Simulating Language Assessment 1
Deadline: 27th of October at noon

Answer all the questions below. There is not a word limit, but your answers must be concise - typically only a few sentences will be necessary, and unnecessarily lengthy answers will be penalised. Some of the answers require a figure, which may include one or more graphs[1] - if you are unsure of how to save a graph from a jupyter notebook, look back at Lab 1.

Please start your answer to each question on a new page in the document you submit on turnitin, and make sure you indicate on each page what question you are answering!

Each question has an allocated number of marks to indicate its importance; the entire assignment is marked out of 30, and we will scale the resulting marks to lie on the University's Common Marking Scheme.

This work should be completed on your own. **Do not discuss your answers with other students!** Note that you should not need to include any code in your answer, only text and figures which clearly illustrate your results; however, keep a copy of the code you used to answer the questions, we may ask to see it.

Please note that we will **only be answering questions about this assignment until noon on Friday 21st October, and all questions must be posted in the assignment channel on Teams so everyone can see the answers**.[2] This is due to past experience that answering questions for too long after the assignment is released introduces unfairness in the results.

[1] Also, you can check out matplotlib.org for more details of how to create various different kinds of plots if you wish.

[2] This rule does not apply if you have a specific adjustment on your student record that allows for clarification on assignment questions.

Questions:

1. Using the *Lab 2 code*, demonstrate the size principle in action.

    a. Present your results in a figure, and;

    b. Provide an accompanying caption which explains how the figure illustrates the main implications of the size principle. **[8 marks**]

2. In Bayesian models, learning is shaped by data and the prior. Using the Lab 3 frequency learning/regularisation code, illustrate:
    - *one case* where you can see the influence of the prior on learning, **and**
    - *one case* where the influence of the prior is hard to spot.

    a. Present your results in a figure or figures, and;
    b. Provide a caption explaining what the figure shows. If you need several figures, make sure each has an accompanying caption. **[8 marks]**

3. The simulation in Lab 4 is based on the Reali & Griffiths (2009) paper on regularisation in iterated learning. In section 5 of that paper they present results for a word learning experiment with human participants. If you were to simulate as closely as possible what happens in their experiment using our Lab 4 code:

    a. What values would you provide to the iterate function to closely match Condition 3 of their word learning experiment? Specify values for `alpha`, `n_productions`, `starting_count_w1`, & `generations`. **[2 marks]**

    b. Run iterated learning simulations using these parameters and present your results in a similar way to Condition 3 in figure 3a in Reali & Griffiths (2009). **[5 marks]**

    c. Provide a caption for your figure, explaining what these results show, and why. **[3 marks]**

4. Natural languages are transmitted by iterated learning. Based on what you know about iterated learning, if you observed that a certain type of language is relatively frequent in the languages of the world, what would you predict about its prior probability, and why? **[4 marks]**