Falk, R., & Konold, C. (1998). Subjective randomness. In S. Kotz, C.B. Read, & D.L. Banks (Eds.) Encyclopedia of statistical sciences (pp. 653-659). NY, NY: J. Wiley & Sons.

ENCYCLOPEDIA OF STATISTICAL SCIENCES

UPDATE VOLUME 2

EDITOR-IN-CHIEF

SAMUEL KOTZ
The George Washington University

EXECUTIVE EDITOR

CAMPBELL B. READ
Southern Methodist University

ASSOCIATE EDITOR

DAVID L. BANKS
National Institute of Standards and Technology

A WILEY-INTERSCIENCE PUBLICATION

John Wiley & Sons, Inc.

(PSYCHOLOGICAL DECISION MAKING
RANDOMNESS AND PROBABILITY
STATISTICS IN PSYCHOLOGY)

RUMA FALK
CLIFFORD KONOLD

SUBJECTIVE RANDOMNESS

The study of subjective randomness is of interest to psychologists exploring people's judgment of regularities in everyday and scientific contexts. Despite profound difficulties involved in defining randomness* (Hacking [19], Lopes [31], Ford [14], Kac [21], Ayton et al. [3], Gardner [16], Sheynin [36], Zabell [46]), psychologists have studied subjective randomness extensively since the 1950s, mainly using sequences as stimuli. People's responses have been evaluated by comparing them with the sampling distributions of major statistics derived from the sequences.

Early generalizations concerning conceptions of randomness were based on "probability-learning" experiments in which subjects predicted successive elements of random sequences, receiving trial-by-trial feedback. The conclusion was that humans are incapable of perceiving randomness. Convinced there was some pattern in the stimuli, most subjects believed the oncoming event depended on preceding ones (Lee [29]). They predicted sequences that deviated systematically from randomness. However, evidence concerning people's notion of randomness in these experiments is indirect. The produced sequences, which are influenced by various feedback contingencies, may largely reflect subjects' hypotheses concerning the goal of the experiment and their problem-solving strategies.

TWO TYPES OF SUBJECTIVE-RANDOMNESS EXPERIMENTS

In the first, and larger, class of subjective randomness studies, subjects *generate* random sequences under standard instructions to simulate a series of outcomes of a typical random process such as tossing a coin. These experiments (reviewed in Tune [38] and Wagenaar [41]) vary in procedure and instructions, in number of symbol types (possible outcomes) and sequence length. The second class of studies investigates people's spontaneous judgment, or *perception,* of randomness. Subjects rate the degree of randomness of stimuli or select the most random of several sequences (Wagenaar [40], Falk [10], Lopes and Oden [32]). Both classes of subjective-randomness research are reviewed in Bar-Hillel and Wagenaar [5]. The perception experiments reflect subjective concepts of randomness more directly, since people might find it difficult to express in generation what they can recognize in perception.

EXPERIMENTAL FINDINGS

The most systematic bias in subjective randomness is the notorious *gambler's fallacy.* People act as if they believe that in flipping a coin, as a run of heads grows longer, the probability of tails on next trial increases. Thus human-generated sequences are characterized by *negative recency,* a tendency to *overalternate,* and hence too many short runs*. In perception, people identify sequences with an excess of alternations as most random, while "truly" random sequences (those containing the modal number of runs) are judged as less random because the runs appear too long to occur by chance [5, 6, 9, 10, 12, 32, 40, 41].

The sequences mostly used in generation- and perception-of-randomness tasks were binary, and in judgment tasks the two symbol types were generally of equal frequency. Table 1 presents, for a variety of studies employing binary sequences, the *probability of alternation* Pr(A) (a) generated under standard instructions, and (b) perceived as most random among sequences with a range of Pr(A)'s. When the number of runs (r) in a sequence of n symbols was reported in the study, we computed Pr(A) as the ratio $(r - 1)/(n - 1)$.

As can be seen in Table 1, Pr(A) of about .6 recurs across many studies and experimental variations. The expectation of the sampling distribution of Pr(A) is $\frac{1}{2}$, or negligibly greater than $\frac{1}{2}$ in the constrained case of two symbol types of equal frequencies. The longer the constrained sequence, the closer to $\frac{1}{2}$ is $E[\Pr(A)]$ and the more extreme the percentile of the preferred subjective Pr(A) of .6 (see Johnson and Kotz [20] for the sampling distribution and moments of the number of runs, which is linearly related to Pr(A)).

Subjects' responses in tasks involving two-dimensional binary grids show a similar bias. Pr(A) of a grid is computed by dividing the number of color changes along rows and

Table 1 Mean Probability of Alternation [Pr(A)] Generated or Perceived as Most Random in Different Studies[a]

Reference	Randomness task	Size of set	Pr(A)
	(a) Generation		
Bakan [4]	Sequence	300	.59
Falk[b] [9]	Sequence (constrained)	(20 of each type)	.61
Falk[b] [9]	Two-dim. grid (constrained)	10 × 10 (50 of each type)	.63
Wiegersma [44]	Sequence	120	.56[c]
Budescu[d] [6, Tables 2, 3]	Sequence	20–40	.59
		60	.58
Rapoport and Budescu [35]	Sequence	150	.59
Kareev[b] [23]	Sequence	10	.61
Budescu and Rapoport [7, Exhibit 6]	Sequence	150	.58
	(b) Perception		
Wagenaar [40]	Select most random	Not reported	.6
Falk[b] [9]	Rate sequences	21	.6
Falk[b] [9]	Rate two-dim. grids	10 × 10	.6
Wiegersma [43]	Select most random	Not reported	.65
Diener and Thompson [8]	Rate sequences	20	.58
Gilovich et al. [18]	Classify as "chance," "streak," "alternate"	21	.7–.8
Wiegersma [45, Expts.1–3]	Select most random	40	.63; .64; .57[c]

[a]The expected and most probable Pr(A) in *random* productions is .5. Differences in decimal accuracy partly reflect reported differences in the original works.
[b]Averaged over different age and sophistication levels.
[c]As read from Fig.1 in Wiegersma [44, 45].
[d]Two different estimates based on the same data (of subjects exhibiting negative recency).

columns by the total number of possible changes. For 10 × 10 grids with 50 cells of each kind, as used by Falk [9], $E[\Pr(A)]$ is .51. Such a grid, however, was *not* perceived as maximally random by Falk's subjects. Grids judged as most random had Pr(*A*)'s of about .6 (Fig. 1).

The grids in Fig. 1 appear in the same order as their mean rated randomness and illustrate the negative skewness of perceived randomness as a function of Pr(*A*) (Falk [9, 10]). In contrast, the random variable Pr(*A*) is approximately normally distributed around .5. The peak of perceived randomness (.6) exceeds the 99th percentile in the sampling distribution of 10 × 10 grids with 50 cells of each kind (the distributions and moments of random variables such as the number of color changes in grids were worked out by Moran [33], Krishna Iyer [26, 27], and others).

Binary sequences whose Pr(*A*)'s vary from .1, .2, .3, through to 1.0 were presented to subjects for judgment of randomness in three studies (Falk [9] and Falk and Konold [12, Experiments 1 and 3]). The sequence length was 21 in the first two studies and 41 in the third. Despite procedural differences, remarkably similar outcomes were obtained. These results were pooled to obtain mean subjective randomness as a function of Pr(*A*). In each study, subjects' randomness ratings were averaged for each Pr(*A*), and the 10 means were standardized. The weighted average of the three standardized means was then computed for every Pr(*A*). The *subjective randomness* (SR) function was obtained by linearly transforming these 10 averages to range from 0 to 1, to allow comparison with information theory's second-order *entropy* (EN), a measure of a sequence's objective degree of randomness (Attneave [2, Chaps. 1, 2]). As can be seen in Fig. 2, EN is symmetric around .5, whereas SR is negatively skewed as a function of Pr(*A*).

EXAMPLES OF BIASED PERCEPTION

In the World War II rocket attack on London, people believed that the hits tended to cluster in specific zones. However, the distribution of hits per section showed a remarkably good correspondence to the expected (Poisson) distribution under the assumption of randomness (Feller [13, pp. 160–161]). Similarly, Gilovich et al. [18] describe a pervasive belief in the effect of a "hot hand" in basketball. Players, coaches, and fans all believe that when a player makes a basket, the conditional probability of making the next shot is greater than it is after a miss. However, the authors analyzed massive records of real games and showed that the hand of a basketball player is not any hotter than that of a coin flipper. Actual sequences of hits and misses were largely compatible with the expected output of a Bernoulli process, regardless of the player's overall hit rate.

In a casino setting, where sequences of wins and losses are characterized by sequential independence, gamblers attributed outcomes to a factor they called *luck*, which operates independently of chance [42]. Good (bad) luck produces longer streaks of wins (losses) than those

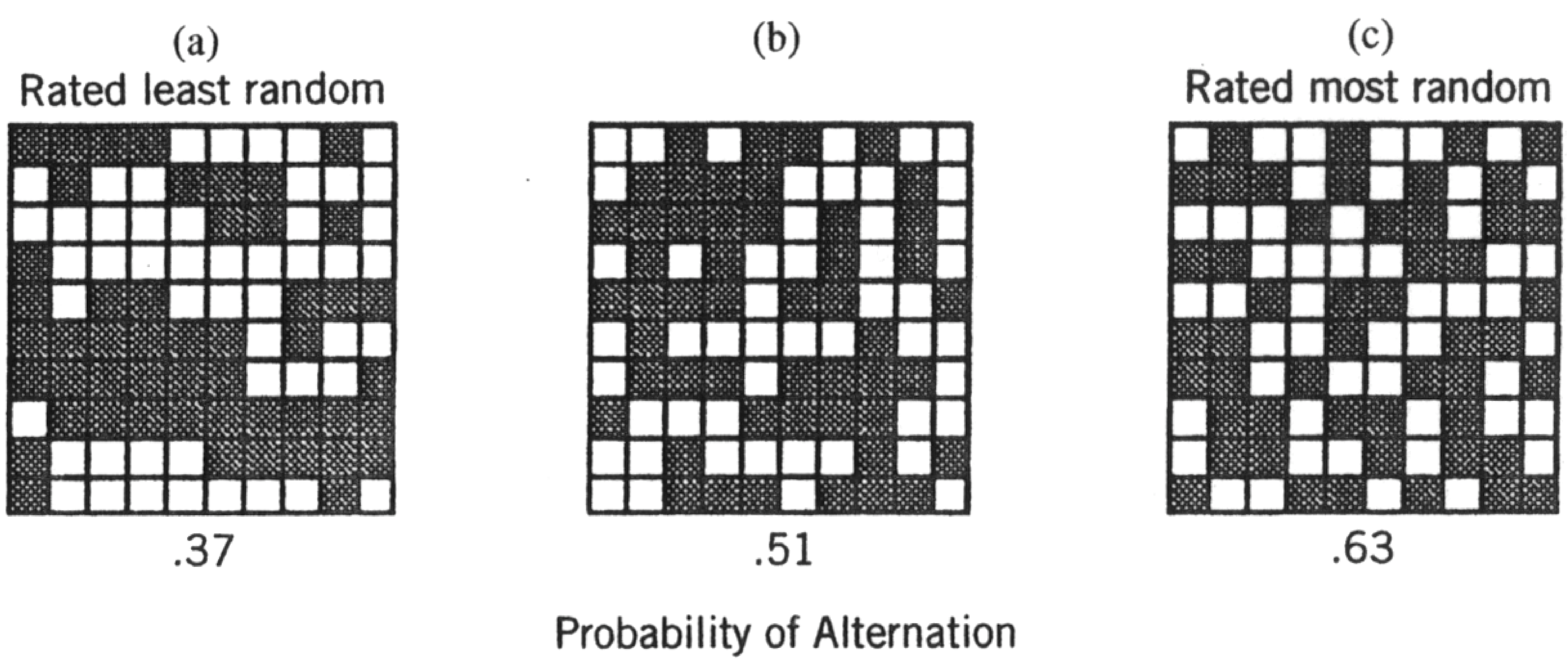

Figure 1 Three grids presented for randomness judgment by Falk [9], ordered according to their probability of alternation and mean perceived randomness ($N = 341$).

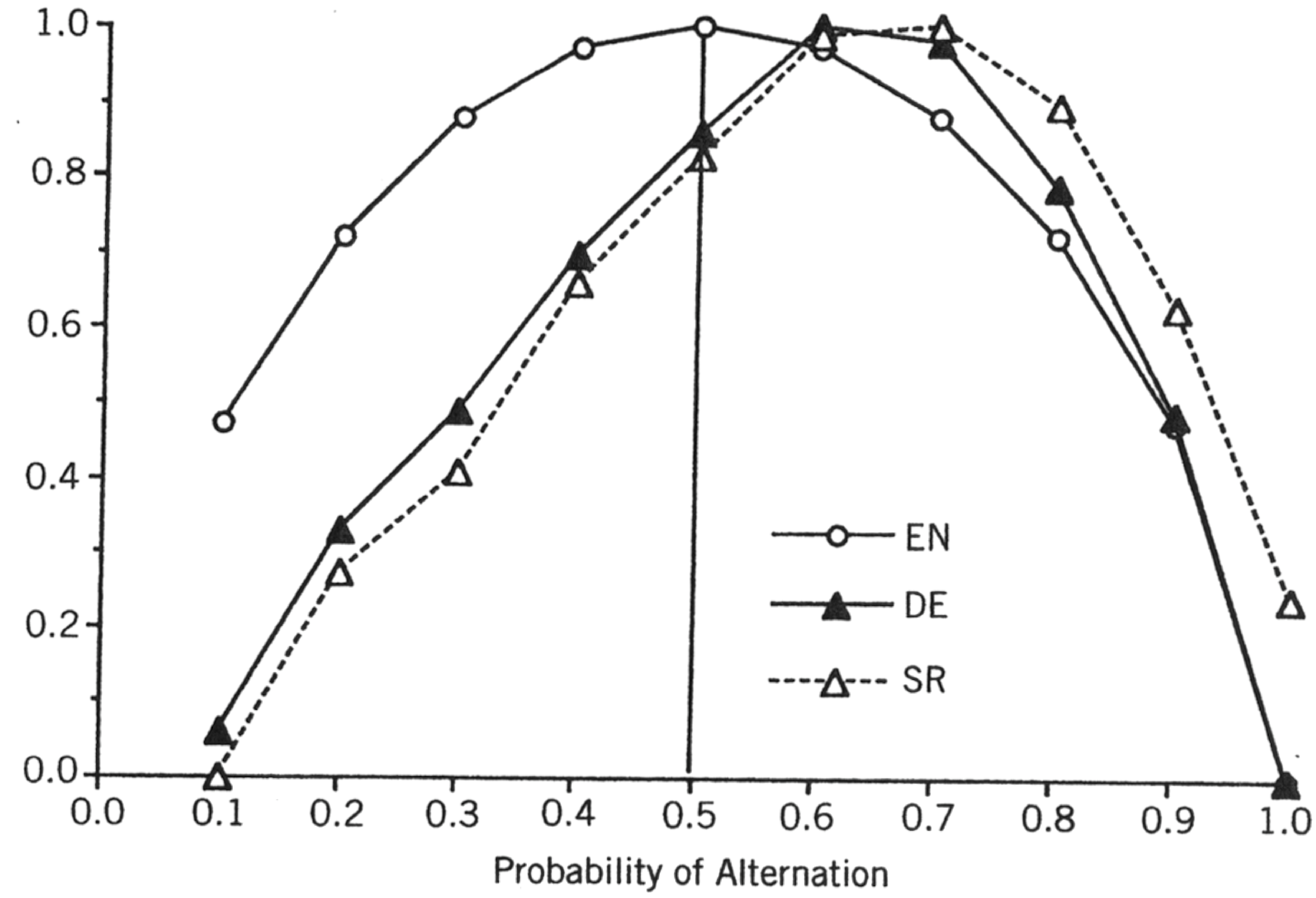

Figure 2 Second-order entropy, EN, and linearly transformed means of subjective randomness, SR ($N = 491$), and of difficulty of encoding, DE ($N = 160$), as functions of the sequence's probability of alternation (based on pooled results of Falk [9] and Falk and Konold [12]).

expected by chance. When luck is at work, the conditional probability of winning given a previous win is greater than it is given a previous loss. The analogy between the lay concepts of hot hand and luck is evident.

In all these cases, people are impressed by clusters that appear too large to be random. However, instead of adjusting their ideas about chance, they invent an idle "theory" to account for the apparent deviations from randomness. They mistakenly reject chance and thus commit a "Type I error." The other way of going wrong, "Type II error," occurs when one overlooks some structure in the stimuli. This is what happened in the many generation and perception studies when subjects viewed overalternating sequences as most random.

Researchers are not immune to these fallacies. Alberoni [1] presented a (supposedly random) sequence of 24 blue and 25 red beads to subjects who unanimously perceived it as random. However, the sequence contained 40 runs, which translates to Pr(A) of .81. This places the sequence above the 99.9th percentile in the sampling distribution of number of runs for sequences of this type. Apparently, Alberoni selected that sequence as a "good example" of a random sequence, and thus committed a Type II error.

ACCOUNTING FOR THE BIASES

Functional Factors or Concept?

A class of explanations attributes suboptimality in randomization to factors such as motor tendencies, boredom, and limitations of attention and short-term memory (reviewed by Tune [38], Wagenaar [41], and Bar-Hillel and Wagenaar [5]). However, the similarity of people's responses across generation and perception tasks and in the face of experimental variations argues against such functional limitations and suggests an underlying biased concept of randomness [40, 9]. This view is further supported by findings of individual consistency in people's performance of diverse tasks involving randomness [9, 6].

Local Representativeness

According to Kahneman and Tversky [22], subjects regard a sequence as random if it is *locally representative* of the salient features of its parent population and the process by which it was generated. Thus, a sequence of coin tosses should include about the same number of heads and tails in its entirety and in its various subsequences. At the same time, the sequence should display some irregularity. A sequence that satisfies local representativeness contains exaggerated alternations. People seem to regard chance as a self-correcting mechanism, which takes care to restore the balance whenever it is disrupted, as if they believe in "the law of small numbers" [39].

Although compelling as an account of subjects' responses, explaining *subjective randomness* by claiming that people expect *irregularity* is somewhat circular. Indeed, local representativeness succeeds in predicting excessive alternations, but it fails to predict the extent of this bias. It specifies neither how local our span of consideration is, nor how representative the local segments should be. (See Kubovy and Gilden [28] and Kareev [23] for promising attempts to delineate subjects' span of localness and the type of representativeness they try to attain in generation.)

Apparent Randomness as Subjective Complexity

Konold and Falk [25] and Falk and Konold [11, 12] examined the hypothesis that people judge the randomness of a sequence by assessing its complexity of structure. Subjects presumably attempt to make sense of the sequence, for example, by encoding it. The harder this task, the more random the sequence is perceived as. This hypothesis was inspired by Kahneman and Tversky [22], who suggested that apparent randomness is a form of complexity and the most random-appearing sequence would be the one whose description is longest. The idea accords with complexity theory's approach, which identifies randomness with *incompressibility*.

In principle, the *complexity* of a sequence (also known as *Kolmogorov complexity* or *algorithmic information content*) is the length of the shortest computer program that can reproduce the sequence. A technical outline of algorithmic information theory's* approach to randomness (including references to the founders of the theory in the 1960s) is given by Gács [15]. A random sequence cannot be condensed and thus has maximal complexity [14, 30, 17]. This definition of randomness is intuitively appealing, since strings that are incompressible must be patternless. A pattern could have been used to reduce the description length.

Subjective complexity has been studied by psychologists independently of the study of subjective randomness. Simon's review of theories and behavioral tasks [37] indicates that different measures of sequences' subjective complexity—such as number of errors in recall, length of description, and rated "goodness" of pattern—correlate highly with each other and are essentially interchangeable.

Falk and Konold [12] obtained, in a between-subjects design, randomness ratings and several measures of difficulty of encoding (subjective complexity) for the same sets of sequences. Pr(A) varied from .1 to 1.0 in .1 intervals in each set. SR values (as explained above) are plotted against Pr(A) in Fig. 2, alongside a composite measure of the *difficulty of encoding* (DE) and EN of these sequences. DE incorporates two encoding tasks: *memorizing* the sequence in as short a time as possible, and *copying* it while minimizing both viewing time and number of viewings. For every Pr(A), the standardized mean measures of difficulty in the two tasks were averaged. The 10 DE values were obtained by linearly transforming these averages to [0, 1].

As can be seen in Fig. 2, DE behaves much like SR. Both variables are negatively skewed as functions of Pr(A), and the highest two points of the two functions are at .6 and .7. Subjective randomness is better predicted by the sequence's encoding difficulty than by its objective randomness. SR's correlation coefficient with DE is .95, whereas with EN it is only .62.

The negative skewness of DE as a function of Pr(A) and its resemblance to SR are somewhat surprising. Overalternating sequences contain some cues which theoretically could be used to encode them more easily than sequences of Pr(A) = .5. Likewise, sequences of Pr(A) = .5 $\pm$ d, whose complexity and entropy are the same, should in theory be equally easy to encode. It is therefore instructive that one type of dependency (overrepetitions) is largely utilized, whereas the same degree of dependency of the other type (overalternations) is apparently not detected and seems even to impair performance. (Kareev [24] considers this asymmetry in judgment a rational predisposition for early detection of potentially more informative relationships.)

The similarity of DE and SR supports the hypothesis that tacit encoding mediates the judgment of how random a sequence is. On this account randomness is perceived when encoding fails. That subjective randomness results from people's failure to make sense of their observations is not a new idea. Piaget and Inhelder [34] attribute the origin of the idea of chance in children to their realizing the impossibility of predicting oncoming events or finding causal explanations. The experience of randomness is thus construed as an admission of failure of our intellectual operations.

End of "Subjective Randomness"