

A certain uncertainty:
Random number generation from
Karl Pearson to the Monte Carlo method

Alex Wellerstein
Department of the History of Science
Harvard University
August 14, 2005 draft — not for circulation

At the heart of the history of statistics—if one were to generalize about such things—is a story of bringing of order to a natural world which appears disorderly to the human eye. Concepts of disorder, chance, and randomness have been, since the very foundations of statistical or probabilistic thinking, the problem to be solved, the phenomena to be tamed, and a reflection of human ignorance or error in the ultimate understanding of the complexities of the world.¹ Put another way, one way to view the discipline of statistics is as a way of bring a form of order into existence, and for much of its history the idea of "disorder" has been defined primarily in the *negative*, that is, as the morass out of which order is teased. But there have also been, at various times, attempts to take randomness and disorder out of the background and turn them into *positive* concepts as well, with their own precise definitions, for the purpose of producing a particular coordinated form of *uncertainty*—an idea which sits as a mirror image to the rest of this previous history. But what does it mean to define concepts of disorder in a *positive* sense, much less attempting to *actively produce* objects of randomness, much less a *calculable* understanding of randomness?

¹ This is, at least, the general theme in most of the general surveys on the subject. See, for example, Ian Hacking, *The taming of chance* (Cambridge, England: Cambridge University Press, 1990); Gerd Gigerenzer, et al., eds., *The empire of chance: how probability changed everyday life* (Cambridge, England: Cambridge University Press, 1989); Theodore Porter, *The rise of statistical thinking, 1820-1900* (Princeton, New Jersey, Princeton University Press, 1986); Theodore Porter, *Trust in numbers: the pursuit of objectivity in science and public life* (Princeton, New Jersey: Princeton University Press, 1995).

The subject of this paper is just such an activity, namely the generation of lists of “random numbers,” which first appeared as statistical tools at the end of the 1920s but soon proliferated into a variety of other disciplines as one of many newly-developed statistical tools and objects. Over the course of the century, these lists and the methods for generating them were put to a multitude of uses—primarily in the aid of sampling, simulation, and statistical experimentation—and were created and used by specialists in a wide variety of disciplines, including statisticians, mathematicians, physicists, biologists, bureaucrats, and computer scientists (reflective perhaps of the nomadic quality of “statistics” as a field for much of this time). The lists themselves varied in length and, as tools built on a positive notion of “randomness,” served as focal points for latent questions about the nature of “chance” itself, especially when the question of their *quality* was raised as a salient concern. Some thirty years later, the intellectual discussions which had originally begun with the intention of harnessing some sort of native disorder present in the world had reached the point where scientists were more-or-less satisfied with producing “random numbers” from entirely deterministic algorithms, fully aware of the apparent paradox.

The history of “randomness” is usually told as a story of paradoxes and difficulties—as statisticians so desire more randomness, they find their current approaches more and more unsatisfactory, and reach towards further and more refined approaches. Sometimes this is told as a series of successes (the generation of larger and larger sets of successful “random” digits, and the simultaneous achievement of more and more understanding of “randomness”), sometimes as a series of failures (the inability to achieve “truly random” digits). Most draw a strict line between the “true randomness” of

the pre-computer period, and the “pseudo-randomness” born out of the computer work of the Second World War, by which deterministic algorithms simulate randomness.²

My argument in this paper is that there is something else at work here: that the forging of statistical “randomness” out of a common concept entailed a number of conceptual shifts from early on, and that the key conceptual shift which allows for “pseudo-randomness”—an instrumental approach to randomness, whereby “randomness” is defined by how well it passes a “test” for “randomness” rather than any situation related to its initial parameters—was entered into use before the computer was employed. In outlining these conceptual changes, I am hoping to find a new way of framing the issue of “randomness” in the history and philosophy of statistics, and give it a better standing than that of historical oddity.³

PART I: EXPECTATION AND REALITY

Grounds for doubt

Before looking at the early generation of “random numbers,” it is important to look at two separate historical movements which set the stage for both the *necessity* of such an activity as well as the *grounds for doubt* which would eventually plague it.

² One of the more comprehensive (and sensitive) treatments by a statistician is Daniel Teichroew, “A history of distribution sampling prior to the era of the computer and its relevance to simulation,” *Journal of the American Statistical Association* 60:309 (Mar 1965), 27-49. A very quick technical overview of the relevant literature (with an excellent bibliography) is Thomas G. Newman and Patrick L. Odell, *The generation of random variates* (London: Charles Griffin & Co., 1971). Typical examples of “popular” approaches include chapter 8, “Wanted, random numbers” in Deborah J. Bennett, *Randomness* (Cambridge, MA: Harvard University Press, 1998); and chapter 9, “Gambling with numbers,” in Ivars Peterson, *The jungles of randomness: A mathematical safari* (New York: Penguin Books, 1998).

³ A primary area of information almost completely omitted from this study—for purposes of both space and coherency—has been the theories of “randomness” put forward by philosophers such as Venn, von Mises, Hume, Peirce, Mill, and many others. It seems, though, that most of my historical actors did not spend much time reading these discussions either, so I am not sure how relevant they are to the historical questions I am interested in asking. A partial bibliography and one-line summaries of each of their contributions of view can be found in Bennet (ref. 2), chapter 9, though I do not trust it to reflect much in the way of their arguments.

Since the early days of formal probability theory, games of chance provided a source of experimental inspiration as well as being a guiding metaphor for statisticians. Many early approaches to theories of chance were based on experimentation and observation of relatively equiprobable devices seems like an act of pragmatism as much as anything else.⁴ Dice, roulette wheels, and drawing lots from urns could be broken down into their component likelihoods, robbing the gods of capriciousness and whim, and whether the reasons for “chance” were transformed into mere “error,” an ignorance of the full behavior of an essentially deterministic world, or conceived in a more inherently non-deterministic sense, hardly mattered in the choice of models and metaphors by the early theorists.

The question of whether these models actually produced *true* “randomness” or not came into sharp focus in the last years of the nineteenth century and the early years of the twentieth century. The work of the British statistician, socialist, and eugenicist Karl Pearson is especially prominent in this realm, beginning with a popular essay written on the “randomness” of the roulette wheels in Monte Carlo, the famed Monacan casino city. As Pearson told it, the idea came to him as he was attempting to create a set of examples of the applications of statistics for a popular audience based on “real-world” information. Generating his own data proved laborious: “25,000 tosses of a shilling occupied a good portion of my vacation, and, being conducted frequently in the open air, gave me, I have little doubt, a bad reputation in the neighbourhood where I was staying.”⁵ The idea of using data from “the most sacred shrine” of the goddess Chance, appealed to him—the

⁴ Lorraine Daston, *Classical probability in the Enlightenment* (Princeton, NJ: Princeton University Press, 1988).

⁵ Karl Pearson, “The scientific aspect of Monte Carlo roulette,” chapter 2 in *The chances of death and other studies in evolution* (London: Edward Arnold, 1897): 42-62, on 44.

casino was “clearly a scientific laboratory preparing material for the natural philosopher.”⁶ The difficulty of obtaining the data—surely, he reasoned, the British Association or Royal Society would balk at funding “an agent engaged in such a novel form of scientific investigation”—abated when he found that a French journal, *Le Monaco*, recorded the weekly results of the tables, which he duly tallied. But the results did not accord with theory; calculating the odds of the divergence from his theoretical expectations, Pearson found them to be roughly two million to one:

At this result I felt somewhat taken aback. I did not immediately assume that the laws of chance did not apply to Monte Carlo roulette, but I considered myself very unfortunate to have hit upon a month of roulette which was so improbable in its characteristics that it would only occur, on the average, *once* in 167,000 years of continuous roulette-playing. Such were clearly not the most suitable returns for illustrating the laws of chance!⁷

Further analysis was unrelentingly damning to the casino: “Roulette as played at Monte Carlo is not a scientific game of chance. Such results as those published in *Le Monaco* give the deathblow to the mathematician’s theory of the random spinning of a mechanically accurate roulette. The man of science may proudly predict the results of tossing halfpence, but the Monte Carlo roulette confounds his theories and mocks at his laws!”⁸

If the Monte Carlo results were valid, Pearson concluded, they would require a rethinking of all of the theory of probability. It was more likely that the tables were rigged, and Pearson suggested that the French government “be urged by the hierarchy of science to close the gaming-saloons” and donate their holdings to the Académie des Science “for the endowment of a laboratory of orthodox probability.”⁹ Jokes aside, Pearson’s work, published originally in the *Fortnightly Review* in 1894 and three years

⁶ Ibid., 44-45.

⁷ Ibid., 52-53.

⁸ Ibid., 55-56.

⁹ Ibid., 62.

later in his compilation *The chances of death* (named after an essay demonstrating the statistical regularity of death rates), served to call explicitly into question reliance on the classical metaphors of “chance,” and was key in his own development of methods for determining what was and was not an acceptable experimental deviation from theory.

It was work along these lines which would lead Pearson to develop the chi-squared “goodness-of-fit” test, the statistician's workhorse for determining deviation from prediction. Begun in about the same period as his examination of Monte Carlo, Pearson turned his analytical eye to 26,306 throws of dice performed and tabulated by W.F.R. Weldon (and his poor wife). Finding Weldon's results not in good accord with theoretical expectations, Pearson judged *the dice*—not Weldon—to be in *error*. Pearson speculated that the “pips” carved out of the surfaces of the dice to mark their value had made the “five” and “six” sides weigh slightly less than their corresponding “one” and “two” sides, resulting in a slight “bias” towards higher rolls. In Pearson's formulation, the chi-squared “goodness-of-fit” test indicated the *probability* that a given set of data came from a given probability—in this case, whether the events Weldon was testing for had a probability of $1/3$, as one would ideally expect, or 0.3377, the result of dice weighted slightly toward higher values. Pearson's test found the likelihood as being 62,499 to one in the first case, eight to one in the second.¹⁰ The ultimate question as to whether Weldon's dice actually *were* loaded, or whether the results were simply an unlikely (but not impossible) run of

¹⁰ Karl Pearson, “On the criterion that a given system of derivations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling,” *Philosophical Magazine* 5:50 (1900): 157-175. See also, Adrienne W. Kemp and C. David Kemp, “Weldon's dice data revisited,” *The American Statistician* 45:3 (Aug 1991): 216-222; and R.L. Plackett, “Karl Pearson and the Chi-Squared Test,” *International Statistical Review* 51:1 (Apr 1983): 59-72.

independently random events, was rejected in favor of an approach that held *expectation* as being the ultimate determinant of *reality*.

Although statistical “experiments” of the sort which fueled Pearson had been performed since at least the time of Buffon, there seems to have been a definite rise in their use in the early years of the twentieth century.¹¹ One interesting example, continuing the theme of rising doubts about classical models of reality was an “experiment” undertaken in April 1900: William Thompson (Lord Kelvin), in the course of an attempt to experimentally refute the Maxwell-Boltzmann doctrine on the partition of energy (which he believed mathematically unproven and incorrect), attempted to construct a statistical experiment which could be used to simulate particle motion. He staged a “lottery,” as he put it, of “837 small squares of paper with velocities written on them and mixed in a bowl.”¹² He found the experiment “very unsatisfactory,” as the drawing of lots—a mainstay metaphor of probability theory since the urns of Laplace and Bernoulli—could not be made sufficiently variable: “The best mixing we could make in the bowl seemed to be quite insufficient to secure equal chances for all the billets.”¹³ There were also less theoretical annoyances with the method, as he warned the reader in his write-up: “in using one’s fingers to mix dry billets of card, or of paper, in a bowl, very considerable disturbance may be expected from electrification.”¹⁴ He was forced to change his experiment to one based on playing cards instead.

¹¹ The question these sorts primarily mathematical and theoretical activities are comfortably labeled as “experiments” is taken up in detail in chapter 8 of Peter Galison, *Image and logic: A material culture of microphysics* (Chicago: University of Chicago Press, 1997), 689-780.

¹² Lord Kelvin [William Thompson], “Nineteenth century clouds over the dynamical theory of heat and light,” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, Series 6*, 2:1 (1901): 1-40, on 35.

¹³ *Ibid.*, 27.

¹⁴ *Ibid.*, 27.

But if Pearson and Thompson had shown marked distrust at the simplicity of the classical analog as a practical approach to “randomness,” there were also those who were not disturbed by the idea. A study to turn to in this regard is the work of William Sealey Gosset (“Student”) which is usually cited as the first formal instance of what is known as a “random sampling experiment,” whereby a certain amount of data is taken “at random” and used to extrapolate a whole, specifically his investigation into the behavior of small samples (which would eventually lead to his and Fisher’s so-called “t-statistic”).¹⁵ Taking a “population” of data which he knew the distribution of beforehand—in this case, the heights and middle-finger lengths of 3,000 criminals (one wonders how literally his talk of “individuals” and “populations” could be read)¹⁶—Gosset was able to produce a “random” set with which to experiment:

The measurements were written out on 3000 pieces of cardboard, which were then very thoroughly shuffled and drawn at random. As each card was drawn its numbers were written down in a book which thus contains the measurements of 3000 criminals in random order.¹⁷

Once his “book” was created, Gosset then could take samples of data from different points, comparing the way smaller samples interacted with larger samples, with a goal of throwing light onto where “the limit between ‘large’ and ‘small’ samples is to be drawn,”¹⁸ which he published in 1908. This “empirical” approach, according to Gosset, aided him before he had solved his problem “analytically.”

Unlike Thompson, Gosset did not question the “randomness” of his sample. This is not terribly surprising: drawing lots from an urn was, as has been previously mentioned, one of the canonical approaches to visualizing “chance” since the beginning

¹⁵ Egon S. Pearson, “‘Student’ as statistician,” *Biometrika* 30:3/4 (Jan 1939): 210-250, 223.

¹⁶ Gosset took the measurements from an earlier article in *Biometrika*. It is a telling thing that in even the most theoretical works of early British statistics, one only has to scratch the surface a little bit for the criminal bodies to come falling out!

¹⁷ “Student” [William S. Gosset], “The probable error of a mean,” *Biometrika* 6:1 (Mar 1908): 1-25, on 13.

¹⁸ *Ibid.*, 2.

of statistical investigations, and the data conformed with his later theoretical models. In the wake of Gosset's work, a variety of statistical methods which relied upon "experimentation" with "random" data would proliferate.

Tippett's table

Pulling lots from a bowl was good enough for Gosset, and Weldon might have felt throwing a dozen dice twenty thousands times was a pleasant way to spend an evening (one hopes his wife knew what she was in for when they married), but for the most part these approaches are *slow* and not particularly adaptable to *general* problems.¹⁹ But alternatives neither existed nor were searched for—when a statistician desired "random" data of this sort, he devised methods for producing it, usually specific to the experiment at hand. Most of the records of this approach I have seen are for "illustrating" various laws, theories, and methods, such as G. Udny Yule's experiments in 1916 and 1917 in throwing a hundred beans onto a roulette wheel to demonstrate a particular application of the chi-squared method.²⁰ By dropping his beans onto an analog of Galton's quincunx held over his device—a wheel divided into sixteen sectors separated by raised divisions,

¹⁹ Francis Galton had described, in an article in *Nature*, a generalized method of producing generalizable "random" data for statistical "experimentation" (using a specially modified set of dice) as early as 1890, his approach does not seem to have had much influence on future investigations. Francis Galton, "Dice for statistical experiments," *Nature* 42:1070 (1890), 13-14. See also their coverage in Stephen M. Stigler, "Stochastic simulation in the nineteenth century," *Statistical Science* 6:1 (1991): 89-97, on 94-96. In his article talking about the dice, which contains a wonderful photograph of them, Stigler expresses some confusion at why people like Weldon and Pearson never found any use in this method. But looking at Pearson's biography of Galton makes it quite clear that he knew about them, but felt that the paper "contained a good many pitfalls." As will be discussed later, when Pearson did want "random numbers," he turned to other sources. Karl Pearson, *The life and letters of Francis Galton*, Vol. 2 (Cambridge, England: Cambridge University Press, 1924), 405.

²⁰ G. Udny Yule, "On the application of the chi-squared method to association and contingency tables," *Journal of the Royal Statistical Society* 85:1 (Jan 1922): 92-102, description of device on 94-95.

“kindly constructed” by a friend—Yule’s beans “scattered through the layers of wedges as they fell, and the result of which could hardly be other than random.”²¹

In 1925, however, a somewhat different trajectory began. A masters student working under Karl Pearson at his Biometric Laboratory, Leonard Henry Caleb Tippett, was in the process of work dealing with the question of statistically determining whether or not certain “extreme individuals” (again, I wonder just how figurative this was) could be rejected from an overall distribution.²² In the tradition of Gosset, he considered it “worth while to make some sampling experiments in order to illustrate and confirm the results” of his paper, at the suggestion of Pearson.²³ To begin with, he created a set of 1,000 “very small cards,” numbered sequentially, “which were drawn one at a time from a bag, the card being replaced and the contents of the bag were mixed between each draw.” Making 5,000 such draws and recording the results, he then took a number of samples from this set (such were the specifics of his experiment) and took the mean range from these. He was dissatisfied with the results, though, and felt a pang of doubt about how “random” his data was:

The discrepancies between theory and practice are too great to be attributed to ordinary errors of random sampling, so it was concluded that the mixing between each draw had not been sufficient, and there was a tendency for neighboring draws to be alike, leading to a low value of the mean range. The fact that later and more careful experiments removed the discrepancies make this explanation plausible. The mean range thus supplies a fairly sensitive test of the randomness of a series of observations.²⁴

²¹ Ibid., 25.

²² L.H.C. Tippett, “On the extreme individuals and the range of samples taken from a normal population,” *Biometrika* 17:3/4 (Dec 1925), 364-387. Very little biographical information on Tippett exists in the published literature; I have found a single obituary on him, and he seems to have made most of his way in applying statistical approaches in the private sector. His primary contribution to statistics was in developing the Fisher-Tippett distribution (usually known as extreme value distribution), and work relating to statistical quality control. He authored a number of statistics textbooks, including three volumes of *Methods of statistics* in 1931, 1948, and 1952, and co-authored a book on *Statistical methods for textile technologists* in 1960 and 1979. J.E. Ford, “L.H.C. Tippett, 1902-1985,” *Journal of the Royal Statistical Society, Series A (General)* 149:1 (1986): 44.

²³ Tippett (ref. 22), 378.

²⁴ Ibid., 378.

Two things are worth noting here: first, his decision regarding the *cause* of his original “discrepancies” as a function of the *method of sampling* (poor mixing); second, his suggestion that the failure of his numbers to correspond with his theory was a possible *test* of “randomness.” The numbers didn’t deliver what he expected them to, and so the error had to lie either in his theory or the numbers; he chose the numbers, and specifically the method.

Rather than suffer the difficulties of shuffling, Tippet resorted to a different method for his second, larger experiment, at the suggestion of Pearson, who was still suspicious of “mechanical” sampling methods. Instead of filling a bag with each of the many potential values for his experiment, he constructed first a “key” which would allow him to convert any four-digit number (from 0000 to 9999) into the sorts of values usable for his testing (itself a range between 0.05 and 6.45, broken into 69 ranges). Once he had established this system of translation, he created a set of 40,000 individual digits—“taken at random from census reports”—grouped into sets of fours. With these numbers, he was translate them into values he could use for his experiment with his “key,” and thus have a “random sample of the original population.” This time, he was satisfied with the way his “random sample” handled under calculation. As a footnote, he added that, “By means of a suitable key, these numbers can be used to construct samples from any population. It is hoped that they will be published later.”²⁵

Like Galton, Tippet saw his method as being *generally* useful, but unlike Galton, he didn’t want to tell you *how* to make such samples—he’d already done all the hard work, he’d just give you the digits now that he was done with them. With the right translation key, you could use his list of “random numbers” towards *any* problem, and

²⁵ Ibid., 378.

could skip all of those tedious (and potentially unsatisfactory) hours of drawing lots from a bag (or rolling dice, or throwing beans, etc.). And so in 1927, Tippet's list of 40,000 digits—with an additional 1,600 digits, for an even twenty six pages of numbers—appeared as volume fifteen of Karl Pearson's series, "Tracts for Computers," under the title of *Random Sampling Numbers* (that is, numbers for random sampling).²⁶

Pearson wrote a foreword to the 1927 tract, starting off with the observation that "in the case of many problems we may both check theory and prevent the pure algebraist, as he is apt to do, outrunning statistical experience by an appeal to the test of artificial sampling."²⁷ But even with small studies, "the work of testing is very laborious," and "it is complicated by the fact that drawing balls or tickets from a bag or urn, however pleasing in theory to the mathematician, transcends the powers of the practical statistician," through the sheer amount of work required, and the large number of balls and tickets needed for certain problems. More to the point, Pearson made explicit again his distrust in such methods *in principle*:

Practical experiment has demonstrated that it is impossible to mix the balls or shuffle the tickets between each draw adequately. Even if marbles be replaced by the more manageable beads of commerce their very differences in size and weight are found to tell on the sampling. The dice of commerce are always loaded, however imperceptibly.²⁸

Along with an amusing jab at the free market (what else from the British socialist?), Pearson translated his previous assault on classical metaphors of chance into *practice*: a fear of "imperceptible" biases, laden in the structures of the physical devices themselves.

But of course, the validity of Tippet's alternative method (which he did not go into any detail on) still required some establishment. After a series of three "illustrations"

²⁶ L.H.C. Tippet, *Random sampling numbers* (London: Cambridge University Press, 1927). The digits themselves were written out *by hand* (impeccably, I might add) by Pearson's "draftswoman," Ida McLaren, and then lithographed.

²⁷ Karl Pearson, "Foreword," in Tippet (ref. 26), iii.

²⁸ *Ibid.*, iii.

of the way one could use Tippett's tables for statistical sampling experiments (one of which involves a table correlating "Familial Income" with the "Intelligence of Child," ranging from "Very Able and Capable" to "Very Dull and Mentally Defective"), Pearson concluded his foreword by affirming his utmost faith in Tippett's numbers:

Of course the value of the present series of numbers depends on their being truly a random selection. Any such series may be a random sample, and yet a very improbable sample. We have no reason whatever to believe that Tippett's numbers are such, they have conformed to the mathematical expectation in a variety of cases, and we would suggest to the user who finds a discordance between the results provided by the table and the theory of sampling he has adopted, first to investigate whether his theory is really sound, and if he be certain that it is, only then to question the randomness of the numbers.²⁹

Pearson's conviction that Tippett's numbers do not constitute an "improbable sample," that is, having a low probability of being "truly random," is based on the fact that under a number of experiments, the numbers have provided results *as expected by the theory*.

That is, the way one knows whether random numbers are "random," in Pearson's formulation, is whether or not they can *perform* as random numbers ought, and Tippett's had done so well enough to inspire enough confidence that Pearson would suggest that contradictory results reflect upon the theory, not the numbers themselves. The numbers were made reliable by their accordance with certain "sound" theory, to the point that they could by themselves call into doubt theories in which they disagreed. This is not yet an explicit "test" for "randomness," but rather more akin to the reliability of certain theories over others: the numbers were vindicated by their accordance with "sound" theory, and thus became a reliable tool against which to test new (and potentially un-sound) theory.

Pearson ends the piece with a final note further related to an establishment of what "randomness" would mean in a positive sense:

But the reader must remember that in taking hundreds of samples he must expect to find improbable samples occurring or even runs of such samples; they will be in their place in the

²⁹ Ibid., viii.

general distribution of samples, and he must not conclude, from their isolated occurrence, that Tippet's numbers are not random.³⁰

This simple warning conceals a much larger philosophical problem in the question of “random numbers”: the question of whether “runs” of them will behave in apparently “non-random” ways. As an example of this, on Tippet's first page of numbers, there is a run of four 0s in row, as well as a run of four 7s in a row. Of course, the probability of any individual sequence occurring is, ideally, the same as any other individual sequence, however the odds of probability greatly favor sequences of heterogeneous composition over ones of homogenous composition. Illustrated simply, if we enumerated a list of 10,000 unique four-digit numbers (from 0000 to 9999), it is easy to see that only ten of them will be runs of four of the same digits (0000, 1111, 2222, and so on). But Pearson urges us not to worry about such things: all of these apparent “improbable samples” are allowable in the “general distribution of samples,” so we ought not doubt the numbers—it would all work out in the long run. With a little faith, the “law of large numbers” would run its course if allowed, because the *method* was what provided the guarantee of “randomness.”

³⁰ Ibid., viii.

PART II: INSTRUMENTAL RANDOMNESS

“Randomness” from the non-random

In the decades following Tippet’s initial table, techniques of statistical sampling became a standard approach in the ever-widening array of fields touched by the work of the British biometricians. Tippet’s numbers no doubt aided in this by removing the need to reinvent the (roulette) wheel every time a random sample was required, and was one of the more popular tracts produced by Pearson’s laboratory, going through at least four separate printings. So standard had “random numbers” become to statistical research that when Ronald A. Fisher and Francis Yates compiled their *Statistical Tables for Biological, Agricultural, and Medical Research* in 1938—a standard reference work for many years, published with the stated goal of popularizing statistical tables to audiences other than statisticians—they included, alongside tables of normal distribution values, natural sines, and common constants, a table of 15,000 of their own “random numbers.”³¹

Fisher and Yates used somewhat different method of construction of their table than Tippet, though. In a manner similar to Tippet’s, they too took their numbers from an existing “supply,” but rather than basing it on a selection of “real world” variability, they chose instead pages from a book of logarithm tables. Setting up for themselves a large table to be filled in with the digits, they chose sets of 50 digits at a time (“rearranged in a random order”) from the logarithm tables, reading downwards, and copied them into their own table, selecting the column to begin entry “at random.” All of their “random” decisions were done by the use of two packs of playing cards.³² A need

³¹ R.A. Fisher and F. Yates, *Statistical tables for biological, agricultural, and medical research* (London: Oliver and Boyd, 1938), 82-87.

³² *Ibid.*, 18.

was clearly felt to justify their choice in taking as their input numbers from a decidedly non-random source:

It will be seen that this method of construction can hardly fail to give a table of numbers which is the equivalent of a random selection made by an ideal mechanical contrivance. Any slight systematic element that may occur in parts of the logarithmic table will be effectively obliterated by the method of distribution of the digits.³³

However, in the very next sentence, they revealed that a preliminary analysis of their digits had revealed a problematic trend:

The use of even a perfect randomising process does not, of course, preclude the possibility of obtain a somewhat exceptional sample. This appears to have occurred in this case, for when, as a matter of interest, the digits making up the six 50 X 50 pages were counted... a somewhat large preponderance of 6's is apparent.... This preponderance of 6's appears to be due to chance selection of columns containing a large number of 6's, rather than to any systematic feature of the figures used.³⁴

Fisher and Yates found by running a chi-squared calculation on the distribution of their digits, assuming a null result with an equal distribution of 1's, 2's, 3's, etc., that they had failed in their first attempt, with over a hundred 6's more than the expected amount. Still, they stood by their *method*, though they were clearly aware of the possible objections which might have been made to it. But what should be done? If they really believed their method to be “random,” wouldn't this deviation actually be representative of the sorts of “flux” possible in the world of chance? That is, the chi-squared test only measured the likelihood that their sample was created using a “random” method which would assure an equal distribution—it was a test of what was more likely, not what actually *had* happened. Convinced that their methodology was sound, they wrote it off to chance, not error—unlike Tippett with his first attempt at drawing numbers from a bag. In the end, they opted to change the table:

It is somewhat debatable whether a set of figures exceptional in one of its main features should be retained unaltered to form a standard table of random numbers. On the whole, however, it seemed

³³ Ibid., 18.

³⁴ Ibid., 18.

better to reduce the number of 6's so as to give a more normal sample. This was done by picking out 50 of the 6's strictly at random and replacing each of them by one of the other 9 digits selected at random.³⁵

They do not specify it here, but I think it is reasonable to assume that they again used their playing cards when making decisions “at random” (they seem too concerned with methodology to let it slide at this point). The final distribution, with their changed 6's, corresponded much better with the null hypothesis, and a table of the “distribution of occurrences of each digit,” showed that the set was “clearly reasonably regular.”³⁶ Along with this, they also examined the number of two-digit pairs of the same values (00, 11, 22, etc.), and compared these counts with the theoretical expectations. Whether they took this extra step because their suspicions were raised, or because they felt a need to again display their “randomness” to the world, is up for speculation. Before giving a number of examples for using the table, they laid out one additional methodological statement:

There are, of course, many other features of the table that might be tested. In general the values of χ^2 obtained from a large number of such tests may be expected to be evenly distributed over the probability scale, provided the tests are chosen without prior inspection of the table, so that for instance one in 20 tests will indicate a significant departure (at the 5 per cent. level) from randomness. If features which on inspection of the table appear exceptional are tested the proportion of significant deviations from randomness will clearly considerably exceed 1 in 20.³⁷

What we have here is the beginning of an impulse to *test* for “deviations from randomness,” which, while not quite the same thing as testing *for* “randomness” in a *positive* sense, is a considerable step closer to this. If we look at the steps to this point, the departure will start to become clear: Lord Kelvin decided that his lot-drawing wasn't random enough, but seemed to be relying more on gut instinct (or the fact that the results just weren't vindicating his theory) than anything else; Pearson felt Tippett's numbers were vindicated by their correspondence with “sound” theory; and Fisher and Yates

³⁵ Ibid., 19.

³⁶ Ibid., 19.

³⁷ Ibid., 20.

applied two systematic analyses to their numbers to measure their “deviation” from an expected “randomness.” Even a “perfect randomising process,” which they believed their procedure to be, could produce certain numbers of *errors*, lacks of “randomness,” and these could be *corrected*—the table itself became the goal of the process, and “randomness” became a property *disconnected* from the method of producing it.

“Testing” “randomness”

In 1938, an article appeared in the *Journal of the Royal Statistical Society* which would significantly re-frame the question of “randomness” in all future investigations of “random” numbers. “Randomness and Random Sampling Numbers,” grew out of an address Maurice G. Kendall had given to a study group of the Society a year before, and, with the co-authorship of Bernard Babington Smith, had developed into a twenty page treatise on the subject.³⁸ “The ideas of randomness and probability are inseparable,” Kendall and Smith put forward, but the interpretations of what they meant in a logical sense varied widely between different schools of thought.³⁹ But rather than trying to consider the two as sides of the same coin, or even different names for the same thing, they single out “randomness” as a distinct object of study: “Whatever may be the proper relationship between the two concepts in logic, in a discussion of the technique of random sampling randomness must be considered in its own right.”⁴⁰ “Randomness” could be no longer “neglected” as simply subservient to notions of probability; there was a need for “scrutiny of the concept of randomness itself, particularly in view of the

³⁸ M.G. Kendall and B. Babington Smith, “Randomness and random sampling numbers,” *Journal of the Royal Statistical Society* 101:1 (1938): 147-166. I suspect Kendall's voice to be speaking primarily in the text, but I do not know this for a fact.

³⁹ Ibid., 147-148.

⁴⁰ Ibid., 148.

practical importance of being able to judge whether any given method of sampling is random or not.”⁴¹

The orientation of Kendall and Smith to the question was one of pure pragmatism: a reevaluation of what the *goals* of “randomness” were for the purposes of statistical sampling. This was explicitly different from the approach of Tippett, Fisher, and Yates—it was the beginning of a conceptualization of “randomness” along lines of purely *operational* characteristics—“randomness” was desirable so far as it was useful for “random sampling.” Not all “randomness” fit this criterion, though:

Any set of digits whatsoever is random in the sense that it might arise in random sampling from a infinite universe of digits; but such a set is not necessarily one which can be used for random sampling, as, for instance, if it consists of a block of zeros. A set of Random Sampling Numbers (by which is meant a set of numbers which can be used for random sampling, not necessarily a set obtained by random methods) must therefore conform to certain requirements other than that of having been chosen at random.⁴²

After all of that discussion of “randomness” and “random methods,” they were now inclined to conclude that numbers for random sampling no longer needed to be “random” in any true sense. Questions of true philosophical “randomness” were “in fact, of an abstract metaphysical character verging at times on the theological,” and they had a practical purpose to fulfill.⁴³

Such is the basis of what I see as the second period of thought on “randomness,” one which was founded on an *instrumental* view of what “random” meant, and if this was not clear enough, Kendall and Smith go so far as to separately capitalize Random Sampling Numbers in order to emphasize their not necessarily literal meaning (much as I have endeavored to put almost every possible problematic term in quotation marks, perhaps to a ridiculous degree). Fisher and Yates were leaning in this direction when they

⁴¹ Ibid., 148.

⁴² Ibid., 153. All questionable grammar is in the original text.

⁴³ Ibid., 150.

removed their fifty errant 6's in preference of a more "normal" table, but Kendall and Smith made it explicitly the basis of their own efforts. But a philosophical concept cannot be turned into an operational principle by willing it alone; what Kendall and Smith endeavored to do was to consider what *properties* of "randomness" were useful for random sampling, and then find a way to see if a given set of "random numbers" actually fulfilled them.

What would these be? If numbers were selected "at random" from a set of equally probable alternatives, Kendall and Smith reasoned, in the long run each possible number should be represented roughly an equal number of times, according to the standard "law of large numbers." A list of Random Sampling Numbers should conform to this same ideal, they reasoned, though it was clear that not every set of numbers, even those produced with a "random method" would necessarily conform to this at all times. Whether or not the sampling method would prove to be equiprobable in the long run was less important than the "local randomness" of any given set of Random Sampling Numbers, that is, conforming to *expectations* of "randomness" within certain minimum ranges.⁴⁴

With these conditions in mind, Kendall and Smith proposed four tests towards establishing "the existence of local randomness" in a given set of digits. Like Fisher and Yates, they would use the chi-squared method to determine deviations from expectations, but, having focused now on the creation of specific properties desirable for their Random Sampling Numbers, they had a new set of expectations. The "first and most obvious" test was the same as Fisher and Yates's, "that all the digits shall occur an approximately equal number of times," which they named "the *frequency test*." But this was extended further:

⁴⁴ Ibid., 153.

true “normal randomness” would mandate as well that “no digit shall tend to be followed by any other digit,” and so they would check the frequency of each possible *pair* of digits as well; this they named “the *serial test*.” Why stop at checking pairs, though? Their third test would check blocks of five, but was somewhat different: rather than testing for any particular set of digits, they would test for the odds of any given five-digit set being the same digit, or being four of the same and one of another, and so forth. This test they named “the *poker test*, from an analogy with the card game.” The last test was of a somewhat different nature, testing only the arrangement of the set, rather than its contents, by looking for “gaps occurring between the same digits in the same series,” simplified by looking only at the gaps between zeros. This “*gap test*” would rely on the *expectation* that a zero would only immediately follow another zero one tenth of the time, and only nine out of a hundred times would there be two zeros separated by a different number, and so forth.⁴⁵ One could imagine finding ways of constructing series of numbers which could “evade” any of these tests taken individually, or even in pairs, they noted, but all four tests “taken together are very powerful.” To construct a series which could evade all four “would, it appears, have to a very peculiar bias indeed, such as would hardly ever rise in practice.”⁴⁶

But how “local” is “local randomness”? If sets of Random Sampling Digits are “locally random” on the whole, what if certain subsections of them are not “locally random”—that is, do not conform to theoretical expectations, as measured by the four tests? Such a thing was thoroughly imaginable to Kendall and Smith, depending on which subsection one used and how large the original sample was (once again, the concern

⁴⁵ Ibid., 154.

⁴⁶ Ibid., 155.

about the behavior of different sample sizes returns). This possibility complicated things considerably:

It follows that a Random Sampling set cannot be used indiscriminately to draw any given number of samples of any given size from any given universe until it has been tested for local randomness not only as a whole, but in the parts which it is proposed to use separately. It follows further from the previous paragraph that a perfect Random Sampling set is impossible—there can never be compiled a set of Random Sampling Numbers which are adequate for all requirements. A set may be useful over a certain range—for instance, a set of a million might be locally random for sampling which required a batch of ten digits or more; but as the size of the set increases, there are bound to appear bad patches which are in themselves, not locally random.⁴⁷

Instrumental randomness, as I am calling the product of this second shift in attitudes, is thus a dramatic re-conception of “randomness” from the negative definition traditionally used by the statistical discipline. Rather than simply being the backdrop for extracting order, or a set of theoretical expectations, “randomness” became a positive property, enforced and established by a barrage of “testing.”

In the Foucauldian framework, tests and examinations serve as tools for enforcing normality, rewarding various forms of conformity, and executing a power over the structure and worth of the subject.⁴⁸ Foucault spoke in terms of testing *people* primarily, but the model can be useful for thinking about *numbers* as well, especially so in the field of statistics, which has long held that there were little differences between the two subjects—by the nineteenth century, the numbers usually *were* people, and people were expected to behave according to the numbers, and the lack of boundaries between the two in the minds of the practitioners is seen nowhere so clearly as in the work of the biometricians. So when our statisticians talk about their sequences as if they had their own moral system—always possibly infected with “bias” and “bad patches”—it should not surprise us much.

⁴⁷ Ibid., 155-156.

⁴⁸ See especially the discussion of “the examination” in Michel Foucault, *Discipline and punish: The birth of the prison* (New York: Viking Books, 1979), chapter III/2.

But I want to push the Foucauldian model still further: for Foucault, the examination was employed for the rendering of the “individual” and of “the subject,” the creation of *objects* out of *people*, the “subjectification of those who are perceived as objects and the objectification of those who are subjected.” In the numerical realm, processes of examination were crucial for a positive construction of “randomness”—one could not have the property without a way to test for it. By constructing Random Sampling Numbers as sequences expected to pass certain qualifications, Kendall and Smith created a new phenomena of “randomness,” and transformed the informational content of the numbers themselves into a tool, one which would transcend the mere paper they were printed on. Pearson and Tippett—and even Fisher and Yates—had wanted to distribute their numbers only because they thought it would save time for others; their reproduction was simply a process of distribution. The Kendall and Smith understanding of the numbers was that they were only valuable if they were up to par; they became desirable for their *quality*, a property which could only be determined by *testing*.⁴⁹

Mechanical chance

Kendall and Smith were not simply interested in “randomness” for purely theoretical reasons (however pragmatic their theory was): they were interested in producing Random Sampling Numbers themselves, ones which were rich in quality as well as quantity. But how to produce numbers which would pass the tests? With reference to Pearson’s work in showing the biases in dice and roulette tables, they noted that mechanical methods “had long been held” to “not give satisfactory results.” The only other table of “random

⁴⁹ Ian Hacking’s small section on testing and the emergence of phenomena (in his case, “child abuse”) has been provoking on this subject as well; see Ian Hacking, “Kind-making: The case of child abuse,” in *The social construction of what?* (Cambridge, MA: Harvard University Press, 1999): 125-162.

numbers” they knew of at the time was Tippett’s—Fisher and Yates published theirs in the same year this one was published—which had used digits, “taken at random” from census reports. Kendall and Smith saw this as “an abandonment of the mechanical method in favour of one which may be reasonably be supposed to be free from bias,” but in the end, its utility came down to one thing: “The reliability of such numbers must, however, depend on the results which they give.”⁵⁰ Simply selecting one’s digits from “apparently random distributions” (such as census reports) was not enough: Kendall and Smith had attempted to construct a set of sampling numbers from numbers taken from the London Telephone Directory (pages were selected “haphazardly,” and only the two right-most digits of any given number were taken), and found the resulting series to be “significantly biased.”⁵¹

Their own (non-extensive) tests on Tippett’s numbers seemed to confirm them as “locally random,” though a paper by Yule immediately following theirs in the journal concluded, with his own, different set of tests, that there was some “patchiness” (a notion he himself puts in quotes, but does not at any point elaborate on).⁵² Still, for their own tables, they wanted to return to a mechanical approach, fearful of hidden biases which might lie in wait in taking numbers “at random” from lists in the fashion of Tippett, and Fisher and Yates, though they were unaware of their work at this point. In a paper published a year later, they gave the Fisher-Yates table the stamp of approval from their

⁵⁰ Kendall and Smith (ref. 38), 156.

⁵¹ Ibid., 156-157.

⁵² Yule’s tests were also frequency tests, looked at someone different properties, such as the sums of five digits in a row. He concluded that “no one of the preceding results leads to a *highly* improbable divergence from the expectation on random sampling, some of the values of *P* are not as high as could be wished, and [a table looking at the standard deviations of sets of digits appears to confirm the impression of ‘patchiness,’ especially in the earlier part of [Tippett’s] Tables, which alone I had been using.” G. Udny Yule, “A test of Tippett’s Random Sampling Numbers,” *Journal of the Royal Statistical Society* 101:1 (1938): 167-172, on 172.

tests, though noted that replacement of 6's was a procedure which had caused them "some misgiving."⁵³

With this framework, they designed a "randomising machine," an apparatus which would produce "locally random sequences of random digits." In the spirit of the roulette wheel, they first drew up a set of ten discs, ten inches in diameter, with ten equal sections numbered with the digits 0 to 9, and selected one disc "haphazardly" from the set to mount on the axle of an electric motor (the multiple discs were to avoid any systematic biases due to defects in the discs or the drawing of their sections).⁵⁴ While the motor spun the disc at a constant speed, the operator moved a stylus around a "maze" of circuits, which would create a circuit, charge a condenser, and set off a flash from a neon lamp, making the disc appear momentarily stationary and allowing the operator to read off a number indicated by a pointer. Because the condenser took a variable amount of time to charge, the movement along the stylus would not correspond perfectly with the light flashes, though one end of the "maze" would result in a shorter duration between flashes than the other. Kendall and Smith recognized that they might have been able to make the system work without the operator's use of the stylus, but "we felt it best to take extra trouble to be on the safe side. It would have been devastating to find, after the numbers had all been run off, that the machine had developed a rhythm which deprived them on local randomness."⁵⁵

⁵³ M.G. Kendall and B. Babington Smith, "Second paper on Random Sampling Numbers," *Supplement to the Journal of the Royal Statistical Society* 6:1 (1939): 51-61, on 60.

⁵⁴ Kendall and Smith's 1938 paper describes the "randomising machine" somewhat differently than their later, more technically precise 1939 paper. It is the latter description I am primarily using. Kendall and Smith (ref. 38), 157; Kendall and Smith (ref. 53), 51-53.

⁵⁵ Kendall and Smith (ref. 53), 52.

The machine generated on 1,500 digits per hour on average, and, without “forcing the pace” (fearing operator error and fatigue), Kendall and Smith managed to produce 100,000 such digits. Each thousand was first subjected to their first three tests (frequency, serial, and poker)—five sets were “rejected” by either one or two of the tests. This did not daunt them—they figured that some of the thousands ought to be rejected by the tests, on average, and none of the “rejections” were large enough “to cause any suspicion of absence of local randomness.”⁵⁶ The digits were then looked at in sets of five thousand with better results—all passed all four tests. The digits were then divided into groups of 25,000 and barraged by all four tests—again all passed—and finally the tests were run against the entire set, which passed. From this Kendall and Smith concluded that their 100,000 were locally random, though if only a single thousand from the set was to be used, the specified five which failed their tests should be avoided.⁵⁷

It was Smith who performed the actual work of operating the machine and running the 100,000 numbers used in the final table (about 67 hours worth of work, according to their speed estimates). Another (unnamed) assistant had also produced a run of 10,000 digits which were supposed to be included in the final set, but, to the surprise of Kendall and Smith, were “rejected by the tests in a decisive way.” Six of the ten “disobeyed the tests” (one by three of the tests) and two of the others “were near the border line.”⁵⁸ They found in general a “strong bias” against the numbers 1, 3, and 9, and a bias in favor of even numbers in general, especially 0 and 6, and the poker test revealed too many instances of numbers being clustered in pairs of twos or threes (a tendency

⁵⁶ In three of the cases it was the serial test which was failed, in another it was the poker test, and in the other it was both the serial and the frequency tests. *Ibid.*, 54.

⁵⁷ *Ibid.*, 54, 57.

⁵⁸ *Ibid.*, 58.

confirmed by the gap test). The explanation, they decided, lay in psychology of the poor assistant:

It thus appears that the observer suppressed certain digits, mainly the odd ones, and got too much clustering of digits of the same time. We questioned him about the methods he had used in obtaining the numbers. He used the machine in the way we have described, and the bias cannot, we think, be attributed to the machine. The explanation of the bias then seems to be that he had a strong number-preference—*i.e.* that either he actually mis-saw the numbers—or that his brain controlled his ocular impression and censored them. That individuals preferences for certain digits is well known... But the effect we are now discussing is more peculiar still. There is here no question of estimation. The observer has merely to record on paper something that he thinks he has seen only a few seconds before... the results are sufficient to show that the personal equation is a very dangerous thing in random sampling, and should be eliminated wherever possible.⁵⁹

The objectivity of the machine thus insisted upon, the tests resurrected the ugly “personal equation” which had plagued nineteenth-century astronomy, and the misbehaving 10,000 digits were discarded.⁶⁰

That same year, the pair published their 100,000 digits again as part of the “Tracts for Computers” series, now edited by Karl Pearson’s son, Egon S. Pearson.⁶¹ The introduction, written by Kendall, began with a warning about creeping order:

It was recognized by some workers early in the development of statistics that many of the methods which might reasonably be expected to yield random samples are, in fact, biased. The evidence which has since accumulated supports that view. It seems that wherever any human element of choice is allowed free play, as, for instance, when an observer selects “at random” by eye a number of plants in a field, or draws cards “haphazardly” from a pack, bias inevitably creeps in. There may be individuals whose psychological processes are so finely balanced that they can deliberately choose random samples; but they are the exception, not the rule, and the ordinary statistician dare not use his own nervous mechanism as a random selector.⁶²

Their machine had been designed, they explained, “to eliminate sources of bias which could reasonably occur,” but more importantly, its results could pass their battery of tests, which were the result of a “more firmly founded” theory of random sampling that the

⁵⁹ Ibid., 59.

⁶⁰ On the “personal equation,” see, e.g., Simon Schaffer, “Astronomers mark time: discipline and the personal equation,” *Science in Context* 2 (1988): 101-131.

⁶¹ M.G. Kendall and B. Babington Smith, *Tables of Random Sampling Numbers* (London: Cambridge University Press, 1939).

⁶² Ibid., vii. I suspect that Kendall is the voice being heard most of the time in their co-published work, only because his individual work uses the same tone, but this has been difficult to confirm (Smith published very little individually).

"intuitive" one used in Tippet's time. Tippet's tables had been "a great advance," but more digits were needed: "The present tables, consisting of 100,000 digits, are intended to fill that need. They will, I hope, be sufficient for any sampling inquiry likely to be encountered in practice at the present time."⁶³ It might not have been such a bad prediction, if they had not been publishing it on the eve of the Second World War.

PART III: ABANDONING "REALITY"

Gambling with neutrons

It was out of the massive Allied effort to produce atomic weaponry during the Second World War that the next shift in conceptions of "randomness" emerged. In order to model the complicated physical processes involved in the implosion of the plutonium bomb, large amounts of funding and work were poured into the use of punched card computers for solving large equations for which there were neither the time nor manpower to solve in the more traditional fashion.⁶⁴ The first punched card machines used during the war were essentially used to scale up traditional approaches to calculating the phenomena in question, leaning closer and closer to approaches which would attempt to "model" the phenomena in question, as opposed to producing "simple" calculations. The initial plutonium bombs had proven difficult and time-consuming projects in and of themselves; applying this work to the far more complicated mixture of phenomena included in the plans for the then still speculative hydrogen bomb would push these methods to their extremes. Calculating even relatively small problems related to thermonuclear physics

⁶³ Ibid., vii.

⁶⁴ These historical developments have been chronicled in a number of sources; I am in particularly fond of the treatment in Paul N. Edwards, *The closed world: Computers and the politics of discourse in Cold War America* (Cambridge, MA: MIT Press, 1994).

took days on end for the first electronic computer, the ENIAC (Electronic Numerical Integrator and Calculator).

As the standard story goes, one day at Los Alamos in 1946, Polish mathematician Stanislaw Ulam was home sick and playing a game of Solitaire with a deck of cards and began to wonder about the odds of winning. Solitaire is, in its ideal state, a game which requires no skill, only the ability to see how best to play the cards one was dealt, though not all combinations of starting card positions allow for success, a consoling fact to those who lose at this essentially deterministic game. Ulam realized that computing the combinatorial possibilities involved in the solution was impractically large and laborious, but by playing a reasonably large number of games and tallying the outcomes, he could come to an approximation of the likelihood of winning. He then realized that this same statistical approach could be applied to the diffusion of neutrons inside the fissioning weapon, solving the statistical outcome of only a selection of neutrons and extrapolating from this the behavior of the entire set. This he presented to computer pioneer John von Neumann, who, after some initial skepticism, took up the idea enthusiastically, and worked to elaborate the basic approach into a systematic method which Nicholas Metropolis suggested they call “Monte Carlo,” in part a reference to the utilization of “chance” and a reference to a gambling uncle of Ulam’s.⁶⁵ Ironically, the casinos which exemplified “chance” for the physicists were the same ones which Karl Pearson had far earlier declared as rigged.

⁶⁵ Nicholas Metropolis, “The beginning of the Monte Carlo method,” *Los Alamos Science* (Special Issue 1987): 125-130. A thoughtful and analytical source on the method’s origins is Galison, (ref. 11), 689-780. See also William Aspray, *John von Neumann and the origins of modern computing* (Cambridge, MA: MIT Press, 1990), 110-115.

The Monte Carlo method has never been given a very precise or distinct definition; it usually refers to a host of techniques for mathematical estimation and simulation of very complex problems, united primarily in the fact that they use “random numbers” to drive their approximations.⁶⁶ The primary tool for these calculations was the computer, and this, along with the common identity of “Monte Carlo” (the writings of many statisticians of the period indicate that its association with postwar bomb physics contributed at least partially to the term’s allure) seem to define these techniques more than any strict philosophical or mathematical parameters.⁶⁷ This need for computerized “random numbers” provides the most immediate link to this present study.

Deterministic “randomness”

Where to get these numbers for these insatiable machines? The previously published “tables” of “random numbers”—by Tippett, Fisher and Yates, Kendall and Smith, and a few other miscellaneous thousands published by sources not mentioned—numbered less than 150,000 digits total and were available only in print sources. Tabulating them onto punch cards could be done—indeed, a Mayo Clinic statistician did just this in a 1943 reevaluation of Tippett’s numbers (which passed another battery of “tests”)⁶⁸—but this was time consuming and did not seem to be a good long-term solution. As Von Neumann put it in at a conference in 1949, “In manual computing methods today random numbers

⁶⁶ Ulam and Metropolis defined it once as: “a statistical approach to the study of differential equations, or more generally, of integro-differential equations that occur in various branches of the natural sciences,” which is probably as good as any other. Nicholas Metropolis and S. Ulam “The Monte Carlo method,” *Journal of the American Statistical Association* 44:247 (Sep 1949): 335-341, on 335.

⁶⁷ A detailed overview of many different types of “Monte Carlo” mathematical techniques can be found in Newman and Odell (ref. 2). The difficulty in definitions is said very explicitly through the early literature on the subject.

⁶⁸ Robert Gage, “Contents of Tippett’s ‘Random Sampling Numbers’,” *Journal of the American Statistical Association* 38:222 (Jun 1943): 223-227.

are probably being satisfactorily obtained from tables. When random numbers are to be used in fast machines, numbers will usually be needed faster.”⁶⁹

Von Neumann judged that there were two ways of generating sequences of “random numbers.” The first he termed *physical processes*—devices which took advantage of some sort of latent “randomness” in the world at large, such as the time between atomic disintegrations in radioactive substances (though samples decay on the whole at a predictable rate—hence of the applicability of half-lives—quantum physics dictates that there is no way of knowing at what time any particular atom will fall apart). These could be hypothetically fed into a high-speed computer, but if they were not being recorded they could never be tested for their “randomness” after the fact; if they were recorded, it would overtax the computer’s memory.⁷⁰

A workable solution would be to generate the numbers using what Von Neumann termed *arithmetical processes*—deterministic algorithms which would generate sequences which could pass the tests for “randomness,” a potential antithesis to any notion of true “randomness.” As Von Neumann famously put it: “Any one who considers arithmetical methods of producing random digits is, of course, in a state of sin.”⁷¹ This quote has often been taken to imply that Von Neumann was *against* “arithmetical methods,” but I do not believe this is the case. The quote in context clearly refers to the error of believing that “arithmetical methods” produce *random* numbers; Von Neumann is warning for the scientist to not misunderstand what they were dealing with and its

⁶⁹ John von Neumann, “Various techniques used in connection with random digits,” in A.S. Householder, G.E. Forsythe, and H.H. Germond, eds., *Monte Carlo Method*, National Bureau of Standards Applied Mathematics Series, 12 (Washington, D.C.: U.S. Government Printing Office, 1951): 36-38.

⁷⁰ *Ibid.*, 36.

⁷¹ *Ibid.*, 36.

potential limitations. There were, Von Neumann continued, no true “random numbers,” only methods to produce them, “and a strict arithmetic procedure is not such a method.”

We are here dealing with mere “cooking recipes” for making digits; probably they can not be justified, but should merely be judged by their results. Some statistical study of the digits generated by a given recipe should be made, but exhaustive tests are impractical. If the digits work well on one problem, they seem usually to be successful with others of the same type.⁷²

Computer scientists and physicists have reproduced the above Von Neumann quotes as a virtual paradigm shift in the realm of “random number” generation, but our path so far has shown that it is not without its strong affinities. The idea that “random numbers” should be judged by how well they work in problems is one which goes back to Tippett’s tables, and the notion that sequences should be judged by their “results” even more so than their generative methods is one which is traceable back to the work of the late 1930s.

Von Neumann’s preferred “cooking method” for random digits was exceedingly simple. Known as the “middle-square” method, a given ten-digit number was taken as a “seed” and squared, its ten middle digits were used as the random sequence and as the next “seed” when the calculation was run again. The method has a number of easily identifiable limitations: there are a finite number of numbers which can be generated by it, even at best, before it “repeats,” and some seeds would quickly “degrade” into rapidly repeating sequences (if the ten middle digits were all zeros, for example, there is no way it could continue). In much of the literature on this subject, this has been viewed as a blunder on the part of Von Neumann—the great mathematician, architect of the computer, and theoretical physicist renowned for his ability to do incredibly complicated

⁷² Ibid., 36.

mathematical calculations in his head had apparently not seen the limitations of this incredibly simplistic method.

In reality, of course, Von Neumann made no such slip; he was well aware of the limitations of the “middle square” method. In fact, what others identify as a limitation—the possibility for sequences to be catastrophically “destroyed” by self-perpetuating zeros—was seen by him as a benefit. That “in some cases the zero mechanism is the major mechanism destroying the sequences in encouraging, because one always fears the appearance of undetected short cycles.”⁷³ A spectacular failure would at least be easy to detect, unlike more subtle errors. In any event, generating sequences was easy—testing was the laborious part:

Hence the degree of complication of the method by which you make them is not terribly important; what is important is to carry out a relatively quick and efficient test. Personally I suspect that it might be just as well to use some cooking recipe like squaring and taking the middle digits, perhaps with more digits than ten.⁷⁴

All “cooking recipes,” in Von Neumann’s eyes, were fairly arbitrary, and none had anything much to do with a true sense of “randomness,” so better to quickly find ways of testing sequences which could easily be churned out.

The advantages of these “pseudo-random” methods, as they would later become known in an attempt to distinguish their deterministic qualities from any notion of “true” randomness, were many.⁷⁵ For one, all you needed to know in order to reproduce a given set of numbers later (for verification, checking, reproduction, etc.) was its initial “seed,” reducing greatly the amount of actual effort needed to record any given sequence (imagine all of the effort that poor assistant at the Biometric Laboratory would have been

⁷³ Ibid., 37.

⁷⁴ Ibid., 37.

⁷⁵ The third edition of the *Oxford English Dictionary* cites the first use of “pseudorandom” to an IBM seminar publication from 1949. “Quasi-random” was a term used to similar effect.

saved had Tippet only required the hand-transcription of a single number). Additionally, their effect on early computing cannot be underestimated: it took the ENIAC 600 milliseconds to read a “random number” from a punched card, while it required only 3 to 4 milliseconds to generate one using the “middle square” method. Von Neumann’s “blunder” was to choose a method 200 times faster than any others feasible at the time. Using 38-bit numbers with the “middle square” method, it should be noted, Nicholas Metropolis was able to generate sequences of around 750,000 digits before “degeneracy”—a sequence many times longer than the tediously-copied methods of Kendall and Smith.⁷⁶

Von Neumann was under no delusions. What mattered was expediency and practicality, not “randomness” in an ideal sense. This marked the beginnings of the third shift, moving “randomness” even further away from the philosophical questions posed at the beginning of the century when there was an attempt to formulate a definition of “randomness” based entirely on the *end-product*. As Von Neumann wrote to a physicist at Oak Ridge in 1948, this “randomness” was just “random for practical purposes.”⁷⁷

RAND’s million

Not everybody was convinced by Von Neumann’s “cooking recipe” approach, though. Though it seems as though many of those “high-speed” computers of the 1940s and 1950s used “pseudo-random” processes of different sorts for their Monte Carlo methods, the rest of the world (scientific and otherwise) was another question. As the idea of the Monte Carlo method (perhaps more than any single application of it) spread beyond the

⁷⁶ Cited in chapter 3, section 3.1 of Donald E. Knuth, *The art of computer programming*, Vol. 2, *Seminumerical algorithms* 2nd edn. (Reading, MA: Addison-Wesley, 1981).

⁷⁷ Letter from John von Neumann to Alston Householder (3 Feb 1948), cited in Galison (ref. 11), 703.

realm of the ENIAC, the idea of using deterministic algorithms to generate “random” sequences was viewed with suspicion by more than a few practitioners. Lists of numbers would, for a few decades at least, still be preferable for problems which required less urgency than beating the Soviets in the race for thermonuclear weapons.

In 1947, only a few months after the Monte Carlo was first being applied to ENIAC problems, a group of researchers at the RAND Corporation (then still known as Project RAND) set out to create “a large supply of random digits, of sufficiently high quality so that the user wouldn’t have to question whether they were good enough for his particular application in the case of every different application,” as one project member put it.⁷⁸ The effort was in many ways very similar to the project of Kendall and Smith, though pumped up on postwar funding, manpower, and computing resources. Using a random frequency pulse source instead of Kendall and Smith’s operator and stylus, the RAND engineers carried over the “roulette wheel” metaphor to a binary counter, which would then be processed to form their single random digit and fed to an IBM punch. Between April 29 and May 21, half a million digits were generated, which were submitted to Kendall and Smith’s four tests, and a few others, passing without difficulty. Another half a million were created by July 7, but these produced somewhat anomalous results under testing. Further investigation centered on two blocks of a quarter-million digits each, one produced between June 4 and 10, immediately following a “thorough tune up” of the machine, and the other between July 7 and 8, after a month of constant operation without adjustment. The second block did considerably worse than the first, leading to the

⁷⁸ George W. Brown, “History of RAND’s random digits—Summary,” in A.S. Householder, G.E. Forsythe, and H.H. Germond, eds., *Monte Carlo Method*, National Bureau of Standards Applied Mathematics Series, 12 (Washington, D.C.: U.S. Government Printing Office, 1951): 31-32.

conclusion that “the machine had been running down in the month since its tune up”—without servicing, its ability to produce “random” digits could “degrade.”⁷⁹

The million digits on 20,000 punched cards were thus “rerandomized” by taking modulo 10 of the sum of each digit (that is, the remainder of each sum after being divided by 10) with a corresponding digit on the previous card. This final set was subjected to the battery of tests, and was proclaimed to “have a clean bill of health and have been adopted as RAND’s table of random digits.”⁸⁰ They were published as a 600 page tome in 1955 as *A million random digits and 100,000 normal deviates*, and remained a standard reference in the sciences until the proliferation of cheap computing four decades later.⁸¹ For those with computers who were still untrusting of “pseudo-randomness,” RAND also made their digits available as punched cards as well. In a fifteen page (anonymous) introduction, the description of the digit production was discussed in detail, the results of the “randomness” tests (all derived from Kendall and Smith) were given, and instructions for use were provided. How, for example, was one supposed to know where to start with the table, since starting with page one would result in same sequence each time? “In any use of the table, one should first find a random starting position,” the anonymous authors advised. “A common procedure for doing this is to open the book to an unselected page of the digit table and blindly choose a five-digit number,” which then could be used as the line number of the starting digits.⁸² Augustinian-age Biblical scholars had made a similar practice in taking passages from the Bible “at random” (bibliomancy), relying on God to guide their fingers to portentous lines. Our twentieth century scientists would

⁷⁹ Ibid., 31-32.

⁸⁰ Ibid., 32.

⁸¹ RAND Corporation, *A million random digits with 100,000 normal deviates* (Glencoe, IL: Free Press, 1955).

⁸² Ibid., xx-xxi.

throw their final hope of “randomness” onto the whims of the goddess Chance—but just in case this goddess failed them again, they were advised to mark numbers they had already used, so as not to accidentally repeat themselves.⁸³

Beyond “randomness”

The tension inherent in the notion of “deterministic randomness” is one which seems to have caused—and still causes—much suspicion and confusion among many of those whose scientific investigations tapped into “randomness.” A sign of this tension can be seen in the records of a talk given before the Royal Statistical Society in 1959 on the development of a “random number”-generating machine developed by the British General Post Office in the mid-1950s. ERNIE—“Electronic Random Number Indicator Equipment”—took the natural noise fluctuations in a set of ten neon discharge tubes and transformed it into a set of bond numbers, the holders of which would then receive a monetary prize as part of a scheme to encourage bond purchasing.⁸⁴ The hardware in ERNIE was based on another machine the Post Office had developed to simulate traffic conditions (using Monte Carlo simulations), and was used to assure “fairness” in the drawing (a valence of “randomness”—as justice—which will not be discussed in this paper, but is worth considering),⁸⁵ and likely also to add a bit of spectacle to the event.

⁸³ Ibid., xxi. A humorous discussion with Matthew Underwood and Nasser Zakariya was responsible for noting this similarity in style. The switching to the feminine goddess is in reference to Pearson's formulation discussed earlier.

⁸⁴ W.E. Thomson, “ERNIE—A mathematical and statistical analysis,” *Journal of the Royal Statistical Society, Series A (General)* 122:3 (1959): 301-333.

⁸⁵ In 1970, allegations that the Selective Service numbers chosen for the U.S. Army draft were not “random” led to considerable controversy. The problem, it was decided, was similar to Tippett and Kelvin's: inadequate mixing—those numbers most recently added had a higher chance of being chosen than others, which, as it was, correlated with their birthdates. See Stephen E. Fienberg, “Randomization and social affairs: The 1970 draft lottery,” *Science* 171:3968 (22 Jan 1971): 255-261.

After a lengthy presentation on the technical details of ERNIE and how its results fared in the standard “tests of randomness,” the discussion amongst those present was mostly a mixture of lighthearted comments and statistical analysis. A number of speakers suggested that perhaps a modified form of ERNIE could be used for statistical experiments as well, such as one who noted that a machine modeled on ERNIE “might be attached to a general purpose automatic digital computer in order to provide a source of random digits for problems in artificial sampling, operational research and for Monte Carlo methods.”⁸⁶ Another suggested that they could “take digits from ERNIE and put them on magnetic tape” would be an “acceptable solution,” to the need for “random digits” for simulation work, and noted how a number of “determinate random number” generators had failed him and others in the past.⁸⁷ Another concluded that “the existence of ERNIE settles the controversy about the relative merits of random and pseudo-random numbers and reduces the argument for pseudo-random numbers to one of convenience.” This same speaker then asked, “How many computer programmers who cheerfully use them would be happy if their fate in the state lottery was decided on the use of pseudo-random numbers?”⁸⁸

Only one poor fellow chimed in to defend the “pseudo-random” approach: “Some speakers have expressed doubts about pseudo-random numbers... For high-speed calculations, requiring many thousands or millions of random numbers, the multiplicative congruential [pseudo-random] methods are almost indispensable.”⁸⁹ The speaker representing the Post Office, who had given the paper on ERNIE, could only note that

⁸⁶ Comment by Dr. E.S. Page, in Thomson (ref. 84), 325.

⁸⁷ Comment by Mr. Berners-Lee, in Thomson (ref. 84), 326.

⁸⁸ Comment by Dr. K.D. Tocher, in Thomson (ref. 84), 331.

⁸⁹ Comment by Dr. J.M. Hammersley, in Thomson (ref. 84), 331.

“ERNIE was not designed as a general-purpose machine, although many of its features would be suitable.”⁹⁰

Whether this single incident is representative of the attitudes within the communities of statisticians (or whether they varied dramatically from those of the early computer scientists) is not terribly important for the scope of this paper. It is enough to indicate that there was no singular “triumph” of “pseudo-random” methods over “physical” approaches. As even one of the participants in RAND project remarked soon after its completion:

My own personal hope for the future is that we won’t have to build any more random digit generators. It was an interesting experiment, it fulfilled a useful purpose, and one can do it again that way, if necessary, but it may not be asking too much to hope that this addition property, perhaps, or some other numerical process will permit us to computer our random numbers as we need them.⁹¹

“Pseudo-randomness” had, however, brought the question of “randomness” to a new stage. If one were not content (as Von Neumann was) with choosing a method and relying only on the ability of “tests” to determine the worth of its output, the development of a process to generate “random” sequences of digits required a further rethinking of *what was desired* in a “random sequence” in the first place. As one pair of mathematician/computer scientists, Thomas E. Hull and A. Rodney Dobell, put it in an article from the early 1960s:

Our purpose all along has been to obtain sequences of numbers which can be considered to be drawn at random from a uniform distribution. The key phrase here is “can be considered to be.” We know our numbers are *not* drawn at random from a uniform distribution, but for practical purposes it is sufficient that they have the appearance of being so drawn. This is of course not the first time in life where we meet a situation in which it is only the appearance of what we are doing that matters!⁹²

⁹⁰ Thomson (ref. 84), 333.

⁹¹ Brown (ref. 78), 32.

⁹² T.E. Hull and A.R. Dobell, “Random number generators,” *SIAM Review* 4:3 (Jul 1962): 230-254, on 238-239. Dobell would later become an economist.

Hull and Dobell are, in my view, perhaps the most clear-headed writers on their subject, among the few who fully articulated what many were striving for at the time: a realization that all of these efforts, past and previous, were leading to a focus on what sort of sequence was meant by assigning it the term “random,” and that these requirements might deviate dramatically from any notion of “true randomness.” The regimes of testing were simply looking for “various statistical properties of the generated sequences,” and if enough thought was put into their mode of generation, the “performance” of any given generator would be theoretically predictable ahead of time.⁹³ The ultimate goal would be to move *beyond* “randomness”:

Perhaps we can also look forward to a future in which we will be able to provide, on demand, generators to suit specific purposes. The ideal would be to know what statistical properties we required for a particular sequence, and then to design a generator to produce such a sequence. We would use only sequences which were carefully manufactured to suit our purposes. Under such circumstances, how could anyone manage to get along with sequences which were known only to be truly random!⁹⁴

I see no evidence that this understanding—the logical outcome of redefining “randomness” in terms of “practicality” and “purpose”—was ever fully accepted by computer programmers or mathematicians. This is not to say that any of them necessarily believed they were creating “true randomness,” but most stopped at the point in their understanding where they saw that “pseudo-random” generators produced sequences which *appeared* “random,” and some realized explicitly that “appeared” simply meant “satisfied certain statistical tests.” In his treatment of the topic in the second volume of his seminal *The art of computer programming*, first published in 1969, computer scientist

⁹³ Ibid., 244.

⁹⁴ Ibid., 248.

Donald E. Knuth gets as far as concluding that “no sequence of ‘random’ numbers can be adequate for every application.”⁹⁵

In the decades after Von Neumann’s “middle-square” method—continuing all the way into the present—mathematicians, computer programmers, and statisticians have attempted to craft algorithms which would produce sequences of “random” digits along with more and more “tests” to assure their “quality.” Assessing their results is beyond the scope of this paper (and this author), but there seems to be a general consensus amongst the published literature that there is still an amount of “randomness” which shows itself to be elusive. A “new class” of generators developed by George Marsaglia and Arif Zaman in 1991 and heralded as capable of producing “immensely long sequences” passed a new compilation of “DIEHARD” statistical tests. But by December 1992, they were announced to have failed in at least one Monte Carlo simulation, and it was agreed that the results of any particular generator needed to be subjected to scrutiny, “*regardless* of the tests which the generator has passed.”⁹⁶ In a strange combination of historical trends, George Marsaglia generated a list of five billion random digits using one of his generators, subjected them to his “DIEHARD” tests, and has been distributing them on CD-ROMs and the Internet since 1995.⁹⁷

Whether it makes theorists nervous or not, “pseudo-random” generators seem to be accepted as the only practical way of producing such sequences, especially as the gap between how many numbers could be generated or recorded by “physical” methods and

⁹⁵ Knuth, (ref. 76), section 3.5, 145.

⁹⁶ George Marsaglia and Arif Zaman, “A new class of random number generators,” *Annals of Applied Probability* 1:3 (1991): 462-480; Alan M. Ferrenberg, D.P. Landau, and Y. Joanna Wong, “Monte Carlo simulations: Hidden errors from ‘good’ random number generators,” *Physical Review Letters* 69:23 (7 Dec 1992): 3382-3384, on 3384.

⁹⁷ George Marsaglia, *The Marsaglia Random Number CDROM including the Diehard Battery of Tests of Randomness* (Department of Statistics, Florida State University, 1995). Available online at: <<http://www.cs.hku.hk/~diehard/cdrom/>>, accessed 18 Jun 2005.

those which could be generated by ever-faster computers grew even wider. Even Marsaglia's five billion numbers could be used up in a reasonably limited amount of time by modern personal computers, depending on the problem for which they were used.

What makes "pseudo-randomness" particularly interesting in this framework is that the expectations of "randomness" have, by this point, long outstripped any notion of "randomness" being an essential property of "reality," a statement pushed even further by those who state that "randomness" in a true sense may not even be what they desire. A simple example of what this might mean in practice is indicated by Knuth's note that "local nonrandomness is necessary in some applications, but it is disastrous in others." In other words, some problems would benefit from long stretches of apparently "non-random" data, while others would suffer greatly from it, a difference which could affect the "testing" methodology used to validate any given series of digits or generator.⁹⁸ The transformation of "randomness" from an unstated property of "reality" into a mathematical notion which could produce data on command required, in the end, a dramatic move away from the idiosyncratic "reality" which once underpinned it, and "randomness" simply became a property of whatever types of tests used to determine it. The idea of "local randomness" itself began the real push in this direction, for it immediately suggested that one could dictate *how local* this "randomness" need be, implying "randomness" was a sliding scale rather than a strict philosophical concept, and moving quickly beyond the notion that a "random sequence" was composed of digits generated by a "random method" to the idea that the "method" could be entirely deterministic. Peter Galison has outlined how these would become shifted even further in the move from "pseudo-randomness" to "quasi-randomness" by removing "randomness"

⁹⁸ Knuth (ref. 76), 145.

completely, replacing it instead with a definition based on a “clumping” function (a low-discrepancy sequence, in the technical parlance), one which could set standards beyond those which would even be expected in any ideally “random” sequence.⁹⁹

Conclusion

The history of “random numbers” has been told by most as either a tale of linear triumph (more and more numbers) or constant failure (the inability to harness “true randomness”), and generally for the purpose of telling a story, one based on apparently paradoxical notions of “randomness.” I am not sure either narrative is a very accurate one. The cyclotron designers of the 1930s and 1940s built increasingly large machines to more stringent standards which indicated the limitations and inadequacies of previous models—is this a set of triumphs or failures? The only notion of “progress” which seems appropriate is one which states that different uses of numbers emerged, different requirements for these numbers emerged, and most of the applications they were put to seem to have eventually worked out for the purposes of those using them. For all of the apparent “failures” of “random numbers,” the hydrogen bombs still detonated, the simulations still functioned.¹⁰⁰

The versions of these histories by sophisticated commentators have not taken the discussion far beyond this, unfortunately. Most have dwelled on the paradoxical nature of having expectations for “randomness” without the reflection that those who developed

⁹⁹ Peter Galison, “Random philosophy,” in Evandro Agazzi and Massimo Pauri, eds., *The reality of the unobservable: observability, unobservability, and their impact on the issue of scientific realism*, Boston Studies in the Philosophy of Science, no. 215 (Boston: Kluwer Academic Publishers, 2000): 123-128.

¹⁰⁰ On the question of hydrogen bombs, though, it would be exceptionally interesting—were the records ever sufficiently declassified—to see how and to what degree simulations of internal bomb processes failed to predict the yields of a number of explosions which were known to diverge quite impressively from their predictions.

these “expectations” into “requirements” did not do so accidentally. As a well-cited example, the psychologist Lola Lopes used this history as a way of criticizing statisticians for believing that you could “test” for “randomness” at all, seeing it as directly connected to the problem of induction.¹⁰¹ Statisticians who would reject sequences with a million zeroes as “non-random,” Lopes claimed, are committing the same error that they would criticize the “naïve public” of doing when they would reject a sequence of twenty coin flips of “heads” as being non-random, and in any event, she elaborates, the latter case is really just a case of the statisticians misunderstanding the “naïve public” in the first place—while twenty “heads” are indeed just as likely as any other possible sequence, the number of non-homogenous sequences greatly outnumbers the possibility of a homogenous one, which the “public” clearly understands. But by not looking at what it is that the statisticians were actually looking for when they meant “randomness,” and why, Lopes misinterprets them in the same unjust manner that she accuses them of smugly doing with results from the “naïve public.”

The history of “random number” generation is historically interesting not simply because it is full of cute contradictions or frustrated statisticians. The transformation of “randomness” from a property understood as in the background of “reality” to one which could be “generated” at will from entirely self-contained logic is reflective of a number of intellectual trends in the history of statistics, and provides another example of how “common sense” concepts are transformed, through regimes of testing and practical necessities, into “scientific” concepts often quite different-looking than their common meanings. Around 1900, statisticians began to realize that standard metaphors of

¹⁰¹ Lola L. Lopes, “Doing the impossible: A note on induction and the experience of randomness,” *Journal of Experimental Psychology: Learning, Memory and Cognition* 8:6 (1982): 626-636.

“chance” were not in “practice” conforming to “theory.” Sampling methods and experiments were important, but required custom-made studies until the creation of Tippett’s table, the validity of which was proved by its behavior under use.

“Randomness” in this period was still the “background” expectation, what “order” was *not*. Once a tool like Tippett’s table was available, though, it could be submitted to tests and “randomness” itself could be rethought, and in the late 1930s sensibilities were independently emerging that “randomness” was a property of numbers which needed to be “tested” for, reflecting the assumption that the “true randomness” of a given method for producing it was not enough. This regime of “testing” shaped the concept of “randomness” into something *positive*, a set of expectations which in only a very limited sense resembled anything of more philosophical (or even “common”) notions of “randomness.” By the 1940s, with the practical requirements of the first computers, the expectation of “randomness” as a specific property of a sequence itself (rather than its method of generation) was pushed to its extreme and “random numbers” were produced by entirely deterministic algorithms. The burden of “randomness” shifted entirely from the *method* used to produce the numbers, to a *testable property* of the numbers themselves. Once this had been accomplished, the method no longer mattered so long as it produced an acceptable outcome.

In my mind, there is no contradiction here—it is simply a matter of redefinition, and many of those involved in these reconceptualizations were acutely aware that what they were doing was abandoning some definitions of “randomness” in favor of others to suit their practical purposes. In the process of turning “randomness” into a *positive* statement, it was divorced from metaphysics very early on, though it was not the most

amiable split. In much of the literature there is still a longing for a metaphysical notion of “randomness,” a hope that someday a source of “true randomness” will be adaptable to these requirements. But such longing misses the big conclusion of this historical process: that “true randomness” was never really good enough; that reality always fell short. Von Neumann’s “state of sin” was originally a chastisement to any who hoped they could get indeterminacy from determinism, but perhaps we can playfully apply a similar sentiment to those who forget that classical indeterminism was never reliably “random” enough for their practical purposes.