

Alternation Blindness in the Representation of Binary Sequences

Ru Qi Yu
University of British Columbia

Daniel Osherson
Princeton University

Jiaying Zhao
University of British Columbia

Binary information is prevalent in the environment and contains 2 distinct outcomes. Binary sequences consist of a mixture of alternation and repetition. Understanding how people perceive such sequences would contribute to a general theory of information processing. In this study, we examined how people process alternation and repetition in binary sequences. Across 4 paradigms involving estimation, working memory, change detection, and visual search, we found that the number of alternations is underestimated compared with repetitions (Experiment 1). Moreover, recall for binary sequences deteriorates as the sequence alternates more (Experiment 2). Changes in bits are also harder to detect as the sequence alternates more (Experiment 3). Finally, visual targets superimposed on bits of a binary sequence take longer to process as alternation increases (Experiment 4). Overall, our results indicate that compared with repetition, alternation in a binary sequence is less salient in the sense of requiring more attention for successful encoding. The current study thus reveals the cognitive constraints in the representation of alternation and provides a new explanation for the overalternation bias in randomness perception.

Public Significance Statement

Binary information is a basic form of information people encounter on a daily basis. Being able to accurately represent binary information is fundamental to many cognitive operations. How do people process binary information? This study examines the ability to process alternations versus repetitions in binary sequences. Using 4 different paradigms (number estimation, memory recall, change detection, and visual search), we found a robust underrepresentation of alternations compared with repetitions. This result reveals a fundamental limitation in the cognitive system in processing binary sequences, and helps explain the biases people have in perceiving randomness (i.e., overalternation bias).

Keywords: alternation bias, attention, numerosity perception, randomness perception, working memory

An event is called “binary” if it is restricted to two possible outcomes, such as the result of a coin flip or a sports event. A “sequence” of binary events means an array of these outcomes, extended in time or space, such as a string of bits produced by a

compiler. Understanding how people perceive such sequences would contribute to a general theory of information processing.

Research on the perception of binary information has focused on sequences that exhibit different degrees of randomness (Bar-Hillel & Wagenaar, 1991; Julesz, 1962; Lopes & Oden, 1987; Nickerson, 2002). But what is randomness? Despite the difficulty in defining (Beltrami, 1999; Earman, 1986; Fitelson & Osherson, 2012), or even in subjective terms (Ayton et al., 1989; Lopes, 1982; Oskarsson et al., 2009), previous studies have revealed systematic biases in the perception of randomness. These biases include the gambler’s fallacy (Kahneman & Tversky, 1972), and the hot hand fallacy (Gilovich et al., 1985). Another bias that has received considerable attention concerns the tendency to judge sequences as “random” despite alternating more than expected by chance (Bar-Hillel & Wagenaar, 1991; Falk & Konold, 1997; Lopes & Oden, 1987; Nickerson, 2002). Likewise, when people are asked to produce random sequences, their output tends to contain too many alternations and consequently runs that are too short (Baddeley, 1966; Kahneman & Tversky, 1972; Wagenaar, 1972). This overalternation bias is especially present when people process temporal sequences, compared with spatial ones (Yu et al., in press). More

This article was published Online First August 17, 2017.

Ru Qi Yu, Department of Psychology, University of British Columbia; Daniel Osherson, Department of Psychology, Princeton University; Jiaying Zhao, Department of Psychology and Institute for Resources, Environment, and Sustainability, University of British Columbia.

We thank Yi-Chia Chen for her comments on A', and the Zhao Lab for helpful comments. This work was supported by NSERC Discovery Grant (RGPIN-2014-05617 to Jiaying Zhao), the Canada Research Chairs program (to Jiaying Zhao), the Leaders Opportunity Fund from the Canadian Foundation for Innovation (F14-05370 to Jiaying Zhao), and the NSERC Canada Graduate Scholarship Master’s program and Elizabeth Young Lacey Fellowship (to Ru Qi Yu).

Correspondence concerning this article should be addressed to Ru Qi Yu, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC, Canada, V6T 1Z4. E-mail: ruqiyu@psych.ubc.ca

generally, the bias is robust across different stimulus domains, sensory modalities, and presentation modes (Yu et al., in press).

Several accounts have been proposed to explain overalternation bias. One explanation focuses on the limits of working memory (Baddeley, 1966; Kareev, 1992). A sequence in memory is constructed to be random by ensuring that every small segment looks “choppy.” The result is excessive alternation. This account resembles the local equality hypothesis put forth by Tversky and Kahneman (1971). According to this idea, people assume equal frequency of outcomes within a random sequence that fits into short-term memory (STM), and such sequences have a tendency to alternate too often. A recent account argues that biases of randomness reflect people’s limited perceptual experiences with the environment (Hahn & Warren, 2009; Miller & Sanjurjo, 2015).

A different approach is advanced by Falk and Konold (1997). They posit a monotonic relationship between the probability that a given sequence is judged random and the time needed to correctly memorize or copy it. That is, a sequence looks random to the extent that the perceiver experiences cognitive difficulty when encoding it. This hypothesis has been challenged, however, by the finding that for some pairs of stimuli, the one that is easier to distinguish from a copy is judged to be more random on a separate occasion (Zhao et al., 2014).

Here we explore a new explanation for the overalternation bias, focusing on a limitation in how people represent binary sequences. The limitation lies in the relative ability to detect and represent alternations as opposed to repetitions. If alternations are less salient than repetitions, then objectively more alternations are needed for a sequence to look random. Thus, our explanation identifies the overalternation bias as a processing limitation, rather than a conceptual limitation. The perceiver’s conception of randomness might be accurate but applied to faulty messages delivered by perception (for a similar view, see Rapoport & Budescu, 1992). Of course, it is also possible that distortions arise at both perceptual and conceptual levels. Our account is motivated by the distinction between the ability to perceive randomness and the ability to identify randomness (Zhao et al., 2014).

The Present Study

The goal of the present study is to examine how people represent alternations and repetitions in a binary sequence. To generate binary sequences that contain different levels of alternations and repetitions while maintaining equal probability of the two outcomes, we used an algorithm that deviates from stochastic independence by allowing previous bits to influence the next one. This algorithm allowed us to manipulate the probability of seeing a repeat versus a switch. Specifically, for each number p in the unit interval (from 0 to 1), let $D(p)$ generate a sequence of bits consisting of zeros and ones 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 desired length.

This algorithm was first introduced by Zhao, Hahn, and Osherson (2014). It can be seen that $D(.5)$ is a genuinely random device. For $p < .5$, $D(p)$ tends to repeat itself, resulting in long streaks,

whereas for $p > .5$, $D(p)$ tends to alternate. The expected proportion of each bit is 50% for all $p \in [0, 1]$, although empirically, the output might deviate from 50%. For any sequence produced by $D(p)$, the expected proportion of alternation, called the “switch rate” of the generating process, is p . The expected proportion of repetitions, called the generating “repeat rate”, is $1 - p$. See Yu et al., (in press) for more discussion of $D(p)$.

We conducted four experiments using a range of different paradigms to examine how people represent alternations versus repetitions. There was no mention to participants of randomness in any experiment. In Experiment 1, participants viewed a binary sequence and estimated the number of switches or repeats in the sequence, and we measured their estimation accuracy. In Experiment 2, participants briefly viewed a binary sequence and then recalled the sequence bit for bit, and we measured their recall accuracy. In Experiment 3, participants viewed two sequences and judged whether the sequences were the same or different; we measured their sensitivity to change. In Experiment 4, participants searched for a target embedded in a binary sequence; we measured their response time to gauge their attention to switches versus repeats.

Experiment 1

The goal of this experiment was to examine whether there are systematic differences in the estimation of alternations and repetitions in a binary sequence. Specifically, we presented participants binary sequences, either temporally or spatially, and asked them to estimate the number of switches or repeats in each sequence.

Method

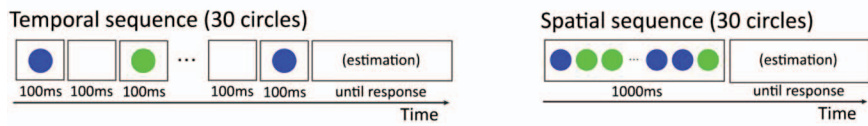
Participants. Forty-five undergraduate students (32 female, mean age = 19.9 years, $SD = 2.3$) from the University of British Columbia (UBC) participated for course credit. Participants in all experiments provided informed consent. All experiments reported here have been approved by the UBC Behavioral Research Ethics Board. We conducted a power analysis in G*Power (Faul, Erdfelder, Lang, & Buchner, 2007), using an effect size of $\eta_p^2 = 0.53$ observed in our prior work using similar methods and analyses (Zhao & Yu, 2016). In the previous work, participants estimated the number of dots on the screen for each trial and the data were analyzed using a repeated-measures ANOVA. Based on the power analysis, a minimum of 38 participants would be required to have 95% power to detect the effect in our paradigm with an alpha level of 0.05.

Apparatus. In this and subsequent experiments, participants were seated 50 cm away from a computer monitor (refresh rate = 60Hz). Stimuli were presented using MATLAB and the Psychophysics Toolbox (<http://psycho toolbox.org>).

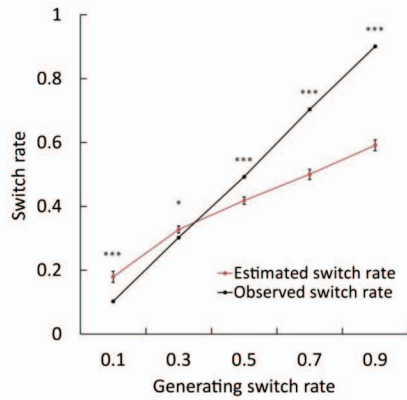
Stimuli. In each trial, participants viewed a 30-bit sequence. Each sequence was binary, containing circles of two different colors: green (RGB value: 0 255 0) and blue (RGB value: 0 0 255). Each circle subtended 0.9° in diameter (Figure 1a). There were five levels of switch rates in $D(p)$ in generating the sequences, where $p = .1, 0.3, 0.5, 0.7$, and 0.9 . Correspondingly, there were five levels of repeat rates $(1 - p) = 0.9, 0.7, 0.5, 0.3$, and 0.1 .

Temporal sequences. For half of the trials, participants viewed a temporal sequence where the 30 circles were presented

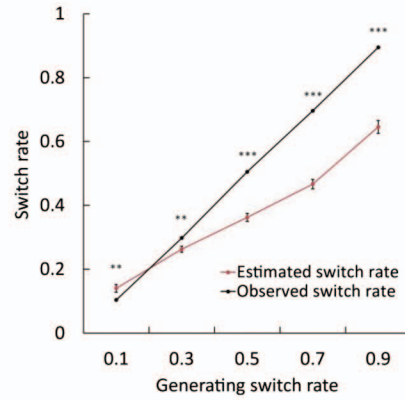
a) Task: estimating the number of color switches or repeats



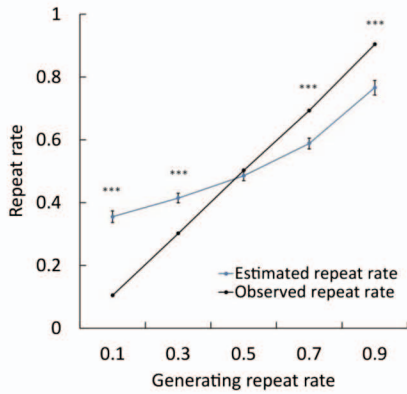
b) Temporal trials: estimation of switches



c) Spatial trials: estimation of switches



d) Temporal trials: estimation of repeats



e) Spatial trials: estimation of repeats

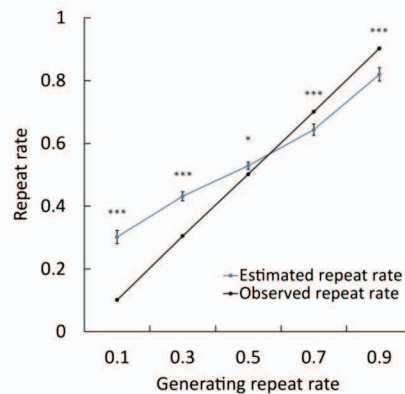


Figure 1. Experiment 1. (a) Participants ($N = 45$) were presented with temporal sequences or spatial sequences of green and blue circles. Each sequence contained 30 circles. In temporal sequences, the circles were presented one at a time, and in spatial sequences the circles were presented simultaneously on the screen. For each sequence, participants were asked to estimate either the number of circles that had a different color from the previous circle (switch) or the number of circles that had the same color as the previous circle (repeat). (b) The estimated switch rate and the observed switch rate were plotted against the five levels of generating switch rates for temporal trials. (c) The estimated switch rate and the observed switch rate were plotted against the five levels of generating switch rates for spatial trials. (d) The estimated repeat rate and the observed repeat rate were plotted against the five levels of generating repeat rates for temporal trials. (e) The estimated repeat rate and the observed repeat rate were plotted against the five levels of generating repeat rates for spatial trials. (Error bars reflect ± 1 SEM; * $p < .05$. ** $p < .01$. *** $p < .001$). See the online article for the color version of this figure.

one after another over time. Each circle was presented at the center of the screen for 100ms, and the interstimulus interval (ISI) was 100 ms with a blank screen (Figure 1a).

Spatial sequences. For the other half of the trials, participants viewed a spatial sequence, where the 30 circles were presented on the screen simultaneously. The circles in the sequence were arranged from left to the right. The space between two adjacent circles in the sequence subtended 0.1° . Each sequence was presented on the screen for 1,000ms (Figure 1a).

Procedure. There were 200 trials in total for each participant. In each trial, participants viewed a sequence with one of the five generating switch rates (0.1, 0.3, 0.5, 0.7, or 0.9). Each level of switch rate contained 40 trials, among which 20 trials were temporal sequences and 20 trials were spatial sequences. After viewing the 30-bit sequence, participants were asked to estimate either the number of the color switches (10 trials), or the number of color repeats (10 trials). Specifically, the instruction for estimating color switches was “How many times did a dot have a DIFFERENT

color from the previous dot in the sequence?” and the instruction for estimating color repeats was “How many times did a dot have the SAME color as the previous dot in the sequence?”. Participants were also told that the range of their estimate was from 0 to 29 (29 was the maximum possible number of switches or repeats in the sequence). Participants typed in their estimate after seeing each sequence. In sum, there were three within-subjects factors: the generating switch rate of the sequence (from 0.1 to 0.9), the presentation of the sequence (temporal vs. spatial), and the estimation type (switches vs. repeats). The order of the trials was randomized for each participant. There was no mention of randomness in all experiments.

Results and Discussion

Before the analyses, we should define three types of switch rates for each sequence. *Estimated switch rate* was the derived by dividing the estimated number of switches by 29 (the maximum possible switches in the sequence). Likewise, *estimated repeat rate* was calculated by dividing the estimated number of repeats by 29 (the maximum possible repeats). For example, if the participant reported that there were 10 switches in the sequence, the estimated switch rate would be $10/29 = 0.34$. If the participant reported that there were 15 repeats in the sequence, the estimated repeat rate would be $15/29 = 0.52$. *Observed switch rate* was the objective switch rate in the sequence presented to the participants in each trial. This was calculated by dividing the objective number of switches in the sequence by 29. Likewise, *observed repeat rate* was the objective repeat rate (number of repeats divided by 29) in the sequence presented in each trial. The *generating switch rate* was the p in $D(p)$ in the algorithm that generated the sequence. The *generating repeat rate* was $1 - p$. To verify that the presented sequence actually exhibited the generating switch rate or repeat rate, we plotted the observed switch rate or repeat rate for each sequence (Figure 1b–1e), which mapped closely to the generating switch rate or repeat rate. The estimated and observed switch rate and repeat rate were plotted in Figure 1b and 1d for temporal sequences, and in Figure 1c and 1e for spatial sequences.

We computed the signed error between the estimated and the observed switch rate or repeat rate at each of the five generating rates, separately for temporal trials and spatial trials. This signed error (estimated—observed) served as our measure. The goal of the analysis was to examine whether the signed error was different when participants were estimating the number of switches or repeats across the five generating rates.¹ For temporal trials (Figure 1b and 1d), a 5 (generating rate: 0.1, 0.3, 0.5, 0.7, and 0.9) \times 2 (estimation type: switches vs. repeats) repeated-measures ANOVA revealed a main effect of generating rate [$F(4, 176) = 162.3, p < .001, \eta_p^2 = 0.79$] and of estimation type [$F(1, 44) = 49.34, p < .001, \eta_p^2 = 0.53$], and a reliable interaction [$F(4, 176) = 10.75, p < .001, \eta_p^2 = 0.20$]. Post hoc Tukey’s HSD analysis showed that across the five generating rates, all pairwise comparisons were significant [p ’s $< .001$]. Pairwise comparisons at each generating rate showed that participants underestimated the number of switches more than repeats at each of the five generating rates [p ’s $< .01$]. As revealed by a linear contrast model, the signed error for both repeat and switch estimation linearly decreased in value as the generating rates increased from 0.1 to 0.9 ($p < .001$). To

further assess the strength of this linear relationship, we correlated the signed error with the generating rates for each participant, and found that the error negatively correlated with the generating rate (across participants, mean $r = -0.88$ for repeat estimation and -0.92 for switch estimation), showing that as the number of repeats or switches increased in the sequence, the greater the underestimation was.

For spatial trials (Figure 1c and e), the same ANOVA revealed a main effect of generating rate [$F(4, 176) = 107.2, p < .001, \eta_p^2 = 0.71$] and of estimation type [$F(1, 44) = 114.2, p < .001, \eta_p^2 = 0.72$], but no interaction [$F(4, 176) = 0.07, p = .99, \eta_p^2 < 0.01$]. Post hoc Tukey’s HSD analysis showed that across the five generating rates, all pairwise comparisons were significant [p ’s $< .001$], except between 0.7 and 0.9 [$p = .62$]. Again, pairwise comparisons at each generating rate showed that participants underestimated the number of switches more than repeats at each of the five generating rates [p ’s $< .001$]. As revealed by a linear contrast model, the signed error for both repeat and switch estimation linearly decreased in value as the generating rates increased from 0.1 to 0.9 ($p < .001$). Moreover, we found that the signed error negatively correlated with the generating rate across participants (mean $r = -0.82$ for repeat estimation and -0.87 for switch estimation), showing that as the number of repeats or switches increased in the sequence, the greater the underestimation was.

To further explore the signed error, we compared the estimated switch or repeat rate with the observed switch or repeat rate. For temporal trials (Figure 1b), participants overestimated the switch rate at 0.1 and 0.3, but underestimated the switch rate at 0.5, 0.7, and 0.9. They also overestimated the repeat rate at 0.1 and 0.3, but underestimated the repeat rate at 0.7 and 0.9 (Figure 1d). For spatial trials (Figure 1c), participants overestimated the switch rate only at 0.1, and underestimated the switch rate at 0.3, 0.5, 0.7, and 0.9. They overestimated the repeat rate at 0.1, 0.3, and 0.5, but underestimated the repeat rate at 0.7 and 0.9 (Figure 1e).

The intersection of the estimated curve and the observed curve signals the point where participants made the most accurate estimation. Interestingly, when estimating the number of repeats, participants were the most accurate around 0.5 where the sequences were truly random. For the same random sequence at 0.5, participants were significantly underestimating the number of switches. In fact, for people to perceive a 0.5 switch rate, the sequence must contain more than 50% switches, with a switch rate of around 0.7 (Figure 1b and 1c). This underestimation of switches may underlie the conceptual overalternation bias of randomness. Taken together, these results suggest that alternations in a binary sequence were consistently underrepresented compared with repetitions.

¹ Theoretically, a fully random sequence with a switch rate of 0.5 contains the maximal level of entropy. As the switch rate approaches 1 or 0, entropy declines. This was confirmed by a Kolmogorov complexity analysis, and we found that as the generating switch rate becomes closer to 0.5, the more entropy the sequences contain ($r = .94$). Given this high correlation, we think that the generating switch rate is an adequate indicator of entropy, and therefore did not directly assess people’s performance against entropy.

Experiment 2

One explanation for the underestimation of switches could be due to a failure in working memory. Specifically, people may not be able to hold alternating bits accurately in working memory, mistaking them for repeating bits, thus leading to underestimation. To examine this possibility, we conducted Experiment 2 where participants were asked to recall each sequence.

Method

Participants. Forty-five undergraduate students (30 female, mean age = 19.6 years, $SD = 1.2$) from UBC participated for course credit. We conducted a power analysis in G*Power (Faul et al., 2007), using the effect size from Experiment 1 ($\eta_p^2 = 0.53$). We found that a minimum of 38 participants would be required to have 95% power to detect the effect in our paradigm with an alpha level of 0.05. Thus, we kept the same sample size as in Experiment 1.

Stimuli. The stimuli presented to the participants were the same as those in Experiment 1, except for the following three differences: (a) there were 10 circles per sequence instead of 30, to circumvent a floor effect in the recall task; (b) each circle was slightly larger, subtending 1.4° in diameter, and the distance between each circle in spatial sequences was 0.2° ; and (c) each spatial sequence was presented for 500ms (Figure 2a).

Procedure. The procedure was identical to that in Experiment 1, except for one critical difference: after seeing each sequence, participants were asked to recall the sequence they just saw as accurately as they could, by pressing two different keys to produce the green circles (the “G” key) and the blue circles (the “B” key). Participants were instructed to recall the dots in the same order as they appeared. To recall each bit in a temporal sequence, participants pressed one key and the corresponding circle was presented on the screen for 100 ms after each press, and then disappeared. To recall each bit in a spatial sequence, participants pressed one key and the corresponding circle was presented from left to right on the screen after each key press, and remained on the screen.

Results and Discussion

As shown in Experiment 1, the observed switch rate of the sequences mapped closely onto the generating switch rates. Thus, for all following experiments task performance was plotted against the five generating switch rates.

To assess the accuracy of participants’ recalled sequences, we divided the exact matches between the presented sequence and the recalled sequence by 10. The accuracy was plotted over the five levels of switch rates. For temporal trials (Figure 2b), a one way repeated-measures ANOVA revealed a significant difference in accuracy across the five switch rates [$F(4, 176) = 75.61, p < .001, \eta_p^2 = 0.63$]. Post hoc Tukey’s HSD analysis showed all pairwise comparisons were significant except between 0.7 and 0.9, and 0.5 and 0.9. As revealed by a linear contrast model, recall accuracy linearly decreased as the switch rate increased from 0.1 to 0.9 ($p < .001$). We also found that across participants recall accuracy negatively correlated with the switch rate (mean $r = -0.86$). For spatial trials (Figure 2c), accuracy was different across the switch rates [$F(4, 176) = 111.5, p < .001, \eta_p^2 = 0.72$], and post hoc Tukey’s HSD analysis showed that all pairwise comparisons were

significant except between 0.7 and 0.9. As revealed by a linear contrast model, recall accuracy linearly decreased as the switch rate increased from 0.1 to 0.9 ($p < .001$). Across participants recall accuracy negatively correlated with the switch rate (mean $r = -0.86$). These results demonstrate that as the switch rate of the sequence increased, recall accuracy decreased.

To obtain a more fine-grained comparison between the recall of switches and repeats, we performed two more analyses. From the second bit on in each sequence, we calculated the recall accuracy of each bit depending on whether the bit repeated the previous bit, or switched from the previous bit. We compared the recall accuracy of switching versus repeating bits. For temporal trials (Figure 2d), a 5 (generating rate: 0.1, 0.3, 0.5, 0.7, and 0.9) \times 2 (bit type: repeating vs. switching) repeated-measures ANOVA showed a main effect of generating rate [$F(4, 176) = 75.61, p < .001, \eta_p^2 = 0.63$] and of bit type [$F(1, 44) = 206.7, p < .001, \eta_p^2 = 0.82$], and a reliable interaction [$F(4, 176) = 37.4, p < .001, \eta_p^2 = 0.46$]. Pairwise comparisons at each generating rate showed that the recall accuracy of repeating bits was consistently higher than that of switching bits [p ’s $< .01$]. For spatial trials (Figure 2e), the same ANOVA showed a main effect of generating rate [$F(4, 176) = 111.5, p < .001, \eta_p^2 = 0.46$] and of bit type [$F(1, 44) = 28.84, p < .001, \eta_p^2 = 0.40$], and a reliable interaction [$F(4, 176) = 7.18, p < .001, \eta_p^2 = 0.14$]. Pairwise comparisons at each generating rate showed that the recall accuracy of repeating bits was higher than that of switching bits [p ’s $< .001$] at switch rates 0.1, 0.3, and 0.5.

One problem with the accuracy measure based on exact matches was that it penalizes cases where participants reversed one bit but were nonetheless accurate. For example, take a presented sequence at switch rate 0.9, 010101011, the participant might encode the overall switchiness of the sequence and produced a recall sequence, 101010100. Based on exact matches, the accuracy would be 0, but the recalled sequence still resembled the presented sequence in its overall switch rate. To circumvent this problem, we conducted another analysis where we calculated the switch rate of the recalled sequence, and compared that to the observed switch rate of the presented sequence (Figure 2f and 2g).

We computed signed error (switch rate of the recalled sequences—observed switch rate) separately for temporal and spatial trials. For temporal trials (Figure 2f), a one-way repeated-measures ANOVA revealed a significant difference in signed error across the five generating switch rates [$F(4, 176) = 140.7, p < .001, \eta_p^2 = 0.76$]. Post hoc Tukey’s HSD analysis showed all pairwise comparisons were significant except between 0.1 and 0.3, and 0.1 and 0.5, suggesting that errors were greater at higher switch rates. For spatial trials (Figure 2g), the same ANOVA revealed a significant difference in signed error across the five switch rates [$F(4, 176) = 92.54, p < .001, \eta_p^2 = 0.68$]. Post hoc Tukey’s HSD analysis showed all pairwise comparisons were significant except between 0.1 and 0.3, and 0.1 and 0.5, suggesting errors were greater at higher switch rates. In addition, comparisons between the switch rate of the recalled sequences and the observed switch rate showed a reliable difference at every generating rate, except at 0.5 for temporal trials and 0.1 and 0.5 for spatial trials. This analysis suggests that at 0.5 when the sequence contained the same amount of switches and repeats, participants could still recall the sequence with the correct switch rate.

A caveat in this experiment was the balance of the two outcomes in the sequence. At switch rate 0.1, there was on

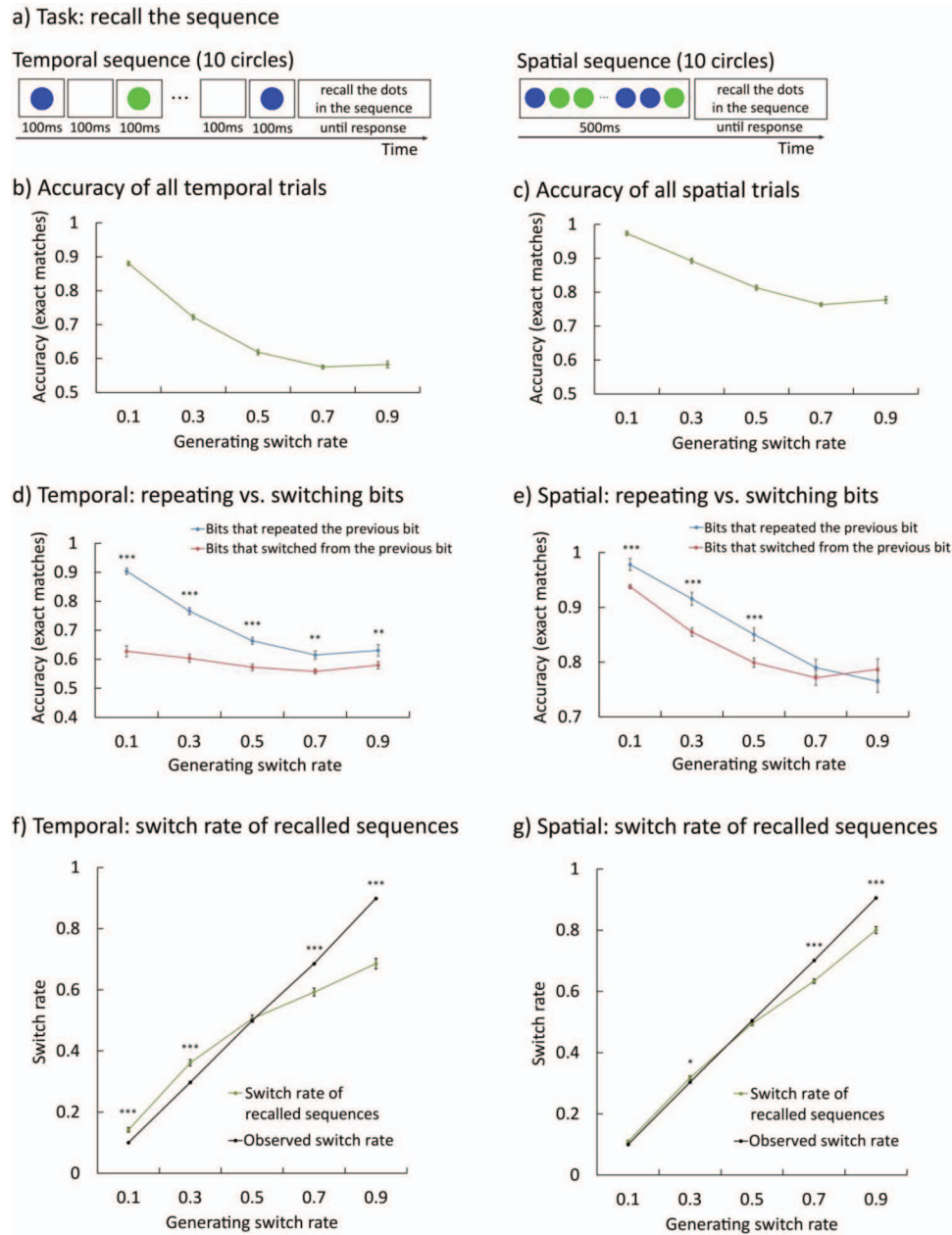


Figure 2. Experiment 2. (a) Participants ($N = 45$) were presented with 10-bit sequences of green and blue circles in temporal or spatial sequences. Participants were asked to recall the dots in each sequence after seeing the sequence. Accuracy was calculated as the proportion of exact matches in the dots between the presented sequence and participants' recalled sequence, for temporal trials (b) and spatial trials (c). From the second bit on in each sequence, we calculated the recall accuracy of each bit depending on whether the bit repeated the previous bit, or switched from the previous bit. This accuracy was plotted across the five generating switch levels for temporal sequences (d) and spatial sequences (e). For participants' recalled sequences, we also calculated the switch rate of the recalled sequences, plotted with observed switch rate of the presented sequences across the five generating switch levels for temporal trials (f) and spatial trials (g). (Error bars reflect ± 1 SEM; * $p < .05$. ** $p < .01$. *** $p < .001$). See the online article for the color version of this figure.

average only 1 switch in the 10-bit sequence. As the switch could occur anywhere in the sequence, the sequence could be 0000011111 or 0000000001 (where 0 and 1 here are blue and green circles). The first sequence had an equal frequency of

outcomes, but the second was highly unbalanced. Thus, at switch rate 0.1 the frequency of the two outcomes could be a potential confound. To address this issue, we separated sequences at switch rate 0.1 into 11 groups, from sequences with

0 blue circles, 1 blue circle, to the sequences with 10 blue circles. For temporal trials, a repeated-measures ANOVA revealed a significant difference in recall accuracy in exact matches across the 11 levels [$F(10, 331) = 7.50, p < .001, \eta_p^2 = 0.18$]. Post hoc Tukey's HSD analysis revealed that the accuracy at 50% (5 blue circles and 5 green circles) was only lower than that at 0%, 90%, or 100% (0, 9, or 10 blue circles), but not different from the other frequencies. For spatial trials, a repeated-measures ANOVA revealed a significant difference in recall accuracy across the 11 levels [$F(10, 331) = 2.67, p = .004, \eta_p^2 = 0.07$]. Post hoc Tukey's HSD analysis revealed that the accuracy at 50% (5 blue circles and 5 green circles) was also lower than that at 0%, 90%, or 100% (0, 9, or 10 blue circles), but not different from the other frequencies. This means that for highly repeating sequences, recall accuracy for balanced sequences was not that different from unbalanced sequences, except for the extremes.

These results revealed three findings: (a) as the sequence became more alternating, recall accuracy diminished; (b) people were better at recalling repeating bits than switching bits in a sequence; and (c) as the sequence became more alternating, the recalled sequence had fewer switches. The greater recall error in switching bits compared with repeating bits suggests that people are more likely to encode switches as repeats, than to encode repeats as switches. This encoding difference could explain the greater underestimation of switches compared with repeats in Experiment 1.

Experiment 3

What explains the encoding difficulty of switching bits? One explanation is that switching bits may be less salient than repeating bits, and thus are more difficult to process. To examine salience, in Experiment 3 we used a change detection task where participants detected changes in two binary sequences that were presented one after another.

Method

Participants. Forty-five undergraduate students (24 female, mean age = 20.6 years, $SD = 1.8$) from UBC participated for course credit. We conducted a power analysis in G*Power (Faul et al., 2007), using the effect size from Experiment 2 ($\eta_p^2 = 0.63$). We found that a minimum of 22 participants would be required to have 95% power to detect the effect in our paradigm with an alpha level of 0.05. Thus, we kept the same sample size for the subsequent experiments as in Experiment 2.

Stimuli and procedure. There were 200 trials in total. In each trial, participants were presented with two back-to-back sequences of 15 green and blue circles (Figure 3a). The color and size of the circles were identical to those used in Experiment 2. The sequences were generated with one of the five switch rates (0.1 to 0.9) as before. There were 40 trials per switch rate, 20 of which contained a change where the color of one randomly selected circle was different between the two sequences, and 20 of which contained no change where the two sequences were the same. In each

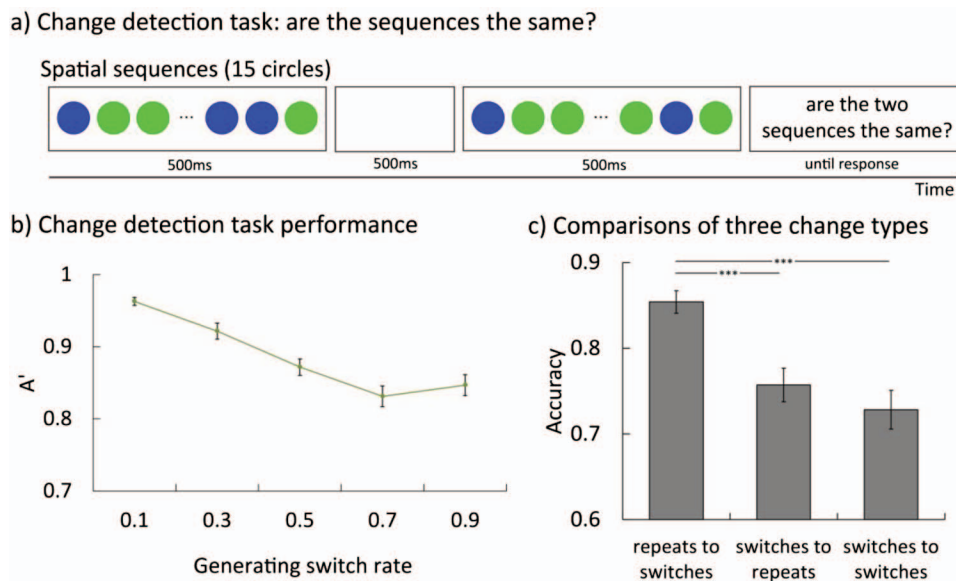


Figure 3. Experiment 3. (a) Participants ($N = 45$) were presented with two back-to-back sequences. There were 15 blue and green circles in each sequence. In half of the trials, the two sequences differed in the color of one circle, and for the other half the two sequences were the same. Participants were asked to judge whether the two sequences were the same or different after seeing the sequences. (b) Each participants' performance was assessed using A' across the 5 generating switch rates. (c) Trials with changes were categorized into three change groups: 1. repeats to switches (e.g., 000 to 001, 010, or 100), 2. switches to repeats (e.g., 010, 001, or 100 to 000), and 3. switches to switches (e.g., 001 to 011 or 101, 010 to 110 or 011, 100 to 101 or 110). The accuracy in the change detection task was compared across the three types of changes. (Error bars reflect ± 1 SEM; *** $p < .001$). See the online article for the color version of this figure.

trial, all circles in the first sequence were presented simultaneously at the center of the screen for 500 ms, with an ISI of 500 ms, followed by the second sequence also presented for 500 ms. Participants had to judge whether the two sequences were the same or different by pressing the “Y” key or the “N” key, respectively. The trials were presented in a random order.

Results and Discussion

To examine the performance of the change detection task, we calculated A' based on the nonparametric method proposed by Pollack and Norman (1964). A' was plotted across the five generating switch rates (Figure 3b). There was a reliable difference in A' across the five rates [$F(4, 176) = 40.64, p < .001, \eta_p^2 = 0.48$]. Post hoc Tukey's HSD analysis showed all pairwise comparisons were significant except for between 0.5 and 0.9, or 0.7 and 0.9. As revealed by a linear contrast model, A' in the change detection task quartically decreased as the switch rate increased from 0.1 to 0.9 ($p < .001$). Across participants, A' negatively correlated with the switch rate (mean $r = -0.72$).

In addition, we examined change detection accuracy depending on the local environment where the change occurred. For all trials with a change, we categorized them into three groups: repeats to switches (e.g., 000 to 001, 010, or 100), switches to repeats (e.g., 010, 001, or 100 to 000), and switches to switches (e.g., 001 to 011 or 101, 010 to 110 or 011, 100 to 101 or 110). Because we only considered trials where a change occurred, there was no false alarm. Therefore, we used accuracy as the measure here (Figure 3c). Among the three types changes, there was a reliable difference in accuracy [$F(2, 88) = 55.95, p < .001, \eta_p^2 = 0.56$]. Post hoc Tukey's HSD analysis showed that accuracy in the repeats to switches group was reliably higher than that in the switches to repeats and switches to switches groups [p 's $< .001$].

As in Experiment 2, we examined whether the frequency of the outcomes influenced performance for sequences at switch rate 0.1. We separated these highly repeating sequences into 16 groups, from sequences with 0 blue circles, 1 blue circle, to the sequences with 15 blue circles. A repeated-measures ANOVA did not reveal a significant difference in A' across the 16 levels [$F(15, 313) = 1.56, p = .08, \eta_p^2 = 0.07$]. Thus, at switch rate 0.1 there was no difference in change detection performance due to the frequency of the two outcomes.

These results showed that as the sequence became more alternating, a change in the sequence was harder to detect. This suggests that repetitions were more salient than alternations. Moreover, a change was more salient when a streak was interrupted, than when an alternating pattern became streaky or remained alternating. This differential performance suggests that people may have paid more attention to the streak presented in the first sequence, than to the switches presented in the first sequence.

Experiment 4

Experiment 3 provided some evidence suggesting that alternations were less salient than repetitions. To provide further support for this account, in Experiment 4 we used a visual search task to measure attention to switching versus repeating sequences. If participants were faster at finding the target in a repeating sequence than in a switching sequence, this would suggest that

repetitions draw more attention than alternations. In contrast, if participants were faster at finding target in a switching sequence than a repeating sequence, this would suggest that alternations draw attention more than repetitions.

Method

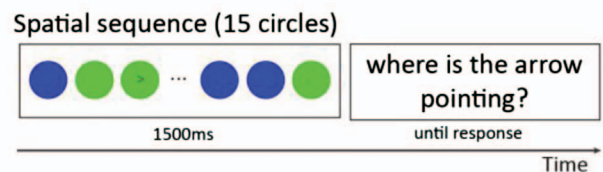
Participants. Forty-five undergraduate students (33 female, mean age = 19.6 years, $SD = 2.1$) from UBC participated for course credit.

Stimuli and procedure. As in Experiment 3, there were 200 trials, and in each trial, a sequence containing 15 colored circles were presented simultaneously on the screen. One of the randomly selected circles contained a target (a red arrow pointing left “<” or right “>”) superimposed on the circle. The target appeared at the same time as the sequence. As before, the sequences were generated with one of the five switch rates, and there were 40 trials per switch rate. For each trial, participants had to search for the target and to identify the direction at which the arrow was pointing as fast and as accurately as they could (Figure 4a). Half of the trials contained an arrow pointing left, and the other half contained an arrow pointing right. Each sequence was presented for 1500ms. The trials were presented in a random order.

Results and Discussion

The accuracy of the target search task was high ($M = 97.5%$, $SD = 2%$). Thus, we only examined the response times of correct

a) Task: target search



b) Target search performance

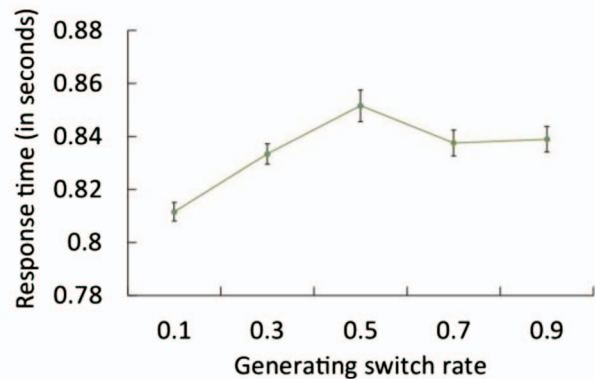


Figure 4. Experiment 4. (a) Participants were presented with 15-bit sequences of green and blue circles. The target was a small red arrow, pointing either to the left or right, in one of the circles. Participants were asked to report the direction of the arrow as fast and as accurately as they could. (b) Response time of correct trials was plotted over the 5 generating switch rates. (Error bars reflect ± 1 SEM). See the online article for the color version of this figure.

trials as our measure of attention (Figure 4b). There was a reliable difference in response time across the five switch rates [$F(4, 176) = 2.55, p < .05, \eta_p^2 = 0.05$]. Post hoc Tukey's HSD analysis showed a reliable difference in response times only between switch rates 0.1 and 0.5. As revealed by a linear contrast model, the response time in the visual search task quartically increased as the switch rate increased from 0.1 to 0.9 ($p < .001$). Across participants, the response time positively correlated with the switch rate (mean $r = .22$). This result showed that participants were faster to find the target in sequences with more repetitions than with more switches. This suggests that repeating sequences may draw attention more strongly than switching sequences.

As in previous experiments, we examined whether the frequency of the outcomes influenced performance for sequences at switch rate 0.1. We separated these highly repeating sequences into 16 groups, from sequences with 0 blue circles, 1 blue circle, to the sequences with 15 blue circles. A repeated-measures ANOVA revealed a significant difference in response times across the 16 levels [$F(15, 592) = 2.19, p = .006, \eta_p^2 = 0.05$], but post hoc Tukey's HSD analysis did not reveal any difference between individual levels. Thus, at switch rate 0.1 there was no difference in visual search performance due to the frequency of the two outcomes.

General Discussion

The goal of the current study was to examine how people represent alternations versus repetitions in a binary sequence. Across four experiments, we found that the number of alternations was underestimated more strongly than the number of repetitions (Experiment 1). This underestimation of switches could be explained by constraints in working memory, since recall accuracy diminished as the sequence became more alternating (Experiment 2). The greater encoding difficulty of alternations could be explained by the possibility that alternations are less salient than repetitions. This was supported by the finding that changes were harder to detect as the sequence became more alternating (Experiment 3). Finally, visual targets were slower to be found as the sequence became more alternating, suggesting that alternating sequences draw attention less strongly than repeating sequences (Experiment 4). Overall, these results from four different paradigms (i.e., estimation, working memory, change detection, and visual search tasks) converge to the same finding that people are more blind or insensitive to alternations than to repetitions, which suggests that alternations are underrepresented compared with repetitions.

It is important to note that our findings do not speak to a strictly perceptual phenomenon, since our measures involve a combination of perceptual and memory processes. Specifically, estimating the number of alternations from a spatial sequence involves an immediate assessment of the sequence which is perceptually available, but estimation from a temporal sequence involves enumerating over information held in working memory (Experiment 1). Recalling the sequence involves retrieving information from working memory (Experiment 2). Change detection involves a comparison between the second sequence which is perceptually available and the first sequence held in working memory (Experiment 3). Visual search involves the detection of a target in the sequence which is perceptually available (Experiment 4). Thus, our findings reveal

biases in the representation of binary sequences, which encapsulates both perceptual and memory processes.

The current findings provide evidence for the new account on the overalternation bias. Specifically, there is a limitation in the ability to accurately represent alternations as opposed to repetitions in a binary sequence. This means that for people to represent a 0.5 switch rate, the sequence must contain more than 50% alternations (in fact around 70% given our current data). For production tasks, the overalternation bias can be explained by the possibility that people underestimate the alternations in their own productions of random sequences, and therefore may try to compensate by generating more alternations. However, it remains likely that people may produce sequences by matching their productions to a truly random sequence (e.g., tosses of a fair coin) they have experienced before. If this were the case, then they would end up generating a random sequence matching the one they have experienced, because the underestimation of alternations should occur in both the experienced random sequence and the produced sequence. Thus, the production should in theory match people's external experience. This is supported by the finding in Experiment 2, where the switch rate of the recalled sequence at 0.5 was also 0.5.

Why are alternations underrepresented compared with repetitions? We offer two explanations. First, two alternating bits (e.g., 10) may be perceptually more complex than two repeating bits (e.g., 11), and this higher complexity in an alternation could be more difficult to encode. Second, people may implicitly chunk an alternation into a unit (e.g., perceiving 101010 as three chunks of 10, Zhao & Yu, 2016), but rely on numerosity perception for repetitions (e.g., perceiving 111111 as 1 repeating five times). Both accounts would lead to an underrepresentation of alternations compared with repetitions.

We should note that we were careful not to mention randomness or related concepts at all in our experiments. This is because when primed with the notion of randomness, people process the binary sequences differently than if they were not, producing shorter streaks (Olivola & Oppenheimer, 2008). We wanted to ensure that our experiments were strictly assessing how people represent binary sequences, not how people conceive randomness.

The results from Experiments 1 and 2 suggest that participants might encode distinct types of summary statistics at different levels of accuracy from a binary sequence. For example, at switch rate 0.5, the sequence contained the same amount of alternations and repetitions. Yet, participants severely underestimated the switch rate but were more or less accurate at estimating the repeat rate (Experiment 1). When recalling the sequence, they were very accurate at producing the same switch rate 0.5 in their produced sequence (Experiment 2). This suggests that participants could implicitly encode a global switch rate of the binary sequence, but when verbally asked, they were less accurate at reporting the number of switches in the sequence.

The current study reveals a basic limitation in the representation of alternations. The study is important in several ways: first, it provides a new explanation of the overalternation bias in the subjective concept of randomness; second, it reveals new insights on the limits in the representation of binary information; and finally, the same finding was replicated in four different paradigms using different measures. The current findings shed light on how

people process binary information, which is fundamental to understanding the limits of the cognitive system.

References

- Ayton, P., Hunt, A. J., & Wright, G. (1989). Psychological conceptions of randomness. *Journal of Behavioral Decision Making*, 2, 221–238. <http://dx.doi.org/10.1002/bdm.3960020403>
- Baddeley, A. D. (1966). The capacity for generating information by randomization. *The Quarterly Journal of Experimental Psychology*, 18, 119–129. <http://dx.doi.org/10.1080/14640746608400019>
- Bar-Hillel, M., & Wagenaar, W. A. (1991). The perception of randomness. *Advances in Applied Mathematics*, 12, 428–454. [http://dx.doi.org/10.1016/0196-8858\(91\)90029-1](http://dx.doi.org/10.1016/0196-8858(91)90029-1)
- Beltrami, E. (1999). *What is random? Chance and order in mathematics and life*. New York, NY: Springer-Verlag. <http://dx.doi.org/10.1007/978-1-4612-1472-4>
- Earman, J. (1986). *A primer on determinism*. Dordrecht, the Netherlands: D. Reidel. <http://dx.doi.org/10.1007/978-94-010-9072-8>
- Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review*, 104, 301–318. <http://dx.doi.org/10.1037/0033-295X.104.2.301>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191. <http://dx.doi.org/10.3758/BF03193146>
- Fitelson, B., & Osherson, D. (2012). *Remarks on random sequences*. Retrieved from <http://arxiv.org/abs/1205.5865>.
- Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, 17, 295–314. [http://dx.doi.org/10.1016/0010-0285\(85\)90010-6](http://dx.doi.org/10.1016/0010-0285(85)90010-6)
- Hahn, U., & Warren, P. A. (2009). Perceptions of randomness: Why three heads are better than four. *Psychological Review*, 116, 454–461. <http://dx.doi.org/10.1037/a0015241>
- Julesz, B. (1962). Visual pattern discrimination. *I. R. E. Transactions on Information Theory*, 8, 84–92. <http://dx.doi.org/10.1109/TIT.1962.1057698>
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430–454. [http://dx.doi.org/10.1016/0010-0285\(72\)90016-3](http://dx.doi.org/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. <http://dx.doi.org/10.1037/0096-1523.18.4.1189>
- 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. <http://dx.doi.org/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. <http://dx.doi.org/10.1037/0278-7393.13.3.392>
- Miller, J. B., & Sanjurjo, A. (2015). *Is it a fallacy to believe in the hot hand in the NBA three-point contest?* Innocenzo Gasparini Institute for Economic Research, Bocconi University, Working Paper number 548.
- Nickerson, R. S. (2002). The production and perception of randomness. *Psychological Review*, 109, 330–357.
- Olivola, C. Y., & Oppenheimer, D. M. (2008). Randomness in retrospect: Exploring the interactions between memory and randomness cognition. *Psychonomic Bulletin & Review*, 15, 991–996. <http://dx.doi.org/10.3758/PBR.15.5.991>
- 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. <http://dx.doi.org/10.1037/a0014821>
- Pollack, I., & Norman, D. A. (1964). A non-parametric analysis of recognition experiments. *Psychonomic Science*, 1, 125–126. <http://dx.doi.org/10.3758/BF03342823>
- Rapoport, A., & Budescu, D. V. (1992). Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology: General*, 121, 352–363. <http://dx.doi.org/10.1037/0096-3445.121.3.352>
- Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76, 105–110. <http://dx.doi.org/10.1037/h0031322>
- Wagenaar, W. A. (1972). Generation of random sequences by human subjects: A critical survey of the literature. *Psychological Bulletin*, 77, 65–72. <http://dx.doi.org/10.1037/h0032060>
- Yu, R., Gunn, J., Osherson, D., & Zhao, J. (in press). The consistency of the subjective concept of randomness. *Quarterly Journal of Experimental Psychology*.
- Zhao, J., Hahn, U., & Osherson, D. (2014). Perception and identification of random events. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 1358–1371. <http://dx.doi.org/10.1037/a0036816>
- Zhao, J., & Yu, R. Q. (2016). Statistical regularities reduce perceived numerosity. *Cognition*, 146, 217–222. <http://dx.doi.org/10.1016/j.cognition.2015.09.018>

Received February 17, 2017

Revision received June 3, 2017

Accepted June 11, 2017 ■