1 just some thoughts

Let's say engine A plays $n \geq 1$ games against engine B and scores $+w, = d, -l$, where w =number of wins, d =number of draws and l =number of losses, with $w + d + l = n$.

Then we may assume the expected score of A against B to be normally distributed around the mean

$$\mu = \frac{w}{n} + \frac{d}{2n} \text{ (win rate + half the draw rate)},$$

with 'small variance', if n is high, or 'larger variance' if not many games have been played. For example, let's use $\sigma = \frac{1000}{n}$ (?):

$$E_A \sim \mathcal{N}\left(\frac{w}{n} + \frac{d}{2n}, \frac{1000}{n}\right).$$

Given E_A, the rating difference $R_B - R_A$ can be calculated to be

$$R_B - R_A = \frac{400\ ln(\frac{1}{E_A} - 1)}{ln(10)}.$$

So the rating difference is a random variable that we expect to be around $\frac{400\ ln(\frac{1}{\mu} - 1)}{ln(10)}$ (?)

Our goal is to somehow set up the prior (=our initial belief, what we think the ratings are according to the data from the .csv-file) where the ratings are centered around 2000, using STAN ...