Detecting deviations from randomness

Jiaying Zhao (jiayingz@princeton.edu)
Department of Psychology, Green Hall,
Princeton University, NJ 08540 USA

Daniel Osherson (osherson@princeton.edu)
Department of Psychology, Green Hall,
Princeton University, NJ 08540 USA

Abstract

We explore the ability to distinguish random from non-random events without invoking the randomness concept. Randomness is defined in terms of radioactive decay whereas non-randomness is quantified by excess repetitions (i.e., repeat) or alternations between successive bits (i.e., switch). In four experiments, participants completed tasks including identifying the boundary between random and non-random textures, distinguishing random from non-random movement, learning to classify patterns, and tracking changes in successive matrices. Importantly, in task instructions, no mention was made of randomness, probability, or related concepts. We found superior performance in distinguishing random stimuli from repeat stimuli compared to switch stimuli. Moreover, memory for repeat stimuli declined as stimuli became more random, whereas memory for switch stimuli did not vary with the degree of non-randomness.

Keywords: Randomness; perception; alternation; texture

Introduction

Random processes are fundamental and ubiquitous in nature. While the definition of randomness remains controversial among philosophers and mathematicians, psychologists have been interested in people's subjective intuitions of randomness. The concept of randomness has been typically examined in two ways. In one method, participants are presented with a sequence of binary events and are asked to judge how random the sequence appears. It has been found that sequences of bits that alternate more than expected on the basis of random generation are more likely to be labeled as random (Lopes & Oden, 1987; Bar-Hillel & Wagenaar, 1991; Falk & Konold, 1997; Nickerson, 2002). In the other method, participants are instructed to produce a sequence as if it is produced by a random process such as a fair coin. The results from these studies demonstrate that sequences produced as random tend to result in too many alternations, thus, in runs that are too short (Wagenaar, 1972; Kahneman & Tversky, 1972; Baddeley, 1966).

These studies have greatly illuminated the ordinary conception of "random" but leave unexplored the perceptual consequences of deviations from randomness in the absence of explicit reference to the randomness concept. In this paper, we will perform four critical experiments to assess the ability to distinguish random from non-random stimuli without any explicit mention of randomness, probability, or their cognates. In Experiment 1, we will examine the ability to distinguish random from non-random bit matrices. This is followed by Experiment 2 where we will examine the ability to distinguish random from non-random events when they are temporally presented. In Experiment 3, we will look at the ability to learn to classify random vs. non-random matrices. Finally, in Experiment 4, we investigate the ability to encode random vs. non-random matrices.

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). Reliance on this prototypically random 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, and Weinstein (2006); Osherson and Weinstein (2008).) Regarding "non-random," there are many ways to distort a random process. We rely on two of the simplest. Assume the existence of an unlimited source of random bits (i.e., hotbits).

> <u>Definition of a repeat(x) sequence</u>: 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 nth 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 nth 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.

> <u>Definition of a switch(x) sequence</u>: For $x \in [0,1]$, a switch(x) sequence S is constructed as follows. The first bit is set equal to the next hotbit. Suppose that the nth 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 nth bit.

It can be seen that for $x < 1$, repeat(x) have longer runs than expected from a random source, whereas the runs in a switch(x) are too short. In both cases we expect the string to be more compressible than a (fully) random string, in the sense of being generated by shorter programs in an intuitively reasonable programming language (Li & Vitányi, 2008). There is no guarantee, of course, that such differences in compressibility can be detected or put to use by human observers. Indeed, the experiments described below probe the psychological boundary between full and partial randomness, asking in each case:

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

We begin with the elementary task of distinguishing bit matrices created from random versus repeat(x) or switch(x) sequences.

Experiment 1

Participants

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

Materials

Stimuli 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 repeat(x) sequence or a switch(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.

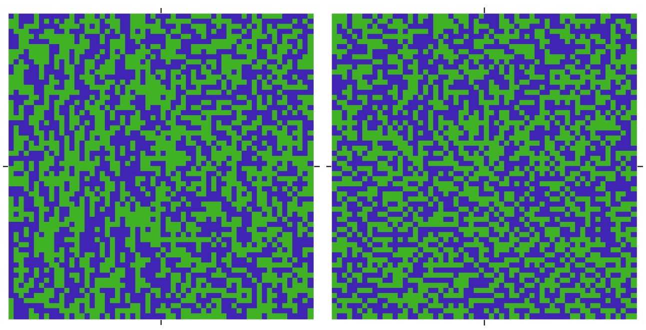

Figure 1: The matrix on the left can be divided vertically into a random half-matrix (right) and a repeat(0.7) half-matrix (left). The latter half-matrix was created from a repeat(0.7) sequence by traversing columns. The matrix on the right can be divided horizontally into a random half-matrix (top) and a switch(0.7) half-matrix (bottom). The latter half-matrix was created from a switch(0.7) sequence by traversing columns.

Procedure

Thirty participants (17 female) took part in the study. Each participant 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 *switch* condition only switch(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. Matrices appeared on the screen for 4000 ms. 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 could involve no feedback. Thus, a new set of 30 participants took part in the "no feedback" version, same as above but without feedback.

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.

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 switch(x) halves of matrices. Specifically, for a given x we tallied the numbers of correct and incorrect trials at or above x, 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 switch trials, yielding repeat and switch *thresholds*.

With feedback, the average repeat threshold was 0.85 ($SD = 0.05$, median = 0.85) and the average switch 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 switch 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 $p < .01$]. These results suggest that participants were better at discriminating random matrices from matrices biased towards repeats compared to switches.

Experiment 2

Since events can also occur in a sequential order (e.g., basketball shoots), in this experiment we examined the ability to distinguish random from non-random stimuli that were temporally presented.

Participants

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

Materials

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 100 ms. followed by a 50 ms. inter-movement interval. The movements in a given quadrant will be called its "walk." The walk in one quadrant (the "oddball") followed a repeat(x) or a switch(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.

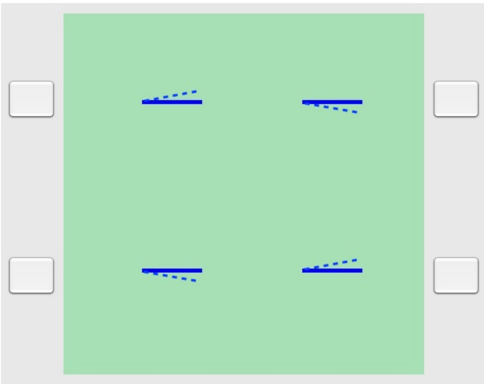

Figure 2: Every line starts horizontally and rotates according to its bit string. One line follows a repeat(x) or a switch(x) sequence, while the others move randomly. Dashed lines (not present in the stimuli) show possible first movements. The four buttons (left and right margins) registered the participant's guess about the oddball.

Procedure

Forty participants took part in the study; twenty were randomly assigned to the *repeat* condition, and twenty to *switch*. Each condition consisted of 100 trials. A given trial in the repeat condition presented random walks in three quadrants and a repeat(x) walk in the other; for the switch condition, the latter quadrant presented a switch(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.

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 1: 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.

Results and Discussion

Just as in Experiment 1, we identified repeat and switch thresholds for every participant. The average repeat threshold was 0.80 ($SD = 0.10$, median = 0.83) and the average switch 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 switch-based walks. The results are congruent with the findings of Experiment 1 except that both repeat and switch 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.

Experiment 3

Whereas the preceding experiments involved discriminating random from non-random stimuli, the third 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 switch(x) matrices. There was no mention of randomness or related notions.

Participants

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

Materials

Stimuli were the half-matrices used in Experiment 1, generated anew in the present setting. Each was either 30×60 or 60×30, randomly determined for each trial.

Procedure

Thirty-five participants served in both repeat and switch 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 switch condition, either a random or a switch(x) matrix was presented. The participant was informed that each matrix was drawn from one of two distinct categories. The image was projected for 1000 ms. 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 switch) 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.

Results and Discussion

We identified repeat and switch thresholds for every participant as in Experiment 1. The average repeat threshold was 0.90 (*SD* = 0.03, median = 0.91) and the average switch 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 switch. In turn, this implies greater facility in discriminating repeat-biased than switch-biased matrices from random ones.

At the same time, both the repeat and the switch thresholds were reliably higher than those in Experiment 1 (.90 vs. .85 and .86 vs. .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 the just the first hundred trials of Experiment 3, they are the same as in Experiment 1 (.85 for repeat and .82 for switch).

Experiment 4

Performance in the preceding experiments was better with repeat compared to switch stimuli (except for the variant procedure in Experiment 2). Does the difference reflect easier encoding of repeats compared to switches? This study was motivated by the finding that the encoding difficulty of a bit sequence increases with its degree of randomness (Falk & Konold, 1997). The present experiment examines this issue by requesting participants to detect changes in serially presented matrices.

Participants

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

Materials

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

Procedure

Forty participants completed the experiment individually. There were nine kinds of trial corresponding to (a) repeat levels 0.2, 0.4, 0.6, and 0.8, (b) the same switch 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 200 ms. followed by a 500 ms. gap (no mask); then the second matrix was presented for 200 ms. 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 were presented in individualized random order to each participant.

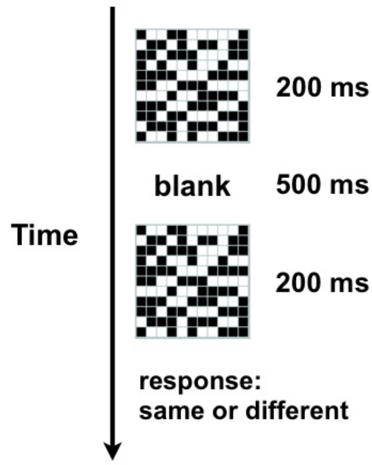

Figure 3: Sample trial in Experiment 4.

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(.2) matrices is 91% whereas it is just 70% for switch. For each level, repeat accuracy was compared to switch accuracy via paired t-test, also summarized in the table. Repeat accuracy was reliably higher than switch only for levels 0.2 and 0.4.

Table 1: Mean percent accuracy (and SD) in Experiment 4. p-values reflect repeat/switch differences in accuracy.

x	Repeat	Switch	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)	$p = 1$
1.0	65 (10)		

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 switch. There was a significant difference among the levels in the repeat condition [$F(4,191) = 8.14$, $p < .01$], but not in the switch 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). These results suggest that participants were better at encoding repeats than switches 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 switch, performance seemed not to vary with degree of non-randomness. One potential explanation for the difference in encoding ability between repeat and switch 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.

General Discussion

In four experiments, we examined the ability to distinguish genuine random from non-random stimuli without invoking randomness or related concepts in task instructions. Using different stimuli and a range of methods, we found superior performance in distinguishing random stimuli from stimuli that are biased towards repeats compared to switches. In Experiment 1, the ability to distinguish repeat matrices from random matrices was slightly but reliably better than the ability to distinguish switch matrices from random. Participants were likewise better at detecting non-random walks biased towards repeat in a field of random walks (Experiment 2). When learning to classify matrices into nominal categories (Experiment 3), participants were better at classifying repeat versus random matrices compared to switch versus random; the thresholds were close to those seen in Experiment 1. Finally, participants were better at perceiving changes embedded in non-random matrices biased towards repeats compared to switches (Experiment 4); moreover, memory for repeat matrices declined as matrices became more random whereas memory for switch matrices did not vary with the degree of non-randomness.

Across experiments, we have consistently observed that repeat thresholds are higher than switch thresholds. To explain the gap it is tempting to invoke the perceptual difference between streaks and alternations/checkerboards. The former, 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.) Switch 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 2 (random vs. non-random walks) whose stimuli were presented temporally. Further investigation is plainly needed.

The higher discrimination threshold with repeat stimuli compared to switches could in part be explained by the superior ability to encode repeat stimuli compared to switches. This result is consistent with the encoding hypothesis (Falk & Konold, 1997). In their study, participants were presented with binary bit sequences and were asked to copy the sequence from memory. It was found that the difficulty of encoding the sequence increased with its degree of randomness. We found similar results for repeat stimuli but not for switch since performance did not vary with degree of non-randomness.

There were several limitations of the current studies. The discrimination threshold could vary depending on a number of parameters, such as the size of the matrix, the presentation duration, the number of trials, the starting point of x, and whether the current bit was dependent on the previous bit or the second last bit. However, despite these possible variations, we expect that the difference between repeat threshold and switch threshold remains. Another limitation was that the current tasks were very abstract. Future study could use more real-world stimuli in order to generalize the current results.

Finally, further investigation could compare performance on perceptual discrimination with conceptual identification of random vs. non-random stimuli. For example, one could ask participants to identify which matrix looks more random in Experiment 1 and see whether their performance is consistent with their perceptual ability to distinguish. In addition, the following question also strikes us as meriting inquiry.

> What is the relationship between (a) the threshold difference for repeat and switch stimuli when discriminating them from randomness, and (b) the well-documented finding that over-alternating stimuli are more likely to be labeled as "random"?

Acknowledgments

We would like to thank John White, Nick Turk-Browne, Danny Oppenheimer, and Chris Olivola for helpful comments. Osherson acknowledges support from the Henry Luce Foundation.

References

Baddeley, A. D. (1966). The capacity for generating information by randomization. *Quarterly Journal of Experimental Psychology*, *18*, 119-129.

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

Brady, T. F., & Tenenbaum, J. B. (2010, August). Encoding higher-order structure in visual working memory: A probabilistic model. In *Proceedings of the thirty-second annual* *conference of the cognitive science society* (p. 411-416). Austin, TX.

Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review*, *104*, 301-318.

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

Li, M., & Vitányi, P. (2008). *An Introduction to Kolmogorov Complexity and Its Applications (2nd edition).* New York NY: Springer.

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

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

Nickerson, R. S. (2002). The production and perception of randomness. *Psychological Review*, *109*, 330-357.

Osherson, D., & Weinstein, S. (2008). Recognizing strong random reals. *Review of Symbolic Logic*, *1*(1), 56 - 63.

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

Walker, J. (2006). *Hotbits: Genuine random numbers, generated by radioactive decay.* (http://www.fourmilab.ch/hotbits/how3.html)