ADVANCES IN APPLIED MATHEMATICS 12, 428–454 (1991)

The Perception of Randomness*

MAYA BAR-HILLEL

Department of Psychology, The Hebrew University, Jerusalem 91905, Israel

AND

WILLEM A. WAGENAAR

Department of Psychology, University of Leiden, 2300 RA Leiden, The Netherlands

Psychologists have studied people's intuitive notions of randomness by two kinds of tasks: judgment tasks (e.g., "is this series like a coin?" or "which of these series is most like a coin?"), and production tasks (e.g., "produce a series like a coin"). People's notion of randomness is biased in that they see clumps or streaks in truly random series and expect more alternation, or shorter runs, than are there. Similarly, they produce series with higher than expected alternation rates. Production tasks are subject to other biases as well, resulting from various functional limitations. The subjectively ideal random sequence obeys "local representativeness"; namely, in short segments of it, it represents both the relative frequencies (e.g., for a coin, 50%–50%) and the irregularity (avoidance of runs and other patterns). The extent to which this bias is a handicap in the real world is addressed.

Randomness is a concept which somehow eludes satisfactory definition. Devices which are random by definition, such as fair coins, can nonetheless generate series of outcomes which lack the appearance of randomness (e.g., a very long string of *heads*), while some digit series, although clearly patterned, define normal numbers, namely, numbers whose decimal form provably passes all tests for randomness (e.g., the infinite series obtained from writing down all the counting numbers in order:

$$12345678910111213141516171819202122 23 \ldots).$$

*Paper based on a lecture given in a conference on Randomness, Columbus, Ohio, April 1988.

0196-8858/91 $7.50

In effect, randomness is an unobservable property of a generating process. Theory can assume this property, but in practice it can only be inferred indirectly, from properties of the generator's output. The inspection of outputs for "randomness" involves subjecting them to various statistical tests of these necessary, but not sufficient, properties. The conclusions based on these tests are thus inherently statistical in nature—there are no logical or physical proofs of randomness.

Why Study the Perception of Randomness?

Where people see patterns, they seek, and often see, meaning. Regarding something as random is attributing it to (mere, or blind) chance. Perceiving events as random or non-random has significance for the conduct of human affairs, since matters of consequence may depend on it. The market price of stocks is essentially a random walk, but people see trends in it, with the help of which they attempt to predict future prices. People may be promoted (or demoted) for strings of job related successes (or failures) that are in effect no more than chance results. Coincidences are given significant, and often even mystical, interpretations, because their occurrence seems to transcend statistical explanation.

The categorization of events into random or non-random is often done intuitively. Moreover, the emergence of "patterns" from what is essentially "noise" is so powerful, that people may reject the statistical analysis, even when it is available, in favor of the intuitive feeling. The perception, whether visual or conceptual, is so compelling as to withstand analysis. Much as the famous Müller–Lyer illusion is not dispelled by measurement (see Fig. 1), the perceived "clumping" of random events is not dispelled by statistical analysis (see Fig. 2). Indeed, it has led to the development of what statisticians call *clumping theory*.

One of the best known examples of the misperception of randomness concerns the pattern of German flying-bombs that hit London during World War II. "Most people believed in a tendency of the points of

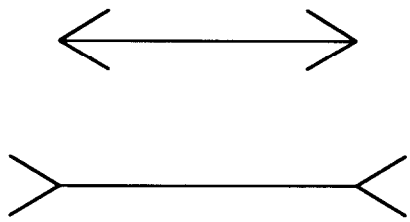

FIG. 1. The Muller–Lyer illusion. The arrows are of equal length.

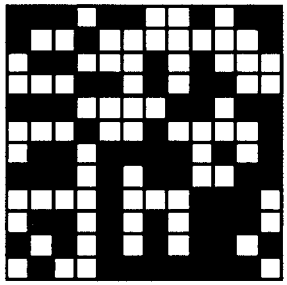

FIG. 2. A binary matrix with a 0.5 alternation rate.

impact to cluster," though analysis showed the "fit of the Poisson distribution [to be] surprisingly good" indicating "perfect randomness . . . ; we have here an instructive illustration of the established fact that to the trained eye randomness appears as regularity or tendency to cluster" (Feller, 1950, p. 120). Another well-known example is people's incredulity at the fact that among a sample of 40 people, chances are in excess of 90% that two will share a birthday. A perhaps less well-known example is often exploited by so-called "psychics," whose success at predicting people's supposedly "random" generations ("pick a number between 50 and 99, made of two different, and odd, digits") results from the combined facts that these generations are not really random and that the probability of both the "psychic" and the "medium" choosing the same number, even if it were really done at random, is highly underestimated (Marks and Kammann, 1980). As a final example, clumping causes people's experience often to bear out the superstition that "bad luck comes in threes."

The motivation for studying perceptions of randomness could also come from an interest in intuition. Intuition is accorded an important and respectable role in many types of judgments. Modern linguistics regards (a competent native speaker's reflective) intuition to be the final arbiter of grammaticality. In aesthetics (e.g., judging the quality of a piece of art) or ethics (e.g., judging the fairness of an allocation rule) intuition is often all there is to go by. When judging the pitch of a tone, the temperature of a tub of water, or the mean of a set of data points—intuitive judgments are found to often provide acceptable, yet quick and ready, approximations to objective measurement. The study of "intuitive tests of randomness," so to speak, thus acquires some interest also in comparison with normative tests thereof, even though—or because—there is no single statistical test sufficient to establish randomness.

This paper examines some of the evidence collected by psychologists about people's intuitions regarding randomness. Psychologists have not

themselves endeavored to define randomness. Rather, they have studied people's judgments of randomness, and their ability to "generate," or simulate, randomness. Often, though not always, they bypassed the problematics of the concept by omitting any mention of "randomness" in their instructions to subjects and, instead, talked directly about standard and familiar "random devices," such as coins or dies. For example, Bakan [3] instructed subjects "to produce a series of 'heads' and 'tails' such as they might expect to occur if an unbiased coin were tossed in an unbiased way." The subjects' output is then compared with either a formal analysis or a simulation of the device in question on the properties of choice.[1]

The paper is organized according to: a. What—the descriptive properties of people's judgments and/or productions; b. How—modelling the judgmental process and/or the cognitive production mechanism; c. Why —what sustains the erroneous subjective concept of randomness, and why it is not unlearned with experience.

We attempted a thorough, though not exhaustive, survey of the psychological literature on the perception of randomness. A small number of studies was selected for special, and more detailed, attention. In the course of surveying these studies we have digressed occasionally into tangential, but relevant, issues in general cognitive psychology. We hope the reader finds these digressions to be of interest.

A. What?

It is perhaps unfortunate that early laboratory studies of the perception of randomness relied on production tasks ("do like a coin") rather than judgment tasks ("is this like a coin?"), and the former still form a majority of the existent studies. Insofar as systematic biases are found in such tasks, it is apparent that they could be either accurate reflections of biased notions of randomness, or biased reflections of accurate notions of randomness (or both). By analogy, people who have linguistic competence may still produce ungrammatical sentences in actual speech, and people who have a good musical ear may fail to carry a tune properly. Hence, judgment tasks are a purer way of studying the perception of randomness. Nonetheless, the basic biases in subjective perceptions of randomness were discovered already by the early production tasks.

[1]Occasionally, instructions have been biased. For example, Baddeley's [2] told subjects that the kind of random sequence they were expected to generate "would be completely jumbled and would not therefore be likely to comprise [meaningful patterns]" (p. 119). A critique of other instructions can be found in Ayton, Hunt, and Wright [1].

TABLE I
Description of Ten Pre-1970 Random Production Studies

Author & year	# of symbols	Symbols	Length	Medium	Pace	*N*
Baddeley '66	26	Letters	2 × 100	Saying	4, 2, 1, .5	12
	26	Letters	16 × 100	Saying	2	12
	2, 4, 8, 16, 26	Letters	3 × 120	Writing	1	124
	2, 4, 8, 16, 26	Letters	3 × ?	Writing	—	120
	2, 4, 8, 16, 26	Digits	3 × ?	Writing	—	92
Bakan '60	2	H/T	2 × 150	Writing	—	70
Chapanis '53	10	Digits	1 × 2520	Writing	—	13
Rath '66	2	Digits	10 × 250	Writing	0.5	20
	10	Digits	10 × 250	Writing	0.75	20
	26	Letters	10 × 250	Writing	1	20
Ross '55	2	Digits	1 × 100	Stamping	?	60
Ross & '58	2	H/T	8 × 20	Writing	?	15
Levy '58	2	H/T	4 × 20	Writing	?	15
Teraoka '63	5	Digits	1 × 1252	Saying	—	4
	5	Letters	1 × 1252	Saying	—	4
	5	Nonsense Syllables	1 × 251	Saying	—	2
	5	Digits	1 × 751	Saying	1	3
	5	Digits	1 × 752	Saying	—	3
Warren & Morin '65	2, 4, 8	Digits	6 × 500	Saying	0.25, 0.5, 0.75	2
Weiss '64	2	Buttons	1 × 600	Pushing	1	28
Wolitzky & Spence '68	10	Digits	1 × 45	Writing	2.45	20

The Basic Findings

Table I (from Wagenaar [48]) shows the many varieties of tasks that were used in several early studies of random productions (an even earlier review can be found in Tune [42, 43]). These studies were startingly unsystematic. The number of alternatives varied from 2 (e.g., heads–tails, digits, card suits) to 26 (letters), produced in from 1 to 16 series per subject, of length from 20 up to 2520 each. Mode of production (e.g.,

TABLE II
Results of Ten Pre-1970 Random Production Studies

Author & Year	Measure	Order	Results
Baddeley '66	Stereotyped and repeated digrams	1	stereotyped digrams
	redundancy	0	unbalanced frequencies of 1- and 2-grams
Bakan '60	number of runs	1	avoidance of symmetric response patterns
	alternation and symmetry in trigrams	2	
	frequency of singletons	0	
Chapanis '53	frequency of singletons	0	unbalanced frequencies of 1, 2, and 3-grams
	frequency of digrams and trigrams	1, 2	preference to decreasing over increasing series
	autocorrelation	1,?	
Rath '66	frequency of singletons	0	preference for symbols adjacent in the natural order
	frequency of digrams, corrected; modified	1	
	frequency of trigrams, corrected	2	
Ross '55	frequency of singletons	0	
	number of alternations	1	
Ross & Levy '58	number of alternations	1	overuse of run length with expected frequency ≥ 1
	occurrence of runs	1–?	
Teraoka '63	frequency of singletons	0	stringing responses in their natural order
	conditional probabilities	1	dependencies over gaps of 5
	frequency of digrams, modified	1	periodicity with cycle 5
	frequency of runs	1–4	
Warren & Morin '65	redundancy	0–3	
Weiss '64	frequency of n-grams, corrected	1–9	preference for symmetric trigrams
	frequency of trigrams		
Wolitzky & Spence '68	frequency of singletons	0	

writing, calling out), as well as speed (from no limitation to 4/s), availability of memory aids (from none to complete record), and other factors, also varied across the studies. With the exception of number of alternatives and the required production rate, these variables were not varied within a single study.

Table II gives the measures or tests for nonrandomness that these studies employed. Deviations from randomness can appear in various forms. First, it is possible that the various alternatives be chosen with unequal frequencies. This is called a zero-order sequential effect, because

it does not involve any sequential property of the response series. Then, it is possible that the first-order transition probabilities reveal a sequential dependency. This first-order sequential dependency would appear as a deviation from the expected frequencies of pair combinations (or, equivalently, from the expected number of either alternations or runs). First-order effects extend across a distance of one in the sequence. Higher order effects test for dependencies between elements that are two, three, etc., places apart in the sequence. Some statistical measures of randomness confound effects of various orders (e.g., a paucity of runs of length n subsumes a first-order alternation effect), and some require longer sequences than others in order to apply meaningfully (e.g., testing runs of length $n + m$ versus of length n). Statistically equivalent measures may correspond to different psychological processes, so the distinctions between them should be kept in mind for interpretative purposes.

The findings of these studies were reported in terms of the measures selected, and sometimes invented, by their authors (see Table II). In spite of the variety of tasks and measures used, two fairly robust findings emerged from these studies, that have since withstood the test of time. The first concerns what are essentially motor biases, specific to the production task, and to the medium within which it is performed. It has no parallel in judgment tasks and is not a reflection of notions of randomness. This is overpreponderance of short strings (2 or 3) of symbols that are adjacent in some "natural" sequence (e.g., consecutive numbers or letters) or some artifactual one (e.g., on the keyboard). We shall have little further to say about this finding.

The second is that "human produced sequences have too few symmetries and long runs, too many alternations among events, and too much balancing of event frequencies over relatively short regions" [29, p. 392]. The alternation bias is also called the *negative recency* effect, and it confirms an hypothesis which Reichenbach had stated as early as 1934 [35]. The balancing over short regions is also called the *local representativeness* effect [22]. These closely related effects have been found to extend up to the sixth order of dependency. They are regarded as essentially cognitive biases, and as such, they have direct counterparts in judgment tasks.

In typical judgment tasks, sequences are presented to subjects, who are requested to select the sequence that is "more likely to have been produced by a fair coin" or "most random," or—conversely—which is "most patterned." Few researchers thought to study the perception of randomness in matched judgment and production tasks. An exception is Falk [14], who asked subjects to rate for "randomness" exactly the same kinds of binary strings or binary matrices that her other subjects had been required to generate. In the production phase, Falk gave subjects 20 green and 20 yellow cards and asked them to line them up "the way they would

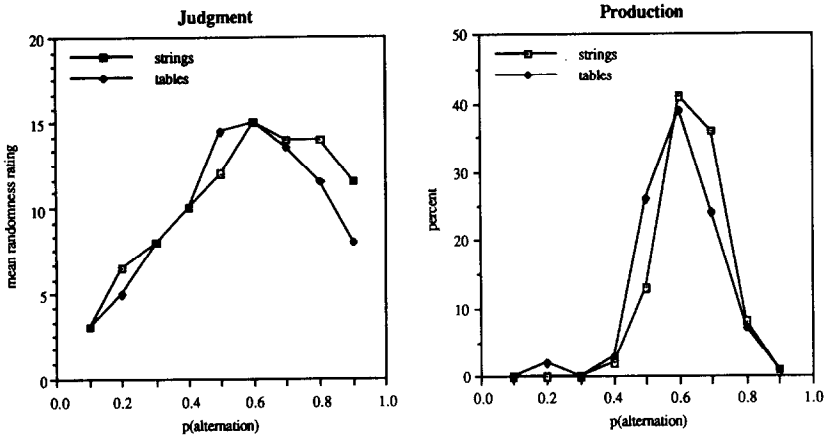

FIG. 3. Results of Falk's 1981 study, separated for the judgment and the production tasks.

be if they were well shuffled."[2] She found the preferred alternation rate to be not 0.5, but 0.6. Falk extended her study to two-dimensional arrays—specifically, 10 × 10 matrices—instructing her subjects to color 50 of the 100 cells a single color "in a random way." She then generalized the notion of "alternation" to the matrix, defining it as two different-colored cells with a shared side. Falk found that the preferred alternation rate in the two-dimensional arrays was again 0.6, "equal to the 99th percentile in the mathematical sampling distribution of random binary tables" of this kind [15]. In the analogous one- and two-dimensional judgment tasks, Falk found the same alternation rate (i.e., 0.6) to be rated as most random, in preference even to strings (or matrices) with an alternation rate of 0.5—which is the statistically expected value (see Fig. 2). Indeed, ask yourself how random you find the matrix in Fig. 2, which was generated by coloring each cell of the array with $p = 0.5$.

Wagenaar [45] found the same bias. Presenting subjects with seven binary sequences at a time of white and black dots on a neutral gray background, he too found that "sequences with conditional probabilities [of repetition] around 0.4 were judged as most random" (p. 348). Some studies have reported higher alternation rates as subjectively "most random" (0.7–0.8 in Gilovich, Vallone, and Tversky's, 1985 "Hot hand" study described below [18]).

In between-subject designs, the correlation between production biases and judgment biases is rather weak, around 0.3 for first-order deviation from randomness, dropping to about 0.2 for second-order effects, and

[2]This is not, of course, a Bernoulli process.

virtually disappearing for higher order effects [46; also 54, 55]. These tasks, however, were not as well matched as Falk's.

Other Judgment Studies

a. *Detecting random from non-random strings.* Lopes and Oden [29] studied judgments of randomness within a totally different paradigm—that known as *signal detection*. Their stimuli were short binary strings (length 8) —much shorter than the strings which production tasks ask for. These strings were computer-generated on-line for each subject (so they could be different for each subject). A subject viewed 250 strings generated by a Bernoulli process with $p = 0.5$. These were randomly interspersed with 250 strings generated by a process with an alternation probability of 0.8 (or, for other subjects, with an alternation probability of 0.2). On each trial, subjects guessed which process had generated the presented string and were given a small monetary reward if they guessed correctly. Note that "correct" here is literal, not to be confused with "optimal."

Subjects' performance (i.e., percent correct) was compared with the actual (rather than expected) percent correct of an "optimal" rule based on maximum likelihood. The rule averaged about 82% correct. The subjects' hit rate depended on the conditions under which they were guessing, namely: were they in the 0.8 or in the 0.2 alternation rate group?; were they informed about the direction of the alternation bias, or not?; if not, did they get trial-by-trial feedback about the correct source of the string? The results appear in Table III. All subjects performed at better than chance level, and those operating under the most favorable conditions did not do much worse than the optimal rule. There was considerable variability in percent correct as a function of the type of string (i.e., did the string exhibit a cyclic or mirror symmetry, or had it no particular pattern?)—the hit rate was much higher for the non-patterned strings, in line with familiar biases noted earlier. But even though performance for some types of strings fell considerably short even of chance, the

TABLE III
Probability of Making Correct Choice between Random and Non-random Source[a]

Uninformed; p(repetition) = 0.8	65%	Uninformed; p(alternation) = 0.8	52%
Informed; p(repetition) = 0.8	71%	Informed; p(alternation) = 0.8	67%
Feedback; p(repetition) = 0.8 [changed from 64% to 74%]	71%	Feedback; p(alternation) = 0.8 [changed from 56% to 76%]	67%

[a]Results of Lopes and Oden's [29] study.

ecological rarity of these "trick" strings prevented them from having much of an effect on overall performance. Lopes and Oden concluded that "biases in people's conceptions of randomness, although real enough, are less important to performance in the detection task than [whether or not judges know the direction of the bias in the non-Bernoulli process]" (p. 398). These results are compatible with previous findings (e.g., [8]) that showed people to be more successful in correctly selecting the more "random" stimulus when the distractors happen to be highly patterned (e.g., regular or symmetric) or frequency biased (i.e., contain a preponderance of one of the two symbols).

b. *The "hot hand" in basketball.* A provocative judgment study, whose conclusions raised much passionate debate was done by Gilovich *et al.* [18]. Rather than only manufacture artificial stimuli and ask subjects for their judgments, this study took as its point of departure a pre-existing naturally formed judgment and checked for its validity. The judgment in question is the belief, shared by basketball spectators and participants alike, in the existence of a "hot hand" or "streak shooting" phenomenon.

A corollary of the negative recency effect (and an example of the clumping effect) is that strings exhibiting the mathematically expected number of alternations will contain runs that appear to people to be inordinately long, hence "non-random." In a basketball context this could mean that a player's string of shots in a game will occasionally exhibit noticeable "streaks" of hits even if it is generated by a stationary process. Indeed, "analyses of the shooting records of the Philadelphia 76ers [including reputed streak shooter Andrew Toney] provided no evidence [for streak shooting]" (Gilovich *et al.*, [18, p. 295]). Contrary to the belief of 100 basketball fans who were surveyed for their beliefs regarding sequential dependence among shots, the very data forming the basis for this belief shows unequivocally that the probability that a player will make a given shot is not higher after having made a previous shot than after having missed it. Significantly, when Gilovich *et al.*'s subjects were shown on paper strings of 11 X's and 10 O's that were created using an alternation probability of 0.5, they saw streaks there, too. Apparently, the "streaks" that basketball aficionados report seeing in the game of some player are no more streaky than those exhibited by a random device whose p matches that player's hit rate. Hence players, even when they streak-shoot, are actually no "hotter" than random coins.

c. *The gambler's fallacy.* The *gambler's fallacy* is another name for the negative recency effect. It refers to gamblers' reputed tendency in a game of roulette to gamble on red after a run of blacks. In an ingenious study, Gold and Hester [19] demonstrated that this belief extends to an expectation that random devices somehow "intentionally monitor" their own output to make sure it balances out. Their subjects were shown a (secretly

manipulated) sequence of coin tosses. On each trial, a "winning side" was specified, and subjects chose whether or not to gamble on it. If they did, they earned 100 points on the winning side, 0 otherwise. If they did not, they simply got 70 points. 70 points was chosen because this is the value around which subjects usually split evenly between preferring the gamble or the sure thing. However, following a run of *heads* (*tails*), the subjects clearly preferred the gamble when the winning side was specified as *tails* (*heads*) and the sure thing when it was specified as *heads* (*tails*). This behavior is consistent with the gambler's fallacy. This was the case even for subjects whose verbally expressed beliefs countermanded the gambler's fallacy. Gold and Hester's most piquant finding is that either switching coins before the target toss, or allowing the coin to "rest" awhile prior to it, reversed this preference for the gamble almost completely. It is as if subjects believe that the "mechanism" that "causes" the coin to exhibit negative recency is a memory-type one: it does not carry over from coin to coin, and it decays over time.

Other Production Studies

a. *Production tasks with a large number of symbols.* Whereas judgments are not biased in terms of zero-order effects (i.e., relative frequencies), nor are productions which utilize a small number (e.g., two) of elements, production frequencies are seldom uniform when a large number of symbols is involved (see, e.g., Tables I and II). As a case in point, Triesman and Faulkner [41] found that, when asked to produce random sequences of the digits, subjects did not always produce uniform distributions. For example, 0 and 1 were typically underrepresented, while 3 was overrepresented (see Fig. 4). Moreover, this bias represents a consistent and systematic individual difference between subjects. The mean correlation between the distributions of digit frequencies given by subjects in two sessions conducted on different days was 0.79. Other tasks where responses allegedly drawn at random from a uniform distribution are not uniformly distributed will be mentioned in the section *Can people generate a single response "randomly"?* below.

b. *Avoidance of patterns.* Kubovy and Gilden [25] gave subjects an answer sheet designed for multiple choice tests and asked them to fill in the 240 circles in sequence, according to whether they imagine a coin coming up *heads* or *tails*. They found their subjects careful to maintain a balance of 50%–50% within short sequences (4 to 11), but found little evidence that subjects avoid "patterned" sequences, such as 000111, or 010010. By their own admission, however, such patterns have low salience when embedded in larger series (e.g., how salient are these two patterns in

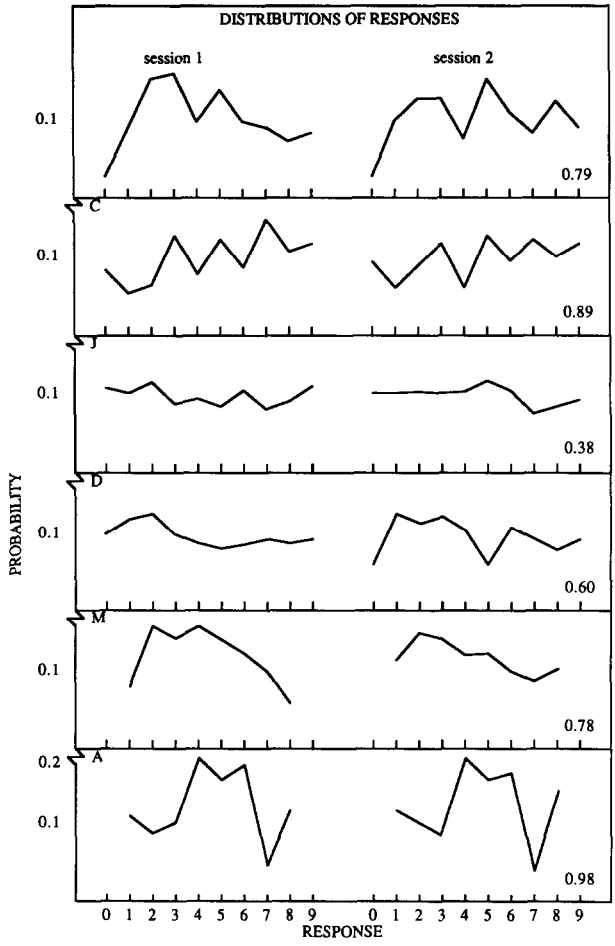

FIG. 4. Distribution of responses for six subjects in Triesman and Faulkner's [41] study. The numbers on the left are the cross-session correlations.

the sequence: 10100100011101101?), and are harder to control than the short-range balance.

c. *Stationariness in production*. Budescu [5] tested for, and found, stationariness in the series produced by his subjects. His subjects generated 60 series at most, however, and few if any of the subject-generated binary sequences extended to more than a few hundred elements. This is too short to test an hypothesis such that tolerance for, therefore the frequency of, longer runs or symmetric patterns will increase as the series extends—even if the rate of this increase will not be nearly enough. If this hypothesis is true, then the familiar biases can exist even in the absence of

stationariness. Indeed, Wagenaar [47] reported that "non-randomness decreases with time spent at the task" (p. 78).

d. *Probability learning: Predicting random sequences.* A paradigm that utilized production for purposes of prediction generated much research in the 50's and 60's (see [13, 21]). Subjects predict a randomly generated binary sequence trial by trial, with feedback given after every trial. The proportions of the two symbols were typically not equal. The major finding in the probability learning paradigm is called probability matching: subjects' prediction frequencies eventually came to match the observed frequency in the sequence. Though this is a clear manifestation that subjects learned the production probabilities, matching has a lower expected accuracy than merely predicting the majority category throughout (since $p \times p + (1 - p) \times (1 - p) < p$, for every $\frac{1}{2} < p < 1$). More pertinent to present purposes, subjects' predictions "seemed to have preferences for patterns... reflecting their ideas of chance" [42, p. 294], namely, avoidance of long runs.

A Learning Study

Although people's productions have failed many different types of tests for randomness, suitable feedback and training can apparently teach them how to overcome the biases they tend to exhibit. The most ambitious and thorough study of this kind was carried out by Neuringer [33]. In the Before phase, 11 students generated 60 series of 100 binary responses each (by self-paced typing of two digits on a computer keypad "as randomly as possible"). In the Feedback phase, feedback from several statistical descriptors was given just after each additional series of 100 responses was generated. There were two Feedback conditions, utilizing either five or ten descriptors (with 7 subjects in the first, 4 in the second condition). The set of descriptors is described in Table IV.

Feedback was given as follows: for each descriptor, five equiprobable categories were found by computer simulation, and subjects were shown a tabulation of the quintile values which their series of 100 responses had scored on the descriptors. They were told to try to produce an equal number of series in each of the five categories (quintiles). Subjects were started out with feedback on descriptor 1 only. As soon as they had managed to generate at least one series in each quintile, feedback on the second descriptor was added, until the same criterion was achieved on it, too, at which point the third descriptor was added for feedback purposes, etc.

Sessions lasted about an hour, in which time subjects typically produced 60 series (100 long) each. Subjects were free to ask questions and were given some explicit suggestions on how they might improve their perfor-

TABLE IV
The Statistical Descriptors Used in Neuringer's [33] Study

The first batch of five descriptors, in order of their introduction[a]	
1. RNG1	Based on the deviation of the observed frequencies of the 4 kinds of digrams from their expectancies
2. RNG2	Likewise, but alternate (rather than contiguous) responses are paired
3. C1	Likewise, but response i of trial n is paired with response i of $n + 1$
4. C2	Likewise, but response i of trial n is paired with response i of $n + 2$
5. Number of alternations	
The second batch of five descriptors, in order of their introduction	
6. Relative frequency of the binary symbols	
7. RUNS1	Number of singletons
8. RUNS2	Number of doubletons
9. RUNS3	Number of triplets
10. RUNS4	Number of quartets

[a]The notations and terminology are Neuringer's.

mance (i.e., how they might generate a series whose appropriate descriptor would fall into a chosen category). Feedback sessions were piled up as long as was necessary for a subject to reach a criterion of 60 consecutive series (or 2×60, in the 10-descriptor condition), none of which deviated significantly (according to the Kolmogorov–Smirnoff statistic) from the computer-generated "random" series on any of the five (or 10) descriptors. Although there was considerable variance in learning speed, all subjects eventually reached this criterion (e.g., after a maximum of 483 feedback trials when five measures were used, and 1771 when 10 measures were used).[3]

The last 60 series were additionally analyzed according to eight new descriptors on which subjects had not been trained (e.g., binomial test, one-sample runs test, some chi-square tests, and some auto-correlations). These are tests which naive subjects typically fail. However, two of the four trained subjects passed all eight new tests, and the other two passed six of them (though the *combined* responses of the last 60 series—i.e., 6000 responses per subject—failed some tests for all subjects).

In addition to these subjects, Neuringer himself served as a lone subject receiving feedback on 30 statistical descriptors in a 10-symbol generation task and eventually achieved criterion performance [32].

Neuringer was quick to admit that "for both a priori and empirical reasons, we cannot conclude that subjects learned to behave 'truly ran-

[3]Needless to say, no such improvements were noted in a control group receiving no feedback.

domly' " [33, p. 73]. Moreover, whatever it was that subjects did learn was not permanent in that it seemed to deteriorate markedly as soon as feedback was discontinued.

B. How?

Is There a Random Device in the Mind?

There is much evidence to show that people perform certain cognitive tasks by constructing or consulting mental images, little "pictures in the head," on which mental operations are performed in analogy to the operations that the represented object would have been subjected to in the physical world. For example, when asked on which side of his head George Bush parts his hair, many people report that they form a mental image of George Bush, and "look" at it to "see" the answer [23]; when asked what letter is formed when the letter N is rotated 90° clockwise, they report "rotating" a little mental picture of N in their head [9]; when asked at what speed a car was going when it crashed, people consult a mental video of the accident [27].

It thus could, in principle, have been the case that when people are asked to produce "heads" and "tails" as if they were tossing a coin, they form a mental image of a coin in their head, "toss" it, and read out the observed results. Alas, that seems not to be the case. Might there, nonetheless, be some "random device" in the mind, or in the nervous system, that generates the responses subjects produce, but, unlike mental imaging, is opaque to introspection? The observed biases could be introduced at a later stage by the "reporter" of the random device's outcomes.

There are a priori grounds against the plausibility of "random generators in the mind." First, when good random devices are so notoriously difficult to find in the external physical world—why would there be any in the mind? Second, an internal random device subject to censorship, instead of accounting for the presence of biases, merely adds the question: "how does the internal random generator work?" Third, the model smacks of homunculism. It is a little like explaining the creations of a mediocre composer as the result of some unfortunate mental meddling in the creations of a little Mozart residing in the mediocre composer's head. Nonetheless, some have argued for this possibility (e.g., [41]).

Kubovy and Gilden [25] tested a production model according to which people have an internal random device which generates strings, but those strings that turn out insufficiently representative are blocked or censored. They subjected Bernoulli strings to various kinds of "representativeness filters" (such as: "eliminate strings with runs longer than 4"; "eliminate strings with perfect alternation"; etc.) and compared this truncated set to

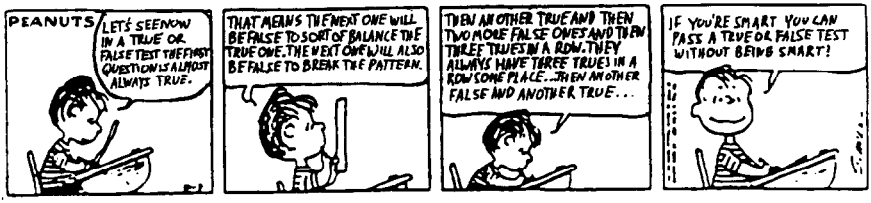

FIG. 5. The perfectly ordered true–false test, according to Linus (PEANUTS character © 1952, UFS, Inc.).

a set of human productions. The deviations between the simulation and the real thing was too large and systematic to be chalked to chance. Strangely, Kubovy and Gilden concluded that the observed biases in "subjective randomness" cannot be the result of representativeness filters, rather than concluding that there cannot be a random generator in the head.

The Account by Local Representativeness

In the accompanying "Peanuts" strip (Fig. 5), Linus is taking a true-or-false exam. Linus is assuming, presumably, that true-or-false test developers randomize the order of correct answers to foil the testees, yet he attempts to predict this order. His paradoxical conclusion: "If you're smart you can pass a true or false test without being smart!" nicely captures the essence of the paradox of randomness: in order to be perceived as random, sequences cannot be afforded to be constructed at random.[4] "[A]s a series of digits... comes closer and closer to satisfying all the tests for randomness it begins to exhibit a very rare and unusual type of statistical regularity that in some cases even permits the prediction of missing portions" [17, pp. 164–165].

Linus's sequence of *True*'s (0) and *False*'s (1) is nicely compatible with Kahneman and Tversky's [22] conclusion that, up to symbol exchange and left–right exchange, there is a single binary sequence of length 6 (01101-0) that is "ideally random." Indeed, Linus's string (01101-1000-10) exactly agrees with Kahneman and Tversky's in the first five places, an agreement also shared with Popper [57], who proposed an algorithm for constructing binary strings that start "randomly" and stay "random" throughout (the first 11 symbols in Popper's string are 01101-0111-10).[5]

Kahneman and Tversky's [22] series derives from their representativeness notion, according to which people judge the probability of events by

[4]This brings to mind Richard's paradox: "The smallest number that is undescribable in less than twelve words."

[5]We leave it to the reader to judge how closely these "ideals" match each other—and the reader's. The strings were broken into segments to facilitate the comparison.

the extent to which they represent the essential characteristics of their generating source. People also believe in a "law of small numbers" [44]; namely, they expect even small samples to be representative. Combining these two yields the simplest and most intuitive account of subjective randomness, that of local representativeness. By this account, when asked to judge which of a set of sequences is "most random," people look to see which captures the essential features of the random generating device best. In the case of a fair coin, say, these features are equiprobability of the two outcomes, along with some irregularity in the order of their appearance; these are expected to be manifest not only in the long run, but even in relatively short segments—as short as six or seven. The flaws in people's judgments of randomness in the large is the price of their insistence on its manifestation in the small.

Local representativeness readily accounts for the main results of the judgment tasks, namely, the high alternation rate and the rejection of pattern and symmetry. Nonetheless, some investigators have rejected it on the grounds that it is inconsistent with some of the production tasks results (e.g., the observation of non-uniform distributions in tasks using large numbers of alternatives) and because of the weak correlation observed between judgment and prediction tasks.

Clearly, however, production tasks make different cognitive demands than judgment tasks. For example, it takes a degree of musicality to hear with your inner ear how a piece of piano music should sound, but playing it to sound that way calls for technique as well. Thus, it is possible that in a judgment task, most subjects would judge the productions of Triesman and Faulkner's subject *J* (see Fig. 4) as more random than those of subject *A* —including subject *A*, their producer, himself.

The Account by Elimination of Alternative Nonrandom Hypotheses

Diener and Thompson [11] asked their subjects to guess which of 50 binary strings of 20 elements had been "produced by a random process similar to tossing a fair coin [and which] were generated by some other, nonrandom, methods" (pp. 438–439) and later to give their probability that each string was generated by a fair coin. The results showed that for strings classified as "nonrandom," reaction time (i.e., the time it took to give the response) was shorter the lower the subject's confidence in the string's randomness, whereas for strings classified as "random," the reaction time was on average higher, but no such relationship was found between it and the confidence ratings. The authors interpreted their results as showing that people compare each sequence to an ordered and fixed list of hypotheses which are alternative to randomness. Nonrandom agents high in the hierarchy produce both short reaction times (for saying

"nonrandom") and low probabilities (for "random"). Only when all alternatives are eliminated is a judgment of randomness given. The absence of a correlation between reaction time and confidence for the strings classified as "random" was taken by the authors to disconfirm an account relying on direct judgments of representativeness. However, many methodological weaknesses in this study render this conclusion dubious.

Can People Generate a Single Response "Randomly"?

Of all the factors that can account for the biases which are unique to production tasks, we shall focus on the role played by memory, a purely cognitive variable. We shall examine the body of production tasks results in light of the following hypothesis: people attempt to produce series that will match their subjective notion of randomness (the flaws in which are apparent from judgment tasks results), and the extra biases unique to production tasks are added by functional limitations on this attempt.

Memory for previous responses is as difficult to eradicate altogether as it is to stretch indefinitely. Non-random productions can be blamed on the fact that memory for previous responses is limited, as well as on the fact that previous responses are remembered at all. Both accounts uphold that productions are guided by a mental image of randomness. According to the former, if memory could be stretched indefinitely, people would not experience difficulty in keeping a mental tally of previous responses, and we would not obtain non-uniform (unrepresentative) distributions.[6] According to the latter, if previous responses were totally lost to memory, new ones could hardly be dependent on them, so we would not, for example, observe the high alternation rates. Memory operates differently on different measures of randomness.

The issue of previous responses is bypassed when we ask if people can produce a single response "at random." This is hard to test directly, because it is not even clear how to instruct people what it is we want them to do. However, since lay people often assume that what comes to mind first does so at random, or that spontaneity is tantamount to unpredictability, it is worth looking at some of the evidence that demonstrates that people's first, or single, spontaneous responses cannot be treated as random.

In a pertinent study, Kubovy and Psotka [26] asked people to report the first digit that came to their mind. The distribution of these digits is far from uniform. By far the most popular response is 7, which is typically given by almost 30% of respondents. In contrast, 0, 1, 2, and 9 are each

[6]Our discussion throughout this paper has assumed that all random variables are uniform, which of course they need not be, but everything said can be readily generalized to the non-uniform case.

given by considerably less than 10% of the respondents. "Why does 7 appear spontaneous?," ask the authors, and they reply: "Perhaps it is unique among the numbers from 0 to 9 because it has no multiples among these numbers, and yet it is itself not a multiple of any of these numbers. The numbers fall into groups: 2, 4, 6, 8 form one group; 3, 6, 9 form another. Only 0, 1, 5, 7 remain. One can rule out 0 and 1 for being endpoints, and perhaps 5 for being a traditional midpoint. This leaves us with 7 in the unique position of being, as it were, the "oddest" digit [26, p. 294].[7,8]

Additional evidence that digits do not come to mind at random can be found in Triesman and Faulkner's [41] results (Fig. 4 above), which showed high across-session within-subject correlations between digit frequencies.

Categories such as color, furniture, and fruits do not have a natural ordering, the way numbers do. For that reason, the determinants of the order in which they come to mind are opaque to naive respondents, who are therefore more compliant, and exercise less censorship, in reporting the first response that comes to their mind. Here again responses are highly predictable. When asked to say the first color, or piece of furniture, or fruit, that comes to mind, the most likely responses are, respectively, red, chair, and apple. These are also the most prototypical members of the respective category, as confirmed by other, independent observations [36]. Moreover, in these categories and others, three to four prototypical category members account for 80% or more of the category responses [4].

A related finding is that about 80% of subjects give "heads" as their first response when simulating a coin (e.g., [3, 20]). It is hard to argue that "heads" is more prototypical than "tails," so this is more likely a response bias reflecting the linguistic convention of preceding "heads" to "tails" in expressions such as "heads or tails?" Finally, Teigen [39] used a prediction task, rather than either a judgment or a production task, asking subjects to guess the outcome of a lottery where a number between 1 and 12 was

[7]Remember Linus?

[8]In this task, people might actually be disobeying the instruction. The first digit that comes to their mind may well be the first digit in the natural sequence of numbers. But it is censored (1 is the given answer in less than 3% of the cases) as not looking spontaneous enough. If so, then the reaction time of subjects responding with a "7" should be on average longer than that of subjects responding with a "1," because a higher proportion of the former rather than of the latter would have thought previously of some other digit and censored it. In Triesman and Faulkner's study, the second author, who served as one of the subjects (*A*, in Fig. 4), produced 7 with a frequency significantly lower than expected and lower than that given by other subjects, presumably because "This is known to be a favored random response... and is therefore one a sophisticated subject [such as Faulkner] might be wary of overproducing" [41, p. 341].

allegedly sampled at random. He too found a nonuniform distribution of guesses, with subjects clustering in the center, especially at 7, and avoiding the extremes.

The Role of Memory in Production Tasks

The previous section suggests that the role of memory in production tasks is by and large to assist people in producing sequences that capture their subjective notion of randomness. A well known fact about short term memory is that its span is 7 ± 2 items. This is also the span of immediate attention. This "Magical number 7" [31] may account for the size of the "windows" within which subjects try to achieve "representative randomness."

Subjects who wish to ensure that they are producing alternatives with equal frequencies must keep a mental tally of previous responses. For large sets of alternatives and long sequences this can be extremely difficult (as anyone who has tried to count cards at the blackjack table knows). One cognitive solution is to monitor the tally only within the limited segment of six or seven previous responses that can be readily remembered. Indeed, Wagenaar [49, Chap. 6] reported a cycling tendency (i.e., a tendency to use each alternative once in a cycle of A responses, with A being the number of alternatives) which peaked for $A = 6$, and a tendency to match relative frequencies within segments of 6–7 items, which was independent of A. Cycling was virtually absent for smaller A's (e.g., $A = 2$), where local representativeness (i.e., frequency matching plus absence of "pattern") can take care of everything. For larger A's (e.g., the alphabet), where the number of alternatives far exceeds the short term memory span, cycling is done over smaller subsets of letters at a time.

The strategies of cycling and matching in "moving windows" of size 6–7 can account also for some of the effects of the other factors that have been studied, factors such as the length of the sequence, the number and nature of the symbols, the rate of production, and the presence or absence of external aids for retaining previous responses. Unfortunately, only one study manipulated these various factors systematically [49]. In that study, however, the effects were analyzed in terms of deviations from the *mathematical* properties of random variables, not in terms of deviations from the *subjective* notion of randomness, which is what interests us here. Be that as it may, the study suggests that factors which strain memory (larger A's; no aids for keeping a record of previous responses) bring about more "randomness," presumably because they interfere with the attempt at local tallying and introduce more "noise." Ironically, since the subjective "ideal" of a random sequence is flawed, obstructions to carrying it out turn out here to be beneficial.

Other factors that have been studied are monotony, age, intelligence, mathematical sophistication, speededness, etc. (for a partial review see Tune, 1964b). Results, and especially conclusions, are mixed, no doubt because memory both aids the task (e.g., making it possible to keep a tally of first-order frequencies) and hinders it (making it possible for responses to rely on their predecessors). Thus, when writers describe the effects of various variables as increasing or decreasing randomness, it depends what particular property they are attending to.

Kubovy and Gilden's subjects [25, described above] also did their "bookkeeping" within windows, whose size varied (4–11—a range that slightly exceeds 7 ± 2). There were no memory requirements in that task, because all previous responses were available on a page, but the "windows" may have been necessitated by attention limitations.

C. Why?

The opportunities for observing and experiencing randomness, or at least unpredictability, would seem to be myriad (stock market prices, newborns' sexes, accidents occurrences, etc.), and the disadvantages of erroneous beliefs are self-evident. Yet in the course of time people either acquire an erroneous concept of randomness, or fail to unlearn it. In this section, we will try to see what in the nature of our experience, its observable properties, and the feedback it confers, might sustain people's erroneous concept of randomness, and why it is not unlearned. We will also speculate what advantages, if any, accrue to this biased concept.

Is it important to have a correct notion of randomness?

Insofar as being unpredictable is sometimes advantageous (as it demonstrably is, say, in various gaming and conflict situations) it would seem important to have a correct notion of randomness. Yet since one can use external aids to devise unpredictable strategic schemes, one need not rely on intuition for that goal. Moreover, though the kind of "unpredictability" that people regard as the epitomy of randomness differs *systematically* from the real thing, it does not differ *radically* from it. For example, a subjectively random binary sequence with an alternation rate of 0.6 rather than 0.5, can be predicted on average 52% of the time ($60\% \times 60\% + 40\% \times 40\%$), as compared to the 50% predictability of perfect chance. To have enough power (say, more likely than not) to detect a difference between 52% and 50% at customary significance levels (0.05 or less) requires large samples indeed—of at least 2500 observations.

Another advantage of immediate ability to recognize randomness (or unpredictability) is economical, insofar as it saves one the (futile) effort of attempting prediction. On the other hand: a. phenomena which yield phenotypically "random" observations can nonetheless be "caused" (in the commonsensical sense of the term) and, hence, potentially predicted, so the effort of seeking to explain or predict them can pay off. The distribution of newborns' sexes is nicely captured by the binomial distribution, but this is not to say that a newborn's sex cannot therefore be predicted at better than chance level; b. the ability to make lay predictions and causal attributions comes very easily to people. Indeed, people readily, and spontaneously, engage in attempts to predict—even paradigms of randomness, such as coins [19] and roulette wheels [51]. Hence, the cognitive gain in obviating such attempts may be paltry.

The Biased Nature of Feedback

One reason why experience with random sequences does not teach us the true nature of randomness (or, to put it more modestly, the accurate statistical properties of random variables) is because once a sequence exhibits properties that differ from our internal prototype for randomness, we often cease to perceive it as random. Thus, when people encounter longer runs in a binary sequence than they expected to "by chance alone," rather than saying to themselves "Aha! Long runs DO occur in random sequences," they say, "Aha! This seems NOT to be a random sequence" (recall, for example, the belief in the hot hand[9]; see also [10, p. 310]).

The self-fulfilling nature of this biased belief is also apparent in the occasionally biased nature of the feedback generated by erroneous judgments. Consider a decision maker who has made, under conditions of uncertainty, a chain of normatively optimal, but practically unsuccessful decisions (namely, wise decisions that have turned out badly). The decision maker can be a business person, a physician, whatever. Decision making under uncertainty is tantamount to gambling, and in gambling there are no guarantees of success. Suppose the unfortunate outcomes were just a coincidence, namely, were brought about "by chance." Though the decision maker cannot be faulted, her superiors erroneously perceive the long run of failures as evidence of incompetence. Unfair as this may sound, to the extent that it happens, it alters the nature of the distributions to which

[9]A similar phenomenon was noticed by Wagenaar and Keren [51], who showed that casino gamblers interpret long sequences of winning (losing) as streaks of good (bad) luck. As a consequence gambling acquires a skill aspect: a skillful player is one who recognizes a lucky streak when it occurs, exploits it, and can predict when it will end.

observers are then exposed: if decision makers are replaced after a bad run, they do not get the chance to have their bad run diluted by more typical future runs. Thus, their final track record, being based on a truncated career, is indeed poorer, on average, than that of a luckier, though not better, decision maker who did not have such a bad run. Thus, action based on a decision maker's possibly erroneous judgment generates evidence which ultimately confirms the erroneous judgment.

A different, though related, example of how actions based on erroneous judgments generate evidence confirming the judgment is discussed in Einhorn and Hogarth [12]. Imagine a waiter who prides himself on the ability to tell good from poor tippers at first sight. Having judged some customer to be a good tipper, the waiter gives her prompt and friendly service. Another customer, judged to be a poor tipper, gets the corner table and curt service. Lo and behold—at the end of the meal, the first indeed tips more generously than the second (for further examples, see [7]).

Arguably, the practice of removing random generators (decision makers, roulette wheels, ball players) after a bad run, though it confounds the nature of subsequent feedback, might ofttimes be a good idea, because truly poor performers are weeded out along with those who were just unlucky. While there is a cost to removing a "good" generator because of an accident of bad performance, there is also a cost in holding on to a "bad" generator who gets the benefit of the doubt.[10] The point is, that in a typical real-life context, a judgment of randomness versus non-randomness is made in the presence of an alternative source for the observed event, in a sense similar to Lopes and Odens' [29] signal detection paradigm. Of course, a string of 20 heads *could* have been generated by a fair coin, and indeed, as theoreticians take glee in pointing out, is precisely as likely as any other ordered string of outcomes. On the other hand, it is even *more* likely to have been generated by a trick coin or a prankster—so much so, that in our kind of world, rejecting the hypothesis of "fair coin" on that kind of evidence could be a very sensible thing to do.

When randomizing is deemed important—for example, in experimental design, as when selecting subjects to serve in the treatment versus placebo groups of an experiment—the conscientious researcher uses a random device. But if the random device just happened, as a fluke occurrence, to divide the subjects into groups in a patently unbalanced manner (e.g., all

[10] If a roulette wheel in Las Vegas exhibits an unusual run of reds, the House does not wait to see if it will stop, but rather the wheel is changed. Wheels are changed more frequently than warranted, in the sense that some of the removed wheels are still operating perfectly fine ("randomly"); but then again, some are not—and it is too costly to wait until enough evidence is in to make a confident assessment.

males in the control group, all females in the placebo), few researchers would serenely abide by that dictum. "The generation of random numbers is too important to be left to chance" (R. R. Coveyou, as quoted by Gardner [17, p. 169]). In that sense, even sequences generated artifactually in real life by physical random devices end up, due to doctoring, as biased representatives of their source.

Advantages of the Subjective Notion of Randomness

The perception of events as random or nonrandom is most important when it serves as a guide to action in real life. The major bias in the subjective notion of randomness is the overpreponderance of alternations (or underpreponderance of runs). However, "to evaluate fairly whether [this bias is] helpful or harmful over a lifetime . . . , one would have to know whether in the world nonrandom events are more often biased towards alternation or towards repetition" [28, p. 633]. Although Lopes declined to speculate on the answer, Ayton, Hunt, and Wright [1] were bolder. "[P]eople's apparently biased concepts may perhaps be . . . tuned to capitalise on properties of our environment. So, from an ecological viewpoint, perhaps repetition of outcomes is actually *correctly* considered to be more likely that alternation in non-random sequences. Or, it may be that the utilities associated with non-random events are structured so that it is more cost-effective to notice those non-random processes biased towards repetition at the expense of missing some of those non-random processes that are biased towards alternation" (p. 223, italics there).

Whereas the mathematical properties of random sequences are typically global properties, manifest in the limit of infinitely long runs, local representativeness is a local judgment—often the only kind that our cognitive apparatus can afford. Since experience, after all, is finite, the difference between local representativeness and the actual measures commonly employed by statisticians to test randomness would seem to be largely one of degree—local representativeness is a sort of poor man's goodness-of-fit, goodness-of-fit in the (very) small. Yet it is cheap, swift, widely applicable, and though slightly biased with respect to the normative standard, it "protects" against the gross departures from expectation that a fully honest and incorruptible randomizer must on rare occasions contend with.

ACKNOWLEDGMENTS

We thank David Budescu, Robyn Dawes, and Amos Tversky for useful and supportive comments on an earlier draft.

REFERENCES

1. P. AYTON, A. J. HUNT, AND G. WRIGHT, Psychological conceptions of randomness, *J. Behav. Decis. Making* **2** (1989), 221–238.
2. A. D. BADDELEY, The capacity for generating information by randomization, *Quart. J. Exp. Psychol.* **18** (1966), 119–129.
3. P. BAKAN, Response tendencies in attempts to generate random binary series, *Am. J. Psychol.* **73** (1960), 127–131.
4. W. F. BATTIG, AND W. E. MONTAGUE, Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms, *J. Exp. Psychol. Monog.* **80**, No. 3, part 2 (1969).
5. D. BUDESCU, A Markov model for generation of random binary sequences, *J. Exp. Psychol.: Hum. Percept. Perform.* **13**, No. 1 (1987), 25–39.
6. A. CHAPANIS, Random-number guessing behavior, *Amer. Psychol.* **8** (1953), 332.
7. M. D. COHEN AND J. G. MARCH, "Leadership and Ambiguity: The American College President," McGraw–Hill, New York, 1974.
8. A. COOK, Recognition of bias in strings of binary digits, *Percept. Mot. Skills* **24** (1967), 1003–1006.
9. L. A. COOPER AND R. N. SHEPARD, Chronometric studies of the rotation of mental images, *in* "Visual Information Processing" (W. G. Chase, Ed.), Academic Press, New York, 1973.
10. R. M. DAWES, "Rational Choice in an Uncertain World," Harcourt, Brace, Jovanovitch, New York, 1988.
11. D. DIENER AND W. B. THOMPSON, Recognizing randomness, *Amer. J. Psychol.* **98** (1985), 433–447.
12. H. J. EINHORN AND R. M. HOGARTH, Confidence in judgment: Persistence in the illusion of validity, *Psychol. Rev.* **85** (1978), 395–416.
13. W. K. ESTES, Probability learning, *in* "Categories of Human Learning" (A. W. Melton, Ed.), Academic Press, New York, 1964.
14. R. FALK, The perception of randomness, unpublished doctoral dissertation, The Hebrew University, 1975. [Hebrew]
15. R. FALK, The perception of randomness, *in* "Proceedings, Fifth International Conference for the Psychology of Mathematical Education, Grenoble, France, 1981," 222–229.
16. W. FELLER, "An Introduction to Probability Theory and Its Applications," Vol. I, Wiley, New York, 1950.
17. M. GARDNER, "Mathematical Carnival," Vintage, New York, 1977.
18. T. GILOVICH, R. VALLONE, AND A. TVERSKY, The hot hand in basketball: On the misperception of random sequences, *Cognit. Psychol.* **17** (1985), 295–314.
19. E. GOLD AND G. HESTER, The gambler's fallacy and the coin's memory, unpublished manuscript, Carnegie Mellon University, 1989.
20. L. D. GOODFELLOW, The human element in probability, *J. Gen. Psychol.* **23** (1940), 201–205.
21. M. R. JONES, From probability learning to sequential processing: A critical review, *Psychol. Bull.* **76** (1971), 153–185.
22. D. KAHNEMAN AND A. TVERSKY, Subjective probability: A judgment of representativeness, *Cognit. Psychol.* **3** (1972), 430–454.
23. S. KOSSLYN, "Image and Mind," Harvard Univ. Press, Cambridge, MA, 1980.
24. M. KUBOVY AND D. L. GILDEN, More random than random: A study of scaling noises, paper presented at the "Psychonomic Society Meeting, Seattle, 1988."

25. M. KUBOVY AND D. L. GILDEN, Apparent randomness is not always the complement of apparent order, *in* "The Perception of Structure" (G. Lockhead and J. R. Pomerantz, Eds.), Amer. Psychol. Assoc., Washington, DC, 1990.
26. M. KUBOVY AND J. PSOTKA, Predominance of seven and the apparent spontaneity of numerical choices, *J. Exp. Psychol.: Hum. Percept. Perform.* **2**, No. 2 (1976), 291–294.
27. E. LOFTUS, "Eyewitness Testimony," Erlbaum, Hillsdale, NJ, 1979.
28. L. L. LOPES, Doing the impossible: A note on induction and the experience of randomness, *J. Exp. Psychol.: Learn. Mem. Cognit.* **8** (1982), 626–636.
29. L. L. LOPES AND G. C. ODEN, Distinguishing between random and nonrandom events, *J. Exp. Psychol.: Learn. Mem. Cognit.* **13**, No. 3 (1987), 392–400.
30. D. F. MARKS AND R. KAMMANN, "The Psychology of the Psychic," Prometheus, Buffalo, 1980.
31. G. A. MILLER, The magical number 7 plus or minus two: Some limits on our capacity in processing information, *Psychol. Rev.* **63** (1956), 81–97.
32. A. NEURINGER, Melioration and self-experimentation, *J. Exp. Anal. Behav.* **42** (1984), 397–406.
33. A. NEURINGER, Can people behave "randomly"? The role of feedback. *J. Exp. Psychol.: Gen.* **115** (1986), 62–75.
34a. K. R. POPPER "The Logic of Scientific Discovery," Basic Books, New York, 1959.
34b. G. J. RATH, Randomization by humans, *Amer. J. Psychol.* **79** (1966), 97–103.
35. H. REICHENBACH, "The Theory of Probability," Univ. of California Press, Berkeley, 1949.
36. E. ROSCH, Principles of categorization, *in* "Cognition and Categorization," (E. Rosch and B. B. Lloyd, Eds.), Erlbaum, Hillsdale, NJ, 1978.
37. B. M. ROSS, Randomization of a binary series, *Amer. J. Psychol.* **68** (1955), 136–138.
38. B. M. ROSS AND N. LEVY, Patterned prediction of chance events by children and adults, *Psychol. Rep.* **4** (1958), 87–124.
39. K. H. TEIGEN, Studies in objective probability: Prediction of random events, *Scand. J. Psychol.* **24** (1983), 13–25.
40. T. TERAOKA, Some serial properties of "subjective randomness," *Japan. Psychol. Res.* **5** (1963), 120–128.
41. M. TRIESMAN AND A. FAULKNER, Generation of random sequences by human subjects: Cognitive operations or psychophysical process? *J. Exp. Psychol.: Gen.* **116**, No. 4 (1987), 337–355.
42. G. S. TUNE, Response preferences: A review of some relevant literature, *Psychol. Bull.* **61** (1964), 286–302.
43. G. S. TUNE, A brief survey of variables that influence random generation, *Percept. Mot. Skills* **18** (1964), 705–710.
44. A. TVERSKY AND D. KAHNEMAN, The belief in the law of small numbers. *Psychol. Bull.* **76** (1971), 105–110.
45. W. A. WAGENAAR, Subjective randomness and the capacity to generate information, *Acta Psycol.* **33** (1970), 233–242.
46. W. A. WAGENAAR, Appreciation of conditional probabilities in binary sequences, *Acta Psyco.* **34** (1970), 348–356.
47. W. A. WAGENAAR, Serial non-randomness as a function of duration and monotony of a randomization task, *Acta Psycol.* **35** (1971), 78–87.
48. W. A. WAGENAAR, Generation of random sequences by human subjects: A critical survey of the literature, *Psychol. Bull.* **77** (1972), 65–72.
49. W. A. WAGENAAR, "Sequential Response Bias: A Study on Choice and Chance," unpublished doctoral dissertation, Leiden University, 1972.
50. W. A. WAGENAAR, "Paradoxes of Gambling Behavior," Erlbaum, Hillsdale, NJ, 1989.

51. W. A. WAGENAAR AND G. KEREN, Chance and luck are the same, *J. Behav. Decis. Making* **1** (1988), 65–75.
52. P. A. WARREN AND R. E. MORIN, Random generation: Number of symbols to be randomized and time per response. *Psychonom. Sci.* **3** (1965), 557–558.
53. R. L. WEISS, On producing random responses. *Psychonom. Rep.* **1**, No. 4 (1964), 931–941.
54. S. WIERGESMA, A control theory of sequential response production, *Psychol. Res.* **44** (1982), 175–188.
55. S. WIERGESMA, Can repetition avoidance in randomization be explained by randomness concepts? *Psychol. Res.* **44** (1982), 189–198.
56. D. L. WOLITZKY AND D. P. SPENCE, Individual consistencies in the random generation of choices, *Percept. Mot. Skills* **26** (1968), 1211–1214.