

**From the holdings of the Department of Special Collections , Princeton University Library,
Princeton, NJ**

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyright law. We do not charge any permission or usage fees for the publication of images of material in our collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff. For more information regarding copyright, credit, and citation please review the information on the Copyright, Credit and Citation Guidelines page of our website.

<https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines>. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. <https://library.princeton.edu/special-collections/ask-us>

Doctoral Dissertations, 2011-2026

Psychology

2013

Randomness Cognition: Linking Perception and Conception

Zhao, Jiaying

Princeton, NJ : Princeton University

<http://arks.princeton.edu/ark:/88435/dsp01mg74qm18h>

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. <https://www.copyright.gov/title17/title17.pdf>

**RANDOMNESS COGNITION:
LINKING PERCEPTION AND CONCEPTION**

Jiaying Zhao

A DISSERTATION PRESENTED TO THE
FACULTY OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF PSYCHOLOGY

Adviser: Daniel Osherson, Ph.D.

JUNE 2013

© Copyright by Jiaying Zhao, 2013. All Rights Reserved.

Abstract

The cognition of randomness consists of perceptual and conceptual components. One might be able to discriminate random from non-random stimuli, yet be unable to identify which is which. We show that the ability to distinguish random from non-random stimuli is superior to the ability to correctly identify which given stimuli are “random.” We further evaluate the encoding hypothesis according to which the tendency of a stimulus to be labeled “random” varies with the cognitive difficulty of encoding it. For some stimuli, the difficulty of encoding fails to predict the probability of being labeled “random,” providing evidence against one version of the encoding hypothesis (Chapter 2). Using different methods, we next replicate these findings. Moreover, we observe that the ability to distinguish repeating stimuli from random is reliably better than the ability to distinguish alternating stimuli from random, presumably due to the ease of encoding of repeating stimuli compared to alternating (Chapter 3). In addition, we examine the ability to produce stimuli as random, and find that the produced sequences are more biased towards alternations than sequences labeled as random (Chapter 4). Finally, we provide a set of experimental paradigms measuring the sensitivity to randomness without invoking the notion of randomness or related concepts (Chapter 5). Overall, we present a comprehensive investigation of randomness cognition and findings suggesting a gap between the perception and the conception of randomness.

Acknowledgements

This dissertation documents the research on randomness cognition which I have been pursuing over the last few years. None of the work reported here has been published. A portion of the work (Chapter 3) has been presented at the Cognitive Science Conference.

I am tremendously indebted to my advisor Dan Osherson. You have provided constant guidance, support, and encouragement over the last five years, and contributed key ideas and insights to the work report here. Beyond the bounds of this dissertation, you have sharpened my thinking as a researcher and made me a better and happier person in many ways throughout graduate school. I'm immensely grateful for how available you always are, for your prompt feedback, and for the humorous chats that sparked numerous ideas. And don't worry, I'll get your jokes eventually.

I also owe a great intellectual debt to Eldar Shafir. Our work on poverty is probably the most challenging, stimulating, and fulfilling project I've ever pursued. You have constantly expanded my horizon and pushed me to new limits. To illustrate, running two participants on two computers at once is not easy, even in the lab. You requested to run four at a time in the shopping mall in order to collect more data. I laughed and cried. In the end, I managed to run eight participants on eight computers at the same time in the mall. I will keep pushing and reaching new heights after my time at Princeton.

I'm deeply thankful for the help and advice from Nick Turk-Browne. I have learned so much from our collaborations. You have inspired me to be a productive researcher, seeking fundamental connections between different areas in cognition. I will always remember our multiple 5+ hours marathon writing sessions, which are directly responsible for my back problems.

Moreover, I'd like to thank Ulrike Hahn who helped to improve the ideas and the analyses in Chapter 2. Thanks also go to John Myles White, a good friend and collaborator, who worked on an fMRI experiment on randomness perception with me.

Finally, I'd like to thank my wonderful friends here at Princeton, in Dublin, and in China, who have provided me with their endless heart-warming support over the years. And my family, mom and dad, I dedicate this work to you with love. You have always believed in me. Everything not owed to someone else, I owe to you.

Contents

Abstract	iii
Acknowledgements	iv
Chapter 1 Introduction	1
1.1 What is randomness?	2
1.2 Randomness cognition	3
1.3 Theoretical explanations	5
1.4 Current investigation	9
Chapter 2 Perception and identification of randomness	11
2.1 Methodological overview	12
2.2 Experiment 1: Random vs. non-random matrices	14
2.2.1 Method	14
2.2.2 Results and discussion	16
2.2.3 Robustness of results	19
2.2.4 Feedback	19
2.2.5 An ideal observer analysis	21
2.3 Experiment 2: Random vs. non-random walks	22
2.3.1 Method	23
2.3.2 Results and discussion	24
2.3.2 Robustness: Rating task	27
2.4 Experiment 3: The encoding hypothesis	29
2.4.1 Method	31
2.4.2 Results and discussion	33
2.5 Chapter summary	34

Chapter 3 Replications using different methods	36
3.1 Methodological overview.....	37
3.2 Experiment 4: Distinguishing random vs. non-random matrices.....	38
3.2.1 Method	38
3.2.2 Results and discussion.....	40
3.2.3 Identifying random matrices	41
3.3 Experiment 5: Distinguishing random vs. non-random walks.....	42
3.3.1 Method	42
3.3.2 Results and discussion.....	44
3.4 Experiment 6: Classifying random vs. non-random matrices	44
3.4.1 Method	45
3.4.2 Results and discussion.....	46
3.5 Experiment 7: Encoding random vs. non-random matrices.....	46
3.5.1 Method	46
3.5.2 Results and discussion.....	48
3.6 Chapter summary	49
 Chapter 4 Human production of random stimuli	 50
4.1 Experiment 8: Producing random sequences	50
4.1.1 Method	50
4.1.2 Results and discussion.....	51
4.1.3 Comparing one's own productions with random sequences	52
4.2 Experiment 9: Linking discrimination, identification, and production.....	53
4.2.1 Method	53
4.2.2 Results and discussion.....	53
4.3 Chapter summary	54
 Chapter 5 Implicit measures of randomness	 56
5.1 Experiment 10: Predicting random vs. non-random bits.....	57
5.1.1 Method	57

5.1.2	Results and discussion.....	58
5.2	Experiment 11: Time estimation of random vs. non-random matrices.....	59
5.2.1	Method	60
5.2.2	Results and discussion.....	61
5.3	Experiment 12: Scene perception with random vs. non-random masks	63
5.3.1	Method	64
5.3.2	Results and discussion.....	65
5.4	Chapter summary	66
Chapter 6	Concluding remarks	67
6.1	Summary of results.....	67
6.2	General discussion.....	69
6.3	Conclusion.....	71

Chapter 1

Introduction

How people understand randomness has been a question of longstanding interest in psychology. The ability to apprehend randomness is fundamental to cognition since it involves noticing structure, or the lack thereof, of events in the world (Bar-Hillel & Wagenaar, 1991). Such ability underlies a variety of cognitive processes, including basic perception of textures and visual patterns (e.g., Julesz, 1981; Landy & Graham, 2004), and high-level processes of decision-making and judgments of chance (e.g., Edwards, 1961; Gilovich et al., 1985; Tversky & Kahneman, 1971).

Psychological literature on randomness has predominantly focused on beliefs or subjective intuitions about randomness. In most studies, participants are often instructed to assess the degree of perceived randomness of a given bit string (i.e., sequence), or to produce a sequence as if it is generated by a fair coin. Before reviewing the findings on randomness cognition, the definition of randomness should be first discussed.

1.1 What is randomness?

There is little agreement among philosophers and mathematicians as to what defines randomness. It is argued that randomness should be conceived as a property of the underlying processes, not of products, and that random processes cannot be observed directly (Lopes, 1982). Thus, to say something is random we can only sample the output from the process and test whether it has properties of random sequences. When asked to assess the randomness of a sequence, people tend to rely on the properties of the sequence produced by a random process.

Randomness typically invokes concepts such as equiprobability, sequential independence, unpredictability, irregularity, and incompressibility (see review by Nickerson, 2002). Equiprobability refers to the idea that in the process that selects repeatedly (with replacement) from a finite set, every member of the set has the same chance of being selected. This is not a sufficient condition for randomness because equiprobability can be obtained by a non-random process that samples in a highly regular and deterministic way. Sequential independence means that each item is selected independently from any other items. It implies that the preceding output cannot provide a basis for subsequent selections. This is related to the idea of unpredictability, which is that the knowledge of any portion of a sequence does not provide any information about other portions. Brown (1957) notes that both individual items and groups of individual items in a random series are unpredictable. Irregularity means that a random sequence lacks any comprehensible pattern, and is therefore indistinguishable from another random sequence. Irregularity is neither necessary nor sufficient for randomness because a random process can sometimes produce outputs that are highly regular, and a non-random process can produce a seemingly irregular sequence.

A final property of randomness is incompressibility: A sequence is said to be random if it cannot be compressed. That is, the description of the sequence in an intuitively reasonable programming language cannot be shorter than the sequence itself (Chaitin, 1975; Kolmogorov, 1965; Li & Vitányi, 2008). Importantly, the compressibility of a given bit string depends on the programming language at issue; shortest programs for generating a given string within different languages have the same length only up to a finite but arbitrary constant. It is also worth recalling a well-known fact about standard programming languages like JAVA, C, or Python (Blum, 1967). In such a language, specifying a minimal length program for generating a given finite bit string is an uncomputable task in a strong sense (see Osherson & Weinstein, 2011, §3 and §8 for discussion and proof).

1.2 Randomness cognition

Randomness has been argued to be important to both humans and non-human animals for at least two reasons: First, common survival goals (finding food, avoiding predators, selecting mates) are facilitated by the ability to identify patterns and relevant structure in the world. Correctly identifying randomness, as the absence of structure (Beltrami, 1999), is the flipside to that ability. Mistaking random events as non-random can lead to various cognitive biases and distortions (Gilovich et al., 1985; Tversky & Kahneman, 1971), which may lead to suboptimal behaviors — as illustrated in gambling problems (e.g., Michalczuk et al., 2011; Reuter et al., 2005; Steeves et al., 2009; Terrell, 1998; Toneatto et al., 1997; Wagenaar, 1988), as well as financial losses in the stock market (e.g., Johnson & Tellis, 2005). Second, there may be contexts in which it is important to be able to generate random, unpredictable responses in order to evade

predators or thwart competition in the game theoretic context (Camerer, 2003; Dorris & Glimcher, 2004; Glimcher, 2005; Rapoport & Budescu, 1992).

Given the importance of randomness, in past research, however, the verdict on people's understanding of randomness has tended to be negative. Reichenbach (1934/1949) speculated that people untrained in probability would be unable to generate random sequences and would produce series with too many alternations. This speculation was later confirmed by Tune (1964) and Wagenaar (1972). In general, people tend to associate randomness with too many alternations and short runs of repetitions (Kahneman & Tversky, 1972, Wagenaar, 1972; but see Nickerson & Butler, 2009; for reviews see Bar-Hillel & Wagenaar, 1991; Oskarsson et al., 2009).

In randomness perception studies, people tend to perceive genuine random sequences to be non-random because the runs appear too long to be random (Gilovich et al., 1985). Rather, they judge over-alternating sequences as maximally random (Falk, 1975, 1981). To quantify randomness, Falk (1975) developed a measure of switch rate defined as the ratio between the number of switches in the sequence and the length of the sequence minus 1. It was found that the sequences that were judged to be most random contained a switch rate around .6 (Falk, 1975; Wagenaar, 1970; Lopes & Oden, 1987).

In randomness production experiments, people often generate series with too many alternations, a bias termed negative recency (Kahneman & Tversky, 1972). They also tend to balance the relative frequencies of responses according to the required probabilities in small samples.

Both perception and production paradigms rely on the subjective conception of randomness: In perceptual tasks, people are asked to judge how random a sequence or a pattern

looks; and in production tasks, they are asked to generate a sequence as if it is produced by a fair coin. It is argued that the perception paradigm is more appropriate for examining the subjective conceptions of randomness (Falk & Konold, 1997), since a person could perceive randomness accurately but are unable to produce randomness. An advantage of the perception paradigm is that the experimenter can systematically manipulate statistical characteristics of the sequence and determine the point of subjective randomness (Falk, 1981; Wagenaar, 1970).

1.3 Theoretical explanations

There are a number of explanations proposed to account for people's performance in the randomness perception and production tasks. One view is that human generated sequences reflect their intuitive and subjective concept of randomness. The produced sequences do not match the probabilistic concept of randomness because the subjective concept of randomness is faulty (Chapanis, 1953; Ross & Levy, 1958). The faulty concept explanation cannot predict or explain why people associate randomness with more alternations (Wagenaar, 1972). An interesting question that remains is whether the faulty concept of randomness is due to an inability to perceive randomness (this account will be tested in the next chapter).

One explanation draws on the functional limitations of the cognitive system. Tune (1964) attributes the inability to randomize to memory limitations. When producing a random series, people tend to remember the frequencies of relative events in the set. Since people can hold only limited items in working memory, they cannot produce a random sequence. On the other hand, Weiss (1964) argues that poor performance is not due to limitations in memory, but rather it is because people try to remember past sequences. Since people often fail to maintain the necessary level of distraction, they tend to generate responses that are related rather than independent.

While the memory explanation offers a reasonable account of performance in production tasks, it cannot be easily applied to the findings in perception studies.

A compelling account is offered by Kahneman and Tversky (1972) who suggest that people generally believe characteristics of randomness (e.g. equiprobability) should manifest not only globally in the sequence but also locally in parts of the sequence. As a result, people tend to balance the relative frequencies of the responses in local samples according to the required probabilities. Thus, a locally representative sequence will contain more alternations than expected by chance.

In addition, people may hold a mistaken belief that random sources provide an equilibrium process, whereby equiprobable outcomes must “balance out” and hence runs must self-correct, thus giving rise to the gambler’s fallacy. Such a belief could stem from a failure to distinguish between sampling with replacement and sampling without replacement (e.g., Ayton et al., 1989; Bar-Hillel & Wagenaar, 1991). When sampling without replacement, each repeat of an outcome makes a switch more likely, due to depletion, thus biasing against longer runs. It has been argued not only that people may have more experience of sampling without replacement in the real world, but Hahn and Warren (2009) have demonstrated recently that, given a likely model of how people experience random sequences, sampling with and without replacement are surprisingly difficult to distinguish on the basis of experience alone.

Specifically, Hahn and Warren (2009) demonstrated how short-term memory limitations may impact both perception and generation of random sequences. While all possible sequences of length n are equally likely as outcomes of n flips of a coin, they are not equally likely as (local) subsequences within a longer (global) sequence: If one starts flipping a coin, the average number of coin tosses one has to wait before encountering the sequence HHHH is considerably

longer than the average wait time for the sequence THHH. Given human short-term memory limitations, the actual experience of unfolding sequences will be akin to a fixed-length sliding window moving through the overall data stream, both in sequence production and perception. The local subsequences that appear in that moving window differ in how they occur. Moreover, the corresponding experience of sampling with and without replacement is remarkably similar.

Furthermore, Kareev (1992) demonstrated that short sequences have higher alternation rates than do long sequences. Kareev (1992) used this fact as the basis for explaining over-alternation in random sequence generation tasks: An agent with limited short-term memory will effectively be generating a long sequence of “random outputs” by stringing together short sequences which can be held in working memory (see Rapoport and Budescu (1997) for one way of modeling such a process), and consequently will exhibit over-alternations for long sequences. In other words, people may look like they exhibit the gambler’s fallacy even when they do not, purely due to information processing limitations. In support, Kareev (1992) found correlations between alternation rates in sequence generation and working memory capacity. Hahn and Warren (2009) argued that the same fact may underlie over-alternation bias in sequence judgment tasks on the assumption that people are seeking to match their experience. If experience is based on statistics compiled over a short, sliding window, then over-alternations will characterize that experience.

Another prominent model is provided by Falk and Konold (1997) based on compression and algorithmic complexity. The more structure or regularity a sequence contains, the more compactly it can be represented (e.g., 1111111111 as 10 1’s), and hence, the greater the compression afforded relative to enumeration of the sequence itself. This forms the basis of formal definitions of randomness as incompressibility. For $x \neq .5$, a $\text{switch}(x)$ sequence will often

be more compressible than a fully random string [switch(.5)], in the sense of being generated by shorter programs in an intuitively reasonable programming language (Li and Vitányi, 2008).¹

This can also be used to motivate a psychological process account of randomness judgment. Specifically, Falk and Konold (1997) propose the encoding hypothesis that the tendency for a given pattern to be classified as “random” varies with the cognitive difficulty the judge experiences in encoding the pattern. Sequences with greater internal structure, such as repetition or symmetry will be easier to encode, and thus perceived as less random. It is also these properties that make sequences more compressible, a connection that has been drawn out in several areas of cognition (see Chater, 1999; van der Helm & Leeuwenberg, 1996). Falk and Konold (1997) also provided experimental evidence for this account. In their studies they found a direct relation between the time participants needed in order to memorize a given bit string and participants’ judgment of “how likely it is that such a sequence was obtained by flipping a fair coin.”

Finally, Neuringer (1986) attributes people’s suboptimal performance to the lack of skill, arguing that people do not understand all the requirements of randomness and have no experience in dealing with random series. In the experiment, when given feedback from statistical descriptors participants learned to generate sequences that were indistinguishable from computer generated random sequences. This account is problematic since it cannot explain results in the perception paradigm.

As Wagenaar (1972) notes, there are several difficulties in theorizing about subjective conceptions of randomness. It is almost impossible to compare and synthesize the various studies due to the vast methodological differences. These differences vary in the number of alternatives

¹ There is no guarantee, however, that such differences in compressibility can be detected or put to use by human observers.

in the choice set (from 2 to 26), the nature of alternatives (letters or digits), the number of responses (from 32 to 2520), the mode of production (calling out or writing down), self-paced or not, and whether previous responses were available. Measures of randomness also vary tremendously. In addition, researchers have employed not only a range of judgment tasks, but also prediction tasks (Edwards, 1961), memory tasks (Olivola & Oppenheimer, 2008), and competitive games (Rapoport & Budescu, 1992). Thus, there is no canonical way of examining people's understanding of randomness.

Research on randomness has been plagued by conceptual difficulties, which have often undermined empirical results (Ayton et al., 1989; Lopes, 1982; Nickerson, 2002). Inspecting a pattern does not allow its qualification as random or non-random: A random source could produce a thousand 0's in a row and a non-random source might "look random" by chance. Indeed, all strings of the same length have identical probability of being randomly generated (but see Hahn & Warren, 2009). In fact, making sense of the evidential grounds for a given stimulus being randomly produced is a surprisingly subtle affair (Beltrami, 1999; Fitelson & Osherson, 2012). These conceptual difficulties raise questions about what kind of intuitive knowledge people are expected to possess, and how that conceptual knowledge can be studied empirically.

1.4 Current investigation

As discussed above, previous research on randomness has focused on two types of processes: randomness perception and randomness generation, with some consistency of results between the two (Bar-Hillel & Wagenaar, 1991). Both processes rely on the subjective conception of randomness. What has been given little attention in either of these contexts, however, is the fact that there are potentially different ways in which randomness cognition can be examined. In

particular, it seems important to distinguish a conceptual understanding of randomness from the ability to respond appropriately to random sources in everyday life.

The ability to apprehend the randomness of events includes both perceptual and conceptual components. Perceptually, people have some ability to distinguish random from non-random stimuli. Conceptually, they may be capable of identifying which stimulus looks “random” — by correctly recognizing which of two stimuli is randomly generated, where only one is. These are separate abilities because one might be able to discriminate random from non-random stimuli, yet systematically misidentify the random stimulus, in the same way that one might be able to distinguish different hues, yet be mistaken about their labels. The primary goal of the present work is to examine the relationship between the perceptual and conceptual sides of randomness cognition.

Chapter 2

Perception and identification of randomness

The goal of the current chapter is to examine randomness cognition by comparing the ability to distinguish random from non-random stimuli to the accuracy with which given stimuli are identified as “random.” In addition, the encoding hypothesis is evaluated, which states that the tendency of a stimulus to be labeled “random” varies with the cognitive difficulty of encoding it.

The theoretical definition of randomness has been difficult (Beltrami, 1999; Nickerson, 2002), because genuine random generating sources can give rise to outputs that do not look random (e.g., uniform runs), and non-random sources can generate outputs that look random. Our terminology is faithful to a “process” rather than “product” conception of randomness (see Eagle, 2012; Earman, 1986 for discussion). In our usage, a “random” stimulus (or pattern) is an object that has been produced by a random process. Non-random stimuli are defined as productions from a distorted random source. Importantly, our instructions to participants in the tasks examining their perception of randomness make no reference to the terms random or randomness. In probing their conceptual understanding of randomness, we provide no definitions or examples, leaving it entirely up to the participants how they interpret the term, in order to ensure that it is their own conception of randomness that we are studying.

2.1 Methodological overview

We used the following operationalization. Assuming a random source R and a non-random source N, we take an observer's proficiency at *discriminating* R from N to be the probability of correctly affirming whether a given pair of stimuli were both produced by R versus one by R the other by N. The observer's proficiency at *identifying* R and N is taken to be the probability of correctly labeling the sources of two stimuli, one from R the other from N.

A natural hypothesis is that discrimination proficiency predicts identification proficiency, in other words:

- (1) Hypothesis: The probability of correctly characterizing stimuli R and N (as “random” and “non-random,” respectively) coincides with the ease of distinguishing between the two sources.

The accuracy of (1) might be compromised by a pervasive bias in people's concept of “random.” As discussed in Chapter 1, it is well known that strings of bits that alternate slightly more than expected on the basis of random generation are likely to be labeled as random (Bar-Hillel & Wagenaar, 1991; Falk & Konold, 1997; Lopes & Oden, 1987; Nickerson, 2002). Likewise, attempts by untutored individuals to produce random sequences result in too many alternations, thus, in runs that tend to be too short (Baddeley, 1966; Kahneman & Tversky, 1972; Wagenaar, 1972). The alternation bias, however, might have a discrimination counterpart if strings with short runs are relatively difficult to distinguish from products of R. The two biases might thus be

balanced, ensuring the truth of (1). In sum, only direct comparison of discrimination versus identification can determine the relation between them.

Our experiments involved binary outcomes. In order to produce random bits (each with equal probability of being 0 or 1), we exploited the MATLAB (pseudo)random number generator. Reliance on this popular (albeit imperfect) process allows us to avoid difficult questions about the definition of “random.” (Even the most popular theories of “infinite random sequence” are open to objection; see Lieb et al., 2006; Osherson & Weinstein, 2008.) Regarding “non-random,” there are many ways to distort a random process. To explain the approach taken here, let $x \in [0, 1]$ be given. A bit string S is called “switch(x)” if and only if it was generated by the following stochastic algorithm:

Algorithm: Set the first bit of S randomly. Suppose that the n th bit of S has been constructed. Then with probability x the $n + 1$ st bit of S is set equal to the opposite of the n th bit; with probability $1 - x$ the $n + 1$ st bit of S is set equal to the n th bit.

The sequence S may be carried out to any length.

Thus, a switch(1) sequence consists of perfectly alternating bits, and a switch(0) sequence is homogeneous. In both cases, the first bit (chosen randomly) controls the rest. A switch(0.5) sequence is fully random. The expected proportion of alternations in a switch(x) sequence — called the “switch rate” — is obviously x . It can be seen that for $x < 0.5$, switch(x) sequences have longer runs than expected from a random source, whereas for $x > 0.5$ the runs are too short. Note that the expected proportion of 1’s (and of 0’s) in a switch(x) sequence of sufficient length is one half, for all $x \in (0, 1]$. Thus, the first-order entropy of a switch(x) string does not depend

on x (except if $x = 0$), and is maximal. In contrast, second-order entropy — defined over the relative frequencies of 00, 01, 10, and 11 in a given string — declines as x deviates from 0.5. But this does not lead us to rely on second-order entropy as a measure of objective randomness since we adopt the view (described above) that a bit string is random provided that it is produced by a random process, whatever the character of the string (e.g., all zeros). Note, moreover, that the string 001100110011001100110011 has maximal second-order entropy (the four binary patterns are equally represented) yet exhibits a non-random-looking regularity. The same kind of example can be given for n th-order entropy for any n (for discussion, see Attneave, 1959).

Finally, to foreshadow the experiments in this chapter, we first compare the discrimination versus identification of random and non-random stimuli. The results demonstrate systematic discrepancies between the two, revealing bias that exists in people's conceptual understanding of randomness but not in their perception. The final study seeks to probe conceptual understanding via a critical test of Falk and Konold (1997)'s encoding hypothesis, which claims that people base their judgments of randomness on the ease of memory encoding.

2.2 Experiment 1: Random vs. non-random matrices

This experiment examined the ability to distinguish random from non-random matrices and the ability to accurately identify which given matrix is random.

2.2.1 Method

Participants: Forty undergraduates (28 female, mean age = 19.6 years) from Princeton University participated in exchange for course credit. In all experiments reported in this

dissertation, participants provided informed consent. The protocol approved by the Princeton Institutional Review Board. All participants had normal or corrected-to-normal vision.

Stimuli: Stimuli were 60×60 matrices made up of green and blue dots. Each matrix (subtending ~ 8.5 visual degrees) could be divided either horizontally or vertically into equal halves (the orientation was randomly determined). One of the halves was fully random whereas the other was created from a sequence with a given switch rate x . The latter sequence was used to populate either successive rows or successive columns of the half-matrix (counterbalanced). All matrices were generated separately (“on the fly”). Figure 1a, below, provides six examples. In one condition of the experiment, the two halves were separated by a visible 50-pixel gap, and were presented either side by side (vertical division) or top and bottom (horizontal division).

Procedure: There were two conditions in the experiment: *discrimination* and *identification*. Participants were randomly assigned to each. Matrices in the discrimination condition were presented without a gap between the two halves, whereas those in the identification condition contained a gap. In the discrimination condition ($N = 20$), participants received the following instruction:

Each matrix can be divided into two halves either horizontally or vertically. The two halves are generated from different processes. Your task is to judge the orientation of the boundary between the two halves, by pressing “v” key for vertical or “h” key for horizontal.

Critically, no mention was made of randomness, probability, or related concepts in the task instructions. In the identification condition ($N = 20$), participants received the following instruction:

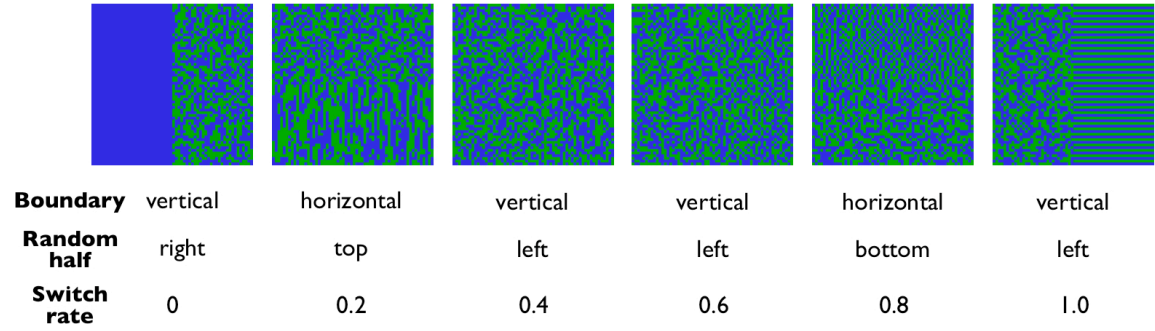
Each matrix is divided into two halves either horizontally or vertically. The two halves are separated by a gap. One half is generated from a random process and the other from a non-random process. Your task is to identify which half is more likely to be produced by a random process than a non-random process. Press the top or bottom arrow key if the division is horizontal, and left or right if vertical.

In both conditions, the switch rate used for half of the matrix varied from 0 (homogeneous) to 1 (checkerboard) by 0.02, resulting in 51 levels, while the other half of the matrix was random (i.e., $\text{switch}(0.5)$). Each level was repeated 10 times, making up 510 trials. The order of the trials was randomized for every participant. The same algorithm was used to generate the matrices in the two conditions. For each trial, the matrix (or the two separate halves) appeared on the screen for 1500ms; the screen then remained blank until response. No feedback regarding judgment accuracy was provided.

2.2.2 Results and discussion

For each participant in both conditions, the average discrimination and identification accuracy at every switch rate was calculated, then grand means were calculated by averaging across participants. Results are shown in Figure 1b. Overall, as the non-random half of the matrix became more random (i.e., switch rate closer to 0.5), both discrimination and identification accuracy decreased.

(a) Matrices



(b) Results

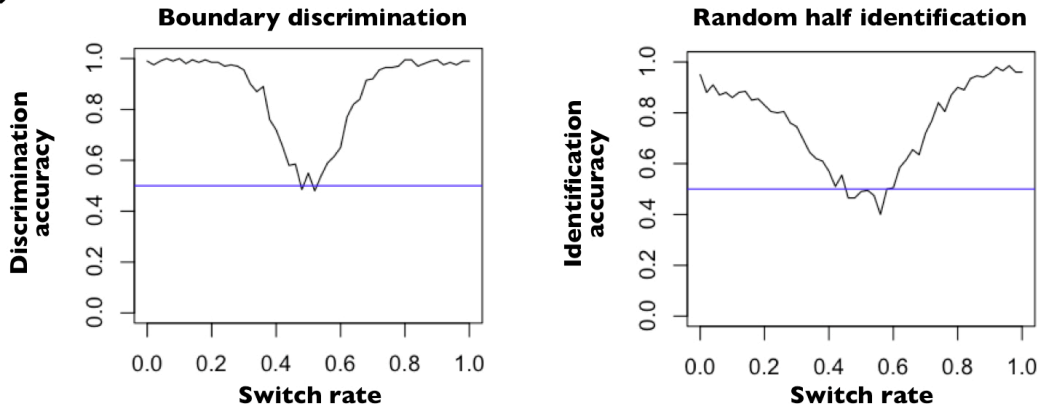


Figure 1: Experiment 1. (a) Six sample matrices (not drawn to scale). Each matrix consisted of two halves, one fully random, and the other with a given switch rate. The boundary between the two halves (vertical versus horizontal) was randomly determined, as was the choice of random half. (b) Discrimination and identification accuracy (defined as percentage of correct response) plotted as a function of switch rate. The blue horizontal line indicates chance performance (50%).

Across 51 levels of switch rate, discrimination accuracy was correlated with identification accuracy. There was a high correlation [$r(49) = 0.91$, $p < .001$], and it was maintained when considering just low switch rates ($\leq .5$) or high switch rates ($\geq .5$), with $r(23) = 0.92$, and $r(23) = 0.91$, respectively, $p < .001$.

Pooling all switch rates, discrimination accuracy (87.4%) was reliably higher than identification accuracy (75.4%) [$t(100) = 3.56$, $p < .001$]. This held for switch rates $\leq .5$ (88.7% vs. 74.8%; $t(48) = 3.20$, $p < .01$), and for switch rates $\geq .5$ (87.3% vs. 77.1%; $t(48) = 2.00$, $p = .05$). For 32 out of 51 levels of switch rate, discrimination accuracy was reliably higher than

identification accuracy [for each of these 32 levels, $t(38) > 2.12$, $p < .05$]. This suggests that the ability to distinguish random from non-random stimuli is superior to the ability to identify which stimulus is random.

At the same time, discrimination performance was reliably above chance (50%) for switch rates ≤ 0.42 (accuracy = 65.5%, $t(19) = 4.61$, $p < .001$), as well as for switch rates ≥ 0.56 (accuracy = 59.0%, $t(19) = 2.54$, $p = .02$). By contrast, identification performance was reliably above chance only for switch rates ≤ 0.38 (accuracy = 61.0%, $t(19) = 2.17$, $p = .04$), and ≥ 0.62 (accuracy = 58.5%, $t(19) = 2.82$, $p = .01$). Thus, participants experienced difficulty for a wider range of switch rates when performing identification compared to discrimination. Specifically, for x between 0.38 and 0.42, and between 0.56 and 0.62, participants could reliably distinguish random from non-random stimuli but nonetheless were unable to identify which was random. These results are inconsistent with Hypothesis (1) — the probability of correctly identifying stimuli as random coincides with the ease of distinguishing random from non-random stimuli.

Moreover, when switch rate was 0.56, identification accuracy was 40.0%, which is reliably below chance [$t(19) = 3.01$, $p < .01$]. Thus, participants consistently chose the switch(0.56) matrices as more random than truly random matrices whereas discrimination performance for the same switch rate was reliably above chance.² More generally, Figure 1b reveals that the identification curve was shifted to the right compared to discrimination (which was easier than identification, and also less biased by alternation). Once again, these results are contrary to Hypothesis (1).

² Note that discrimination below chance (which never occurred) would be anomalous.

2.2.3 Robustness of results

One follow-up study is briefly noted here. With 20 new participants (13 female, mean age = 19.4 years), we performed a replication of the identification procedure of Experiment 1 except that participants were asked to identify which half is more likely to be produced by a *non-random* process rather than a random one (thereby reversing the framing of the question). Mean identification accuracy across participants and switch rates was 75.3%, that is, virtually indistinguishable from the identification condition in Experiment 1 (75.4%, $t(38) = 0.04$, $p = .96$). There was also a strong correlation of identification accuracy across switch rates between the two experiments [$r(49) = 0.97$, $p < .001$]. The results suggest that framing has minimal influence on identification.

2.2.4 Feedback

It has been found that performance in both randomness generation and evaluation can be improved by feedback. For example, Rapoport and Budescu (1992) devised a competitive game that led participants to generate outputs that were increasingly random. Similarly, trial-by-trial feedback can eliminate certain biases associated with the conception of randomness; these include the Gambler's Fallacy and more generally, exaggerated expectation of alternation between outcomes (Edwards, 1968; Neuringer, 1986). However, the benefits of feedback have not been universal, for example, Budescu (1987) failed to find evidence for learning in a generation task.

It is thus unclear whether feedback can improve the conception of randomness, and it is equally unclear whether it will improve perceptual discrimination as has been found in other areas of perceptual learning (e.g., Goldstone, 1998; Pevzow & Harnad, 1997). We therefore

conducted an additional follow-up study with 80 new participants (51 female, mean age = 20.3 years) to determine whether feedback would alter discrimination and identification performance.

Twenty participants performed the boundary discrimination task and 20 performed the identification task without feedback, as in Experiment 1. Another 20 participants performed the discrimination task with feedback on every trial; the remaining 20 performed the identification task with feedback. Results are shown in Figure 2.

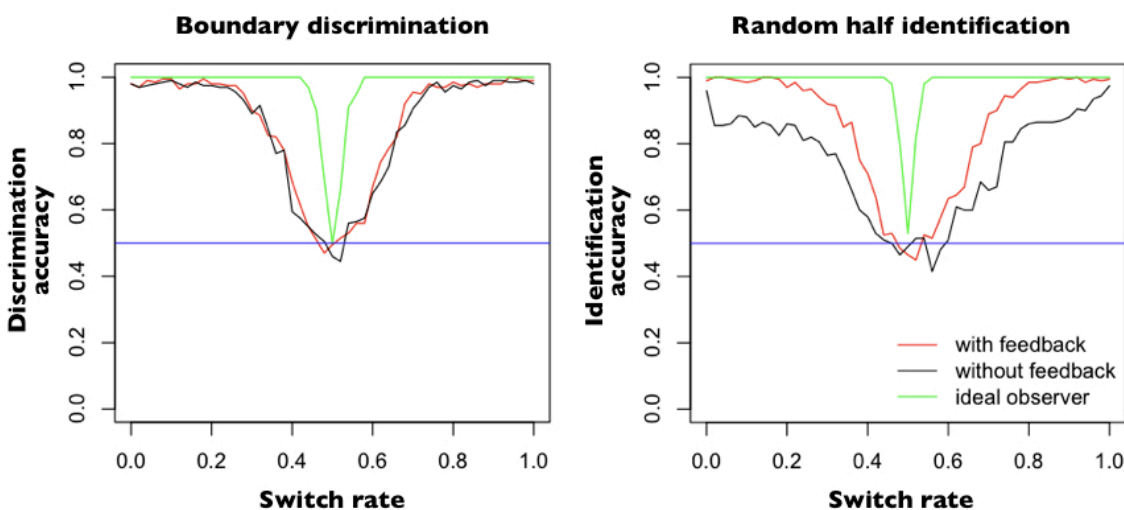


Figure 2: Feedback. Boundary discrimination accuracy and identification accuracy are plotted as a function of switch rates. The red line indicates performance with feedback, and the black line indicates performance without feedback. The green line indicates performance of the ideal observer. The blue horizontal line indicates chance performance (50%).

The results of the no-feedback conditions again closely replicate the findings of Experiment 1. More important, though, are the effects of feedback. For discrimination, there was no difference between feedback and no-feedback groups across switch rates [$t(100) = 0.30, p = .77$]. This was true for switch rates ≤ 0.5 [$t(50) = 0.17, p = .86$], and ≥ 0.5 [$t(50) = 0.24, p = .81$]. For identification, however, accuracy was reliably higher when feedback was provided than when no feedback was provided [$t(100) = 3.21, p = .002$]. This was true for switch rates ≤ 0.5 [$t(50) = 2.47, p = .02$], and marginal for switch rates ≥ 0.5 [$t(50) = 1.90, p = .06$]. Moreover,

identification performance rose to the level of perceptual discrimination when feedback was provided [$t(100) = 0.22, p = .82$]. Thus, Hypothesis (1) appears to be true for trained participants. These results further highlight the difference between discrimination and identification. The latter can be tuned to the expression “random;” no tuning is apparent for discrimination.

2.2.5 An ideal observer analysis

We ran a simulation to gauge how well an ideal observer would perform on the matrices used in our tasks. For discrimination, the switch rate for each of the four potential halves of a given matrix was computed (once traversing across rows and again traversing across columns). The ideal observer decided how the matrix was divided based on the maximum difference in switch rates between halves. For example, if the difference between left and right halves was greater than that between top and bottom (using the traversal direction that maximizes this difference), the boundary was declared vertical.

For identification, the switch rate for the two (separated) halves was computed. The ideal observer decided which half looked more random based on the minimum deviation of switch rate from 0.5 (fully random). This comparison was made using both horizontal and vertical traversal. That is, within a given half, we average the switch rate using horizontal traversal and the switch rate using vertical traversal. The half whose average switch rate showed minimum deviation from 0.5 was declared random. For example, given left and right halves, if the average switch rate of the left was closer to 0.5, then the left half was declared random. The ideal observer performance is plotted across switch rates in Figure 2. The mean accuracy for discrimination was 97.2% (SD = 9.4%) and the accuracy for identification was 98.2% (SD = 7.4%). The two

performances were not reliably different [$t(100) = 0.62, p = .54$]. In other words, the two tasks are objectively “equally hard.”

We then computed each participant’s *efficiency*: the ratio of the participant’s performance to that of the ideal observer. For discrimination, the efficiency was 88.5% (SD = 14.4%) with feedback and 87.3% (SD = 14.7%) without feedback, not reliably different [$t(100) = 0.39, p = .70$]. For identification, the efficiency was 86.7% (SD = 16.5%) with feedback and 75.9% (SD = 15.0%) without feedback, which were reliably different [$t(100) = 3.47, p < .001$]. Moreover, there was a reliable difference between discrimination and identification efficiency without feedback [$t(100) = 3.89, p < .001$], but not with feedback [$t(100) = 0.56, p = .58$].

2.3 Experiment 2: Random vs. non-random walks

The second experiment aimed to generalize the findings of Experiment 1 (and follow-ups thereto) by using a very different kind of stimulus: a dynamic random walk. This stimulus has a temporal component in that information about past outcomes is present only to the extent that it has contributed to the current, global state of the walk. Information about the local, step-by-step outcomes are available only via memory. This makes the stimulus interesting in light of recent emphasis on the role of short-term memory in judgment and generation of random sequences (Hahn & Warren, 2009; for older work emphasizing the role of short-term memory in sequence generation, see Baddeley, 1966; Kareev, 1992; Rabin, 2002; Rapoport & Budescu, 1997; Wagenaar, 1972). The dynamic nature of the stimulus is also of interest because the majority of past work on randomness judgments has used static displays. Dynamic stimuli not only extend the range of ecologically relevant stimuli, but also connect well with many of the key tasks in

which randomness is presumed to play a functional role in the real world such as predator evasion or foraging behaviors.

2.3.1 Method

Participants: A new group of forty undergraduates (27 female, mean age = 19.8 years) from Princeton University participated in exchange for course credit.

Stimuli: A horizontal line (subtending ~ 2.5 visual degrees) was presented in each quadrant of a computer screen. A given segment could rotate clockwise or counterclockwise with respect to its fixed left end (like an hour hand started at 3 o'clock). The direction of the rotation was determined by the next member of a given bit string (10° clockwise versus 10° counterclockwise). The movements in a quadrant will be called its “walk.” The four quadrants could be divided horizontally or vertically (randomly determined) into two halves. The two walks in one half were fully random, and those in the other half followed $\text{switch}(x)$ sequences at a given switch rate x . For the identification condition, the boundary between the two halves was marked by a black line. See Figure 3a, below. All walks were freshly generated for each trial. Sample trials can be seen at <http://www.princeton.edu/~jiayingz/experiments.html>.

Procedure: As in Experiment 1, there were two conditions: discrimination and identification, to which participants were randomly assigned. In the discrimination condition ($N = 20$), participants were informed that the four walks could be divided into left and right halves, or top and bottom halves, based on the way they moved. Specifically, either the left two walks would move in a different fashion from the right two walks, or the top two walks would move

differently from the bottom two walks. The participants had to judge the orientation of the boundary between the two halves. No mention was made of randomness in the instructions.

In the identification condition ($N = 20$), participants were informed that one half of the walks — indicated by the visible boundary — would move differently from the other half. Specifically, it was explained that for two of the lines whether the line moved clockwise or counter-clockwise on a given movement was randomly determined; for the other two lines, the walks would be non-random. The participant was asked to identify which half moved randomly.

As before, there were 51 levels of switch rate, varying by 0.02 from 0 (uniform sweep around the dial) to 1 (jiggling around the initial position). Each level was repeated twice, making up 102 trials.³ The order of the trials was randomized for every participant. For each trial, the line started at the 3 o'clock position. Each successive position (10° displacement) was presented for 100ms followed by a 50ms inter-movement interval. Each walk contained 100 movements, occurring simultaneously in the four quadrants. A trial lasted 15 seconds followed by a 1-second interval. If the participant responded within the 1-second interval, the trial was allowed to finish; otherwise, the screen remained blank until response. No feedback regarding judgment accuracy was provided.

2.3.2 Results and discussion

For each participant, the average discrimination or identification accuracy at every switch rate was computed, and these were averaged into grand means. Results are shown in Figure 3b. There was a reliable correlation between discrimination and identification accuracy [$r(49) = 0.67$, $p <$

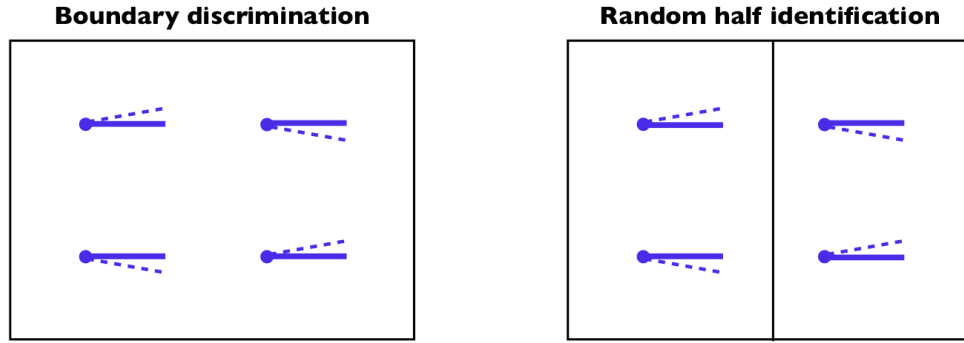
³ We used two repetitions in the present experiment (compared to 10 repetitions in Experiment 1) to keep duration manageable.

.001]. This correlation was maintained for switch rates ≤ 0.5 [$r(23) = 0.73, p < .001$], and ≥ 0.5 [$r(23) = 0.77, p < .001$].

Averaging across all switch rates, discrimination accuracy (71.6%) was reliably higher than identification accuracy (57.1%; $t(100) = 4.52, p < .001$). The difference appears when isolating switch rates ≤ 0.5 (74.3% vs. 53.9%; $t(48) = 4.69, p < .001$), as well as for switch rates ≥ 0.5 (69.8% vs. 60.3%; $t(48) = 2.00, p = .05$). For 19 of the 51 levels of switch rate, discrimination accuracy was reliably higher than identification accuracy [for each of these 19 levels, $t(38) > 2.04, p < .05$].

Finally, discrimination performance was reliably above chance for switch rates ≤ 0.38 and also for rates ≥ 0.80 (coincidentally, the same accuracy was achieved in the two cases: accuracy = 70.0%, $t(19) = 2.99, p < .01$). In comparison, identification performance was reliably above chance only for switch rates ≤ 0.04 (accuracy = 77.5%, $t(19) = 3.58, p < .01$), and for rates ≥ 0.92 (accuracy = 72.5%, $t(19) = 2.65, p = .02$). Thus, participants experienced difficulty for a much wider range of switch rates when performing identification compared to discrimination. Specifically, for switch rates between 0.04 and 0.38, and between 0.80 and 0.92, participants could reliably distinguish random from non-random walks but were unable to identify which was random. Once again, the results are inconsistent with Hypothesis (1) that the ability to identify random from non-random stimuli corresponds to the ability to distinguish between the two.

(a) Tasks



(b) Results

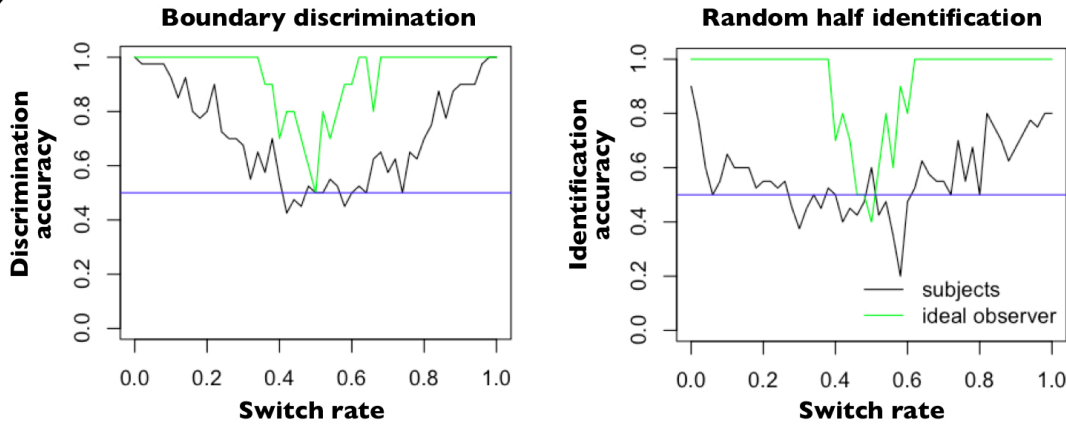


Figure 3: Experiment 2. (a) Each line starts horizontally and rotates according to its bit string. Dashed lines (not present in the stimuli) show possible first movements. In both conditions, the four quadrants could be divided into two halves: in one half the two walks were fully random, and in the other the two walks followed a sequence at a given switch rate. In the identification condition, a black line indicated the boundary between the two halves. (b) Discrimination and identification accuracy are plotted as a function of switch rate. The blue line indicates chance performance (50%).

The worst identification performance was at switch rate 0.58, yielding an accuracy of 20.0%, reliably below chance [$t(19) = 3.94, p < .001$]. Notice the proximity of switch rates with the lowest identification accuracy in Experiments 1 and 2 (0.56 and 0.58, respectively). Both are consistent with past findings, revealing that switch rates around 0.60 are most likely to attract the “random” label (Bakan, 1960; Budescu, 1987; Diener & Thompson, 1985; Falk, 1975).

As in Experiment 1, we ran an ideal observer analysis on the walks. For discrimination, the switch rate for each walk was first computed and the switch rates for two walks making up

each of the four potential halves were then averaged. The ideal observer discriminated the boundary based on the maximum difference between the two halves. For identification, the switch rate for each walk was computed and then averaged across two walks in one half. The ideal observer chose the half which looked random based on the minimum deviation of switch rate from 0.5 (fully random). The ideal observer performance is plotted across switch rates in Figure 3b. The mean accuracy for discrimination was 93.7% (SD = 12.0%) and the accuracy for identification was 92.7% (SD = 15.6%). They were not reliably different [$t(100) = 0.36, p = .72$].

Overall, the results of Experiment 2 replicated those in Experiment 1, despite the disparity in stimuli used. One notable difference is that the discrimination curve in Experiment 2 was flatter, and identification failed for a much wider range of switch rates (see Figure 3b). The difference may hinge on the dynamic vs. static contrast, the former making more demands on working memory.

2.3.3 Robustness: Rating task

The identification task based on binary choice (i.e., labeling one half as “random”) in our experiments has been used in other studies (e.g., Hsu et al., 2010). More common are tasks that ask participants to rate the randomness of stimuli on a scale. It is thus interesting to also examine randomness ratings for both the matrices and the walks.

In a follow-up study, 20 participants (11 female, mean age = 19.8 years) were presented with half-matrices exhibiting a given switch rate, and asked to rate how random the matrix looked on a scale from 0 to 100 (0 being completely non-random, 100 being completely random). Another 20 participants were presented with two walks governed by a given switch

rate, and asked to rate how random the two walks appeared (using the same scale). The average ratings of the matrices and the walks are shown in Figure 4.

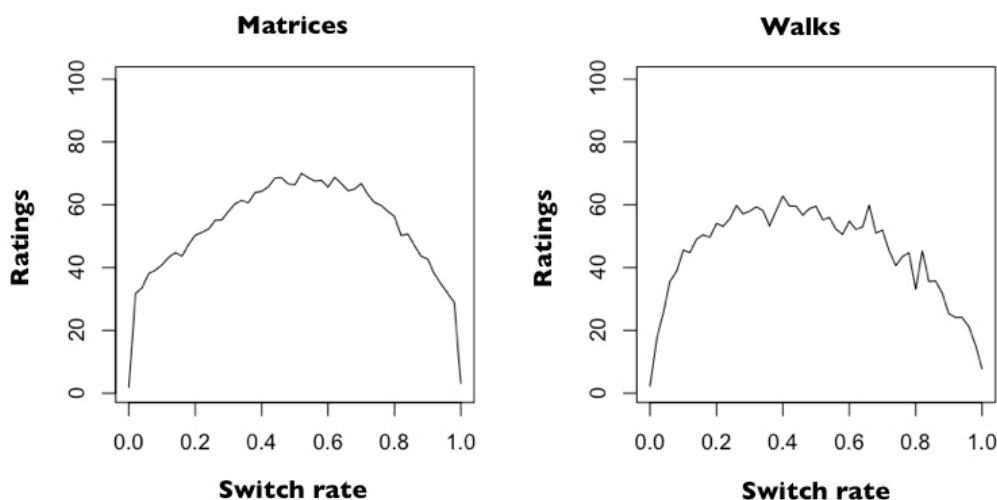


Figure 4: Ratings. Mean ratings of how random the half-matrices (left) and the walks (right) look on a scale from 0 (completely non-random) to 100 (completely random).

For matrices, we correlated the average rating at a given switch rate with the absolute difference between that switch rate and .5 (perfect randomness). The correlation was $r(49) = -0.90, p < .001$, suggesting that ratings carry considerable information about degree of randomness. For walks, the same correlation was $r(49) = -0.83, p < .001$. The correlation between the matrix rating versus identification performance at the same switch rate in Experiment 1 was $r(49) = -0.79, p < .001$; for walks, the correlation was $r(49) = -0.75, p < .001$. These strong correlations suggest good correspondence between the two types of measure, revealing the sensitivity of the randomness identification paradigm in use. The data for the random walk are much noisier (as they were for both our other tasks with this stimulus as well). The graph for the matrix data, however, shows qualitative evidence of over-alternation bias through the asymmetric, leftward shift along the horizontal axis, and visually corresponds to the findings of Falk and Konold (1997, see e.g., their Figures 3, 5, and 6).

2.4 Experiment 3: The encoding hypothesis

It seems that participants possess a conceptual bias of randomness which extends across very different types of stimulus. Consistent with past work (Bar-Hillel & Wagenaar, 1991), we find that people tend to identify over-alternating stimuli as “random.” Importantly, our results indicate that this bias is not present in perception. In this experiment, we sought to test the encoding hypothesis provided by Falk and Konold (1997), which stated the following:

- (2) Encoding hypothesis: The probability that a given bit string is judged “random” varies directly with the time needed to memorize or copy it.

As just noted, they found that memorization and copying difficulty predicted the “random” label quite well. Hypothesis (2) thus invites extension to other kinds of stimuli and additional measures of encoding difficulty. Specifically, in the present experiment we consider matrices rather than strings and detection of change rather than memorization. Same-different judgments provide a natural tool for probing memory encoding and are widely used to this effect (Grimes, 1996; Simons, 2000). Hypothesis (2), may thus be extended as follows:

- (3) Extended encoding hypothesis: The probability that a given matrix is judged “random” varies directly with the difficulty of detecting a change in the matrix. (By change we mean some bits of the matrix are inverted.)

To put Hypothesis (3) to a critical test, we constructed a type of matrix one half of which was fully random, the other half a mirror image of the first half. We reasoned that the precedence of global features in perception (Navon, 1977; Poljac et al., 2012) might lead participants to reject the “random” label for symmetrical matrices, but nonetheless to experience difficulty in processing individual bits (i.e., the matrix’s local features). Specifically, the role of symmetry may change as a function of stimulus and memory task. Symmetry increases compressibility (through redundancy) which can be exploited in encoding, at least in circumstances where all aspects of the stimulus are faithfully remembered. However, much of everyday memory is arguably not like this. Memory encoding is not like taking a snapshot and there is a wealth of research to suggest that stimuli are typically encoded partially (see Rensink, 2002 for a review). Moreover, most day-to-day tasks such as recognition and identification arguably require only encoding only up to a level of detail that supports discrimination.

These considerations reveal a possible defect in Falk and Konold (1997)’s encoding hypothesis. Symmetry is a property that is detected seemingly effortlessly and automatically in a wide variety of conditions (Wagemans, 1995), and it is likely to have a strong influence on perceptions of randomness (see e.g., Hsu et al., 2010). Yet in the same way that it is only one of many structural features that support data compression (and hence algorithmic complexity), it is only one of many stimulus aspects that potentially influence ease of encoding. It is not enough, vis-à-vis the encoding hypothesis, to show that each of these features individually influences both randomness perception and ease of encoding. Rather, to be successful as the process theory as Falk and Konold (1997)’s encoding account is intended to be, it must be the case that the specific *degree* to which such structural features influence encoding difficulty is matched by their

degree of influence on judgments of randomness. Only then is the claim plausible that the one serves as a (process level) proxy for the other.

Moreover, if the account is to apply beyond the confines of the lab, randomness judgments must track encoding difficulty not only in the time course of faithful copying, but for other plausible measures of encoding as well, since we do not, in real-world circumstances, typically memorize every aspect of a stimulus. In more naturalistic memory tasks, such as sequential same-different judgment, the ease of encoding may draw more strongly on some structural aspects than others, allowing potential dissociation between judgments of randomness (based on the presence or absence of structure) and memory performance. The final experiment sought explicitly to test this.

2.4.1 Method

Participants: A new group of forty undergraduates (27 female, mean age = 20.5 years) from Princeton University participated in exchange for course credit.

Stimuli: Stimuli were 16×16 matrices (each subtending ~ 5.7 visual degrees). Two kinds of matrices were generated. One kind was constructed identically to the switch(x) matrices of Experiment 1, for x ranging from 0.1 to 0.45 and from 0.55 to 0.9 in steps of 0.05. The second kind (called “mirror matrices”) were generated by filling the upper triangle submatrix (including its diagonal) with random bits, and then reflecting the upper triangle around the diagonal to fill the bottom triangle with corresponding bits (see Figure 5a, below). The choice of diagonal (top-left to bottom-right or top-right to bottom-left) was made randomly. For both kinds of matrix

[switch(x)⁴ and mirrored], whether tiling proceeded horizontally or vertically was determined randomly. All matrices were generated individually “on the fly” for each trial.

Procedure: Participants were assigned randomly to one of two conditions ($N = 20$ in each): change detection and identification. In the change detection condition, two matrices were presented serially and participants judged whether the matrices were the same or different. No mention was made of randomness. Half the time the two matrices were different (diversity trials), and half the time they were the same (identity trials). For diversity trials, the colors of 10 bits (randomly chosen) in the second matrix (of the two in a given trial) were inverted. The first matrix was either switch(x) for x in the range described above, or a mirror matrix. In individually randomized order, there were 10 diversity trials for every level of x plus 10 mirror diversity trials, and likewise 10 identity trials for every level of x plus 10 mirror identity trials — 340 trials in total. Each trial started with a blank screen for 500ms. Then the first matrix was presented until the participant pressed a button to proceed.⁵ Five hundred milliseconds later the second matrix was presented until response. See Figure 5a.

In the identification condition, each trial contained a switch(x) matrix and a mirror matrix, presented left to right (positions counterbalanced). Participants were asked to judge which matrix “looked random.” The same levels of x as in the change detection condition were used for identification trials, each repeated 10 times. There were thus 160 trials, each comparing a switch to a mirror matrix. For each trial, a blank screen appeared for 500ms after which the two matrices were presented until response.

⁴ We did not include fully random matrices ($x = 0.5$) in the experiment because they may be equally difficult to encode compared to mirror matrices. To demonstrate that encoding may not predict randomness judgment, we used non-random matrices which were presumably easier to encode but were identified as more random compared to mirror matrices.

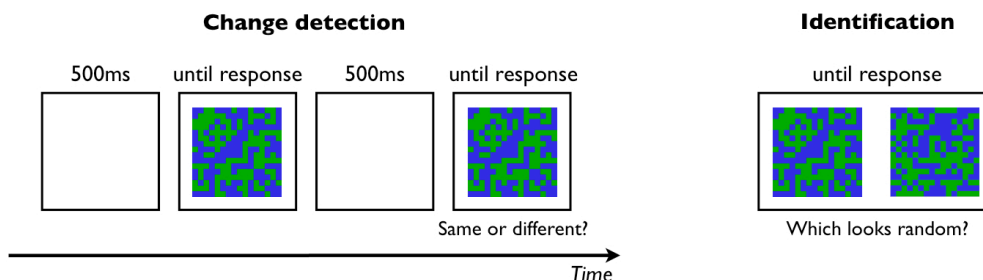
⁵ In a separate experiment we presented the first matrix for 500ms, and obtained similar results.

2.4.2 Results and discussion

To examine the accuracy of change detection performance, d' was computed for each participant for every switch rate as well as for the mirror matrices; averages were then computed over individual d' 's. The value of d' for a given level of x measures the encodability of switch(x) matrices, and likewise for the mirror matrices. The probability of choosing the mirror matrix as random in the identification task was also computed for every switch rate. Results are shown in Figure 5b.

In change detection, for every switch rate x , d' for mirror matrices was reliably lower than d' for switch(x) matrices [t 's(19) ≥ 2.14 , p 's $\leq .04$]. This suggests that mirror matrices were more difficult to encode than switch(x) matrices for all levels of x . In contrast, the probability of choosing the mirror matrix as more “random” was reliably below the chance level of 0.5 [t 's(19) ≥ 2.63 , p 's $\leq .01$], except for $x = 0.85$ and 0.9 . Compared to mirror matrices, switch(x) matrices (except for extreme values of x) were perceived as more random. Indeed, in debriefing, every participant evoked symmetry as their reason not to label a matrix as “random.” Thus, mirror matrices were harder to encode but less random in appearance compared to switch(x) matrices. The results are inconsistent with the extended encoding hypothesis (3).

(a) Tasks



(b) Results

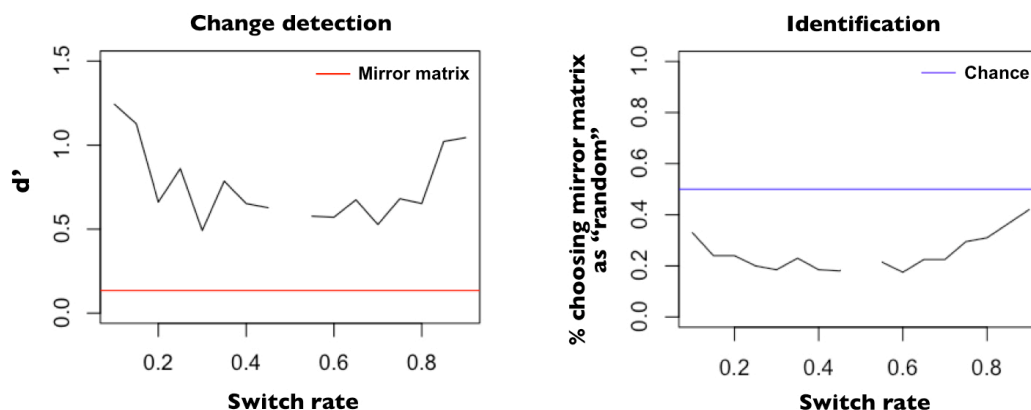


Figure 5: Experiment 3. (a) The left side shows the timeline for a change detection trial (matrices not drawn to scale). The right side shows an identification trial, with sample mirror and switch(x) matrices. (b) The left side plots change detection performance (d') for every switch rate and also for mirror matrices (the red line shows d' for mirror matrices). The right side shows the probability, for each switch rate, of choosing a mirror matrix “as random.” The blue line indicates chance performance (50%).

2.5 Chapter summary

In Experiments 1 and 2, we tested Hypothesis (1) — that the ability to identify the provenance of a stimulus drawn from a random source versus a non-random source can be predicted by the ability to distinguish between the products of these two sources. In other words, if you can see the difference between random and non-random stimuli, then you can tell which is random; discrimination entails identification.

Our results demonstrated that Hypothesis (1) is not correct, whether bit strings are rendered statically (Experiment 1) or dynamically (Experiment 2). In both cases, participants

were better at discrimination than identification. In our experiments, participants were not informed of the generating process, but they nonetheless performed reasonably close to an ideal observer. Apparently, the lay concept of randomness does not fully exploit the perceptual resources available for discriminating random from non-random sources. The alternation bias (Baddeley, 1966; Kahneman & Tversky, 1972; Wagenaar, 1972) thus appears to be fundamentally a conceptual phenomenon. The faulty human concept of randomness is not due to an inability to perceive randomness.

In Experiment 3, we evaluated the hypothesis of Falk and Konold (1997) according to which the tendency to label given stimuli as “random” can be predicted by their encoding difficulty. Matrices with a range of switch rates were compared to mirror matrices. We used a natural measure of encoding wherein participants detected a change between two matrices. The mirror matrices were more difficult to encode yet less likely to be labeled “random,” in contradiction with the extended encoding hypothesis (3). Thus, the encoding hypothesis due to Falk and Konold (1997) does not appear to generalize to detecting changes in bit matrices.

Chapter 3

Replications using different methods

In this chapter, we aim to replicate and extend the findings on the perception of randomness using different methods. We will also use genuine random numbers produced from radioactive decay, rather than pseudo-random numbers. As before, we explore the ability to distinguish random from non-random events without invoking the randomness concept. Non-randomness is quantified by excess repetitions (i.e., repeat) or alternations between successive bits (i.e., alternate). The first two experiments are replications of the first two studies in Chapter 1, with several methodological changes. We also report two additional experiments examining the perception of randomness without invoking the randomness concept. Specifically, participants learned to classify random and non-random patterns, and tracked changes in successive matrices. To foreshadow the results, we found superior performance in distinguishing random stimuli from repeat stimuli compared to alternate stimuli. Moreover, memory for repeat stimuli declined as stimuli became more random, whereas memory for alternate stimuli did not vary with the degree of non-randomness.

3.1 Methodological overview

To produce random bits (each with equal probability of being 0 or 1), we exploit the on-line service *Hotbits* that generates them from radioactive decay (see Walker, 2006). To generate “non-random” bits, we use a slightly different algorithm from that in Chapter 1. Assume the existence of an unlimited source of random bits (i.e., hotbits).

Definition of a repeat(x) sequence: Let $x \in [0, 1]$ be given (x is a probability). Then a repeat(x) sequence S is constructed via the following algorithm. The first bit is set equal to the next hotbit. Suppose that the n th bit of S has been constructed. Then with probability x the $n + 1$ st bit of S is the next hotbit; with probability $1 - x$ the $n + 1$ st bit of S is set equal to the n th bit. The sequence S may be carried out to any length.

Thus a repeat(1) sequence is fully random (namely, a string of hotbits), and a repeat(0) sequence is all 1's or all 0's depending on the hotbit drawn for the first bit. Note that in constructing repeat(x) sequences, the probability x is also computed using hotbits. (For example, three hotbits are used for the $1/8$ probability of repeating the last bit.) The same remarks apply to the second distortion of randomness.

Definition of an alternate(x) sequence: For $x \in [0, 1]$, an alternate(x) sequence S is constructed as follows. The first bit is set equal to the next hotbit. Suppose that the n th bit of S has been constructed. Then with probability x the $n + 1$ st bit of S is the next hotbit; with probability $1 - x$ the $n + 1$ st bit of S is set equal to the *opposite* of the n th bit.

It can be seen that for $x < 1$, $\text{repeat}(x)$ have longer runs than expected from a random source, whereas the runs in an $\text{alternate}(x)$ are too short. In both cases we expect the string to be more compressible than a (fully) random string. The experiments described in this chapter probe the psychological boundary between full and partial randomness, asking in each case:

What is the greatest $x \in [0, 1]$ such that $\text{repeat}(x)$ sequences are treated differently from random sequences, and similarly for $\text{alternate}(x)$ sequences?

We begin with a task of distinguishing bit matrices created from random versus $\text{repeat}(x)$ or $\text{alternate}(x)$ sequences.

3.2 Experiment 4: Distinguishing random vs. non-random matrices

This experiment examined the ability to distinguish random from non-random matrices with a staircasing method. We aimed to replicate Experiment 1 with different methods.

3.2.1 Method

Participants: Sixty undergraduates (43 female, mean age = 19.8 years) from Princeton University participated for course credit.

Stimuli: The stimuli were similar to those in Experiment 1. They were 60×60 matrices made up of green and blue dots. Each could be divided either horizontally or vertically into equal halves (the orientation was randomly determined). One of the halves was filled with hotbits (i.e., it was fully random) whereas the other was created from either a $\text{repeat}(x)$ sequence or an $\text{alternate}(x)$

sequence; the sequence was used to populate either successive rows or successive columns of the half-matrix (counterbalanced). Each matrix subtended 14.2×14.2 degrees of visual angle. All matrices were generated separately (“on the fly”). Figure 1 provides two examples.

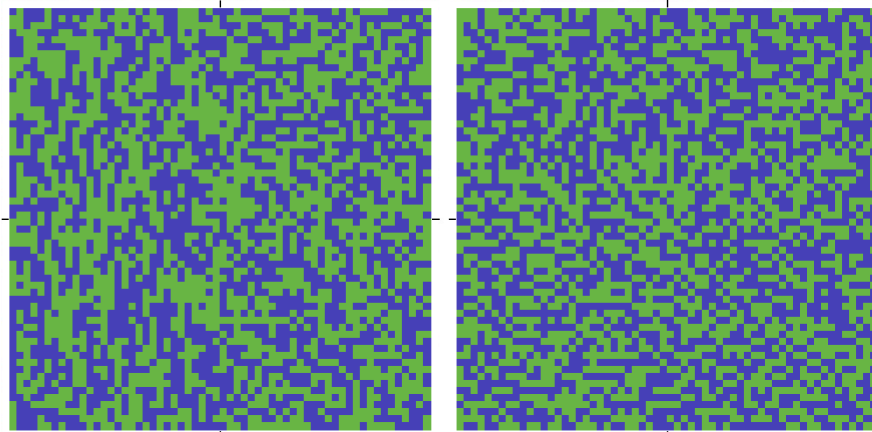


Figure 6: Experiment 4. The matrix on the left can be divided vertically into a random half-matrix (right) and a repeat(0.7) half-matrix (left). The latter half-matrix was created from a repeat(0.7) sequence by traversing columns. The matrix on the right can be divided horizontally into a random half-matrix (top) and an alternate(0.7) half-matrix (bottom). The latter half-matrix was created from an alternate(0.7) sequence by traversing columns.

Procedure: Each of the thirty participants served in two conditions (order counterbalanced). In the *repeat* condition only repeat(x) sequences were used to compose non-random halves of matrices whereas in the *alternate* condition only alternate(x) sequences were used. In each condition a total of 100 matrices were presented. The participant was informed that the two halves were different and was asked to guess the orientation (horizontal versus vertical) of the boundary between the two halves. Each matrix appeared on the screen for 4000ms, following which the participant guessed the boundary. Preliminary practice trials included clear cases based on low values of x . There was no mention of randomness. Feedback regarding judgment accuracy was provided after every trial. By receiving feedback participants could learn to distinguish the difference between random and non-random matrices. A cleaner way to assess

discrimination ability would involve no feedback. Thus, a new set of 30 participants took part in the “no feedback” version, same as above but without feedback.

We used a staircasing method to gauge the ability of discrimination. For every participant, the first matrix was generated with x at 0.7 since pilot studies indicated near perfect discrimination up to $x = 0.7$. Subsequent matrices were generated contingent on performance with the previous one. Specifically, for every incorrect answer x decreased by 0.01, whereas x increased by 0.01 in case of two successive correct answers; otherwise, x was left unchanged.

3.2.2 Results and discussion

For each participant, we identified the highest x value at which the participant could reliably discriminate between the random and repeat(x) or alternate(x) halves of matrices. Specifically, for a given x we tallied the numbers of correct and incorrect trials at or above x , and then determined via a binomial test whether the difference is significant in the direction of accuracy. The greatest such x was retained. This calculation was carried out separately for repeat and alternate trials, yielding repeat and alternate *thresholds*.

With feedback, the average repeat threshold was 0.85 ($SD = 0.05$, median = 0.85) and the average alternate threshold was 0.82 ($SD = 0.04$, median = 0.82), $N = 30$ in both cases. The difference was reliable via a paired t -test [$t(29) = 2.36$, $p < .05$] and via Wilcoxon test [$p < .05$]. Without feedback, the average repeat and alternate thresholds were 0.85 ($SD = 0.05$, median = 0.85) and 0.78 ($SD = 0.05$, median = 0.80), respectively. The difference was again reliable [$t(29) = 5.42$, $p < .01$, Wilcoxon $p < .01$]. Feedback improved discrimination performance for alternate matrices but not for repeat matrices. These results reveal that participants were better at discriminating random matrices from matrices biased towards repeats compared to alternates.

3.2.3 Identifying random matrices

Here we examined the ability to identify which half of the matrix looked random. Stimuli were 60×60 matrices, each consisting of two halves: one half generated by a random sequence and the other by a repeat(x) or an alternate(x) sequence. Here, however, the two halves were separated by an easily visible 50-pixel gap. The two half-matrices were presented either side by side (vertical division) or top and bottom (horizontal division). A new group of thirty participants (18 female) served in both repeat and alternate conditions (order counterbalanced). For each matrix, the participant identified the half that “looked more random.” No feedback was given regarding judgment accuracy. The first trial started with $x = 0.7$. Subsequent trials were generated according to the usual schedule: For every incorrect answer x decreased by 0.01 whereas it increased by 0.01 in case of two successive correct answers; otherwise, x was left unchanged.

The average threshold was 0.80 ($SD = 0.14$, median = 0.84) for repeat and 0.72 ($SD = 0.11$, median = 0.69) for alternate. The difference is reliable via a paired t -test [$t(29) = 2.23$, $p < .05$] and via Wilcoxon test [$p < .01$]. Thus, compared to repeat, alternate stimuli had a greater tendency to be considered “random.” This result is consistent with previous findings, as well as the results reported in Chapter 2, that stimuli that alternate more than expected are more likely to be labeled as random (e.g., Lopes & Oden, 1987). Moreover, the average repeat and alternate thresholds were reliably lower than those reported above (Section 3.2.2) in the discrimination paradigm [$p < .01$]. Hence, consistent with Experiment 1, the ability to distinguish random from non-random matrices was superior to the ability to identify random matrices in the presence of non-random distracters.

3.3 Experiment 5: Distinguishing random vs. non-random walks

Since events can also occur in a sequential order (e.g., basketball attempts), in this experiment we examined the ability to distinguish random from non-random temporal stimuli. Here we hoped to replicate Experiment 2 with different methods.

3.3.1 Method

Participants: Forty undergraduates (24 female, mean age = 20.1 years) from Princeton University participated for course credit.

Stimuli: The stimuli were similar to those in Experiment 2. A horizontal line was presented in each quadrant of a computer screen. Each line could rotate clockwise or counterclockwise with respect to its fixed left end (like an hour hand started at 3 o'clock). The direction of the rotation was determined by the next member of a given bit-string (10° clockwise versus 10° counterclockwise). See Figure 2. The new position of the line was presented for 100ms followed by a 50ms inter-movement interval. The movements in a given quadrant are called its *walk*. The walk in one quadrant (the “oddball”) followed a *repeat(x)* or an *alternate(x)* sequence, whereas the walks in the three others were (fully) random. A walk consisted of 100 successive movements; they occurred simultaneously in the four quadrants.

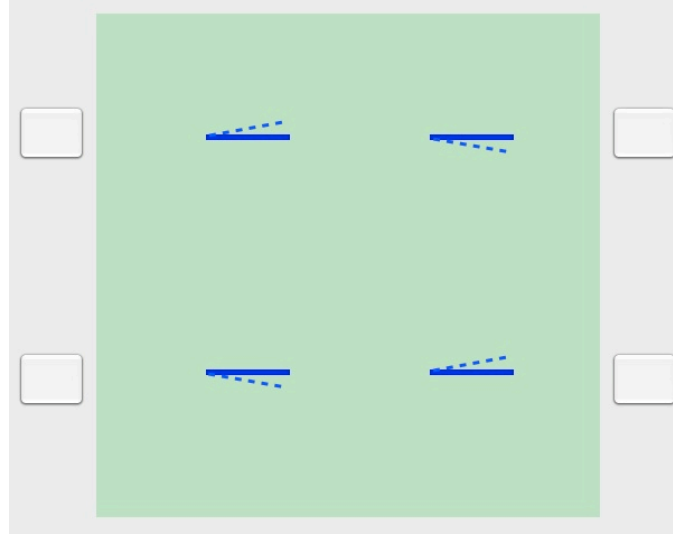


Figure 7: Experiment 5. Every line starts horizontally and rotates according to its bit string. One line follows a $\text{repeat}(x)$ or an $\text{alternate}(x)$ sequence, while the other three lines move randomly. Dashed lines (not present in the stimuli) show possible first movements. The four buttons (left and right margins) registered the participant's guess about the oddball.

Procedure: Twenty participants were randomly assigned to the *repeat* condition, and twenty to *alternate*. Each condition consisted of 100 trials. A given trial in the repeat condition presented random walks in three quadrants and a $\text{repeat}(x)$ walk in the other; for the alternate condition, the latter quadrant presented an $\text{alternate}(x)$ walk. The choice of oddball quadrant was determined randomly for each trial. The participant was informed that one line would move in a distinctive way compared to the other lines. After viewing the walks, the participant was invited to choose the oddball quadrant by clicking its button. There was no mention of randomness. Feedback regarding judgment accuracy was provided after every trial.

A staircasing method was used to examine the ability to discriminate between random and non-random walks. The first trial started with the easy value $x = 0.3$ (the value was determined from pilot studies). Subsequent trials were generated as in Experiment 4: For every incorrect answer x decreased by 0.05 whereas it increased by 0.05 in case of two successive correct answers; otherwise, x was left unchanged.

3.3.1 Results and discussion

Just as in Experiment 4, we identified repeat and alternate thresholds for every participant. The average repeat threshold was 0.80 ($SD = 0.10$, median = 0.83) and the average alternate threshold was 0.63 ($SD = 0.20$, median = 0.67). The difference was reliable via an independent-samples t -test [$t(38) = 3.32$, $p < .01$] and via Mann-Whitney U test [$p < .01$]. It thus seems easier to discriminate repeat-biased walks from random walks than to discriminate alternate-based walks. The results are congruent with the findings of Experiment 4 except that both repeat and alternate thresholds are lower in the temporal compared to the spatial context [0.80 vs. 0.85, $t(48) = 2.29$, $p < .05$ and 0.63 vs. 0.82, $t(48) = 5.02$, $p < .01$, respectively]. The difference may be due to the smaller number of bits exploited in creating four walks compared to a matrix (400 vs. 1800). Alternatively, it may hinge on the temporal character of the walks, or the need for divided attention in viewing them.

3.4 Experiment 6: Classifying random vs. non-random matrices

Whereas the preceding two experiments involved discriminating random from non-random stimuli, the current experiment examined the ability to learn distinct responses to each. With the help of feedback, participants attempted to press one button when presented with a random matrix and another button in response to repeat(x) or alternate(x) matrices. There was no mention of randomness or related notions.

3.4.1 Method

Participants: Thirty-five undergraduates (22 female, mean age = 20.3 years) from Princeton University participated for course credit.

Stimuli: Stimuli were the half-matrices used in Experiment 4, generated anew in the present setting. Each was either 30×60 or 60×30 , randomly determined for each trial.

Procedure: Each participant served in both repeat and alternate conditions (order counterbalanced). Each condition consisted of 200 trials. A given trial in the repeat condition presented either a random or a repeat(x) matrix (the choice was randomly determined); likewise, in the alternate condition, either a random or an alternate(x) matrix was presented. The participant was informed that each matrix was drawn from one of two distinct categories. The image was presented for 1000ms, following which the participant guessed the category to which it belonged. The categories were represented by two unlabeled buttons below the image, left and right. One was meant for random matrices and the other for non-random (consistently for a given participant, determined randomly across participants). Feedback was provided after every trial indicating whether the correct button was selected.

At the start of the experiment every participant sampled five matrices from the random, repeat (or alternate) categories, and the associated buttons were indicated. Then the first trial started with the easy value $x = 0.55$ (the value was determined from pilot studies). Subsequent trials were generated according to the usual schedule: For every incorrect answer x decreased by 0.01 whereas it increased by 0.01 in case of two successive correct answers; otherwise, x was left unchanged.

3.4.2 Results and discussion

We identified repeat and alternate thresholds for every participant as in Experiment 4. The average repeat threshold was 0.90 ($SD = 0.03$, median = 0.91) and the average alternate threshold was 0.86 ($SD = 0.07$, median = 0.87). The difference is reliable via a paired t -test [$t(34) = 2.94$, $p < .01$] and via Wilcoxon test [$p < .01$]. Thus, comparably to the first two experiments, classification was easier in the repeat context compared to alternate. In turn, this implies greater facility in discriminating repeat-biased than alternate-biased matrices from random ones.

At the same time, both the repeat and the alternate thresholds were reliably higher than those in Experiment 4 (0.90 vs. 0.85 and 0.86 vs. 0.82; $t > 2.73$, $p < .01$ in both cases). The difference between the two experiments, however, might be due to the greater number of trials in the present setting (200 vs. 100). If thresholds are calculated using just the first hundred trials of Experiment 6, they are the same as in Experiment 4 (0.85 for repeat and 0.82 for alternate).

3.5 Experiment 7: Encoding random vs. non-random matrices

So far performance was better with repeat compared to alternate stimuli. Does the difference reflect easier encoding of repeat-based compared to alternate-based stimuli? The current experiment examined the encoding difficulty of repeat vs. alternate matrices by requesting participants to detect changes in serially presented matrices.

3.5.1 Method

Participants: Forty undergraduates (20 female, mean age = 19.7 years) from Princeton University participated for course credit.

Stimuli: Stimuli were 11×11 matrices made up of black and white dots. Each matrix was created from a repeat(x) or an alternate(x) sequence by traversing rows or columns (counterbalanced). All matrices were generated on the fly.

Procedure: There were nine types of trials corresponding to (a) repeat levels of 0.2, 0.4, 0.6, and 0.8, (b) the same alternate levels, and (c) full randomness. In a given trial, a black and white matrix was constructed for one of the nine levels. A second matrix was constructed as follows. With 50% probability the second matrix was identical to the first, and with 50% probability the second matrix differed at five randomly chosen cells (whose colors were reversed). The first matrix was presented for 200ms followed by a 500ms gap (no mask was presented); then the second matrix was presented for 200ms. See Figure 3. At the end of a trial, the participant judged whether the two matrices were the same or different. There were 30 trials for each of the nine levels. The resulting set of 270 trials was presented in individualized random order to each participant.

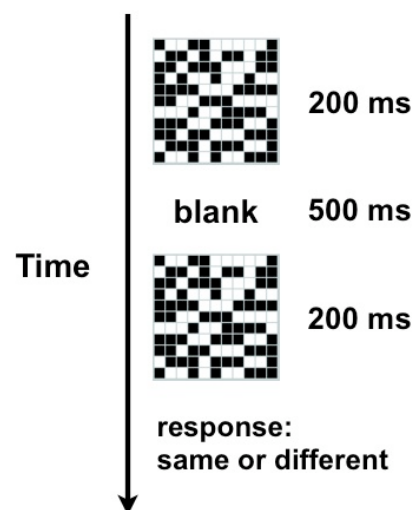


Figure 8: Experiment 7. Timeline of a sample trial in Experiment 7.

3.5.2 Results and discussion

For each participant and each of the nine levels, we tabulated the number correct out of 30 trials. The means (over forty participants) are shown in Table 1. For example, the mean accuracy for repeat(0.2) matrices is 91% whereas it is just 70% for alternate(0.2) matrices. For each level, repeat accuracy was compared to alternate accuracy via paired *t*-test, also summarized in the table. Repeat accuracy was reliably higher than alternate only for levels 0.2 and 0.4.

A one-way repeated measure ANOVA was performed on the four repeat levels plus full randomness as a fifth level; the same ANOVA was performed for alternate. There was a significant difference among the levels in the repeat condition [$F(4,191) = 8.14, p < .01$], but not in the alternate condition [$F(4,191) = 0.40, p = 0.81$]. Post-hoc *Tukey HSD* tests revealed that accuracy for $x = 0.2$ and $x = 0.4$ was reliably higher than for the other levels, whereas accuracy for $x = 1.0$ (full randomness) was reliably lower ($t > 2.69, p < .01$ in all cases).

x	Repeat	Alternate	Sig.
0.2	91% (7%)	70% (10%)	$p < .01$
0.4	77% (10%)	68% (10%)	$p < .01$
0.6	71% (10%)	69% (9%)	$p = .32$
0.8	70% (10%)	70% (10%)	$p = 1.0$
1.0	65% (10%)	65% (10%)	$p = 1.0$

Table 1. Mean percent accuracy (and SD) in Experiment 7. *p*-values reflect repeat/alternate differences in accuracy.

These results suggest that participants were better at encoding repeats than alternates but only for highly non-random sequences ($x = 0.4$ and below). Moreover for repeat, participants showed weaker encoding as matrices became more random. For alternate, performance seemed not to vary with degree of non-randomness. One potential explanation for the difference in encoding

ability between repeat and alternate matrices could be that it is easier to extract streaks than alternations (Brady & Tenenbaum, 2010). Alternations may require more visual working memory resources compared to repeats.

3.6 Chapter summary

In four experiments, we examined the ability to distinguish genuine random from non-random stimuli. Using different algorithms to generate non-random stimuli and a different staircasing method, we replicated the results of Experiment 1. That is, the ability to distinguish random from non-random matrices was superior to the ability to identify which matrix was random in the presence of non-random distracters. In addition, we found superior performance in distinguishing random stimuli from stimuli that are biased towards repeats compared to alternates. The ability to distinguish repeat matrices from random matrices was slightly but reliably better than the ability to distinguish alternate matrices from random. Participants were likewise better at detecting non-random walks biased towards repeat in a field of random walks. When learning to classify matrices into nominal categories, participants were better at classifying repeat versus random matrices compared to alternate versus random. Finally, participants were better at perceiving changes embedded in non-random matrices biased towards repeats compared to alternates; moreover, memory for repeat matrices declined as matrices became more random whereas memory for alternate matrices did not vary with the degree of non-randomness.

Chapter 4

Human production of random stimuli

We have examined the ability to distinguish and to identify random from non-random stimuli. In this chapter, we examine another aspect of randomness cognition: the ability to produce stimuli as random. It is well known from previous literature that attempts by untutored individuals to produce random bits result in too many alternations, thus, in runs that tend to be too short (Baddeley, 1966; Kahneman & Tversky, 1972; Wagenaar, 1972). We would like to know whether such bias for alternations is related to the ability to perceive or to identify random from non-random stimuli.

4.1 Experiment 8: Producing random sequences

This experiment invited participants to produce “random sequences,” and quantified the degree of non-randomness of their productions.

4.1.1 Method

Participants: Forty undergraduates (25 female, mean age = 20.1 years) from Princeton University participated for course credit.

Stimuli and procedure: Each participant was asked to produce ten sequences of 100 heads (H's) or tails (T's), “as if the sequences were produced by a fair coin.” They were required to construct a given sequence within 50 seconds. For each sequence, the bits were reproduced and remained on the computer screen as they were entered. The time and number of bits remaining were also made visible.

4.1.2 Results and discussion

For each sequence produced in the experiment, we tallied the number of alternations and divided it by 99 (the sequence length – 1). This yields the alternation rate r of the sequence. In all 400 sequences (ten per participant), r was above 0.5. For every participant, the mean r of the participant's ten sequences was computed. Across participants, the grand mean of the average r was 0.68 ($SD = 0.12$, median = 0.67).

To quantify non-randomness, we converted r for a given sequence into the (unique) value x such that the expected number of alternations in an alternate(x) sequence equals r [via $r = 0.5 \times x + (1 - x)$]. For each participant, we tallied the average of this x value for her 10 sequences. Across participants, the grand mean of x was 0.64 ($SD = 0.24$, median = 0.64). This value was lower than the alternate threshold for identification (0.72) in Experiment 4 [$t(68) = 1.89$, $p < .05$]. The grand mean $x = 0.64$ was also significantly lower than the average alternate threshold for discrimination (0.82 with feedback, 0.78 without) for Experiment 4 [$t > 4.07$, $p < .01$ in both cases]. These results suggest that the sequences produced as “random” were more biased towards alternations than sequences judged as “random”, or sequences that could be distinguished from truly random ones. In the predator-prey context, this suggests that the prey may generate a

temporal sequence that is biased towards alternations ($x = 0.64$), but the prey can still evade the predator since the predator cannot discriminate the same sequence ($x = 0.64$) from random, and the sequence is identified as “random” (see Experiment 2, where discrimination was reliably above chance for $x \geq 0.80$, and identification was reliably above chance for $x \geq 0.92$).⁶

4.1.3 Comparing one’s own productions with random sequences

Here we tried to determine whether participants could discriminate their productions from truly random sequences. Twenty participants (11 female, mean age = 19.5 years) produced ten sequences of length 100 (meant to be random, as above). Then each sequence was matched with a genuine random sequence, and each was used to control a train of stimuli: a blue circle (for H) or a yellow square (for T). The two trains were made available side by side on the screen; participants chose which to view first (both could be viewed as many times as desired). Each stimulus in a train was shown for 100ms, followed by a 100ms blank interval. The left-right position of the random versus produced sequence was randomized, as was the order of the ten productions. Subsequent to examining the two sequences, participants chose the one that “looks more random.”

Across the 20 participants, the average x of produced sequences was 0.67 ($SD = 0.06$, median = 0.67), very close to the production result reported above. For each participant, we tabulated the percentage of trials (out of 10) in which the genuinely random sequence was chosen. The average (over participants) was 45% ($SD = 0.23$), not reliably different from chance (50%) [$t(19) = 0.96$, $p = .35$]. It thus seems that participants could not discriminate their own produced sequences from random ones.

⁶ We thank Eldar Shafir for this observation.

4.2 Experiment 9: Linking discrimination, identification, and production

The purpose of this experiment was to compare discrimination, identification, and production performance on a within-subject basis. We therefore asked each participant to perform the paradigms of Experiments 4 and 8.

4.2.1 Method

Participants: Thirty undergraduates (17 female, mean age = 19.4 years) from Princeton University participated for course credit.

Stimuli and procedure: Each participant first discriminated between random and repeat(x) matrices, or between random and alternate(x) matrices. They then identified which half of the matrix looked more random as in Experiment 4. No feedback was provided. Finally, the participant produced random sequences as in Experiment 8. The order of the identification and production tasks was counterbalanced. On the other hand, the discrimination task always went first so that it was completed before any mention of “randomness.”

4.2.2 Results and discussion

For every participant, we calculated the repeat and alternate thresholds for the discrimination task, and the same thresholds for the randomness identification task of Experiment 4. We also calculated the alternate(x) value of the produced sequences, as in Experiment 8.

For randomness discrimination, the average repeat threshold was 0.82 ($SD = 0.06$, median = 0.83) and the alternate threshold was 0.78 ($SD = 0.05$, median = 0.78); the latter value is reliably lower than for repeat [$t(29) = 3.14$, $p < .01$; Wilcoxon $p < .01$]. The repeat and

alternate thresholds were not significantly different from those computed in Experiment 4 [$p > .37$].

For randomness identification, the average repeat threshold was 0.85 ($SD = 0.07$, median = 0.87) and the alternate threshold was 0.72 ($SD = 0.15$, median = 0.66); once again, the latter value is reliably lower than for repeat [$t(29) = 4.11$, $p < .01$; Wilcoxon $p < .01$]. The comparable thresholds from Experiment 4 are 0.80 and 0.72, not significantly different from those calculated here [$p > .07$ in both cases].

For randomness production, the grand mean for x values was 0.64 ($SD = 0.21$, median = 0.63). The mean is identical to the one seen in Experiment 8. Thus, the results from previous experiments were replicated here.

We next correlated the thresholds of perceptual discrimination with those of identification, across the 30 participants. The correlation was insignificant for both repeat ($r = 0.08$, $p = .69$) and alternate ($r = -0.05$, $p = .81$). Hence, the ability to identify matrices as “random” seems not to co-vary with the ability to discriminate them from non-random matrices. Likewise, the average x value of productions was not significantly correlated with the discrimination alternate threshold [$r = 0.08$, $p = .69$]. Finally, average x values in production are uncorrelated with alternate thresholds in identification ($r = 0.31$, $p = .09$). In sum, the ability to produce random sequences seems to co-vary neither with the ability to discriminate random from non-random matrices, nor with the ability to pick the random matrix from a pair.

4.3 Chapter summary

We found that the sequences produced by participants as “random” were biased towards alternations, consistent with previous findings. Moreover, these productions were more biased

towards alternations than sequences judged as “random”, or sequences that could be distinguished from truly random ones. By comparing performance within the same individual, we found that the ability to produce random sequences seems not to be related with the ability to discriminate random from non-random matrices, or the ability to identify the random matrix from a pair. In sum, the results suggest the production of random stimuli may be a process different from those in the perception and identification of randomness.

Chapter 5

Implicit measures of randomness

How we understand randomness can be examined in multiple ways. In previous chapters, randomness was studied in discrimination tasks (e.g., what's the boundary between the two halves of the matrix?), identification tasks (e.g., which half looks more random?), and production tasks (e.g., produce a sequence as if it's generated by a fair coin). The latter two paradigms explicitly measure the subjective concept of randomness, which does not match that of genuine randomness. On the other hand, the discrimination paradigm does not invoke people's ideas of randomness. Rather, it measures the perceptual sensitivity to the difference between fully random and non-random stimuli. Several studies reported in the previous chapters have consistently demonstrated that discrimination performance is superior to identification or production performance. This suggests that tasks that do not rely on the label "randomness" may reveal other aspects of randomness cognition which are untapped by the more explicit paradigms.

In this chapter, we will explore a set of implicit measures of randomness, which are designed to examine the cognitive consequences of randomness without explicitly invoking the label "randomness." We first report some preliminary findings.

5.1 Experiment 10: Predicting random vs. non-random bits

Here we examine the ability to predict either a random or a non-random binary sequence. No mention was made of randomness or related concepts in the experiment.

5.1.1 Method

Participants: Sixty undergraduates (41 female, mean age = 19.7 years) from Princeton University participated for course credit.

Stimuli and procedure: There were four conditions with 15 participants in each: visual repeat, visual alternate, auditory repeat, and auditory alternate. In the two visual conditions, a square may appear either in the top or bottom position on the screen. The positions of the square followed either a repeat(x) sequence (visual repeat condition), or an alternate(x) sequence (visual alternate condition). There were five levels of x : 0.6, 0.7, 0.8, 0.9, and 1 in both conditions. For each trial, the participant first made a prediction of whether the square would appear on the top or bottom by clicking on the ‘high’ or ‘low’ button. See Figure 9. The square then appeared on the screen, either confirming or rejecting the participant’s prediction. In the two auditory conditions, a high-frequency sound or a low-frequency sound may be presented to the participant through loud speakers. The frequency of the sound followed either a repeat(x) sequence (auditory repeat condition), or an alternate(x) sequence (auditory alternate condition). There were five levels of x : 0.6, 0.7, 0.8, 0.9, and 1 in both conditions. For each trial, the participant first made a prediction of whether there would be a high-frequency or low-frequency sound by clicking on the ‘high’ or ‘low’ button. The sound was then presented to the participant. In all conditions, the

participant made a total of 50 predictions in the experiment. The number of total and correct predictions was visible on the screen throughout the experiment.

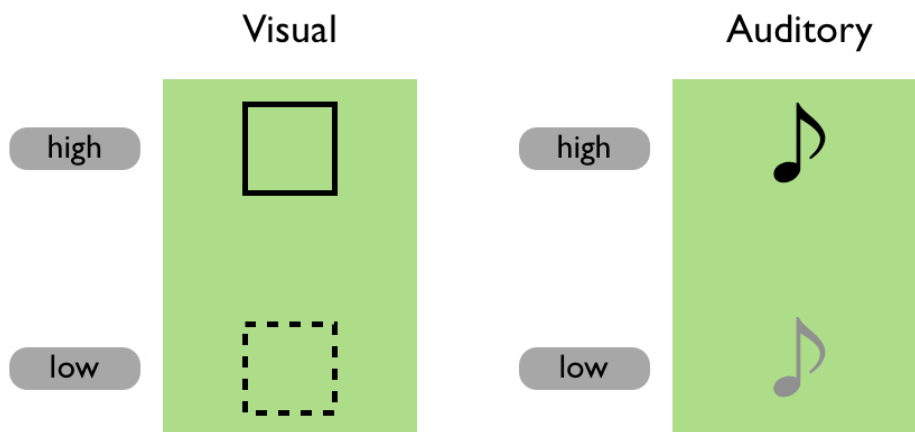


Figure 9: Experiment 10. Experimental setup for making predictions of a binary sequence. The sequence determined the positions of a square in the visual conditions (left), or the frequency of a sound in the auditory conditions (right).

5.1.2 Results and discussion

For every participant, we computed the prediction accuracy (number of correct predictions divided by the total number of predictions). The results were shown in Figure 10. In all conditions but one (visual alternate), the prediction accuracy decreased as x increased. Prediction accuracy was above chance (50%) for $x = 0.6, 0.7$, and 0.8 for the visual repeat and auditory alternate conditions. This suggests that the more random the sequence is, the less likely one is to predict the next bit.

We also ran a three-way mixed-design ANOVA: visual vs. auditory (between-subjects factor), repeat vs. alternate (between-subjects factor), and the five levels of x (within-subjects factor). Overall, prediction accuracy was marginally higher for repeat conditions (mean = 60.9%) than for alternate conditions (mean = 58.6%, $F(1,56) = 3.53, p = .06$). This suggests that overall, repeat(x) sequences were easier to predict compared to alternate(x) sequences. There was also a

main effect for x levels [$F(4,224) = 20.08, p < .001$]. This may be due to the fact that sequences of low levels of x were easier to predict than those of high x . There was a significant three-way interaction [$F(4,224) = 2.55, p = .04$]. There were no other significant main effects or interactions [$p > .15$].

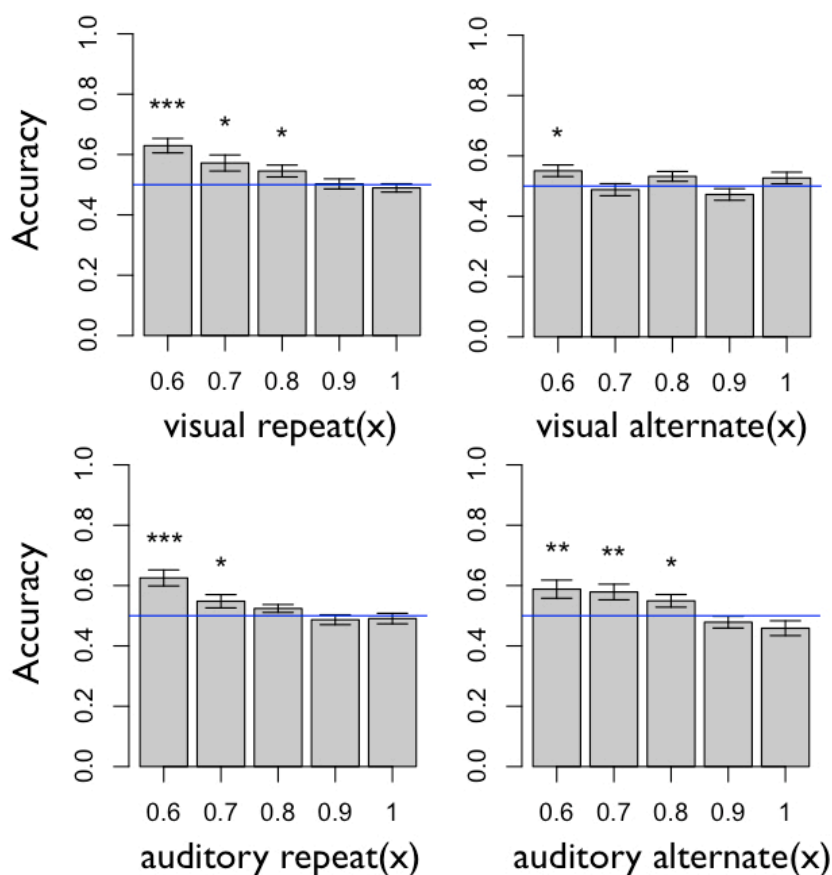


Figure 10: Results of Experiment 10. Prediction accuracy for each x level in the four conditions. The blue line indicates chance performance (0.5). Error bars indicate ± 1 SEM. Asterisks indicate significant difference from chance performance (* $p < .05$, ** $p < .01$, *** $p < .001$).

5.2 Experiment 11: Time estimation of random vs. non-random matrices

Another task measuring sensitivity to randomness is time estimation. It's been found that stimulus complexity can influence perceived duration of the stimulus (see Levin & Zakay, 1989, for a review). Specifically, the judged duration of a stimulus increases with the complexity of the stimulus (e.g., Ornstein, 1969). In this experiment, we investigate how perceptual complexity

(randomness) influences judgments about time. No mention was made of randomness or related concepts in the experiment.

5.2.1 Method

Participants: Ten undergraduates (6 female, mean age = 20.5 years) from Princeton University participated for course credit.

Stimuli and procedure: For every trial, one matrix consisting of 20×20 green and blue cells was presented in the middle of the screen. The matrix was presented for n seconds (n varied randomly between 1s and 6s across trials). The matrix was constructed using a $\text{switch}(x)$ sequence (the $\text{switch}(x)$ algorithm described in Chapter 2). As a reminder, the x in $\text{switch}(x)$ represents the switch rate of the sequence, ranging from 0 to 1—with 0 being complete streaks (repeats), 0.5 being fully random, and 1 being complete alternations. For every trial, x was randomly determined between 0 and 1. After the matrix, the participant was presented with some instruction. For 50 trials, they were asked to estimate the duration of the matrix (i.e., how long the matrix was presented on the screen). The participant was told to first press the space bar, upon which a gray circle appeared on the screen. The participant could press the space bar for the second time, and the gray circle would disappear. The gap between the two presses was recorded. The goal was to match duration of the gray circle to the duration of the matrix.

To ensure that participants paid attention to the matrix, 10 more trials were administered. For each of these trials, a matrix appeared on the screen as before. Then the participant was presented with two matrices, one on the left and one on the right side of the screen. One matrix was identical to the one just shown, and the other was the same except 50 cells in the matrix

(randomly chosen) switched colors. The task was to identify which matrix the participant had just seen by pressing the left (1) key or the right (0) key. These 10 trials were excluded from the analysis. In total, each participant completed 60 trials in a random order.

5.2.2 Results and discussion

For every trial, we computed four measures: the error in duration judgment (estimated duration – actual duration), relative error ((estimated duration – actual duration)/actual duration), absolute error (the absolute difference between estimated and actual durations), and relative absolute error (the absolute difference divided by actual duration).

Figure 11 shows all 500 data points collapsed across participants (10 participants $\times$ 50 trials per participant) for each of the four errors as a function of actual duration of the matrix. To quantify the relationship, for every participant we correlated each of the four errors with the actual duration of the matrix. Across 10 participants, the average correlations were: $r = -0.63$ (SD = 0.19) between error and time was, $r = -0.31$ (SD = 0.25) between relative error and time, $r = 0.63$ (SD = 0.26) between absolute error and time, and $r = 0.16$ (SD = 0.30) between relative absolute error and time. These results suggest that the longer the matrix was presented, the more likely the participant was to underestimate the duration, and the greater the extent of underestimation it was.

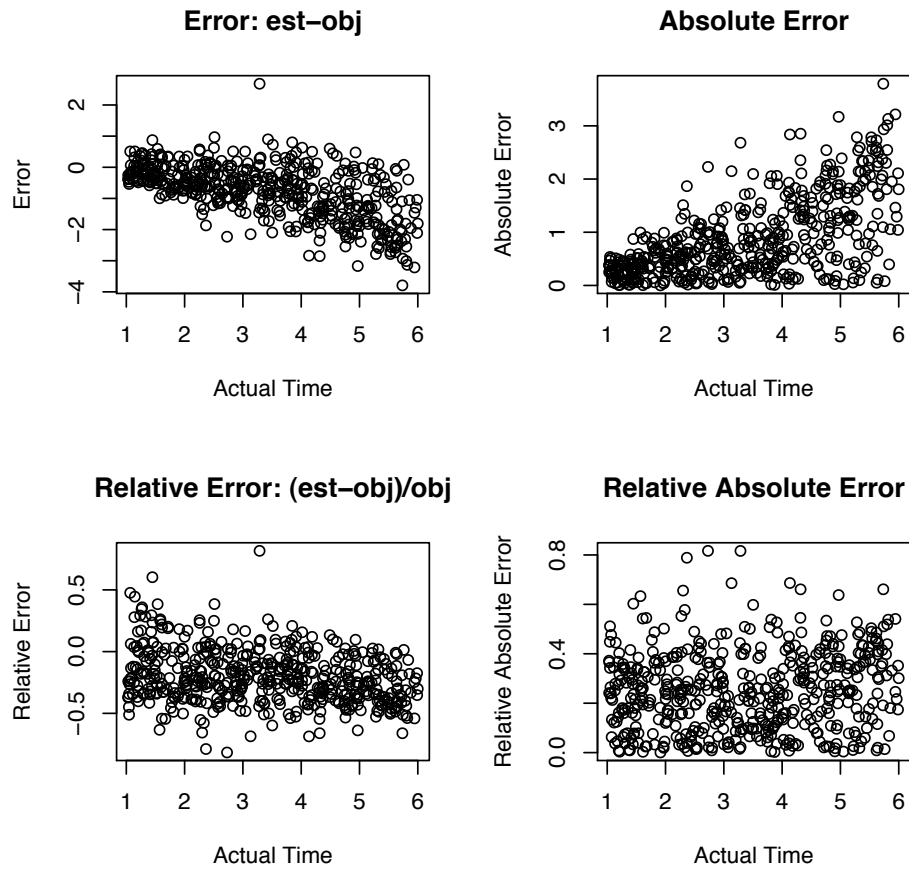


Figure 11: Experiment 11. Four types of judgment errors as a function of actual (i.e., objective) time of the matrix which varied from 1s to 6s. Each graph presented a total of 500 data points collapsed across participants. The error was the difference between estimated and objective durations of the matrix; the relative error was the error divided by the objective duration; absolute error was the absolute difference between estimated and objective durations of the matrix; and relative absolute error was the absolute difference divided by the objective duration.

To examine the relationship between estimation of matrix duration and the randomness of the matrix, Figure 12 shows all 500 data points collapsed across participants for each of the four errors as a function of the switch rate of the matrix. For every participant, we correlated each of the four errors with the switch rate of the matrix. Across 10 participants, the average correlations were: $r = -0.03$ ($SD = 0.16$) between error and switch rate was, $r = -0.04$ ($SD = 0.15$) between relative error and switch rate, $r = -0.04$ ($SD = 0.13$) between absolute error and switch rate, and $r = -0.06$ ($SD = 0.07$) between relative absolute error and switch rate. These results suggest there

was a weak relationship between people's estimation of duration and the randomness of the matrix. Further studies were needed to clarify such relationship.

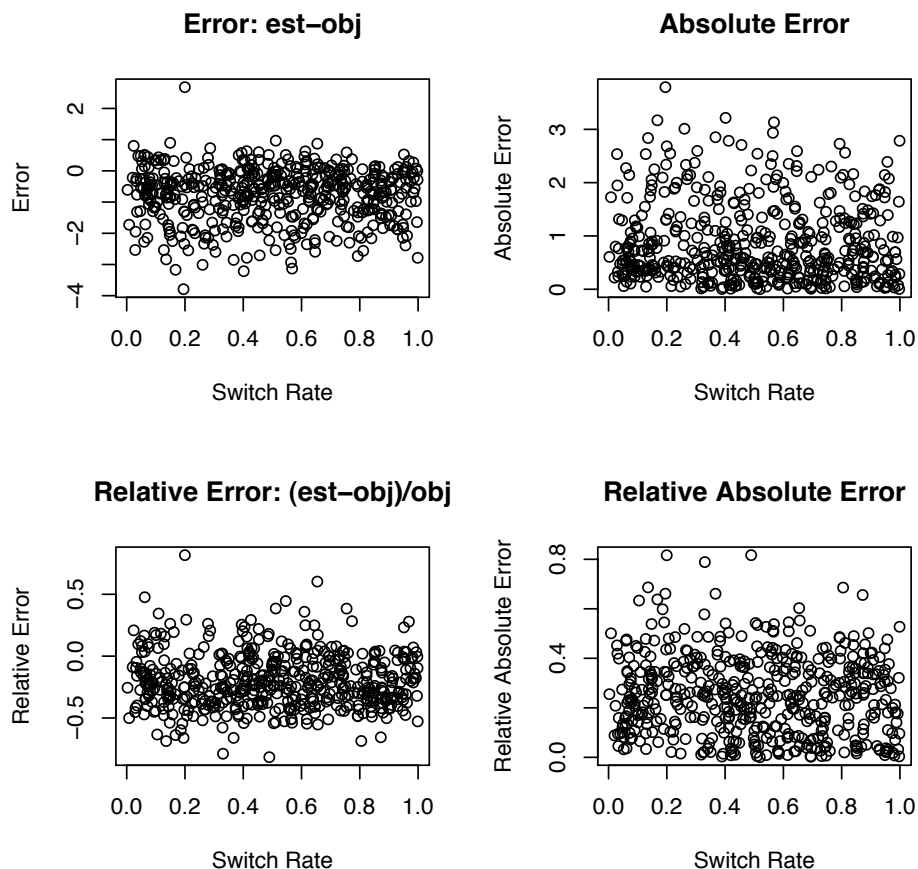


Figure 12: Experiment 11. Four types of judgment errors as a function of switch rate of the matrix which varied from 0 to 1. Each graph presented a total of 500 data points collapsed across participants.

5.3 Experiment 12: Scene perception with random vs. non-random masks

This experiment aimed to examine the ability to perceive natural scenes, which are covered by random or non-random masks. In other words, how perception of natural scenes varies as a function of the degree of non-randomness of the masks. No mention was made of randomness or related concepts in the experiment.

5.3.1 Method

Participants: Thirteen undergraduates (8 female, mean age = 19.8 years) from Princeton University participated for course credit.

Stimuli and procedure: There were 10 categories of natural scenes: bar, bedroom, castle, garden, gym, kitchen, library, playground, seaport, and street. All exemplars in each category were black and white and had the same brightness and contrast. For the mask, there were 10 levels of switch rate x , from 0.1 to 1 by steps of 0.1. For each level of x and for each category, three distinct exemplars were used. This resulted in a total of 300 distinct exemplars (10 categories $\times$ 10 x levels $\times$ 3 exemplars) for every participant. The order of the exemplars was randomized. Each exemplar was covered by a mask. A mask contained 64×64 cells, and was generated by tiling a $\text{switch}(x)$ sequence, either vertically or horizontally. Specifically, if the bit is 1, the cell was opaque and was colored blue; if the bit is 0, the cell was transparent and revealed the underlying exemplar image.

For a trial, the exemplar with the mask was presented for 300ms, after which the labels for the 10 categories were presented on the screen. The response keys (1 to 10) were also listed below the category labels, each corresponding to a category. The participant was asked to judge the category of the exemplar by typing a number key corresponding to that category.

Ten participants completed the experiment. As a baseline measure for scene category perception, another three participants completed the same experiment, except here the exemplars were never covered by masks.

5.3.2 Results and discussion

We first computed the accuracy of categorization for participants who viewed exemplars without masks as a baseline measure. The average accuracy was 93.4% (SD = 3.7%), which was sufficient to make the experimental task meaningful (see Figure 13). We then computed the accuracy for participants who viewed masked exemplars. Across all switch rates, the average accuracy was 67.1% (SD = 8.3%), reliably lower than that without masks [$t(11) = 5.23, p < .001$].

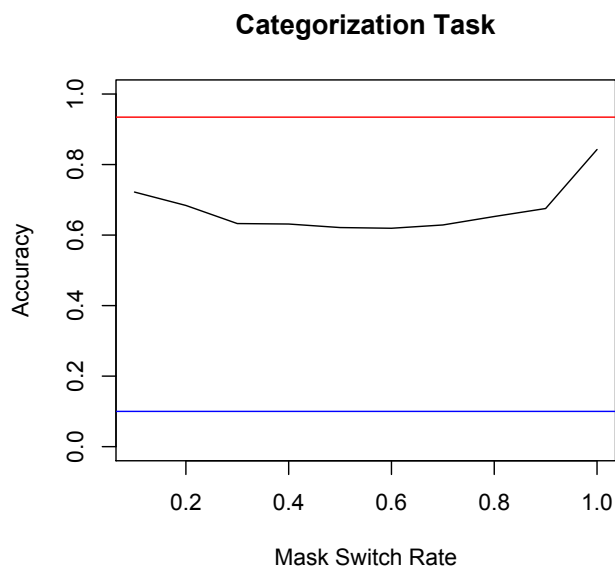


Figure 13: Experiment 12. Categorization of natural scenes. The black line indicates average performance of categorization over switch rates. The red line shows the performance of categorization of exemplars without masks. The blue line indicates chance performance (10%, 1 out of 10 categories).

The goal of the experiment was to see how scene perception varies with the randomness of the masks. To address this question, we analyzed the categorization performance for each participant and for every level of switch rate, collapsing across 10 scene categories. The average categorization performance across participants for each switch rate was plotted in Figure 13.

A one-way within-subjects ANOVA revealed a reliable difference in the performance among the 10 levels of switch rate [$F(9) = 6.75, p < .001$]. Planned comparisons showed that the performance at $x = 0.1$ was reliably higher than those at x from 0.3 to 0.7 [$t's(9) > 2.30, p's < .05$]. Moreover, performance at $x = 1$ was reliably higher than that for all the other x 's [$t's(9) > 5.30, p's < .001$]. These results suggest that as the masks become less random (either more streaky or switchy), it was easier to categorize the scenes. Future studies are needed to address the underlying mechanism.

5.4 Chapter summary

As implicit measures of randomness, three experiments examined the cognitive consequences of randomness without explicitly invoking the notion of randomness. Experiment 10 tested the ability to predict a random or a non-random binary sequence, and we found that the prediction accuracy decreased as the sequence became more random. Experiment 11 examined how perceptual complexity of the matrix influences judgments about duration. We failed to find a strong relationship between the degree of randomness and the errors in estimation of duration. Experiment 12 tested the ability to categorize natural scenes covered by random or non-random masks. We found that the ability of categorizing natural scenes improved as the masks became less random. Overall, these experiments implied that varying levels of randomness can impact cognition without explicitly eliciting the notion of randomness.

Chapter 6

Concluding remarks

6.1 Summary of results

The current work aims to provide a comprehensive investigation of randomness cognition by examining the perception and conception of randomness. In Chapter 2, we found that the ability to distinguish random from non-random static matrices (Experiment 1) or dynamic walks (Experiment 2) was superior to the ability to identify which stimuli were random. Interestingly, when feedback was provided, identification performance improved to the level of discrimination, while discrimination performance did not change with feedback. These results further highlight the difference between discrimination and identification of randomness. We further examined the encoding hypothesis (Experiment 3), which states that the tendency of a stimulus to be labeled random can be predicted by the cognitive difficulty of encoding it. Using mirror matrices, we found that these matrices were more difficult to encode and yet they were less likely to be labeled random, compared to other non-random matrices. This finding is in contradiction with the encoding hypothesis.

Chapter 3 involved an effort to replicate and extend the findings using different algorithms to generate non-random stimuli and a staircasing method to gauge performance.

Consistent with the findings of Experiment 1, we found the ability to distinguish random from non-random matrices to be superior to the ability to identify which matrix was random in the presence of non-random distracters (Experiment 4). Moreover, we found that the ability to distinguish repeat matrices (Experiment 4) or walks (Experiment 5) from random was reliably better than the ability to distinguish alternate matrices from random. When learning to classify matrices into nominal categories, participants were better at classifying repeat versus random matrices compared to alternate versus random (Experiment 6). This could be due to the possibility that repeat stimuli were easier to encode than alternate stimuli (Experiment 7).

We examined the ability to produce stimuli as random in Chapter 4. Consistent with prior research on randomness production, we found that the sequences produced by participants as random were biased towards alternations. Interestingly, these productions were more biased towards alternations than sequences judged as random, or sequences that could be distinguished from truly random ones (Experiment 8). By comparing performances within the same individual, we found that the ability to produce random sequences seems not to be related with the ability to distinguish random from non-random stimuli, or the ability to identify the random stimuli (Experiment 9).

In Chapter 5, three experiments were designed to investigate implicit measures of randomness, exploring the cognitive consequences of randomness without explicitly invoking the notion of randomness. We found that the ability to predict bits in a sequence worsened as the sequence became more random (Experiment 10). We did not find a relationship between the ability to estimate duration of the matrix and the degree of non-randomness of the matrix (Experiment 11). Finally, the ability of perceiving natural scenes improved as the masks covering the scenes became less random (Experiment 12).

6.2 General discussion

Our work is motivated by the fact that randomness cognition has both perceptual and conceptual components. Perceptually, one has some ability to distinguish random from non-random stimuli. Conceptually, one is sometimes capable of identifying which stimulus looks random. To fully understand randomness cognition, we need to consider both processes. A natural hypothesis is that if you can see the difference between a random and a non-random source, you can say which is random. Our findings suggest otherwise. The ability to distinguish between random and non-random stimuli is superior to the ability to identify which stimulus looks random. In Experiments 1 and 2, participants were not informed of the generating process, but nonetheless performed reasonably close to an ideal observer. Discrimination accuracy was symmetrically distributed around genuine randomness. That is, the probability of distinguishing a stimulus with switch rate $.5 + \delta$ was about equal to the probability for $.5 - \delta$ (see Figure 1). In contrast, identification was worst (and reliably below chance) for switch rate 0.56 in Experiment 1. Remarkably, it was worst (and far below chance) at essentially the same switch rate (0.58) in Experiment 2, involving radically different stimuli. This suggests that the lay concept of randomness does not fully exploit the perceptual resources available for discriminating random from non-random sources. In other words, the faulty human concept of randomness is not due to an inability to perceive randomness.

A different hypothesis was put forth by Falk and Konold (1997), which suggests that the tendency to label given stimuli as random can be predicted by the encoding difficulty of these stimuli. We found, on the contrary, that there was a certain type of stimuli (mirror matrices) that were more difficult to encode, and yet less likely to be labeled random. To explain this

contradiction, note that success in the encoding task requires local processing, namely, to determine whether a given cell in the matrix has changed bits. In contrast, judgments of “randomness” need to integrate the global feature of matrix symmetry, which strikes participants as non-random. Indeed, recent studies have highlighted the difference between global and local processing (Poljac et al., 2012; Zhao et al., 2011). Thus, the encoding hypothesis due to Falk and Konold (1997) does not appear to generalize to detecting changes in bit matrices. Future studies are needed to identify predictors of randomness judgments in addition to perceptual measures.

Another interesting aspect of randomness perception is that it is easier to distinguish repeat stimuli from random ones than to distinguish alternate stimuli from random. This difference could be explained by the ease of encoding of repeat stimuli compared to alternate stimuli. To further explain differences in encoding, it is tempting to invoke the perceptual difference between repeats/streaks and alternations/checkerboards. The former, typically thought of as contours, are central to object recognition, whereas the latter seem less common to our visual experience. The perceptual system might therefore be more sensitive to streaks, able to process them better, and detect their relative absence from random stimuli.

To test this idea, we examined whether streaks are more common in our visual environment. We collected 2800 natural scene images [<http://cvcl.mit.edu/database.htm>]. Each image was converted to a black-white (binary) scale with the boundary at the mean brightness value. The images remain remarkably interpretable in the face of such alteration. Alternation rate was computed by traversing horizontally through columns or vertically through rows. The average switch rates for horizontal and vertical traversals were 0.07 and 0.08, resulting in corresponding repeat(x) values of 0.14 and 0.16, respectively. Such values support the claim that our visual environment is highly streaky. This line of explanation, however, is insufficient in

view of the results of Experiment 5 (random vs. non-random walks) whose stimuli were presented temporally. Further investigation is obviously needed. Another question also merits inquiry: What is the relationship between (a) the threshold difference for repeat and alternate stimuli when discriminating them from randomness, and (b) the well documented finding that over-alternating stimuli are more likely to be labeled as random?

Several researchers have proposed a different explanation for the repeat vs. alternate discrepancy (Hahn & Warren, 2009; Sun & Wang, 2010). They note that when observing a random source, the waiting time for switch sequences tends to be less than that for streak sequences. For example, the probability of observing the consecutive bits 101010101 prior to 111111111 is 0.6. (The calculation can be made using an algorithm developed by John Conway; see Gardner, 2001.) This tendency might ultimately be assimilated through experience with random sources, and then mistakenly taken as evidence that the former string is more likely than the latter to be constructed starting from a given bit (like the first bit seen). The greater attraction of alternation compared to repetition might thereby find an explanation. Whether this idea can be extended to account for the other phenomena reported here remains to be seen.

Another comparison between the perception and conception of randomness draws on the ability to produce stimuli as random. We found that sequences produced as random were more biased towards alternations than sequences which were judged as random. In other words, the bias for alternations was more pronounced in productions of randomness. Moreover, we found a lack of correlations between discrimination, identification, and production performances, suggesting that the production of random stimuli seems to be unrelated to the perception or identification of randomness. Future studies need to identify the processes underlying the

production of random stimuli and how they are different from those involved in the perceptual or judgment tasks.

Finally, we suggest that there may be a rich set of implicit measures examining randomness cognition other than the more explicit judgment or production tasks. In life, it is perhaps more important and more common that we respond to random or non-random stimuli than we understand the word “random,” a theoretical term that itself is extremely complex. We thus tested three paradigms which did not involve the concept of randomness. It is worth noting that during debriefing, all participants in these tasks reported that they were not aware that the stimuli were random, and they did not know about the purpose of the experiments. Although subjects’ performance on these tasks was partly modulated by the degree of non-randomness, these paradigms need to be fine-tuned in future studies.

6.3 Conclusion

Although much past work has examined the lay conception of randomness in various ways, the present work provides a more nuanced assessment of randomness cognition. Importantly, we distinguish between perceptual and conceptual understanding of randomness. The ability to distinguish random from non-random stimuli was superior to the ability to identify which stimuli were random, and the ability to produce stimuli as random. The discrepancy between discrimination and identification performances cannot be explained by the encoding difficulty of the stimuli.

Bibliography

- Attneave, F. (1959). *Applications of Information Theory to Psychology: A Summary of Basic Concepts, Methods, and Results*. New York: Holt, Rinehart & Winston.
- Ayton, P., Hunt, A. J., & Wright, G. (1989). Psychological conceptions of randomness. *Journal of Behavioral Decision Making*, 2, 221–238.
- Baddeley, A. D. (1966). The capacity for generating information by randomization. *Quarterly Journal of Experimental Psychology*, 18, 119–129.
- 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.
- Beltrami, E. (1999). *What Is Random? Chance and Order in Mathematics and Life*. New York: Springer-Verlag.
- Blum, M. (1967). On the size of machines. *Information and Control*, 11, 257–265.
- Brown, G. S. (1957). *Probability and Scientific Inference*. London: Longmans Green.
- Budescu, D. V. (1987). A markov model for generation of random binary sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 25–39.
- Camerer, C. F. (2003). Strategizing in the brain. *Science*, 300, 1673–1675.
- Chaitin, G. J. (1975). Randomness and mathematical proof. *Scientific American*, 232, 47–52.
- Chapanis, A. (1953). Random number guessing. *American Psychologist*, 8, 332.

- Chater, N. (1999). The search for simplicity: A fundamental cognitive principle? *Quarterly Journal of Experimental Psychology*, 52, 273–302.
- Diener, D., & Thompson, W. B. (1985). Recognizing randomness. *American Journal of Psychology*, 98, 433–447.
- Dorris, M. C., & Glimcher, P. W. (2004). Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron*, 44, 365–378.
- Eagle, A. (2012). Chance versus randomness. In Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/win2003/entries/chance-randomness/>
- Earman, J. (1986). *A Primer on Determinism*. Dordrecht Holland: D. Reidel.
- Edwards, W. (1961). Probability learning in 1000 trials. *Journal of Experimental Psychology*, 62, 385–394.
- Edwards, W. (1968). Conservatism in human information processing. In Kleinmuntz, B. (ed.) *Formal Representation of Human Judgment*. New York: Wiley.
- 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*, pp. 405–406.
- Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review*, 104, 301–318.
- Fitelson, B., & Osherson, D. (2012). Remarks on random sequences.
<http://arxiv.org/abs/1205.5865>
- Gardner, M. (2001). *The Colossal Book of Mathematics*. New York: W. W. Norton.

- Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, 17, 295-314.
- Glimcher, P. W. (2005). Indeterminacy in brain and behavior. *Annual Review of Psychology*, 56, 25-56.
- Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, 49, 585-612.
- Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In Akins, K. (ed.) *Vancouver Studies in Cognitive Science: Vol. 2: Perception*. Oxford: Oxford University Press.
- Hahn, U., & Warren, P. A. (2009). Perceptions of randomness: Why three heads are better than four. *Psychological Review*, 116, 454-461.
- Hsu, A. S., Griffiths, T. L., & Schreiber, E. (2010). Subjective randomness and natural scene statistics. *Psychonomic Bulletin & Review*, 17, 624-629.
- Johnson, J., & Tellis, G. J. (2005). Blowing bubbles: Heuristics and biases in the run-up of stock prices. *Journal of the Academy of Marketing Science*, 33, 486-503.
- Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature*, 290, 91-97.
- 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.
- Kolmogorov, A. N. (1965). Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 1, 1-7.

- Landy, M. S., & Graham, N. (2004). Visual perception of texture. In L. M. Chalupa, & J. S. Werner, (Eds.), *The Visual Neurosciences* (pp. 1106-1118). Cambridge, MA: MIT Press.
- Levin, I., & Zakay, D. (1989). *Time and Human Cognition: A Life-Span Perspective*. New York: Elsevier.
- Li, M., & Vitányi, P. (2008). *An Introduction to Kolmogorov Complexity and Its Applications (2nd Edition)*. New York: Springer.
- 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.
- Michalczuk, R., Bowden-Jones, H., Verdejo-Garcia, A., & Clark, L. (2011). Impulsivity and cognitive distortions in pathological gamblers attending the UK National Problem Gambling Clinic: A preliminary report. *Psychological Medicine*, 41, 2625–2635.
- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9, 353–383.
- Neuringer, A. (1986). Can people behave randomly? The role of feedback. *Journal of Experimental Psychology: General*, 115, 62–75.
- Nickerson, R. S. (2002). The production and perception of randomness. *Psychological Review*, 109, 330–357.
- Nickerson, R. S., & Butler, S. F. (2009). On producing random binary sequences. *American Journal of Psychology*, 122, 141–151.

- Olivola, C. Y., & Oppenheimer, D. M. (2008). Randomness in retrospect: Exploring the interactions between memory and randomness cognition. *Psychonomic Bulletin & Review*, 15, 991–996.
- Ornstein, R. E. (1969). *On the Experience of Time*. Middlesex: Harmondsworth Penguin.
- Osherson, D., & Weinstein, S. (2011). Notes on random reals. <http://arxiv.org/abs/1209.2875>.
- Oskarsson, A. T., van Boven, L., McClelland, G. H., & Hastie, R. (2009). What's next? Judging sequences of binary events. *Psychological Bulletin*, 135, 262–285.
- Pevtsov, R., & Harnad, S. (1997). Warping similarity space in category learning by human subjects: The role of task difficulty. In Ramscar, M., Hahn, U., Cambouropoulos, E., and Pain, H. (Eds.) *Proceedings of SimCat 1997: Interdisciplinary Workshop on Similarity and Categorization*. Edinburg: Edinburgh University.
- Poljac, E., de Wit, L., & Wagemans, J. (2012). Perceptual wholes can reduce the conscious accessibility of their parts. *Cognition*, 123, 308–312.
- Rabin, M. (2002). Inference By Believers in the Law of Small Numbers. *Quarterly Journal of Economics*, 117, 775–816.
- Rapoport, A., & Budescu, D. V. (1992). Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology: General*, 121, 352–363.
- Rapoport, A., & Budescu, D. V. (1997). Randomization in individual choice behavior. *Psychological Review*, 104, 603–617.
- Reichenbach, H. (1934/1949). *The Theory of Probability: An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability*. Berkeley: University of California Press.
- Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, 53, 245–277.

- Reuter, J., Raedler, T., Rose, M., Hand, I., Galscher, J., & Buchel, C. (2005). Pathological gambling is linked to reduced activation of the mesolimbic reward system. *Nature Neuroscience*, 8, 147–148.
- Ross, B. M., & Levy, N. (1958). Patterned prediction of chance events by children and adults. *Psychological Reports*, 4, 87–124.
- Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition*, 7, 1–15.
- Steeves, T. D. L., Miyasaki, J., Zuerowski, M., Lang, A. E., Pellecchia, G., Van Eimeren, T., Rusjan, P., Houle, S., & Strafella, A. P. (2009). Increased striatal dopamine release in Parkinsonian patients with pathological gambling: a [11C] raclopride PET study. *Brain*, 132, 1376–1385.
- Sun, Y., & Wang, H. (2010). Perception of randomness: On the time of streaks. *Cognitive Psychology*, 61, 333–342.
- Terrell, D. (1998). Biases in assessments of probabilities: New evidence from greyhound races. *Journal of Risk and Uncertainty*, 17, 151–166.
- Toneatto, T., Blitz-Miller, T., Calderwood, K., Dragonetti, R., & Tsanos, A. (1997). Cognitive distortions in heavy gambling. *Journal of Gambling Studies*, 13, 253–266.
- Tune, G. S. (1964). Response preferences: A review of some relevant literature. *Psychological Bulletin*, 61, 286–302.
- Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76, 105–110.
- van der Helm, P. A., & Leeuwenberg, E. L. J. (1996). Goodness of visual regularities: A nontransformational approach. *Psychological Review*, 3, 429–456.
- Wagemans, J. (1995). Detection of visual symmetries. *Spatial Vision*, 9, 9–32.

- Wagenaar, W. A. (1970). Appreciation of conditional probabilities in binary sequences. *Acta Psychologica*, 34, 348–356.
- Wagenaar, W. A. (1972). Generation of random sequences by human subjects: A critical survey of the literature. *Psychological Bulletin*, 77, 65–72.
- Wagenaar, W. A. (1988). *Paradoxes of Gambling Behaviour*. Hillsdale, NJ: Erlbaum.
- Walker, J. (2006). Hotbits: Genuine random numbers, generated by radioactive decay.
<http://www.fourmilab.ch/hotbits/how3.html>.
- Weiss, R. L. (1964). On producing random sequences. *Psychological Reports*, 14, 931–941.
- Zhao, J., Ngo, N., McKendrick, R., & Turk-Browne, N. (2011). Mutual interference between statistical summary perception and statistical learning. *Psychological Science*, 22, 1212–1219.