Project Proposal

# Hypothesis

We aim to reproduce the results in the article "Focal colors are universal after all", written by Regier, Kay and Cook (2005). The paper can be found at <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1149449/?tool=pmcentrez&report=abstract>.

Since two studies were done in "Focal colours are universal after all," we will have two hypotheses in this study. Our first hypothesis is that colour foci from different languages should cluster around the English foci for white, black, red, green, yellow, and blue, irrespective of the center of the colour cluster that the foci belonged to. We will also hypothesize that the observed foci would cluster more closely together than the centres of the different colour extensions.

If our data supports both of the above hypotheses, then we have evidence for the idea that colour foci are universal.

# Methodologies, Models, and Analyses

The first study investigated the colour chips that were most often chosen as a foci. We'll analyse the data by calculating how many people chose foci on each chip. The model we'll use in this study is a variation of the prototype model we used in class. We will take the English speaker's foci for white, black, red, green, yellow, and blue as prototypes and compare it to the peaks in the foci distribution obtained using the WCS data. If we find that the best examples of colour-naming choices from 110 languages of the WCS assemble close to or at the best examples of English colour terms for white, black, red, green, yellow, and blue, then we have supporting evidence for the proposed universal (i.e. apply to both industrialized and nonindustrialized societies) foci.

The second study tested if colour foci were more clustered than the centers of colour extensions. We will use the CIEL\*a\*b\* colour space for this analysis. For each speaker and colour that they named, we will find the centroid of the colours in the CIEL\*a\*b\*. Then we will take the average of the centroids and match it to a chip in the WCS stimulus array. The focus for each colour is determined by the colour chip that the most number of speakers had chosen for a particular colour. Ties were broken by randomly selecting a colour chip with the most number of speakers that had chosen it, where it gave us a centroid and a focus. Then we calculate the centroid separation (CS) and the focus separation (FS) using the formulas listed below:



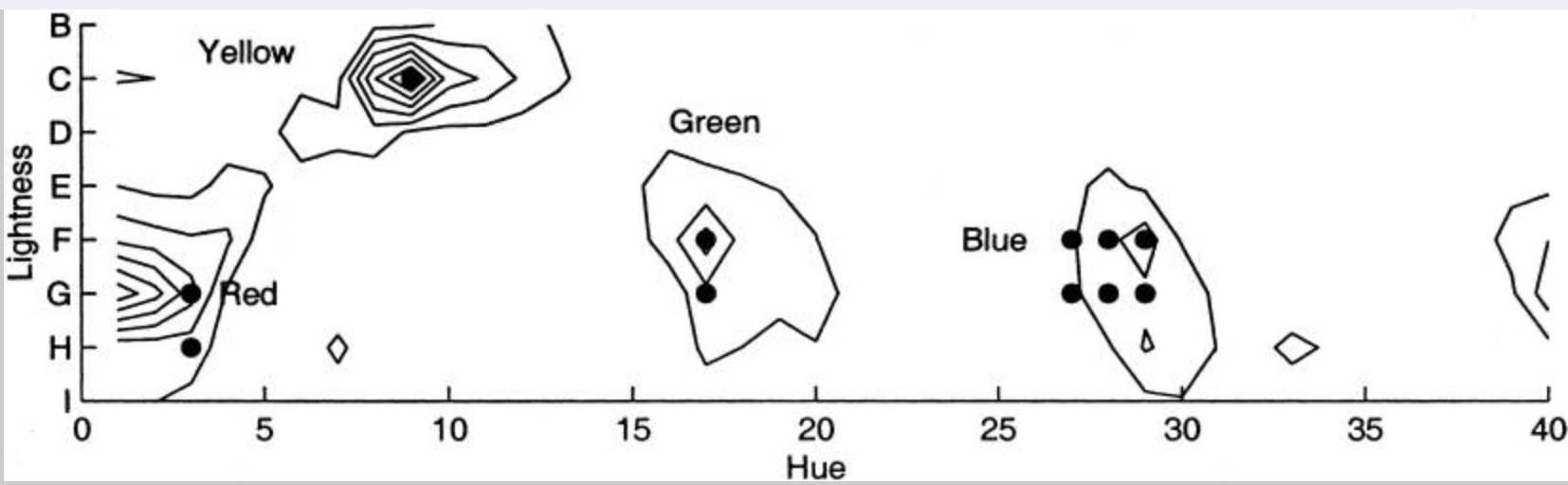


Where c refers to the centroid, and f refers to focus, l refers to each language in the WCS, l\* refers to Berlin and Kay language, t refers to the term, t\* refers to terms from Berlin and Kay’s data, BK refers to Berlin and Kay.

We will then analyze the data by comparing the focus separation and the centroid separation using a paired t-test. If we find that the foci are tightly clustered than the centroids, then this study supports the existence of universal foci as it shows that foci are related to the colour categories formation, and is not a byproduct of the categories.

# Expected Results (Figures/Tables)

For study 1, we expect to recreate figure 1 from "Focal colours are universal after all" (shown below)



We will expect to find that colour foci cluster around the English foci for white, black, red, green, yellow, and blue, as that is what Regier, Kay and Cook found.

For study 2, Regier, Kay and Cook found that the focus separation was 5,596.98, and the centroid separation M was 6,391.78. Since we're replicating the study, we expect similar results. On top of the calculations, we expect to generate a table for the Paired Sample T-Test correlations.

# Division of Labour

Weiran will be in charge of replicating study one, and Huiyan will be in charge of replicating study two.