Psychological Review
1997, Vol. 104, No. 2, 301–318

Copyright 1997 by the American Psychological Association, Inc.
0033-295X/97/$3.00

Making Sense of Randomness: Implicit Encoding as a Basis for Judgment

Ruma Falk
Hebrew University

Clifford Konold
University of Massachusetts at Amherst

People attempting to generate random sequences usually produce more alternations than expected by chance. They also judge overalternating sequences as maximally random. In this article, the authors review findings, implications, and explanatory mechanisms concerning subjective randomness. The authors next present the general approach of the mathematical theory of complexity, which identifies the length of the shortest program for reproducing a sequence with its degree of randomness. They describe three experiments, based on mean group responses, indicating that the perceived randomness of a sequence is better predicted by various measures of its encoding difficulty than by its objective randomness. These results seem to imply that in accordance with the complexity view, judging the extent of a sequence's randomness is based on an attempt to mentally encode it. The experience of randomness may result when this attempt fails.

Judging a situation as more or less random is often the key to important cognitions and behaviors. Perceiving a situation as nonchance calls for explanations, and it marks the onset of inductive inference (Lopes, 1982). Lawful environments encourage a coping orientation. One may try to control a situation by predicting its outcome, replicating, changing, or even by avoiding it. In contrast, there seems to be no point in patterning our behavior in a random environment.

Although people feel that they know what they mean when speaking of *randomness* (Kac, 1983) and they communicate in everyday and professional affairs using their shared intuitive understanding of the term, it is one of the most elusive concepts in mathematics. Randomness resists easy or precise definition, nor is there a decisive test for determining its presence (Ayton, Hunt, & Wright, 1989, 1991; Chaitin, 1975; Falk, 1991; Lopes, 1982; Pollatsek & Konold, 1991; Wagenaar, 1972a, 1991; Zabell, 1992). Attempted definitions of randomness involve intricate philosophical and mathematical problems (Ayer, 1965; Fine, 1973; Ford, 1983; Gardner, 1989, chap. 13; Gilmore, 1989; Kac, 1983; Spencer Brown, 1957).

Despite these difficulties, psychologists have been carrying out, since the early 1950s, extensive research on people's subjective sense of randomness. This research uses our knowledge of the sampling distributions of several major statistics characterizing sequences of any length to serve as the normative background against which people's responses are evaluated.

Conclusions concerning participants' conceptions of randomness have been drawn from experiments employing diverse tasks. The earliest were probability-learning tasks consisting of successive predictions (with feedback) of elements of random sequences (e.g., Hake, 1955; Hake & Hyman, 1953; Jarvik, 1951; Nicks, 1959; Ross & Levy, 1958) and psychophysical and ESP studies (see review in Tune, 1964). It was generally claimed that participants cannot perceive sets of stimuli as random. Cohen (1960) summarized a series of experiments (mostly of the probability-learning paradigm), saying that "nothing is so alien to the human mind as the idea of randomness" (p. 42). Most of the sequences predicted by participants in these experiments deviated from randomness, mainly by alternating too frequently between different outcomes. However, the claim that participants' performance is a direct reflection of their distorted image of randomness has been seriously contested. The sequences produced in probability-learning experiments may be affected by participants' own previous responses and the obtained feedback. They may reflect participants' hypotheses concerning the nature of the experiment (Peterson, 1980) and their problem-solving strategies.

The majority of studies in this area looked at participants' generation of randomness. Participants were required to simulate the outcomes of some random mechanism, such as tossing a coin (see reviews in Tune, 1964, and Wagenaar, 1972a; studies by Kubovy & Gilden, 1991; Neuringer, 1986; Teigen, 1984; Wagenaar, 1970b; Wiegersma, 1982; and references to generation studies in Table 1). Other studies asked directly for partici-

Ruma Falk, Department of Psychology and School of Education, Hebrew University, Jerusalem, Israel; Clifford Konold, Scientific Reasoning Research Institute, University of Massachusetts at Amherst.

This study was supported by National Science Foundation Grant MDR-8954626 and by the Sturman Center for Human Development, Hebrew University, Jerusalem, Israel.

We are grateful to the late Amos Tversky for his helpful comments on a draft of this article. For their cooperation and assistance in collecting data, we thank Lisa Kainan, Jill Lohmeier, Joan Garfield, Meirav Arieli, and graduate students of the 1993–1994 seminar on "The Psychology of Chance" at Hebrew University, Jerusalem, Israel. We benefited greatly throughout all the stages of this study from the continuous advice and dedicated help of Raphael Falk.

Correspondence concerning this article should be addressed to Ruma Falk, Department of Psychology, Hebrew University, Jerusalem, 91905 Israel. Electronic mail may be sent via Internet to rfalk@shum.cc.huji.ac.il.

pants' perception or judgment of randomness by obtaining their ratings or by having them select (classify) random sets of stimuli (see Falk, 1981; Lopes & Oden, 1987; and references to perception studies in Table 1). A recent comprehensive review of subjective-randomness research can be found in Bar-Hillel and Wagenaar (1991).

Despite the important lessons gained from generation-of-randomness research, we regard the perception–judgment studies as more appropriate for revealing subjective concepts of randomness (Wagenaar, 1970a, 1972a). A person could perceive randomness "accurately" and still be unable to reproduce it, just as many people cannot satisfactorily draw a scene that they can recognize as one they have observed (Falk, 1975, 1981). Generation of randomness may confound participants' concepts with various performance variables. We use *perception* and *judgment* interchangeably because we have in mind an intuitive, nonanalytic response concerning the randomness of the experimental stimuli.

In this article, we first summarize the basic findings of and views on generation and perception of randomness. We focus mainly on binary sequences in which the symbol types are equiprobable and on sequential dependencies of the first order (i.e., probabilities conditioned only on one preceding event). Then we describe the algorithmic definition and quantification of randomness by the mathematical theory of complexity, which is based on the length of the most concise description of the sequence. Finally, we report our three experiments that suggest that people's perception of randomness may depend on a similar approach. We construe judging a sequence's degree of randomness as based on a covert act of encoding the sequence. Perceiving randomness may, on this account, be the consequence of failure to encode.

Subjective Randomness Research

Main Findings

The most prominent and consistent finding of the research on both generation and perception of randomness of binary sequences (and two-dimensional [2-D] grids) is that people identify randomness with an excess of alternations between symbol types (Bar-Hillel & Wagenaar, 1991; Budescu, 1987; Falk, 1975, 1981; Lopes & Oden, 1987; Wagenaar, 1970a, 1970b, 1972a, 1972b). Typical random sequences—those containing the modal number of alternations expected by chance—are not perceived by participants as maximally random because the runs appear too long to be random (Gilovich, Vallone, & Tversky, 1985; Wagenaar & Keren, 1988). The sequences that participants produce when they attempt to simulate a random process contain too many short runs, relative to randomness. This bias, often termed *negative recency,* is an expression of the *gambler's fallacy.* Some recent studies (Budescu & Rapoport, 1994; Kareev, 1992; Rapoport & Budescu, 1992) also report overalternations when participants generate randomness under standard instructions, but more randomlike results under modified and more motivating instructions.

These biases characterize people's average responses to judgment and production tasks involving randomness. Substantial individual differences were, however, reported by Budescu (1987) and Falk (1975). A minority of participants usually exhibit a positive-recency bias (see also Wagenaar's [1972a] summary of subjective-randomization results).

Before proceeding to summarize additional findings, we need to define a sequence's degree of alternation. The alternation rate of a binary sequence containing n symbols is greater, the greater the number of runs (r) in the sequence. There are $n - 1$ transitions between successive symbols, and the number of actual changes of symbol is $r - 1$. We characterize every sequence by its *probability of alternation,* $P(A)$, which is computed as $(r - 1)/(n - 1)$. For a binary grid, $P(A)$ is computed similarly by counting changes along horizontal and vertical transitions.

When the frequencies of the two symbol types are equal, the expected $P(A)$ of a random sequence (grid) is close to .5, and deviations of magnitude $\pm d$ from $P(A) = .5$ are about equally probable. The sampling distribution of the statistic $P(A)$ for all sequences (grids) of the same size is thus nearly symmetric around the modal point at .5. This distribution may provide researchers an anchor displaying the behavior of objectively random binary sequences when first-order dependencies are considered.

Likewise, a unimodal and symmetric function is obtained when second-order *entropy* (*EN*) is plotted as a function of $P(A)$, as shown in Figure 1. The computation of second-order *EN,* which is a measure of the sequence's (objective) randomness, is based on the relative frequencies of all ordered pairs (digrams) of symbols. This is a measure of the new information, in bits, contributed by the second member of the pair. Because every member of the sequence (except the first, which is negligible) is the second member of some pair, *EN* may be conceived as the new information (that was not contained in the immediately preceding symbol) added by each symbol in the sequence. Second-order *EN* is maximal (1) when all the four digrams are equiprobable, that is, when $P(A) = .50$ and no (first-order) dependencies exist between successive symbols. It is minimal

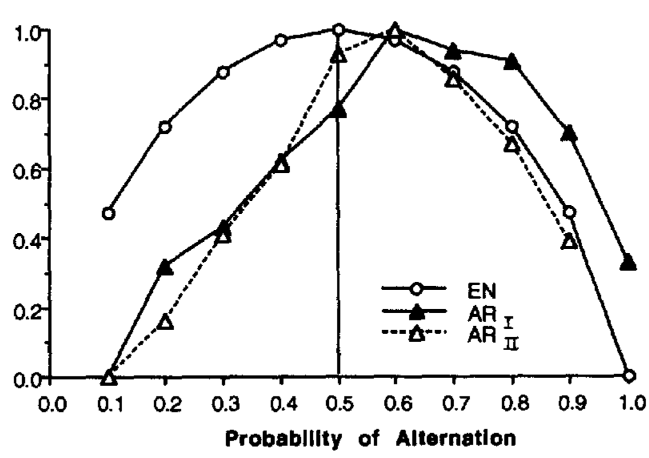

Figure 1. Entropy (*EN*) and linearly transformed mean apparent randomness (*AR*) as functions of probability of alternation. AR_I is the mean ($n = 219$) for sequences, and AR_{II} is the mean ($n = 341$) for grids (based on the results of Falk, 1975).

(0) in the case of perfect alternations. In this case $P(A) = 1.00$, every symbol is completely determined by its predecessor, and the frequencies of the four digrams deviate maximally from equality. An excellent exposition of the rationale and computation of entropy of different orders is given in Attneave (1959, pp. 19–26). Though the general problem of defining and measuring randomness is still unsolved because of the limitation of considering only digrams and ignoring redundancies of higher orders, second-order *EN* affords a workable measure of objective randomness once we focus on dependencies on one preceding symbol.

Falk (1975) obtained participants' randomness ratings (on a scale from 1 to 20) for sequences of length 21 whose $P(A)$s went from .10, .20, .30, through to 1.00 and also for binary grids of 10×10 cells with $P(A)$s varying from .08 to .92. Participants' responses were averaged for each $P(A)$. Figure 1 presents the *apparent randomness* (*AR*) function, which consists of these means linearly transformed to range from 0 to 1 to allow comparison with *EN*. (AR_I denotes apparent randomness for sequences, and AR_{II} denotes that of grids.) As can be seen, the maximum of the two subjective functions is at $P(A) = .60$ instead of at .50 (the mode of *EN*). Sequences and grids with $P(A)$s equally distant from .50 are not judged equally random. *AR* is negatively skewed as a function of $P(A)$. Note, however, that these functions present results averaged over many participants, thus suppressing the noise that characterizes individual responses.

Figure 2 presents three 10×10 grids of 50 white cells and 50 black cells, selected for illustration from 46 grids used by Falk (1975). The grids appear in the same order as their mean rated randomness. The $P(A)$ of Grid B (.51) equals the expected value of such grids (the slight increase over .50 is due to the deviation from Bernoulli process because of the constraint of having 50 cells of each color). Yet, this grid was perceived as less random than Grid C, which has a higher than expected $P(A)$ of .63. As in the case of runs, the contiguous same-color zones that occur by chance make the grid appear too clustered to be random. Feller's (1968, pp. 160–161) classic example of people's illusory beliefs that the pattern of hits during the bombing of London in World War II could not be random is a case in point.

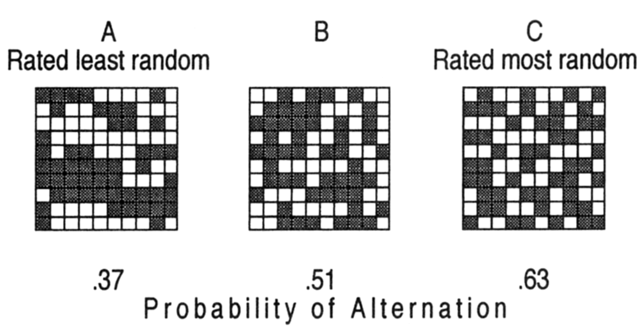

Figure 2. Three grids presented for randomness judgment by Falk (1975), ordered according to their probability of alternation and perceived randomness.

Table 1 displays, for a variety of studies, the mean alternation rate (a) generated under standard instructions and (b) perceived as most random. Whenever necessary and possible, we converted reported summary measures into $P(A)$. As can be seen, the preferred subjective $P(A)$ of about .6 is stable over many studies and experimental variations.

Implications of Biased Randomness Judgment

Cognitive illusions documented in several contexts outside the laboratory match the biases found in studies of people's subjective randomness. For example, Gilovich et al. (1985) studied beliefs concerning sequential dependencies in successive shots in basketball (see also Tversky & Gilovich, 1989a). They found a robust and widely shared belief in the phenomenon of the "hot hand" or streak shooting. Players, coaches, and fans, all believed that once a player makes a basket, his chances of making the next shot increase. However, massive records of individual players in real games analyzed by these authors show that actual sequences of hits and misses are largely compatible with the expected output of a Bernoulli process (see Larkey, Smith, & Kadane, 1989, and Tversky & Gilovich, 1989b, for the hot-hand debate). Apparently, the perception of a hot hand in sequences of basketball shots and of definite clustering in the 2-D distribution of the bombing of London (Feller, 1968) are real-world equivalents of the experimental results in which chance binary sequences and grids are rated as less than maximally random.

In a similar vein, casino gamblers interviewed by Wagenaar and Keren (1988) attributed outcomes of their games not only to chance and skill but to another factor called luck. Good (bad) luck is believed to produce longer streaks of wins (losses) than would be obtained at random. When luck is at work, there is an increase in the conditional probability of winning at the roulette wheel given a previous win. The correspondence between the lay theories of hot hand and luck is striking. Only a substitution of terms is necessary. In both cases, people confronted with random sequences of successes and failures are struck by the subjectively overlong runs. Rather than adjusting their idea of what typically happens by chance, they invoke an idle theory to account for these apparent deviations.

When people invent superfluous explanations because they perceive patterns in random phenomena, they commit what is known in statistical parlance as *Type I error*. The other way of going awry, known as *Type II error*, occurs when one dismisses stimuli showing some regularity as random. The numerous randomization studies in which participants generated too many alternations and viewed this output as random, as well as the judgments of overalternating sets as maximally random in the perception studies, were all instances of Type II error in research results. The error sometimes creeps deviously into the experimental design as well. Alberoni (1962) presented his participants with a supposedly random binary sequence of length 49. Indeed, they unanimously perceived the sequence, whose $P(A)$ was .81, as random. We suspect, however, that Alberoni generated that sequence himself as a "good example" of a random sequence.

Table 1
Mean Alternation Rate, P(A), Generated or Perceived as Most Random in Different Studies

Reference	Randomness task	Size of set	$P(A)$
	Generation		
Bakan (1960)	Sequence	300	.59
Falk[a] (1975)	Sequence (constrained)	40 (20 of each type)	.61
Falk[a] (1975)	2-D grid (constrained)	10 × 10 (50 of each type)	.63
Wiegersma (1986)	Sequence	120	.56[b]
Budescu[c] (1987, Table 2 &	Sequence	20; 40; 60	.59
Table 3)	Sequence	20; 40; 60	.58
Rapoport & Budescu (1992)	Sequence	150	.59
Kareev[a] (1992)	Sequence	10	.61
Budescu & Rapoport (1994, Exhibit 6)	Sequence	150	.58
	Perception		
Wagenaar (1970a)	Select most random	Not reported	.6
Falk[a] (1975)	Rate sequences	21	.60
Falk[a] (1975)	Rate 2-D grids	10 × 10	.6
Wiegersma (1982)	Select most random	Not reported	.65
Diener & Thompson (1985)	Rate sequences	20	.58
Gilovich et al. (1985)	Classify as "chance," "streak," "alternate"	21	.7–.8
Wiegersma (1987, Experiment 1	Select most random	40	.63
Experiment 2		40	.64
Experiment 3)		40	.57[b]

Note. The expected and most probable probability of alternation, $P(A)$, in random productions is .5. Differences in decimal accuracy partly reflect differences in the reported accuracy of the original works (plus occasional transformations by our calculations). 2-D = two-dimensional.
[a] Averaged over different age and sophistication levels.
[b] As read from Figure 1 in Wiegersma (1986, 1987).
[c] Two different estimates based on the same data (of participants exhibiting negative recency).

Explanatory Mechanisms

Many different accounts for the biases of subjective randomness have been proposed. The similarity of people's biases across generation and perception tasks and experimental variations suggests an underlying biased *concept of randomness* (Falk, 1975; Wagenaar, 1970a). This view is supported by Budescu's (1987) results indicating consistent individual styles in randomization and by Falk's (1975) findings of moderate correlations between responses of the same participants to different randomness tasks. Similar responses in the face of varied tasks seem to indicate that the participant's performance is guided by the same underlying image of randomness. In addition, Lisanby and Lockhead (1991) reported that human-generated strings were judged by another group of participants to be random significantly more often than were random-generated strings.

Psychologists are, however, not unanimous in presuming one underlying subjective notion of randomness. Wiegersma (1982, 1986, 1987) found either no correlations or low correlations between participants' performances of different randomness tasks. A class of explanations attributes suboptimality in randomization to variables such as motor tendencies, response sets, boredom, and limited attention (Tune, 1964). See Wagenaar (1970b, 1972b) and Budescu (1987) for a critique of these explanations. Treisman and Faulkner's (1987) suggestion of an internal aleatory generator, affected by the same variables as psychophysical responses, also belongs to the category of explanations of generation results that reject the idea of a faulty concept of randomness. The same is true for Wiegersma's (1982) *control hypothesis,* which attributes repetition avoidance to participants' attempts to overcome perseveration of previous responses, a tendency considered undesirable in speech and everyday situations.

Memory-based accounts are of two apparently opposing types. Kareev (1992) maintains that participants' attempts to produce typical sequences are subject to the limitations of short-term memory. The shorter the subsequence within which the participant tries to balance the two frequencies, the greater the resultant alternation rate. He thus attributes suboptimality in generation of randomness to shortcomings of our memory. According to the other approach, good memory interferes with participants' optimal ability: A certain level of distraction is required to attain maximal independence between responses (Weiss, 1965). Wagenaar (1972b) found that factors that introduce diversion or increase the memory load improve the randomness of productions. Budescu and Rapoport (1994) interfered with their participants' memory by having them compete in a two-person, zero-sum game (where the optimal strategy is to act randomly). These participants produced sequences that more closely approximated randomness than did a control group under standard instructions.

These two views of the role of memory in people's failure to mimic randomness—one blaming the finiteness of our memory (Kareev, 1992) and the other its very existence (Budescu & Rapoport, 1994)—can be subsumed under one contention: The culprit is our intermediate memory span. A creature with infinite memory capacity would achieve perfection by taking into account all of its previous responses, whereas a creature with zero capacity would also produce random responses by starting from scratch on each step, which is the very definition of statistical independence (see also Bar-Hillel & Wagenaar, 1991).

The explanations involving memory cannot, however, be easily applied to perception of randomness. In particular, they cannot be applied to tasks in which participants view several entire sequences simultaneously and compare them with each other, as in Wagenaar (1970a), Wiegersma (1982, 1987), and our experiments reported later in this article. Kareev (1995a) suggests an interesting and elaborate thesis claiming that limited capacity of working memory, together with participants' expectation of typical subsequences, account for the biases observed in perception-of-randomness studies. Typical subsequences, in which the two symbols appear in the same percentages as in the entire sequence, are noted by overalternations, relative to the complete sequence; whereas atypical segments (subsequences) contain excessively long runs. The notorious negative-recency bias occurs because typical segments serve as a norm with which other segments are compared. However, Kareev's experiments, which support his thesis, did not present tasks of perception of randomness of complete sequences but rather probability-learning-type tasks. These, indeed, depend strongly on participants' memory span.

Lopes (1982) and Kareev (1995a) puzzle over the persistence of people's misperceptions of randomness in spite of experience and evolutionary pressures. They wonder whether these biases could be adaptive. People show what Kareev (1995a) calls *positive bias* in their perception of the sequence's degree of serial correlation: They view overalternating sequences (whose serial correlation is negative) as random (zero serial correlation), and they view random sequences as containing streaks that are too long (positive serial correlation). Kareev's analysis reveals that when the binary events in a sequence are not equiprobable (which occurs frequently in the real world), positive correlations are inherently more informative than negative ones. Positive serial correlations may be profitably used to improve sequential prediction at lower absolute values than negative correlations. A tendency toward repetitions may reach the point of allowing practically perfect prediction when the correlation is close to 1. In contrast, excessive alternations cannot attain the minimal serial correlation of -1. Hence, he claims that people's positive bias is a rational predisposition for early detection of a potentially more informative relationship.

Moreover, Kareev (1995b) points out that the sampling distribution of the product-moment correlation (Pearson's r) is negatively skewed when the population correlation is positive, and the more so, the smaller the sample size. It follows that most samples (of positively related variables) one encounters in real life indicate a greater correlation than that in the population. Thus, the limited capacity of working memory may serve as an amplifier that helps people to avoid missing positive relationships.

A compelling account of people's intuitions of randomness is offered by Kahneman and Tversky (1972). To be considered random, a sequence should be representative of its parent population and the process by which it was generated. Thus, a sequence of coin tosses should comprise about equal numbers of heads and tails and display an irregular order. Furthermore, these essential characteristics should be manifest not only globally in the entire sequence, but also locally in each of its parts, thus satisfying what the authors call *local representativeness* (LR). A locally representative sequence contains too many alternations relative to chance. The gambler's fallacy is a corollary of viewing chance as a self-correcting mechanism, which promptly takes care to restore the balance whenever disrupted. Tversky and Kahneman (1971) describe participants as applying the law of large numbers too hastily, as if they believe in "the law of small numbers." The presence of LR in participants' produced sequences and grids (under instructions of randomness) was tested and confirmed by Falk (1975) and was reconfirmed for sequences by Budescu (1987).

Although some testable implications of representativeness have been repeatedly confirmed experimentally by Kahneman and Tversky, and the concept was fruitful in inspiring a host of related studies, there is no established procedure for deducing how the heuristic will be implemented in a specific task. It requires a new interpretation in every context (Kubovy & Gilden, 1991; Teigen, 1983). There is some circularity in LR's description because the irregularity, which is supposed to be manifest even in short segments, is as undefined as randomness.

Although LR is convincing as an account of what many participants actually do when judging and generating random sequences, the concept's predictive power is weak. This heuristic seems to offer little more than an insightful redescription of the phenomenon it purports to explain (Gigerenzer, 1991). Though one can infer from LR that overalternations are expected in subjectively random sequences, the extent of the bias cannot be predicted. The LR heuristic specifies neither how local participants' span of consideration is nor how representative these local subsequences are supposed to be[1] (Falk & Konold, 1994).

In Search of a Subjective Definition of Randomness

Despite insights gained from past attempts to account for people's biases in perception of randomness, the question that interests us most—"What is the basis of the subjective experience of randomness?"—has not been answered satisfactorily. If we understood the processes involved in judging stimuli for randomness, or what implicit definition of randomness participants employ, we might then know why they judge sequences

[1] Promising attempts to delineate participants' span of localness and the type of representativeness they try to implement when generating randomness are reported by Kareev (1992) and Kubovy and Gilden (1991).

as they do. However, this raises again difficulties inherent in defining randomness.

One major source of confusion is the fact that randomness involves two distinct ideas: *process* and *pattern* (Zabell, 1992). It is natural to think of randomness as a process that generates unpredictable outcomes (stochastic process according to Gell-Mann, 1994). Randomness of a process refers to the unpredictability of the individual event in the series (Lopes, 1982). This is what Spencer Brown (1957) calls *primary randomness*. However, one usually determines the randomness of the process by means of its output, which is supposed to be patternless. This kind of randomness refers, by definition, to a sequence. It is labeled *secondary randomness* by Spencer Brown. It requires that all symbol types, as well as all ordered pairs (digrams), ordered triplets (trigrams), . . . , n-grams in the sequence be equiprobable. This definition could be valid for any n only in infinite sequences, and it may be approximated in finite sequences only up to ns much smaller than the sequence's length. The entropy measure of randomness (Attneave, 1959, chaps. 1 and 2) is based on this definition.

These two aspects of randomness are closely related. We ordinarily expect outcomes generated by a random process to be patternless. Most of them are. Conversely, a sequence whose order is random supports the hypothesis that it was generated by a random mechanism, whereas sequences whose order is not random cast doubt on the random nature of the generating process.[2]

Which definition do participants adopt when they assess the randomness of sequences (grids)? Our experience is that a small minority of participants focus on the generating process. These participants rate all the sequences equally, maintaining that all sequences of the same length are equiprobable under randomness. They are right, of course, with respect to ordered sequences.[3] Most participants, however, seem to realize that they do not have access to the unobservable source and must rely on properties of the output.

What do participants attend to in this output? Are they sensitive to the information (entropy) of the sequences? There are some indications in the literature that behavioral variables such as reaction time (RT) apparently covary with the informational content of the stimuli. Hyman (1953) manipulated sequences' entropy in three ways: by changing the number of symbol types (zero order), the probabilities of different symbol types (first order), and by introducing sequential dependency of increased alternation rate between successive symbols (second order). He found that participants' RT in a choice task was linearly related to the sequence's entropy and indifferent to the method by which this measure of information was varied. However, Hyman did not investigate equal degrees of overalternations and overrepetitions.

As previously described, perceived randomness is sensitive to different manipulations of second-order entropy. Sequences (and grids) with equal degrees of overalternations and underalternations are rated very differently (see the *AR* functions in Figure 1). This means that participants' sense of randomness is not based on deviations of n-grams from equiprobability, not even for $n = 2$. To find out whether some other principle may be guiding people's judgments, we turn to a mathematical notion of randomness known as *complexity*.

The Algorithmic Definition of Randomness

Sequences comprising a simple pattern can usually be described concisely. This is not the case with random sequences. Consider, for example, the two following binary sequences of length 21:

a. O X O X O X O X O X O X O X O X O X O X O
b. X O X X X X O O O X O O O O X O X X X O X.

Sequence a can be described by "start with O, then X, and alternate all the way" or by "O X O X . . . ." No comparable condensed description can be given to Sequence b, whose $P(A) = .50$. The meaning of ". . ." would not be clear following the first few characters of Sequence b, which do not allow extrapolation.

The length of the most efficient description that enables reconstruction of a sequence conveys its complexity, a concept that became the focus of interest in recent years in various scientific and technical fields (Li & Vitányi, 1993). A popular exposition can be found in Gell-Mann (1994). The definition and quantification of complexity led to a new conceptualization of randomness of different degrees. It is intuitively clear that patterned sequences can be easily handled in many ways: They are easy to compress, memorize, dictate to a secretary, or copy, whereas all these tasks are much harder with a complex sequence. When typing the two sequences above, we gave Sequence a one quick look and reproduced it successfully in one attempt, whereas Sequence b required careful attention and repeated viewings and was typed in several chunks.

In the 1960s, Kolmogorov (1965), Chaitin (see his 1975 and 1988 articles), and Solomonov (1964) independently defined the complexity of a string by relying on its *minimal description* (Koppel, 1988). The concept rapidly became popular and is known by a variety of names (Gell-Mann, 1994, p. 35; Li & Vitányi, 1993, p. vi). Whereas Gell-Mann prefers *algorithmic information content* and Li and Vitányi *Kolmogorov complexity*, the labels most suitable for our discussion are *algorithmic randomness* or, simply, *complexity*.

The algorithmic randomness of a binary-digit sequence is the bit length of the shortest computer program that can reproduce the sequence. The subjectivity or arbitrariness inherent in the choice of computer or language can be avoided by reference to

[2] But there are exceptions. Seemingly patternless sequences, such as the decimal expansion of π, are sometimes generated by a deterministic process and are therefore utterly predictable (they are considered pseudorandom, see Gács, 1986, and Gell-Mann, 1994, pp. 46–47). On the other hand, primarily random processes may occasionally yield some finite patterned sequences (Falk, 1975; Gell-Mann, 1994, pp. 44–50; Zabell, 1992).

[3] This reminds us of an episode told by Gell-Mann (1994, p. 44). On one of his early visits to the RAND Corporation in Santa Monica, California, he was handed a stack including the "RAND Table of Random Numbers." A small piece of paper fluttered out of it and fell to the floor. When he picked it up, he found it was an errata sheet to the tables.

an idealized computer, or a universal Turing machine (see Koppel, 1988). In fact, the strength of the algorithmic approach lies in the essential invariance of this measure of complexity, despite different hardware and software (Chaitin, 1975; Gács, 1986; Li & Vitányi, 1993).

Although it cannot always be determined easily (Gell-Mann, 1994, pp. 38–41), algorithmic randomness can be viewed as an intrinsic, or objective, attribute of a sequence. A finite string is random when it has maximum complexity. (Infinite sequences having maximum complexity are incalculable by any finite algorithm.) A sequence of maximal complexity cannot be calculated by any algorithm whose bit length is appreciably less than the bit length of the sequence itself. The simplest way to specify a random sequence is to provide a copy of it. Such a sequence is incompressible.

This definition of randomness has strong intuitive appeal because strings that are incompressible must be patternless. A pattern could have been used to reduce the description length (Ford, 1983; Li & Vitányi, 1993). The definition also has the advantage of applying to outputs rather than to processes and to finite and infinite sequences as well. The complexity measure provides quantification of different degrees of randomness on a continuum between the extremes of complete predictability and total randomness, as does entropy. Unlike entropy, however, the complexity definition is not limited to a certain order of dependency. Additionally, complexity is the easier of the two to understand.

Another way to quantify the randomness of a sequence, described in Garner (1970), is by using the size of the set of similar sequences. All sequences of size n, with equal frequencies of Xs and Os, may be sorted into disjoint sets according to their $P(A)$. Sequences within each equivalence set are similar in having the same number of runs, which is linearly related to $P(A)$. The smallest sets comprise the most redundant, or least random, sequences such as OXOX . . . , which belongs to a set of Size 2 because only XOXO . . . has the same $P(A)$. Similarly, there are only two sequences of two runs, corroborating the title of Garner's (1970) article, "Good Patterns Have Few Alternatives." In contrast, the most random sequences, whose $P(A) = .50$, can be made up in many different ways, and their set size is maximal (see also Kubovy & Gilden, 1991; Teigen, 1984). The same picture emerges by considering the sampling distribution of the number of runs (Feller, 1968, p. 62; Siegel, 1956, p. 138). This distribution peaks over the number of runs, which renders $P(A)$ of about .5 and is lowest at the two ends.

However, judging the randomness of a sequence by its location on the sampling distribution of the number of runs is based only on dependencies of the first order, whereas algorithmic randomness captures (in principle) all types of regularities in the sequence. For instance, the sequence OOXXOOXX . . . is clearly patterned to permit a concise description (given a digram, the next symbol is determined). However, $P(A)$ of the sequence is .50, and its number of runs is as expected under randomness. A runs test is "fooled" by this sequence (Mogull, 1994), as is second-order entropy, which is maximal (i.e., 1) for this sequence.

When the complexity measure is used to sort long sequences into random and nonrandom (based on some reasonable cutoff point), the results usually agree with those obtained by means of statistical tests (which are based on a predetermined level of significance). A detailed exposition of the relation between the complexity and the statistical definitions of randomness can be found in Fine (1973). The mathematician Martin-Löf (1966) found that sequences that are incompressible can be shown to possess the various properties of randomness (stochasticity) known from the theory of probability. On the whole, we can satisfactorily identify *incompressibility* with randomness (Ford, 1983; Li & Vitányi, 1993, chap. 2).[4]

The concept of complexity and its algorithmic definition aroused interest beyond their technical use in mathematics and computer science. Physicists have written on complexity (Ford, 1983), and it has recently found its way into popular scientific literature (Gardner, 1989, pp. 169–170; Gell-Mann, 1994; Paulos, 1991, pp. 47–51, 1995, pp. 120–125) as well as into philosophical writings (Dennett, 1991). Recently, Chater (1996) based his argument, that the simplicity and likelihood principles in perceptual organization are equivalent, on mathematical results in complexity theory. Psychologists who study subjective randomness have also described randomness as complexity (Gilmore, 1989; Kahneman & Tversky, 1972; Kubovy & Gilden, 1991; Lopes, 1982). However, they did not investigate the possibility that mathematical complexity's approach could provide an explanation of the way people view randomness intuitively. In what follows, we present our recent attempts to do so. We first describe the background and rationale of this research and briefly review a preliminary experiment. Then we report three new experiments that probe the relationship between apparent randomness and subjective complexity.

Apparent Randomness as Subjective Complexity

We suggest that participants asked to judge the randomness of a sequence attempt to make sense of the sequence in some way. For example, they might implicitly try to encode the sequence before passing judgment. Kahneman and Tversky (1972) raised this possibility. They argued that "random-appearing sequences are those whose verbal description is longest" (p. 436), and that "apparent randomness, therefore, is a form of complexity of structure" (p. 437). However, they did not pursue this psychological hypothesis experimentally.

Subjective complexity, in turn, has been studied by psychologists independently of the study of perceived randomness (e.g., Garner, 1970; Glanzer & Clark, 1962, 1963; Vitz & Todd, 1969). A comprehensive review of theories and behavioral tasks involving patterned sequences can be found in Simon (1972). Tasks in which participants have to deal with a sequence's complexity included judgment of "goodness," verbal description of the sequence, reproduction, memorization, and discrimination

[4] The equality between incompressible and what is ordinarily considered random is, however, not perfect. According to the algorithmic approach, the random-number generator built into most computers is not properly named. It is really a program describable in a few bits (Dennett, 1991), although its productions have the appearance of randomness relative to a certain set of tests (Lopes, 1982).

between sequences. Variables such as accuracy of pattern recall, description length, and rated complexity all correlated considerably with each other, suggesting that diverse measures of subjective complexity are essentially interchangeable.

The Research Hypothesis

Our goal was to test the psychological validity of the complexity approach to randomness. In particular, we tested the hypothesis that randomness judgments are arrived at through some tacit attempt to encode the sequence. A close alternative formulation is that participants tacitly assess the sequence's difficulty of encoding (complexity) in order to judge its randomness.

In our experiments, we measured the *subjective complexity* of binary sequences of the type used for perception of randomness. We correlated these measures with independent ratings of the sequences' randomness. We used several alternative methods to quantify subjective complexity, all based on measuring the *difficulty of encoding* these sequences. Participants were given two different tasks of encoding sequences: either memorizing or copying. In both cases, we obtained measures of their difficulty in performance of these tasks and of their assessment of how difficult the task would be without actually doing it. All these measures were meant to quantify participants' difficulty in encoding a given sequence.

Two different predictions, both based on plausible arguments, can be made. According to the first prediction, difficulty of encoding would be highly correlated with the sequence's entropy (*EN*). Irrespective of the direction of deviation of the sequence's $P(A)$ from .50, overalternating and underalternating sequences whose $P(A)$ is equally distant from .50 are objectively as redundant. Regardless of how they might rate sequences, participants, in principle, should use this redundancy equally well for encoding the two kinds of sequences.

According to the second prediction, difficulty of encoding would be highly correlated with the sequences' apparent randomness (*AR*) and less so with *EN*. This prediction differs from the first one in expecting encoding difficulty to be negatively skewed as a function of $P(A)$ and to peak at values greater than .50. This would obtain if our hypothesis that the judgment of randomness is mediated by an implicit attempt to encode the sequence is true.

While Hyman's (1953) study (reported earlier) seems to support the first prediction, we found in two studies some indications that support the second. Glanzer and Clark (1962) asked participants to decide whether two binary sequences were the same or not. An accuracy score was computed for each sequence and plotted against the number of runs in the sequence (Glanzer & Clark, 1962, Figure 2). Converting number of runs into $P(A)$, and low (high) accuracy scores into high (low) difficulty of encoding, yields a negatively skewed function peaking over $P(A) = .71$. Diener and Thompson (1985) found that in making a decision of whether a given sequence was generated by a random process, RT was longer for "yes" than for "no" responses. This result is compatible with the hypothesis that judging randomness entails some form of processing (encoding) the sequence and that encoding is hardest when randomness is perceived to be maximal (i.e., when the sequences overalternate to some degree).

Preliminary Experiment

In our first small-scale study (Falk & Konold, 1994), we used the same two sets of sequences of length 21 (10 Xs and 11 Os or vice versa) for which Falk (1975) had obtained randomness judgments from 219 participants. A set comprised 10 sequences with $P(A)$s of .10, .20, . . . , 1.00. We chose the task of memorization for measuring subjective complexity. Each of 10 participants viewed in random order all 10 sequences of one set on a computer's screen. Participants were instructed to study each sequence until they could type it out from memory. Memorization time was recorded; typing time was not (see Konold & Falk, 1992, for details of the method).

It turned out that participants failed to utilize the cues, or the redundancy, inherent in some of the overalternating sequences. *Mean memorization time* (*MT*) was maximal for $P(A)$ of .70, and it was negatively skewed as a function of $P(A)$, much like *mean apparent randomness* (*AR*; see Figure 1). The correlation between *AR* (as obtained by Falk, 1975) and *MT* (as obtained by Falk & Konold, 1994, for the same sequences) was .89, whereas between *AR* and *EN* it was .54. These results accord with the second prediction. Perceived randomness was better predicted by the sequence's difficulty of encoding than by its objective degree of randomness. The hypothesis that participants base their judgment of randomness on an implicit attempt to encode the sequence is compatible with the results.

Difficulty of encoding, which we suggest accounts for perceived randomness, is a subjective variable. To find a simple objective way to predict sequences' difficulty of encoding and judged randomness, we adopted a technique developed by Falk (1975). This method assigns each sequence a numerical score based on the number of pure runs and runs of alternations.

Kahneman and Tversky (1972) suggest that when trying to dictate a sequence of binary symbols, one uses shortcut expressions such as "four Xs," or "XO three times." The number of these chunks might provide an index of the difficulty of encoding of the sequence. However, forming the chunk XO three times is probably more difficult than four Xs. The former requires more counting and checking. Therefore, the proposed score, which we call the *difficulty predictor*, assigns double weight to alternating runs. To quantify the subjective complexity of a sequence, one adds twice the number of alternating runs to the number of pure runs. Weighing runs of alternations twice as heavily makes sense considering the fact that the unit repeating itself in these runs is twice as long as in pure runs.

The procedure does not render a unique score for a given sequence, because there are no clear boundaries between pure and alternating runs. For instance, the sequence "X X X O X O" could be assigned a score of 4 (X X X O X O) or 3 (X X X O X O) depending on how the sequence is partitioned. A unique score for the sequence may be obtained if we agree to partition the sequence to achieve the lowest possible number.

Consider a sequence of $P(A) = .20$:

X X X X X X O O O X X O O O O O O O X X X.

The score of this sequence is 5, as obtained by listing five uniform runs. Another sequence,

X X O X O X O X O X O O O X X O X O X O X

whose $P(A) = .80$, is partitioned into five chunks, two of which are doubly weighted because they are runs of alternations. The score of this sequence is 7. This may explain why, although these two sequences deviate equally from the expected $P(A) = .50$, the overalternating one is harder to encode and is also perceived to be more random.

The difficulty predictor was computed retrospectively for all the sequences used by Falk (1975) in an attempt to account for the *AR* function of these sequences. She found a correlation coefficient of .90 between *mean difficulty predictor* (*DP*) and *AR* (as compared with .54 between entropy and *AR*). Prospectively, *DP*, as predictor, yielded a correlation coefficient of .96 with *MT* of these sequences (Falk & Konold, 1994). Thus, whatever strategies participants employ in memorizing the sequences, the difficulty of the task is highly predictable by the above weighted sum.

Overview of the Experiments

The number of participants in the preliminary experiment was small, and only one task (memorization) was used to measure difficulty of encoding. To replicate and extend these findings, we carried out three experiments. The first was an extended replication of Falk and Konold's (1994) pilot study. In addition, we examined the relationship between assessments of difficulty of memorization and apparent randomness in order to test the hypothesis that implicit assessments of encoding difficulty mediate the judgment of randomness. In the second experiment, we used a different operational definition of subjective complexity based on performing a copying task.

Sequences of length 21 were used in the first two experiments. In the third experiment, we extended the inquiry to longer sequences. We also used a more concrete method of eliciting participants' assessments of the difficulty in copying a sequence.

A between-subjects design was used in all the experiments. All the results were summarized at group level and analyzed by using group means. In all cases, we tested the efficacy of our a priori difficulty predictor.

Experiment 1

In this experiment, memorization time measured a sequence's difficulty of encoding (subjective complexity), and participants' ratings assessed the randomness of sequences and their difficulty of encoding.

Method

Materials. Four alternative sets of sequences of length 21 were used (two of the four were the same as the sets used by Falk, 1975, and Falk & Konold, 1994). Each set comprised 10 sequences whose $P(A)$s ranged from .10 to 1.00 in intervals of .10. Half the sequences had 11 Os and 10 Xs, and the other half vice versa (in a counterbalanced design). All sequences began and ended with the character of frequency 11. In this way, there were 10 transitions after Xs and 10 after Os, and the conditional probability of alternation following either X or O was the same as the sequence's total $P(A)$. Moreover, these transition probabilities obtained whether a participant read the sequence from left to right or from right to left (the first language of some of our participants was Hebrew).

Barring the constraints mentioned, the structure of each sequence was randomly determined. We computed the difficulty predictor score for every sequence. The score was 3 when $P(A)$ was .10 or 1.00, and 5 for $P(A)$ of .20. When $P(A)$ was in the range .30 to .90 there was some variation in the scores of sequences of the same $P(A)$ in the four different sets. Linearly transformed mean difficulty predictors (*DP*) are displayed as a function of $P(A)$ in Figure 5, together with the results of Experiments 1 and 2.

For the judgment of randomness, the 10 sequences of every set were printed, in random order, in 10 lines on a one-page form. This allowed participants to view the sequences simultaneously and compare them with each other. Each set of sequences appeared in four different random orders, which resulted in 16 forms. These same 16 sets of sequences, in the same format but with different instructions, served for the task of assessing how difficult a sequence would be to memorize. For the actual memorization task, the sequences were displayed, one at a time, on a computer screen.

Participants. Ninety-seven volunteers participated in the randomness judgment part of the experiment. About the same number of participants received each of the 16 forms. Another 80 participants completed the memorization task; they were equally divided among the four alternative sets of sequences. These participants included high school and college students from the United States (Massachusetts and Minnesota) and Israeli students mostly from the Hebrew University of Jerusalem. Another group of 136 participants assessed sequences' difficulty of memorization. They were roughly evenly distributed among the 16 forms. These were all students (high school, teachers college, and university) from Israel. None of our participants were students majoring in statistics or mathematics.

Procedure. The judgment-of-randomness forms were administered to small groups of participants in a classroom setting. They were instructed to "rate each sequence on a scale of 0 to 10 according to your intuition of how likely it is that such a sequence was obtained by flipping a fair coin." They were advised to inspect all 10 sequences before assigning any rating and then to give a value of 10 to the sequence or sequences most likely and a value of 0 to the sequence or sequences least likely to have been obtained by flipping a coin. Only then was it recommended that they rate the remaining sequences. The procedure and instructions for assessing memorization difficulty were essentially the same. Instead of rating randomness, the participants were asked to rate the difficulty of memorizing the sequence.

For the actual memorization task, participants were individually presented with the sequences on a Macintosh computer. Each sequence was displayed separately on a line numbered from 1 to 21 (Figure 4A). Participants were instructed to study each sequence until they could reproduce it from memory. When a participant was ready, he or she hit the return key. This caused the target sequence to be masked. The participant then typed the response sequence on a numbered line provided on the screen. When finished, the participant again hit the return key. If the response sequence was correct, the participant could go on to the next sequence. If it was incorrect, the display showed the target sequence again and informed the participant where the first error occurred and how many errors were made. Then the participant got another chance to view and then type the sequence. This process could be repeated until the sequence was correctly reproduced. The participants had the option to skip to the next sequence after five failed attempts.

The computer recorded the total time the target sequence was displayed. Participants were informed that time spent typing and editing the sequence was not being recorded and that the objective was to try to minimize their total viewing time.

The order of presentation of the 10 sequences was randomly determined for each participant by the computer program. The 10 experimental sequences were preceded by four practice sequences (from another set) whose $P(A)$s were .20, .90, .50, and .30 (always presented in that order). Participants were not informed that these were practice trials. The experiment lasted about 45 minutes. Completing the task required effort and concentration, but most participants appeared highly motivated by the task.

Results and Discussion

No differences were evident between the results of U.S. and Israeli participants or between college and high school students. We therefore pooled the results across all participants in a given task.

We used the following abbreviations for the three response variables measured in this experiment: *apparent randomness: AR; memorization time: MT;* and *assessed difficulty of memorization: AD.* Technically, these labels designate group means of the respective raw measurements, subject to some additional transformations as explained below.

For every $P(A)$, we computed the mean randomness rating across the four sets and 97 participants and the mean rating of memorization difficulty across the four sets and 136 participants. The 10 *AR* values were obtained by linearly transforming these means so as to range from 0 to 1. The same was done to obtain the 10 *AD*s. This permits comparison of *AR* and *AD* with second-order entropy (*EN*), which is an objective measure of the randomness of these sequences, satisfying $0 \leq EN \leq 1$.

In the memorization task, 19 participants who had skipped at least one sequence were deleted from the analysis, leaving 80 participants who memorized all 10 sequences. Because there were substantial differences among participants in memorization times, the 10 memorization times of every participant were standardized prior to averaging across all participants for each $P(A)$. This procedure gives equal weight to every participant. The 10 *MT* values were obtained by transforming these means linearly to range from 0 to 1.

Figure 3 presents *AR, MT,* and *AD* alongside *EN* as functions of the sequences' $P(A)$. Comparing the three response functions with the sequences' (objective) randomness (*EN*), we see that the response variables peak at $P(A)$s greater than .50, where *EN* is maximal. The two highest points of all three functions occur at $P(A)$s of .60 and .70. As in previous studies, *AR* is a negatively skewed function. Moreover, difficulty of encoding, whether measured by participants' performance (*MT*) or by their assessments (*AD*), behaves very much like *AR*. These results fit the second prediction concerning the difficulty-of-encoding function.

Some incidental performance results replicate the asymmetry of the *MT* function. Participants were not instructed to minimize the number of times they viewed a sequence when memorizing it (they were only instructed to minimize total viewing time). Nevertheless, the number of times a sequence was viewed is another indicator of its difficulty because every viewing beyond the first is a result of an error in reproducing the sequence. Mean number of viewings was negatively skewed as a function of $P(A)$, peaking over .60. In addition, the 48 sequences skipped by the 19 participants deleted from the analysis were similarly distributed as a function of $P(A)$ with a mode of .70. The correlations with *MT* were .96 for mean number of viewings and .93 for number of skipped sequences.

Table 2 presents the correlations between all the pairs of variables involved in this experiment and in Experiment 2. The pattern of the correlations clearly shows that perceived randomness (*AR*) is correlated more strongly with memorization difficulty (*MT* and *AD*) than with the sequence's objective degree of randomness (*EN*). Furthermore, the difficulty predictor (*DP*) is highly correlated with the response variables (*MT, AD,* and *AR*). The respective correlations of entropy (*EN*) with these response variables are lower.

Difficulty of encoding (whether measured by performance or by assessments) thus predicts perceived randomness better than does the sequence's degree of randomness. This result is compatible with the hypothesis that implicit encoding (or an assessment of the difficulty thereof) mediates the judgment of a sequence's randomness.

It should be noted that participants had to concentrate and persist to memorize some of the sequences of length 21. It may seem that a participant would take advantage of any kind of regularity in the stimuli to succeed in such a demanding task. The first prediction, that participants' encoding difficulty would closely follow the informational content of the sequences, was based on such a premise. Yet, participants were oblivious to slight-to-moderate degrees of deviations from randomness in overalternating sequences. Moreover, these sequences were even harder for them to memorize than were the most random sequences. This was a real performance difficulty. No judgment of randomness or of local representativeness in the sequence could have been involved in performing this task. Although the LR heuristic could have been used to judge the sequences'

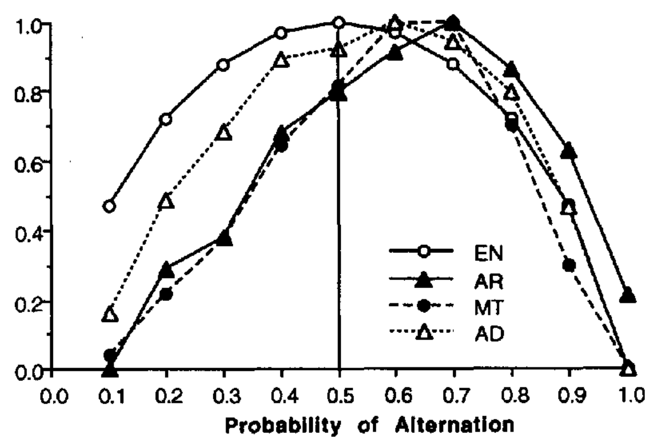

Figure 3. Entropy (*EN*) and linearly transformed means of apparent randomness (*AR*; $n = 97$), memorization time (*MT*; $n = 80$), and assessed difficulty (*AD*; $n = 136$), as functions of the sequence's probability of alternation (Experiment 1).

Table 2
Intercorrelations Between Entropy, Difficulty Predictor, and Response Variables: Experiments 1 and 2

Variable	*DP*	*MT*	*AD*	*CD*	*AR*
EN: entropy; determined by *P*(*A*)	.81	.79	.92	.82	.63
DP: difficulty predictor; computed, based on the actual sequences	—	.99	.97	.97	.95
MT: memorization time; standardized; mean of 80 participants; Experiment 1		—	.94	.91	.94
AD: assessed difficulty of memorization; mean of 136 participants; Experiment 1			—	.96	.87
CD: copying difficulty; standardized no. of chunks plus standardized copying time; mean of 80 participants; Experiment 2				—	.94
AR: apparent randomness; mean rating of 97 participants; Experiment 1 and 2					—

Note. All variables are means, computed for each probability of alternation, *P*(*A*), across four alternative sets of sequences of length 21.

randomness, the similarity between difficulty of memorization and *AR* supports the underlying mechanism of tacit encoding.

Experiment 2

In this experiment, we compared the randomness ratings of the 97 participants obtained in Experiment 1 with the difficulty of other participants in copying these sequences. The copying could be conducted in stages by breaking the sequence into chunks to ease the encoding task. This might presumably give the participant a better chance of spotting helpful cues in the sequences and thus perform more in accord with the sequences' informational content (*EN*).

Method

Materials. The same four sets of sequences used in Experiment 1 were used for the copying task.

Participants. Eighty volunteer high school and college students from Massachusetts performed the copying task. Each set of sequences was given to 20 participants.

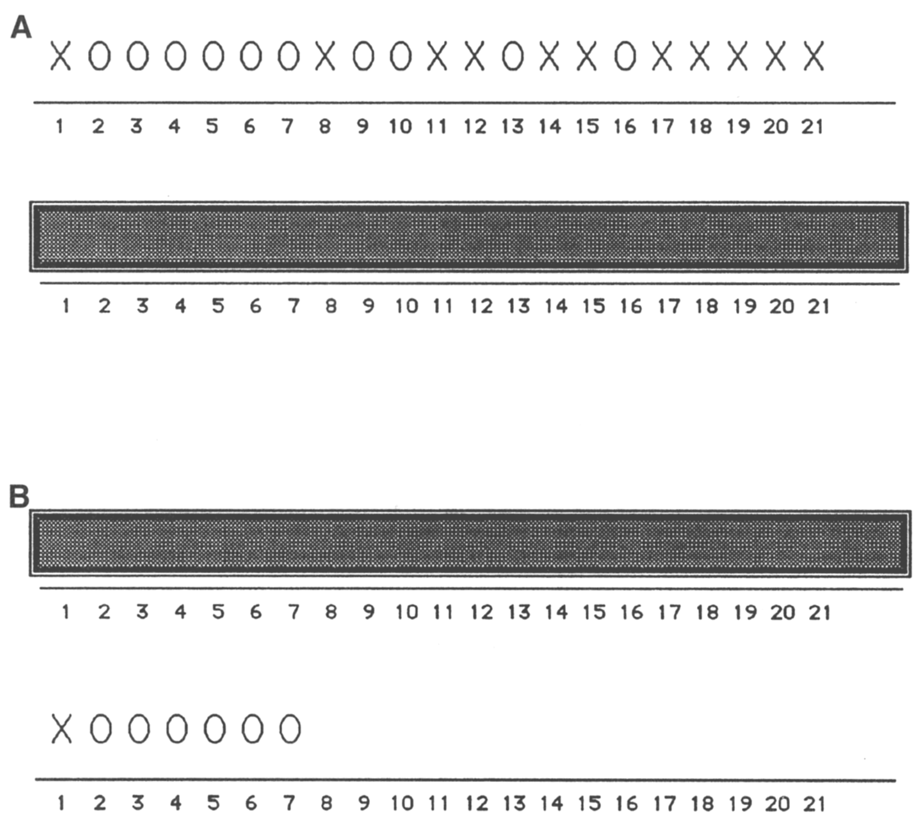

Figure 4. A: Screen display of the target sequence in Experiments 1 and 2, prior to masking. B: Screen display for Experiment 2, in which the target sequence is masked (top) while the participant enters the first chunk.

Procedure. Every participant performed the copying task individually. Each sequence was displayed separately on the computer, as in Experiment 1 (see Figure 4A). Participants were instructed to look at each sequence carefully and copy it making no errors while trying to do so as efficiently as possible. Pressing the return key caused the target sequence to be masked. The participant then started typing the sequence as far as they remembered on a numbered line provided on the screen just below the target sequence. Figure 4B shows the first chunk typed by a participant while the target sequence is masked. After entering as much of the sequence as they remembered, the participant hit the return key.

If the first chunk was copied correctly, the rest of the sequence, beginning just to the right of the copied chunk, would reappear. The participant would then proceed to learn and copy another segment of the sequence. The process continued until the entire sequence was correctly copied. If there was a mistake in any chunk, a note would appear on the screen. That chunk would reappear for extra viewing and retyping until copied correctly. A practice sequence of $P(A) = .20$ was used to demonstrate the task. Participants were told that time spent viewing the target sequence was recorded by the computer, whereas time spent copying the sequence was not. They were instructed to minimize total viewing time and the number of chunks without making errors.

Following the first practice trial, three additional practice sequences whose $P(A)$s were, in turn, .90, .50, and .30 were presented without informing the participants that these were practice trials. These were followed by the 10 experimental trials. The order of presentation of the 10 sequences was randomly determined for each participant. All the participants completed the task, which lasted about 15 minutes.

Results

Many participants copied highly patterned sequences completely in one chunk. In this case, the task reduced to the memorization task of Experiment 1, and viewing time tended to be short. In the case of harder sequences, most participants accomplished the task in several chunks, and the overall viewing time tended to be longer. However, the association between number of chunks and viewing duration is far from perfect. The correlation coefficient between the two sets of 800 raw scores (80 participants, each responding to 10 sequences) is .12. After standardizing the 10 scores of each variable for every participant, this correlation becomes .44 (individual correlations range from .02 to .90) indicating that each of the two variables contributes some new information. Indeed, there is possibly some trade-off between these variables. We, therefore, used a composite measure of both variables to quantify a sequence's copying difficulty.

For every participant, we standardized the 10 viewing times and the 10 numbers of chunks. Then we added the two standard scores to obtain a difficulty score for each sequence. These difficulty scores were averaged across the 80 participants for each $P(A)$. The function of *copying difficulty* (CD) was obtained by linearly transforming these means to range from 0 to 1. Figure 5 presents CD as a function of the sequence's $P(A)$ alongside the difficulty predictor (DP) of the same sequence and AR and MT of Experiment 1.

In places, it is not easy to distinguish between the curves of the different functions in Figure 5. These four curves, which display means of group responses (or of several sequences' scores), are very similar to each other. Indeed, CD as well as MT—the measure of encoding difficulty in Experiment 1—behave much like the apparent-randomness function, as does the (a priori computed) DP. The two highest points of all four functions correspond to $P(A)$s of .60 and .70. The vertical line at $P(A) = .50$ (the axis of symmetry of the objective-randomness function) highlights the negative asymmetry of these functions.

In addition, as can be seen in Table 2, CD is highly correlated with DP and with AR. The results of Experiment 2 thus replicate the findings of Experiment 1: Difficulty of encoding, operationally defined by CD in the present experiment, is largely predicted by the DP variable, and it deviates from objective randomness in the same direction and to the same degree as does apparent randomness. Copying the sequence in parts does not seem to reduce the bias characterizing the one-stage memorization task of Experiment 1.

Experiment 3

In this experiment, we asked participants to assess the difficulty of copying a sequence by breaking their assessments into stages. Conceivably, this more analytic procedure increases the chance of discovering statistical regularities. To extend the generality of our findings, we used longer sequences and obtained randomness ratings for them as well.

Method

Materials. Eight new alternative sets of sequences of length 41 were created at random, subject to the following constraints. As before, a set comprised 10 sequences of $P(A)$s ranging from .10 to 1.00. Half the sequences had 20 Xs and 21 Os and the other half vice versa, the character of frequency 21 always being the first and last in the sequence. All these features were counterbalanced with respect to each other. We computed the difficulty-predictor score (number of pure runs plus twice the number of alternating runs) for each sequence and averaged these scores across the eight sets for each $P(A)$. The function DP was obtained

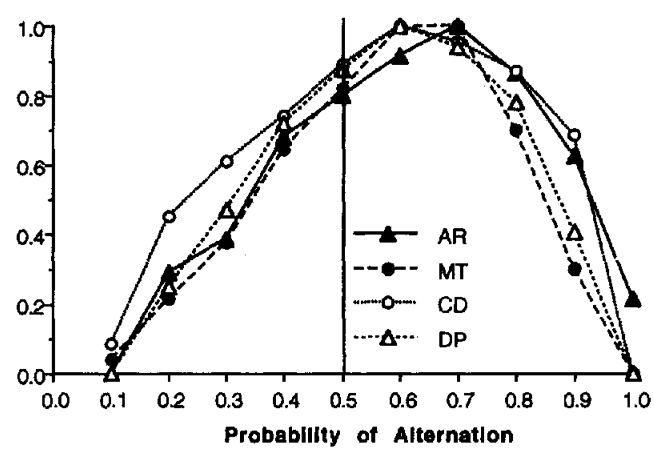

Figure 5. Linearly transformed means of difficulty predictor (DP; across 4 sets), copying difficulty (CD; $n = 80$), and apparent randomness (AR) and memorization time (MT; repeated from Figure 3), as functions of the sequence's probability of alternation (Experiments 1 and 2).

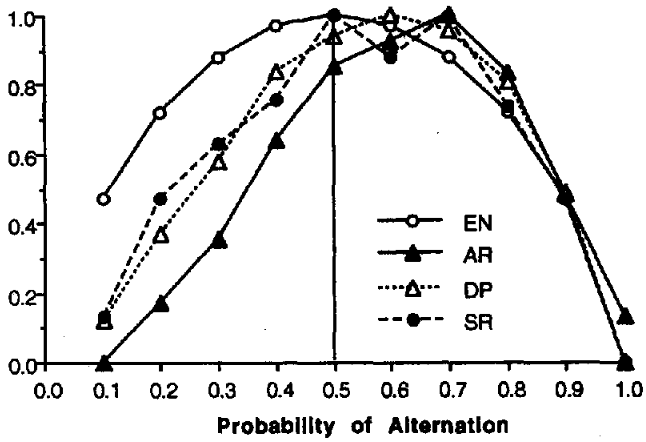

Figure 6. Entropy (*EN*) and linearly transformed means of difficulty predictor (*DP*; across 8 sets), apparent randomness (*AR*; n = 175), and segmentation and rating (*SR*; n = 60 or 58), as functions of the sequence's probability of alternation (Experiment 3).

by linearly transforming these 10 means to range from 0 to 1 (see Figure 6).

For the judgment of randomness, the 10 sequences of every set were printed on one form, as in Experiment 1. Each set appeared in two different random orders, thus yielding 16 different forms. For the assessment-of-copying-difficulty task, the sequences of each of the eight sets were split into two halves, yielding 16 different forms. One half contained sequences with odd $P(A)$s (.10, .30, . . . , .90) and the other contained even $P(A)$s. This was done to reduce the time required to complete the task. The five sequences of each half were printed spaciously, in a random order, below the task instructions. A second page with additional instructions was stapled to each form.

Participants. Randomness ratings were obtained from 175 volunteer, undergraduate biology students from the Hebrew University of Jerusalem. About the same number of participants responded to each of the 16 forms. Another 118 volunteer participants from Israel of a range of ages and variety of occupations participated in the difficulty-assessment task. They were divided fairly evenly among the 16 forms. No statisticians, mathematicians, or students majoring in these areas were included. Sixty participants responded to forms with odd $P(A)$s and 58 to even $P(A)$s.

Procedure. Participants responded to the randomness-judgment forms in groups in a classroom setting. The instructions were as in Experiment 1 except that the scale ranged from 1 to 10 instead of from 0 to 10. The assessment-of-copying-difficulty task was administered individually to each participant. The instructions were:

> Below are 5 sequences of 41 characters each. Imagine that you have to copy the sequences on the back side of the page. Usually, such sequences are copied in chunks. Try to divide each sequence as you would if you were to copy it in the fastest and most efficient way. On one hand, the number of segments (no. of page turnings) should be minimized, and on the other hand, one has to remember them as accurately as possible. Mark your division by perpendicular lines that cut the sequence into chunks.

When the participants finished the task of dividing the sequences, they turned to the next page where they were instructed:

> Now, please go back to the sequences that you have divided. Write *underneath each segment* a number from 1 to 3 signifying how difficult it would be to copy (1 = easy; 2 = medium; 3 = difficult).

Table 3
Intercorrelations Between Entropy, Difficulty Predictor, and Response Variables: Experiment 3

Variable	*DP*	*SR*	*AR*
EN: entropy	.87	.90	.68
DP: difficulty predictor; computed on actual sequences	—	.98	.95
SR: segmentation and rating; mean of 60 or 58 participants[a]		—	.91
AR: apparent randomness; mean rating of 175 participants			—

Note. All variables are means, computed for each probability of alternation, $P(A)$, across eight alternative sets of sequences of length 41.
[a] Sixty participants received odd (.10, .30, . . . , .90) $P(A)$s, and 58 received even $P(A)$s.

Results

As before, randomness ratings were averaged across all participants (175) and sets (8) for each $P(A)$. The function *AR* in Figure 6 shows these 10 means transformed linearly to [0, 1]. In the difficulty-assessment task, we first summed the ratings of all segments to obtain a difficulty score for each sequence as rated by every participant. These scores were averaged across all participants for each $P(A)$. The *segmentation and rating* (*SR*) function was obtained by linearly transforming these means to [0, 1].

Figure 6 presents *AR, SR, DP* (the sequence's computed difficulty predictor), and the objective randomness index *EN* as functions of $P(A)$. Clearly, the negatively skewed *AR* function is replicated for sequences of length 41 (the two highest points are again over .60 and .70). The *SR* function, although peaking over .50, is also negatively skewed (its second-highest point is at .70). Apparent randomness (*AR*) is better approximated by the assessed difficulty of encoding, *SR*, than it is by *EN* (see also correlations in Table 3). As in Experiment 1, *DP* is highly

Table 4
Some Features of Three Studies of the Apparent-Randomness (AR) Function

Study's feature	Source: Falk (1975)	Present research: Experiment 1	Present research: Experiment 3
N	219	97	175
Sequence length	21	21	41
Alternative sets	2	4	8
Rating scale	1–20	0–10	1–10
Procedure of rating	Sequential[a]	Simultaneous[b]	Simultaneous[b]

[a] Sequences (shown one by one) comprised cards of two colors. Judgment required: likelihood that the deck of cards had been well shuffled.
[b] See *Method* section for Experiments 1 and 3.

predictive of the assessed-difficulty measure (*SR*) and of judged randomness (*AR*).

General Discussion

The Apparent-Randomness Function

Three studies (Falk, 1975; and our Experiments 1 and 3) independently obtained the complete mean-rated-randomness function of sequences of equiprobable binary symbols for a wide range of *P*(*A*)s. The experiments differed in several respects including the length of sequences, rating scale, and format of sequence presentation (see Table 4). Despite these variations, the results of the studies are remarkably uniform. The three functions have the same negatively skewed shape (see the *AR* functions in Figures 1, 3, and 6). The pairwise correlation coefficients between the *AR* values of the three studies are .96, .98, and .99. These coefficients are exceptionally high, even when allowing for the inflation of correlations when means are used (Freedman, Pisani, Purves, & Adhikari, 1991, pp. 140–141). The high correlations attest to the robustness and prevalence of the bias toward alternations in the perception of randomness.

We computed weighted means of *AR*s of the three studies for each *P*(*A*) and linearity transformed the results to obtain the aggregate variable *AR*, which ranges from 0 to 1. Table 5 shows these aggregated *AR* values ($N = 491$) along with second-order entropy (*EN*) of the same sequences. The correlation coefficient between *AR* and *EN* is only .62.

One should bear in mind that the *AR* function in Table 5 is averaged over many participants. Individual participants' functions vary considerably (Budescu, 1987; Wagenaar 1972b). Still, this function is universal in the sense of holding, on the average, for varied groups of participants (American, Israeli, different ages, educational levels, etc.).

Our studies are restricted to binary sequences in which each alternative appears about equally often. Future research should generalize the apparent-randomness results and the findings concerning encoding difficulty to two or three dimensions and to more than two symbols in different relative frequencies.

Table 5
Summary: Apparent Randomness (AR) and Entropy (EN) as Functions of the Sequence's Probability of Alternation, P(A)

P(*A*)	*AR*	*EN*
.1	0	.47
.2	.27	.72
.3	.41	.88
.4	.66	.97
.5	.83	1.00
.6	.99	.97
.7	1.00	.88
.8	.90	.72
.9	.62	.47
1.0	.24	0

Note. *AR* values are (linearly transformed) weighted means of results of three studies (Experiments 1 and 3 and Falk [1975]; $N = 491$).

The Difficulty-of-Encoding Function

The results regarding the peak of the apparent-randomness function substantiate a phenomenon previously observed by psychologists. The results concerning the different functions (*MT*, *CD*, *AD*, and *SR*) that measure *difficulty of encoding* (*DE*) are more surprising. We refer to these four variables collectively as *DE*. In all three experiments, *DE* is negatively skewed as a function of *P*(*A*) regardless of task and of whether the measure of difficulty is based on performance or on assessment.

Results of the performance tasks, in particular, cannot easily be explained as artifacts of task interpretation. When participants must efficiently and accurately transcribe a sequence, whether by memorizing it as a whole or by copying it in parts, they probably take advantage of any sequence regularities they notice. It is therefore surprising that one type of regularity (overrepetitions) is readily utilized, whereas the same degree of regularity of the complementary type (overalternations) is apparently not detected and seems even to interfere with performance (see Kareev, 1995a).

That *DE* would not closely follow the informational content (*EN*) of the sequence was not a priori self-evident. On the contrary, the alternative hypothesis—that *DE* would be maximal for maximally random sequences and that it would decrease to the same extent when sequences depart from randomness to the same degree—is compelling. This makes all the more instructive the finding that *DE* (in its different manifestations) is negatively skewed as a function of *P*(*A*). To quote Dawes (1992), "Unless we can reject something that we previously believed to be true, we have not learned anything new" (p. 2).

It seems that the mathematician's view of randomness as maximal complexity captures people's intuitive approach to this concept. This can be interpreted as a reprieve, of sorts, of the soundness of intuitive thinking. The high similarity found between the *DE* and *AR* functions suggests that participants equate randomness with difficulty in encoding a sequence. It is possible that they attempt some kind of instantaneous (perhaps preverbal) encoding and use the difficulty of that attempt to decide the randomness of a sequence. Their point of maximal difficulty is, however, misplaced relative to the mathematician's point of maximal complexity. Kahneman and Tversky (1972) were thus right in contending that apparent randomness is a form of complexity, that is, if one refers to subjective complexity. The latter, as measured by encoding difficulty, differs from the objective, or mathematical concept in a systematic way, just as apparent randomness differs from entropy.

A scrutiny of Tables 2 and 3 reveals that the various *DE* functions are somewhat closer to *EN* than is *AR*. This suggests that perceived randomness is somewhat more biased toward alternations than difficulty of encoding would lead us to believe. Our anticipation that actual and perceived encoding-difficulty measures would serve as equivalent indexes of subjective complexity is confirmed by the high correlations between either *MT* or *CD* and *AD*. Performance *DE* variables (*MT* and *CD*) are perhaps slightly better predictors of *AR* than are assessed *DE* variables (*AD* and *SR*). We do not understand the reason for this difference, which is negligible relative to the high correlations of all four *DE* variables with *AR* (they range from .87 to .94).

Implicit Encoding Versus Judgment of Representativeness

Both the hypothesis that participants tacitly encode a sequence to evaluate its randomness and the hypothesis that they judge its local representativeness (LR) for making that judgment are compatible with the obtained *AR* results. Our experiments do not enable us to decide between these two accounts of people's perception of randomness.[5] The crux of the problem is that locally representative sequences are the most difficult to encode. It is not impossible that judgments of LR involve implicit encoding as well. In that case, LR and encoding difficulty are not alternative accounts but rather two aspects of the same phenomenon.

The LR heuristic and the mechanism of implicit encoding both predict the direction of deviation from perfect randomness. The hypothesis of implicit encoding, however, has the experimental advantage that *DE*, measured independently, succeeds in parametrically predicting the peak of perceived randomness and the shape of the function. It is hard to imagine a task in which participants could be instructed to judge sequences for LR, without reference to randomness. It would be difficult in such a case to explain what it is one is supposed to judge, that is, what is meant by *local* and of what the judged stimuli are supposed to be representative.

Quantifying Encoding Difficulty and Apparent Randomness

Our results, and those of other experimenters, indicate that sequences with overalternations are perceived as more random than their entropy and algorithmic randomness warrant. From a mathematical point of view, a run of alternations is as redundant as a uniform run. Both allow perfect prediction within the boundaries of the run. Psychologically, however, they appear not to be equivalent. These two kinds of deviations from randomness weigh differently both in judging how random a sequence is and in the difficulty in encoding it. Thus, one should not treat overalternations and underalternations equally when trying to predict human perceived randomness or encoding difficulty.

Our difficulty predictor (*DP*) takes only first-order dependencies into consideration and assigns runs of alternations twice the weight of pure runs. This measure seems to fit our experimental sequences, which were created by manipulating only first-order dependencies.[6] The double weight for alternating runs is a rough commonsensical guess of the increased cost in terms of difficulty due to the double length of the recurring unit in such runs.

DP was first computed (in Falk, 1975) as a post hoc attempt to account for the *AR* function. It was then cross-validated by Falk and Konold (1994) and was found to be highly correlated with memorization time as well. In the present research, *DP* was highly predictive of judgments of randomness and encoding difficulty. The success of *DP* as an a priori predictor was replicated in all three experiments[7] (see Figures 5 and 6 and Tables 2 and 3).

Like other variables in our analysis, *DP* is averaged over several alternative sequences for every $P(A)$. It remains to be established to what extent variations in *DP*, for a constant $P(A)$, affect participants' responses. Although our study was not designed to answer this question, our results suggest that measures of encoding difficulty (such as *MT* and *CD*) follow *DP* somewhat more closely than does apparent randomness.

Concluding Comments

The subjective experience of encoding difficulty seems to account for the perception of completed sequences. It was not shown to describe the way in which participants attempt to generate random sequences. However, our results, combined with those of generation of randomness, including the evidence that human-generated sequences are virtually indistinguishable from those judged by participants as being most random (see also Table 1), suggest that people generate just the kind of sequences that they perceive to be random. If one assumed that people are capable of reproducing their image of randomness (an assumption that we doubted on a priori grounds in the introduction), one could speculate that during the generation process, participants evaluate their productions by the subjective encoding-difficulty criterion. For example, a long uniform run would stand out as too easy to memorize or to communicate. The same, only to a lesser degree, would be true for a long run of alternations. Both would therefore be relatively avoided. Long same-symbol runs, in particular, would not be tolerated. Applying the LR criterion would obviously bring about similar results.

Repetitions, even those occurring by chance, are readily utilized for encoding a sequence and are perceived as violations of randomness, whereas streaks of alternations have to be much more pronounced to be put to use and induce the impression of nonrandomness. "More of the same" seems to be the easiest to encode. Conceivably, change is harder to encode and is more suggestive of randomness[8] than is perpetuation of the status

[5] See Diener and Thompson's (1985) interpretation of their results in favor of the model of perceiving randomness after eliminating all tenable alternative hypotheses, as opposed to the model of judging by representativeness.

[6] There are several indications in the literature that pure and alternating runs capture most of participants' attention when dealing with binary sequences. Vitz and Todd (1969) reported that the most common coded elements in human processing of sequential binary events were runs of like events and runs of alternations. A close scrutiny of the accent points chosen by Garner's (1970) participants as a start for organizing a melody of binary tones that they heard played cyclically, shows that they preferred the beginning and the first note after either a long pure run or a long alternating run. Budescu's (1987) results also supported the first-order Markovian model for the random series-generation task.

[7] Despite the predictability of participants' performance afforded by *DP*, we could not confirm the use of pure and alternating runs as encoding units by inspecting the chunks that participants formed in the copying task of Experiment 2 and in the segmentation-and-rating task of Experiment 3. Participants cut mostly larger segments, lumping together runs of both kinds.

[8] It seems as if people's mental anchor for randomness is the state of perfect alternations. They sense, however, that this could not be right because it is too regular, so they disrupt it a little. But their adjustment is insufficient, much like that of Tversky and Kahneman's (1974) participants whose anchor was determined experimentally.

quo. Indeed, only change of outcomes can make the sequence locally representative. Yet, excessive alternations do help an encoder by affording a long (doubly weighted) encoding chunk, thereby decreasing the value of *DP*. This explains why perfect or near-perfect alternations are judged as nonrandom. The double difficulty in encoding an alternating chunk also suggests an explanation of why subjective complexity differs from objective complexity and why the former is negatively skewed as a function of $P(A)$.

Diener and Thompson (1985) describe the evolvement of the perception of randomness after eliminating all tenable alternative hypotheses for the sequence's production. Their description shares an important feature with our tacit-encoding hypothesis: The judgment of randomness results from the repeated failure of attempts to find regularities (confirm hypotheses; detect encoding shortcuts). The idea that subjective randomness is built on our failure to make sense of the world is not new. Piaget and Inhelder (1951/1975) attribute the emergence of the idea of chance to children's realization of the impossibility of predicting oncoming events or of offering causal explanations. In the same vein, Alberoni (1962) showed that university students resorted to chance only after admitting their failure to discover a regular pattern in the stimuli they were shown. To use Holden's (1985) words in his article on chaos, "To call this irregularity 'randomness' is a confession of ignorance: it is not an answer" (p. 15).

Our interpretation of the experimental results is that people tacitly endorse the same principle as in the mathematical approach to randomness and apply the criterion of compressibility. When data fail to meet this criterion, they are considered random. Participants' application of this criterion is, however, biased relative to mathematical prescription. In one sense, though, these misperceptions of randomness could be considered valid and perhaps even adaptive. People demonstrate an accurate intuition for which sequences are hardest for them to encode. Their biased judgment of randomness would prove beneficial if it spared them futile attempts to encode information they would find too complex to handle.

The view of perceived randomness as a result of tacit encoding is compatible with a picture of mind that grounds cognition in action in the world. Action is a key feature in Piaget's and others' developmental theories of cognition. A host of data indicate that in early development, cognition and physical activity cannot be easily disentangled. In an attempt to mark the limitations of the representational-computational view of mind, Shanon (1993) suggests that the principles of operation underlying language, memory, perception, reasoning, and problem solving are akin to those used negotiating the world and manipulating objects. "Even when confined to the internal domain, cognitive activity may be carried out through the simulation of action in the theater of one's mind" (Shanon, 1993, p. 268).

In analogy to Piaget's operations, which are conceived as internalized actions, perceived randomness might emerge from hypothetical action, that is, from a thought experiment in which one describes, predicts, or abbreviates the sequence. The harder the task in such a thought experiment, the more random the sequence is judged to be. Several accounts of tasks involving sequences (Glanzer & Clark, 1962, 1963; Vitz & Todd, 1969) stress the primacy of covert verbalization. Perceptual organization of the sequence and an assessment of its randomness would depend, according to these accounts, on the length of a silent verbalization.

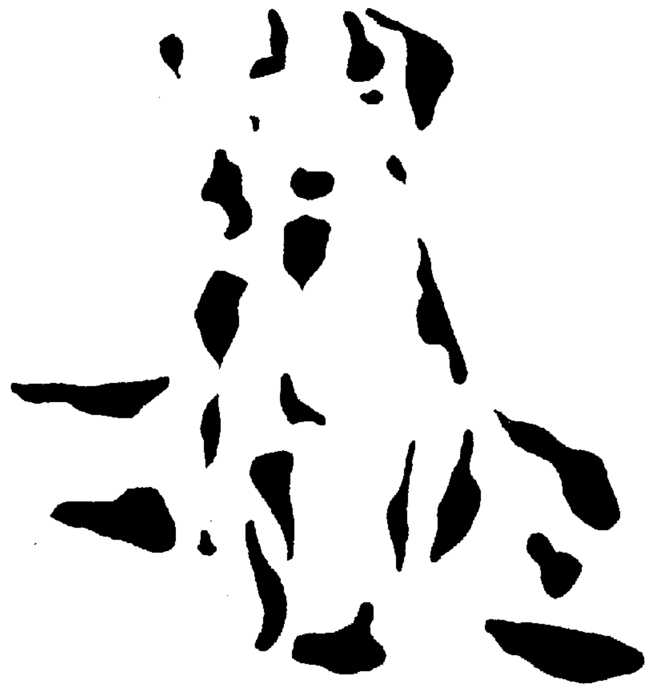

Figure 7. Random blobs or a familiar figure? (Based on Street, 1931).

The boundaries between tacit verbal encoding and (the Gestalt) perceptual organization are, however, not always clear. In Figure 7, the initial apparent randomness of the blobs dissipates once the pattern of a dog is noticed. Pattern recognition is closely connected, in this case, with reducing the length of the verbal description to one word.

Examples abound of the interconnectedness of seeing the underlying structure (i.e., removing randomness) and hitting upon an efficient encoding.[9] Learning a foreign language is one such instance; forming a scientific hypothesis is often another. In 1946, E. B. Lewis set out to investigate mutations in a cluster of genes that affect wing development in *Drosophila* flies: Some mutations transformed the normally degenerate second wing pair to appear more winglike, others caused degeneration of the wings proper, and still others caused the development of winglike and leglike structures on abdominal segments of the flies. More and more difficult-to-relate details accumulated, and accounts of his observations became longer and more cumbersome from decade to decade (reaching a length of 31 pages in Lewis, 1967). Finally, in Lewis (1978), he came forward with a mere six-page article that parsimoniously presented a model of the bithorax complex organization. Once a pattern has been recognized, the description of the same phenomena can be considerably condensed.

[9] We thank David Gilden for suggesting the dog-pattern and foreign-language examples.

References

Alberoni, F. (1962). Contribution to the study of subjective probability: I. *Journal of General Psychology, 66,* 241–264.

Attneave, F. (1959). *Applications of information theory to psychology: A summary of basic concepts, methods, and results.* New York: Holt, Rinehart & Winston.

Ayer, A. J. (1965, October). Chance. *Scientific American, 213,* 44–54.

Ayton, P., Hunt, A. J., & Wright, G. (1989). Psychological conceptions of randomness. *Journal of Behavioral Decision Making, 2,* 221–238.

Ayton, P., Hunt, A. J., & Wright, G. (1991). Randomness and reality: Reply. *Journal of Behavioral Decision Making, 4,* 222–226.

Bakan, P. (1960). Response tendencies in attempts to generate random binary series. *American Journal of Psychology, 73,* 127–131.

Bar-Hillel, M., & Wagenaar, W. A. (1991). The perception of randomness. *Advances in Applied Mathematics, 12,* 428–454.

Budescu, D. V. (1987). A Markov model for generation of random binary sequences. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 25–39.

Budescu, D. V., & Rapoport, A. (1994). Subjective randomization in one- and two-person games. *Journal of Behavioral Decision Making, 7,* 261–278.

Chaitin, G. J. (1975, May). Randomness and mathematical proof. *Scientific American, 232,* 47–52.

Chaitin, G. J. (1988, July). Randomness in arithmetic. *Scientific American, 259,* 80–85.

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review, 103,* 566–581.

Cohen, J. (1960). *Chance, skill, and luck: The psychology of guessing and gambling.* Baltimore, MD: Penguin Books.

Dawes, R. M. (1992). The importance of alternative hypotheses—and hypothetical counterfactuals in general—in social science. *The General Psychologist, 28,* 2–7.

Dennett, D. C. (1991). Real patterns. *The Journal of Philosophy, 88,* 27–51.

Diener, D., & Thompson, W. B. (1985). Recognizing randomness. *American Journal of Psychology, 98,* 433–447.

Falk, R. (1975). *Perception of randomness.* Unpublished doctoral dissertation (in Hebrew, with English abstract), Hebrew University, Jerusalem, Israel.

Falk, R. (1981). The perception of randomness. *In Proceedings of the Fifth International Conference for the Psychology of Mathematics Education* (Vol. 1, pp. 222–229). Grenoble, France: Laboratoire I.M.A.G.

Falk, R. (1991). Randomness—an ill-defined but much needed concept. *Journal of Behavioral Decision Making, 4,* 215–218.

Falk, R., & Konold, C. (1994). Random means hard to digest. *Focus on Learning Problems in Mathematics, 16,* 2–12.

Feller, W. (1968). *An introduction to probability theory and its applications* (3rd ed., Vol. 1). New York: Wiley.

Fine, T. L. (1973). *Theories of probability: An examination of foundations.* New York: Academic Press.

Ford, J. (1983, April). How random is a coin toss? *Physics Today, 36,* 40–47.

Freedman, D., Pisani, R., Purves, R., & Adhikari, A. (1991). *Statistics* (2nd ed.). New York: Norton.

Gács, P. (1986). Randomness and probability—complexity of description. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences* (Vol. 7, pp. 551–555). New York: Wiley.

Gardner, M. (1989). *Mathematical carnival* (Rev. ed.). Washington, DC: Mathematical Association of America.

Garner, W. R. (1970, January–February). Good patterns have few alternatives. *American Scientist, 58,* 34–42.

Gell-Mann, M. (1994). *The quark and the jaguar: Adventures in the simple and the complex.* New York: Freeman.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases." *European Review of Social Psychology, 2,* 83–115.

Gilmore, J. B. (1989). Randomness and the search for psi. *Journal of Parapsychology, 53,* 309–340.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17,* 295–314.

Glanzer, M., & Clark, W. H. (1962). Accuracy of perceptual recall: An analysis of organization. *Journal of Verbal Learning and Verbal Behavior, 1,* 289–299.

Glanzer, M., & Clark, W. H. (1963). The verbal loop hypothesis: Binary numbers. *Journal of Verbal Learning and Verbal Behavior, 2,* 301–309.

Hake, H. W. (1955). The perception of frequency of occurrence and the development of expectancy in human experimental subjects. In H. Quastler (Ed.), *Information theory in psychology: Problems and methods* (pp. 257–277). Glencoe, NY: Free Press.

Hake, H. W., & Hyman, R. (1953). Perception of the statistical structure of a random series of binary symbols. *Journal of Experimental Psychology, 45,* 64–74.

Holden, A. (1985, April 25). Chaos is no longer a dirty word. *New Scientist, 106,* 12–15.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology, 45,* 188–196.

Jarvik, M. E. (1951). Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *Journal of Experimental Psychology, 41,* 291–297.

Kac, M. (1983, July–August). What is random? *American Scientist, 71,* 405–406.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3,* 430–454.

Kareev, Y. (1992). Not that bad after all: Generation of random sequences. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1189–1194.

Kareev, Y. (1995a). Positive bias in the perception of covariation. *Psychological Review, 102,* 490–502.

Kareev, Y. (1995b). Through a narrow window: Working memory capacity and the detection of covariation. *Cognition, 56,* 263–269.

Kolmogorov, A. N. (1965). Three approaches to the quantitative definition of information. *Problems in Information Transmission, 1,* 1–7.

Konold, C., & Falk, R. (1992). Encoding difficulty: A psychological basis for "misperceptions" of randomness. *Proceedings of the Sixteenth Annual Meeting of the International Group for the Psychology of Mathematics Education, 2,* 9–16.

Koppel, M. (1988). Structure. In R. Herken (Ed.), *The universal Turing machine: A half-century survey* (pp. 435–452). Hamburg, Germany: Kammerer & Unverzagt.

Kubovy, M., & Gilden, D. (1991). Apparent randomness is not always the complement of apparent order. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure* (pp. 115–127). Washington, DC: American Psychological Association.

Larkey, P. D., Smith, R. A., & Kadane, J. B. (1989). It's okay to believe in the "hot hand." *Chance: New Directions for Statistics and Computing, 2*(4), 22–30.

Lewis, E. B. (1967). Genes and gene complexes. In R. A. Brink (Ed.), *Heritage from Mendel* (pp. 17–47). Madison: University of Wisconsin Press.

Lewis, E. B. (1978). A gene complex controlling segmentation in *Drosophila. Nature, 276,* 565–570.

Li, M., & Vitányi, P. (1993). *An introduction to Kolmogorov complexity and its applications.* New York: Springer.

Lisanby, S. H., & Lockhead, G. R. (1991). Subjective randomness, aesthetics, and structure. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure* (pp. 97–114). Washington, DC: American Psychological Association.

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

Lopes, L. L., & Oden, G. C. (1987). Distinguishing between random and nonrandom events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 392–400.

Martin-Löf, P. (1966). The definition of random sequences. *Information and Control, 9,* 602–619.

Mogull, R. G. (1994). The one-sample runs test: A category of exception. *Journal of Educational and Behavioral Statistics, 19,* 296–303.

Neuringer, A. (1986). Can people behave "randomly?": The role of feedback. *Journal of Experimental Psychology: General, 115,* 62–75.

Nicks, D. C. (1959). Prediction of sequential two-choice decisions from event runs. *Journal of Experimental Psychology, 57,* 105–114.

Paulos, J. A. (1991). *Beyond numeracy: Ruminations of a numbers man.* New York: Knopf.

Paulos, J. A. (1995). *A mathematician reads the newspaper.* New York: Basic Books.

Peterson, C. (1980). Recognition of noncontingency. *Journal of Personality and Social Psychology, 38,* 727–734.

Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children* (L. Leake, Jr., P. Burrell, & H. D. Fishbein, Trans.). New York: Norton. (Original work published 1951)

Pollatsek, A., & Konold, C. (1991). Randomness is well enough understood to be misunderstood. *Journal of Behavioral Decision Making, 4,* 218–220.

Rapoport, A., & Budescu, D. V. (1992). Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology: General, 121,* 352–363.

Ross, B. M., & Levy, N. (1958). Patterned predictions of chance events by children and adults. *Psychological Reports, 4,* 87–124.

Shanon, B. (1993). *The representational and the presentational: An essay on cognition and the study of mind.* New York: Harvester Wheatsheaf.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.

Simon, H. A. (1972). Complexity and the representation of patterned sequences of symbols. *Psychological Review, 79,* 369–382.

Solomonov, R. J. (1964). A formal theory of inductive inference. *Information and Control, 7,* 1–22.

Spencer Brown, G. (1957). *Probability and scientific inference.* London: Longmans Green.

Street, R. F. (1931). *A Gestalt completion test: A study of a cross section of intellect.* New York: Bureau of Publication, Columbia University.

Teigen, K. H. (1983). Studies in subjective probability: I. Prediction of random events. *Scandinavian Journal of Psychology, 24,* 13–25.

Teigen, K. H. (1984). Studies in subjective probability: V. Chance vs. structure in visual patterns. *Scandinavian Journal of Psychology, 25,* 315–323.

Treisman, M., & Faulkner, A. (1987). Generation of random sequences by human subjects: Cognitive operations or psychophysical process? *Journal of Experimental Psychology: General, 116,* 337–355.

Tune, G. S. (1964). Response preferences: A review of some relevant literature. *Psychological Bulletin, 61,* 286–302.

Tversky, A., & Gilovich, T. (1989a). The cold facts about the "hot hand" in basketball. *Chance: New Directions for Statistics and Computing, 2*(1), 16–21.

Tversky, A., & Gilovich, T. (1989b). The "hot hand": Statistical reality or cognitive illusion? *Chance: New Directions for Statistics and Computing, 2*(4), 31–34.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76,* 105–110.

Tversky, A., & Kahneman, D. (1974, September 27). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

Vitz, P. C., & Todd, T. C. (1969). A coded element model of the perceptual processing of sequential stimuli. *Psychological Review, 76,* 433–449.

Wagenaar, W. A. (1970a). Appreciation of conditional probabilities in binary sequences. *Acta Psychologica, 34,* 348–356.

Wagenaar, W. A. (1970b). Subjective randomness and the capacity to generate information. *Acta Psychologica, 33,* 233–242.

Wagenaar, W. A. (1972a). Generation of random sequences by human subjects: A critical survey of literature. *Psychological Bulletin, 77,* 65–72.

Wagenaar, W. A. (1972b). *Sequential response bias: A study on choice and chance.* Unpublished doctoral dissertation, Leiden University, Leiden, The Netherlands.

Wagenaar, W. A. (1991). Randomness and randomizers: Maybe the problem is not so big. *Journal of Behavioral Decision Making, 4,* 220–222.

Wagenaar, W. A., & Keren, G. B. (1988). Chance and luck are not the same. *Journal of Behavioral Decision Making, 1,* 65–75.

Weiss, R. L. (1965). "Variables that influence random-generation": An alternative hypothesis. *Perceptual and Motor Skills, 20,* 307–310.

Wiegersma, S. (1982). Can repetition avoidance in randomization be explained by randomness concepts? *Psychological Research, 44,* 189–198.

Wiegersma, S. (1986). Avoidance of repetition in produced sequences and the concept of randomness. *Perceptual and Motor Skills, 62,* 163–168.

Wiegersma, S. (1987). The effects of visual conspicuousness and the concept of randomness on the recognition of randomness in sequences. *The Journal of General Psychology, 114,* 157–165.

Zabell, S. L. (1992). The quest for randomness and its statistical applications. In F. Gordon & S. Gordon (Eds.), *Statistics for the twenty-first century* (pp. 139–150). Washington, DC: Mathematical Association of America.

Received July 7, 1995
Revision received June 25, 1996
Accepted June 26, 1996 ■