Original Article

QJEP

Quarterly Journal of Experimental Psychology
2018, Vol. 71(4) 906–916

DOI: 10.1080/17470218.2017.1307428
qjep.sagepub.com
SAGE

The consistency of the subjective concept of randomness

Ru Qi Yu[1], Jordan Gunn[2], Daniel Osherson[2] and Jiaying Zhao[1,3]

Abstract

A pervasive bias in the subjective concept of randomness is that people often expect random sequences to exhibit more alternations than produced by genuine random processes. What is less known is the stability of this bias. Here, we examine two important aspects of the over-alternation bias: first, whether this bias is present in stimuli that vary across feature dimensions, sensory modalities, presentation modes and probing methods, and, second, how consistent the bias is across these stimulus variations. In Experiment 1, participants adjusted sequences until they looked maximally random. The sequences were presented as temporal streams of colors, shapes, auditory tones or tiled as spatial matrices. In Experiment 2, participants produced random matrices by adjusting the color of each cell. We replicated the findings using a within-subjects design in Experiment 3. We found that participants judged and produced over-alternating stimuli as the most random. Importantly, this bias was consistent across presentation modes (temporal vs spatial), feature dimensions (color vs shape), sensory modalities (visual vs auditory), speed (fast vs slow), stimulus size (small vs large matrices) and probing methods (adjusting the generating process vs individual bits). Overall, the results suggest that the subjective concept of randomness is highly stable across stimulus variations.

Keywords

Randomness; modalities; alternation bias; anchoring effect; features

Received: 23 February 2016; revised: 30 September 2016; accepted: 10 January 2017

Introduction

The concept of randomness is instrumental to the understanding of human cognition. That is, the ability to detect regularities and patterns in the environment supports basic learning processes, from simple conditioning (Rescorla & Wagner, 1972), statistical learning (Fiser & Aslin, 2001), to language acquisition (Kelly & Martin, 1994). However, despite the significance of this ability, the cognitive system does not always represent the concept of randomness accurately, and such misconceptions often lead to systematic biases in judgments of randomness (Kahneman & Tversky, 1972; Tversky & Kahneman, 1971), which can cause suboptimal behaviors in gambling (Wagenaar, 1988). Therefore, elucidating how people understand randomness not only helps inform basic learning processes but also provides insight into problematic behaviors.

In common parlance, the term "random" is applied to sequences of events that appear sufficiently disorderly or unstructured. For example, the string *hthhtthttht* of heads and tails from coin tosses might qualify as random, whereas *hhhhhhtttttt* would not. A contrasting usage, adopted here, applies the term to certain mechanisms for generating events, namely, whose successive outputs are independent and unbiased. A standard example is a device *D* that tosses a fair coin repeatedly (ignoring issues about predictability assuming all forces were known). Any sequence of heads and tails produced by such a device is qualified as "randomly generated (by *D*)" regardless of its pattern. Thus, the terminology in this article follows a "process" rather than "product" conception of randomness (see Eagle, 2014; Earman, 1986 for an extended

[1]Department of Psychology, The University of British Columbia, Vancouver, BC, Canada
[2]Department of Psychology, Princeton University, Princeton, NJ, USA
[3]Institute for Resources, Environment and Sustainability, The University of British Columbia, Vancouver, BC, Canada

Corresponding author:
Jiaying Zhao, Department of Psychology, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada.
Email: jiayingz@psych.ubc.ca

discussion between the two approaches). Specifically, in our usage, a "random" stimulus (or pattern) is an object that has been produced by a random process. Non-random stimuli are defined as productions from a distorted random source.

The randomness of the coin flipper D can be compromised in various ways, for example, by making one of its two outcomes more likely than the other. Instead of introducing a bias to the probabilities of the two outcomes, here we consider deviations from stochastic independence by allowing previous flips to influence the next one while maintaining the equal frequency of the two outcomes. This allows us to identify the lay conception of randomness along a continuum. Specifically, for each number p in the unit interval (from 0 to 1), let $D(p)$ generate a sequence of bits consisting of 0s and 1s as follows:

> Sequence generation using the device $D(p)$: An unbiased coin toss determines the first bit. Suppose that the n^{th} bit has been constructed (for $n \geq 1$). Then with probability p the $n+1^{st}$ bit is set equal to the opposite of the n^{th} bit; with probability $1-p$ the $n+1^{st}$ bit is set equal to the n^{th} bit. Repeat this process to generate a sequence of any length.

This procedure was first introduced by Zhao, Hahn, and Osherson (2014). It can be seen that $D(0.5)$ is a genuinely random device. For $p<0.5$, $D(p)$ tends to repeat itself, resulting in long runs, whereas for $p>0.5$, $D(p)$ tends to alternate. In particular, $D(0)$ is uniform, either 0000 . . . or 1111 . . . , while $D(1)$ consists of perfectly alternating bits, either 0101 . . . or 1010 . . . The expected proportion of each bit is 50%, for all $p \in (0, 1)$, although, empirically, the output might deviate from 50%; however, such deviations should be small and random. For any sequence produced by $D(p)$, the expected proportion of alternation—called the "switch rate"—is p. The switch rate of any sequence is calculated by the number of switches between two successive bits divided by the total number of bits in the sequence minus one. For example, the switch rate of the sequence of 11111 or 00000 is 0, and the switch rate of 01010 is 1.

Admittedly, there may be discrepancies between the observed switch rate of the sequences produced by the generating process $D(p)$ and the intended switch rate (p). Here, we ran two simulations demonstrating the nature of this discrepancy. For each of the five intended switch rates—0.1, 0.3, 0.5, 0.7 and 0.9—we produced 1000 binary sequences. The length of each sequence was 100 bits for the first simulation (Figure 1a) and 6400 bits for the second simulation (Figure 1b). We calculated and plotted the observed switch rate of the produced sequences.

Figure 1 shows the distribution of the observed switch rates for 100-bit sequences and 6400-bit sequences across the five levels of intended switch rates. For both types of sequences, there was a normal distribution of the observed switch rates, suggesting that the average observed switch rate matched the intended switch rate, and it was equally likely for the observed switch rate to be higher or lower than the intended switch rate. Moreover, the variance in the observed switch rate was smaller for 6400-bit sequences than for 100-bit sequences.

A large body of research on randomness perception has revealed a pervasive bias in people's concept of "random." That is, people often expect random sequences to exhibit greater alternations than typically produced by random devices, or in our terms, $p>0.5$ (Falk & Konold, 1997; Kahneman & Tversky, 1972; Lopes & Oden, 1987; Nickerson & Butler, 2009; Wagenaar, 1972; for reviews on theoretical analysis of major experimental findings, see Bar-Hillel & Wagenaar, 1991; Oskarsson, van Boven, McClelland, & Hastie, 2009).

The over-alternation bias has been typically examined using judgment or production paradigms (e.g., Kahneman & Tversky, 1972; Wagenaar, 1972), where participants either judged how random a sequence appeared or produced a sequence as if it were generated by a random process. The over-alternation bias is consistent with the gambler's fallacy (Kahneman & Tversky, 1972; Reuter et al., 2005; Wagenaar, 1988) and could be driven by limitations in working memory (Baddeley, 1966; Hahn & Warren, 2009; Kareev, 1992).

However, most studies in the past have primarily used visual stimuli (e.g., 0011 or HHTT) to represent random sequences. Thus, the over-alternation bias found in these studies may be driven by constraints in the visual system. Currently, it is unknown how this bias manifests in different modalities and how the expression of the bias is influenced by specific task demands. Thus, to assess whether the over-alternation bias is a domain-general phenomenon, it is crucial to examine the bias across different sensory modalities (e.g., visual or auditory), stimulus feature dimensions (e.g., color or shape), presentation modes (e.g., temporal or spatial) and tasks (e.g., judgment or production). If the bias is consistent across different domains, this would suggest that people have a stable concept of randomness, robust to different task requirements and sensory processing constraints. On the other hand, any inconsistency would imply that the expression of the randomness concept depends on the elicitation method and the modality in which it is expressed.

The primary goal of this study is, thus, to examine the subjective concept of randomness across multiple domains. This will offer insights into the stability or consistency of the over-alternation bias. In Experiment 1, participants observed a sequence of bits produced by $D(p)$ for as long as they wanted. They were asked to adjust the sequence by increasing the likelihood of alternations or the repetitions of the bits to make the sequence look maximally random. In other words, they manipulated the switch rate p until they were satisfied that the sequence was being generated "randomly." We then measured how close p was to 0.5 (genuine randomness). Of principal interest was the stability of the

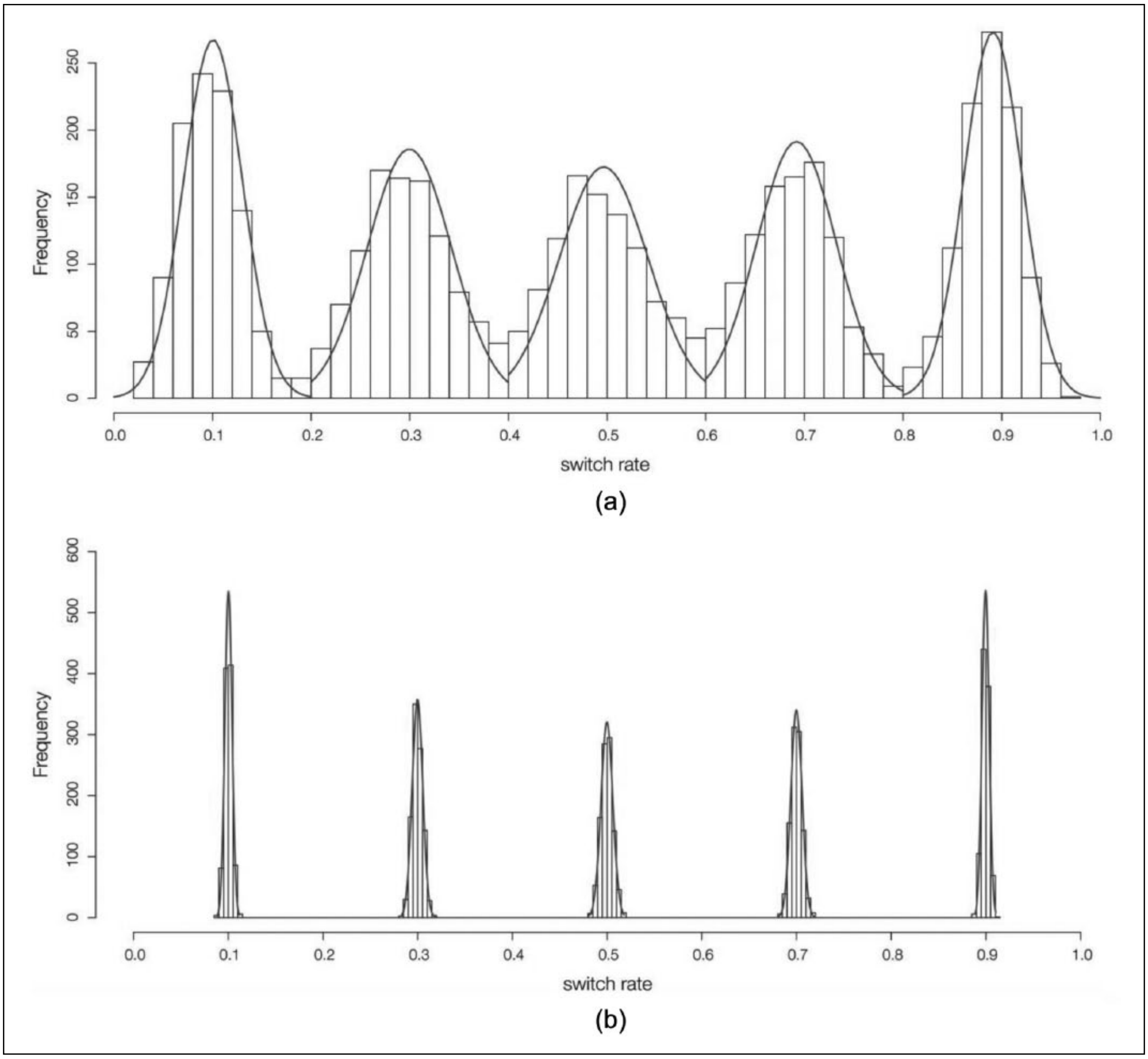

Figure 1. Histograms of switch rates in the simulated sequences. (a) Each sequence was 100 bits in length. At each switch rate (0.1, 0.3, 0.5, 0.7 and 0.9), we simulated 1000 sequences. The observed switch rates of the produced sequences showed a normal distribution at each level. (b) Each sequence was 6400 bits in length. At each switch rate (0.1, 0.3, 0.5, 0.7 and 0.9), we simulated 1000 sequences. The observed switch rates of the produced sequences showed a normal distribution at each level with smaller variances.

switch rate across variations in the stimuli that represent the bits generated by $D(p)$. In Experiment 2, we examined the stability of the switch rate in a randomness production task. We replicated the tasks in Experiments 1 and 2 using a within-subjects design with more statistical power in Experiment 3. The three experiments investigated the consistency of people's randomness concept across stimulus features, modalities, presentations and tasks. A summary of the stimulus conditions is shown in Table 1.

Experiment 1

The goal of this experiment was to examine the subjective concept of randomness across stimulus feature dimensions, sensory modalities and presentation modes.

Participants

In total, 46 undergraduate students (29 females, mean age = 20.7 years, standard deviation [*SD*] = 2.1) from the University of British Columbia (UBC) participated for course credit. Participants in all experiments provided informed consent. All experiments reported here have been approved by the UBC Behavioral Research Ethics Board.

Apparatus

In all experiments, participants were seated 50 cm from a computer monitor (refresh rate = 60 Hz) and used stereo headphones for auditory stimuli. Stimuli were presented

Table 1. Summary of stimulus conditions in Experiments 1 and 2.

Experiment	Probing method	Starting point	Presentation	Features
1	Adjust the switch rate in the generating process until the stimulus looks maximally random	The starting point of the stimulus switch rate was randomly determined from 0 (fully repeating) to 1 (fully alternating)	Temporal	Colored squares: green and blue
			Temporal	Black shapes: squares and circles
			Temporal	Tones: high and low pitches
			Spatial	10 × 10 matrices
			Spatial	80 × 80 matrices
2	Change the color of each cell until the matrix looks maximally random	The starting point of the switch rate was 0 (fully repeating)	Spatial	10 × 10 matrices
		The starting point of the switch rate was 1 (fully alternating)	Spatial	10 × 10 matrices
		The starting point of the switch rate was 0.5 (fully random)	Spatial	10 × 10 matrices
3	Replicating Experiments 1 and 2 using a within-subjects design	The same as described in Experiments 1 and 2	The same as described in Experiments 1 and 2	The same as described in Experiments 1 and 2, except the matrix dimensions are 11 × 11 and 81 × 81

and responses were collected using JAVA and Python interfaces.

Stimuli

Temporal sequences. There were six temporal trials, each containing a binary sequence. The six trials consisted of two color trials, two shape trials and two auditory trials (Figure 2a). In each color trial, the two bits were represented by a green square (RGB values: 3, 254, 82) and a blue square (RGB values: 6, 32, 244). In each shape trial, the two bits were a black square and a black circle. The square width and the circle diameter subtended 5.1°. In each auditory trial, the two bits were a high tone (pitch: 392 Hz) and a low tone (pitch: 262 Hz). One bit was presented at a time. For each type of trial, there was one fast trial and one slow trial. In a fast trial, each bit was presented for 400 ms and the inter-stimulus interval (ISI) was a blank screen for 200 ms. In a slow trial, each bit was again presented for 400 ms, but the ISI was 1000 ms.

Spatial matrices. There were six spatial trials, each containing a matrix. The six trials consisted of three small matrices and three large matrices (Figure 2b). Each matrix was constructed by tiling either vertically or horizontally a sequence of green and blue squares. The large matrix was 80×80, with 6400 bits subtending 14°. The small matrix was 10×10, with 100 bits subtending 11.3°.

Thus, there were 12 trials in total for each participant, allowing the following comparisons in presentation mode temporal versus spatial, color versus shape, visual versus auditory, fast versus slow and small versus large matrices. These exhaustive comparisons permitted the examination of consistency of randomness judgments across various stimulus domains within the same individual participant.

Procedure

Temporal trials. For each temporal trial, the starting switch rate of the sequence was randomly determined from 0 to 1. Each bit was presented on the screen for 400 ms, and the ISI was 200 ms for a fast trial and 1000 ms for a slow trial. As participants viewed each bit, they were asked to adjust the sequence by clicking on two buttons, one labeled "more repeating" and the other "less repeating," in order to make the sequence maximally random. As soon as the participant pressed a button, the switch rate in the generating process $D(p)$ changed by a constant amount. Specifically, if the participant pressed on the "more repeating" button, the switch rate in the generating process $D(p)$ decreased by 0.025; if the participant pressed on the "less repeating" button, the switch rate in the generating process $D(p)$ increased by 0.025. It was made clear to the participants that by clicking the buttons they were directly altering the generating process $D(p)$ which determined the overall likelihood of a repeat in the sequence, rather than immediately changing the number of repeats in the sequence. Their goal was to change the generating process $D(p)$ such that the sequence resembles a fully random sequence. This instruction closely followed the "process" definition of randomness, rather than the "product" definition. Every participant in all three experiments clicked on the buttons to adjust the likelihood of seeing a repeat in the sequence.

While it was possible, although unlikely, for a random process $D(0.5)$ to produce an output that looked non-random, we encouraged the participants to take as much time as they needed, to click as many times as they needed and to view as many bits as possible before they made their decision.[1] It is important to note that by altering the generating process, participants did not immediately increase the run length but rather increased the probability of the next bit to repeat the previous bit since run length and

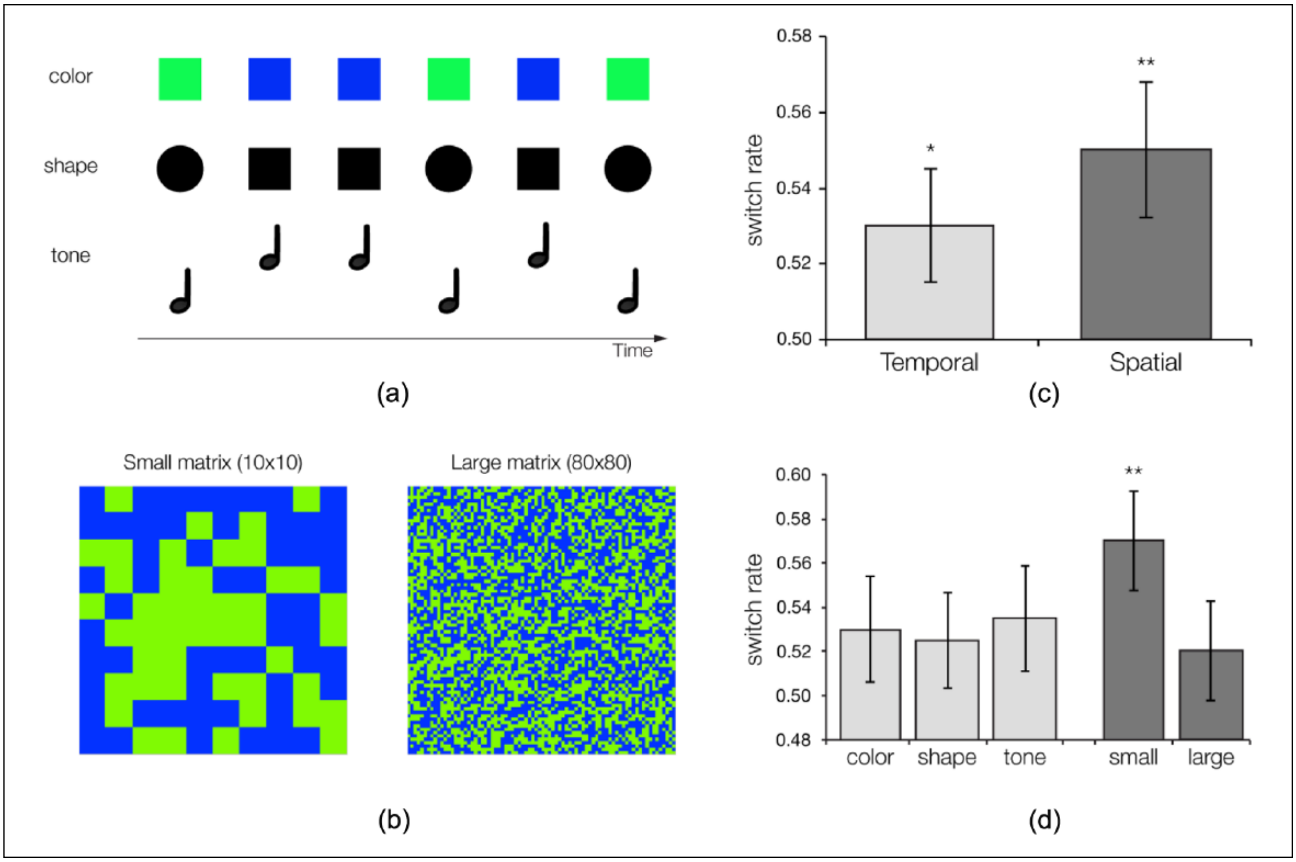

Figure 2. Stimuli and results for Experiment 1. (a) Three types of temporal sequences were presented. In the color sequence, the two bits were green and blue squares. In the shape sequence, the two bits were circles and squares. In the auditory sequence, the two bits were high and low tones. Each sequence started with a random switch rate, and participants adjusted the switch rate until the sequence looked maximally random. (b) Two types of matrices were presented, a small 10×10 matrix and a large 80×80 matrix. A fully alternating sequence tiled horizontally or vertically would result in a striped pattern, rather than a checkerboard pattern. Each matrix was generated from a sequence of blue and green squares, tiled vertically or horizontally. Participants adjusted the switch rate of the matrix until it looked maximally random. (c) The switch rate of temporal and spatial trials. (d) The switch rate of each type of temporal and spatial trials. Error bars indicate ±1 between-subjects standard error of the mean (SEM).
$^{*}p < 0.05$; $^{**}p < 0.01$.

switch rate independently impact the perception of randomness (Scholl & Greifeneder, 2011). There was no time limit for any trial. When the sequence was rendered maximally random in the judgment of the participant, he or she pressed another button to end the trial, and the switch rate (p) of the generating process was recorded; the same was true for the spatial trials described below. The order of the temporal trials was randomized for each participant.

Spatial trials. For each spatial trial, the starting switch rate of the sequence was randomly determined from 0 to 1. The tiling (vertical or horizontal) of the sequence in the matrix was randomly determined. Participants were instructed to process the matrix holistically, and we never mentioned horizontal or vertical tilings. We should point out that even though the switch rate along one traversal (e.g., horizontal) was not fully random (e.g., $p = 0.7$), the switch rate along the other traversal (e.g., vertical) was expected to be close to 0.5, except for fully repeating or alternating matrices.

Participants were asked to adjust the matrix by clicking on either the "more repeating" or the "less repeating" button to make the matrix look maximally random. As soon as the participant pressed a button, the switch rate of the sequence changed by 0.025 as in the temporal trials, and a new matrix was presented with a randomly determined tiling direction. As with the temporal trials, participants were instructed that by clicking the buttons they were directly altering the generating process $D(p)$ which determined the overall likelihood of a repeat in the matrix, rather than immediately changing the number of repeats in the matrix. Their goal was to change the generating process $D(p)$ such that the matrix resembles a fully random matrix.

Participants were encouraged to view each matrix for as long as desired and to make as many adjustments as they wanted. There was again no time limit for any trial. If the matrix looked maximally random, participants pressed another button to end the trial. Participants always completed the temporal trials before the spatial trials. The order of the small or large matrices was randomized for each participant.

Results and discussion

Across all trials, participants viewed 97.8 ($SD=55.4$) bits on average of a sequence before making a decision. They adjusted the sequence 34.4 ($SD=59.0$) times on average before making a decision.

For each trial, the switch rate of the sequence that the participant judged to be maximally random was recorded. Across participants, the average switch rate of the temporal trials was 0.53 ($SD=0.10$), reliably above 0.5 (fully random) ($t(45)=2.03$, $p<0.05$, $d=0.30$). The average switch rate of the spatial trials was 0.55 ($SD=0.12$), again reliably above 0.5 ($t(45)=2.71$, $p<0.01$, $d=0.40$), but not different from that of the temporal trials ($t(45)=0.80$, $p=0.43$, $d=0.14$). This reveals an over-alternation bias for both temporal and spatial trials (Figure 2c). To confirm that the initial switch rate of each sequence was not biased, we found that the starting switch rate was not different from 0.5 for the temporal trials ($M=0.49$, $SD=0.12$, $t(45)=0.42$, $p=0.68$, $d=0.06$) or for the spatial trials ($M=0.50$, $SD=0.12$, $t(45)=0.14$, $p=0.89$, $d=0.02$). These results suggest that over-alternating sequences were judged to be maximally random, regardless of the temporal or the spatial presentation mode.

First, for the temporal trials, a 3 (three feature dimensions: color, shape and tone) × 2 (fast vs slow speed) repeated-measures analysis of variance (ANOVA) revealed no main effect of feature dimension ($F(2, 90)=0.06$, $p=0.94$, $\eta_p^2 < 0.09$), speed ($F(1, 45)=0.08$, $p=0.78$, $\eta_p^2 < 0.01$) or interaction ($F(2, 90)=0.03$, $p=0.94$, $\eta_p^2 < 0.01$). There was no difference in the switch rate among the color, shape and auditory trials ($F(2, 90)=0.06$, $p=0.94$, $\eta_p^2 < 0.01$). Then, the specific types of temporal trials were compared. The average switch rate was 0.533 ($SD=0.16$) for the color trials and 0.525 ($SD=0.15$) for the shape trials, and the two were not different ($t(45)=0.30$, $p=0.77$, $d=0.05$). The average switch rate for the auditory trials was 0.535 ($SD=0.16$). Moreover, the average switch rate was 0.528 ($SD=0.13$) for the fast trials and 0.534 ($SD=0.12$) for the slow trials, and the two were not different ($t(45)=0.28$, $p=0.78$, $d=0.05$). No specific trial type produced a switch rate reliably above 0.5 ($t(45)<1.90$, $p>0.06$, $d<0.29$). Nonetheless, the switch rates of all trial types were remarkably similar (between 0.525 and 0.535; Figure 2d). This suggests that the randomness judgment was consistent across feature dimensions, sensory modalities and presentation speed.

For spatial trials, the average switch rate was 0.57 ($SD=0.15$) for small matrices, reliably above 0.5 ($t(45)=3.28$, $p<0.01$, $d=0.48$). The average switch rate was 0.52 ($SD=0.15$) for large matrices, not different from 0.5 ($t(45)=0.89$, $p=0.38$, $d=0.13$) or from that of the small matrices ($t(45)=1.84$, $p=0.07$, $d=0.35$). There was a significant correlation in the switch rates between small matrices and temporal trials ($r(44)=0.39$, $t=2.79$, $p<0.01$), but not between large matrices and temporal trials ($r(44)=0.13$, $t=0.86$, $p=0.40$). This suggests that the over-alternation bias was more prominent when the sample that the participants experienced at a given moment in time was small (100 bits) than when the sample was large (6400 bits).

Finally, we observed a strong anchoring effect for all types of trials in the experiment. Specifically, the switch rate was highly correlated with the starting switch rate for both temporal trials ($r(44)=0.50$, $t=3.85$, $p<0.001$) and spatial trials ($r(44)=0.44$, $t=3.23$, $p<0.01$), and also for color ($r(44)=0.58$, $t=4.67$, $p<0.001$), shape ($r(44)=0.46$, $t=3.44$, $p<0.01$), auditory trials ($r(44)=0.59$, $t=4.85$, $p<0.001$), as well as for small matrices ($r(44)=0.55$, $t=4.33$, $p<0.001$) and large matrices ($r(44)=0.49$, $t=3.72$, $p<0.001$).

Taken together, these results suggest that over-alternating sequences were judged as maximally random. Importantly, this bias was consistent across presentation modes (temporal vs spatial), feature dimensions (color vs shape), sensory modalities (visual vs auditory), speed (fast vs slow) and stimulus size (small vs large matrices).

Experiment 2

In Experiment 1, any adjustment made by the participants altered the underlying process that generated the stimuli and thus resulted in an entirely new and different sequence. A different method to probe people's concept of randomness is to have participants produce each bit in the sequence as if the bits are generated by a random process. Thus, Experiment 2 employed this paradigm to see whether the over-alternation bias was consistent with that observed in Experiment 1.

Participants

In total, 65 undergraduate students (47 females, mean age = 21.5 years, $SD=2.9$) from UBC participated for course credit.

Stimuli and procedure

Participants completed 12 trials in total. In each trial, they were first presented with a matrix and then asked to adjust the cells in the matrix to make it maximally random. The initial matrix was fully uniform, alternating or random.

Participants were encouraged to change as many cells in the matrix as they wanted in order to make the matrix maximally random. This method was comparable to that in Experiment 1 where a strong anchoring effect was observed.[2]

Each matrix was 10×10 with 100 cells in total, subtending 15°. Each cell could be either black or white, representing the two possible bits. Of the 12 trials, there were three types of matrices which were initially presented to participants: uniform matrices with all black or white cells (probability of a switch was 0 in the generating process), fully alternating matrices with a sequence (probability of a switch was 1 in the generating process) tiled horizontally or vertically, and fully random matrices with a random sequence (probability of a switch was 0.5 in the generating process) tiled horizontally or vertically (Figure 3a). The trials were presented in a random order.

In each trial, participants first viewed the initial matrix and then clicked on any cell in the matrix to reverse its color. They were told to produce a maximally random matrix as if all the bits were generated by a truly random process (e.g., a fair coin). They were encouraged to change as many cells as they like and also take as much time as they needed until the matrix looked fully random. As in Experiment 1, there was no time limit. The observed switch rate of the matrix was recorded.

Results and discussion

To compute the switch rate of a produced matrix, the matrix was transformed into two binary sequences, one by extracting the bits across columns horizontally through the matrix and another by traversing across rows vertically through the matrix. The switch rates of the two sequences were computed and then averaged. The average switch rate was 0.53 (SD=0.07), reliably above 0.5 ($t(64)=3.48$, $p<0.001$, $d=0.43$). This again shows an over-alternation bias in the production of random matrices (Figure 3b). The current switch rate was not reliably different from that of the small matrices (0.57) in Experiment 1 ($t(109)=1.78$, $p=0.08$, $d=0.36$) or from the switch rate of temporal trials (0.53) in Experiment 1 ($t(109)=0.06$, $p=0.95$, $d=0.01$). This reveals a consistent concept of randomness between the two experiments using different probing methods.

Among the three types of matrices, there was a reliable difference in the switch rate via a one-way repeated-measures ANOVA ($F(2, 128)=26.18$, $p<0.001$, $\eta_p^2 = 0.29$). Specifically, when the initial matrix was uniform, the switch rate of the produced matrix was 0.48 (SD=0.13), not different from 0.5 ($t(64)=1.14$, $p=0.26$, $d=0.14$), but reliably different from that when the initial matrix was fully alternating ($M=0.55$, $SD=0.05$, $t(64)=5.45$, $p<0.001$, $d=0.70$) or when the initial matrix was fully random ($M=0.56$, $SD=0.05$, $t(64)=5.19$, $p<0.001$, $d=0.74$). The latter two switch rates were not different from each other ($t(64)=0.62$, $p=0.53$, $d=0.06$) but were both reliably above 0.5 (alternating: $t(64)=7.74$, $p<0.001$, $d=0.96$; random: $t(64)=8.88$, $p<0.001$, $d=1.10$).

These results suggest that the produced matrix was over-alternating and biased toward the initial matrix, showing the same anchoring effect as in Experiment 1. Importantly, the over-alternation bias was consistent with that in Experiment 1, despite the differences in the tasks.

Experiment 3

The goal of this experiment was to replicate Experiments 1 and 2 with a more powerful design. In Experiment 1, each participant only completed one temporal trial for each type of trials. To reduce the variance within a participant, we here increased the number of trials for each type of trials within a participant. Moreover, each participant completed the tasks in both Experiments 1 and 2. Thus, Experiment 3 used a within-subjects design with more statistical power to examine the over-alternation bias.

Participants

In total, 50 undergraduate students (39 females, mean age=19.9 years, SD=2.3) from UBC participated for course credit.

Stimuli and procedure

Each participant completed 64 trials in total. The first 40 trials were analogous to the temporal trials and spatial trials described in Experiment 1. Each temporal trial described in Experiment 1 was run four times, resulting in 24 temporal trials. Following the temporal trials, participants completed 16 spatial trials, where participants either adjusted a smaller 11×11 matrix subtending 10° (8 trials) or a larger 81×81 matrix subtending 9° (8 trials) to allow for a checkerboard rather than a stripy pattern for fully alternating matrices. The order of the direction of tiling (vertical vs horizontal) was randomized and counterbalanced. For both the temporal and spatial trials, participants adjusted the temporal sequences or spatial matrices by clicking two buttons—"more repeating" or "less repeating"—for as long as they would like, until the sequence/matrix looked maximally random to them. In addition, participants were explicitly told that a maximally random sequence is the one that is most likely to be generated by a random process (e.g., a fair coin). Once again, participants were told that they were adjusting the overall likelihood of seeing a repeat in the sequence, rather than immediately changing the number of repeats after each adjustment. Participants then clicked a separate button to end the trial, and the switch rate in the generating process for the temporal sequences was recorded, but the observed switch rate of the matrices was recorded. This is because there was no way to identify the length of the temporal sequence over which the participant deemed most random.

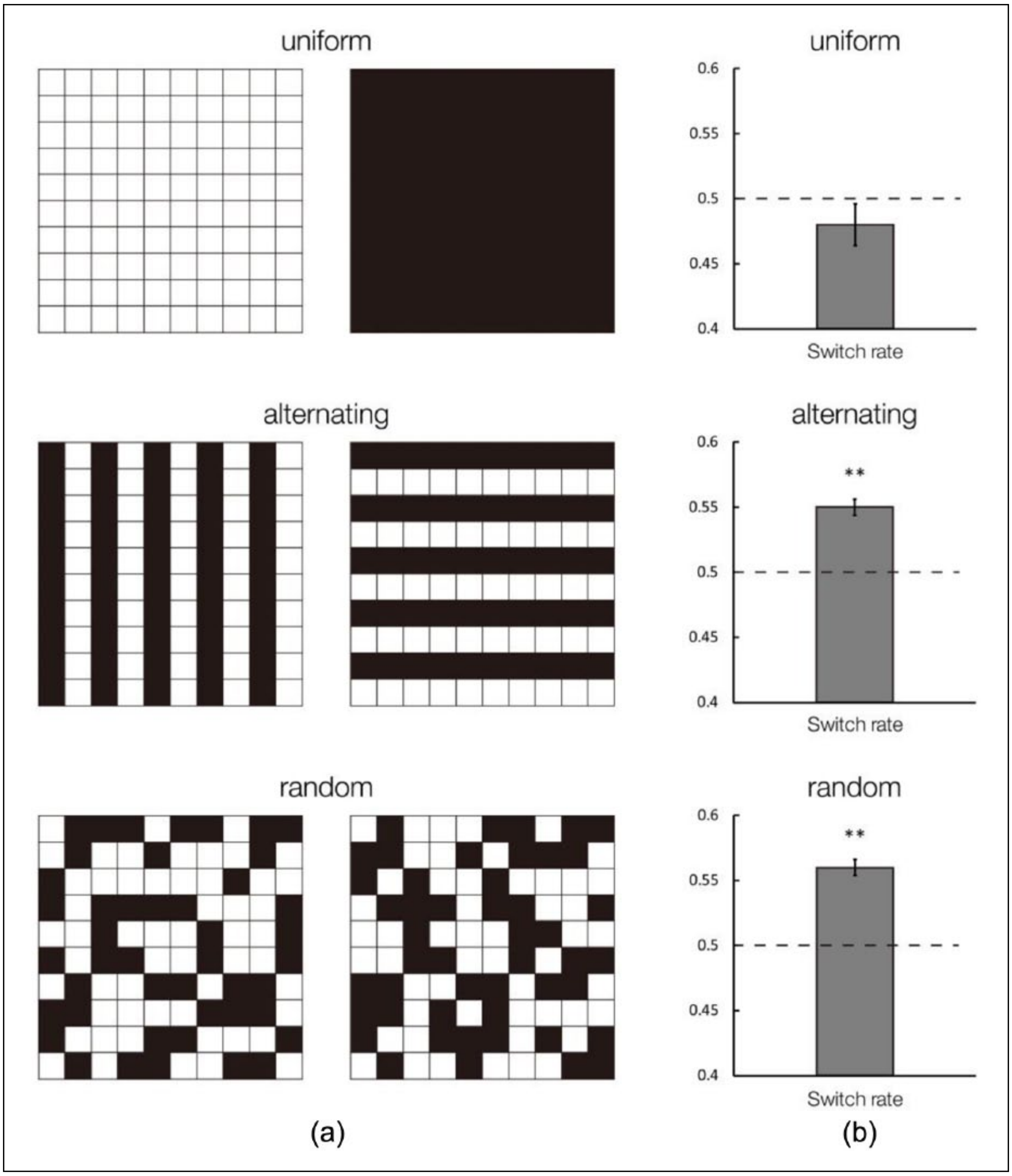

Figure 3. Stimuli and results for Experiment 2. (a) Three types of the initial 10×10 matrices were presented to participants. Two examples of each type are shown in the figure. Uniform matrices started with all black or white cells, fully alternating matrices with a fully alternating sequence tiled horizontally or vertically, and fully random matrices with random sequences tiled horizontally or vertically. Participants clicked on the cells in the matrix to change its color, until the matrix appeared maximally random as if the matrix was determined by a random process. (b) The switch rate of participants' produced matrices was presented for each of the three initial matrices. The switch rates were compared against the truly random point 0.5.
$^{**}p < 0.01$.

The last 24 trials in the experiment were the same as the 12 trials described in Experiment 2, except for two critical differences: (1) the number of trials doubled, and (2) each matrix was 11×11 to allow for a checkerboard rather than a stripy pattern for fully alternating matrices.

Results and discussion

In this experiment, the observed switch rate of the adjusted matrices (analogous to the spatial trials in Experiment 1) and the observed switch rate of produced matrix (analogous to Experiment 2) were computed in the same way as in Experiment 2. That is, to compute the switch rate of a matrix, the matrix was transformed into two binary sequences, one by extracting the bits across columns horizontally through the matrix and another by traversing across rows vertically through the matrix. The switch rates of the two sequences were computed and then averaged. On average, participants viewed 33.9 ($SD = 14.1$) bits in a sequence before making a decision. On average,

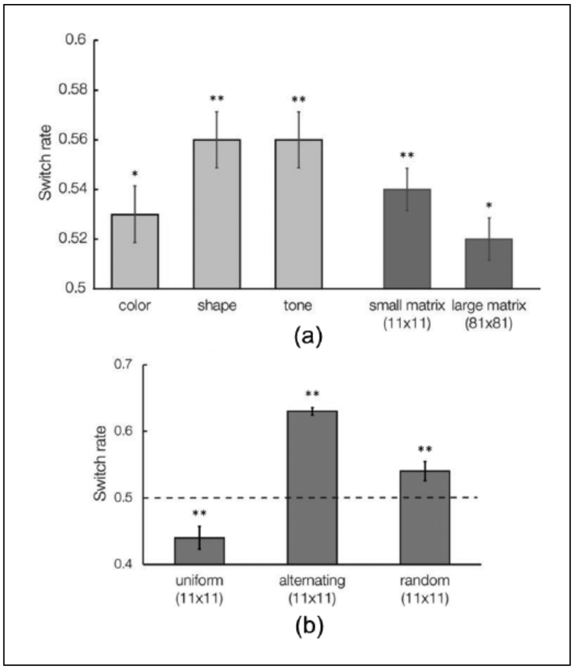

Figure 4. Results of Experiment 3. (a) Replication of Experiment 1. (b) Replication of Experiment 2. The switch rates were compared against the truly random point 0.5. *$p<0.05$; **$p<0.01$.

participants made 14.4 ($SD=9.6$) adjustments before making a decision.

For the temporal trials, a 3 (three feature dimensions: color, shape and tone) × 2 (fast vs slow speed) repeated-measures ANOVA revealed a main effect of feature dimension ($F(2, 98)=4.74, p=0.01, \eta_p^2=0.09$), but no effect of speed ($F(1, 49)=0.05, p=0.83, \eta_p^2<0.01$) or interaction ($F(2, 98)=0.13, p=0.88, \eta_p^2<0.01$). The switch rates were reliably above 0.5 for color ($M=0.53, SD=0.08, t(49)=2.33, p=0.02, d=0.53$), shape ($M=0.56, SD=0.08, t(49)=5.21, p<0.001, d=0.74$) and tone dimensions ($M=0.56, SD=0.08, t(49)=5.49, p<0.001, d=0.78$), revealing a robust over-alternation bias (Figure 4a). Among the three types of temporal trials, there was a reliable difference ($F(2, 98)=4.74, p=0.01, \eta_p^2=0.09$); both shape and tone trials were higher than color trials ($p<0.05$), but there was no difference between shape and tone trials ($p=0.99$).

For the spatial trials analogous to those in Experiment 1, the switch rates of both the large and small matrices were reliably above 0.5 (for large matrices: $M=0.52, SD=0.06, t(49)=2.31, p=0.03, d=0.33$; for small matrices: $M=0.54, SD=0.06, t(49)=5.61, p<0.001, d=0.79$). Consistent with the findings in Experiment 1, the switch rate of small matrices was reliably higher than those of large matrices ($t(49)=2.19, p=0.03, d=0.38$).

For the spatial trials analogous to those in Experiment 2, among the three types of matrices (with the initial matrix being uniform, random or alternating), there was a reliable difference in the switch rate via a one-way repeated-measures ANOVA ($F(2, 98)=64.59, p<0.001, \eta_p^2=0.57$), with all pair-wise comparisons being significant ($p<0.001$). This provides further evidence for the anchoring effect of the initial matrix as found in Experiment 2.

Specifically, when the initial matrix was uniform, the switch rate of the produced matrix was 0.44 ($SD=0.12$), reliably below 0.5 ($t(49)=3.61, p<0.001, d=0.51$), but when the initial matrix was random, the switch rate of the produced matrix ($M=0.54, SD=0.04$) was reliably above 0.5 ($t(49)=6.43, p<0.001, d=0.91$). When the initial matrix was fully alternating, the switch rate of the produced matrix ($M=0.63, SD=0.10$) was reliably above 0.5 ($t(49)=10.01, p<0.001, d=1.42$).

Additionally, the average switch rate of the produced matrices in this task was not different from that of the temporal task ($t(49)=1.16, p=0.25, d=0.23$) or spatial task ($t(49)=0.47, p=0.64, d=0.07$) shown in Figure 4a. This further provides evidence for the consistency of the over-alternating bias across different task domains and probing methods.

General discussion

The goal of this study was to examine consistency in the subjective concept of randomness across different domains. Across three experiments, we found a highly stable over-alternation bias across presentation modes (temporal vs spatial), feature dimensions (color vs shape), sensory modalities (visual vs auditory), speed (fast vs slow), stimulus size (small vs large matrices) and probing methods (adjusting the generating process vs individual bits). These results suggest that the subjective concept of randomness is consistent in the face of vast stimulus variations. In addition, we found a strong anchoring effect in all experiments. Specifically, the switch rate of the sequences that were deemed as most random was correlated with the starting switch rate of the sequence (Experiment 1). Moreover, the switch rate of the produced matrix was lower when the starting matrix was fully uniform than when the starting matrix was fully alternating or random (Experiments 2 and 3). Despite the anchoring effect, the over-alternation bias was consistent across all experiments.

The over-alternation bias observed in this study was less pronounced than that in previous studies on randomness judgments (Bar-Hillel & Wagenaar, 1991; Falk & Konold, 1997; Lopes & Oden, 1987; Nickerson & Butler, 2009; Wagenaar, 1972; Zhao et al., 2014). The switch rate of the stimuli that were deemed maximally random in our tasks ranged from 0.52 to 0.63, whereas in most previous studies the switch rate was above 0.6. The relatively low switch rate might be driven by the anchoring effect. Since the starting switch rate of the stimuli was around 0.5 for all experiments, this initial anchor may have weakened the over-alternation bias toward true randomness, lowering the final switch rate of the stimuli that were judged as maximally random. However, it is worth noting that the anchoring effect automatically alters perception without subjective awareness (Tulving & Schacter, 1990). In our experiments, such bias to anchor to the initial sequence affected the produced sequence, but bias might be largely implicit in our participants which could automatically influence their conscious decision of how random the stimulus looked, although we did not have direct evidence on their awareness.

The most noteworthy finding of this study was that the over-alternation bias was consistently observed across various stimulus and task domains. This consistency suggests that people's concept of randomness is immune to differences in the stimuli used to embody randomness and in the elicitation methods used to express randomness. People's conception of randomness must therefore have a stable abstract character, applying similarly to distinct physical domains. There were, however, variations in the strength of the over-alternation bias. For example, the bias was smaller in the color dimension (in Experiment 3) than in the shape or the tone dimensions. One possible explanation is that the contrast between the two colors may be more salient and thus better encoded than a switch in shapes or tones, resulting in a more accurate expression of randomness.

What explains the consistent over-alternation bias? One explanation focuses on the limitations of working memory (Baddeley, 1966; Kareev, 1992). People can only hold a limited number of items in working memory, which means that the amount of bits processed at a given moment in time is constrained. This is especially true when people process temporal sequences. On the other hand, the over-alternation bias was less prominent with large spatial matrices, where people can sample large amounts of information simultaneously. This finding is consistent with Hahn and Warren's (2009) explanation, that is, within finite sequences (less than 500 bits), alternations are more likely to occur than streaks. Thus, the over-alternation bias could be explained by people's limited perceptual experiences with the environment. This also suggests that as people sample more information in large matrices (with 6400 bits), the over-alternation bias should be reduced, which was supported by our findings.

In addition to the working memory account, people may assume equal frequency of outcomes within a local sequence they can sample due to local representativeness (Tversky & Kahneman, 1971). Such emphasis on local equality for a limited number of bits held in working memory can cause the over-alternation bias. If local equality is assumed for short sequences (as in small matrices or temporal sequences in our experiments), the enforced balance of frequency of each outcome would increase the switch rate of the sequence. This account is supported by the findings in all experiments where the switch rate was consistently above 0.5.

In conclusion, this study demonstrates a highly consistent over-alternation bias in people's randomness concept across presentation modes, feature dimensions, sensory modalities, stimulus speed, stimulus size and probing methods.

Acknowledgements

For assistance with data collection, we thank Yu Luo. We also thank the Zhao Lab for helpful comments.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by NSERC Discovery Grant (RGPIN-2014-05617 to J.Z.), the Canada Research Chairs program (to J.Z.), the Leaders Opportunity Fund from the Canadian Foundation for Innovation (F14-05370 to J.Z.) and the NSERC Canada Graduate Scholarship Master's program and Elizabeth Young Lacey Fellowship (to R.Q.Y.).

Notes

1. The encouragement to observe long stretches of the sequence was a hedge against unlucky outputs from $D(p)$. It is possible even for $D(0.5)$ to relentlessly produce the same bit over and over; however, this is unlikely. (Notice that all sequences of the same length have the same probability of random generation.) There is thus no guarantee that bad luck does not infect the results reported below, only low probability that such is the case. It was made clear to the participants that a maximally random sequence is most likely generated by a random process (e.g., a fair coin).
2. It is difficult to compare our paradigms with a temporal production task where participants generate a random sequence. This is because participants are not exposed to any initial sequence in the temporal production task, and thus, their produced sequence is free from any anchoring effect. Therefore, we did not include a temporal production task in this experiment. Nonetheless, to provide evidence for our concern, we did run a temporal production task where each participant produced a 100-bit sequence of Ts and Hs. The average switch rate of the produced sequence was 0.68 (standard deviation [SD]=0.12), reliably above 0.5 (t=11.54, p<0.001, d=1.43). This switch rate is consistent with previous studies on randomness production (Bar-Hillel & Wagenaar, 1991), but much higher than the switch rate in our tasks.

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.

Eagle, A. (2014). Chance versus randomness. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved from http://plato.stanford.edu/archives/spr2014/entries/chance-randomness

Earman, J. (1986). *A primer on determinism*. Dordrecht, Holland: D. Reidel.

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

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*, 499–504.

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

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

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

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.

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., & Butler, S. F. (2009). On producing random binary sequences. *American Journal of Psychology*, *122*, 141–151.

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.

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.

Scholl, S. G., & Greifeneder, R. (2011). Disentangling the effects of alternation rate and maximum run length on judgments of randomness. *Judgment and Decision Making*, *6*, 531.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, *247*, 301–306.

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

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

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

Zhao, J., Hahn, U., & Osherson, D. (2014). Perception and identification of random events. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 1358.