# Baby Names Selection Models

## Introduction

The naming of a child may depend on several different reasons. I want to know if one can model the frequency of a name in the population based solely on its frequency in the previous generation. My goal is to try and create a model that successfully describes the naming phenomena as a selection model, similar to a biological selection model.

**Hypothesis .** The naming phenomena can be modelled as a selection model, and a frequency-dependant model will fit the data better than a simple selection model with a constant selection bias.

## Models & Methods

**Model 1 .** Consider a population of size $N(t)$, where $t$ is the year. Every baby born can be named either $A$ or any other name, which is categorised as $B$, where the relative fitnesses are: $w_A = 1 + s, w_B = 1$, where $s > -1$ and is a constant.

**Model 2 .** Consider a population of the same size as Model 1, with the same names types $A$ or $B$. In this model $s_{t+1} = f(p_t)$, where $p_t$ is the frequency of the name $A$ in year $t$, and $s_{t+1}$ is the selection bias in year $t + 1$ of $A$. The relative fitnesses in year $t$ are therefore: $w_{t,A} = 1 + s_{t-1}, w_{t,B} = 1$, where $s > -1$. My initial selection coefficient function $f(p_t) = a \cdot p_t^2 + b \cdot p_t + c$, where $a, b, c$ are the model's parameters. During research I may use other frequency dependant functions that may better fit the data.

**Model Fitting.** I intend to fit my models to the data using the **Maximum Likelihood Estimation** approach, using *scipy* packages like *optimize*. I will then use model selection using the **F-test** approach.

## Expected Results

I expect to see a better fit of the frequency-dependant model to the data in more than 75% of the names tested. I will only fit names with frequency above a certain threshold, and with minimal amount of "noise" (i.e many sharp slope changes between years).

*Willing to present my work if requested