# Monkey Text Populations

There are a variety of different kinds of Monkey text populations that can be defined that in some significant way bear some statistical similarity to the Torah text. Each is created by taking the Torah text or its ELSs and performing some kind of randomly shuffling, making whatever compactness relationships that might occur in these texts due to pure chance.

## 1 Real Text Populations

### 1.1 Letter Shuffling

The Torah text has five books and 304,805 letters. Monkey texts can be created by randomly permuting (shuffling) the letters of the Torah to make a monkey text. Each monkey text is created by an independent letter shuffle. The letter shuffling can be done globally, book by book, chapter by chapter, verse by verse, or word by word.

### 1.2 Word Shuffling

The five books of the Torah text have 79,976 words. Monkey texts can be created by randomly permuting (shuffling) the words of the Torah text to make a monkey text. Each monkey text is created by an independent word shuffle. The word shuffle can be done globally, book by book, chapter by chapter, or verse by verse.

### 1.3 Verse Shuffling

The five books of the Torah text have 5,846 verses. Monkey texts can be created by randomly permuting (shuffling) the verses of the Torah text to make a monkey text. Each monkey text is created by an independent verse shuffle. The verse shuffle can be done globally, book by book, or chapter by chapter.

### 1.4 Chapter Shuffling

The five books of the Torah text have 187 chapters. Monkey texts can be created by randomly permuting the order of the chapters, globally, or book by book.

### 1.5 Constant Word Shuffle

The five books of the Torah text have 79,976 words. The number of distinct (different) words is 12,831. In the constant word shuffle, the letters of each of the 12,831 distinct words is randomly shuffled. Each time a word appears in the Torah text the same letter shuffling of that word appears in the shuffled text.

## 2 Virtual Text Populations

The real text shuffling schemes produce Monkey texts in which the number of ELSs for each key word will differ from that in the Torah text. The virtual text populations described here produce for each text exactly the same number of ELSs.

Suppose that for a given key word set, the Torah text has some statistical advantage over texts in say a letter permuted text population because the Torah text has substantially more low rank skip ELSs of some of the key words than expected by chance. In this case an experiment might succeed mainly due to such an ELS distribution in the Torah text, rather than because of the relationship between ELSs of the key words. The ELS random placement text population can be said to be a conservative one because in this case, each ELS random placement text has the identical statistical advantage as the Torah text, and therefore, no text has any advantage.

### 2.1 ELS Random Placement Text Population

The ELS random placement text population is always with respect to a given text and its set $\mathcal{E}$ of ELSs of the given set of key words and skip specification. A text of the ELS random placement text population does not consist of a text as a long string of letters. Rather, each text of the population is represented as a set of ELSs where each ELS keeps the same skip, length, and characters as it had in the original ELS set $\mathcal{E}$. However the beginning (and therefore the ending) positions of each ELS are randomly translated. Each translation that keeps the span of the ELS entirely within the text length has the same probability of occurring. This translation happens independently for each ELS. So if there were $N$ ELSs and ELS $n$ had $X_n$ possible translations, then the number of texts in the ELS random placement text population would in effect be

$$\prod_{n=1}^{N} X_n$$

## 2.2 ELS Random Skip Text Population

In the ELS random skip text population, each text has the same number of ELSs and the beginning position of each ELS is at the same position as it has in the text. So the placements of each ELS are the same. What is randomly sampled is the skip of the ELS. The skip of each ELS is randomly sampled from among the skips that keep the ELS entirely within the text given its starting position.

## 2.3 Global ELS Random Placement Text Population

The text is considered circular, meaning that the end of the Torah text is connected to the beginning. The Global ELS random placement text population is the only virtual text population that in a circular way preserves the spatial relationship between the ELSs of the same key word. In the global ELS random placement text population for a $K$ key word set, $K$ translations are randomly chosen, one for the ELSs of each of the $K$ key word sets. The same randomly chosen translation is applied to all the ELSs of the same key word. The restriction on the translation is that no ELS is allowed to have some of its beginning characters at the end of the text and some of its ending characters at the beginning of the text.

## 2.4 Connected ELS Random Placement Text Population

In some experiments, some key words may be substrings of other key words. From this it follows that some ELSs will be substrings of other ELSs. In the ELS random placement population, each ELS is randomly moved independent of every other ELS. But if one ELS is a substring of another ELS in the Torah text, and these ELSs are each moved independently, then the monkey text population would statistically differ from the Torah text in ELS substring dependence.

In the connected ELS Random Placement text population, a graph is formed between the ELSs of the Torah text that are substrings of one another. Each node of the graph is an ELS. Two ELSs are connected in the graph if one ELS is a substring of the other. The connected components of the graph is computed and all ELSs in the same connected component are randomly moved with the same random translation. In this way the ELS substring dependence between ELSs in the monkey text population is identical to the ELS substring dependence in the Torah text.