Journal of Experimental Psychology:
Human Perception and Performance
2014, Vol. 40, No. 4, 1358–1371

0096-1523/14/$12.00 http://dx.doi.org/10.1037/a0036816

Perception and Identification of Random Events

Jiaying Zhao
University of British Columbia

Ulrike Hahn
Birkbeck College, University of London

Daniel Osherson
Princeton University

The cognition of randomness consists of perceptual and conceptual components. One might be able to discriminate random from nonrandom stimuli, yet be unable to identify which is which. In a series of experiments, we compare the ability to distinguish random from nonrandom stimuli to the accuracy with which given stimuli are identified as "random." In a further experiment, we also evaluate the *encoding hypothesis* according to which the tendency of a stimulus to be labeled *random* varies with the cognitive difficulty of encoding it (Falk & Konold, 1997). In our experiments, the ability to distinguish random from nonrandom stimuli is superior to the ability to correctly label them. Moreover, for at least 1 class of stimuli, difficulty of encoding fails to predict the probability of being labeled random, providing evidence against the encoding hypothesis.

Keywords: randomness, perception, texture, alternation bias

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, in the environment (Bar-Hillel & Wagenaar, 1991). Pavlovian conditioning, for example, depends on the distinction between random and nonrandom pairings of conditioned and unconditioned stimuli (Rescorla & Wagner, 1972). Likewise, the successful isolation of contingent relations between stimuli, distinguished from random co-occurrence, is essential to language acquisition (see Kelly & Martin, 1994, for a review). In evaluating the probabilities of uncertain events, people often exhibit biases in their judgments, which stem from their erroneous beliefs about randomness (Kahneman & Tversky, 1972; Gilovich, Vallone, & Tversky, 1985; Tversky & Kahneman, 1971). In naturalistic settings, people are capable of extracting information from data in the financial market, and they can distinguish real market returns from randomized returns (Hasanhodzic, Lo, & Viola, 2010). Perceptions of randomness and chance have been suggested to underlie economic behaviors such as investing in stocks, making bets in prediction markets, and gambling (e.g., Dreman, 1977; Ladouceur, Sylvain, Letarte, Giroux, & Jacques, 1998; Manski, 2006; Michalczuk, Bowden-Jones, Verdejo-Garcia, & Clark, 2011). Experience with randomness may also influence religious beliefs (Kay, Moscovitch, & Laurin, 2010). Of course, many of the statistical tests that psychologists rely on are intended to discover systematicity in the data, instead of pure randomness.

More generally, randomness has been argued to be important to both humans and nonhuman 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 of that ability. Mistaking random events as nonrandom can lead to suboptimal behaviors—as illustrated in gambling problems (e.g., Reuter et al., 2005; Steeves et al., 2009; Terrell, 1998; Toneatto, Blitz-Miller, Calderwood, Dragonetti, & Tsanos, 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, for example, evade predators or thwart competition in the game theoretic context (Camerer, 2003; Dorris & Glimcher, 2004; Glimcher, 2005; Rapoport & Budescu, 1992).

Prior studies have offered a negative verdict on people's understanding of randomness, finding that misconceptions such as the gambler's fallacy and the overalternation bias are widespread (Kahneman & Tversky, 1972; Wagenaar, 1972; but see Nickerson & Butler, 2009; for reviews see Bar-Hillel & Wagenaar, 1991; Oskarsson, van Boven, McClelland, & Hastie, 2009). But research on randomness has been plagued by conceptual difficulties, which have often undermined empirical results (Ayton, Hunt, & Wright, 1989; Lopes, 1982; Nickerson, 2002). Notably, inspecting a pattern does not allow its qualification as random or nonrandom: A random source could produce a thousand 0s in a row and a

This article was published Online First May 12, 2014.

Jiaying Zhao, Department of Psychology, Institute for Resources, Environment and Sustainability, University of British Columbia; Ulrike Hahn, Department of Psychological Sciences, Birkbeck College, University of London; Daniel Osherson, Department of Psychology, Princeton University.

The research was conducted primarily at Princeton University, where the first author was a graduate student. We acknowledge funding from the Henry Luce Foundation.

Correspondence concerning this article should be addressed to Jiaying Zhao, Department of Psychology, Institute for Resources, Environment and Sustainability, University of British Columbia, 2329 West Mall, Vancouver, BC, V6T 1Z4 Canada. E-mail: jiayingz@psych.ubc.ca

nonrandom 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.

Research on randomness has focused on two types of processes: random sequence judgment (e.g., Kahneman & Tversky, 1972) and random sequence generation (e.g., Wagenaar, 1972), with some consistency of results between the two (Bar-Hillel & Wagenaar, 1991). What has been given little attention in either of these contexts, however, is the fact that there are potentially very different ways in which these challenges may be met. 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 randomness includes both perceptual and conceptual components. Perceptually, people have some ability to distinguish random from nonrandom stimuli. Conceptually, they may be capable of identifying which stimulus looks "random"—for example, by correctly recognizing which of two stimuli was randomly generated, where only one was. These are separate abilities because one might be able to discriminate random from nonrandom 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. Indeed, the distinction between discrimination and identification has been shown using many types of stimuli, including vowels (Fry, Abramson, Eimas, & Liberman, 1962), pitch (Houtsma & Smurzynski, 1990), speech (Godfrey, Syrdal-Lasky, Millay, & Knox, 1981), motion (Tan, Srinivasan, Reed, & Durlach, 2007), and objects (Kunst-Wilson & Zajonc, 1980; Overman, Bachevalier, Turner, & Peuster, 1992). Past research on randomness, however, has only focused on identification in two ways: generation of "random" strings, and evaluation of given strings as "random" (Bar-Hillel & Wagenaar, 1991). There has been much less work on the perceptual ability to distinguish random from nonrandom stimuli. The primary goal of the present work is thus to examine the relationship between the perceptual and conceptual sides of randomness cognition. This will also offer insights on whether people's poor understanding of randomness is driven by a perceptual difficulty in distinguishing random from nonrandom events.

Methodological Overview

There has been considerable methodological variation in studies on randomness. Even within the early literature that focused largely on the generation of random sequences by participants, the specific details of the task varied from study to study. The variation concerned, for example, the number of possible outcomes, length of the sequence to be generated, degree to which the sequence remains accessible to participants, as well as the tests for randomness applied in analysis. These methodological differences led Wagenaar (1972) to conclude that there was no way of combining the results into one coherent theory (p. 69). In addition, researchers have employed not only 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 to examine people's understanding of randomness, though judgment tasks have been the most prevalent.

In the context of judgment tasks, there has also been considerable variation with regard to the materials used; in particular, how they were generated and what kind of random process they were taken to represent. Many studies have focused on unbiased coins as the prototypical random process, and presented participants with sequences of coin tosses. The present studies make use of two very different kinds of stimuli: randomly tiled visual arrays (Julesz, 1962) and dynamic random walks.

The theoretical definition of randomness is difficult (Beltrami, 1999; Nickerson, 2002), because random sources can give rise to outputs that do not look random (e.g., uniform runs), and nonrandom sources can generate outputs that look random. As a consequence, there are definitions of randomness based on the nature of the generating process, and ones based on the nature of the output itself. In light of this, past studies have not always operationalized randomness in reasonable ways (for extensive critiques, see Ayton et al., 1989; Nickerson, 2002). For example, asking participants to generate a random sequence such as "one might see from an unbiased coin" requires participants to both mirror a random generating source and produce an output that looks random—although the two notions are not identical. Such instructions are at best ambiguous, and at worst incoherent; hence, it is unclear whether resultant errors and biases should be attributed to participants or experimental instructions.

Our terminology in this paper 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. Nonrandom stimuli are defined as productions from a distorted random source. It is important to note that 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.

Specifically, we used the following operationalization: Assuming a random source R and a nonrandom 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. Here we consider *identifying* a stimulus as random the same as *categorizing* or *classifying* the stimulus as random. Thus, identification, categorization, and classification are interchangeable.

Our rationale for comparing performance on the discrimination task and the identification task with the same stimuli is premised on the assumption that the discrimination task will measure the ability to perceive boundaries between two different textures without the need for classification or categorization of either texture (e.g., Julesz, 1981; Landy & Graham, 2004), whereas success in the identification task will require the formation of a category concerning the concept of randomness (e.g., Rosch, 1973; Smith & Medin, 1981).

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

> *Hypothesis 1:* The probability of correctly identifying stimuli from R and N coincides with the ease of distinguishing between the two sources.

An accurate conceptualization of randomness would ensure the truth of Hypothesis 1. For example, knowing Jane would lead to success not only in distinguishing Jane's face from someone else's, but also in identifying which face is Jane. If this does not hold, then it points to a flawed conceptual system for creating and maintaining a representation of Jane, as in the neuropsychological condition of prosopagnosia. In contrast, if the concept of randomness is inaccurate, then Hypothesis 1 is false. To illustrate, suppose that the participant is shown two lines and is asked either (a) "which is longer?", or (b) "which is exactly 6 in. in length?" Since humans do not have a well-developed conceptual system for specific absolute length measurements, we expect that the participant will do better at (a) than at (b). This would reveal that the inaccuracy of the "6 in." concept is not due entirely to discrimination failure. Likewise, for the abstract and difficult concept of randomness, we expect that discrimination performance does not fully align with identification performance. Our study precisely explores the relationship between discrimination and identification, as a way to understand the conceptualization of randomness.

Previous studies have shown that bit sequences 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). Thus, the accuracy of Hypothesis 1 may be compromised by people's erroneous beliefs about "random." 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 Hypothesis 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, Osherson, & Weinstein, 2006; Osherson & Weinstein, 2008.) Regarding "nonrandom," 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 nth 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 nth bit; with probability $1 - x$ the $n + 1$st bit of S is set equal to the nth 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(.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 1s (and of 0s) 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 .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 0011001100110011001100110 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 nth-order entropy for any n (for discussion, see Attneave, 1959).

To foreshadow the experiments in this paper, we first compare the discrimination versus identification of random and nonrandom 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. Our 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.

Experiment 1

Participants

Forty undergraduates (28 female, mean age 19.6 years, $SD = 0.7$) from Princeton University participated in exchange for course credit. In this and later experiments, participants provided informed consent (protocol approved by the Princeton IRB), and received course credit.

Materials

Stimuli were 60×60 matrices made up of green and blue dots. Each matrix (subtending $\sim$ 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 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).

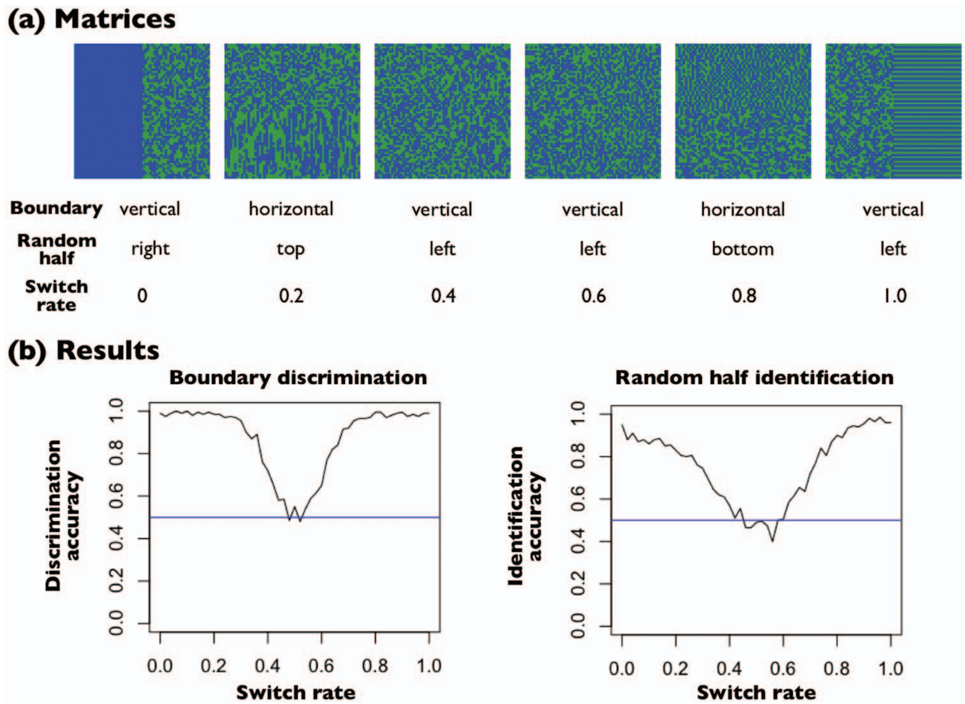

Figure 1. Experiment 1: (a) Six sample matrices (not drawn to scale). Each matrix consisted of two halves, one fully random, and the other with a given switch rate. The boundary between the two halves (vertical vs. horizontal) was randomly determined, as was the choice of random half. (b) Discrimination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

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 nonrandom process. Your task is to identify which half is more likely to be produced by a random process than a nonrandom process. Press the top or bottom arrow key if the division is horizontal, and left or right if vertical.

We note that both discrimination and identification tasks involve relative judgments. In the discrimination task, participants need to consider the horizontal and the vertical boundaries, and determine which boundary looks relatively more salient. In the identification task, participants need to determine which half looks relatively more random. Only the identification task relies on the participant's notion of randomness. The discrimination task does not elicit their conceptual knowledge, and in fact during debriefing, none of the participants was aware that the discrimination task was about randomness.

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., switch(.5)]. Each level was repeated 10 times, making up 510 trials. The order of the trials was randomized for every participant. For each trial, the matrix (or the two separate halves) appeared on the screen for 1,500 ms; the screen then remained blank until response. No feedback regarding judgment accuracy was provided.

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 nonrandom half of the matrix became more random (i.e., switch rate closer to .5), both discrimination and identification accuracy decreased. The correlation between discrimination and identification accuracy across all levels of switch rate was $r(49) = 0.91$, $p < .001$. The correlation 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, $p < .001$.

A two-way mixed design ANOVA was conducted (between-subjects factor: discrimination vs. identification condition; within-subject factor: 51 levels of switch rates). There was a main effect for condition [$F(1, 38) = 25.43$, $p < .001$, $\eta_p^2 = .40$], and a main effect for switch rate [$F(50, 1900) = 55.78$, $p < .001$, $\eta_p^2 = .59$]. There was also a reliable interaction between condition and switch rate [$F(50, 1900) = 2.72$, $p < .001$, $\eta_p^2 = .07$]. For each switch rate, we obtained the average performance of participants in each condition. Across 51 levels of 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, $t(38) > 2.12$, $p < .05$].

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 nonrandom stimuli but were at chance at identifying 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 nonrandom 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(.56) matrices as more random than truly random matrices whereas discrimination performance for the same switch rate was reliably above chance.[1] Once again, these results are contrary to Hypothesis 1. More generally, Figure 1b shows that the identification curve was not symmetrical. Rather, it was shifted to the right, with the lowest point at $x = 0.56$. This confirms the alternation bias. However, for discrimination, the curve was symmetrical in general and was not biased toward alternation. Taken together, the results suggest that a failure to identify random stimuli cannot be exclusively attributed to a failure to perceive the difference between random and nonrandom stimuli.

Robustness of Results

One follow-up study can be briefly noted here. With 20 new participants 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 nonrandom 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.

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, including 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 here, and it is entirely unclear whether it will improve perceptual discrimination as has been found in other areas of perceptual learning (e.g., Goldstone, 1998; Pevtzow & Harnad, 1997). We therefore conducted an additional follow-up study with 80 new participants 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.

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 [$t(50) = 1.90$, $p = .03$]. 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 highlight the difference between discrimination and identification; identification can be tuned based on feedback, whereas no tuning is apparent for discrimination. Further research is needed to find out what ability is tuned by feedback, whether the ability to identify a random event, or the ability to detect structure in one half and infer that the other half is random. Moreover, the fact that only 510 feedback trials were needed to shift the identification curve shows how malleable the intuitive randomness concept is. On the other hand, for discrimination, the lack of improvement from feedback suggests that there may be hard limits to perceptual discrimination between random and nonrandom stimuli. The same limits apply for identification after feedback, since the curve rose to that of discrimination but did not exceed the discrimination curve. Future research is needed to examine what causes the limits.

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

[1] Note that discrimination below chance (which never occurred) would be anomalous.

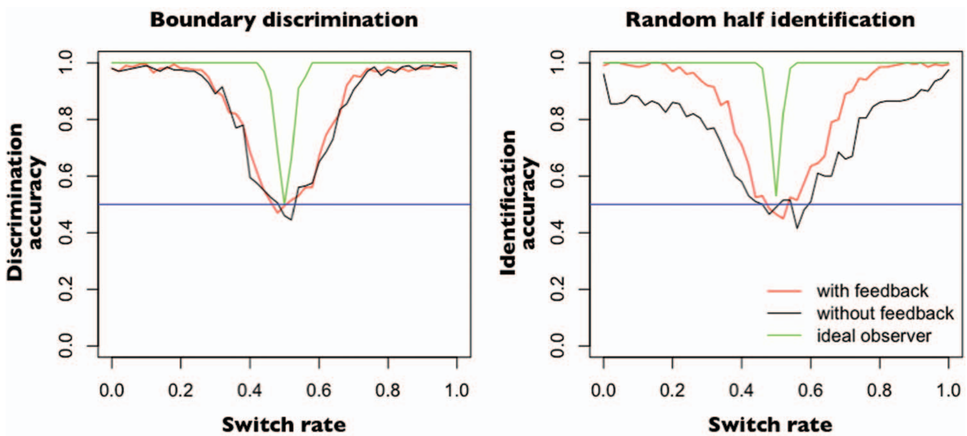

Figure 2. Feedback: Boundary discrimination accuracy and identification accuracy are plotted as a function of switch rates. The red line indicates performance with feedback, and the black line indicates performance without feedback. The green line indicates performance of the ideal observer. The blue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

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 [$t(100) = 0.56$, $p = .58$].

Experiment 2

The second experiment aimed to generalize the findings 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 (STM) in judgment and generation of random sequences (Hahn & Warren, 2009; for older work emphasizing the role of STM 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.

Participants

A new group of 40 undergraduates (27 female, mean age 19.8 years, SD = 0.8) from Princeton University participated in exchange for course credit.

Materials

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 vs. 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 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. All walks were freshly generated for each trial.

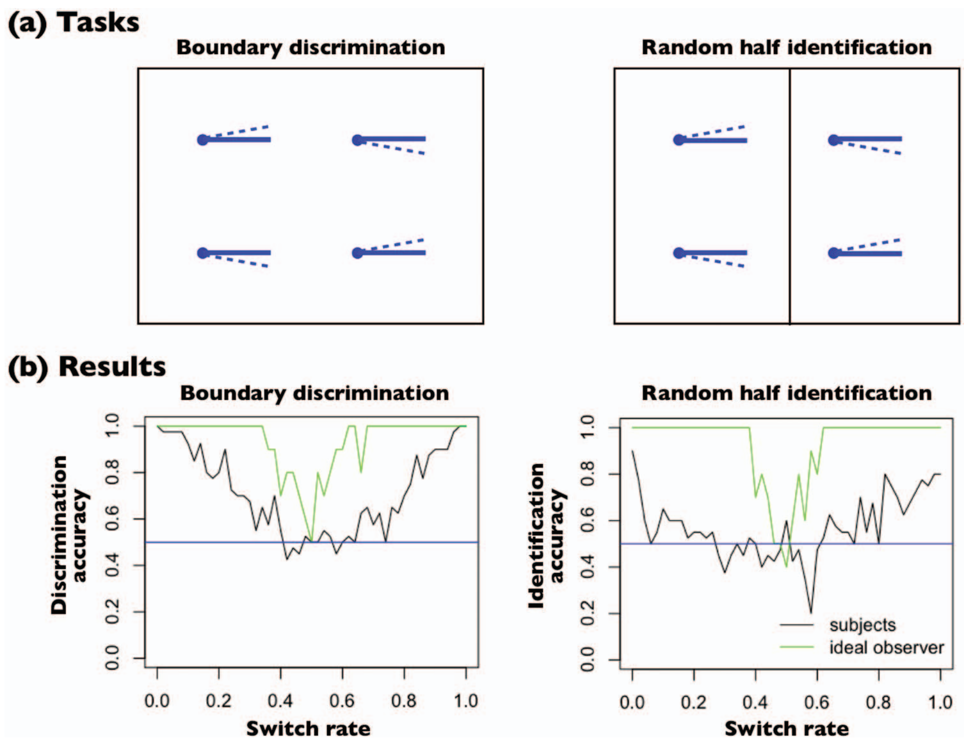

Figure 3. Experiment 2: (a) Each line starts horizontally and rotates according to its bit string. Dashed lines (not present in the stimuli) show possible first movements. In both conditions, the four quadrants could be divided into two halves: in one half the two walks were fully random, and in the 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 color version of this figure appears in the online article only.

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 counterclockwise on a given movement was randomly determined; for the other two lines, the walks would be nonrandom. 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.[2] 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 100 ms followed by a 50 ms intermovement interval. Each walk contained 100 movements, occurring simultaneously in the four quadrants. A trial lasted 15 seconds followed by a 1-s interval. If participants responded within the 1-s interval, the trial finished; otherwise, the screen remained blank until response. No feedback regarding judgment accuracy was provided.

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 [$r(49) = 0.67$, $p < .001$]. This correlation was maintained for switch rates ≤ 0.5 [$r(23) = 0.73$, $p < .001$], and ≥ 0.5 [$r(23) = 0.77$, $p < .001$].

A two-way mixed design ANOVA was conducted (between-subjects factor: discrimination vs. identification condition; within-subject factor: 51 levels of switch rates). There was a main effect for condition [$F(1, 38) = 33.03$, $p < .001$, $\eta_p^2 = .46$], and a main effect for switch rate [$F(50, 1900) = 8.03$, $p < .001$, $\eta_p^2 = .17$]. There was also a reliable interaction between condition and switch

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

rate [$F(50, 1900) = 1.74$, $p = .001$, $\eta_p^2 = .04$]. 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 nonrandom 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 nonrandom stimuli corresponds to the ability to distinguish between the two.

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). Kareev (1992) shows that a switch rate of .57 is considered as most random for strings of seven or eight items.

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 [$t(100) = 0.36$, $p = .72$]. The ideal observer performance is below that in Experiment 1, presumably because there were fewer bits presented in each trial (200 bits in Experiment 2 vs. 1,800 bits in Experiment 1).

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 versus static contrast, the former making more demands on working memory. It might also be due to the smaller number of bits controlling the dynamic displays of Experiment 2 compared to the matrices of Experiment 1 (200 vs. 1,800).

Robustness: Alternative Randomness Judgment Tasks

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, Griffiths, & Schreiber, 2010). More common, however, have been tasks that ask participants to rate the randomness of stimuli on a scale. It is thus interesting to examine randomness ratings for both the matrices and the walks.

In a follow-up study, 20 new participants 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 nonrandom*, 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.

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

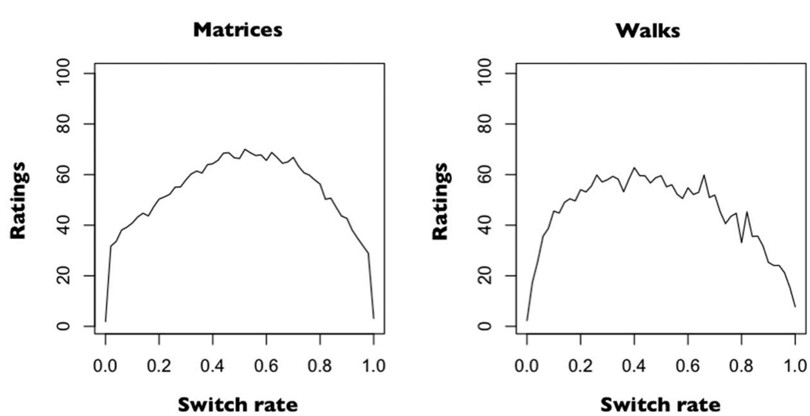

Figure 4. Ratings: Mean ratings of how random the half-matrices (left) and the walks (right) look on a scale from 0 (*completely nonrandom*) to 100 (*completely random*).

$r(49) = -0.83$, $p < .001$. 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. 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 overalternation bias through the leftward shift along the horizontal axis; visually correspond to previous findings (e.g., Falk & Konold, 1997, see their Figures 3, 5, and 6).

Experiment 3

Experiments 1 and 2 showed that the ability to identify a random stimulus cannot be fully predicted by the ability to distinguish between random and nonrandom stimuli. Consistent with past work (Bar-Hillel & Wagenaar, 1991), we find that people tend to identify overalternating stimuli as random. Our results also show that this bias is not present in perception. What, then, underlies the overalternation bias? Clarification of this matter would help explain why our participants in the preceding experiments used the label random as they did.

There are several proposals in the literature. First, 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 (Kahneman & Tversky, 1972; Gilovich et al., 1985; Tversky & Kahneman, 1971). More recently, Hahn and Warren (2009) demonstrated how STM limitations might impact the perception 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 HHHT. Given human STM 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 often they occur, alternating sequences being more likely to appear first in a sliding window compared to a uniform run of heads or tails (see also Kareev, 1992, for a related argument).

Finally, Falk and Konold (1997) provide a prominent psychological process account of randomness judgment. Specifically, they propose the encoding hypothesis:

> *Hypothesis 2:* Encoding Hypothesis—The probability that a given bit string is judged "random" varies directly with the time needed to memorize or copy it.

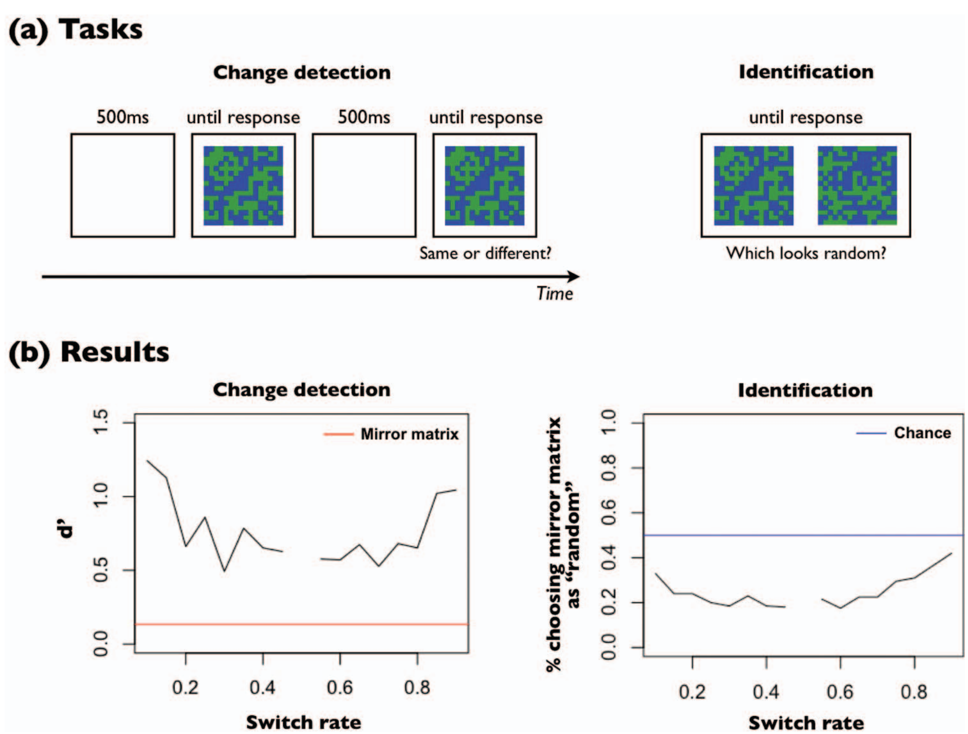

Figure 5. Experiment 3: (a) The left side shows the timeline for a change detection trial (matrices not drawn to scale). The right side shows an identification trial, with sample mirror and switch(x) matrices. (b) The left side plots change detection performance (d') for every switch rate and also for 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%). The color version of this figure appears in the online article only.

This clear and plausible hypothesis suggests that the tendency for a given pattern to be classified as random can be predicted by the cognitive difficulty the judge experiences in encoding the pattern. Sequences with greater structure or regularity, 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). This forms the basis of formal definitions of randomness as incompressibility. For $x \neq .5$, a 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 & Vitányi, 2008).[3]

Falk and Konold (1997) provided experimental evidence for this account. In their studies, they found that the time participants needed in order to memorize a given bit string, or copying difficulty, predicted participants' judgment of "how likely it is that such a sequence was obtained by flipping a fair coin" quite well. Hypothesis 2 invites extension to other kinds of stimuli and additional measures of encoding difficulty.

We thus sought to test the encoding account using 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:

> *Hypothesis 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.

To Test Hypothesis 3, 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, de Wit, & Wagemans, 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). Symmetry is a property that seems to be detected seemingly effortlessly and automatically in a wide variety of conditions (Wagemans, 1995), and it is likely to have a strong influence on randomness judgments (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-á-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's (1997) 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. Memory encoding is not like taking a detailed snapshot and there is a wealth of research to suggest that stimuli are typically encoded partially (see Rensink, 2002 for a review). Most day-to-day tasks such as recognition and identification arguably require only encoding only up to a level of detail that supports discrimination. 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.

Participants

A new group of 40 undergraduates (27 female, mean age 20.5 years, $SD = 1.6$) from Princeton University participated in exchange for course credit.

Materials

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). 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: change detection and identification. In the change detection condition ($n = 20$), 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

[3] There is no guarantee, however, that such differences in compressibility can be detected or put to use by human observers. More generally, the relative compressibility of two strings depends on the programming language at issue; shortest programs for generating a given bit 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). So there may be no practical way to compare the compressibility of strings.

[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 nonrandom matrices, which were presumably easier to encode but were identified as more random compared to mirror matrices.

(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 500 ms. 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 ($n = 20$), 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 500 ms after which the two matrices were presented until response.

Results and Discussion

To examine change detection performance, d' was computed for each participant for every switch rate, and also 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 [$ts(19) \geq 2.14$, $ps \leq .04$]. This suggests that mirror matrices were more difficult to encode than were switch(x) matrices for all levels of x. In contrast, except for $x = 0.85$ and $x = 0.9$, the probability of choosing the mirror matrix as more random than the switch(x) matrix was reliably below the chance level of 0.5 [$ts(19) \geq 2.63$, $ps \leq .01$]. In other words, 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. These results are inconsistent with the extended encoding hypothesis 3.

By the same token, the results point broadly toward an experiential basis for the overalternation bias (e.g., Hahn & Warren, 2009), in keeping with recent findings by Hsu et al. (2010) that participants' judgments of randomness for simple 4-by-4 binary matrices reflected adjacency statistics derived from real world scenes.[6]

General Discussion

Our overall goal was to determine what predicts the subjective judgment of randomness. Two hypotheses were put forward. Experiments 1 and 2 examined Hypothesis 1—that the ability to identify the provenance of a stimulus drawn from a random versus a nonrandom source is 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 nonrandom stimuli then you can see which is random; discrimination entails identification. Experiment 3 examined the encoding hypothesis by Falk and Konold (1997), which suggests that the tendency to identify given stimuli as random can be predicted by the encoding difficulty of the stimuli.

Our results reveal Hypothesis 1 to be inexact, whether bit strings are rendered statically (Experiment 1) or dynamically (Experiment 2). In both cases, participants were better at discrimination than identification. In other words, participants could reliably distinguish random from nonrandom stimuli but nonetheless were unable to identify which was random. This result may seem intuitive, because to be able to identify an event as random, one must be able to discriminate it from nonrandom events. In testing the relationship between identification and discrimination, the possible outcomes are that either discrimination entails identification as in Hypothesis 1, or discrimination is superior to identification. Our findings confirm the latter possibility, ruling out Hypothesis 1. Thus, the lay concept of randomness does not fully exploit the perceptual resources available for discriminating random from nonrandom sources. This also suggests that the lay conceptual difficulty for understanding randomness is not due to a perceptual inability to distinguish random from nonrandom events.

In our experiments, participants were not informed of the generating process, but they nonetheless performed reasonably close to an ideal observer. Discrimination accuracy was symmetrically distributed around true 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 significantly below chance) for switch rate 0.56 in Experiment 1. This is indicative of the alternation bias. Remarkably, it was worst (and far below chance) at essentially the same switch rate (0.58) in Experiment 2, involving radically different stimuli (see Figure 3). For discrimination, however, performance was reliably above chance for switch rate 0.56 in Experiment 1. Thus, the alternation bias noted by other investigators (Baddeley, 1966; Kahneman & Tversky, 1972; Wagenaar, 1972) appears to be fundamentally a conceptual phenomenon.

This is further emphasized by the lack of effect that feedback had on discrimination performance. When feedback was provided, identification performance improved and matched discrimination performance. By contrast, the perceptual ability to distinguish random from nonrandom events was unaffected by feedback. This differential impact of feedback is more evidence for the psychological gulf between randomness perception from conception.

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 turned out to be more difficult to encode yet less likely to be labeled random, in contradiction with the extended encoding hy-

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

[6] It should be mentioned, however, that given the local nature of those adjacency statistics, symmetry posed a challenge also in that study (see Hsu et al., 2010).

pothesis 3. To explain this finding, note that success in the encoding task requires local observation, namely, to determine whether a given cell in the matrix has switched bits. In contrast, judgments of randomness need to integrate the global feature of matrix symmetry, which strikes participants as nonrandom. Indeed, recent studies have highlighted the difference between global and local processing (e.g., Poljac et al., 2012). Thus, the encoding hypothesis 2 due to Falk and Konold (1997) does not appear to generalize to detecting changes in bit matrices.

In summary, although much past work has bemoaned the lay conception of randomness, the present paper provides a more nuanced assessment of people's capabilities in two ways. First, and most importantly, we distinguish between perceptual and conceptual understanding of randomness, and second, we compare human performance to that of an ideal observer. With regard to perception, our findings lead to a fairly positive picture. Moreover, a failure to further improve discrimination performance through feedback suggests that the limits of human performance in our tasks are reached. People are weaker in their conception of randomness. They benefited, however, from feedback, and performance rose to the level possible given the underlying discrimination ability.

While the distinction between discrimination and identification is not unique to randomness, how judgments of randomness reflect judgments of other stimuli (e.g., faces, objects) is an open and interesting question. On one hand, participants in Experiment 1 could be trained to identify random stimuli using feedback, in similar ways as people can be trained to recognize a given object or learn the identity of a person in previous studies. On the other hand, there is an objectively correct answer about randomness—a sequence is random if it's produced by a random source. However, there may not be an objectively correct answer about other objects or events. For example, there is no objective standard for judgments of the beauty of a painting; rather, we expect some individual variation. Future studies are needed to examine the extent to which randomness judgments reflect judgments of other stimuli.

In the real world, there may be cases where it is more important for us to distinguish between random and nonrandom sources than to understand the word *random*, a theoretical term that itself is extremely complex. This is because the former bears on the critical ability of perceiving predictable structure in the environment. In light of this, it would seem overly negative to consider people to be misguided about randomness.

References

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

Ayton, P., Hunt, A. J., & Wright, G. (1989). Psychological conceptions of randomness. *Journal of Behavioral Decision Making, 2,* 221–238. doi:10.1002/bdm.3960020403

Baddeley, A. D. (1966). The capacity for generating information by randomization. *Quarterly Journal of Experimental Psychology, 18,* 119–129. doi:10.1080/14640746608400019

Bakan, P. (1960). Response tendencies in attempts to generate random binary series. *The American Journal of Psychology, 73,* 127–131. doi:10.2307/1419124

Bar-Hillel, M., & Wagenaar, W. A. (1991). The perception of randomness. *Advances in Applied Mathematics, 12,* 428–454. doi:10.1016/0196-8858(91)90029-I

Beltrami, E. (1999). *What is random? Chance and order in mathematics and life*. New York, NY: Springer-Verlag. doi:10.1007/978-1-4612-1472-4

Blum, M. (1967). On the size of machines. *Information and Control, 11,* 257–265. doi:10.1016/S0019-9958(67)90546-3

Budescu, D. V. (1987). A Markov model for generation of random binary sequences. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 25–39. doi:10.1037/0096-1523.13.1.25

Camerer, C. F. (2003). Strategizing in the brain. *Science, 300,* 1673–1675. doi:10.1126/science.1086215

Chater, N. (1999). The search for simplicity: A fundamental cognitive principle? *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 52,* 273–302. doi:10.1080/713755819

Diener, D., & Thompson, W. B. (1985). Recognizing randomness. *The American Journal of Psychology, 98,* 433–447. doi:10.2307/1422628

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. doi:10.1016/j.neuron.2004.09.009

Dreman, D. N. (1977). *Psychology and the stock market: Investment strategy beyond random walk*. New York, NY: Amacom.

Eagle, A. (2012). Chance versus randomness. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved from http://plato.stanford.edu/archives/win2003/entries/chance-randomness/

Earman, J. (1986). *A primer on determinism*. Dordrecht, Netherlands: D. Reidel. doi:10.1007/978-94-010-9072-8

Edwards, W. (1961). Probability learning in 1000 trials. *Journal of Experimental Psychology, 62,* 385–394. doi:10.1037/h0041970

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17–52). New York, NY: Wiley.

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

Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review, 104,* 301–318. doi:10.1037/0033-295X.104.2.301

Fitelson, B., & Osherson, D. (2012). *Remarks on random sequences*. Retrieved from http://arxiv.org/abs/1205.5865

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech, 5,* 171–189.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17,* 295–314. doi:10.1016/0010-0285(85)90010-6

Glimcher, P. W. (2005). Indeterminacy in brain and behavior. *Annual Review of Psychology, 56,* 25–56. doi:10.1146/annurev.psych.55.090902.141429

Godfrey, J. J., Syrdal-Lasky, K. S., Millay, K. K., & Knox, C. M. (1981). Performance of dyslexic children on speech perception tests. *Journal of Experimental Child Psychology, 32,* 401–424. doi:10.1016/0022-0965(81)90105-3

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49,* 585–612. doi:10.1146/annurev.psych.49.1.585

Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (Ed.), *Vancouver studies in cognitive science: Vol. 2*: *Perception. Oxford,* UK: Oxford University Press.

Hahn, U., & Warren, P. A. (2009). Perceptions of randomness: Why three heads are better than four. *Psychological Review, 116,* 454–461. doi:10.1037/a0015241

Hasanhodzic, J., Lo, A. W., & Viola, E. (2010). *Is it real, or is it randomized? A financial turing test*. Retrieved from SSRN. doi:10.2139/ssrn.1558149

Houtsma, A. J. M., & Smurzynski, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *The Journal of Acoustical Society of America, 87,* 304. doi:10.1121/1.399297

Hsu, A. S., Griffiths, T. L., & Schreiber, E. (2010). Subjective randomness and natural scene statistics. *Psychonomic Bulletin & Review, 17,* 624–629. doi:10.3758/PBR.17.5.624

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. doi:10.1177/0092070305277381

Julesz, B. (1962). Visual pattern discrimination. *IRE Transactions on Information Theory, 8,* 84–92. doi:10.1109/TIT.1962.1057698

Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature, 290,* 91–97. doi:10.1038/290091a0

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3,* 430–454. doi:10.1016/0010-0285(72)90016-3

Kareev, Y. (1992). Not that bad after all: Generation of random sequences. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1189–1194. doi:10.1037/0096-1523.18.4.1189

Kay, A. C., Moscovitch, D. A., & Laurin, K. (2010). Randomness, attributions of arousal, and belief in God. *Psychological Science, 21,* 216–218. doi:10.1177/0956797609357750

Kelly, M. H., & Martin, S. (1994). Domain-general abilities applied to domain-specific tasks: Sensitivity to probabilities in perception, cognition, and language. *Lingua, 92,* 105–140. doi:10.1016/0024-3841(94)90339-5

Kunst-Wilson, W. R., & Zajonc, R. B. (1980) Affective discrimination of stimuli that cannot be recognized. *Science, 207,* 557–558. doi:10.1126/science.7352271

Ladouceur, R., Sylvain, C., Letarte, H., Giroux, I., & Jacques, C. (1998). Cognitive treatment of pathological gamblers. *Behaviour Research and Therapy, 36,* 1111–1119. doi:10.1016/S0005-7967(98)00086-2

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.

Li, M., & Vitányi, P. (2008). *An introduction to Kolmogorov complexity and its applications* (2nd ed.). New York, NY: Springer. doi:10.1007/978-0-387-49820-1

Lieb, E. H., Osherson, D., & Weinstein, S. (2006). *Elementary proof of a theorem of Jean Ville*. Retrieved from http://arxiv.org/PS_cache/cs/pdf/0607/0607054v1.pdf

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. doi:10.1037/0278-7393.8.6.626

Lopes, L. L., & Oden, G. C. (1987). Distinguishing between random and nonrandom events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 392–400. doi:10.1037/0278-7393.13.3.392

Manski, C. F. (2006). Interpreting the predictions of prediction markets. *Economics Letters, 91,* 425–429. doi:10.1016/j.econlet.2006.01.004

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. doi:10.1017/S003329171100095X

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9,* 353–383. doi:10.1016/0010-0285(77)90012-3

Neuringer, A. (1986). Can people behave randomly? The role of feedback. *Journal of Experimental Psychology: General, 115,* 62–75. doi:10.1037/0096-3445.115.1.62

Nickerson, R. S. (2002). The production and perception of randomness. *Psychological Review, 109,* 330–357. doi:10.1037/0033-295X.109.2.330

Nickerson, R. S., & Butler, S. F. (2009). On producing random binary sequences. *The 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. doi:10.3758/PBR.15.5.991

Osherson, D., & Weinstein, S. (2008). Recognizing strong random reals. *The Review of Symbolic Logic, 1,* 56–63. doi:10.1017/S1755020308080076

Osherson, D., & Weinstein, S. (2011). *Notes on random reals*. Retrieved from 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. doi:10.1037/a0014821

Overman, W., Bachevalier, J., Turner, M., & Peuster, A. (1992). Object recognition versus object discrimination: Comparison between human infants and infant monkeys. *Behavioral Neuroscience, 106,* 15–29. doi:10.1037/0735-7044.106.1.15

Pevtzow, R., & Harnad, S. (1997). Warping similarity space in category learning by human subjects: The role of task difficulty. In M. Ramscar, U. Hahn, E. Cambouropolos, & H. Pain (Eds.), *Proceedings of SimCat 1997: Interdisciplinary Workshop on Similarity and Categorization* (pp. 189–195). Edinburgh, Scotland: Edinburgh University.

Poljac, E., de Wit, L., & Wagemans, J. (2012). Perceptual wholes can reduce the conscious accessibility of their parts. *Cognition, 123,* 308–312. doi:10.1016/j.cognition.2012.01.001

Rabin, M. (2002). Inference by believers in the law of small numbers. *Quarterly Journal of Economics, 117,* 775–816. doi:10.1162/003355302760193896

Rapoport, A., & Budescu, D. V. (1992). Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology: General, 121,* 352–363. doi:10.1037/0096-3445.121.3.352

Rapoport, A., & Budescu, D. V. (1997). Randomization in individual choice behavior. *Psychological Review, 104,* 603–617. doi:10.1037/0033-295X.104.3.603

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology, 53,* 245–277. doi:10.1146/annurev.psych.53.100901.135125

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current theory and research* (pp. 64–99). New York, NY: Appleton-Century-Crofts.

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. doi:10.1038/nn1378

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology, 4,* 328–350. doi:10.1016/0010-0285(73)90017-0

Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition, 7,* 1–15. doi:10.1080/135062800394658

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press. doi:10.4159/harvard.9780674866270

Steeves, T. D. L., Miyasaki, J., Zuerowski, M., Lang, A. E., Pellecchia, G., Van Eimeren, T., . . . Strafella, A. P. (2009). Increased striatal dopamine release in Parkinsonian patients with pathological gambling: A [11C] raclopride PET study. *Brain, 132,* 1376–1385. doi:10.1093/brain/awp054

Tan, H. Z., Srinivasan, M. A., Reed, C. M., & Durlach, N. I. (2007). Discrimination and identification of finger joint-angle position using active motion. *ACM Transactions on Applied Perception, 4,* article 10. doi:10.1145/1265957.1265959

Terrell, D. (1998). Biases in assessments of probabilities: New evidence from greyhound races. *Journal of Risk and Uncertainty, 17,* 151–167. doi:10.1023/A:1007771613236

Toneatto, T., Blitz-Miller, T., Calderwood, K., Dragonetti, R., & Tsanos, A. (1997). Cognitive distortions in heavy gambling. *Journal of Gambling Studies, 13,* 253–266. doi:10.1023/A:1024983300428

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76,* 105–110. doi:10.1037/h0031322

van der Helm, P. A., & Leeuwenberg, E. L. J. (1996). Goodness of visual regularities: A nontransformational approach. *Psychological Review, 103,* 429–456. doi:10.1037/0033-295X.103.3.429

Wagemans, J. (1995). Detection of visual symmetries. *Spatial Vision, 9,* 9–32. doi:10.1163/156856895X00098

Wagenaar, W. A. (1972). Generation of random sequences by human subjects: A critical survey of the literature. *Psychological Bulletin, 77,* 65–72. doi:10.1037/h0032060

Wagenaar, W. A. (1988). *Paradoxes of gambling behaviour*. Hillsdale, NJ: Erlbaum.

Received May 31, 2013
Revision received March 19, 2014
Accepted March 25, 2014 ■