Supplemental Information

This document provides supplementary analyses to "Random behavior is stable across tasks and time" by Boger, Yousif, McDougle, and Rutledge. Here, we perform two sets of additional analyses. First, we confirm that our results hold even when we do not exclude any participants. Second, we compare our results to a second-order Markov chain baseline.

Results on full data

We re-ran all our models and analyses on our full dataset — without exclusions — to confirm that our results hold regardless of our exclusions. This entailed running exactly the analyses and models reported in our *Results*. Thus, we present tables of all those results, and compare them to the results on the cleaned data ("Excluded data", as reported in our main text).

Experiment 1

Cross-task correlations of emergent properties (*r* values):

	Un-excluded data	Excluded data
Number of repeats	0.69 ($p < 0.001$)	0.42 ($p < 0.001$)
Average numerical distance	0.66 ($p < 0.001$)	0.54 ($p < 0.001$)
Number of direction switches	0.70 ($p < 0.001$)	0.38 ($p < 0.001$)

Model parameter correlations across tasks (*r* values):

	Un-excluded data	Excluded data
Stay parameter	0.51 ($p < 0.001$)	0.28 ($p < 0.001$)
Side-switch parameter	0.61 ($p < 0.001$)	0.45 ($p < 0.001$)
Direction-switch parameter	0.49 ($p < 0.001$)	0.33 ($p < 0.001$)

Performance relative to Markov chain in generating sequences that match true sequences in terms of emergent properties (# of simulations where model outperforms Markov chain / 100):

	Un-excluded data (number model)	Un-excluded data (1D model)	Excluded data (number model)	Excluded data (1D model)
Average numerical distance	100/100	100/100	100/100	100/100
Number of direction switches	100/100	100/100	100/100	100/100

Top-3 accuracy in predicting next choices within-tasks (p value is from a Wilcoxon signed-rank test comparing the model's top-3 accuracy to the permutation baseline):

	Un-excluded data (number model)	Un-excluded data (1D model)	Excluded data (number model)	Excluded data (1D model)
Top-3 accuracy	44.2% ($p < 0.001$)	46.4% ($p < 0.001$)	41.8% ($p < 0.001$)	43.7% ($p < 0.001$)
Permutation baseline	34.8%	33.5%	33.8%	33.2%

Top-3 accuracy in predicting next choices across-tasks (p value is from a Wilcoxon signed-rank test comparing the model's top-3 accuracy to the leave-one-out baseline):

	Un-excluded data (number model predicting 1D data)	Un-excluded data (1D model predicting number data)	Excluded data (number model predicting 1D data)	Excluded data (1D model predicting number data)
Top-3 accuracy	39.2% ($p < 0.001$)	39.5% ($p < 0.001$)	37.8% ($p < 0.05$)	38.2% ($p < 0.001$)
Leave-one-out baseline	32.1%	33.4%	36.6%	34.0%

<u>Experiment 2</u>

Cross-task correlations of emergent properties (r values):

	Un-excluded data	Excluded data
Number of repeats	0.66 ($p < 0.001$)	0.51 ($p < 0.001$)
Average numerical distance	0.64 ($p < 0.001$)	0.44 ($p < 0.001$)
Number of direction switches	0.58 ($p < 0.001$)	0.32 ($p < 0.001$)

Model parameter correlations across tasks (*r* values):

	Un-excluded data	Excluded data
Stay parameter	0.60 ($p < 0.001$)	0.48 ($p < 0.001$)
Side-switch parameter	0.58 ($p < 0.001$)	0.41 ($p < 0.001$)

Performance relative to Markov chain in generating sequences that match true sequences in terms of emergent properties (# of simulations where model outperforms Markov chain / 100):

	Un-excluded data (number model)	Un-excluded data (2D model)	Excluded data (number model)	Excluded data (2D model)
Average numerical distance	100/100	100/100	100/100	100/100
Number of direction switches	100/100	100/100	100/100	100/100

Top-3 accuracy in predicting next choices within-tasks (*p* value is from a Wilcoxon signed-rank test comparing the model's top-3 accuracy to the permutation baseline):

	Un-excluded data (number model)	Un-excluded data (2D model)	Excluded data (number model)	Excluded data (2D model)
Top-3 accuracy	43.1% ($p < 0.001$)	44.1% ($p < 0.001$)	41.9% ($p < 0.001$)	41.7% ($p < 0.001$)
Permutation baseline	33.8%	33.8%	33.6%	33.2%

Top-3 accuracy in predicting next choices across-tasks (*p* value is from a Wilcoxon signed-rank test comparing the model's top-3 accuracy to the leave-one-out baseline):

	Un-excluded data (number model predicting 2D data)	Un-excluded data (2D model predicting number data)	Excluded data (number model predicting 2D data)	Excluded data (2D model predicting number data)
Top-3 accuracy	38.2% ($p = 0.05$)	37.8% ($p < 0.001$)	36.0% ($p < 0.05$)	36.8% ($p < 0.001$)
Leave-one-out baseline	38.8%	34.2%	38.1%	33.2%

Experiment 3

Within-task, cross-time correlations of emergent properties (*r* values):

	Un-excluded data	Excluded data
Number of repeats (number task)	0.67 ($p < 0.001$)	0.42 ($p < 0.01$)
Average numerical distance (number task)	0.70 ($p < 0.001$)	0.56 ($p < 0.001$)
Number of direction switches (number task)	0.70 ($p < 0.001$)	0.51 ($p < 0.001$)
Number of repeats (2D task)	0.60 ($p < 0.001$)	0.42 ($p < 0.01$)
Average numerical distance (2D task)	0.63 ($p < 0.001$)	0.57 ($p < 0.001$)
Number of direction switches (2D task)	0.68 ($p < 0.001$)	0.61 ($p < 0.001$)

Model parameter correlations across time (*r* values):

	Un-excluded data	Excluded data
Stay parameter (number model)	0.55 ($p < 0.001$)	0.33 ($p = 0.01$)
Side-switch parameter (number model)	0.66 ($p < 0.001$)	0.66 ($p < 0.001$)
Direction-switch parameter (number model)	0.57 ($p < 0.001$)	0.48 ($p < 0.001$)
Stay parameter (2D model)	0.60 ($p < 0.001$)	0.38 ($p < 0.01$)
Side-switch parameter (2D model)	0.60 ($p < 0.001$)	0.65 ($p < 0.001$)

Performance relative to Markov chain in generating sequences that match true sequences in terms of emergent properties across time (# of simulations where model outperforms Markov chain / 100):

	Un-excluded data (number model)	Un-excluded data (2D model)	Excluded data (number model)	Excluded data (2D model)
Average numerical distance	98/100	99/100	66/100	99/100
Number of direction switches	100/100	100/100	99/100	93/100

Top-3 accuracy in predicting next choices within-tasks, across-time (p value is from a Wilcoxon signed-rank test comparing the model's top-3 accuracy to the leave-one-out baseline):

	Un-excluded data (number model)	Un-excluded data (2D model)	Excluded data (number model)	Excluded data (2D model)
Top-3 accuracy	39.5% ($p = 0.01$)	41.0% ($p < 0.001$)	37.9% ($p < 0.05$)	38.8% ($p < 0.05$)
Leave-one-out baseline	34.8%	30.9%	36.4%	35.2%

Top-3 accuracy in predicting next choices across tasks and time (p value is from a Wilcoxon signed-rank test comparing the model's top-3 accuracy to the leave-one-out baseline):

	Un-excluded data (number model predicting 2D data)	Un-excluded data (2D model predicting number data)	Excluded data (number model predicting 2D data)	Excluded data (2D model predicting number data)
Top-3 accuracy	36.8% ($p = 0.01$)	37.5% ($p = 0.48$)	36.1% ($p = 0.19$)	36.6% ($p = 0.76$)
Leave-one-out baseline	31.0%	34.6%	34.8%	36.9%

Results relative to second-order Markov chain baseline

Throughout our paper, we compare our model to a simple Markov chain baseline. A Markov chain serves as a reasonable baseline because it is (definitionally) a random process, and it captures some aspects of participants' behavior without needing any parameters. However, some tendencies in random generation cannot be captured by a simple Markov chain that considers only first-order transition probabilities. For example, perhaps direction switches can be easily captured without a parameter if one looks at second-order transition probabilities. Thus, we conducted additional analyses in which we compared our model to a second-order Markov chain.

The Markov chain baseline is only used for comparing correlations between model-simulated sequences and real participant sequences. So, we present the results of our full model (on post-exclusion data) relative to a first-order Markov chain (i.e., as reported in the main text of the paper), and relative to a second-order Markov chain (as discussed here) for each experiment. Note that, as an additional sanity check, we compared the first-order and second-order Markov chains under various model fit metrics like AIC. We found the second-order Markov chain was a better fit than the first-order Markov chain, which is what we should expect; this shows that the second-order Markov chain indeed captures more information than the first-order Markov chain. However, the key comparison is still whether our model outperforms this stronger baseline.

Experiment 1

Model performance relative to first-order and second-order Markov chains in generating sequences that match true sequences in terms of emergent properties (# of simulations where model outperforms given baseline / 100):

	Number model (relative to first-order Markov chain)	Number model (relative to second-order Markov chain)	1D model (relative to first-order Markov chain)	1D model (relative to second-order Markov chain)
Average numerical distance	100/100	100/100	100/100	100/100
Number of direction switches	100/100	100/100	100/100	100/100

Experiment 2

Model performance relative to first-order and second-order Markov chains in generating sequences that match true sequences in terms of emergent properties (# of simulations where model outperforms given baseline / 100):

	Number model (relative to first-order Markov chain)	Number model (relative to second-order Markov chain)	2D model (relative to first-order Markov chain)	2D model (relative to second-order Markov chain)
Average numerical distance	100/100	100/100	100/100	100/100
Number of direction switches	100/100	100/100	100/100	100/100

Experiment 3

Performance relative to Markov chain in generating sequences that match true sequences in terms of emergent properties across time (# of simulations where model outperforms given baseline / 100):

	Number model (relative to first-order Markov chain)	Number model (relative to second-order Markov chain)	2D model (relative to first-order Markov chain)	2D model (relative to second-order Markov chain)
Average numerical distance	66/100	61/100	99/100	98/100
Number of direction switches	99/100	90/100	93/100	84/100