose its connection with the starting configuration. At this point one says that the deck has attained a random state.

Shuffling a deck of cards is an example of a group action. Group actions provide one way of mixing things up, but not the only way. Imagine a gas concentrated in one corner of a box. The particles that make up the gas are in motion. Over time the gas will reach an equilibrium state, filling the entire box uniformly. This is an example of a dynamical system from statistical mechanics. The system starts out in a low entropy state in which all the particles are concentrated in one corner, and eventually reaches a maximal entropy state (= equilibrium state) in which all the particles are evenly distributed throughout the box. The system is said to be random once it reaches the maximal entropy state.

The preceding examples illustrate several features that are common to systems which attain randomness via mixing: (1) such a system starts out from a specified configuration that is highly ordered or constrained (i.e., the opposite of what we would intuitively want to call random); (2) a mixing process (e.g., a group action) acts on the system, over time continually transforming the configuration of the system; (3) eventually an equilibrium state is reached after which further mixing does not affect the equilibrium. When the equilibrium state is reached, the system is said to be random.

It’s worth noting that uniform probabilities frequently characterize the equilibrium states signaling randomness. What it means for a deck of playing to be thoroughly shuffled is that no configuration of the deck is more likely than any other. Shuffling has therefore randomized the deck only if each possible configuration of the deck is equiprobable. Similarly, a gas within a box has reached equilibrium if temperature throughout the box is uniform and the particles are evenly distributed. Uniform probabilities are therefore intimately connected with this understanding of randomness.

4 Complexity

Consider the following two sequences of a hundred coin tosses (heads = 1, tails = 0):

(R) 1100001101011000110111111010001100011011001110111
0001100100001011110111011001111101001010010101110

and

(N) 1111111111111111111111111111111111111111111111111111
1111111111111111111111111111111111111111111111111111

It seems clear that any theoretical account of randomness had better make (R) more random than (N). For instance, since (R) was obtained by actually flipping a coin whereas (N) was artificially contrived, according to the causal account of randomness sketched in §1, (R) would be random, but (N) non-random.

If, however, we prescind from our knowledge of the causal process responsible for these sequences, is it still possible to say that (R) is random and (N) nonrandom? What if we don't know the causal story behind (R) and (N)? Could we still distinguish these sequences in terms of randomness? We could, for instance, try to find a statistical test whose rejection region includes (N) and excludes (R), and thereby justify calling (N) nonrandom and (R) random. But for every such test it is possible to find a corresponding test whose rejection region includes (R) and excludes (N). Nor do probabilities help distinguish the sequences since both (R) and (N), and indeed all such sequences of length a hundred, have the same small probability of occurring by chance, viz., 2^{-100} or approximately 1 in 10^{30}.

Starting in the 1960's a group of researchers that included Gregory Chaitin and Andrei Kolmogorov proposed a way around these difficulties. Instead of characterizing randomness probabilistically, they took the very different tack of characterizing randomness computationally. What they said was that a string of 0's and 1's becomes increasingly random as the shortest computer program that generates the string becomes increasingly long. In the 1980's cryptographers proposed a variant of this characterization: a string of 0's and 1's becomes increasingly random as the most efficient computer program that generates the string requires

increasingly long computation times. The first approach characterizes randomness in terms of space complexity (i.e., the amount of memory the program occupies); the second in terms of time complexity (i.e., the computation time the program requires). The space complexity approach to randomness is referred to in the literature as 'algorithmic information theory'.

It is now intuitively obvious why (R) is more random than (N). The shortest program that computes (N) has the form 'repeat "1" 100 times'. On the other hand, (R) seems to have no shorter description than the string itself. (N) can be drastically compressed, (R) cannot. Thus from the point of view of algorithmic information theory (R) is more random than (N).

Although complexity approaches to randomness represent a genuine advance in the theoretical study of randomness, there is a limitation to these approaches that is often lost in the initial enthusiasm: All complexity approaches to randomness are relativized to a given computational environment. What this means is that even though a sequence may be random when its generating program is running in PASCAL on a standard mainframe computer, with respect to another computational device it may be non-random, and vice versa. In fact, since mappings between finite sets are always computable (recursion theory on finite sets is trivial), any finite string will be random with respect to certain programming environments, non-random with respect to others.

5 Pattern-Breaking

Having now surveyed four distinct approaches to randomness, one is tempted to ask whether a common thread runs through these approaches? There is a common thread, but one that at first sight will seem counterintuitive. If one looks at a dictionary definition of randomness, one finds that the term characterizes objects or events brought about without method, plan, purpose, forethought, pattern, principle, order, or design. Random objects are supposed to be higgledy-piggledy, evincing no patterns.

But what does it mean for an object to evince no patterns? Consider a spy who eavesdrops on a communication channel in which encrypted messages are being relayed. If the

spy has yet to break the cryptosystem, the encrypted messages traversing the communication channel may, as far as the spy is concerned, fail to display any patterns. Yet as soon as the cryptosystem is broken, all the patterns hidden by the cryptosystem become apparent.

The point to recognize is this: what determines the patterns that must be broken for an object to be random is not some objective feature of the world--randomness is not a natural kind. Rather, what is random depends on the patterns that are specified within a given context and that must then be broken for an object to be random. What is counterintuitive about this approach to randomness is that randomness becomes a thoroughly parasitic notion on patterns with respect to which it is defined. Randomness on this view does not make sense until a given collection of patterns is specified.

How then does this pattern-breaking approach to randomness relate to the four preceding approaches? For the computational complexity approach to randomness, the low complexity programs specify the patterns. For the mixing approach to randomness, far from equilibrium states specify the patterns. For the simulation approach to randomness, statistical tests specify the patterns. The pattern-breaking approach to randomness also makes clear why chance is so often a safe route to randomness: in many applications the patterns specified in advance identify a set of very small probability (e.g., a full complement of the statistical tests used to vet a RNG will typically designate as non-random only a tiny proportion of possible numerical sequences). Since small probability events are rare, chance will typically deliver objects or events that break all the patterns, i.e., objects that are random in the pattern-breaking sense.

References and further reading

Borel, E. (1963) Probability and Certainty, translated by D. Scott, New York: Walker. (Cf. §5. Treats the problem of small probabilities.)

Dembski, W. A. (1990) 'Uniform Probability', Journal of Theoretical Probability 3 (4): 611-626. (Cf. §3. A general account of the uniform probabilities.)

------ (1991) 'Randomness by Design', Nous 25 (1): 75-106. (Cf. § 5. An exposition of the pattern-breaking approach to randomness.)

Diaconis, P. (1988) Group Representations in Probability and Statistics, Hayward, Calif.: Institute of Mathematical Statistics. (Cf. § 3. The road to randomness via group actions.)

Garey, M. R. and Johnson, D. S. (1979) Computers and Intractability: A Guide to the Theory of NP-Completeness, New York: Freeman. (Cf. § 4. An introduction to computational complexity in which computation time serves as the measure of complexity.)

Hacking, I. (1965) Logic of Statistical Inference, Cambridge: Cambridge University Press. (Cf. §§ 1. A good discussion of what statisticians mean by chance and hence randomness.)

Knuth, D. E. (1981) Seminumerical Algorithms, 2nd edition, Volume 2 of The Art of Computer Programming, Reading: Addison-Wesley. (Cf. § 2. Includes Knuth's classic treatment of RNGs.)

Kranakis, E. (1986) Primality and Cryptography, Stuttgart: Wiley-Teubner. (Cf. §§ 2 and 4. Treats RNGs as well as the time complexity approach to randomness.)

Lasota, A. and Mackey, M. C. (1985) Probabilistic Properties of Deterministic Systems, Cambridge: Cambridge University Press. (Cf. § 3. The road to randomness via dynamical systems.)

Yockey, H. (1992) Information Theory and Molecular Biology, Cambridge: Cambridge University Press. (Cf. § 4. Includes a brief treatment of algorithmic information theory along with extensive references to the literature.)