Supplementary Materials

Here we report the extended results for the experiments. Specifically, we report the discrete bimodal conditions in Experiments 1-3 separately, and each individual condition in Experiments 4 and 5. We aggregate across these conditions in the main text as we found no difference between them.

Detailed analysis code and individual histograms for each participant can be found on the Open Science Framework and attached in the Supplementary Materials (https://osf.io/wc35k/?view_only=56ea5099e5c14ca8a33d240bd259376b).

Experiments 1-3

In the main paper, we plotted participants aggregate data across Experiments 1 to 3 for the discrete bimodal distribution. Here we present them separately: Figure S1 shows the histograms for each of the experiments with a discrete condition and Figure S2 shows the individual CDFs. Participants showed consistent learning across the three experiments and there was no discernible effect of the instructional manipulations.

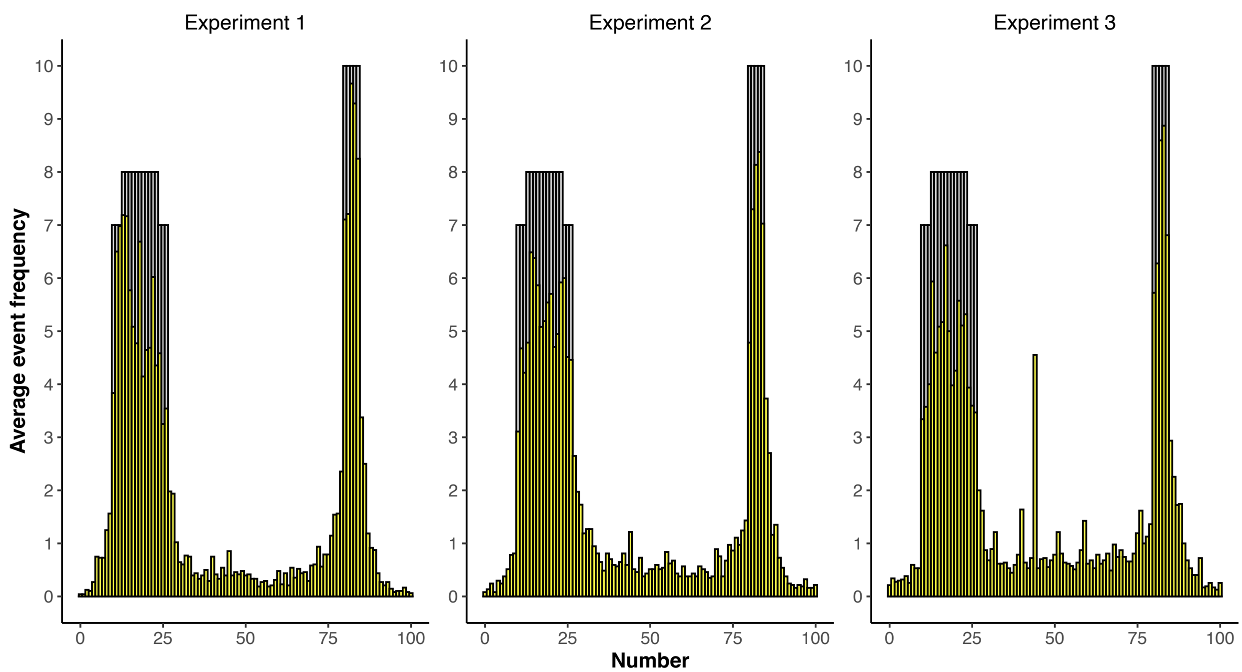

Figure S1. Normalised histogram of the average distribution produced by participants (yellow) and the target distribution (grey) for the discrete bimodal distribution in Experiments 1 to 3. The average event frequency is calculated by multiplying the density by the number of trials.

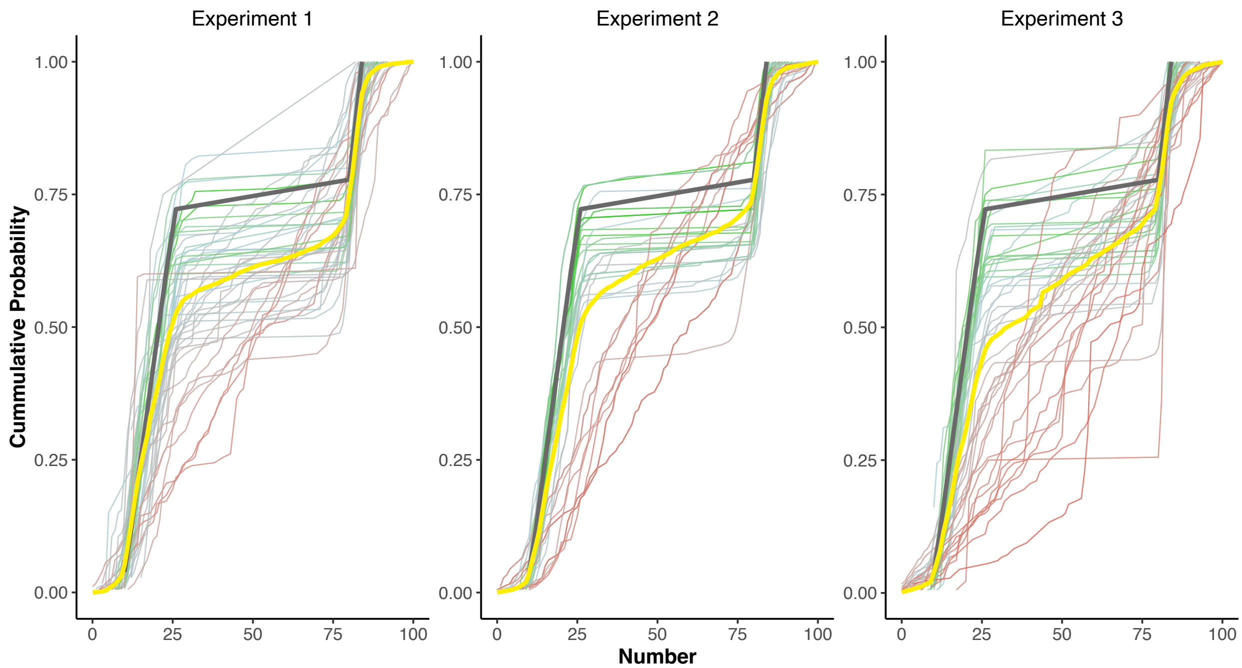

Figure S2. Cumulative density plots for the discrete bimodal distribution in Experiments 1 to 3.

The KS values were similar across Experiments 1-3 (means and standard errors for these experiments respectively were: 0.24 ± 0.016; 0.25 ± 0.023; 0.30 ± 0.028). A linear regression indicated that the experimental conditions had no effect on the KS values ($F(2, 129) = 1.69$, $p = .19$).

Experiment 4

In Experiment 4 we used colour cues and incentives to aid participants learning. For colour cues, in the learning phase of Experiment 4, the modes and noise were presented in two different colours (orange and blue counterbalanced), such that the modes (i.e., numbers between 10 and 26 and numbers between 80 and 84) would appear in one colour and all other numbers would appear in the other colour. We then varied which numbers were rewarded across four conditions: a) the modes, b) the noise c) both the modes and the noise, and d) nothing. Figure S3 shows the average responses for participants in each of the reward conditions, and the individual CDFs across conditions are shown in Figure S4. Learning across conditions was relatively stable.
The KS values were similar across the four reward conditions: modes rewarded [0.17 ± 0.007]; noise rewarded [0.17 ± 0.007]; modes and noise rewarded [0.17 ± 0.009]; no rewards [0.19 ± 0.011]. A linear regression indicated that there was no effect of reward condition on the KS values ($F(3, 203) = 1.30$, $p = .28$).

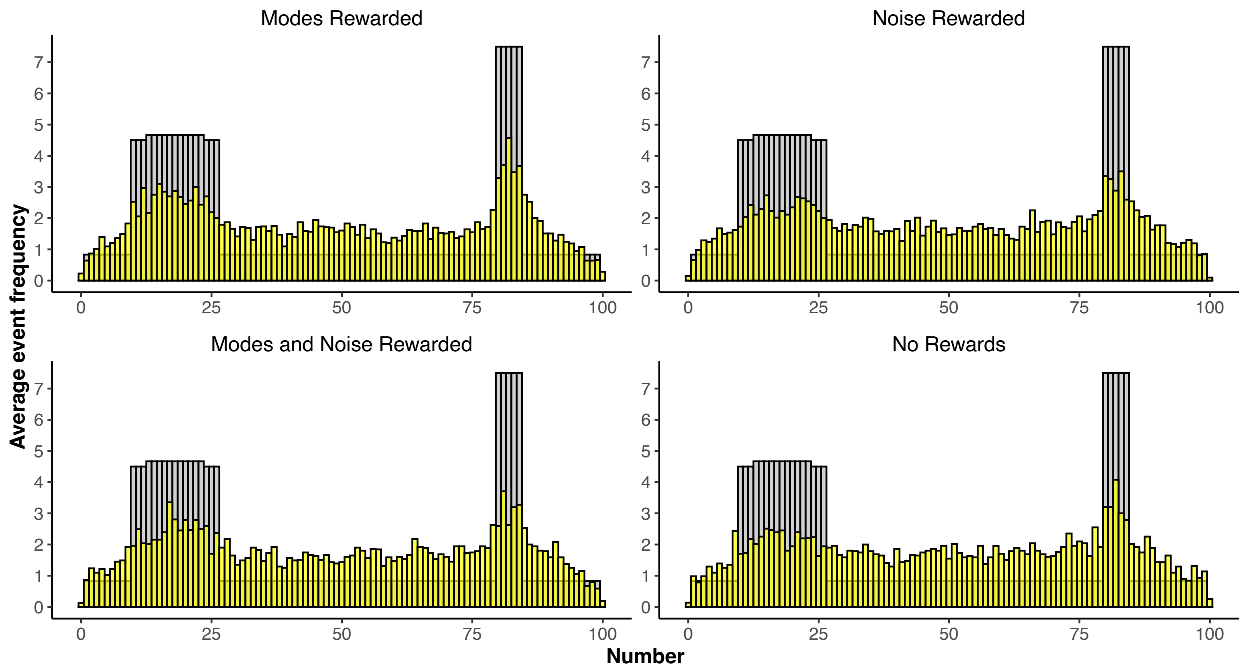

Figure S3. Normalised histogram of the average distribution produced by participants (yellow) and the target distribution (grey) across each of the reward conditions in Experiment 4.

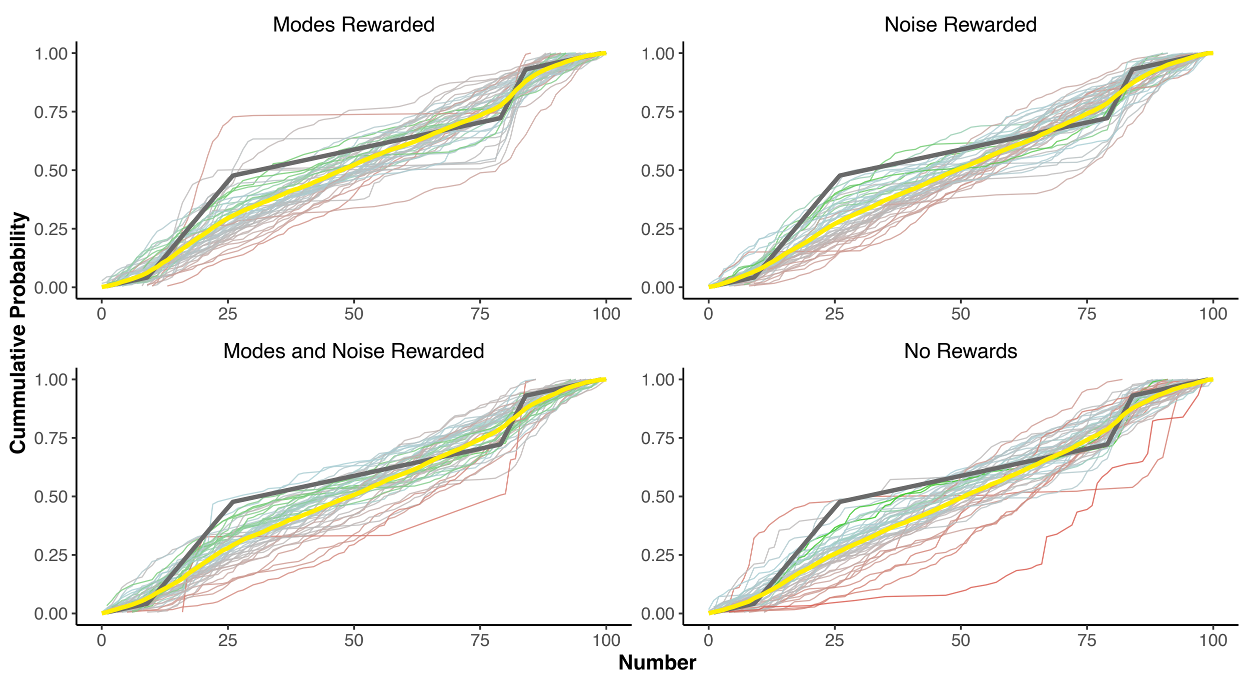

Figure S4. Cumulative density plots for each of the reward conditions in Experiment 4.

Experiment 5

In Experiment 5, we used colour to differentiate between the lower and the upper mode as well, such that the lower mode, the upper mode, and noise were all presented in different colours (orange, blue, and grey). There were two reward conditions: In the rewarded condition participants earnt points when they saw a modal number but did not earn any points in the no rewards condition. Normalised density plots (Figure S5) and individual CDFs (Figure S6) show that participants learn the distribution equally well in both conditions. The KS values were similar across the two conditions: rewards [0.16 $\pm$ 0.012] and no rewards [0.15 $\pm$ 0.010]. A linear regression indicated that there was no effect of reward condition on the KS values ($F(1, 92) = 0.49$, $p = .49$).

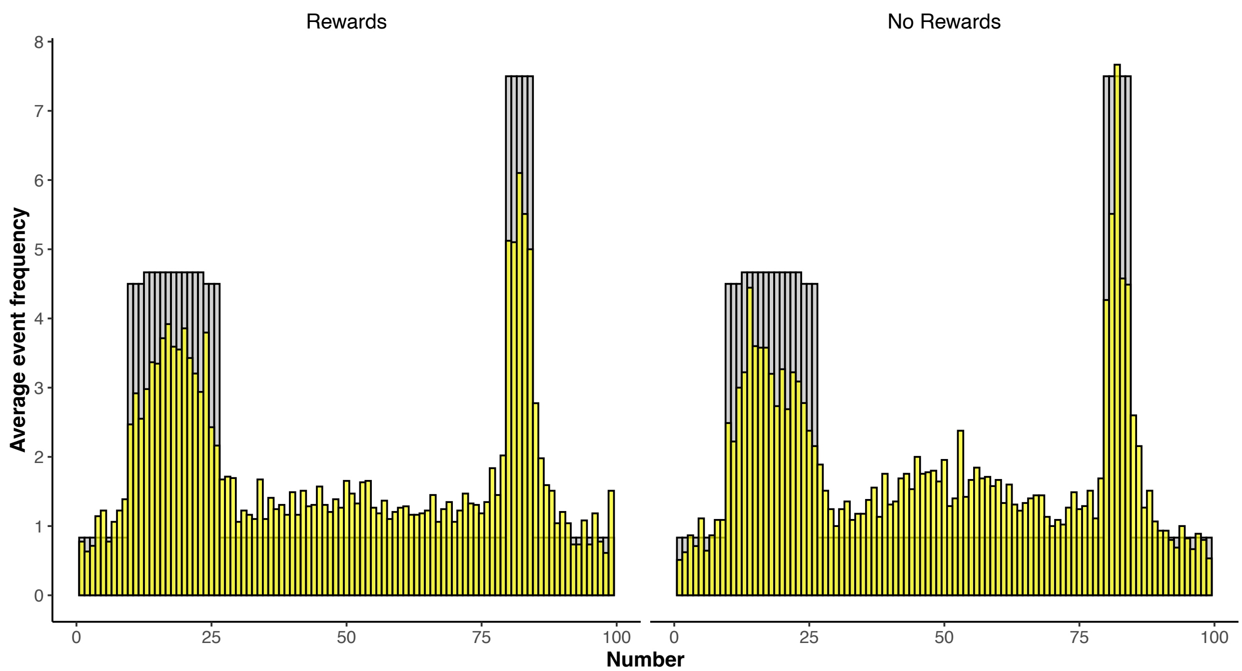

Figure S5. Normalised histogram of the average distribution produced by participants (yellow) and the target distribution (grey) across each of the reward conditions in Experiment 5.

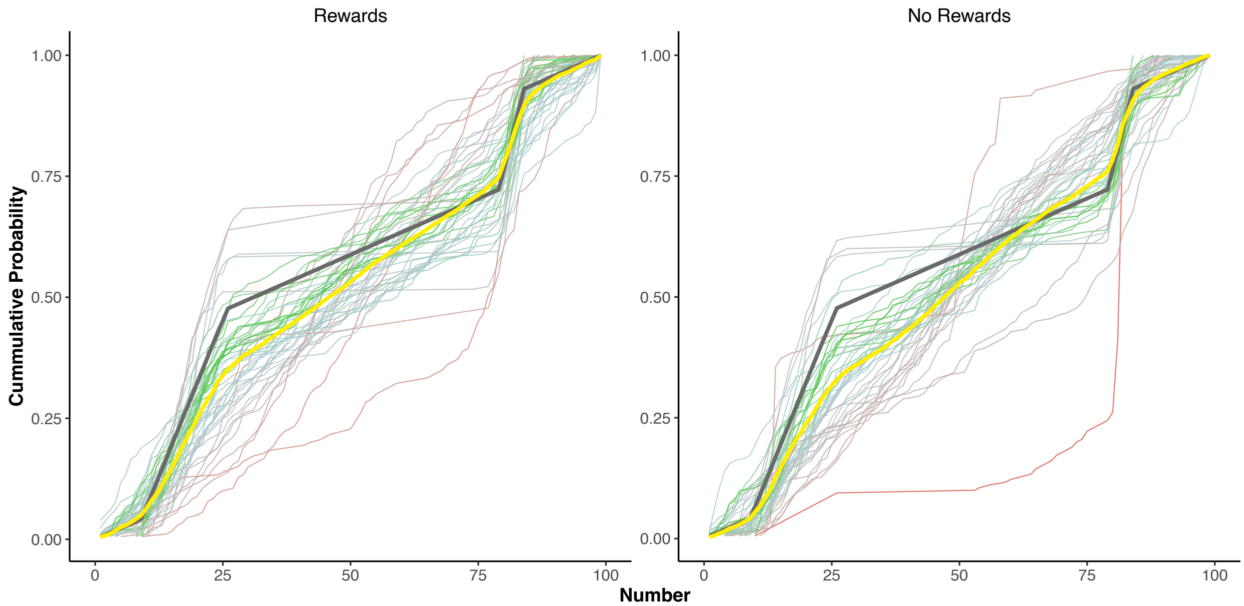

Figure S6. Cumulative density plots for each of the reward conditions in Experiment 5.

Anderson-Darling Test

As a robustness test we also conducted our analysis using the **Anderson-Darling** (AD) test. Figure S7 plots the AD test for each participant and their summary statistics for the noisy bimodal condition of each experiment. We observed similar AS statistics in Experiments 1-4 (means and standard errors for these experiments respectively were: 4344.1 ± 701.7; 4050.9 ± 708.6; 3855.1 ± 552.5 ; 3226.2 ± 207.5) but much lower values in Experiment 5 (2949 ± 349.3). Consistent with our visual inspection of the data, this indicates that participants in Experiment 5 were more accurately reproducing the underlying distribution than participants in Experiments 1-4.

Inferential statistical analyses were consistent with this interpretation. A linear regression indicated that a model accounting for the effect of experiment was not significantly better compared to the intercept-only model ($F(4, 442) = 1.86$, $p = 0.117$). The beta coefficients indicate that, compared to Experiment 1 (reference group), participants had significantly lower AD values in Experiment 5 ($\beta_{Exp5} = -1395.5$, $p = .029$) and marginally significant in Experiment 4 ($\beta_{Exp4} = -.1117.9$, $p = .052$), but not in the other experiments ($\beta_{Exp2} = 293.2$, $p = 0.69$; $\beta_{Exp3} = -489.0$, $p = 0.51$).

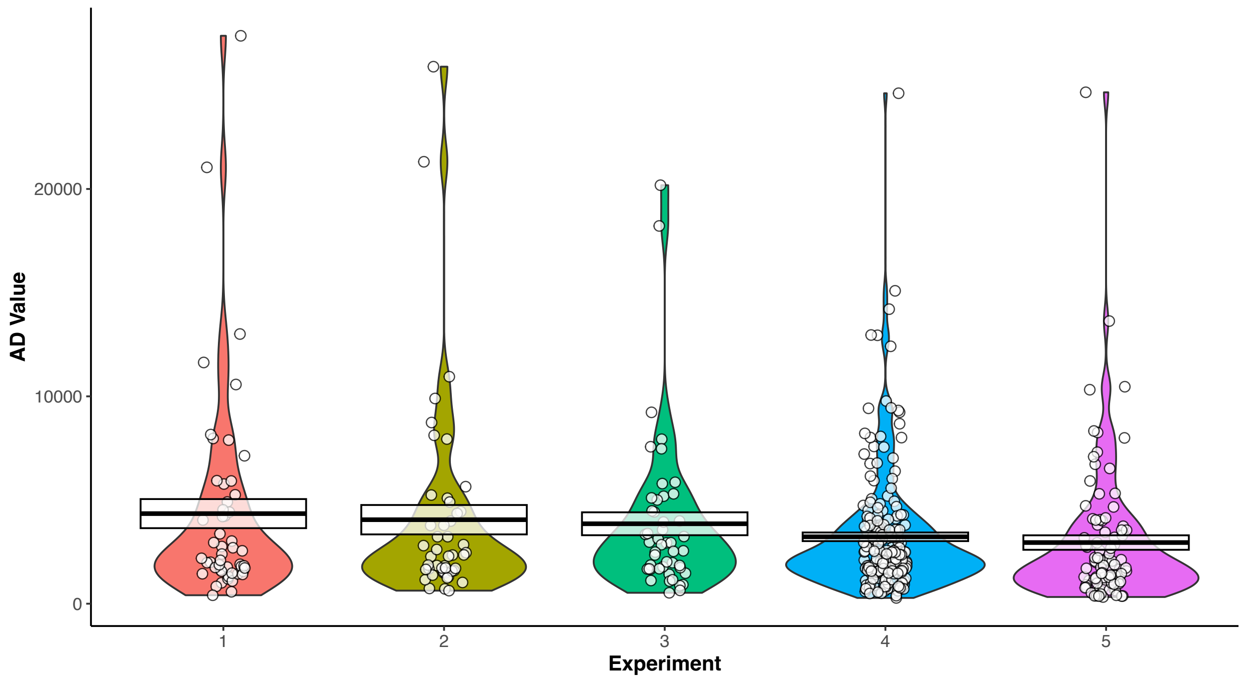

Figure S7. Anderson-Darling (AD) test statistics comparing the distribution each participant produced to the target distribution across the noisy bimodal condition of the five experiments. Smaller the AD test values indicate a smaller difference between distributions (and therefore better learning). Light grey dots represent individual AD test values for each participant. Boxplots indicate the mean and the standard error for the noisy bimodal condition of each experiment.

Sequential Analysis

As per the instructions used by Tran et al. (2017) participants were instructed to "Try to produce a sequence of numbers which is as much like the original sequence as you can make it". We therefore wanted to see if there was a correlation between the target and generated sequences. We observed similar Pearson correlation statistics (close to zero) in Experiments 1-5 (means and standard errors for these experiments respectively were: 0.0059 ± 0.012; 0.0058 ± 0.012; -0.0032 ± 0.0089; -0.0028 ± 0.0049; 0.00038 ± 0.0087.

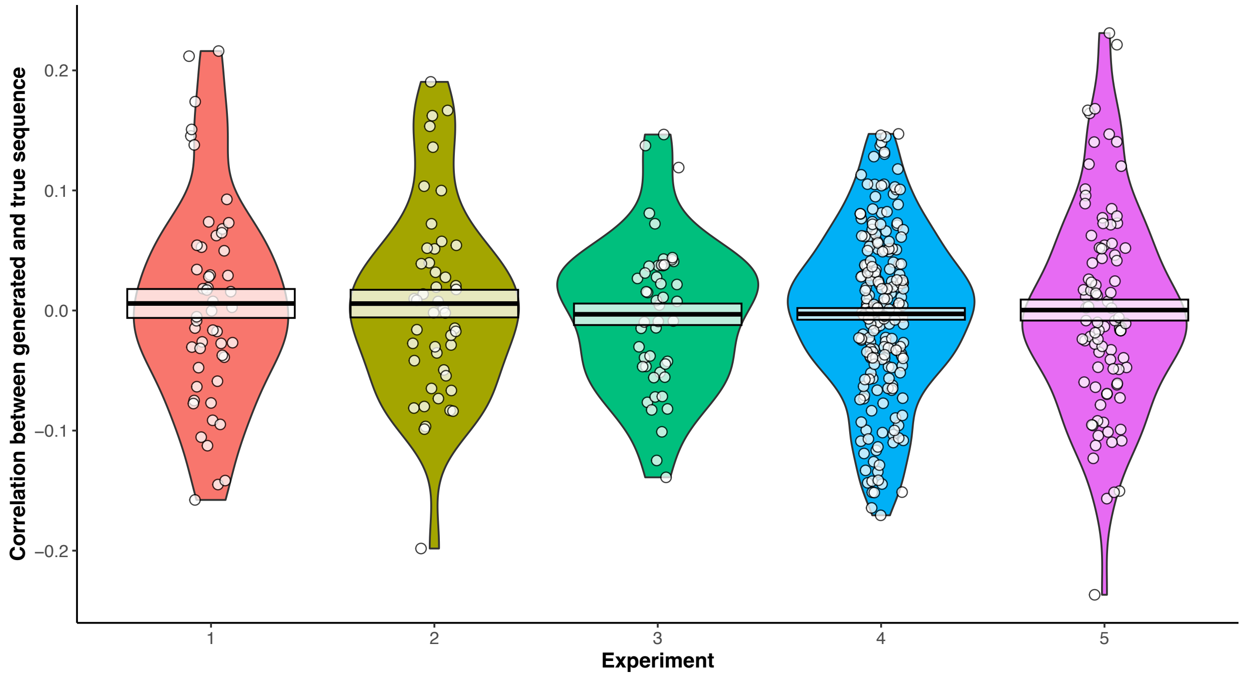

Figure S8. Correlation between the generated and the true sequence for each participant across each of the five experiments. Light grey dots represent individual Pearson correlation values for each participant. Boxplots indicate the mean and the standard error for the noisy bimodal condition of each experiment.