In many psycholinguistic experiments, a fixed set of $n$ items is repeated $m$ times. The total number of $n \times m$ stimuli is presented in randomized order and participants’ responses are recorded. Due to the randomized presentation order, it is assumed that participants cannot predict upcoming stimuli - and that therefore any significant effects in the experiment result from the intended experimental conditions (category type, word duration, etc.). We investigated a potentially confounding effect that may emerge from the random distribution itself.

We first inspected the distribution of unigram, bigram, trigram and four-gram sequences in a randomly generated sequence of 2000 stimuli containing 500 instances of four items (A, B, C, D). For example, the sequence ABCDABCD contains the bigrams AB, BC, CD and DA and the trigrams ABC, BCD, CDA and DAB. The four upper panels in Figure 1 illustrate the rank-frequency distributions (frequency of occurrence of the most frequent to the least frequent) of the n-grams in this sequence. Even when relatively short sequences of 2000 stimuli are inspected, frequency structures emerge. The imbalance between rare and frequent short sequences increases with sequence length.

Frequency effects are among the most robust effects in psycholinguistic research (see e.g. Ellis, 2002, for a review). This suggests that the frequency structures in our randomly generated stimuli may potentially be used to predict upcoming stimuli, as expectation increases with sequence frequency. We tested this hypothesis using the Rescorla-Wagner learning equations (Rescorla and Wagner, 1972; Ramscar, Dye, and McCauley, 2013). The bottom panels of Figure 1 illustrate results of two learning simulations. Trial number is on the x-axis; strength of expectation based on previous trials is on the y-axis. In the first simulation (left panel), the training set consisted of trigram cues (ABC), whose last instance (C) was the outcome of the training. In the second simulation (right panel), the training sequence consisted of bigram cues (AB) with C as the outcome. In both simulations the prediction strength increases over the experiment for both the frequent (red line) and rare sequences (blue line). However, prediction strength is greater for the frequent sequence. That is, the frequent sequence predicts an upcoming item better than a rare sequence. The prediction strength is lower when the target stimulus is not part of the cue structure (right panel).

The simulations suggest that randomization of stimuli results in frequency structures that may be used by participants to form expectations about upcoming trials. We tested whether the predictions of these simulations bear out and participants indeed learn such emerging structures. By means of these different cue structures it is possible to investigate the number of cues participants integrate (e.g AB vs. ABC) to learn to predict upcoming events. If participants learn to use these structures, this should be reflected in faster responses to target stimuli preceded by more frequent sequences than target stimuli preceded by less frequent sequences, as well as faster typing of more frequent sequences than less frequent sequences.

If our predictions bare out, these results would not only have implications for effects in standard categorization tasks. They may also inform us about how humans learn sequences of events such as syntax in the linguistic domain.
Figure 1: Upper two panels: Rank-frequency plots for n-grams in a random sequence of four items. Note: x and y limits differ between plots. Bottom panels: Prediction strength for upcoming stimuli given preceding bigram or trigram sequences.

References

