Judgment and Decision Making, Vol. 6, No. 6, August 2011, pp. 531–541

Disentangling the effects of alternation rate and maximum run length on judgments of randomness

Sabine G. Scholl*   Rainer Greifeneder†

Abstract

Binary sequences are characterized by various features. Two of these characteristics—alternation rate and run length—have repeatedly been shown to influence judgments of randomness. The two characteristics, however, have usually been investigated separately, without controlling for the other feature. Because the two features are correlated but not identical, it seems critical to analyze their unique impact, as well as their interaction, so as to understand more clearly what influences judgments of randomness. To this end, two experiments on the perception of binary sequences orthogonally manipulated alternation rate and maximum run length (i.e., length of the longest run within the sequence). Results show that alternation rate consistently exerts a unique effect on judgments of randomness, but that the effect of alternation rate is contingent on the length of the longest run within the sequence. The effect of maximum run length was found to be small and less consistent. Together, these findings extend prior randomness research by integrating literature from the realms of perception, categorization, and prediction, as well as by showing the unique and joint effects of alternation rate and maximum run length on judgments of randomness.

Keywords: alternation rate, maximum run length, judgments of randomness, binary sequences.

1 Introduction

In everyday life, individuals often encounter binary sequences of outcomes, such as tossing of coins, gains and losses in gambles, black and red outcomes on roulette wheels, birth orders in families, or ups and downs in the stock market (see Oskarsson, Van Boven, McClelland, & Hastie, 2009). Whether such sequences are perceived as either random or lawful and patterned has important implications, both on the level of cognitive processes (e.g., categorizations, inferences) and on the level of behavioral manifestations (e.g., following up on a "streak"). Understanding how individuals form judgments of randomness is therefore critical. The present manuscript contributes to this goal by investigating the effects of two critical features of binary sequences—the length of the longest run and the alternation rate—on judgments of randomness.

Imagine a coin being tossed five times. First, the coin comes up heads. Then, heads again. The same is true for the third, fourth, and fifth toss, resulting in a sequence of five heads (HHHHH). Most individuals who perceive this sequence would judge it as non-random, because they evaluate randomness not based on the generating mechanism—a coin toss—but on the resulting sequence of outcomes, such as HHHHH or HTHTH (Falk & Konold, 1997). To understand how judgments of randomness are formed, it is therefore important to analyze the features of outcome sequences.

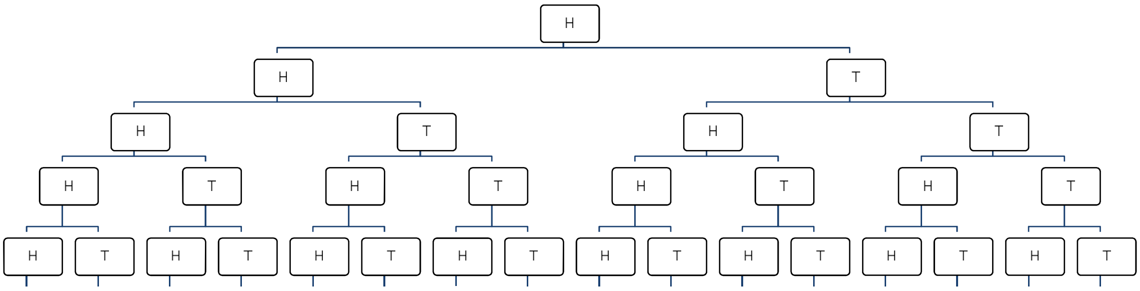
Two characteristics of binary sequences of outcomes, alternation rate and run length, have repeatedly been identified as determining judgments of randomness. Usually, however, these two characteristics have been investigated in separate lines of research, without controlling for the other. Specifically, research on the perception and categorization of randomness primarily has focused on alternation rate (e.g., Ayton & Fischer, 2004; Falk & Konold, 1997; Lopes & Oden, 1987; McDonald, 2009; Wagenaar, 1970; Wiegersma, 1982, 1987), while neglecting run length. In contrast, research on the prediction of randomness has primarily focused on run length (e.g., Altmann & Burns, 2005; Ayton & Fischer, 2004; Croson & Sundali, 2005; Edwards, 1961; Nicks, 1959), while neglecting alternation rate. Because the two features are correlated but not identical, we argue that it is critical to isolate the separate effects of the two characteristics so as to understand more clearly what influences judgments of randomness. Here we offer evidence that the two characteristics may not only have unique effects, but that the effect of one may be contingent on the other. In particular, we argue that the effect of alternation rate is contingent on run length. To lay the foundation for this argument, we first introduce the concepts of alternation rate and (maximum) run length.

We are grateful to Marcia Duriska and Sarah Lohmüller for their help in data collection. We thank the editor and one anonymous reviewer for their helpful comments on an earlier version.

*Sabine G. Scholl (née Czenna), Department of Psychology, School of Social Science, University of Mannheim, 68131 Mannheim, Germany. E-mail: sabine.scholl@uni-mannheim.de.

†Department of Psychology, University of Mannheim.

Figure 1: The probability tree illustrates the possible outcomes of a sequence of 5 coin tosses. For the sake of brevity only the 16 sequences starting with heads and their characteristics—relative frequency of heads and tails, longest run within the sequence, number of runs within the sequence, and alternation rate—are displayed.



Sequence	HHHHH	HHHHT	HHHTH	HHHTT	HHTHH	HHTHT	HHTTH	HHTTT	HTHHH	HTHHT	HTHTH	HTHTT	HTTHH	HTTHT	HTTTH	HTTTT
Rel. frequency	5:0	4:1	4:1	3:2	4:1	3:2	3:2	2:3	4:1	3:2	3:2	2:3	3:2	2:3	2:3	1:4
Longest run	5	4	3	3	2	2	2	3	3	2	1	2	2	2	3	4
Number of runs	1	2	3	2	3	4	3	2	3	4	5	4	3	4	3	2
P(alternation)	0	.25	.50	.25	.50	.75	.50	.25	.50	.75	1.0	.75	.50	.75	.50	.25

1.1 Characteristics of random sequences and their interrelations

Coin tosses often serve as models for random devices (for a discussion, see Ford, 1983). If a coin is tossed 5 times, 2^5 = 32 different sequences of head and tail can result. Figure 1 shows a probability tree illustrating the possible outcomes of a sequence of 5 coin tosses. For the sake of brevity, only the 16 sequences starting with heads are displayed. While the probability of occurrence for each single sequence is the same, namely $1/2^n$ (here: $1/2^5$ = 1/32), the 5-toss sequences differ in their internal structure, which can be described in terms of runs or in terms of alternations.

A *run* is defined as a series of identical consecutive elements preceded or followed by a different element or no element at all, e.g., HHHHT, THHHT, or THHHH. The *run length*[1] is the number of times the repeated element occurs, e.g., HTHTH; HHTHT, HHHTH, HHHHT, or HHHHH. The *number of runs* r within a sequence is the total number of different runs, e.g., the sequence HHHTT contains 2 runs, namely 1 run of length three (HHH) and 1 run of length two (TT).

Finally, an *alternation* is a change between consecutive elements, i.e., HHHHT or TTTTH. The *alternation rate* relates the actual number of alternations between consecutive elements to the maximum number of possible alternations in the sequence. For instance, the sequence HHHTT contains 2 runs (r)—1 run of three consecutive heads (HHH), and 1 run of two consecutive tails (TT)—hence, (r-1) = 2–1 = 1 alternation between head and tail. As the sequence is the result of 5 coin tosses (n), there is a maximum of (n-1) = (5–1) = 4 alternations between head and tail. The ratio of actual alternations to the maximum of possible alternations is assessed by the formula P(A) = (r-1)/(n-1), which results in a rate of alternation of P(A) = (2–1)/(5–1) = .25 for the current example.

Run lengths and alternation rate are negatively correlated in most samples of sequences. The longer the runs in a sequence of a given length, the smaller the alternation rate. Long runs reduce the total number of runs and accordingly the alternation rate. For instance, the sequence HTHTH contains five runs of length 1 and has an alternation rate of 1; in contrast, the sequence HHHHH contains only one run of length 5 and has an alternation rate of 0. Note, however, that alternation rate and run length are not perfectly correlated. Any alternation rate may be

[1]The term "run length" may refer to average run length or to the length of single runs within a sequence. The *average run length* of a sequence is calculated by multiplying each type of run length by its frequency of occurrence, by summing up the resulting products and by dividing this sum by the number of runs r within the sequence. This means relating the number of elements n within a sequence to the number of runs r. As indicated by the formula $M_{run\ length} = n/r$, the sequence HHHTT is characterized by an average run length of 2.5. Averaging, however, neglects the length of the various runs within a sequence. For instance, the sequences HHHTT and HHHHT are both characterized by a mean run length of 2.5, but differ in the length of single runs within the sequence: While the former contains a run of 3 heads (HHH) and a run of 2 tails (TT), the latter contains a run of 4 heads (HHHH) and a run of 1 tail (T). Neglecting the length of single runs is problematic as longer runs are more perceptually salient (Wiegersma, 1987). In the following the term "run length" refers to the length of single runs within a sequence. This is in line with studies on the prediction of randomness which also focus on the length of single runs.

associated with a variety of different run lengths. As illustrated in Figure 1, an alternation rate of .50 may—in our example—contain either one run of length 3 and two runs of length 1 (e.g., HHHTH) or two runs of length 2 and one run of length 1 (e.g., HHTTH). Alternation rate and run length are thus not equivalent, so that—when investigating which characteristics influence judgments of randomness—it is critical to isolate their unique and possibly contingent effects.

1.2 Judgments of randomness

Numerous studies have investigated which characteristics of binary sequences influence individuals' judgments of randomness. Usually, however, these studies have focused on either alternation rate or run length. Specifically, alternation rate has primarily been investigated in studies on the perception and categorization of randomness, while run length has primarily been investigated in studies on the prediction of randomness.

Studies on the *perception of randomness* generally investigate the ability to recognize randomness by using binary sequences of outcomes as stimuli (e.g., XXXOOOOXXOOOXOXOXXOXX). In these studies, various sequences are presented and participants are asked to infer whether a given sequence is random (Altmann & Burns, 2005; Olivola & Oppenheimer, 2008), to order sequences according to their perceived randomness (Falk & Konold, 1997; McDonald, 2009), or to select the only random (Kunzendorf & Pearson, 1984), the most random (Konold, Pollatsek, Well, & Lohmeier, 1993; Wagenaar, 1970; Wiegersma, 1982, 1987), or the least random (Green, 1982; Konold et al., 1993) sequence out of a set of sequences. *Categorization studies* also present several sequences of outcomes to their participants, but ask their participants, for instance, whether a given sequence was produced by a random or nonrandom process (Ayton & Fischer, 2004; Lopes & Oden, 1987). Note that perception and categorization studies usually evaluate responses by focusing on alternation rate (e.g., Ayton & Fischer, 2004; Falk & Konold, 1997; Lopes & Oden, 1987; McDonald, 2009; Wagenaar, 1970; Wiegersma, 1982, 1987) without controlling for the impact of maximum run length or by focusing on maximum run length (Olivola & Oppenheimer, 2008) without controlling for alternation rates.

One key finding of perception and categorization studies is that participants judge sequences with an alternation rate of .60 to .70 as most random or most likely to be produced by a random process. In particular, judgments of randomness usually first increase and then decrease with an alternation rate of .60 to .70 as the transition point. This tendency towards "overalternation" is generally referred to as "negative recency effect" (for overviews, see Bar-Hillel & Wagenaar, 1991; Falk & Konold, 1997; Nickerson, 2002).

Prediction studies typically emphasize the importance of run length. They generally follow a different methodological approach by asking participants to predict single outcomes of random devices such as coin tosses. Studies focus primarily on the length of a run of a particular outcome (e.g., Altmann & Burns, 2005; Ayton & Fischer, 2004; Croson & Sundali, 2005; Edwards, 1961; Nicks, 1959) without controlling for alternation rates.

While the results of manipulating alternation rate have been reliably replicated across different studies, investigations of run length offer a less coherent set of results: many studies identified negative recency, but there are also some studies that reported on positive recency or no effects. For instance, Ayton and Fischer (2004) have identified a negative linear trend for run lengths varying from 1 to 5, indicating that the longer the run of a particular outcome, the less likely participants were to predict that outcome on the next trial. Similarly, investigating gambling behavior in a casino, Croson and Sundali (2005) reported a negative trend for increasing run length. Gamblers were more likely to bet against a run than to bet with it after experiencing a run of 5 or more consecutive repeated outcomes (but not for runs of length 1 to 4). An opposite trend was reported by Edwards (1961), who investigated run lengths varying from 1 to 8. For runs of up to four events, a positive trend occurred in that participants were more likely to predict the same event the longer the run; for runs of four and more events, however, predictions were not influenced by run length. Finally, Olivola and Oppenheimer (2008) did not find an effect of run lengths 5 versus 7 on judgments of randomness.

While most studies focused either on alternation rate or on run length, only a few studies have considered both concepts within one design. None of these earlier studies, however, allows for inferences about the characteristics' unique impact on randomness judgments and whether the influence of one characteristic is contingent on the other characteristic. For instance, McDonald (2009) investigated the effect of causal beliefs on predictions of future outcomes and varied the alternation rate, while controlling only for the length of the run at the very end of each sequence and not the length of the other runs within the sequence (which could be longer than the final run). Although conducive to the study's goals, that approach precludes conclusions about the unique impact of alternation rate versus maximum run length. Similarly, Huettel, Mack and McCarthy (2002) addressed the question of how the disruption of alternation-patterns versus run-patterns influenced reaction times and blood oxygen levels in a reaction task. Alternation patterns were characterized by exhibiting an alternation rate of 1, while run-patterns were characterized by exhibiting an alternation rate of 0. Because that study did not investigate varying degrees of alternation, it can only suggest the relative importance of length of alternation- to run-patterns and cannot disentangle their impact in sequences of varying alternation rates. Finally, Green (1982) asked his participants which characteristics of binary sequences they *thought* they had used in making randomness judgments. He classified participants' answers as referring to relative frequency, run pattern, or alternation rate. Verbalizing reasons, however, does not allow for disentangling the impact of alternation rate and (maximum) run length on randomness judgments. Moreover, verbal reports are often seen as varieties of introspection (Nisbett & Wilson, 1977) and thus less reliable than inferring the effect of different characteristics from participants' responses, as in these experiments.

1.3 The present research

This manuscript aims to integrate results from studies investigating the effect of alternation rate as well as those investigating the effect of run length on judgments of randomness. Extending prior research, we investigate the unique effects of alternation rate and maximum run length, as well as their contingent effect on randomness judgments within a single experimental design. We examine the effect of run length by manipulating the longest run within the sequence. This methodological choice reflects the fact that long (in comparison to short) runs are perceptually more salient. Because *alternation rate* has previously been shown to have consistent effects (see Bar-Hillel & Wagenaar, 1991), we predicted a main effect of alternation rate on randomness judgments. With regard to *maximum run length*, the heterogeneity of previous results does not allow for specific predictions. We therefore investigate the main effect of maximum run length on randomness judgments in an exploratory fashion only.

Going beyond unique main effects, we hypothesized that the effect of alternation rate on judgments of randomness would be contingent on maximum run length. In particular, we assumed that the effect of alternation rate would decrease with increasing maximum run lengths. To substantiate this argument, a conceptual parallel can be drawn to Gestalt Theory (Wertheimer, 1923). Gestalt research has revealed that, in heterogeneous visual fields, individuals tend to group similar elements together (e.g., elements similar in form, size, or color: J. Beck, 1967; Houtkamp & Roelfsema, 2010). and that these clusters of similar elements become apparent against a background of rather dissimilar elements. For instance, when individuals inspect a picture consisting of white and black dots, they perceive a mosaic of white and black clusters instead of a grey picture. In a review on judgments of randomness, Bar-Hillel and Wagenaar (1991) have argued that such clustering tendencies are responsible for misperceptions of randomness in binary matrices (see also Feller, 1950; Wiegersma, 1987). We argue that organizing tendencies apply not only to binary matrices but also to sequences of consecutive elements, as in a series of coin tosses. Specifically, we expected that elements in binary sequences would appear to cluster with identical previous and subsequent elements. We assumed these runs of identical elements would become more apparent against a background of more irregular elements (Wiegersma, 1987; see also D. M. Beck & Palmer, 2002; Palmer & Beck, 2007).

We propose that the perceptual salience of runs is determined by two factors—run length and the context in which the run occurs: First, it has generally been argued that the more identical elements co-occur, the easier grouping is and the more perceptually salient a cluster—here a run—becomes (e.g., see Treisman & Gormican, 1988, for the length of lines). Similar results have been reported for attention (Vitz & Todd, 1969), suspiciousness regarding "true" randomness (Lepley, 1963; Wiegersma, 1987), and the belief that a streak has occurred (Bar-Hillel & Wagenaar, 1991; Carlson & Shu, 2007; Wagenaar, 1970). Second, perceptual salience is also determined by the context in which a run occurs. The longer the run within a sequence of fixed length and fixed alternation rate (the context), the more salient it is.

To summarize, we suggest that, when similar elements co-occur, individuals focus on co-occurrence rather than on change, which leads to certain consequences for the perception of runs in binary sequences: The longer the longest run, the greater the focus on it, and the less focus on both other (shorter) runs and the alternation between these runs. Accordingly, the longer the longest run, the less the influence of alternation on randomness judgments. Statistically, this prediction translates into an interaction effect between maximum run length and alternation rate.

We test hypotheses across two experiments while controlling for potential confounding effects of symbol type and varying the position of the longest run within the sequence. We expected to find a negative recency effect in that randomness judgments are negatively skewed as a function of alternation rate and peak for alternation rates bigger than .50. This is reflected either in a linear or in a linear plus a quadratic relationship between alternation rate and judged randomness in the current experiments. Furthermore, we expected the effect of alternation rate on randomness judgments to be negatively linearly related to maximum run length, such that short (long) runs result in a more (less) pronounced impact of alternation rate on judgments of randomness.

2 Experiment 1

Experiment 1 was designed to disentangle the unique and contingent effects of alternation rate and maximum run length on judgments of randomness. Participants evaluated the randomness of binary sequences consisting of 21 elements in a perception task. Sequences consisted of Xs and Os to minimize conceptual or linguistic influences. Furthermore, the ratio of Xs and Os was close to .50. Alternation rate and the maximum run length within the sequences were manipulated orthogonally.

2.1 Method

Thirty-two University of Mannheim students (16 females; mean age = 23.6 years, *SD* = 4.0) participated in return for 1 EUR in a 3 (alternation rate .40 vs. .50 vs. .60) x 3 (maximum run length 3 vs. 4 vs. 5) x 2 (complement X vs. O) within-subjects design.

In a computer based study, participants were informed that they would see 72 binary sequences and that each sequence was produced by tossing a coin 21 times. Instructions explained that a coin toss was used as the generating device, because prior research has shown that randomness judgments are based on beliefs about the generating process; for instance, whether a sequence was produced by humans or inanimate devices (Ayton & Fischer, 2004), or by the assumed randomness of the generator (Burns & Corpus, 2004; Tyszka, Zielonka, Dacey, & Sawicki, 2008).

Sequences consisted of 11 Xs and 10 Os (or 10 Xs and 11 Os) with sequences starting and ending with the more frequent element (for a similar approach, see Ayton & Fischer, 2004; Falk & Konold, 1997). The 21 elements were presented all at once. Sequences varied in the alternation rate (.40 vs. .50 vs. .60) and in the length of the longest run within the sequence (for a similar approach in a study on the generation of randomness, see Kubovy & Gilden, 1991). Each sequence contained one longest run of length 3, 4, or 5 (plus two runs of length 3; all other runs were of length 2 or 1, e.g., XOOOOXOXOXXXOOXXXOXOX). The longest run was of length 3 or longer because Carlson and Shu (2007) have demonstrated that a run of three repeated consecutive symbols is critical to the perception of a streak (see also Bar-Hillel & Wagenaar, 1991) and because a run length of 3 is the shortest maximum run length possible for sequences characterized by an alternation rate of .40 and equal relative frequencies (e.g., OOXXXOOXXXOOOXXOOXXOO). At the same time, the longest run was of length 5 (or shorter) as a run length of 5 is the longest possible run length for sequences with an alternation rate of .60 and equal relative frequencies (e.g., OXOXXXXXOOOXOOOXOX-OXO). Note that implementing all maximum run lengths and the smallest and highest alternation rates possible for sequences characterized by equal relative frequencies allowed for direct comparisons of the effects of alternation rate and maximum run length.

To control for perceptual confounds, such as the possibility that a run in Os is perceived differently from a run in Xs, we used complementary sequences (e.g., XXOXXOXOXOOOXXOOOOXXX and OOXOOXOX-OXXXOOXXXXOOO). Examples of each of the 18 types of sequences are provided in Table 1. We always used four versions of each type of sequence (e.g., four sequences characterized by an alternation rate of .40, a maximum run length of 3, and complement X). To control further for a potential influence of the position of the longest run within the sequence, the longest run was either at the beginning of the sequence, middle-left, middle-right, or at the end of the sequence. The order in which sequences were presented to participants was determined randomly for every participant. Each sequence was presented individually and remained on the screen until judgments of randomness were made. Participants were asked to indicate how random they perceived each sequence of coin tosses to be. They provided their answers by mouse clicks on a 7-point Likert-type scale (1 = not random at all, to 7 = absolutely random; for a similar approach, see Altmann & Burns, 2005) that was presented underneath each of the sequences. After each judgment, participants started the next trial. At the end of the experiment, participants were debriefed, probed for guessing hypotheses, thanked, and remunerated.

2.2 Results

Preliminary analyses. Falk and Konold (1997) have proposed that, when evaluating randomness, a minority of individuals will focus on the generating device, whereas the majority will focus on the sequence of outcomes. Because those who focus on the generating device should not be influenced by variations in characteristics of the sequence, we first analyzed the variance each individual participant displayed across randomness judgments. Three participants (9.38 %) judged all sequences as equally random, providing only judgments of 7, presumably because they focused on the generating device, and were therefore excluded from further analyses (for a similar approach, see McDonald, 2009). The remaining 29 participants (90.62 %) showed a mean variance in randomness judgments of $M = 2.03$ ($SD = 1.34$).

Judgments of randomness. To disentangle the unique and joint effects of alternation rate and maximum run length on judgments of randomness, participants' judgments were subjected to a 3 (alternation rate .4 vs. .5 vs. .6) x 3 (maximum run length 3 vs. 4 vs. 5) x 2 (com-

Table 1: Examples of each of the 18 types of sequences in Experiment 1.

Type of sequence	Example
Alternation rate .40, max. run length 3, complement O	XXOOOXXOOOXXXOOXXOOXX
Alternation rate .40, max. run length 3, complement X	OOXXXOOXXXOOOXXOOXXOO
Alternation rate .40, max. run length 4, complement O	XXOOOXXXOOXXOOOOXXOXX
Alternation rate .40, max. run length 4, complement X	OOXXXOOOXXOOXXXXOOXOO
Alternation rate .40, max. run length 5, complement O	XXOOOOOXXXOXXOXXOOOXX
Alternation rate .40, max. run length 5, complement X	OOXXXXXOOOXOOXOOXXXOO
Alternation rate .50, max. run length 3, complement O	XXOXXXOOOXXOOXXOOOXOX
Alternation rate .50, max. run length 3, complement X	OOXOOOXXXOOXXOOXXXOXO
Alternation rate .50, max. run length 4, complement O	XXOXXOOOXOXOOOOXXXOXX
Alternation rate .50, max. run length 4, complement X	OOXOOXXXOXOXXXXOOOXOO
Alternation rate .50, max. run length 5, complement O	XXXOOXOOOOOXOXOXXXOXX
Alternation rate .50, max. run length 5, complement X	OOOXXOXXXXXOXOXOOOXOO
Alternation rate .60, max. run length 3, complement O	XXOOOXOXOOOXXXOXOXOXX
Alternation rate .60, max. run length 3, complement X	OOXXXOXOXXXOOOXOXOXOO
Alternation rate .60, max. run length 4, complement O	XXXOXOOXOXOOOOXXXOXOX
Alternation rate .60, max. run length 4, complement X	OOOXOXXOXOXXXXOOOXOXO
Alternation rate .60, max. run length 5, complement O	XOXOOOOOXXXOXXXOXOXOX
Alternation rate .60, max. run length 5, complement X	OXOXXXXXOOOXOOOXOXOXO

plement X vs. O) repeated measures design. Polynomial contrasts were used for testing our hypotheses:[2] specifically, whether participants exhibited a negative recency effect indicated either by a significant linear contrast for alternation rate (-1 0 1) or by a significant linear contrast plus a significant quadratic contrast for alternation rate (-1 0 1) (-1 2 -1). A significant linear contrast indicates that sequences with an alternation rate of .60 are rated as more random than sequences with an alternation rate of .40. A significant quadratic contrast indicates that randomness judgments curve for sequences with an alternation rate of .50. Furthermore, polynomial contrasts were used to test whether the effect of alternation rate decreased linearly with increasing maximum run length. Specifically, the linear effect of alternation rate (-1 0 1) and the quadratic effect of alternation rate (-1 2 -1) were expected to interact with the linear effect of maximum run length.

Analyses revealed the predicted unique effect of alternation rate: Sequences with an alternation rate of .60 were judged as more random than sequences with an alternation rate of .40, as indicated by the significant linear contrast, $F_{linear}(1, 28) = 30.80$, $p < .001$, $\eta_p^2 = .52$. Furthermore, randomness judgments curved for sequences with an alternation rate of .50, as shown by the quadratic contrast, $F_{quadratic}(1, 28) = 8.95$, $p < .01$, $\eta_p^2 = .24$.

As hypothesized, the strength of both the linear and the quadratic effect of alternation rate linearly decreased with increasing maximum run length (for a summary of statistics, see Table 2). This was reflected in a significant interaction between the linear effect of alternation rate and the linear effect of maximum run length, $F_{linear\text{-}linear}(1, 28) = 25.28$, $p < .001$, $\eta_p^2 = .47$, and a significant interaction between the quadratic effect of alternation rate and the linear effect of maximum run length, $F_{quadratic\text{-}linear}(1, 28) = 7.71$, $p < .02$, $\eta_p^2 = .22$. The linear contingency of the effect of alternation rate on maximum run length is illustrated in Figure 2. The interactions of alternation rate with the quadratic effect of maximum run length were not significant, *F*s < 2.56, *p*s > .14.

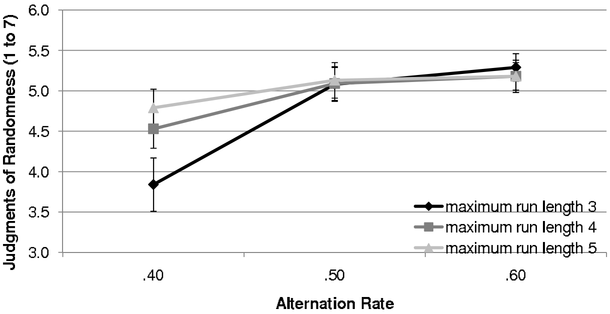
Analyses additionally revealed a tendency towards a unique effect for maximum run length, $F(1.35, 37.79) = 3.27$, $p < .07$, $\eta_p^2 = .10$. Post-hoc polynomial contrasts with a Bonferroni-adjusted alpha-level of .025 showed that randomness judgments tended to increase linearly with increasing maximum run length, $F_{linear}(1, 28) = 3.72$, $p < .07$, $\eta_p^2 = .12$, $F_{quadratic} < 1$. All other *F*s < 1.86, *p*s > .17.

[2]We report contrast statistics for the planned contrasts and repeated measurement statistics for the analyses regarding maximum run length and the perceptual control factor complement X versus O, so as to match hypotheses and statistical tests. In the repeated measures designs in Experiments 1 and 2, we also calculated the main effect for alternation rate and the interaction between alternation rate and maximum run length, which were significant for all reported contrasts, all *F*s > 6.25, *p*s < .001, η_p^2s > .12.

Table 2: Linear and quadratic effects of alternation rate for linearly increasing maximum run lengths of 3, 4, and 5 in Experiment 1.

Maximum run length	Linear effect of alternation rate	Quadratic effect of alternation rate
3	$F(1, 28) = 31.51$, $p < .001$, $\eta_p^2 = .53$	$F(1, 28) = 10.46$, $p < .01$, $\eta_p^2 = .27$
4	$F(1, 28) = 16.56$, $p < .001$, $\eta_p^2 = .37$	$F(1, 28) = 3.49$, $p < .08$, $\eta_p^2 = .11$
5	$F(1, 28) = 9.19$, $p < .01$, $\eta_p^2 = .25$	$F(1, 28) = 1.75$, *ns,* $\eta_p^2 = .06$

Figure 2: Means (with standard errors) of participants' randomness judgments in Experiment 1, separately for a maximum run length of 3, 4 and 5, and an alternation rate of .40, .50, and .60.



2.3 Discussion

To summarize, Experiment 1 shows that alternation rate exerts a unique effect on randomness judgments. Randomness judgments increased for sequences characterized by an alternation rate of .40 to .60, and curved for an alternation rate of .50. Importantly, the effect of alternation rate was contingent on the longest run within the sequence. The longer the longest run in the sequence, the lower the effect of alternation rate on randomness judgments. The unique effect for maximum run length was rather weak or non-existent—despite the fact that the chances of finding an effect were maximized by implementing all possible maximum run lengths for binary sequences of length 21 that are characterized by equal relative frequencies and alternation rates from .40 to .60.

It is noteworthy that Experiment 1 controlled for perceptual factors: specifically, we used complementary sequences to control for whether the longest run is perceived differently depending on whether it is made up of Xs or Os. Additionally, we took care that the position of the longest run within the sequence for each type of sequence varied from the left beginning of the sequence to the right end of the sequence. This high level of control allows for reasonable confidence in the results obtained. However, one may argue that the longest run length might have a systematic differential impact on randomness judgments depending on the position of the longest run within the sequence. Encountering a "run-pattern" or an "alternation-pattern" at the beginning of the sequence might result in a first "impression" of the sequence as being more or less random. To rule out a systematic impact, we ran a replication study (N = 30) which systematically *manipulated* the position of the longest run within the sequence by creating a mirror image of every single sequence (e.g., XXXOOOOXXOOOXOXOXXOXX became XXOXXOXOXOOOXXOOOOXXX)[3]. Substantiating the findings of Experiment 1, the results show that alternation rate exerted a unique impact on randomness judgments. Again, this effect was contingent on the longest run within the sequence. Neither maximum run length nor position of the longest run (see also Olivola & Oppenheimer, 2008) had a reliable impact on randomness judgments.

3 Experiment 2

To control for linguistic and conceptual factors, Experiment 1 employed the symbols X and O and specified that the sequence resulted from coin tosses. While this methodological choice reduced ambiguities in beliefs about the generating device, it may have appeared odd to those few who defined randomness based on the generating device rather than on the observed pattern of the outcome sequence (see Falk & Konold, 1997). To prevent potential misunderstandings and to further substantiate the findings of Experiment 1, no allusion to the generating device is made in Experiment 2. Instead, participants were asked to indicate whether they believe that a given sequence is random or not.

[3]Replication study: Randomness judgments were subjected to a 3 (alternation rate .40 vs. .50 vs. .60) x 2 (maximum run length 4 vs. 5) x 2 (complement X vs. O) x 2 (mirroring longest run left vs. longest run right) within-subjects design. Results replicated the findings of Experiment 1. 27 participants (90 %) showed variance in their randomness judgments and judged sequences with an alternation rate of .60 as more random than sequences with an alternation rate of .40, with judgments curving at .50 ($F_{linear}(1, 26) = 25.08$, $p < .001$, $\eta_p^2 = .49$, $F_{quadratic}(1, 26) = 8.03$, $p < .01$, $\eta_p^2 = .24$). Again, the impact of alternation rate on randomness judgments was contingent on maximum run length: The strength of the linear effect decreased with increasing maximum run length, $F_{linear\text{-}linear}(1, 26) = 8.28$, $p < .01$, $\eta_p^2 = .24$. Maximum run length did not exert a unique effect, $F < 1$.

A second change introduced in Experiment 2 is a new answering format. Specifically, participants were asked to make binary random/not random judgments, which are more easily interpreted and allow for testing whether the observed pattern of results holds for different ways of assessment.

In addition, we introduced two changes to the random sequences presented: alternation rates are varied in five steps from .42 to .74; maximum run lengths of 4 to 6 were implemented. To accommodate these changes, the sequences were extended to 39 elements.

3.1 Method

Forty-six University of Mannheim students (32 females; mean age = 22.5 years, *SD* = 2.5) participated in return for 1 EUR in a 5 (alternation rate .42 vs. .50 vs. .58 vs. .66 vs. .74) x 3 (maximum run length 4 vs. 5 vs. 6) x 2 (complement X vs. O) within-subjects design.

Participants were presented 90 binary sequences consisting of 20 Xs and 19 Os (and vice versa). Materials were the same as in Experiment 1, except for the following changes: First, alternation rate was manipulated on 5 levels to produce alternation rates that varied from .42 to .74. Second, maximum run length was manipulated on 3 levels from 4 to 6, with a length of 6 being the longest possible run length for alternation rates of .74. Third, the position of the longest run was controlled for in that it was either at the beginning, in the middle, or at the end of the sequence. Finally, there were always three (instead of four) versions of each type of sequence.

3.2 Results and discussion

Preliminary analyses. No participant (0 %) answered that all sequences were random.

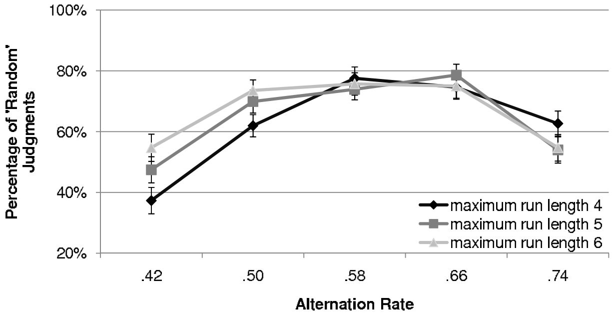
Judgments of randomness. We counted how often participants judged sequences of a given sequence type as random, related this sum to the 3 sequences per sequence type and multiplied the result by 100. This yielded a percentage measure varying between 0 (all sequences not random) and 100 (all sequences random). These percentage measures were subjected to a 5 (alternation rate .42 vs. .50 vs. .58 vs. .66 vs. .74) x 3 (maximum run length 4 vs. 5 vs. 6) x 2 (complement X vs. O) within-subjects design. Polynomial contrasts were used to test hypotheses.[4]

Alternation rate exerted the predicted unique effect: Randomness judgments linearly increased with increasing alternation rates, $F_{linear}(1, 45) = 10.39$, $p < .01$, $\eta_p^2 = .19$, and curved for alternation rates of .50, .58 and .66, $F_{quadratic}(1, 45) = 66.77$, $p < .01$, $\eta_p^2 = .60$. Again, as shown in Figure 3 (for a summary of statistics, see Table 3), the effect of alternation rate on randomness judgments was contingent on maximum run length: The strength of the linear effect of alternation rate decreased with increasing maximum run length, $F_{linear\text{-}linear}(1, 45) = 24.22$, $p < .001$, $\eta_p^2 = .35$. The quadratic-linear interaction between alternation rate and maximum run length was not significant, $F_{quadratic\text{-}linear}(1, 45) = 1.10$, $p > .30$. However, on a descriptive level, the pattern observed in Experiment 1 was replicated: The strength of the quadratic effect of alternation rate tended to decrease with increasing maximum run length. Maximum run length did not exert a unique effect, $F(1.65, 74.08) = 1.01$, $p > .36$. All other Fs < 1.61, ps $> .21$.

In sum, Experiment 2 replicated the results from Experiment 1 despite the fact that characteristics and length of sequences, instructions regarding generating device and task, and answering formats were changed. Alternation rate had a unique effect on randomness judgments. Importantly, this effect was contingent on maximum run length. Specifically, the longer the longest run in the sequence, the lower the effect of alternation rate on randomness judgments. Maximum run length did not exert a unique effect.

[4]Main effects and interactions with cubic and quartic effects of alternation rate were not expected and did not reach conventional levels of significance, Fs < 3.51, ps $> .07$.

Figure 3: Means (with standard errors) of participants' randomness judgments in Experiment 2, separately for a maximum run length of 4, 5, and 6, and an alternation rate of .42, .50, .58, .66, and .74.



4 General discussion

Binary sequences are characterized by various features. Two of these characteristics—alternation rate and run length—were repeatedly shown to influence randomness judgments. To date, the two characteristics have been investigated in different lines of research, usually without controlling for the other characteristic. The present contribution was intended to investigate both alternation rate and maximum run length within one design, so as to isolate their unique impacts as well as to understand

Table 3: Linear and quadratic effects of alternation rate with linearly increasing maximum run lengths of 4, 5, and 6 in Experiment 2.

Maximum run length	Linear effect of alternation rate	Quadratic effect of alternation rate
4	$F(1, 45) = 25.76, p < .001, \eta_p^2 = .36$	$F(1, 45) = 57.69, p < .001, \eta_p^2 = .56$
5	$F(1, 45) = 4.08, p < .05, \eta_p^2 = .08$	$F(1, 45) = 52.73, p < .001, \eta_p^2 = .54$
6	$F(1, 45) = 0.20$, *ns,* $\eta_p^2 = .00$	$F(1, 45) = 41.56, p < .001, \eta_p^2 = .48$

their joint impact on randomness judgments. In two experiments, alternation rate exerted both a linear and a quadratic effect on randomness judgments, independent of maximum run length. Isolating this unique effect extends prior research that has already documented an effect of alternation rate on randomness judgments (e.g., Ayton & Fischer, 2004; Falk & Konold, 1997).

The findings obtained go beyond previous research by demonstrating an interaction between alternation rate and maximum run length on randomness judgments: In two experiments, the impact of alternation rate on randomness judgments was contingent on maximum run length in that the impact on randomness judgments decreased linearly with increasing maximum run length. Specifically, when sequences contained the longest run possible, the difference in randomness judgments between the minimal and the maximal alternation rate was less pronounced than when sequences contained a shorter maximum run length. Furthermore, for sequences containing the longest run length possible, judgments of randomness curved less strongly than for sequences containing a shorter maximum run length. This finding supports the assumption that due to perceptual grouping, longer runs reduce the impact of alternation rate on randomness judgments.

Regarding the main effect for maximum run length on randomness judgments, the results obtained suggest that maximum run length exerts either no effect or only a weak unique effect. While Experiment 1 identified by tendency a unique main effect for maximum run length, no such main effect was found in the replication study or in Experiment 2. Note that all experiments implemented the longest maximum run length possible and therefore maximized the unique effect of maximum run length on randomness judgments.

In addition to our primary findings, several aspects of this evidence deserve short mention. First, the effects of alternation rate and maximum run length on judgments of randomness were comprehensively assessed by controlling for perceptual characteristics of sequences (Experiment 1 & 2: using complementary sequences; Replication Study: using complementary sequences and mirroring sequences in that the longest run was left versus right). This high level of control ensures reliable conclusions about the unique and joint impact of alternation rate and maximum run length on randomness judgments. Second, the effects of alternation rate and maximum run length were directly comparable as experiments implemented all maximum run lengths and the smallest and highest alternation rates possible for sequences characterized by equal relative frequencies. Third, the results were replicated despite the fact that length and characteristics of sequences, instructions regarding generating device and task, and answering formats had been changed.

Of course, the present research does not address all questions. It remains an open question how other characteristics of sequences additionally influence judgments of randomness. For instance, what is the unique impact of correlation, entropy, or relative frequency on randomness judgments? Do joint influences exist? Furthermore, it would be interesting to investigate which characteristics of binary sequences individuals attend to most strongly when forming judgments of randomness, in other words, to investigate whether a hierarchy of characteristics exists. Our finding that the impact of alternation rate has a reliable effect on judgments of randomness, while the unique impact of maximum run length is less pronounced or even unreliable, suggests that alternation rate may be higher in the hierarchy than maximum run length. Also pointing in the direction of a hierarchy, Diener and Tompson (1985) have suggested that individuals form randomness judgments by sequentially eliminating alternative nonrandom hypotheses. The authors suggest that participants judge a given sequence as random only after eliminating all alternative non-random hypotheses. For instance, participants may consider whether the relative frequency of elements is close to 50:50. In the case of severe deviations, participants may conclude that the given sequence is not random without necessarily considering other characteristics of random sequences. In the case of no deviations, participants are assumed to test for other characteristics of randomness, e.g., for alternation rates, and so forth.

Another interesting path of research would be to investigate whether the results obtained generalize to binary sequences reflecting human performance. As pointed out previously, individuals possess different expectations regarding outcome sequences of inanimate devices compared to outcome sequences reflecting human performance (Ayton & Fischer, 2004). While individuals expect sequences of inanimate devices to be characterized by negative recency, they expect sequences reflecting human performance to be characterized by positive recency. These different expectations may influence the focus on characteristics of binary sequences. One might argue that in the case of human generated sequences the focus might be more strongly on run length as reflecting, for instance, a run of good luck or bad luck. Future research should address whether the present findings—the unique impact of alternation rate and the contingency of alternation rate on maximum run length and the small or null-effect of maximum run length—hold when individuals are asked to judge sequences of outcomes reflecting human performance.

To summarize, this manuscript has disentangled the unique and joint impact of alternation rate and maximum run length on randomness judgments. Identifying their unique and joint effects provides a critical piece of evidence towards a more refined understanding of how judgments of randomness in binary sequences are formed.

References

Altmann, E. M., & Burns, B. D. (2005). Streak biases in decision making: Data and a memory model. *Cognitive Systems Research, 6*, 5–16.

Ayton, P., & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory & Cognition, 32*, 1369–1378.

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

Beck, D. M., & Palmer, S. E. (2002). Top-down influences on perceptual grouping. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 1071–1084.

Beck, J. (1967). Perceptual grouping produced by line figures. *Perception & Psychophysics, 2*, 491–495.

Burns, B. D., & Corpus, B. (2004). Randomness and inductions from streaks: 'Gambler's fallacy' versus 'hot hand'. *Psychonomic Bulletin & Review, 11*, 179–184.

Carlson, K. A., & Shu, S. B. (2007). The rule of three: How the third event signals the emergence of a streak. *Organizational Behavior and Human Decision Processes, 104*, 113–121.

Croson, R., & Sundali, J. (2005). The gambler's fallacy and the hot hand: Empirical data from casinos. *The Journal of Risk and Uncertainty, 30*, 195–209.

Diener, D., & Thompson, W. B. (1985). Recognizing randomness. *American Journal of Psychology, 98*, 433–447.

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

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

Feller, W. (1950). *An introduction to probability theory and its applications*. Oxford England: Wiley.

Ford, J. (1983). How random is a coin toss? *Physics Today, 36*, 40.

Green, D. R. (1982). Testing randomness. *Teaching Mathematics and its Applications, 1*, 95–100.

Houtkamp, R., & Roelfsema, P. R. (2010). Parallel and serial grouping of image elements in visual perception. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 1443–1459.

Huettel, S. A., Mack, P. B., & McCarthy, G. (2002). Perceiving patterns in random series: Dynamic processing of sequence in prefrontal cortex. *Nature Neuroscience, 5*, 485–490.

Konold, C., Pollatsek, A., Well, A., & Lohmeier, J. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education, 24*, 392–414.

Kubovy, M., & Gilden, D. (1991). Apparent randomness is not always the complement of apparent order. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure: Essays in honor of Wendell R. Garner.* (pp. 115–127). Washington, DC US: American Psychological Association.

Kunzendorf, R. G., & Pearson, B. (1984). Perception of randomness. *Perceptual and Motor Skills, 59*, 466.

Lepley, W. M. (1963). 'The maturity of the chances': A gambler's fallacy. *Journal of Psychology: Interdisciplinary and Applied, 56*, 69–72.

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

McDonald, F. (2009). *Understanding randomness via the perception and prediction of binary sequences.* Unpublished dissertation, University of New South Wales, Sydney

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

Nicks, D. C. (1959). Prediction of sequential two-choice decisions from event runs. *Journal of Experimental Psychology, 57*, 105–114.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259.

Olivola, C. Y., & Oppenheimer, D. M. (2008). Randomness in retrospect: Exploring the interactions between memory and randomness cognition. *Psychonomic Bulletin & Review, 15*, 991-996.

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.

Palmer, S. E., & Beck, D. M. (2007). The repetition discrimination task: An objective method for studying perceptual grouping. *Perception & Psychophysics, 69*, 68–78.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95*, 15–48.

Tyszka, T., Zielonka, P., Dacey, R., & Sawicki, P. (2008). Perception of randomness and predicting uncertain events. *Thinking & Reasoning, 14*, 83–110.

Vitz, P. C., & Todd, T. C. (1969). A coded element model of the perceptual processing of sequential stimuli. *Psychological Review, 76*, 433–449.

Wagenaar, W. A. (1970). Appreciation of conditional probabilities in binary sequences. *Acta Psychologica, 34*, 348–356.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt II. *Psychologische Forschung, 3*, 301–350.

Wiegersma, S. (1982). Can repetition avoidance in randomization be explained by randomness concepts? *Psychological Research, 44*, 189–198.

Wiegersma, S. (1987). The effects of visual conspicuousness and the concept of randomness on the recognition of randomness in sequences. *Journal of General Psychology, 114*, 157–165.