Can Humans Become More Random?

Sook Mun Wong

Chapman University

Garance Merholz

Université de Paris

Uri Maoz (✉ maoz@chapman.edu)

Department of Psychology, University of California Los Angeles, Los Angeles, CA, USA

Research Article

Keywords: human ability, random-sequence generation (RSG), Network-Modulation Model

Posted Date: February 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-151863/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Can humans become more random?

Sook Mun Wong^{1,2,4}, Garance Merholz^{1,7}, Uri Maoz^{1,2,3,5,6*},

¹ Department of Psychology, University of California Los Angeles, Los Angeles, CA, USA

² Institute for Interdisciplinary Brain and Behavioral Sciences, Chapman University, Irvine, CA

³ Department of Psychology, Crean College of Health and Behavioral Sciences, Chapman University, Orange, CA

⁴ Computational and Data Sciences, Schmid College of Science and Technology, Chapman University, Orange, CA

⁵ Anderson School of Management, University of California Los Angeles, Los Angeles, CA, USA

⁶ Biology and Bioengineering, California Institute of Technology, Pasadena, CA, USA

⁷ Université de Paris, CNRS, Integrative Neuroscience and Cognition Center, F-75006 Paris, France

* Corresponding Author:

Uri Maoz

Chapman University, One University Drive, Orange, CA 92866

Email address: maoz@chapman.edu

Abstract

The human ability for random-sequence generation (RSG) is limited but improves in a competitive game environment with feedback. However, it remains unclear whether RSG during games can improve when explicitly informing people that they must be as random as possible to win the game nor is it known whether any such improvement transfers to outside the game environment. To investigate this, we designed a pre/post intervention paradigm around a rock-paper-scissors game followed by a questionnaire. During the game we manipulated participants' level of awareness of the computer's strategy; they were either (a) not informed of the computer's algorithm or (b) explicitly informed that the computer used patterns in their choice history to beat them, so they must be maximally random to win. Using a novel comparison metric, our results demonstrate that human RSG can reach levels statistically indistinguishable from computer pseudo-random generators in a competitive-game setting. However, our results also suggest that human RSG cannot be further improved by explicitly informing participants that they need to be random to win. In addition, the higher RSG in the game setting does not transfer outside the game environment. Furthermore, we found that the underrepresentation of long repetitions of short patterns explains about a third of the variability in human RSG and discuss what might make up the remaining two thirds of RSG variability. Finally, we discuss our results in the context of the "Network-Modulation Model" and we ponder their potential relation to findings in the neuroscience of volition.

Introduction

Randomness—particularly random series generation (RSG)—is often used in neuroscience and psychology to increase cognitive load—e.g., as a distractor task from a main task¹⁻³. However, in the neuroscience of volition for example, the ability to act randomly is interesting as an upper bound on arbitrary action. In other words, asking someone to make decisions as randomly as possible sets an upper limit on how arbitrary their action can be. This assertion follows from the assumption that arbitrary action, devoid of any reasons or other constraints, reflects the most distilled type of free choice⁴⁻⁷. While this view has been criticized⁸⁻¹¹, it remains central in the field.

In a broader sense, it has been debated whether true randomness exists in nature and whether it can even be clearly defined^{12,13}. Nevertheless, characteristics that have been associated with random series include equiprobability of choices and sequential independence¹². For a binary series $(s_i)_{i=1}^n \in \{0,1\}^n$, for example, equiprobability formally means that for any $b \in \{0,1\}$, $p(s_i = b) = 0.5$. More generally, for any $b \in \{0,1, \dots, m\}$, $p(s_i = b) = \frac{1}{m}$. Sequential independence—also “memorilessness” or negative recency¹⁴—means $p(s_i = b | s_{i-1}, \dots, s_1) = p(s_i = b)$. In other words, sequential independence means that any entry in a series is not influenced by the entries before it.

When instructed to carry out RSG, humans are able to generate series that are empirically equiprobable. However, humans typically fail to generate series that are sequentially independent^{12,15}. To delve deeper into this concept, let us first define a *run* in a series as a sequence of the same entry (e.g., the series (0,0,0,1,1,0) is made up of 3 runs: the three consecutive 0’s, the two consecutive 1’s, and the final 0). Now, humans tend to produce systematically biased series, switching too often between entries, and thus underrepresenting long runs in their generated series^{15,16}. It has nevertheless been found that, with detailed instructions and when they can see their generated series, human RSG becomes more sequentially independent and thus more random^{15,16}. In particular, in a competitive environment with feedback—e.g., a matching-pennies game—humans exhibit higher sequential independence^{13,16}. However, while it appears that humans can be more or less random in certain contexts, it remains unclear how any increase or decrease in randomness in humans compares to non-human objective benchmarks—e.g., how it compares to pseudorandom series generation (pseudo-RSG) by computer algorithms.

Here we aim to integrate three strands of research across multiple disciplines. One pertains to how human behavior is shaped by competitive contexts, also studied in economics and game theory^{13,16}. The second relates to comparing human RSG to computer pseudo-RSG the measurement of random behavior in relation to a comparable random (but non-human) process. And the third contextualizes the implications human RSG has for understanding cognitive processes within multiple fields of psychology related to volition, decision-making, psychopathology, and the underlying brain structures that support these processes. In particular, we aim to answer the following questions. (1) Can human RSG in a game situation be improved when specifically informing people that they must be as random as possible to win the game? (2) Is human RSG in a game environment similarly random to the pseudo-RSG

commonly used in modern programming languages? (3) Is the improved RSG exhibited in a game environment transferrable to a post-game environment? (4) To what extent are changes in RSG ability related to changes in the length of generated runs?

Methods

Participants

153 undergraduate students (112 female, age 20.7 ± 2.2 years (mean $\pm$ SD)) were recruited through the Psychology Department participant pool at the University of California, Los Angeles (UCLA). The study was approved by UCLA's North General Institutional Review Board (NGIRB; IRB#15-001094). All procedures were performed in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants, and all individuals received course credit for participation.

Experimental design and procedure

The experiment as a whole was designed as a 3×1 between-participant study—with the conditions being (1) *Unaware*, (2) *Aware*, and (3) *Control*. Each experimental session was divided into 3 parts, with 200 trials per part, separated into 2 blocks of 100 trials. All participants completed all 3 parts of the experiment. In each trial, participants selected Rock (R), Paper (P), or Scissors (S) by pressing the <G>, <H>, or keys on the keyboard, respectively. In the first—*pre-game*—part of the experiment, participants were instructed to endogenously generate a sequence of R-P-S that was as random as possible. For the second—*game*—part of the experiment, participants were told they would be playing R-P-S with the computer as their opponent, and their goal was to win as many points as possible. During this game part of the experiment, the computer used a specific strategy to predict the participant's next move or not, depending on the experimental condition (see below for details). In the third and final—*post-game*—part of the experiment, the instructions from the baseline part were repeated. The progress of a trial in the experiment is depicted in Figure 1.

***** INSERT FIGURE 1 ABOUT HERE *****

Figure 1. The progress of each trial in the experiment. Following the countdown, participants were required to press the appropriate key for R, P, or S within 500 ms of the onset of the Go signal; otherwise the trial was forfeit. Their selection was then presented on the screen. In game trials (on the right), their selection was accompanied by the computer's selection and by the game score.

The participant and the computer started the game with 100 points each. The winner was the player with more points at the end of the game. Each player gained or lost one point for every trial they won or lost, respectively. Ties left the score unchanged. Participants were further randomly assigned into one of 3 conditions during the game part of the experiment. All were instructed to do their best to win the game. However, participants in the Unaware and Control conditions were told nothing about the computer's strategy. Those in the Aware condition were explicitly informed that the computer would try to predict their moves and it was stressed that to win they must be as random as possible (Table 1).

***** INSERT TABLE 1 ABOUT HERE *****

Pre-game (Baseline)	Game	Post-Game
	<u>Condition 1, Unaware</u> : Computer exploits patterns in participant's choice history against the participant. Participant is not informed of computer's strategy.	
	<u>Condition 2, Aware</u> : Computer exploits patterns in participant's choice history against the participant. Participant is explicitly informed of computer's strategy and told that to win they must be random.	
	<u>Condition 3, Control</u> : Computer follows simple pattern continuously 85% of the time and acts randomly 15% of the time. Participant is not informed of computer's strategy.	

Table 1. Within- and between-participants division of the experiment. *Each participant carried out all three parts of the experiment (pre-game, game, and post-game). Different participant groups were randomized to different conditions in the game part of the experiment.*

In the Unaware and Aware conditions, the computer sought out behavioral patterns in the participant's choice history (e.g., R-R-P-P-S-S-...), their win/lose/tie patterns (e.g., win-stay-lose-switch) and used those against them (for details on the algorithm used see Barraclough, et al. ¹⁷; Maoz, et al. ¹⁸). So, the participant's best strategy was to be as random as possible, minimizing the computer's ability to find patterns in their choice history. In the Control condition, the computer followed a pattern of R-P-S-R-P-S-... with 85% probability on any given trial, with a remaining 15% chance of choosing R, P, or S (with equiprobability) on any given trial. Participants were then given a post-experiment questionnaire asking them about the task, and their performance in the game.

Of our 153 participants 105 answered the post-experiment questionnaire. We asked them to rate how much they agreed with the following 7 statements, each on a Likert scale of 1 to 7 (1 = strongly disagree, 7 = strongly agree):

- 1) Whether they used a strategy against the computer.
- 2) Whether they felt that the computer was predicting their moves.
- 3) Whether they managed to create very random sequences in part 1 (before the game) and part 3 (after the game).
- 4) Whether the sequence they created in part 3 was more random than the one they created in part 1.
- 5) Whether they won most of the trials during the game.
- 6) Whether they tied on most trials during the game.
- 7) Whether they lost most of the trials during the game.

Participant- and trial-inclusion criteria

Figure 2. Visualization of inclusion criterion 1. The solid black line indicates an absolute maximum divergence in the proportion of R, P or S of 0.15. Sequences above this inclusion criterion were subsequently discarded from analyses. Sequences for each experiment part (3 per participant) are sorted by proportion of R, P and S that diverge from 1/3—i.e., diverge from uniform distribution.

***** INSERT FIGURE 2 ABOUT HERE *****

In the post-experiment questionnaire, some participants reported frustration during the game part that resulted in loss of interest in the entire experiment, because they were generally losing and unable to beat the computer. We wanted to only analyze the data of participants who remained engaged with the experiment and continued to make an effort to win throughout the game part and further made an effort to be random during the RSG post-game part. We therefore included only participants who fulfilled both of the following inclusions criteria. (1) Deviation from an even (i.e., 1/3) R-P-S split was no more than 0.15 in any part of the experiment (Figure 2). (2) Participants lost to the computer during the game part by no more than 65 points (Figure 3).

***** INSERT FIGURE 3 ABOUT HERE *****

Figure 3. Visualization of inclusion criterion 2. Difference between human and computer final game scores are shown for the 3 experimental conditions. Participants who lost to the computer by more than 65 points (Human – Computer ≤ -65 ; solid, horizontal line) were discarded from subsequent analysis. Sequences are sorted by the final-score difference between human and computer.

Of the 153 participants, 32 (20.9%) did not meet both these inclusion criteria and were excluded from further analysis. Of those, 21 of 61, 9 of 50, and 2 of 42 participants were excluded from the Unaware, Aware, and Control conditions, respectively. We expected the most participants to be excluded in the Unaware condition followed by the Aware condition, and least in the Control condition. This is because it was rather difficult to beat the computer (in Fig. 3 that would entail a difference >0), and participants in the Unaware condition had to learn for themselves how to beat the computer in the game part, while those in the Aware condition were told how to beat the computer. In contrast, in the Control condition, it was intentionally simple to beat the computer. Given the above, we accordingly assigned more participants to the Unaware condition during randomization. We therefore ended up with 40, 41, and 40 participants in the Unaware, Aware and Control conditions respectively.

There were also two types of errors that could occur on each trial. (1) The participant did not press a key within 500 ms of the Go signal, and (2) the participant pressed a key other than <G>, <H>, or . Both errors resulted in a forfeited trial. Further—in the game part—a point was awarded to the computer in such cases without a loss of a point for the participant. The first error occurred on $3.1 \pm 2.6\%$ of the trials and the second on $0.5 \pm 0.8\%$ of the trials (mean $\pm$ STD). Those trials were removed from further analysis.

Data analyses

Measure of randomness

To quantify the randomness of a given sequence, we used a normalized Lempel-Ziv Complexity (LZC) score^{19,20}, similarly to previous research²¹. We did not use any measures of randomness that directly rely on sequential dependence because, in the game part for the Unaware and Aware conditions, the computer algorithm relied on this measure to play against the participant. Thus, using a similar pattern-matching algorithm to test for randomness could amount to double-dipping, as participants were already evaluated against this measure by feedback from the algorithm in the game.

The LZC score was obtained by compressing each human generated sequence using the variable-rate compression algorithm by Ziv and Lempel²⁰. We then normalized the LZC score by generating 1000 pseudorandom sequences of the same length as the human generated sequence, and divided the human generated sequence score by the mean LZC score of the pseudo-randomly generated sequences²². This was to remove the effect of small variations in sequence length on the raw LZC score, due to the removal of error trials, as a shorter sequence may naturally be less complex than a longer sequence due to its length. Thus, a normalized LZC score closer to 1 indicates that a sequence is more complex, and thus random, whereas a score closer to 0 indicates that a sequence is less random. For example, take a sequence [3 2 1 2 1 2 1 2 2 3] of length 10 as an example. After compressing it with the LZ78 algorithm, the sequence is stored as [3 2 1 257 269 2 3] (see Ziv and Lempel²⁰ for details). The length of this compressed sequence is therefore 7. To normalize it, we take 1000 pseudorandom sequences of length 10, so perhaps $X_1 = [1 2 1 1 3 1 2 2 2 1]$, $X_2 = [2 1 3 1 1 2 1 2 1 2]$, ..., $X_{1000} = [3 3 2 2 3 3 1 2 3 3]$. Then the normalized LZ score is

$$\frac{7}{\text{mean}(\text{length}(\text{LZ78}(x_i)))}$$

Permutation tests

To compare the randomness scores of the subjects to those generated by the computer pseudorandom number generators, we had to first understand the range of the pseudorandom normalized LZ scores. We therefore generated 1000 pseudo-random sequences of R-P-S of length 192 (the length of the average human sequence). Then, for each pseudorandom sequence separately, we ran the identical code that we used on the human data to compute their randomness scores. This gave us 1000 randomness scores to which we could compare the randomness score that we computed on the human data. The pseudo-random sequences were generated by Python's pseudo-random-number generators (using the Mersenne twister algorithm). We then took 5% of the randomness scores as the critical threshold (alpha level) for LZC scores (see below) to designate significant deviations from chance level. The

bootstrapped 2.5th to 97.5th percentile of the normalized LZ score distribution was [0.973, 1.027] for a sequence length of 192. Thus, for a random sequence as random as Python's random-number generator, a sequence's normalized LZ score can be expected to fall between 0.973 and 1.027 with a 5% alpha level in a two-tailed test. Because the sequences were all generated empirically, the upper range of the normalized LZ score can exceed 1 if the numerator sequence's LZ score happened to be larger than the average mean of 1000 sequences other sequences, despite a theoretical upper limit of 1 for an LZ score.

Explanatory variable: average run length

We also computed an explanatory statistic called the run-length—the mean length of a run in a series. For instance, the run-length of the sequence R-R-R-P-P-S-P-P is the mean of [3,2,1,2], which is 2. The run-length is related to the randomness of a sequence in that the longer the run-length of a sequence is, the fewer runs there will be for a sequence of fixed length. We bootstrapped this measure, similarly to the randomness scores above, to obtain an empirical distribution of mean run-lengths. The 2.5th to 97.5th percentile range was [1.363, 1.658]. For a perfectly sequentially independent sequence of 3 items, the ideal run-length is 1.5.

Results

R-P-S game results

Our manipulation in the game part worked well. Participants found it difficult to beat the computer in both the Aware and Unaware conditions, where the winning strategy required them to be random. On the other hand, participants in the Control condition, where the winning strategy was to almost always follow the simple P-S-R pattern, found it easy to win (Fig. 3). This is also evidenced by the fact that, after exclusion of participants who gave up on the task (see Methods), the average final scores for the participants in the Unaware, Aware, and Control conditions were 83.2 ± 10.5 , 83.5 ± 12.3 , and 143.1 ± 49.7 (mean $\pm$ STD), respectively. A final game score of over (under) 100 indicated that the participant won over (lost to) the computer. Overall, the experimental manipulation created a large effect²³. Participants were much more likely to lose to the computer in the Unaware and Aware conditions, than under the Control condition (ANOVA with Welch correction for homogeneity, $F(2, 118) = 27.48$, $p < 0.001$, $\eta_p^2 = 0.47$; post-hoc t-test comparison with Bonferroni correction for Unaware vs Control: $t(39) = -8.84$, $p < 0.001$, Cohen's $d = -1.65$, 95% CI [-75.96, -43.79]; Aware vs Control: $t(40) = -8.84$, $p < 0.001$, Cohen's $d = -1.64$, 95% CI [-75.52, -43.55]).

Similarly, after exclusions (see Methods), 3 out of 40 (7.3%), 4 out of 41 (10.3%), and 31 of 40 (77.5%) participants beat the computer in the Unaware, Aware, and Control conditions, respectively. Hence, instructing the participants that they needed to be as random as possible to beat the computer in the Aware condition resulted in no statistically significant increase in their ability to beat the computer (ANOVA with Welch correction for homogeneity, $F(2, 118) = 43.32$, $p < 0.001$, $\eta_p^2 = 0.49$; post-hoc t-test comparison with Bonferroni correction for Unaware vs Aware, $t(80) = -0.30$, $p = 1.0$, Cohen's $d = -0.08$, 95% CI [-0.2, 0.16]. The significant p-value in the main effect of the ANOVA was due to the differences between the Control condition and the experimental

conditions (post-hoc t-test Unaware vs Control: $t(80) = -9.30$, $p < 0.001$, Cohen's $d = -1.98$, 95% CI $[-0.89, 0.52]$; post-hoc t-test Aware vs Control: $t(79) = -9.06$, $p < 0.001$, Cohen's $d = -1.85$, 95% CI $[-0.86, 0.50]$). Using a Bayesian ANOVA (in JASP ²⁴) the model with condition as an effect versus the null model with no effects had a Bayes Factor (BF) of 6.32×10^{14} , which is overwhelming evidence for the alternative model. A post-hoc comparison corrected for multiple testing between the Unaware and Aware condition showed a Bayes Factor of 0.24, which provides moderate evidence for the hypothesis that there was no difference in wins/losses between the two experimental conditions ²⁵. (The strong evidence for the alternative model in the main comparison was driven by comparisons between the Control and Experimental conditions. Post-hoc Bayesian t-tests for Unaware vs Control had a $BF_{10} = 2.55 \times 10^{10}$, and for Aware vs Control, $BF_{10} = 2.94 \times 10^9$).

Randomness scores

We measured the randomness of the series that our subjects generated using LZC scores (see Methods). The distributions of the LZC scores for each condition and part of the experiment are shown as violin plots in Figure 4.

***** INSERT FIGURE 4 ABOUT HERE *****

Figure 4. LZC scores of human generated sequences. Violin plots of LZC scores by experimental part and condition are shown with the mean and 95% confidence intervals superimposed in black. The means are also indicated above each violin plot. The horizontal, solid red line indicates the mean LZC score of the 2.5th percentile of 1000 bootstrapped pseudo-random sequences, for baseline comparison. Values above this line are statistically indistinguishable from computer-generated pseudo-random sequences.

We separated the analyses for the experimental and control conditions because the control condition was specifically designed to elicit opposite effects (i.e., minimize randomness) during the game part. So, any effects we might find would be obscured and averaged out if pooled together across Unaware, Aware, and Control (see methods and Fig. 4). We therefore conducted two separate repeated measures ANOVAs with a between-subjects factor (condition: Unaware, Aware) and a within-subjects factor (experiment part: baseline, game, post-game) to compare the LZC scores across conditions and across experimental parts. The first ANOVA compared the experimental conditions (Unaware vs. Aware) across experimental parts, while the second looked only at the Control condition.

Our results suggest that for both the Aware and Unaware conditions, the subjects were significantly more random in the game condition compared to the baseline and post-game conditions, on average. At the same time, we found evidence that our subjects were similarly random in the Aware and Unaware conditions in the game part, again on average. For the Unaware vs Aware ANOVA, the LZC scores across the experiment

parts were significantly different and the experimental manipulation created a large effect on the LZC scores²³, ($F(2,158) = 27.89$, $p < 0.001$, $\eta_p^2 = 0.26$). Post-hoc paired t-tests with Bonferroni correction for multiple comparison revealed that participants were significantly more random during the game part than in pre-game and post-game (Baseline vs. Game: $t(80) = -7.31$, $p < 0.001$, Cohen's $d = -0.81$, 95% CI [-0.038, -0.02]; Game vs. Post-game: $t(80) = 4.94$, $p < 0.001$, Cohen's $d = 0.55$, 95% CI [0.01, 0.029]). The baseline however, was not significantly lower than the post-game (Baseline vs. Post-game: $t(80) = -2.37$, $p = 0.057$, Cohen's $d = -0.26$, 95% CI [-0.019, 0.00019]). An overall post-hoc paired t-test with Bonferroni correction between the Aware and Unaware conditions showed no significant difference ($t(79) = -0.88$, $p = 0.38$, Cohen's $d = -0.098$, 95% CI [-0.02, 0.006]). We ran an additional analysis, where we tested a Bayesian repeated-measures ANOVA model with the experiment part only as a factor versus the null model with no effects. We found that the model with the experiment part had the highest Bayes Factor, 2.99×10^8 , which is overwhelming evidence for the alternative model. The model with the experiment part and condition both as factors and the model with both factors including an interaction between the factors had Bayes Factors of 9.13×10^7 and 1.17×10^7 , respectively. Post-hoc comparisons corrected for multiple comparisons found that the Unaware vs. Aware condition had a Bayes Factor of 0.24, again supporting (with moderate evidence) the hypothesis that there was no difference in LZC scores across the experimental conditions. In order to further investigate the lack of significance between the pre-game and post-game LZC scores found earlier, we conducted another post-hoc comparison but found a Bayes Factor of 1.39, which leaves us with inconclusive evidence on whether pre-game and post-game LZC scores came from the same distribution or not.)

We also tested LZC scores on a subject-by-subject basis. In the Unaware condition, 29 of the 40 participants, 72.5%, were more random (according to their LZC scores) during the game than during both pre- and post-game, which is significantly more than expected by chance (binomial test $p < 0.001$; chance level is 25%). Similarly, in the Aware condition, 23 of the 41 participants, 56.1%, were more random during the game than in both pre- and post-game, again more than expected by chance (binomial test $p < 0.001$). Hence, both the average and individual LZC scores suggest a pattern of subjects being more random in the game part compared to the baseline and post-game parts for the Aware and Unaware conditions.

Importantly, we also found that during the game part the LZC scores in the Aware and Unaware conditions were as random as that of an equivalent computer-generated pseudo-random series according (permutation test—Fig. 4).

For the Control condition, subjects' scores versus those of the computer (Fig. 3), suggest that participants realized that the computer was generally using a consistent strategy of R-P-S, with rare deviations (see Methods). Hence, we found that—as per our design—subjects were less random on average during the game part compared to the baseline and post-game parts (repeated measures ANOVA (Greenhouse-Geisser correction), looking at the control condition, LZC scores across the experiment parts significantly different ($F(1.27, 49.43) = 14.52$, $p < 0.001$, $\eta_p^2 = 0.27$)). Post-hoc paired t-

tests with Bonferroni correction showed that participants were significantly less random during the game part of the experiment than during the pre-game and post-game parts, with the baseline and post-game not significantly different from each other (Baseline vs. Game: $t(39) = 3.817$, $p = 0.001$, Cohen's $d = 0.60$, 95% CI [0.02, 0.10]; Game vs. Post-game: $t(39) = -4.137$, $p < 0.001$, Cohen's $d = -0.65$, 95% CI [-0.10, -0.03]; Baseline vs. Post-game: $t(39) = -0.349$, $p = 1$, Cohen's $d = -0.06$, 95% CI [-0.02, 0.01]). As for subject-by-subject results, 8 out of 40 were more random during the game than during both pre- and post-game (binomial test $p = 0.82$) and 24 of 40 were less random during the game than during both pre- and post-game (binomial test $p < 0.001$). So, taken together, these results suggest a pattern of subjects being less random in the game part compared to the baseline and post-game parts for the Control condition.

Run-length measure

Investigating further why participants' LZC scores varied in the above pattern, we looked at the length of the runs that the participants were creating. Thus, we further calculated and compared the average run-lengths of participants' sequences across all conditions (Fig. 5).

***** INSERT FIGURE 5 ABOUT HERE *****

Figure 5. Average run-length scores of human generated sequences. Average run-length scores by experimental part and condition are indicated below each violin plot and in black inside each violin plot with 95% CI. The horizontal red line indicates the mean run-length of the 2.5th percentile of 1000 bootstrapped pseudo-random sequences.

For the experimental conditions, the run-length results generally mirrored those expected from the pattern of LZC scores. (Repeated-measures ANOVA comparing the Unaware and Aware conditions, average run-lengths across the experiment parts were significantly different ($F(2,158) = 28.08$, $p < 0.001$, $\eta_p^2 = 0.26$). Participants also had significantly longer average run-lengths during the game part of the experiment than during the pre-game baseline and post-game parts (post-hoc paired t-tests with Bonferroni correction: Baseline vs. Game: $t(80) = -7.46$, $p < 0.001$, Cohen's $d = -0.83$, 95% CI [-0.19, -0.10]; Game vs. Post-game: $t(80) = 3.15$, $p < 0.01$, Cohen's $d = 0.35$, 95% CI [0.01, 0.11].) Unlike the LZC scores, however, the post-game average run-length was significantly higher than the baseline (Baseline vs. Post-game: $t(80) = -4.32$, $p < 0.001$, Cohen's $d = -0.50$, 95% CI [-0.13, -0.04]), with no significant effect of condition (overall post-hoc paired t-tests with Bonferroni correction: Unaware vs. Aware, $t(79) = -1.54$, $p = 0.128$, Cohen's $d = -0.17$, 95% CI [-0.10, 0.03]).

This trend also held for individual participants. In the Unaware condition, 28 of the 40 participants, 70%, had longer runs during the game than during both pre- and post-game, which is significantly more than expected by chance (binomial test $p < 0.001$). Further, the run-length in the post-game part was longer than in the baseline ($t(39) = -3.39$, $p = 0.01$, 95% CI [-0.18, -0.01]). For the Aware condition, 24 of the 41 participants, 58.5%, had longer runs during the game than during both baseline and post-game, which is once more significantly more than expected by chance (binomial test $p < 0.001$). The run-length in the post-game part was however, not significantly longer than in the baseline ($t(40) = -2.71$, $p = 0.11$, 95% CI [-0.16, 0.008]).

Similar to the situation for the LZC score, in both experimental conditions, the mean run-length was not significantly different from that of an equivalent computer pseudo-RSG only during the game (permutation test—Figure 5).

In the control condition, no significant differences in run-length was found among the 3 parts of the task (repeated-measures ANOVA with Greenhouse-Geisser correction for sphericity, $F(1.18, 46.05) = 0.61$, $p = 0.46$, $\eta_p^2 = 0.015$). The analysis of individual participants did not result in any clearly reliable differences either: while more participants than expected by chance—17 of the 40 participants—had longer runs during the game than during both pre- and post-game (binomial test $p = 0.012$), there was also an almost significant number of participants that had shorter runs during the game compared to both pre- or post-game—15 out of 40 (binomial test $p = 0.054$). This helps explain why, on average, there were no differences between the 3 parts of the experiment.

We wanted to understand the extent to which the variance in randomness might be explained by the run length. We therefore regressed each sequence's average runs-length onto its corresponding LZC score. We found a significant correlation between the two, with the run length explaining almost 3/10 of the variance in randomness across all experiment parts and conditions combined ($R^2=0.28$, $p < 0.001$). We similarly found significant correlations between run length and LZC scores for each experiment part separately ($R^2=0.35$, $R^2=0.33$, $R^2=0.32$, for baseline, game, and post-game, respectively; all $p<0.001$, all OLS regressions).

Questionnaire answers

As part of the post-experiment questionnaire, subjects reported whether they used a strategy against the computer in the game part of the experiment. We thus tested whether there was a difference between their responses in the Aware and Unaware subject groups. Subjects' mean ratings was 3.86 ± 1.61 , 4 for Unaware, 4 ± 1.61 , 4 for Aware and 5.21 ± 2.02 , 6 for Control (mean $\pm$ STD, median). Hence, the Aware and Unaware conditions were both reliably different than the control but not from each other (ANOVA $F(2, 99) = 6.38$, $p<0.001$, $\eta_p^2 = 0.11$. Post-hoc unpaired t-tests: Unaware vs Aware $t(63) = -0.32$, $p=1$; Unaware vs Control $t(64) = -2.78$, $p=0.02$; Aware vs Control $t(71) = -3.28$, $p=0.004$.)

In a similar manner, we tested whether subjects' perceptions of being predicted by the computer differed between the Aware and Unaware condition. Subjects' average score was 4.06 ± 1.93 , 5 for Unaware; 4.62 ± 1.80 , 5 for Unaware; and 3.65 ± 2.16 , 4 for the Control group (mean $\pm$ STD, median), which were not significantly different (ANOVA $F(2, 99) = 1.96$, $p=0.15$, $\eta_p^2 = 0.04$).

We also tested whether participants' perception of their improved ability to generate randomness after the game correlated with their actual improvement in randomness as measured by their LZC score. The correlation between their answer to questionnaire question 4 (whether the sequence they created in part 3 was more random than the one they created in part 1) and the difference in LZC score between the post-game and baseline sequences was small and not statistically significant (Pearson's $R = 0.12$, $p=0.22$). Similarly, we tested the correlation between the LZC score difference and ratings of question 3 (whether they managed to create very random sequences in pre-game and post-game). Again, the correlation was very small and not significant (Pearson's $R = 0.02$, $p=0.76$).

Discussion

It is known that humans are unable to generate highly random sequences^{26,27}. However, previous literature has demonstrated that humans are more random in competitive situations with feedback^{13,16}. Building on that, we set out to examine four research questions. First, we tested the extent to which human RSG in a competitive game situation depended on whether the subjects were specifically informed that they must be as random as possible to win the game. We did not find evidence that there was a difference in randomness when our subjects were aware versus unaware that they must be as random as possible. So, in other words, simply informing subject that

they must be as random as possible did not make them more random. Other factors could have contributed to the lack of difference between the Unaware and Aware groups. For example, Hyman and Jenkin²⁸ have demonstrated for a similar task that subjects' motivation to succeed and their belief in their ability to succeed in the task contribute to their performance. So, our participants may not have been particularly motivated to win in the game context, as their performance was not related to any compensation, unlike in other studies¹⁶. Participants in the Aware and Unaware groups might have also believed that the computer's prediction algorithm works well and that they are unlikely to be able to beat it.

Interestingly, in the post-experiment questionnaire subjects did not report differences in the extent to which they thought that they had used a strategy against the computer between the Aware and Unaware conditions. Hence, being made aware that they need to be random to beat the computer did not make subjects more likely to use a conscious strategy against the computer than when unaware they needed to be random. Similarly, we found no difference in subjects' perceptions of how much the computer predicted them between the Aware and Unaware conditions. In both cases they felt slightly predicted. So, again, being made aware that the computer was predicting them did not make subjects feel more predicted by the computer. Furthermore, the post-experiment questionnaire confirmed that participants' perception of the randomness of their sequences had very little correlation with the actual randomness of those sequences.

Notwithstanding, our second research question was how random can people be in a competitive game environment? We found that human RSG in such an environment, at least when they had to be as random as possible to win, is on average statistically indistinguishable from pseudo-RSG commonly used in modern programming languages. The extent to which computer pseudo-RSG are random is debatable. (Python's default pseudo-random generator relies on the Mersenne twister algorithm—also the default in Matlab, Stata, Ruby, R, Julia, Scilab, SageMath, etc.—which passes the Diehard tests of randomness and the Small Crush battery of the newer TestU01 collection, but fails on the more extensive Crush and Big Crush battery of the TestU01 collection;²⁹⁻³¹. However, our results are nevertheless interesting and set a high benchmark for the ability of humans to be random.

Third, we were wondering whether the improved RSG exhibited in a competitive game environment would transfer to a post-game environment—i.e., are people more random after the game than they were before the game? We did not find evidence for such a transfer of human RSG ability. Put differently, subjects were on average as random after the game as they were before the game. We also asked them explicitly about their RSG ability before and after the game. And we found that, at the group level at least, participants were generally not able to perceive when they were more or less random after the game. Thus, it appears that participants were able to transfer neither the implicit nor the explicit randomness that they exhibited during the game to their RSG after the game.

Last, it is known that humans underrepresent long runs of short sequences during RSG—i.e., that humans over alternate between items in the series—and this is one reason for the limited human RSG ability (Nickerson, 2002; Nickerson & Butler, 2009). We therefore wanted to test whether changes in RSG ability were related to changes in the length of generated runs. The trends we found for runs length generally reflected our level of randomness measure among the different experimental conditions. Importantly, we also found a direct correlation between the level of randomness and the average run length of the sequence among subjects. So, subjects that had longer runs were also more random. Our analysis showed that the run-length explained about one third of the variance in participants' degree of randomness. It would therefore be interesting to investigate what factors make up the remaining two-thirds of the variability in human randomness. For example, giving instructions in a more intuitive manner, or using a different method of eliciting random sequences may reveal different results. Wagenaar²⁷ hypothesized that the difference between verbal instruction and visual display of potential entries in the random sequence, and therefore the difference between an internally and externally represented entries, could affect random sequence generation. Ayton, et al.³² further suggests that some of the observed non-randomness may come from instructional bias from the experiment, by giving explicitly prohibited examples of unlikely or unacceptable sequences (e.g., telling them that the sequence "ABC" or meaningful words are unlikely for random sequences of letters). More concretely, Beach and Swensson³³ showed that specific instructions about the gambler's fallacy to ignore run dependency does not help lengthen participants' run lengths. They suggest that the method of eliciting random sequences may instead be more important. That might explain the results of Peterson and Ulehla³⁴, who found that people had a decreased run-length when the random sequence was generated by throwing dice, as compared to cards. People might expect shorter runs with a die than with cards.

Our results also relate to the "Network-Modulation Model" for randomness proposed by Jahanshahi and colleagues. The model suggests that the superior temporal cortex (STC) is involved in stereotyped responses (such as ordered counting) and that suppression of these habitual responses is necessary for RSG. Further, according to the model, this suppression is achieved via the left dorsolateral prefrontal cortex (DLPFC), which exerts an inhibitory influence on the STC. The model was constructed based on evidence from positron emission tomography, transcranial magnetic stimulation and lesion studies³⁵⁻³⁷. This model more recently gained support from electroencephalography studies as well^{38,39}. However, if participants can upregulate their DLPFC to exert enough inhibitory influence on the STC during game trials, why did they not transfer this ability to the post-game trials of our experiment, which they knew were coming and while instructed to be as random as possible? One possibility is that something in the game environment triggers the DLPFC inhibitory influence, but that such upregulation of the DLPFC is outside of human voluntary, conscious control. If DLPFC activity could be brought under more voluntary control—e.g., via neurofeedback⁴⁰—would up- or down-regulating left DLPFC activity then increase or decrease participants' level of randomness, respectively?

Regardless of the above, it remains unclear what makes our participants more random in the competitive game environment compared to the post-game part. There are at least two differences between the game and post-game trials: the competitive nature and the feedback in the former. A future study could test whether post-game randomness changes after a non-competitive game with constant feedback on one's level of randomness (e.g., via a continuously updated runs-test score), potentially teasing apart the effect of feedback from competition. Competition and feedback could increase motivation, which might further explain the difference in randomness (e.g., through the mechanism of attention). So, it would be good to test whether introducing a monetary reward above a certain randomness threshold, for example, could improve randomness (Hyman and Jenkin ²⁸ suggest it may). In addition, explicitly informing participants that randomness during the game part is higher than during the pre-game part might make them realize that they do better during the game and facilitate transferring of their increased randomness to the post-game part. What is more, a neuroimaging study could test whether the higher degree of randomness during the game part correlates with stronger inhibition of the STC by the DLPFC or whether other circuits also alter their activity, thereby increasing DLPFC activity or decreasing STC activity, in accordance with the Network-Modulation Model.

Recent results suggest that individual RSG ability could potentially be like a fingerprint—a unique biomarker of cognition ⁴¹⁻⁴³. RSG ability also appears to vary amongst different psychopathological populations ⁴⁴⁻⁴⁶. It appears, from our results, among others, that humans cannot easily and consciously modulate their randomness at will. This might explain the randomness-as-a-biomarker phenomenon, providing support for the use of RSG as a clinical diagnostic tool. For example, if RSG is driven by momentary neural noise, such noise patterns might be individually unique. It would be interesting to test whether these individual RSG patterns generalize from self-paced to competitive game environments.

Finally, human RSG appears related to volition. For example, our results that demonstrate that RSG in a competitive environment might be as random as a computer's pseudo-RNG mean that the upper limit on the arbitrariness of human action could be higher than previously realized. Also, previous work has demonstrated the existence of biases that may affect decisions especially when the decision alternatives are of similar value ⁹. Those biases were found in the DLPFC, which has also been implicated in RNG (the Network-Modulation Model). It would therefore be interesting to investigate the extent to which biases in decision making and RNG are supported by the same neural mechanisms, which may involve the DLPFC.

Author Contributions

GM and UM initiated the project. GM collected some of the data. SMW collected the rest of the data and analyzed all the data. SMW and UM wrote the main manuscript text. All authors reviewed the manuscript.

Acknowledgements

We thank Liad Mudrik for helpful discussions about this project. This project was also presented at the Society for Neuroscience 2018 conference, and we are grateful for comments received there that helped improve the manuscript.

Funding acknowledgement

This publication was made possible in part through the support of a joint grant from the John Templeton Foundation and the Fetzer Institute. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of the John Templeton Foundation or the Fetzer Institute.

References

- 1 Baddeley, A. Random generation and the executive control of working memory. *The Quarterly Journal of Experimental Psychology: Section A* **51**, 819-852 (1998).
- 2 Wierchoń, M., Gaillard, V., Asanowicz, D. & Cleeremans, A. Manipulating attentional load in sequence learning through random number generation. *Adv Cogn Psychol* **8**, 179-195, doi:10.2478/v10053-008-0114-0 (2012).
- 3 Brown, S. W. Timing and executive function: Bidirectional interference between concurrent temporal production and randomization tasks. *Memory & Cognition* **34**, 1464-1471, doi:10.3758/BF03195911 (2006).
- 4 Fried, I., Mukamel, R. & Kreiman, G. Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. *Neuron* **69**, 548-562, doi:10.1016/j.neuron.2010.11.045 (2011).
- 5 Libet, B., Gleason, C. A., Wright, E. W. & Pearl, D. K. Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain* **106 (Pt 3)**, 623-642 (1983).
- 6 Libet, B. Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences* **8**, 529-566, doi:10.1017/S0140525X00044903 (1985).
- 7 Soon, C. S., Brass, M., Heinze, H.-J. & Haynes, J.-D. Unconscious determinants of free decisions in the human brain. *Nature Neuroscience* **11**, 543-545, doi:10.1038/nn.2112 (2008).
- 8 Maoz, Yaffe, G., Koch, C. & Mudrik, L. Neural precursors of decisions that matter—an ERP study of deliberate and arbitrary choice. *eLife* **8**, e39787, doi:10.7554/eLife.39787 (2019).
- 9 Maoz, U. *et al.* Predeliberation activity in prefrontal cortex and striatum and the prediction of subsequent value judgment. *Frontiers in Neuroscience* **7**, 1-16 (2013).
- 10 Maoz, U., Ye, S., Ross, I., Mamelak, A. & Koch, C. in *Advances in Neural Information Processing Systems 25* (eds F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger) 872-880 (Curran Associates, Inc., 2012).
- 11 Roskies, A. L. Freedom, Neural Mechanism, and Consciousness. *Free Will and Consciousness: How Might They Work?*, 153 (2010).
- 12 Nickerson, R. S. The production and perception of randomness. *Psychological Review* **109**, 330-357, doi:10.1037/0033-295X.109.2.330 (2002).
- 13 Rapoport, A. & Budescu, D. V. Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology: General* **121**, 352-363, doi:10.1037/0096-3445.121.3.352 (1992).
- 14 Wagenaar, W. A. Randomness and randomizers: Maybe the problem is not so big. (1991).
- 15 Nickerson, R. S. & Butler, S. F. On producing random binary sequences. *The American Journal of Psychology* **122**, 141-151 (2009).
- 16 Budescu, D. V. & Rapoport, A. Subjective randomization in one-and two-person games. *J. Behav. Decis. Making* **7**, 261-278, doi:10.1002/bdm.3960070404 (1994).

- 17 Barraclough, D., Conroy, M. & Lee, D. Prefrontal cortex and decision making in a mixed-strategy game. *Nature Neuroscience* **7**, 404-410 (2004).
- 18 Maoz, U. *et al.* Predeliberation activity in prefrontal cortex and striatum and the prediction of subsequent value judgment. *Front Neurosci* **7**, 225 (2013).
- 19 Lempel, A. & Ziv, J. On the Complexity of Finite Sequences. *IEEE Transactions on Information Theory* **22**, 75-81, doi:10.1109/TIT.1976.1055501 (1976).
- 20 Ziv, J. & Lempel, A. Compression of individual sequences via variable-rate coding. *IEEE transactions on Information Theory* **24**, 530-536 (1978).
- 21 Hornero, R. *et al.* Variability, regularity, and complexity of time series generated by schizophrenic patients and control subjects. *IEEE Transactions on Biomedical Engineering* **53**, 210-218, doi:10.1109/TBME.2005.862547 (2006).
- 22 Rapp, P. E. Quantitative characterization of animal behavior following blast exposure. *Cogn Neurodyn* **1**, 287-293, doi:10.1007/s11571-007-9027-8 (2007).
- 23 Cohen, J. Statistical power analysis. *Curr Dir Psychol Sci* **1**, 98-101 (1992).
- 24 Rouder, J. N., Morey, R. D., Speckman, P. L. & Province, J. M. Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology* **56**, 356-374, doi:10.1016/j.jmp.2012.08.001 (2012).
- 25 Lee, M. D. & Wagenmakers, E.-J. *Bayesian cognitive modeling: A practical course*. (Cambridge University Press, 2013).
- 26 Mookherjee, D. & Sopher, B. Learning Behavior in an Experimental Matching Pennies Game. *Games and Economic Behavior* **7**, 62-91, doi:10.1006/game.1994.1037 (1994).
- 27 Wagenaar, W. A. Generation of random sequences by human subjects: A critical survey of literature. *Psychological Bulletin* **77**, 65-72, doi:10.1037/h0032060 (1972).
- 28 Hyman, R. & Jenkin, N. S. Involvement and Set as Determinants of Behavioral Stereotypy. *Psychol Rep* **2**, 131-146, doi:10.2466/pr0.1956.2.3.131 (1956).
- 29 Sexton, N. J. & Cooper, R. P. An architecturally constrained model of random number generation and its application to modeling the effect of generation rate. *Front Psychol* **5**, doi:10.3389/fpsyg.2014.00670 (2014).
- 30 Matsumoto, M. & Nishimura, T. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput. Simul.* **8**, 3-30, doi:10.1145/272991.272995 (1998).
- 31 Ecuyer, P. L. & Owen, A. B. *Monte Carlo and Quasi-Monte Carlo Methods 2008*. (Springer Science & Business Media, 2010).
- 32 Ayton, P., Hunt, A. J. & Wright, G. Psychological conceptions of randomness. *J. Behav. Decis. Making* **2**, 221-238, doi:10.1002/bdm.3960020403 (1989).
- 33 Beach, L. R. & Swensson, R. G. Instructions about randomness and run dependency in two-choice learning. *Journal of Experimental Psychology* **75**, 279-282, doi:10.1037/h0024979 (1967).
- 34 Peterson, C. R. & Ulehla, Z. J. Sequential patterns and maximizing. *Journal of Experimental Psychology* **69**, 1-4, doi:10.1037/h0021597 (1965).
- 35 Jahanshahi, M. & Dirnberger, G. The left dorsolateral prefrontal cortex and random generation of responses: studies with transcranial magnetic stimulation. *Neuropsychologia* **37**, 181-190 (1999).

- 36 Jahanshahi, M., Dirnberger, G., Fuller, R. & Frith, C. D. The Role of the Dorsolateral Prefrontal Cortex in Random Number Generation: A Study with Positron Emission Tomography. *NeuroImage* **12**, 713-725, doi:10.1006/nimg.2000.0647 (2000).
- 37 Jahanshahi, M. *et al.* Self-initiated versus externally triggered movements. I. An investigation using measurement of regional cerebral blood flow with PET and movement-related potentials in normal and Parkinson's disease subjects. *Brain* **118 (Pt 4)**, 913-933 (1995).
- 38 Joppich, G. *et al.* Brain potentials index executive functions during random number generation. *Neuroscience Research* **49**, 157-164, doi:10.1016/j.neures.2004.02.003 (2004).
- 39 Schneider, S., Joppich, G., Lugt, A. v. d., Däuper, J. & Münte, T. F. Brain potentials and self-paced random number generation in humans. *Neuroscience Letters* **367**, 51-55, doi:10.1016/j.neulet.2004.05.074 (2004).
- 40 Van den Boom, M. A., Jansma, J. M. & Ramsey, N. F. Rapid acquisition of dynamic control over DLPFC using real-time fMRI feedback. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology* **28**, 1194-1205, doi:10.1016/j.euroneuro.2018.08.508 (2018).
- 41 Schulz, M.-A., Baier, S., Böhme, B., Bzdok, D. & Witt, K. A Cognitive Fingerprint in Human Random Number Generation. (PsyArXiv, 2020).
- 42 Schulz, M.-A., Schmalbach, B., Brugger, P. & Witt, K. Analysing Humanly Generated Random Number Sequences: A Pattern-Based Approach. *PLOS ONE* **7**, e41531, doi:10.1371/journal.pone.0041531 (2012).
- 43 Jokar, E. & Mikaili, M. Assessment of Human Random Number Generation for Biometric Verification. *J Med Signals Sens* **2**, 82-87 (2012).
- 44 Horne, R. L., Evans, F. J. & Orne, M. T. Random Number Generation, Psychopathology, and Therapeutic Change. *Arch Gen Psychiatry* **39**, 680-683, doi:10.1001/archpsyc.1982.04290060042008 (1982).
- 45 Williams, M. A., Moss, S. A., Bradshaw, J. L. & Rinehart, N. J. Brief Report: Random Number Generation in Autism. *J Autism Dev Disord* **32**, 43-47, doi:10.1023/A:1017904207328 (2002).
- 46 Jahanshahi, M., Saleem, T., Ho, A. K., Dirnberger, G. & Fuller, R. Random number generation as an index of controlled processing. *Neuropsychology* **20**, 391-399, doi:10.1037/0894-4105.20.4.391 (2006).

Figure Legends

Figure 1. The progress of each trial in the experiment. Following the countdown, participants were required to press the appropriate key for R, P, or S within 500 ms of the onset of the Go signal; otherwise the trial was forfeit. Their selection was then presented on the screen. In game trials (on the right), their selection was accompanied by the computer's selection and by the game score.

Figure 2. Visualization of inclusion criterion 1. The solid black line indicates an absolute maximum divergence in the proportion of R, P or S of 0.15. Sequences above this inclusion criterion were subsequently discarded from analyses. Sequences for each experiment part (3 per participant) are sorted by proportion of R, P and S that diverge from 1/3—i.e., diverge from uniform distribution.

Figure 3. Visualization of inclusion criterion 2. Difference between human and computer final game scores are shown for the 3 experimental conditions. Participants who lost to the computer by more than 65 points (Human – Computer $\leq$ -65; solid, horizontal line) were discarded from subsequent analysis. Sequences are sorted by the final-score difference between human and computer.

Figure 4. LZC scores of human generated sequences. Violin plots of LZC scores by experimental part and condition are shown with the mean and 95% confidence intervals superimposed in black. The means are also indicated above each violin plot. The horizontal, solid red line indicates the mean LZC score of the 2.5th percentile of 1000 bootstrapped pseudo-random sequences, for baseline comparison. Values above this line are statistically indistinguishable from computer-generated pseudo-random sequences.

Figure 5. Average run-length scores of human generated sequences. Average run-length scores by experimental part and condition are indicated below each violin plot and in black inside each violin plot with 95% CI. The horizontal red line indicates the mean run-length of the 2.5th percentile of 1000 bootstrapped pseudo-random sequences.

Table Legends

Table 1. Within- and between-participants division of the experiment. Each participant carried out all three parts of the experiment (pre-game, game, and post-game). Different participant groups were randomized to different conditions in the game part of the experiment.

Figures

Figure 1

The progress of each trial in the experiment. Following the countdown, participants were required to press the appropriate key for R, P, or S within 500 ms of the onset of the Go signal; otherwise the trial was forfeit. Their selection was then presented on the screen. In game trials (on the right), their selection was accompanied by the computer's selection and by the game score.

CRITERIA 1: PROPORTION CUT OFF

Figure 2

Visualization of inclusion criterion 1. The solid black line indicates an absolute maximum divergence in the proportion of R, P or S of 0.15. Sequences above this inclusion criterion were subsequently discarded from analyses. Sequences for each experiment part (3 per participant) are sorted by proportion of R, P and S that diverge from 1/3—i.e., diverge from uniform distribution.

Figure 3

Visualization of inclusion criterion 2. Difference between human and computer final game scores are shown for the 3 experimental conditions. Participants who lost to the computer by more than 65 points (Human – Computer $\leq$ -65; solid, horizontal line) were discarded from subsequent analysis. Sequences are sorted by the final-score difference between human and computer.

Figure 4

LZC scores of human generated sequences. Violin plots of LZC scores by experimental part and condition are shown with the mean and 95% confidence intervals superimposed in black. The means are also indicated above each violin plot. The horizontal, solid red line indicates the mean LZC score of the 2.5th percentile of 1000 bootstrapped pseudo-random sequences, for baseline comparison. Values above this line are statistically indistinguishable from computer-generated pseudo-random sequences.

Figure 5

Average run-length scores of human generated sequences. Average run-length scores by experimental part and condition are indicated below each violin plot and in black inside each violin plot with 95% CI. The horizontal red line indicates the mean run-length of the 2.5th percentile of 1000 bootstrapped pseudo-random sequences.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SRSupplementaryWongMerholzMaozRandomness.docx](#)