



Cognitive Science 45 (2021) e13072

© 2021 The Authors. *Cognitive Science* published by Wiley Periodicals LLC on behalf of Cognitive Science Society (CSS).

ISSN: 1551-6709 online

DOI: 10.1111/cogs.13072

The Effect of Context and Individual Differences in Human-Generated Randomness

Mikołaj Biesaga,^a  Szymon Talaga,^a  Andrzej Nowak^{a,b} 

^a*Robert Zajonc Institute for Social Studies, University of Warsaw*

^b*Department of Psychology, Florida Atlantic University*

Received 19 February 2021; received in revised form 2 November 2021; accepted 17 November 2021

Abstract

Many psychological studies have shown that human-generated sequences are hardly ever random in the strict mathematical sense. However, what remains an open question is the degree to which this (in)ability varies between people and is affected by contextual factors. Herein, we investigated this problem. In two studies, we used a modern, robust measure of randomness based on algorithmic information theory to assess human-generated series. In Study 1 ($N = 183$), in a factorial design with task description as a between-subjects variable, we tested the effects of context and mental fatigue on human-generated randomness. In Study 2 ($N = 266$), in online research, in experimental design, we further investigated the effect of mental fatigue on the randomness of human-generated series and the relationship between the need for cognition (NFC) and the ability to produce random-like series. Results of Study 1 show that the activation of the ability to produce random-like series depends on the relevance of the contextual cues ($\chi^2(2) = 7.9828$, $p = .0192$), whether they activate known representations of a random series generator and consequently help to avoid the production of trivial sequences. Our findings from both studies on the effect of mental fatigue (Study 1 – $t(47, 529.5568) = -18.62$, $p < .001$; Study 2 – $F(\text{edf} = 3.587, \text{Ref. df} = 3.587) = 11.863$, $p < .0001$) and cognitive motivation ($t(180) = 2.66$, $p = .009$) demonstrate that regardless of the context or task's novelty people quickly lose interest in the random series generation. Therefore, their performance decreases over time. However, people high in the NFC can maintain the cognitive motivation for a longer period and consequently on average generate more random series. In general, our results suggest that when

Correspondence should be sent to Mikołaj Biesaga, Robert Zajonc Institute for Social Studies, University of Warsaw, ul. Stawki 5/7, 00-183, Warsaw, Poland. E-mail: m.biesaga@uw.edu.pl

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

contextual cues and intrinsic constraints are in optimal interaction people can temporarily escape the structured and trivial patterns and produce more random-like sequences.

Keywords: Algorithmic complexity; Cognitive motivation; Individual differences; Random numbers generation; Random series production; Working memory capacity

1. Introduction

With the development of tools and methods for tracking and predicting our actions the question of whether people can behave truly unpredictably gains on popularity. Although machine learning algorithms that try to foresee our behavior are already fairly accurate, it is still an open problem whether people can escape these patterns. This is especially crucial for the sake of security and cryptography because people might be an independent source of entropy for encrypting connections (Halprin & Naor, 2009). Although it might seem to be an issue for computers scientists, in fact, it is a question of the origins of human behavior that have been studied since the early 1940s by experimental psychologists. Initially, researchers operationalized unpredictability in terms of the ability to produce responses at random, that is, ordered in a way that could not be predicted by experimenters (see Brugger, 1997; Nickerson, 2002; Oskarsson et al., 2009, for a review). The focus was on generative tasks asking people to produce random n -element sequences. In the experiments, participants were usually asked to reorder a given alphabet of symbols (e.g., Kahneman & Tversky, 1972), or simulate a random process, for example, tossing of a fair coin (e.g., Bar-Hillel & Wagenaar, 1991). The results of such tasks were usually compared against outcomes of the Bernoulli process (see Tune, 1964, for a review). However, this line of research did not bring the expected conclusions. Instead of proving that people can produce random-like series, it showed that human-generated responses are rarely random in the strict mathematical sense (see Brugger, 1997; Nickerson, 2002; Oskarsson et al., 2009, for a review). This inability to generate perfectly random data is currently considered a well-established fact. Moreover, research on the organized correlated nature of human behavior showed that human responses in a vast array of domains are structured in non-trivial and non-pattern ways (Kello et al., 2007). Nonetheless, what remains an open problem in the domain of research on the ability to produce random-like series is the degree to which this (in)ability varies between different people and can be affected by contextual factors. This approach is in line with the modern theories of cognitive control that argue that signals guiding our behavior are organized into two main categories: the ones related to past events and contextual cues. For example, the essence of the cascade model is that these signals are represented and processed in a context- or experience-dependent fashion. Therefore, both contextual and episodic control provide important information for selecting appropriate actions (Koechlin et al., 2003; Koechlin & Summerfield, 2007). In this paper, we investigate this issue and focus on between-subjects variability concerning the level of randomness of generated sequences under different task descriptions (providing different contextual cues). Moreover, we quantify the level of randomness using modern, objective

methods grounded in algorithmic information theory (AIT) that provide better estimates than standard methods based on probability theory.

The article is organized as follows. First, we review various methods to assess the randomness/complexity of human-generated series. We compare measures used in psychological studies with a modern approach grounded in the AIT. Second, we describe the current state of the art in the topic of human-generated randomness. We review the most important effects and formulate hypotheses regarding the effect of context (H1), the effect of fatigue (H2), and the effect of the Need for Cognition (H3), on the complexity of human-generated series. Then, in two studies, we test them using modern and objective methods for quantifying randomness grounded in AIT and based on the notion of algorithmic (Kolmogorov) complexity.

1.1. Measures of randomness

Although in general research indicates that people cannot produce perfectly random series, the methods used for assessing the randomness of human-generated series vary between different studies. Moreover, such studies often suffer from a lack of both a good definition of randomness and an objective and robust measure applicable to human-generated series, which are usually relatively short (Oskarsson et al., 2009). Some researchers tried to overcome this problem by instructing participants that a random sequence is a result of a random process, for example, tossing of a fair coin (e.g., Wagenaar, 1971). Such a comparison was meant to make the task easier. Early works on the subject revealed that people usually consider the term “random” vague, hence giving a more relatable process as an example was aimed to facilitate the task (see Brugger, 1997; Nickerson, 2002; Oskarsson et al., 2009, for a review). On the other hand, some researchers asked to maximize variability (Yavuz, 1963) or to attempt to use all available symbols equally often (Wiegersma, 1982). However, such instructions were criticized by Ayton et al. (1991). They noted that focusing on only one property of random series makes the instruction biased. As an example, they used a study by Wagenaar’s (1971, p. 80 in which “(...) it was checked carefully whether they [participants] understood that any systematic trend would make the sequence too predictable.”

Ayton et al. (1991) argued that it lead to a contradiction as participants were asked to produce random series, while a rule excluding some types of sequences, and hence reducing randomness, was hidden in the instruction. Moreover, they also criticized the assumption that people share the same concept of randomness. Nonetheless, this assumption, as well as randomness tests capturing only specific aspects of randomness, are rather common practices in the literature. Moreover, it seems that there is no general agreement on one definition and measure of randomness. Hence, it is not only hard to compare results from different studies but also decide whether they even measured the same ability. For example, Baddeley (1966) and Rath (1966) based their evaluation on frequencies of the elements in a sequence; Bakan (1960) and Ross (1955) on frequencies of alternations (or repetitions or runs); Baddeley (1966) and Chapanis (1953) on pair, triplet, or n-tuple frequencies; Chapanis (1953) on autocorrelations; Slak and Hirsch (1974), Warren and Morin (1965) on information transmission rates; and Wagenaar (1970) and Neuringer (1986) on RUNS test based on counting the

number of runs (sequences of identical responses), and assessing the probability of obtaining the observed number by chance (cf. Table 1 for a review of the most popular measures).

All the above-mentioned measures including entropy (e.g., Sharifian, 2016), symbol redundancy (e.g., Schuler et al., 2010), or hidden Markov chains methods (e.g., Budescu, 1987; Baena-Mirabete et al., 2020) may only serve to describe how a given series differs from a result of a specific kind of random process. For example, many information theory-based measures such as entropy focus on comparisons with the uniform distribution and are sensitive to problem representation, for example, whether one considers a distribution of symbols or n-tuples of symbols of different lengths. In other words, they measure the degree to which people tend to repeat the same subsequences of elements. Similarly, methods based on hidden Markov chains provide information about the mental model of randomness people preserve, not about a string itself.

Moreover, although traditional tools proved to be useful in detecting common biases, they failed to discriminate sequences produced by simple rules. For example, Copeland–Erdős sequence, arising from the concatenation of prime numbers (2, 3, 5, 7, 11, 13, 17, 19...), or Champernowne sequence, which is simply a concatenation of all positive integers (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, ...; or in base two, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, ...; Champernowne, 1933), they both pass any probabilistic test of randomness despite being produced by simple rules. Therefore, although they are not random at all, they would be considered random from the vantage point of most of the traditional methods based on probability and information theory.

Gauvrit et al. (2014) argued that the remedy for these definitional and methodological problems is AIT with the notion of Kolmogorov–Chaitin (algorithmic) complexity. It allows for the identification of random-like sequences produced by simple rules. This mathematical framework links randomness directly to algorithmic complexity (Li & Vitányi, 2008) and defines a random string to be a string that cannot be generated with any simple rule. In other words, the description length of the most succinct algorithm that generates a given string is not shorter than the string itself (Gauvrit et al., 2014). From the perspective of AIT, a random object is an object that cannot be efficiently compressed. This approach is consistent with the current psychological view on randomness perception. Falk and Konold (1997) suggest that the subjective perception of randomness relies on the perception of complexity. Nonetheless, for years, the applicability of Kolmogorov–Chaitin complexity in psychological research has been limited due to the lack of an algorithmic complexity index applicable to short strings produced in random series generation tasks. However, the recent development of new methods based on coding theorem (Soler-Toscano et al., 2014) solved this problem and therefore allowed overcoming the theoretical and computational problems present in studies on the ability to produce random-like series. Algorithmic complexity appears to be a more sensitive measure than entropy and has some other useful properties. First, unlike many classical measures, it allows comparing any two strings of different lengths. Second, it detects any deviation from randomness understood in the broadest possible sense as a lack of computable regularity. In contrary to most randomness measures, it does not focus on specific mathematical/statistical properties but is directly related to how likely it is that a given sequence would be produced by an algorithm sampled uniformly at random from the set of all possible

Table 1
Summary of popular methods to assess randomness

Name	Examples of use	Description
Frequency of the elements	Baddeley (1966); Rath (1966)	These measures focused on the fact that in a random sequence, the distribution of elements (or n-tuples) is uniform. Therefore, researchers tested how close the human-generated series' distribution was to a theoretical one. These types of measures did not account for the order of elements, for example, such regular sequences as 101010 and 111000 are both considered perfectly random if one considers only zero-order frequencies — frequency of elements. Therefore, assessment of randomness with measures based on the frequency of the elements either require computing multiple indexes for tuples of different length or selecting their length arbitrary.
Frequency of alternations	Bakan (1960); Ross (1955); Wagenaar (1970); Neuringer (1986)	Measures like the RUNS test examined the frequency of non-empty segments of a sequence that consisted of only one type of element (subsequences of uninterrupted identical responses). These types of measures assumed that the distribution of runs is normal. Therefore, the random sequence has an equal number of runs, for example, under this test 110011 is considered equally random as 100001 because both consist of three runs.
Autocorrelation	Chapanis (1953)	Measures based on autocorrelation assumed that the autocorrelation in a perfectly random series for all lags is equal to 0. Therefore, similarly to measures based on frequency of elements, they required either computing multiple indexes to examine autocorrelation with all lags or choosing the lag arbitrary. For example, for a sequence 1110001, the autocorrelation varies between lags. When one considers lag, 1 is $cor(s_t, s_{t-1}) = 0.31$; $cor(s_t, s_{t-2}) = -0.131$; $cor(s_t, s_{t-3}) = -0.671$; $cor(s_t, s_{t-4}) = -0.179$; $cor(s_t, s_{t-5}) = -0.036$; and $cor(s_t, s_{t-6}) = 0.107$.
Entropy	Sharifian (2016)	The average symbol information of the human-generated series s is defined as $H(s) = -\sum_i p_i \log_2(p_i)$, where p_i is the probability of an item i in the sequence. It takes a maximal value when probabilities of all items are equal, then entropy is $H_{\max} = \log_2(n)$. When one considers the probability of a single element in binary series, regular sequences like 101010 and 111000 yield maximal entropy. Therefore, similarly to other mentioned indexes measures based on entropy require either computing multiple entropies for different n-tuples length or choosing an arbitrary length scale.
Symbol Redundancy	Schulter et al. (2010)	Symbol Redundancy arises from entropy. It is defined as $SR(s) = 1 - \frac{H(s)}{H(s)_{\max}}$. Therefore, it suffers from similar constrains.
Hidden Markov Chains	Budescu (1987); Baena-Mirabet et al. (2020)	Measures based on hidden Markov chains assumed that the next element's probability depends on the previous element(s) of a sequence. In a simple case, the empirical probability of 0 after 1 might be different than after 1. Using this approach, it is possible to test for cognitive biases (e.g., tendency to avoid repetitions) in the random generation process. However, it does not allow for comparing randomness itself of the human-generated series.

algorithms. Third, it is based on a solid mathematical theory that provides a comprehensive definition of randomness. Altogether, algorithmic complexity seems to be an objective and universal randomness measure that is now easily applicable to experimental psychology (Gauvrit et al., 2016).

1.2. The role of experimental instruction

The first objective of this paper is to use algorithmic complexity to verify the long-standing assumption that less abstract and more relatable tasks facilitate random-like series production. Many researchers made this claim arguing that although the mechanism behind generating elements regardless of the task is the same, people often consider the notion of randomness vague (see Brugger, 1997; Nickerson, 2002; Oskarsson et al., 2009, for a review). Therefore, to make it more accessible in many studies, instructions additionally listed examples of well-known random processes, for example, tossing of a fair coin (Wiegersma, 1982). It was meant to illustrate an abstract and vague term with a relatable real-life example. Moreover, Neuringer (1986) argued that one of the reasons why people fail when producing random-like series is the fact that the task itself is uncommon. In day-to-day life, people rarely have to generate random series of abstract elements. Therefore, asking them to do so is destined to fail. However, even when a task description incorporates relatable examples, people still tend to fail to produce truly random series (e.g., Brugger, 1997; Nickerson, 2002; Oskarsson et al., 2009). Therefore, it can be concluded that regardless of the context people do not perform well when asked to produce random series. But such results do not necessarily falsify Neuringer's (1986) hypothesis. That is because in the majority of classical studies on the production of random-like series, researchers used different measures to assess the randomness of generated strings/sequences and often used methods capturing only one or few properties expected from random strings (cf. Table 1). Therefore, it is very difficult to decide whether a task description affects the level of randomness of human-generated series. Here, we approach this problem with a robust and universal measure of algorithmic complexity that does not suffer from any of the above-mentioned drawbacks and examine whether there are differences in the level of randomness in human-generated series depending on the cognitive accessibility of a task description. Under hypothesis one (H1), we aim to test whether task descriptions using easy-to-imagine processes such as a fair coin toss or price changes in a stock market facilitate the production of more random-like series than a neutral instruction with no additional hints in which subjects are simply asked to write down random sequences of zeros and ones.

1.3. The effect of fatigue

Using algorithmic complexity as a measure of randomness of human-generated series allows looking at the ability behind the series production from the perspective of between-subjects variability or individual differences. Recently, in an online study on over 3,400 participants, Gauvrit et al. (2017) showed that the developmental curve of the ability to produce random-like series is similar to most cognitive abilities. It peaks at age of 25 and decreases slowly afterward and more quickly after the age of 60. Therefore, the question should not be

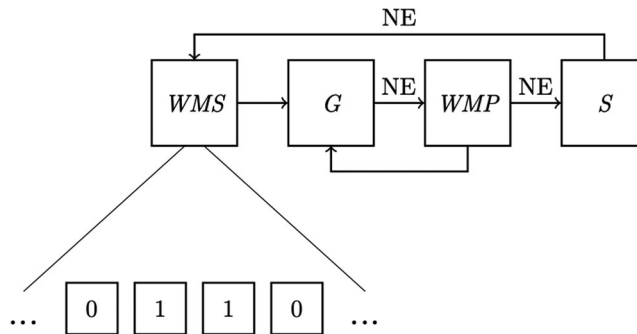


Fig. 1. Schematic model of random series generation. WMS—working memory storage component; G—random series generator; WMP—working memory processing component; S—output sequence; NE—a new element of the series. During the random series generation process, people maintain only the last few elements that have been produced, active in their working memory. When a new element is generated, the subsequence is compared to the random sequence’s prototype and based on the result, a next element is either selected or discarded (then another one is generated). The new element is appended to the sequence in the working memory, while the first one is removed. Therefore, the storage component of the working memory defines the length of the sliding window in which the comparisons are being made.

what makes people unable to produce random-like series of elements or what are the cognitive constraints that prevent people from doing so, but rather what drives the individual differences concerning this ability.

Therefore, to further examine complex cognitive determinants of the ability to produce random-like series, we investigated the working memory capacity, which is often said to be between five and nine chunks of information. This limitation refers to a limited capacity system responsible for maintenance in a highly accessible state, manipulation, and retrieval of task-relevant information that is needed for ongoing cognition (Baddeley, 1986). It is used for temporary storage and carrying out computational processes on mental representations necessary for successful task performance. Kareev (1995) argued that producing a random series is not a static process and people’s poor performance may be due to structural limitations of their working memory. During random series generation, people do not consider subsequent elements as independent but store previous ones in the working memory. Studies in which researchers used the hidden Markov chains approach to investigate the randomness of human-generated series showed that each new element is determined based on the previous values (Budescu, 1987; Baena-Mirabete et al., 2020). Therefore, it is argued that sequence elements are not considered in isolation but rather jointly within a sliding window of width equal to the working memory capacity (see Fig. 1 for the description of the mechanism behind the random-like series production).

Moreover, Warren et al. (2018) reported that human-generated series, when the capacity of working memory is taken into account, resembled sequences produced by a random process. In their study, they compared frequencies of elements in a sliding window of length four in both human-generated series and series produced by the Bernoulli process. The results showed that there was not much difference in terms of frequencies of elements

(Barbasz et al., 2008; Hahn & Warren, 2009). However, in their analysis, they only focused on the frequencies in the whole series not investigating how this resemblance changed over time. Here, we focus on the dynamical aspect of random-like series production and analyze how it changes over the time of series generation. The mechanism behind random-like series production requires constant attention and evaluation of whether the subsequence kept in the working memory meets the criteria of random series (cf. Fig. 1). Therefore, it employs higher level cognitive functions.

Research showed that prolonged execution of cognitively demanding tasks might induce mental fatigue and consequently will lead to the decrease of the task performance (Lorist et al., 2005). Moreover, tasks that require vigilance but do not convey much information or are simply boring also have a similar effect on the performance (Lorist & Faber, 2011). That is because mental fatigue depends not only on the availability of cognitive resources but also on cognitive motivation. Hockey (2011) argues that the mechanism behind the performance decrease is effort-controlled and is only mildly affected by the energy depletion over time. Repetitive tasks do not become more difficult but rather they lose attractiveness. Therefore, to maintain high performance over time, they require more effort and consequently a high level of cognitive motivation. However, before testing the relationship of the ability to produce random-like series with the cognitive motivation, first, under hypothesis two (H2), we expect a decrease of the task performance over time. Similarly to Warren et al. (2018), we take into account the capacity of working memory and examine the randomness of series in a rolling window of a length corresponding to the working memory capacity. However, instead of frequencies of elements, we use algorithmic complexity as a measure of randomness.

1.4. Cognitive motivation

As previously mentioned, according to a motivational control theory of cognitive fatigue, the decrease of the performance in cognitively demanding tasks depends predominately on cognitive motivation (Hockey, 2011). That is because having the potential to produce more random-like series than others does not mean that it is fulfilled. Although it may be argued that producing random-like series is a relatively easy task, it requires engaging cognitive resources (Barbasz et al., 2008). Recognizing patterns and inhibiting simple repetitions require constant attention. Hence, producing a random-like series does not only depend on what people are intellectually able to do but also how they tend to invest their cognitive resources. In other words, whether they are inclined to engage in challenging tasks and whether they can put enough effort to maintain high performance. Therefore, the Need for Cognition (NFC) — a stable tendency to engage in intellectually effortful tasks and enjoy them — should be correlated with the ability to produce random-like series (Matusz et al., 2011). Much of the research on this construct investigated the individual differences in processing information in situations involving persuasion and argumentation (cf. Petty et al., 2009, for a review). In general, people high on the NFC scale tend to think more intensively on a variety of things and consider a wider range of possibilities than people with lower levels of NFC (Levin et al., 2000). Therefore, their judgments and decision-making process tend to be affected by larger and more diverse arrays of thoughts or ideas. Consequently, Cacioppo and Petty

(1982) argued that people with high levels of NFC use different strategies for processing information than people with lower levels. People with high NFC naturally use the central path during information processing and search. Therefore, their judgment is based on a more thoughtful analysis of available information, while people with low NFC base their decisions more directly on the input (whether emotion, image, or intuition).

Regarding the relationship between the NFC and cognitive abilities, Dornic et al. (1991) showed that the tolerance for mental effort is positively correlated with the performance on challenging tasks. Therefore, people who rated themselves as needing mental stimulation were more efficient in stimulating tasks. Similarly, Dai and Wang (2007) reported that individuals with high NFC performed better in such cognitive exercises as text comprehension. Finally, Fleischhauer et al. (2010) showed that NFC is positively associated with intelligence beyond verbal intelligence or a specific domain of reasoning. It is moderately related to fluid intelligence in general. That is because, on the one hand, it is easier to enjoy intellectually challenging tasks when there is a track record of successfully resolved ones. On the other hand, people with a high NFC are more motivated to acquire complex cognitive skills and consequently perform better in tasks requiring them (Day et al., 2007). Therefore, under hypothesis three (H3), we expect that the NFC is positively associated with the ability to produce random-like series. That is because for people with high NFC, cognitive motivation to maintain high performance in effortful tasks lasts longer.

1.5. Hypotheses

Based on the literature review, in this paper, we investigated three hypotheses regarding the effect of contextual cues on the random-like series generation and the relationship of mental fatigue and cognitive motivation with performance in such tasks:

Hypothesis 1 (H1). *Random generation tasks that refer to processes known from day-to-day life allow for producing more random (algorithmically complex) series than under more abstract instructions.*

Hypothesis 2 (H2). *Over time the performance (algorithmic complexity of the produced series) in random generation tasks decreases.*

Hypothesis 3 (H3). *Randomness (algorithmic complexity) of human-generated series is positively associated with the tendency to engage in cognitively effortful tasks and enjoy them (NFC).*

2. Method

Our goal was to test the above-mentioned hypotheses in two studies. In Study 1, we tested hypotheses regarding the effect of task description (H1) on the algorithmic complexity of human-generated series. Additionally, in all experimental conditions, we investigated the dynamics of the randomness of human-generated series to test the potential

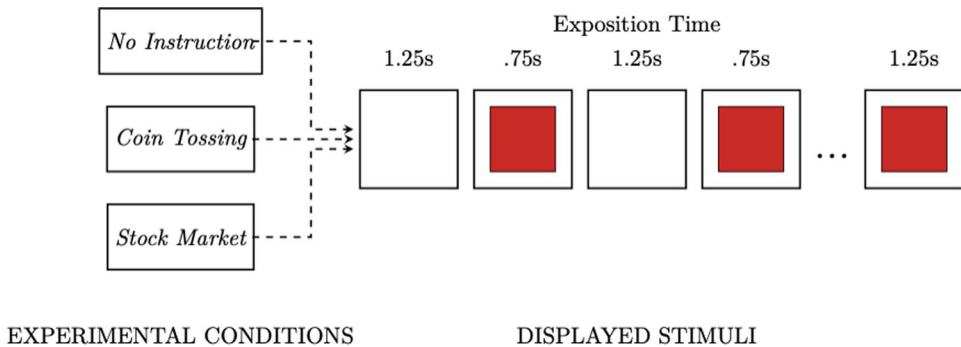


Fig. 2. Flow chart of Study 1 design. Both groups of participants were assigned at random to one of the three experimental conditions: (a) No instruction (NI); (b) coin tossing instruction (CSI); and (c) stock market instruction (SMI). In all three conditions, a custom software displayed every 1.25 s a red square for 0.75 s.

effect of mental fatigue (H2). In Study 2, we further examined the dynamics of randomness in human-generated series (H2) and the relationship between algorithmic complexity of human-generated series with the NFC (H3).

2.1. Study 1

In the first study, participants were tested individually in sessions that lasted about 15 min. They were simply asked to produce a binary series of 300 elements. We gathered the responses from the participants using a custom software¹ that was written in processing (Reas & Fry, 2006). The procedure was run on MacBook Pro 13,3-inch Early 2015 with OS X 10.11. It collected signals with 16 ms delay. The stimuli were displayed on a built-in monitor with 2,560 × 1,600 pixel resolution and 60 Hz refresh rate.

2.1.1. Procedure and design

The experiment followed a factorial design with one between-subjects variable — the task description. Participants were two groups of students recruited either from the psychology or chemistry faculties at the University of Warsaw. Participants from both groups were assigned randomly to one of the three experimental conditions: (a) no instruction (NI); (b) coin tossing instruction (CTI); and (c) stock market instruction (SMI). In all three conditions, we used custom software to run the experiment. Every 1.25 s it displayed a red square for 0.75 s (cf. Fig. 2). Furthermore, in the NI condition, participants were simply asked to select randomly and press one of the two predefined keys every time they saw the red square. In the CTI condition, participants were instructed to imagine a toss of a fair coin whenever they saw the red square and press either key marked as tails or heads. In the SMI condition, participants were asked to imagine a stock market chart and to assume that price fluctuations were random. They were instructed to try to predict whether the price in the next time step will go up or down and to press either key marked “up” or “down.”

2.1.2. Participants

The participants were students from the Faculty of Psychology and Faculty of Chemistry at the University of Warsaw. A total of 183 subjects (129 females), aged from 18 to 30 ($M = 21.54$, $SD = 2.12$), were randomly assigned to one of the experimental conditions. The procedure was approved by the ethics committee of Robert Zajonc Institute for Social Studies at the University of Warsaw. All participants gave informed consent before taking part in the study.

2.1.3. Data preprocessing

Although the participants were instructed to only press relevant keys when they saw the red square, some pressed it more and some less than 300 times. Therefore, the length of the series varied between subjects from 218 to 1016 elements ($Median = 300$). However, there were only 13 people who produced significantly longer series than others. Therefore, in further analysis, we removed observations exceeding the typical length of the series, that is, the last 10% of observations (those with more than 313 elements). This strategy did not affect the results significantly because people who did not follow the task scrupulously were spread almost uniformly between experimental conditions. After the data preprocessing, we had 56 participants in the NI condition; 54 in the SMI condition, and 55 in the CTI condition. The analysis of the whole data set can be found in the Supporting Information.

We used “pybdm”² library to estimate algorithmic complexity of series. All other analyses were performed using R language (R Core Team, 2020) with “nlme” and “lme4” packages for estimating linear mixed models (LMMs) (Bates et al., 2015; Pinheiro et al., 2020) and “quantreg” package for fitting quantile regression (Koenker, 2020).

For each participant, we computed an overall algorithmic complexity of the entire sequence as well as vectors of complexity estimates in rolling windows of different lengths. Both were normalized (using the method described in Zenil et al., 2018) because, even though we removed observations with too long series (longer than 313 elements), the lengths of the remaining sequences still varied. Normalized algorithmic complexity ranges from 0 to 1 where 0 stands for the simplest possible object (a constant series filled with a single symbol) and 1 for the most complex object of a given size. The rolling algorithmic complexity was calculated for window lengths from 5 to 9. This corresponds to the estimated capacity of the working memory that is usually said to be 7 ± 2 elements (Baddeley, 1986). Herein, we present the results only for the window of length 8. The rest can be found in the Supporting Information.

2.1.4. Results

Before testing our main hypotheses, we investigated whether there were differences between the two groups of participants: students from the chemistry and psychology faculties. We wanted to explore whether their knowledge and differences in the curriculum may have affected the outcomes. We used a non-parametric test to assess whether the distributions of algorithmic complexity were systematically different between groups of observations. The Wilcoxon rank sum test revealed that the main difference was not significant, $W = 3,422.5$, $p = .9506$. The further analysis within experimental conditions also did not

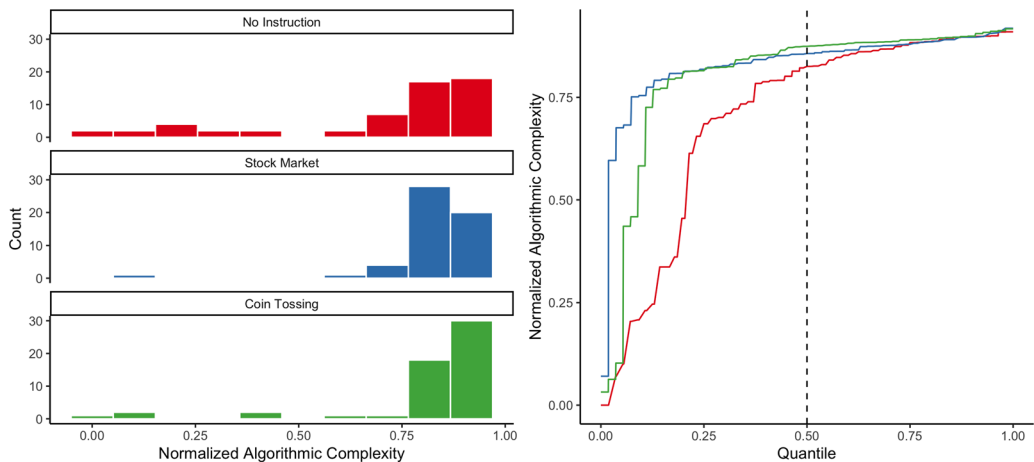


Fig. 3. Left panel: Distributions of normalized algorithmic complexity in the experimental groups. Right panel: The distributions of the normalized algorithmic complexity of series across quantiles. In the first quartile, the distribution of normalized algorithmic complexity of series in NI condition was different than in SMI (95% CI [0.0975 0.4677]) and CTI conditions (95% CI [0.086 0.4631]). Similarly, in the case of the second quartile, significant differences were observed between NI and SMI (95% CI [0.0047 0.071]), and NI and CTI (95% CI [0.0091 0.0875]). There were no significant differences between the distributions in the groups in the case of the third quartile.

yield significant differences in distributions of algorithmic complexity between faculties (NI, $W = 431.55$, $p = .5172$; SMI, $W = 395$, 0.5975 ; CTI, $W = 316$, $p = .3005$). Therefore, there was no reason to further explore these differences in later analyses.

To test the hypothesis regarding the effect of the task description (H1) on the overall algorithmic complexity of human-generated series, we also used a non-parametric test to assess whether the distributions of algorithmic complexity were systematically different between groups of observations. The Kruskal–Wallis test revealed a significant difference between distributions of algorithmic complexity in task description conditions, $\chi^2(2) = 7.9828$, $p = .0192$. The closer visual examination of the histograms revealed that the main difference between distributions was in their left tails (see the left panel of Fig. 3). In the NI condition, the frequency of the lowest results is relatively higher than in the CTI and SMI conditions. Therefore, to better understand this effect, we used quantile regression (Koenker & Hallock, 2001) to estimate the full quantile process in the three experimental groups as well as calculate 95% confidence intervals for the first, second, and third quartiles. As Fig. 3 shows, there is a clear gap between low quantiles (left tail) in the no instruction (NI) condition compared to other conditions. Furthermore, confidence intervals showed that there are significant differences in terms of first quartiles between the NI condition and both SMI, 95% CI [0.097 0.468], and CTI conditions, 95% CI [0.086 0.463]. Similarly, in the case of the second quartile, significant differences were observed between NI and SMI 95% CI [0.005 0.071], and NI and CTI 95% CI [0.009 0.087]. We did not observe significant differences in the case of the third quartile. These results confirm that the gap visible in Fig. 3 is statistically significant. This

Table 2
Estimated parameters of the rolling window model

Predictors	Estimates	Std. Error	df	<i>t</i> -Value	<i>p</i> -Value
(Intercept)	0.5482885832	0.03278906860	66.39290	16.7216883	.0000000
Time step	−0.0002096913	0.00001126191	47,529.55680	−18.6195100	.0000000
Condition (Stock Market)	0.1156103370	0.03382503485	74.99477	3.4178926	.001023
Condition (Coin Tossing)	0.1062608990	0.03914455582	102.85681	2.7145767	.0077835
Faculty (Psychology)	−0.0076733215	0.02208870126	105.83864	−.3473867	.7289907

Marginal $R^2 = 4.01\%$, Conditional $R^2 = 45.21\%$.

Degrees of freedom were adjusted with the Kenward–Roger method.

indicates that both the Stock Market and Coin Tossing conditions prevented participants from producing series of low algorithmic complexity. The differences in the task description conditions seem not to affect the upper parts of the distributions (roughly above median). Thus, the results suggest that the task description does not enhance the ability to generate random-like sequences, as many researchers believed (e.g., Ayton et al., 1991), but rather helps to activate it whatsoever, resulting in significantly less trivially non-random sequences being produced.

In a more detailed analysis, we used the rolling algorithmic complexity. We estimated a linear mixed-effects model to test the second hypothesis that algorithmic complexity decreases over time (i.e., the effect of mental fatigue). We present herein the analysis for a rolling window of length 8 only as this model attained the best goodness-of-fit. The specification of the rest of the models is in the Supporting Information together with a reproducible R script. The dependent variable in the models was (normalized) algorithmic complexity. All models included, as fixed effects, the following factors: time step and task description type (with NI condition used as the reference level). Goodness-of-fit of the models was assessed with marginal R^2 (variance retained by fixed effects only) and conditional R^2 (variance retained by a model as such) as proposed by Nakagawa et al. (2017).

In the rolling window model, random intercepts and slopes for task description conditions were included (clustered by subjects). The model indicated several significant effects. There was a significant negative effect of the time step on the algorithmic complexity, $t(47, 529.5568) = -18.62$, $p < .001$. With each time step, the algorithmic complexity decreased by 0.0002 ± 0.00001 . This result supports the second hypothesis regarding the observed decline in the algorithmic complexity over time. Additionally, there were significant differences between task description conditions (see Table 2). In SMI condition, algorithmic complexity was higher, 0.1156 ± 0.0338 , than in NI condition, $t(74.99) = 3.418$, $d = 0.516$, $p = .001$. Similarly, the algorithmic complexity in CTI condition was 0.1062 ± 0.0391 higher than in NI condition, $t(102.68) = 2.714$, $d = 0.479$, $p = .008$. Although most of the fixed effects were statistically significant, they explained only 4.01% of the variation while after including the random effects the fraction was 45.21%. The closer examination revealed that the model with subject-specific random intercepts and slopes (for task description conditions) fitted the data better than the model with random intercepts only, $\chi^2(1) = 33.075$, $p < .001$. Therefore, the between-subjects variance was the function of the task description conditions. The profiled confidence intervals of standard deviation in all three experimental conditions (cf. Fig. 4) show that in NI condition, between-subjects

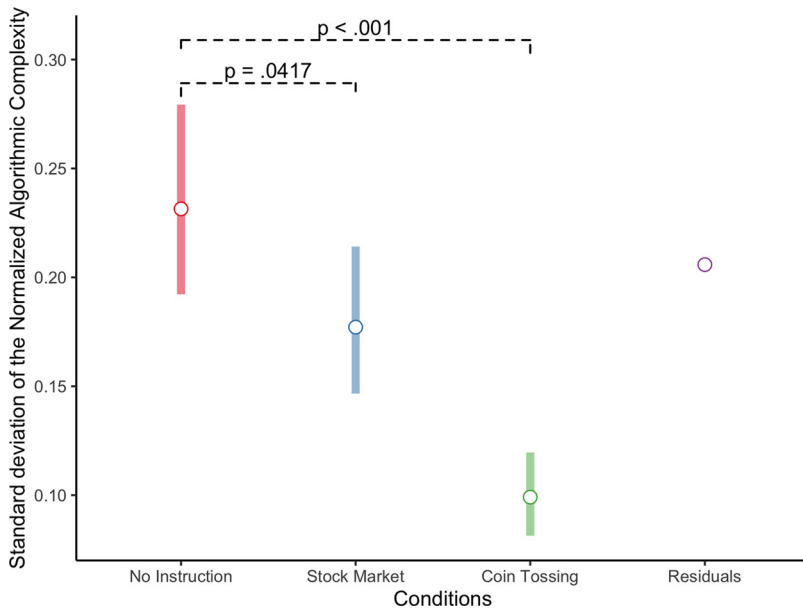


Fig. 4. Estimated standard deviations (with 95% CI based on likelihood profiling) of the distributions of random intercepts in the rolling window model (with length 8) in three experimental conditions. The standard deviation of the residual distribution serves as a reference point. In NI condition, between-subjects variance of the algorithmic complexity was higher than in SMI ($p = .012$) and CTI conditions ($p = .001$).

variance was significantly higher than in CTI (an analysis of deviance of the model that included independent terms for the variance in all three task description conditions against the model that included an independent term for variance CTI condition was significant, $\chi^2(1) = 35.537$, $p < .001$) and SMI conditions (an analysis of deviance of the model that included independent terms for the variance in all three task description conditions against the model that included an independent term for variance in the stock market condition was significant $\chi^2(1) = 4.1442$, $p = .0418$). These outcomes are in line with the quantile regression results for the overall algorithmic complexity. The reason why people perform better in CTI and SMI conditions than in NI condition is the fact that in the former they can usually avoid trivial series with near-zero (normalized) complexity. Hence, on average, the difference between average complexity produced by two different people in these conditions is lower than in NI condition.

Altogether, the results of Study 1 support H1 and H2. The task description affected the randomness of the human-generated series. Subjects produced more random sequences when the instructions included examples of random processes (tossing a fair coin or random fluctuations in a stock market chart) than when there was NI and participants were simply asked to produce random series. The difference was mainly because in these conditions, unlike the NI condition, subjects were usually able to avoid producing trivial series of low randomness (e.g., composed of almost only of 1's or 0's); hence, the between-subjects variance of algorithmic complexity was smaller in these groups. However, we argue that our results, based

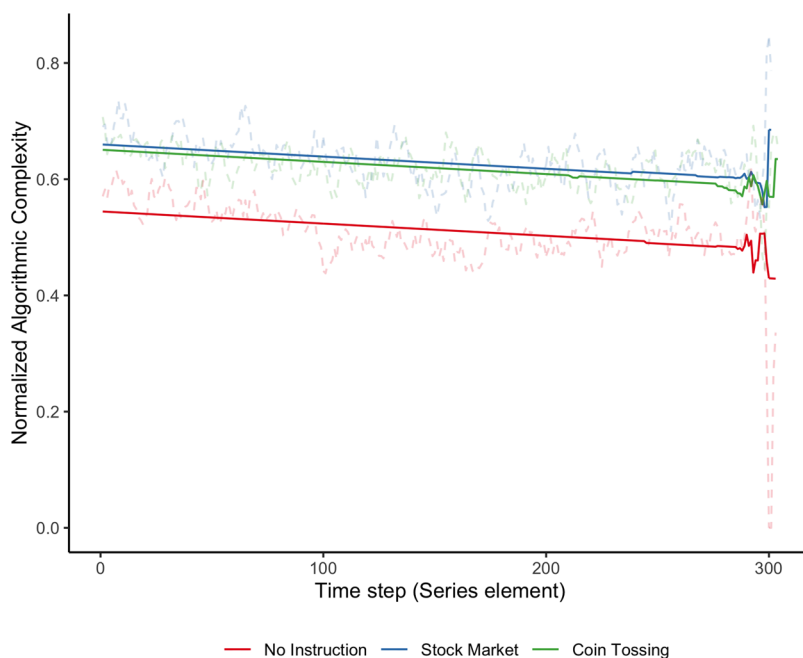


Fig. 5. Solid lines present trend curves for normalized algorithmic complexity and dashed lines depict the average algorithmic complexity as a function of experimental conditions. For each participant, we computed vectors of complexity estimates in a rolling window of length 8. Although the experimental task asked for the creation of 300-long series, the length still varied. Therefore, the uncertainty of both the trend curve and the average algorithmic complexity increased around the 292nd element. Regardless of the condition, with each time step, the algorithmic complexity decreased by 0.0002 ± 0.00001 , $t(47, 529.5568) = -18.62$, $p < .01$. However, in SMI condition, algorithmic complexity was higher, 0.1156 ± 0.0338 , that in NI condition, $t(74.99) = 3.418$, $d = 0.516$, $p = .001$. Similarly, the algorithmic complexity in CTI condition was 0.1062 ± 0.0391 higher than in NI condition, $t(102.68) = 2.714$, $d = 0.479$, $p = .008$.

on a detailed analysis of quantiles of the distributions in all experimental groups, provide more insight into the effects of task instruction in random sequence generation and show that important effects occur not at the level of averages/central tendencies but rather in lower tails of complexity distributions, meaning that the contextual cues have inherently nonlinear effects. Additionally, the results of Study 1 support the second hypothesis that the randomness of the series decreases over time (compare Fig. 5).

2.2. Study 2

The second study aimed to further investigate the observed effect of mental fatigue on the random series production (H2) and to examine the relationship of algorithmic complexity with the need for engaging cognitive resources in challenging tasks (H3). The results of Study 1 showed that the algorithmic complexity of human-generated series decreased over time (see Fig. 5). We argue that this effect was caused by fatigue due to the task requiring constant attention and putting more effort to maintain the high performance over time. Participants in an

almost 15 min-long procedure were asked to produce 300 elements series without any break. However, our main interest in Study 1 was to investigate the effect of context. Therefore, we used linear mixed models (LMMs) approach. It allowed for detailed examination of differences in variance of algorithmic complexity between experimental conditions and estimation of the trend curve at the same time. In Study 2, we wanted to focus more on the shape of the trend curve because the LMM approach did not account for the fact that normalized algorithmic complexity is bounded within $[0, 1]$. We investigated whether the effect of fatigue would also be present between series of shorter independent tasks since many researchers argued that the reason behind people's inability to produce random-like series is the fact that subjects do not perceive each element as independent (e.g., Kareev, 1995). Moreover, Hockey (2011) showed that the mental fatigue-related decrease of the performance in cognitively demanding tasks might be inhibited by novelty. That is because the new task does not require additional motivation that maintains its attractiveness and consequently requires putting less effort. Therefore, in this study, participants were asked to produce 10 independent binary series and to complete the Polish version of the NFC scale (Matusz et al., 2011; Cacioppo and Petty, 1982). To facilitate the notion of independence between produced series and to allow for novelty between tasks, we introduced two experimental conditions. In the homogeneous instruction condition, participants produced 10 independent binary series. Before each series, they were instructed using the CTI from Study 1. In the heterogeneous instruction condition, the CTI and the SMI from Study 1 were alternated every second series. Therefore, in this condition, each participant produced five series under CTI and five under SMI. We expected that participants assigned to the heterogeneous condition would be able to maintain a high level of algorithmic complexity between series for longer periods of time than participants in the homogeneous condition. Participants were recruited through a Polish nationwide opinion poll Ariadna. It is an online poll often used to conduct political surveys or scientific research. Depending on the declared length of a study, users are gratified with points which they can exchange for prizes.

2.3. Procedure and design

We used an experimental design including one two-level between-subjects variable. First, participants were assigned at random to either a homogeneous instruction condition or a heterogeneous instruction condition. In both conditions, similarly to Study 1, they were presented with a red square every 1.25 s for 0.75 s in ten 12 displays-long series (compare Fig. 6). Thus, each subject was supposed to generate 10 sequences of 12 elements. In the homogeneous condition, in all series, participants followed the coin tossing condition from Study 1. In the heterogeneous condition, their task altered every second series. In the odd series, they were asked to follow the coin tossing condition from Study 1. In the even series, their task was identical as in the stock market condition in Study 1. Regardless of the condition, there was 5 s break between each series during which a blank slide with information on how many series were left was presented. After completing the procedure of generating random series, participants in both conditions filled in the Polish version of the NFC scale (Matusz et al., 2011; Cacioppo and Petty, 1982).

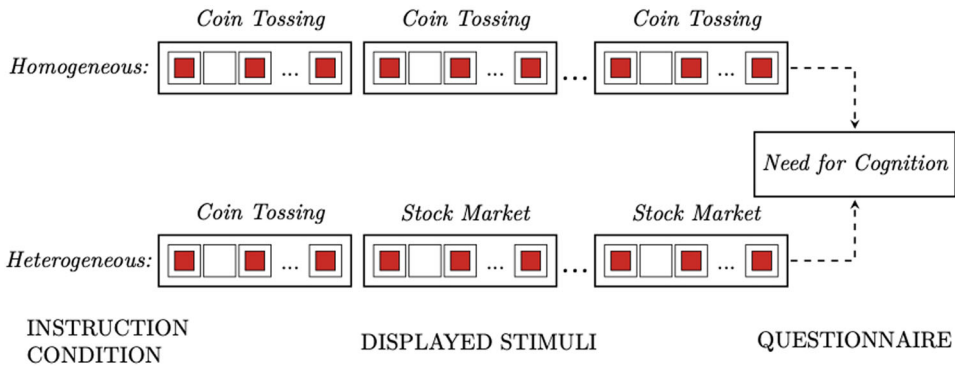


Fig. 6. Flow chart of Study 2 design. Participants were assigned at random to either homogeneous or heterogeneous instruction conditions. In both conditions, they were presented with a red square every 1.25 s for 0.75 s in ten 12 displays-long series. In the Homogeneous condition, in all series, participants were asked to follow the same instruction as in the Study 1 Coin Tossing Condition. In the Heterogeneous condition, their task altered every second series. In the odd series, they were asked to follow the instruction from the coin tossing condition from Study 1. In the even series, participants followed the identical instruction as in the Stock Market Condition in Study 1. Regardless of the condition, there was 5 s break between each series during which they were presented with a blank slide with information on how many series remained. Afterward, in both conditions, participants were asked to complete the Polish version of the NFC scale.

2.3.1. Participants

A total number of 266 subjects agreed to take part in the study. However, due to the unrealistic (too short or too long) time of completion and unfinished surveys, we excluded 80 participants. Therefore, finally, a sample of 186 participants (134 females), aged from 18 to 77 ($M = 39.32$, $SD = 13.08$), was used. They were assigned at random to one of the experimental conditions. The procedure was approved by the ethics committee of Robert Zajonc Institute for Social Studies at the University of Warsaw. All participants gave informed consent before taking part in the study.

2.3.2. Data preprocessing

In online research, it is crucial to measure survey time and exclude participants who complete tasks in a suspiciously short or long amount of time. Therefore, we removed observations with an unrealistic length of completion time of the whole study (both the random series generation task and the NFC scale), that is, 15% of observations with the shortest response times (below 11 min) and 15% with the longest response times (above 28 min). Although the participants were instructed to only press relevant keys when they see a red square some people pressed it less frequently than 120 times (10 sequences of 12 elements). Therefore, the total number of produced elements in all series varied between subjects from 100 to 120 elements ($Median = 114$).

As in Study 1, we used “pybdm” library for Python to estimate the algorithmic complexity of generated series.³ The complexity of each sequence was estimated independently, so each

participant was represented by 10 data points ordered as a time series. The algorithmic complexity of each series was normalized because although the instruction asked to produce 12 elements binary series the lengths varied.

We measured the need for engaging cognitive resources in challenging tasks with the Polish version of the NFC scale developed by Matusz et al. (2011). It contained 36 statements. Participants were asked to make a judgment on a five-point Likert scale on how much they agree with the statement (with 1 — I strongly disagree and 5 — I strongly agree). The total score was calculated as the sum of all answers. The psychometric quality of the Polish version of the NFC (internal consistency, temporal stability, factor analysis, criterion, and experimental validity) was confirmed in several studies (Matusz et al., 2011). Similarly, to the original scale by Cacioppo and Petty (1982), the Polish adaptation has one major factor and an average reliability coefficient $\alpha = 0.895$.

2.3.3. Results

Unlike Study 1, in Study 2, we used a GAMM that can capture more complex nonlinear trends than LMMs. Additionally, this family of models allowed for testing whether changing task descriptions between series production (heterogeneous condition) would diminish the fatigue effect compared to the production of independent sequences of binary series (homogeneous condition). The dependent variable was normalized algorithmic complexity of a currently generated series (10 data points per subject). The mean difference between experimental conditions was represented with a single fixed parametric effect with the homogeneous condition used as the reference group. The second fixed parametric effect was a linear regression term corresponding to the effect of the NFC. For non-parametric effects, we had a non-linear difference in trends of algorithmic complexity over time between the homogenous and heterogeneous conditions (with the homogeneous condition being the reference level in dummy coding) and a non-linear trend of algorithmic complexity over time. Additionally, we used subject-level random intercepts and slopes for the time trend to model systematic between-subjects differences. The goodness-of-fit of the model was assessed with marginal R^2 (variance retained by fixed effects only) and conditional R^2 (variance retained by the model as such) as proposed by Nakagawa et al. (2017).

The fitted model explained 50.7% of the variance with fixed effects reproducing 10.67%. The non-linear trend of algorithmic complexity over time was significant (cf. Fig. 7), $F(edf = 3.587, Ref.df = 3.587) = 11.863, p < .0001$. The negative effect of time was approximately linear for the first five series and then the curve flattened and remained at approximately the same level for the rest of the series. However, neither linear ($t(180) = -1.69, p = .091$) nor non-linear ($F(edf = 1, Ref.df = 1) = 1.956, p = .162$) differences between the homogeneous and heterogeneous conditions were significant (See Table 3). The alternation of task description between series did not inhibit the effect of fatigue. Additionally, there was a small significant parametric effect of the NFC, ($t(180) = 2.66, p = .009$). With an average of 0.01 point increase on the NFC scale, there was 0.2 ± 0.093 point change of the algorithmic complexity of the series. This positive relationship between the randomness of series and the NFC scale provided support for H3. People with a tendency to engage in intellectually challenging tasks performed better in random-series generation tasks.

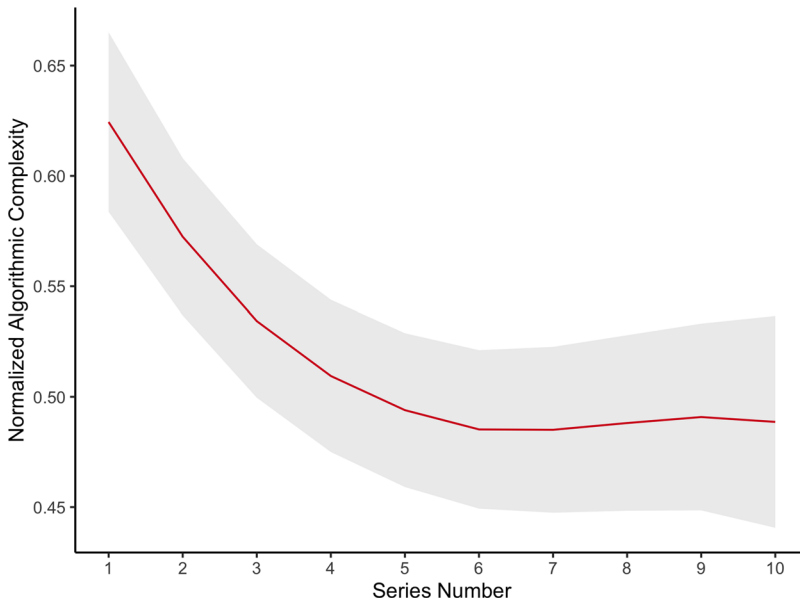


Fig. 7. The trend curve for normalized algorithmic complexity for subsequent generated series with 95% confidence interval. For each series produced by participants, we estimated its algorithmic complexity. Therefore, each participant was represented by 10 data points ordered in time sequences. The algorithmic complexity of each series was normalized because, even though the instruction asked to produce 12-elements binary sequences, the lengths varied.

Table 3
The generalized additive mixed model specification

Parametric coefficients	Estimates	Std. Error	t-Value	df	p-Value
(Intercept)	0.3299843	0.0860422	3.835	1,611	.00013
Need for Cognition	0.0024632	0.0009256	2.661	180	.00786
Linear difference between trend curves (Homogeneous)	-0.0398899	0.0235888	-1.691	180	.091
Smooth terms	edf	Ref.df.	F	p-Value	
Non-parametric trend curve	3.587	3.587	11.863	<.0001	
Non-parametric difference between trend curves (Homogeneous)	1.000	1.000	1.956	0.162	

Marginal $R^2 = 10.67\%$, Conditional $R^2 = 50.7\%$.

3. Discussion

Based on the literature review, in this paper, we assumed that people are not able to produce perfectly random series (cf. Oskarsson et al., 2009; Brugger, 1997; Nickerson, 2002).

Therefore, unlike most research on the topic, we focused on the effect of context on human-generated randomness and individual differences in producing responses at random.

In Study 1, we investigated the effect of context and the effect of mental fatigue on human-generated randomness. Participants produced a 300-elements long binary series under one out of three experimental conditions (cf. Fig. 2): (a) random sequence generation with NI; (b) instruction to imagine results of fair coin tosses and report the outcomes (CTI); and (c) instruction to imagine a stock market chart with its “ups” and “downs” and report the outcomes (SMI). Under hypothesis (H1), we expected that random generation tasks that refer to processes known from day-to-day life would allow for producing more random (algorithmically complex) series than under more abstract instructions. The results showed that there were differences between experimental conditions. Under NI condition, participants produced series less random than under the other two conditions. A detailed analysis showed that the observed effect was driven primarily by the fact that subjects in the CTI and SMI conditions were more capable of avoiding trivially patterned, non-random series than those in the NI group (see Fig. 3). It suggests that, contrary to popular belief (compare Ayton et al., 1991, for a review), instructions incorporating examples of random processes do not facilitate producing more random series in terms of an increase of the average performance but rather help to activate the ability whatsoever and therefore avoid producing trivial series of low complexity (e.g., containing mainly 1’s or 0’s). Such results might be explained in terms of the Relevance Theory (Wilson & Sperber, 2004). In the NI condition, the task description did not yield a positive cognitive effect for some participants and therefore did not activate the right mechanism — the ability to produce random-like series. That is, the situation did not resonate with participants’ previous experience and knowledge about the random process, and as a result, they produced series of low complexity. On the other hand, in the other two conditions, the task descriptions yielded a positive cognitive effect for the majority of the participants.

Furthermore, these results are congruent with the modern theories of cognitive control that argue that the behavior is guided by signals that might be categorized as the ones related to past events or contextual cues. For example, the cascade model (Koechlin et al., 2003; Koechlin & Summerfield, 2007) assumes that cognitive control is organized as multistage hierarchical architecture in three nested processing levels: (a) sensory control responsible for selecting motor responses to stimuli; (b) contextual control involved in premotor representations that are induced by external information accompanying the stimulus; and (c) episodic control involved in premotor representations based on experience or ongoing internal goals. Therefore, under CTI and SMI conditions, contextual cues might have activated known representation of the random series generator. That lead to simulating known process and consequently avoidance of trivial, non-random series. On the other hand, under the NI condition, contextual cues (the instruction) were less likely to be associated with known representations. Hence, on average, people in this condition were less likely to be able to produce non-trivial random-like series than in the SMI and CTI conditions. However, this explanation requires further examination that would indeed show that under more abstract — less likely to be associated with known representations — contextual cues different regions of the prefrontal cortex are activated (Koechlin et al., 2003).

The results of Study 1 allowed also for testing the second hypothesis (H2) that over time the performance (algorithmic complexity of produced series) in random generation tasks decreases. As a consequence, we observed a simple negative effect of time on random-like series production. The prolonged execution of such cognitively demanding tasks induced mental fatigue and consequently resulted in a decrease in task performance. The failure to maintain the same level of randomness of generated series was independent of the contextual cues. We observed a similar negative effect of time under all experimental conditions. That might suggest that tasks in all experimental conditions lost attractiveness at the same rate. Therefore, to maintain the highest level of randomness, the cognitive motivation to perform in the task would have to increase inversely proportional to the rate of decrease of its attractiveness. People either would have to find new levels of intrinsic cognitive motivation to continue performing well in an effortful task or the task itself would have to continuously become more attractive. We further investigated this effect of mental fatigue and cognitive motivation on random series generation in Study 2.

In Study 2, we wanted to test whether the trend observed in Study 1, which showed that over time the randomness of the series decreases, would be also present when people were asked to produce shorter independent series. Therefore, participants generated ten 12-elements long binary series in two experimental conditions (see Fig. 6): homogeneous and heterogeneous. In the former, we used the same instruction (CTI condition from Study 1) over all series, and in the latter, we altered two instructions (therefore each participant produced five series under CTI and five under SMI conditions from Study 1). After completing generative tasks, participants filled in the NFC scale. Although the results did not show differences between experimental conditions, we observed a similar trend in randomness over time as in Study 1. However, unlike Study 1, our analysis in Study 2 using the Generalized Additive Model captured a non-linear nature of the trend. The observed trend of algorithmic complexity over time (cf. Fig. 7) was approximately linear in the beginning but then after the fifth series, the curve flattened regardless of the experimental condition. It suggests that although the effect of fatigue has a negative influence, it does not lead, on average, to series with normalized algorithmic complexity lower than 0.5 (half of the maximum complexity). This indicates that similar to the effect of mental fatigue observed in Study 1, people tend to lose interest in the task very quickly after the first few attempts to generate the most random-like series they can, and then they can generate only markedly more patterned, less random-like sequences. This effect might be explained in terms of the Motivational Control Theory of Cognitive Fatigue (Hockey, 2011) and the decrease of the task attractiveness. People very quickly realize that it is a repetitive task and lose interest in it. However, the lack of differences between conditions and the fact that regardless of a series' number people maintained a certain level of randomness might be attributed to the fact that they were motivated by the experiment itself. Therefore, we suspect that their baseline level of cognitive motivation is higher than it would be if they were meant to complete a similar task in more natural conditions. That could explain why they were motivated enough to complete the study instead of giving up. Additionally, in Study 2, we tested the third hypothesis (H3) that randomness (algorithmic complexity) of human-generated series is positively associated with the tendency to engage in cognitively effortful tasks and enjoy them (NFC). We observed a positive, but weak, association between the randomness of the series and the need for engaging cognitive

resources in challenging tasks. This result is also congruent with the Motivational Control Theory (Hockey, 2011). People who found the effortful experimental task appealing were able to maintain their interest in it for longer periods of time, and thus, the average level of randomness in their series was higher. Hence, the variability of the randomness of human-generated series is not only a function of contextual cues (cf. results of Study 1) but also of intrinsic cognitive motivation to maintain engagement in an effortful task.

Finally, our results might be interpreted from the broader perspective of the correlated nature of human behavior. Studies on this topic show that repeated human responses tend to follow a specific neither highly structured nor random pattern. This pattern of responses referenced in the literature as $1/f$ scaling (or $1/f$ noise) has been observed across multiple domains of human behavior such as mental rotation and transition (Gilden, 1997), visual search (Aks et al., 2002), simple classifications (Kelly et al., 2001) color and shape discrimination (Gilden, 2001), speech (Kello et al., 2008), and even in reaction times and key-press times during random series generation (Kello et al., 2007). Kello et al. (2007) argue that this is a domain-independent phenomenon as it results from how systems' components interact with each other. Therefore, under this framework, cognitive functions emerge as metastable patterns of neural and bodily activity operating flexibly between ordered and disordered phases, that is, with strong or weak coupling between different components of the system. Crucially, in this view, $1/f$ scaling is indicative of an optimal level of synchronization and interactions between different system's components, which is neither too weak nor too strong. Unlike the domain-specific approach that locates specific brain regions responsible for cognitive functions, here the cognitive functions arise as a consequence of both internal and external constraints of the system (Bressler & Kelso, 2001). In more psychological terms, they are a function of contextual and intrinsic variability. Under this theoretical framework, the results of our studies show that the dynamics of random series production might be explained in terms of the interaction between contextual cues and internal constraints. Therefore, the contextual cues (instructions in both studies) and internal constraints (cognitive resources and motivation to perform well in effortful tasks) need to be in an optimal systemic interaction to allow the production of non-trivial, random-like sequences. This would explain why in the shape of the trend curve in Study 2, we observed flattening after the first few series (cf. Fig. 7). This effect might be interpreted in terms of vanishing synchronization between interacting components. At first, people can use their cognitive resources, task-related motivation, and contextual cues provided by a task instruction to generate random-like data but after a while the fatigue kicks in effectively pushing the system out of a parameter region optimal for the task at hand leading to much more regular and patterned series being produced. This interpretation could also explain the non-linear effect (i.e., located in the lower tail of the distribution) of contextual cues observed in Study 1. Without the cues, it is significantly harder for the system to even reach the state optimal for random series generation due to the lack of appropriate external constraints. Such a hypothesis is indirectly corroborated by results, indicating that contextual cues can indeed affect the functioning of different functional components and brain regions such as motor areas and prefrontal cortex (Koechlin et al., 2003; Mugruza-Vassallo et al., 2021). In other words, patterned, non-random series are the typical results of internal dynamics of the cognitive system without additional external constraints and it is such constraints that help to push the system from the ordered to disordered phase. It remains an open question

whether such a shift at the level of the organization of the output, that is, binary series, corresponds to a parallel change, perhaps in the opposite direction, at the level of some different parameters of the cognitive system. Such hypotheses and intuitions need further examination, preferably based on power spectrum methods typical for studies of $1/f$ scaling, a task which we defer to future work.

4. Conclusions

Taken together, our results indicate that by default, people tend to produce non-random series. However, either intrinsic factors or contextual cues might improve the performance in such tasks. Results of Study 2 showed that this effect is only temporal and very quickly the randomness of human-generated series return to what seems to be a stable and typical level of algorithmic complexity. Moreover, the results of Study 1 shed new light on the effect of contextual cues on human-generated randomness. The activation of the ability to produce random-like series depends on the relevance of the contextual cues, whether they activate known representations of a random series generator, and consequently, help to avoid the production of trivial sequences. Although in our approach, we refer to instruction manipulation as contextual cues clearly elements selection (random-like series generation) was a top-down process. It was not related to the physical characteristic of the environment or the stimuli (participants were asked to imagine relevant processes when they saw a red square) but rather to internalized goals or expectations based on the instruction manipulation. Therefore, for future work, it would be interesting to address whether discrepancies in the physicality of the stimulus would affect task performance. Rincon-Gonzalez et al. (2012) showed that integration of both proprioception and tactile apparatus diminished the magnitude of the error in hand location tasks. Similarly, one could expect that adding the stimulus that is more relevant to the random series generation tasks could also affect the variability of the randomness in the tasks.

Notes

- 1 It is available on GitHub under MIT License <https://github.com/MikoBie/Survey>.
- 2 It is a Python package implementing Coding Theorem and Block Decomposition methods for estimating algorithmic complexity (Soler-Toscano et al., 2014; Zenil et al., 2018). It is available as a standard package through PyPI <https://pypi.org/project/pybdm/>.
- 3 <https://pypi.org/project/pybdm/>

Acknowledgments

Project was supported by the Polish National Science Centre under grant agreement No.: 2019/35/N/HS6/04318.

Open Research Badges



This article has earned Open Data and Open Materials badges. Data are available at <https://osf.io/2uqma/> and materials are available at <https://osf.io/2uqma/>.

REFERENCES

- Aks, D. J., Zelinsky, G. J., & Sprott, J. C. (2002). Memory across eye-movements: 1/f dynamic in visual search. *Nonlinear Dynamics, Psychology, and Life Sciences*, 6(1), 1–25.
- Ayton, P., Hunt, A. J., & Wright, G. (1991). Randomness and reality. *Journal of Behavioral Decision Making*, 4(3), 222–226.
- Baddeley, A. D. (1966). The capacity for generating information by randomization. *The Quarterly Journal of Experimental Psychology*, 18(2), 119–129.
- Baddeley, A. D. (1986). *Working memory*. England: Oxford University Press.
- Baena-Mirabete, S., Espinal, A., & Puig, P. (2020). Exploring the randomness of mentally generated head-tail sequences. *Statistical Modelling*, 20(3), 225–248.
- Bakan, P. (1960). Response-tendencies in attempts to generate random binary series. *The American Journal of Psychology*, 73(1), 127–131.
- Bar-Hillel, M., & Wagenaar, W. A. (1991). The perception of randomness. *Advances in Applied Mathematics*, 12(4), 428–454.
- Barbasz, J., Stettner, Z., Wierchoń, M., Piotrowski, K. T., & Barbasz, A. (2008). How to estimate the randomness in random sequence generation tasks? *Polish Psychological Bulletin*, 39(1), 42–46.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Bressler, S. L., & Kelso, J. (2001). Cortical coordination dynamics and cognition. *Trends in Cognitive Sciences*, 5(1), 26–36.
- Brugger, P. (1997). Variables that influence the generation of random sequences: An update. *Perceptual and Motor Skills*, 84(2), 627–661.
- Budescu, D. V. (1987). A Markov model for generation of random binary sequences. *Journal of Experimental Psychology*, 75(1), 25–39.
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42(1), 116–131.
- Champernowne, D. G. (1933). The construction of decimals normal in the scale of ten. *Journal of the London Mathematical Society*, 1–8(4), 254–260.
- Chapanis, A. (1953). Random-number guessing behavior. *American Psychologist*, 8(1995), 1347–1363.
- Dai, D. Y., & Wang, X. (2007). The role of need for cognition and reader beliefs in text comprehension and interest development. *Contemporary Educational Psychology*, 32(3), 332–347.
- Day, E. A., Espejo, J., Kowollik, V., Boatman, P. R., & McEntire, L. E. (2007). Modeling the links between need for cognition and the acquisition of a complex skill. *Personality and Individual Differences*, 42(2), 201–212.
- Dornic, S., Ekehammar, B., & Laaksonen, T. (1991). Tolerance for mental effort: Self-ratings related to perception, performance and personality. *Personality and Individual Differences*, 12(3), 313–319.
- Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review*, 104, 301–318.
- Fleischhauer, M., Enge, S., Brocke, B., Ullrich, J., Strobel, A., & Strobel, A. (2010). Same or different? Clarifying the relationship of need for cognition to personality and intelligence. *Personality and Social Psychology Bulletin*, 36(1), 82–96.
- Gauvrit, N., Singmann, H., Soler-Toscano, F., & Zenil, H. (2016). Algorithmic complexity for psychology: A user-friendly implementation of the coding theorem method. *Behavior Research Methods*, 48(1), 314–329.

- Gauvrit, N., Zenil, H., Delahaye, J.-P., & Soler-Toscano, F. (2014). Algorithmic complexity for short binary strings applied to psychology: A primer. *Behavior Research Methods*, 46(3), 732–744.
- Gauvrit, N., Zenil, H., Soler-Toscano, F., Delahaye, J.-P., & Brugger, P. (2017). Human behavioral complexity peaks at age 25. *PLOS Computational Biology*, 13(4).
- Gilden, D. (1997). Fluctuations in the time required for elementary decisions. *Psychological Science*, 8(4), 296–301.
- Gilden, D. L. (2001). Cognitive emissions of 1/f noise. *Psychological Review*, 108(1), 33–56.
- Hahn, U., & Warren, P. A. (2009). Perceptions of randomness: Why three heads are better than four. *Psychological Review*, 116(2), 454–461.
- Halprin, R., & Naor, M. (2009). Games for extracting randomness. In *Proceedings of the 5th Symposium on Usable Privacy and Security* (pp. 1–12).
- Hockey, G. R. J. (2011). A motivational control theory of cognitive fatigue. In P. L. Ackerman (Ed.), *Cognitive fatigue: Multidisciplinary perspectives on current research and future applications* (pp. 167–188). Washington: American Psychological Association.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3(3), 430–454.
- Kareev, Y. (1995). Through a narrow window: Working memory capacity and the detection of covariation. *Cognition*, 56, 263–269.
- Kello, C. T., Anderson, G. G., Holden, J. G., & Van Orden, G. C. (2008). The pervasiveness of 1/f scaling in speech reflects the metastable basis of cognition. *Cognitive Science*, 32(7), 1217–1231.
- Kello, C. T., Beltz, B. C., Holden, J. G., & Van Orden, G. C. (2007). The emergent coordination of cognitive function. *Journal of Experimental Psychology: General*, 136(4), 551–568.
- Kelly, A., Heathcote, A., Heath, R., & Longstaff, M. (2001). Response-time dynamics: Evidence for linear and low-dimensional nonlinear structure in human choice sequences. *The Quarterly Journal of Experimental Psychology Section A*, 54(3), 805–840.
- Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, 302(5648), 1181–1185.
- Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences*, 11(6), 229–235.
- Koenker, R. (2020). *quantreg: Quantile regression*. R package version 5.70.
- Koenker, R., & Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, 15(4), 143–156.
- Levin, I. P., Huneke, M. E., & Jasper, J. D. (2000). Information processing at successive stages of decision making: Need for cognition and inclusion–exclusion effects. *Organizational Behavior and Human Decision Processes*, 82(2), 171–193.
- Li, M., & Vitányi, P. (2008). *An introduction to Kolmogorov complexity and its applications*, Vol. 3. New York: Springer-Verlag.
- Lorist, M. M., Boksem, M. A., & Ridderinkhof, K. R. (2005). Impaired cognitive control and reduced cingulate activity during mental fatigue. *Cognitive Brain Research*, 24(2), 199–205.
- Lorist, M. M., & Faber, L. G. (2011). Consideration of the influence of mental fatigue on controlled and automatic cognitive processes and related neuromodulatory effects. In P. L. Ackerman (Ed.), *Cognitive fatigue: Multidisciplinary perspectives on current research and future applications* (pp. 105–126). Washington: American Psychological Association.
- Matusz, P. J., Traczyk, J., & Gasiorowska, A. (2011). Kwestionariusz Potrzeby Poznania konstrukcja i weryfikacja empiryczna narzędzia mierzącego motywację poznawczą. *Psychologia Społeczna*, 6(17), 113–128.
- Mugruza-Vassallo, C. A., Potter, D. D., Tsiora, S., Macfarlane, J. A., & Maxwell, A. (2021). Prior context influences motor brain areas in an auditory oddball task and prefrontal cortex multitasking modelling. *Brain Informatics*, 8(1), 1–28.
- Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R² and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, 14(134), 20170213.

- Neuringer, A. (1986). Can people behave “randomly?”: The role of feedback. *Journal of Experimental Psychology*, *115*(1), 62–75.
- Nickerson, R. S. (2002). The production and perception of randomness. *Psychological Review*, *109*(2), 330–357.
- Oskarsson, A. T., Van Boven, L., McClelland, G. H., & Hastie, R. (2009). What’s next? Judging sequences of binary events. *Psychological Bulletin*, *135*(2), 262–285.
- Petty, R. E., Brinol, P., Loersch, C., & McCaslin, M. J. (2009). The need for cognition. In M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior*, (Chapter 21, pp. 318–329). New York: The Guilford Press.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team (2020). *nlme: Linear and nonlinear mixed effects models*. R package version 3.1-148.
- R Core Team (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rath, G. J. (1966). Randomization by humans. *The American Journal of Psychology*, *79*(1), 97–103.
- Reas, C., & Fry, B. (2006). Processing: Programming for the media arts. *Ai & Society*, *20*(4), 526–538.
- Rincon-Gonzalez, L., Naufel, S., Santos, V., & Helms Tillery, S. (2012). Interactions between tactile and proprioceptive representations in haptics. *Journal of Motor Behavior*, *44*(6), 391–401.
- Ross, B. M. (1955). Randomization of a binary series. *The American Journal of Psychology*, *68*(1), 136–138.
- Schulter, G., Mittenecker, E., & Papousek, I. (2010). A computer program for testing and analyzing random generation behavior in normal and clinical samples: The Mittenecker pointing test. *Behavior Research Methods*, *42*(1), 333–341.
- Sharifian, S. (2016). *Random number generation using human gameplay* [PhD thesis]. University of Calgary.
- Slak, S., & Hirsch, K. A. (1974). Human ability to randomize sequences as a function of information per item. *Bulletin of the Psychonomic Society*, *4*(1), 29–30.
- Soler-Toscano, F., Zenil, H., Delahaye, J.-P., & Gauvrit, N. (2014). Calculating Kolmogorov complexity from the output frequency distributions of small turing machines. *PLoS One*, *9*(5).
- Tune, G. (1964). Response preferences: A review of some relevant literature. *Psychological Bulletin*, *61*(4), 286–302.
- Wagenaar, W. (1970). Subjective randomness and the capacity to generate information. *Acta Psychologica*, *33*, 233–242.
- Wagenaar, W. (1971). Serial non-randomness as a function of duration and monotony of a randomization task. *Acta Psychologica*, *35*(1), 78–87.
- Warren, P. A., Gostoli, U., Farmer, G. D., El-Deredy, W., & Hahn, U. (2018). A re-examination of “bias” in human randomness perception. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(5), 663–680.
- Warren, P. A., & Morin, R. E. (1965). Random generation: Number of symbols to be randomized and time per response. *Psychonomic Science*, *3*(1), 557–558.
- Wiegiersma, S. (1982). A control theory of sequential response production. *Psychological Research*, *44*(2), 175–188.
- Wilson, D., Sperber, D., Horn, L. R., & Ward, G. (2004). Relevance theory. *Handbook of pragmatics*. (607–632). Oxford: Blackwell.
- Yavuz, H. S. (1963). The production of random letters sequences in schizophrenics. *The Journal of Psychology*, *56*(1), 171–173.
- Zenil, H., Hernández-Orozco, S., Kiani, N., Soler-Toscano, F., Rueda-Toicen, A., & Tegnér, J. (2018). A decomposition method for global evaluation of Shannon entropy and local estimations of algorithmic complexity. *Entropy*, *20*(8), 605.