

Idiosyncratic random behavior in a massive online sample

Tal Boger¹ and Sami Yousif²

¹ Johns Hopkins University, ² The Ohio State University

tboger1@jhu.edu, yousif.36@osu.edu

Abstract:

What factors govern and predict an individual's random behavior? Under many views, human random behavior arises from a grab bag of group-level heuristics (e.g., the representativeness heuristic, the law of small numbers, etc.). Contrary to this view, we present evidence that random behavior arises from highly idiosyncratic processes, borne out even with minimal amounts of data. Over 50,000 people generated random sequences of (a) coin flips, (b) dice rolls, and (c) locations in a 3-by-3 grid. Importantly, participants submitted only 8-12 responses in each task. This very small window minimizes many factors that might create individual differences in random behavior on their own, such as differences in working memory or attention. Yet, individuals demonstrated consistent behavior across these three different tasks, as reflected both by their sequence features and by a simple computational model of random behavior. Thus, these results indicate that, while some general heuristics guide random behavior, there also exist unique, idiosyncratic patterns that warrant attention — raising broader questions about how random behavior is realized in the human mind.

Keywords: randomness, heuristics and biases, individual differences

Data and materials: <https://osf.io/vs5fp/>

Version: October 18, 2026; in press, *Open Mind*.

Introduction

Among the most pervasive and well-documented biases in human cognition is our supposed inability to behave randomly. When asked to generate sequences of random coin flips, numbers, locations, and so on, humans fall prey to a variety of biases that lead to non-random behavior. For example, ask a person to generate a sequence of coin flips and they will rarely repeat responses on consecutive trials (i.e., they will avoid generating two heads or two tails in a row); ask a person to predict what will come next in a series of coin flips following 7 heads in a row and they will likely guess tails (i.e., the gambler's fallacy); and so on (for review, see Bar-Hillel & Wagenaar, 1991; Falk & Konold, 1997; Nickerson, 2002; Hahn & Warren, 2009).

Canonically, researchers believe such biases reflect a variety of heuristics. Perhaps the most famous example of this is the representativeness heuristic, wherein people judge the randomness of an event or predict its next outcome based on how well it matches their prototype of a random sequence (Kahneman & Tversky, 1972; Reimers et al., 2018; see also the law of small numbers, Tversky & Kahneman, 1971). A variety of later work expands upon this idea, proposing that judgments of local representativeness operate over a subject's short-term memory window (e.g., Kareev, 1992; Rapoport & Budescu, 1997; Kubovy & Gilden, 1991), and that perhaps the limits of working memory explain such biases (Hahn & Warren, 2009; Warren et al., 2018; see also Icard, 2021; Boger et al., 2027).

Although these biases reflect real and meaningful aspects of human random behavior, researchers often imply (and in some cases, explicitly say) that their underlying mechanisms are homogeneous across people. In other words, it is often assumed that random behavior arises from *group-level* heuristics. For example, Falk et al. (2009) take this for granted, writing that “[b]ecause of people's shared subjective-randomness image, human-generated patterns meant to be random may resemble each other more than truly random patterns do”; in a review of random behavior, Nickerson (2002) notes that individuals’ “performance, *as a group*, has been evaluated in terms of the expected properties of a random set of sequences” (emphasis added); and so on.

While the notion of “subjective randomness” often refers to group-level heuristics, recent evidence suggests that random behavior might be idiosyncratic — i.e., that different people show different signatures of random behavior. For example, Schulz et al. (2012) found individual differences in random behavior insofar as (a) a model fit to one person's random sequence performed better at predicting another random sequence from that same person than a model fit to a different person's random sequence, and (b) a model given two random sequences from one person and a third random sequence from a separate person correctly identifies the ‘foreign’ sequence at a high rate. Later work from Schulz et al. (2021) found that such patterns remained stable when re-tested after one week (see also Peters et al., 2007; Maes et al., 2011 for instances of test-retest or simple cross-task reliability in random behavior). More recently, Boger et al.

(2025) demonstrated idiosyncratic random behavior not only across a variety of tasks (e.g., random number generation, one-dimensional random location generation, and two-dimensional random location generation) but also over the span of one full year. Thus, whereas the predominant view in the literature assumes uniformity in random behavior, some recent results point to the possibility of individuality in random behavior. Of course, the existence of individual differences in random behavior is not, in itself, incompatible with the existence of population-level biases in such behavior. Nevertheless, it is worth appreciating that these biases vary considerably across individuals, such that the most effective model of random behavior is one that takes into account an individual's idiosyncratic 'parameters' for each of several distinct biases or heuristics.

Are different people random in different ways? How quickly do idiosyncratic patterns in random behavior emerge? And is there meaningful variability across a much broader cross-section of the population? To approach these questions, we analyzed a massive dataset containing individual random behavior across three tasks. In April 2022, The Pudding — a data journalism website — published an article analyzing a finding (from Gauvrit et al., 2017) about how random behavior varies across the lifespan. This article included a demo where people generate short sequences of coin flips, dice rolls, and locations in a 3-by-3 grid. Impressively, the article collected random sequences across three different tasks for over 50,000 people, making it an ideal dataset for testing idiosyncratic random behavior. Perhaps more importantly, for our purposes, these sequences were at most 12 items long in the case of coin flips, and only 10 items long in the other two tasks (Figure 1). Thus, the tasks present a marked departure from typical laboratory tasks, which often contain several hundred trials for a given participant. These very short sequences should (in theory) minimize individual differences in random behavior. After all, with only a few responses, there is little to analyze in any sequence, and even less to analyze in a fashion that might distinguish one person's random behavior from another's. Furthermore, long laboratory tasks (of the kind in Boger et al., 2025) invite possible individual differences unrelated to randomizing ability itself. For example, in tasks with hundreds of trials, differences in working memory or attention span may manifest as differences in randomization. The present work mitigates these concerns insofar as the short sequences tested here impose minimal memory demands. With these limited data, and with various other sources of individual differences minimized, are individual signatures of random behavior still traceable?

Here, we demonstrate that the answer is yes. Individual differences in random behavior arose not only in simple emergent properties of the sequence (e.g., the frequency of repetition), but also in a simple computational model of random behavior. This model generalized across tasks: For example, observing only a few random coin flips from a given participant was enough to predict idiosyncratic random behavior in the other two tasks. Thus, our work provides unusually large-scale evidence for individual variability in random behavior.

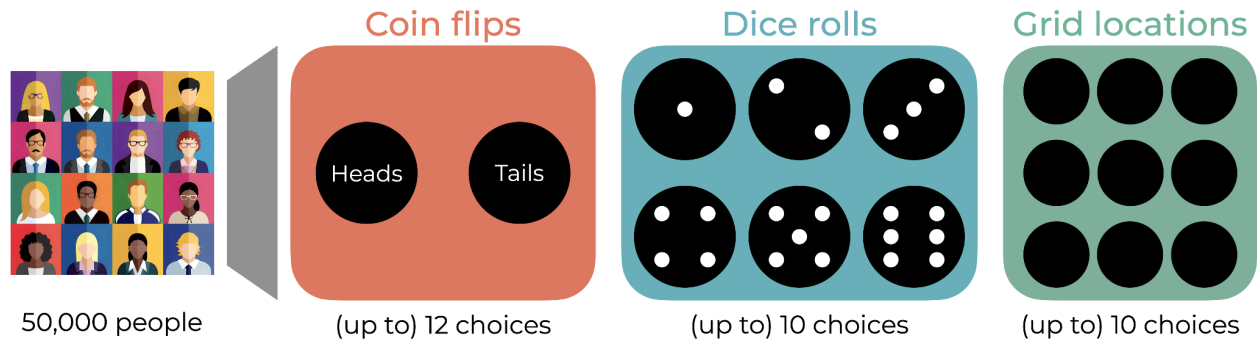


Figure 1: A massive dataset of cross-task randomization behavior. As part of an article on The Pudding, over 50,000 people produced random sequences of coin flips, dice rolls, and locations in a 3-by-3 grid. The resulting data provide a direct window into individual variability in random behavior.

Methods

Data collection

Our data come from demos in The Pudding’s April 2022 article on random behavior (<https://pudding.cool/2022/04/random/>). We thank Russell Samora (a writer for The Pudding and an author of that article) for sharing the raw sequences with us. The Pudding stopped logging people’s random sequences on August 1, 2022 — at which point they had already received over 50,000 responses. Given the nature of this ‘experiment’, the demos did not collect many variables that are often collected in typical experiments. For example, the demos did not collect any identifying information about the ‘participants’; the demos did not log response times for each sequence (or each item in the sequence); and so on. Rather, the data contain only (1) the subject’s self-reported age (this was collected for the purpose of age-prediction in The Pudding’s article), (2) their sequence of random coin flips, (3) their sequence of random dice rolls, and (4) their sequence of random 3-by-3 grid locations. (Given that it is unrelated to our core questions — and that there is no way to verify if this information is correct — we do not report anything related to the participants’ age. Preliminary analyses found no relationship between age and any of the relevant statistics reported below, in line with The Pudding article’s conclusion.)

In total, we received data from 52,708 people.

Preregistration and code availability

We did not preregister any aspects of this study. The data already existed at the time of the idea’s inception. However, our analysis methods and models mirror previous work (e.g., Boger et al., 2025; Boger et al., 2027; see also Angelike & Musch, 2025). All data and analysis scripts are available on our OSF repository (<https://osf.io/vs5fp/>); and readers can experience the ‘tasks’ for themselves on The Pudding (<https://pudding.cool/2022/04/random/>). (We highly recommend

doing this, not least because the article itself is interesting and beautifully done. Raw sequences are also available on The Pudding’s GitHub page for this article (<https://github.com/the-pudding/data/tree/master/random>.)

Procedure

In The Pudding’s article, participants provided a sequence of (up to) 12 random coin flips, 10 random dice rolls, and 10 random locations in a 3-by-3 grid. Each participant completed the tasks in this order. In the coin flipping task, participants had access to their most recent choice (i.e., heads or tails). In the dice rolls task, participants had access to a full history of their choices (i.e., their selection on all previous trials). Finally, in the grid task, participants had no access to any history of their previous choices. (After completing these three tasks, the article guessed the participant’s age, at which point the participant input their real age with a slider.)

Exclusion criteria

First, we excluded participants who did not submit sequences that were at least 8 items in all three tasks. This excluded 2,281 people, leaving us with 50,427 participants.

Next, we excluded participants who exhibited obviously “non-random” behavior in any of the three tasks. This excluded very few people, and all results hold regardless of these exclusions (see also Boger et al., 2025 for discussion of exclusions in studies of individual differences in random behavior). First, we excluded participants whose coin flips contained all heads or all tails; if a fair coin was flipped 8 times, this would occur 0.8% of the time. Next, we excluded participants whose dice rolls contained any given value 6 or more times; if a fair die was rolled 8 times, this would occur 0.3% of the time. Finally, we excluded participants whose grid sequence contained any given value 5 or more times; for 8 uniform draws from a set of 9 items, this would occur 0.6% of the time. We applied these same bounds to all sequences (i.e., even ones with more than 8 responses). If a participant was excluded in any task, we excluded their data from all tasks (as we wanted to examine cross-task reliability). In total, this excluded 1,044 additional participants (2.07%), leaving us with 49,383 participants total.

Computational model

To analyze our data, we fit a simple computational model to subjects’ sequences in the three tasks. Our model adapts the one proposed by Angelike & Musch (2025), which is itself a variation of a prior model (Boger et al., 2025). The model proposed by Angelike & Musch (2025) contains three parameters: (1) a ‘stay’ parameter, which adjusts the probability of repeating a choice, (2) a ‘distance’ parameter, which adjusts the probability of selecting choices that are nearby (or faraway) from the most recent choice in the response space, and (3) a

‘cycling’ parameter, which adjusts the probability of selecting a choice as a function of how recently it was last selected. In this work, we simplified this model and used only the distance parameter. We decided not to use the other two parameters for a variety of reasons. First, using a three-parameter model for sequences as short as 8 items causes instability in parameter estimates (and overfitting). Second, the parameters definitionally overlap in certain cases (e.g., a participant who repeats choices often will have a high ‘stay’ parameter and a low ‘distance’ parameter); longer sequences tease apart this overlap in ways that help distinguish the parameters from one another. Finally, the coin flipping task exacerbates both of these issues: With only two possible responses, every response is either a ‘stay’ with distance 0, or a ‘switch’ with distance 1. In other words, parameter estimates are especially erratic for the coin flips. Thus, we decided to use only the distance parameter here. (We believe other models might produce even stronger performance on the below metrics — this model is just one possible implementation.)

Our one-parameter model worked as follows. Using a nonlinear program solver for maximum likelihood estimation from the MATLAB optimization toolbox, we fit a model to each participant in each task. The model starts with uniform probabilities across all choices (the baseline to which we compare our full model). The model attempts to predict the next item in the sequence as a function of its ‘distance’ to the most recent choice (i.e., the distance between the most recent choice and a candidate next choice), starting from the second response (given that the first choice has no predecessor). For coin flips, distance is always 1 (alternating) or 0 (repeating). In the other two tasks, ‘distance’ means Euclidean distance between choices (e.g., the distance between the top-left and bottom-center locations would be $\sqrt{5}$).¹ The distance matrix is multiplied by the distance parameter, which is bounded [-1, 1]. This bound follows previous approaches (e.g., Boger et al., 2025; Angelike & Musch, 2025; Boger et al., 2027), and indeed most participants’ parameters fell within these bounds (i.e., very few experience compression because of these bounds, as the 5th percentile parameter value is -0.80, and the 95th percentile is 0.94). Positive values of the distance parameter indicate that the subject prefers to choose faraway options; negative values indicate a preference for nearby options; and a value of 0 indicates no preference (i.e., uniform randomness). Finally, a softmax function is applied to convert the predictions into a probability distribution for the next choice.

Results

Do participants display idiosyncratic random behavior, even in these short sequences? First, we examined this with regard to individual stability in emergent properties across tasks. Specifically, we examined whether a participant’s frequency of repeating a choice (i.e., selecting the same choice on consecutive trials) in one task predicted their frequency of repeating in another task;

¹ This applied to the dice task too, as the dice were presented in two rows of 3 options (i.e., one row of 1-3, and one row of 4-6). Thus, in terms of participant responses, the distance between 4 and 1 was the same as the distance between 2 and 1, making Euclidean distance an appropriate metric even here.

and whether the same was true for their average distance between choices (i.e., Euclidean distance between consecutive choices).

We performed 1,000 subsamples where, in each sample, we randomly selected 200 participants (approximately the sample size used in other, similar studies; e.g., Boger et al., 2025) and then computed the cross-task Pearson correlation for that subset. This provides a more stringent test for stability and ensures that any observed patterns hold not only in the full, 50,000-person sample but also in randomly selected slices of that sample (i.e., a sample size one might usually find in an online experiment). We applied this method to all tests below.

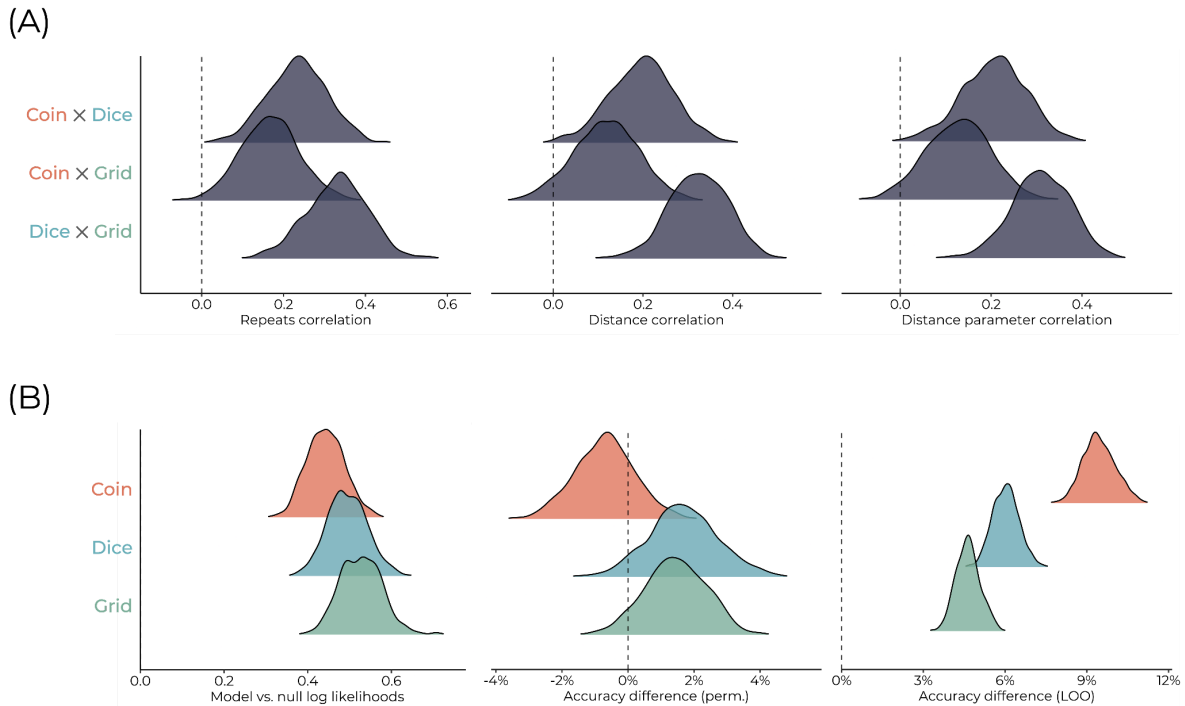
First, we observed within-subject, cross-task correlations in repeats; median $r(\text{coin flips, dice rolls}) = 0.23$, 95% subsampling interval = [0.09, 0.37], median $r(\text{coin flips, grid locations}) = 0.17$, 95% subsampling interval = [0.02, 0.31], median $r(\text{dice rolls, grid locations}) = 0.33$, 95% subsampling interval = [0.18, 0.47]. A similar pattern arose when considering distance between choices; median $r(\text{coin flips, dice rolls}) = 0.20$, 95% subsampling interval = [0.05, 0.34], median $r(\text{coin flips, grid locations}) = 0.12$, 95% subsampling interval = [-0.03, 0.25], median $r(\text{dice rolls, grid locations}) = 0.33$, 95% subsampling interval = [0.19, 0.45].² At first, these results might seem underwhelming, but they are quite strong, especially given (a) how much the tasks differ from one another and (b) how short the sequences are. For example, ‘distance’ in the coin flipping task is functionally equivalent to repeats; still, there is evidence for a relationship between distance in the coin flipping task and distance in the much richer grid locations task. (More generally, the relationships between the coin flipping task and either of the other two tasks may be more informative than the relationship between the dice rolling task and the grid location task, given that these latter two tasks are more similar in response format than the coin flipping task is to either.) Furthermore, with sequences as short as 8 choices, some participants might not repeat at all when generating dice rolls or grid locations — yet, we observed substantive correlations across these tasks.

Next, we fit the computational model described above. As expected given the correlation in average distance, we observed within-subject, cross-task consistency in the model’s distance parameter; median $r(\text{coin flips, dice rolls}) = 0.21$, 95% subsampling interval = [0.06, 0.33], median $r(\text{coin flips, grid locations}) = 0.13$, 95% subsampling interval = [-0.01, 0.26], median $r(\text{dice rolls, grid locations}) = 0.31$, 95% subsampling interval = [0.18, 0.43]³ (Figure 2A).

² Upon the request of a peer reviewer, we also include the full-sample tests, which show the same results. For repeats: $r(\text{coin flips, dice rolls}) = 0.24$, $p < 0.001$, 95% CI = [0.23, 0.25]; $r(\text{coin flips, grid locations}) = 0.17$, $p < 0.001$, 95% CI = [0.16, 0.17]; $r(\text{dice rolls, grid locations}) = 0.33$, $p < 0.001$, 95% CI = [0.33, 0.34]. For distance: $r(\text{coin flips, dice rolls}) = 0.20$, $p < 0.001$, 95% CI = [0.19, 0.21]; $r(\text{coin flips, grid locations}) = 0.12$, $p < 0.001$, 95% CI = [0.11, 0.13]; $r(\text{dice rolls, grid locations}) = 0.33$, $p < 0.001$, 95% CI = [0.32, 0.34].

³ As above, upon the request of a peer reviewer, we report the full-sample tests, which show the same results: $r(\text{coin flips, dice rolls}) = 0.21$, $p < 0.001$, 95% CI = [0.20, 0.21]; $r(\text{coin flips, grid locations}) = 0.13$, $p < 0.001$, 95% CI = [0.12, 0.14]; $r(\text{dice rolls, grid locations}) = 0.31$, $p < 0.001$, 95% CI = [0.31, 0.32].

The model outperformed a null model, which assigns equal probability to all choices in the response space. This was expected given that the null model is nested in the full model; yet we still provide these statistics as a sanity check. First, we found higher log likelihoods for our model as opposed to the uniform null model; coin flips median difference = 0.44, 95% subsampling interval = [0.36, 0.54], dice rolls median difference = 0.49, 95% subsampling interval = [0.41, 0.59], grid locations median difference = 0.53, 95% subsampling interval = [0.44, 0.63]. Furthermore, as a more important test of predictive performance, the fitted model numerically outperformed a model fit to permuted versions of the subjects' sequences in a next-token prediction task (for two of three cases), suggesting that even in very short sequences, the model broadly captures sequential dependencies between choices; coin flips median difference = -0.68%, 95% subsampling interval = [-2.46%, 1.11%], dice rolls median difference = 1.66%, 95% subsampling interval = [-0.39%, 3.84%], grid locations median difference = 1.42%, 95% subsampling interval = [-0.40%, 3.14%].⁴ Finally, in these within-task predictions, subjects displayed idiosyncratic tendencies, as reflected by the fact that the fitted model outperformed the mean accuracy obtained when predicting a subject's sequence using the parameters from each of 50 other randomly sampled subjects (i.e., akin to a leave-one-out baseline); coin flips median difference = 9.41%, 95% subsampling interval = [8.36%, 10.69%], dice rolls median difference = 6.03%, 95% subsampling interval = [5.13%, 6.99%], grid locations median difference = 4.63%, 95% subsampling interval = [3.81%, 5.52%] (Figure 2B).



⁴ In this analysis, the coin flipping model underperforms the permuted baseline. We hypothesize this may be because participants choose each option (heads or tails) with relatively equal frequencies, such that most shuffles result in similar (or perhaps stronger) sequential dependencies than the observed data. Nevertheless, our other analyses demonstrate that the coin flipping model successfully modeled some aspects of the participants' data.

Figure 2: Emergent properties and model fitting. (A) People’s tendency to repeat and select nearby (or faraway) choices was consistent across tasks. Furthermore, the model’s distance parameter was stable within-subject across these tasks. (B) The model successfully fit the subjects’ data. This is first reflected by differences in log likelihood — i.e., the fitted model had a higher log likelihood than a null uniform model. Furthermore, the fitted model was numerically better at next-token prediction than the same model fit to permuted versions of the subjects’ sequences in two of three cases (suggesting that the model captured sequential preferences in random behavior); and it was also better than the same model using a random selection of other participants’ parameters (suggesting that the behavior within-task differed across individuals). Graphs depict the relevant statistics when computed with the subsampling approach described in the main text.

But does the model also predict individual behavior *across* tasks? Our next set of analyses asked this question by comparing the model’s subject-specific, cross-task predictive power to the model’s subject-general, within-task predictive power. In other words, we used each subject’s parameters in one task to predict their choices in another task, and compared this to a similar leave-one-out baseline as described above. However, the leave-one-out baseline used other subjects’ parameters in the to-be-predicted task. For example, we compared the accuracy of the subject-specific coin flipping models in the dice rolling task to the accuracy of other subjects’ dice rolling parameters on the dice rolling task. Surprisingly, even here we observed evidence for idiosyncratic behavior; train dice test coin median difference = 2.14%, 95% subsampling interval = [0.11%, 4.00%], train grid test coin median difference = 1.40%, 95% subsampling interval = [-0.78%, 3.34%], train coin test dice median difference = 1.48%, 95% subsampling interval = [0.21%, 2.70%], train grid test dice median difference = 2.08%, 95% subsampling interval = [0.73%, 3.43%], train coin test grid median difference = 0.81%, 95% subsampling interval = [-0.19%, 1.80%], train dice test grid median difference = 1.66%, 95% subsampling interval = [0.57%, 2.66%] (Figure 3A).

We further tested this cross-task predictability with a forced-choice model identification task, inspired by the approach of Schulz et al. (2012). In our analysis, pairs of subjects — and their corresponding sequences within a given domain — were selected at random. Then, we asked the model to ‘identify’ which subject generated a sequence in a held-out domain. For example, after observing two subjects’ sequences of coin flips, could the model identify who generated a sequence of dice rolls? Identification accuracy followed a simple procedure wherein a ‘correct guess’ occurred when the model (trained on the observed domain’s data) assigned a higher log likelihood to the correct subject (in the held-out domain). Here too, we observed evidence for individual stability; given dice identify coin median accuracy = 54.5%, 95% subsampling interval = [47.5%, 61.5%], given grid identify coin median accuracy = 52.5%, 95% subsampling interval = [46.0%, 60.0%], given coin identify dice median accuracy = 53.0%, 95% subsampling interval = [47.0%, 60.0%], given grid identify dice median accuracy = 56.3%, 95% subsampling interval = [49.0%, 63.0%], given coin identify grid median accuracy = 53.0%, 95% subsampling interval = [46.0%, 59.5%], given dice identify grid median accuracy = 58.0%, 95% subsampling interval = [52.0%, 64.5%] (Figure 3B).

Relative to chance-level performance (of 50%), this may seem underwhelming at first. Indeed, many of these subsampling intervals — both in the forced-choice accuracy and in the cross-task prediction — include chance-level performance (though they would reach significance in a less stringent, full-sample test relative to chance). However, for the forced choice case, Schulz et al. (2012)’s performance provides a reasonable baseline; they reach 88% accuracy, but this is with 300-trial sequences, a richer response space (numbers 1-9), within-task comparisons, and a substantially more complex model. In other words, given the dramatically shorter sequences, smaller response spaces, cross-task comparisons, and simpler model, the above-chance performance we observe here is noteworthy. Together, these two analyses provide evidence for idiosyncratic random behavior in ways that extend across tasks.

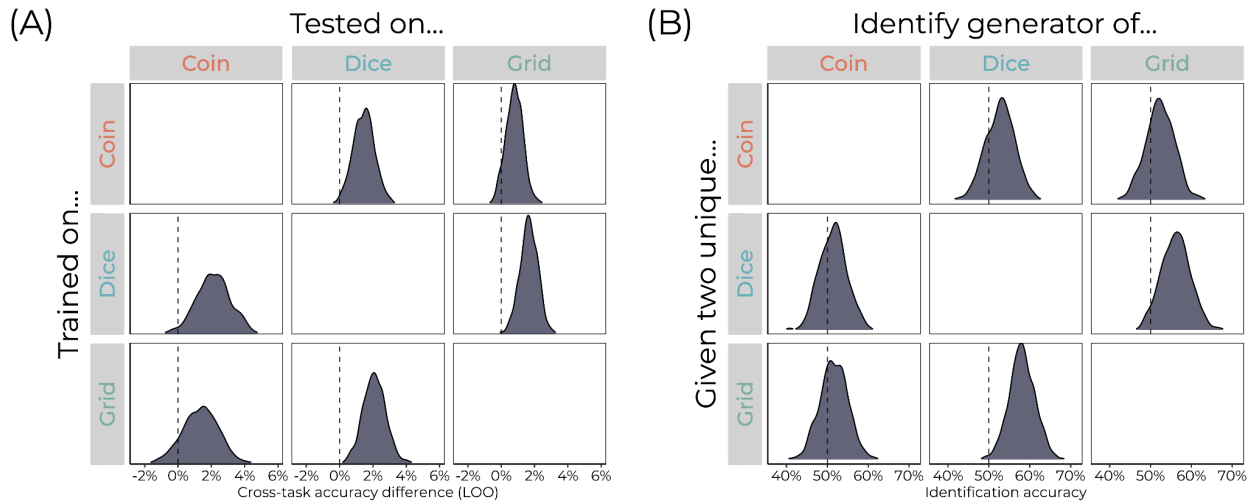


Figure 3: Cross-task model performance. (A) Using a participant’s parameters in one task to predict their behavior in another task (e.g., training on coin flips and testing on dice rolls) leads to better performance than using other participants’ parameters within the same task (e.g., using other participants’ dice roll parameters). (B) When presented with two unique sequences in one domain, the model identified the generator of a new sequence in a different domain at a numerically above-chance rate. Graphs depict the relevant statistics when computed with the subsampling approach described in the main text.

General discussion

Does random behavior emerge from homogeneous group-level heuristics or individual-level preferences? Whereas the extant literature largely assumes the former, some recent studies point to the latter (e.g., Schulz et al., 2012; Schulz et al., 2021; Boger et al., 2025). Here, we build on this possibility by demonstrating that idiosyncratic random behavior emerges (a) in a massive online sample of over 50,000 people and (b) with minimal input from each individual (both in terms of the number of trials, and in some cases, in terms of the number of possible responses).

The tasks examined here stand in stark contrast to the longer laboratory tasks typically used to study individual differences in random behavior (often containing hundreds of trials). Longer laboratory tasks open the door to a variety of individual differences in behavior unrelated to randomness itself. For example, perhaps some subjects simply have longer attention spans than others, leading to strong individual differences later in the tasks; perhaps some subjects have larger working memory capacities than others, allowing them to better monitor their own choices; and so on. In other words, prior results could in principle be explained by the fact that, when generating long sequences, participants vary in their ability to maintain focus over an extended period of time, rather than in their ability to randomize *per se*. In this way, the extremely short sequences tested here provide a minimal test for individual stability. Even with a substantial reduction in the factors that might induce individual differences in random behavior (i.e., differences in attention span or working memory), we found idiosyncratic behavior across tasks. Furthermore, our work presents data from three separate tasks, which allows us to explore the effect of randomness itself over and above task-specific factors.

Along these lines, the observed cross-task stability is especially notable given the fact that the three tasks differ not only in their format, but also in their feedback structure. In the coin flipping task, subjects had access to their most recent choice; in the dice rolling task, subjects had access to *all* of their previous choices; and in the grid locations task, subjects had access to *none* of their previous choices. This, again, presents an opportunity for the tasks to come apart for reasons orthogonal to random behavior itself. Yet, despite this difference, we found evidence for individuality in random behavior — suggesting that individual differences in random behavior are unlikely to be reducible to potential differences in working memory capacity or monitoring.

The results here should not be taken as definitive evidence *against* group-level heuristics at play in random behavior. Indeed, it seems likely that there are general biases and heuristics involved — in line with our proposal. Even so, these results suggest that there may be individual variability in the usage or application of those heuristics. In other words, it may not be that one either does or does not exhibit any given bias but instead that people vary continuously in the extent to which they exhibit that bias — and, thus, that the ideal way to understand and model random behavior may be to treat various known biases as ‘traits’ that vary in magnitude. Such a view would stand in contrast with the extant literature, which usually assumes a relative uniformity in the design of these heuristics.

Concluding remarks

Collectively, these findings invite a shift in how we think about “subjective randomness.” Rather than reflecting a single shared heuristic applied uniformly across people, subjective randomness may instead consist of stable but individually tuned biases — different people approximating randomness in systematically different ways. Even with only a small amount of input, and across

tasks with different structures and feedback, these signatures remain detectable. Randomness, in other words, is not merely something people broadly fail to produce; it is something they fail to produce in their own unique ways. Further understanding these idiosyncratic biases may provide insight into how random behavior is realized in the first place.

Acknowledgments: We would like to thank Dorsa Amir for helpful feedback and discussion. We would like to thank Russell Samora for sharing this data with us, and for creating the article that collected the data in the first place.

References

- Angelike, T., & Musch, J. (2025). An improved modeling approach to investigate biases in human random number generation. *PLOS ONE*, 20(5), e0324870.
- Bar-Hillel, M., & Wagenaar, W. A. (1991). The perception of randomness. *Advances in Applied Mathematics*, 12(4), 428-454.
- Boger, T., Yousif, S. R., McDougale, S. D., & Rutledge, R. B. (2025). Random behavior is stable across tasks and time. *Journal of Experimental Psychology: General*, 154(6), 1571-1582.
- Boger, T., Amir, D., & Yousif, S. R. (2027). Near-random behavior in strategic games. *Cognition*, 278, 106737.
- Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review*, 104(2), 301-318.
- Falk, R., Falk, R., & Ayton, P. (2009). Subjective patterns of randomness and choice: Some consequences of collective responses. *Journal of Experimental Psychology: Human Perception and Performance*, 35(1), 203-224.
- 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), e1005408.
- Hahn, U., & Warren, P. A. (2009). Perceptions of randomness: why three heads are better than four. *Psychological Review*, 116(2), 454-461.
- Icard, T. (2021). Why be random? *Mind*, 130(517), 111-139.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3(3), 430-454.
- Kareev, Y. (1992). Not that bad after all: Generation of random sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 18(4), 1189-1194.
- Kubovy, M., & Gilden, D. (1991). Apparent randomness is not always the complement of apparent order. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure: Essays in honor of Wendell R. Garner* (pp. 115-127).

- Maes, J. H., Eling, P. A., Reelick, M. F., & Kessels, R. P. (2011). Assessing executive functioning: On the validity, reliability, and sensitivity of a click/point random number generation task in healthy adults and patients with cognitive decline. *Journal of Clinical and Experimental Neuropsychology*, 33(3), 366-378.
- Nickerson, R. S. (2002). The production and perception of randomness. *Psychological Review*, 109(2), 330-357.
- Peters, M., Giesbrecht, T., Jelicic, M., & Merckelbach, H. (2007). The random number generation task: Psychometric properties and normative data of an executive function task in a mixed sample. *Journal of the International Neuropsychological Society*, 13(4), 626-634.
- Rapoport, A., & Budescu, D. V. (1997). Randomization in individual choice behavior. *Psychological Review*, 104(3), 603-617.
- Reimers, S., Donkin, C., & Le Pelley, M. E. (2018). Perceptions of randomness in binary sequences: Normative, heuristic, or both? *Cognition*, 172, 11-25.
- Schulz, M. A., Schmalbach, B., Brugger, P., & Witt, K. (2012). Analysing humanly generated random number sequences: a pattern-based approach. *PLOS ONE*, 7(7), e41531.
- Schulz, M. A., Baier, S., Timmermann, B., Bzdok, D., & Witt, K. (2021). A cognitive fingerprint in human random number generation. *Scientific Reports*, 11(1), 20217.
- Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76(2), 105-110.
- 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.