

More trees

The EVB for each single-action update (i.e., taking action a^* at belief state b) is calculated as:

$$\text{EVB}(ba^*) = p(b_\rho \rightarrow b \mid \pi) \times (\mathbb{E}_{b' \sim p(b'|b, a^*)} [v(ba^*)] - v(b)) \quad (1)$$

Where $p(b_\rho \rightarrow b \mid \pi)$ is the Need term – i.e., the probability of reaching belief b from the root of the tree, b_ρ , when following policy π . The policy π here is i) softmax of the MF Q -values at the root of the tree, b_ρ ; and ii) softmax of the immediate reward at every other node within the tree.

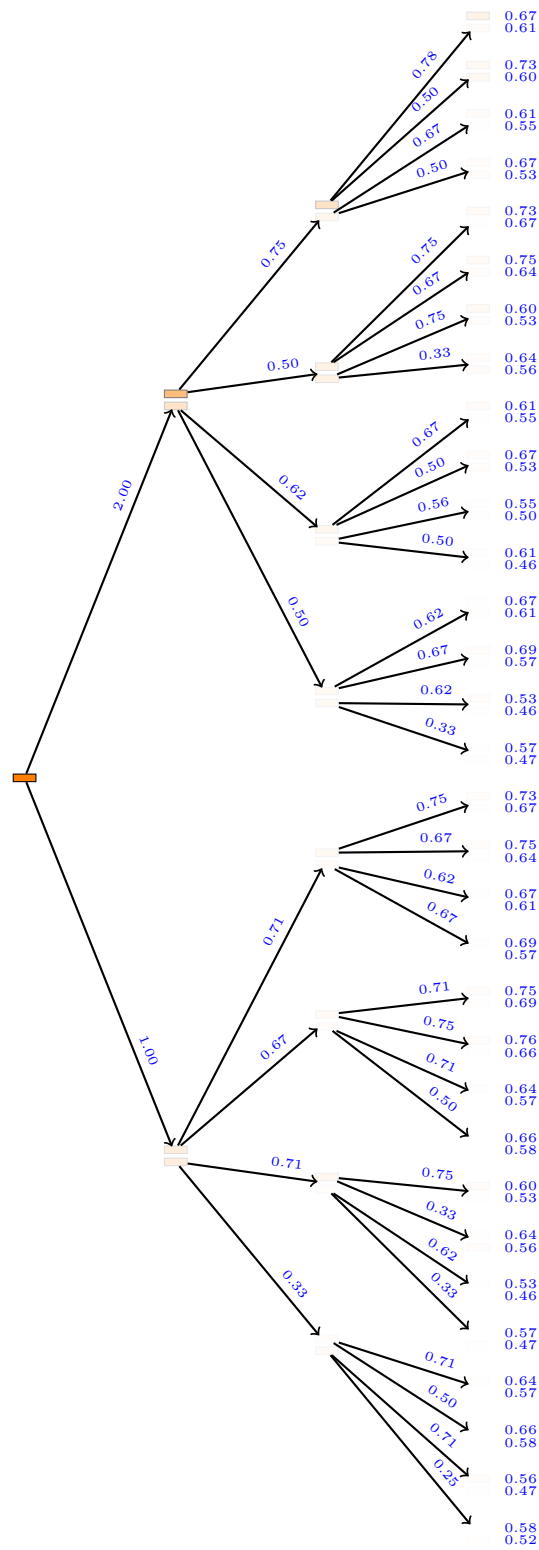
In the trees, each action has a blue number written above it – these are the Q -values. Each orange square is a belief state, and the opacity of each belief state is proportional to the Need term from equation 1. When belief states appear in pairs, the top belief state is always the one which results from obtaining a reward from the corresponding action, whilst the bottom belief state corresponds to zero reward.

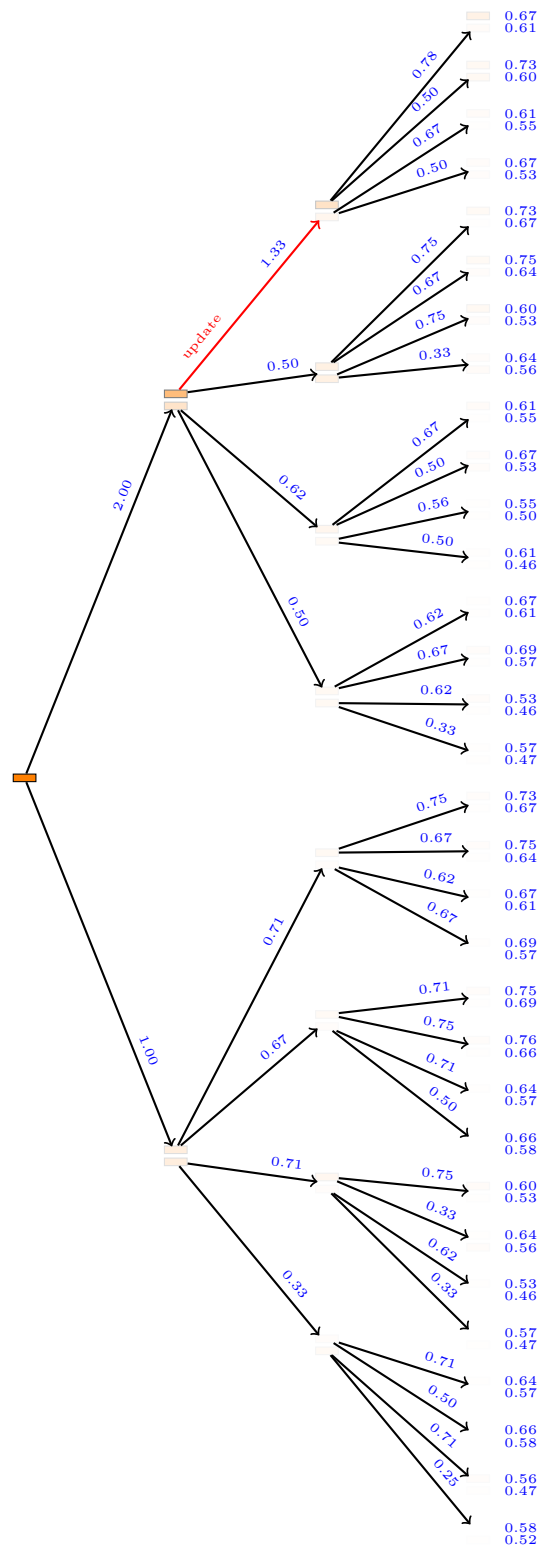
Updates in the tree are always highlighted by red arrows.

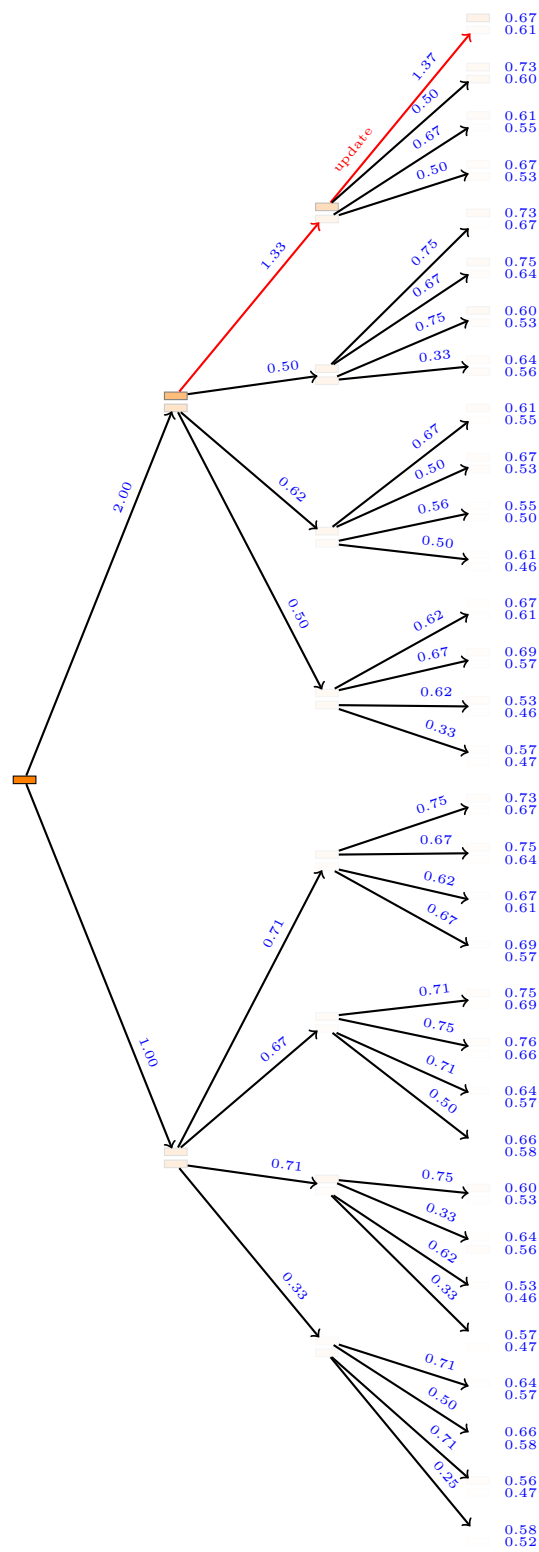
Example 1

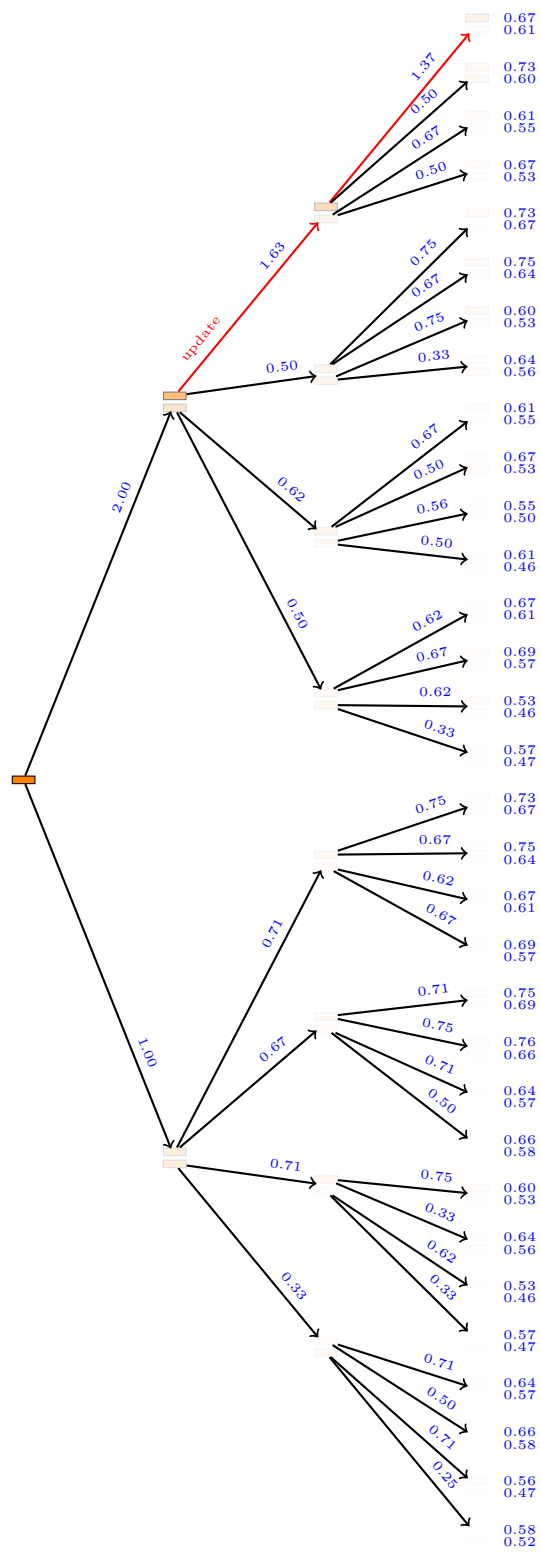
Belief at the root in this example is:

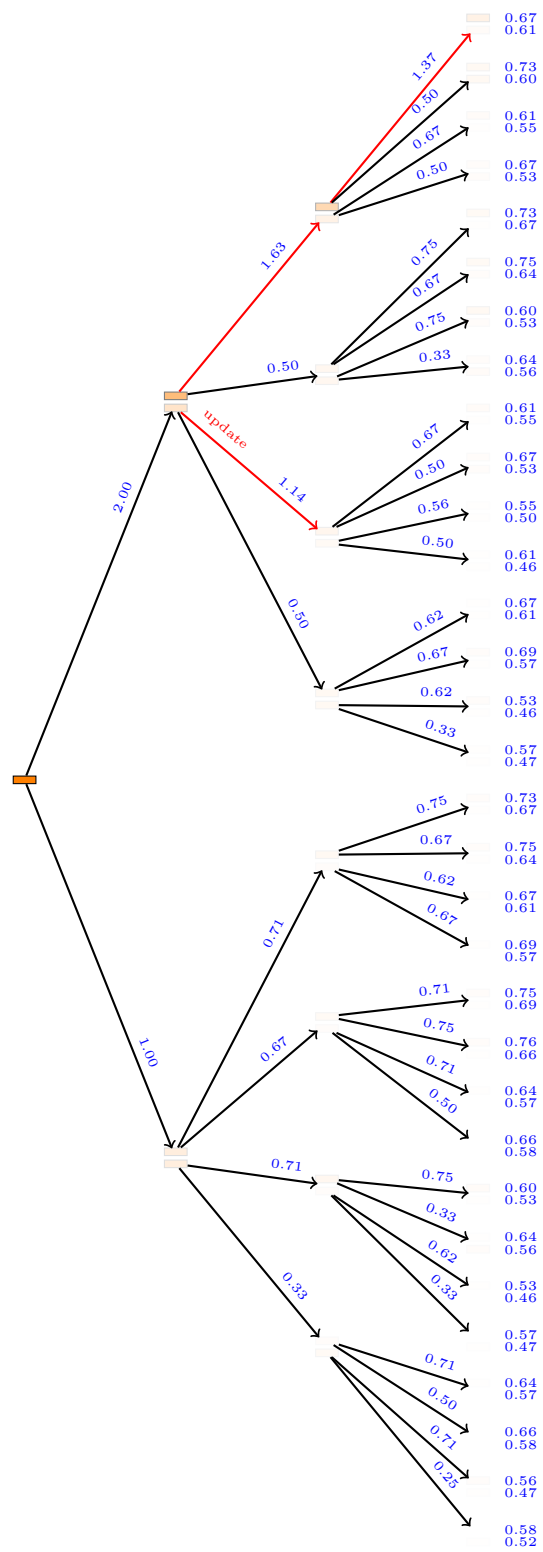
$$\begin{array}{ll} \alpha_0 = 5 & \beta_0 = 2 \\ \alpha_1 = 1 & \beta_1 = 1 \end{array}$$

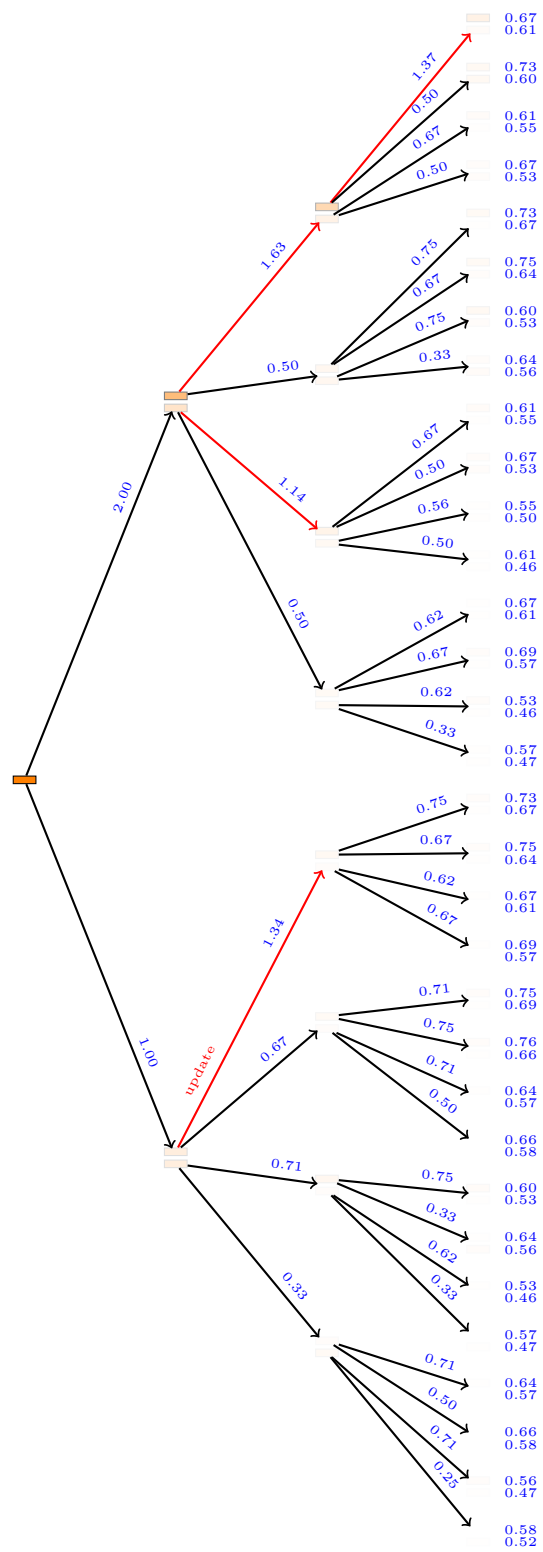


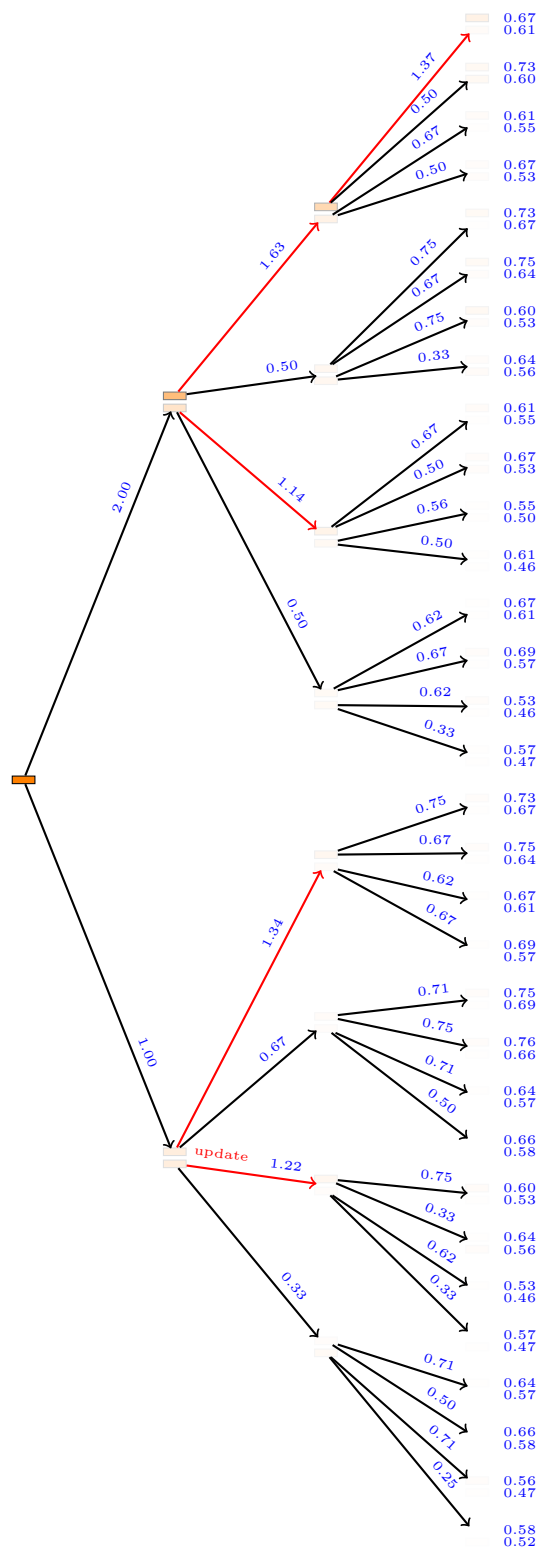


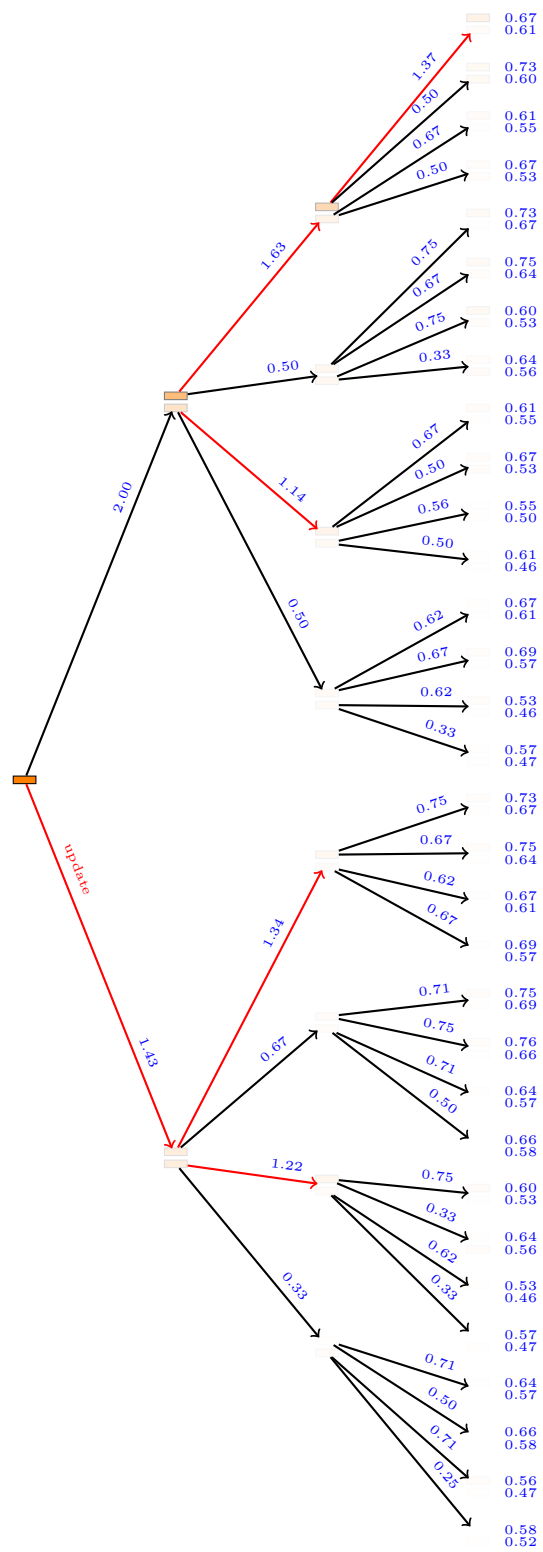


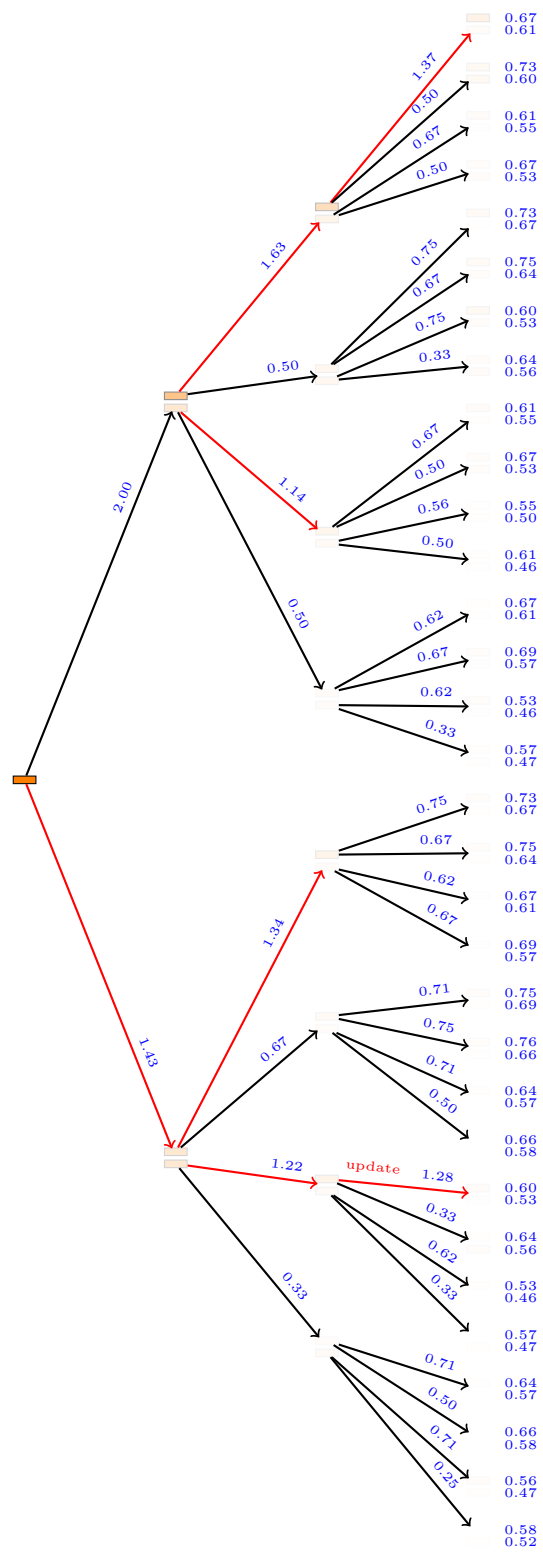


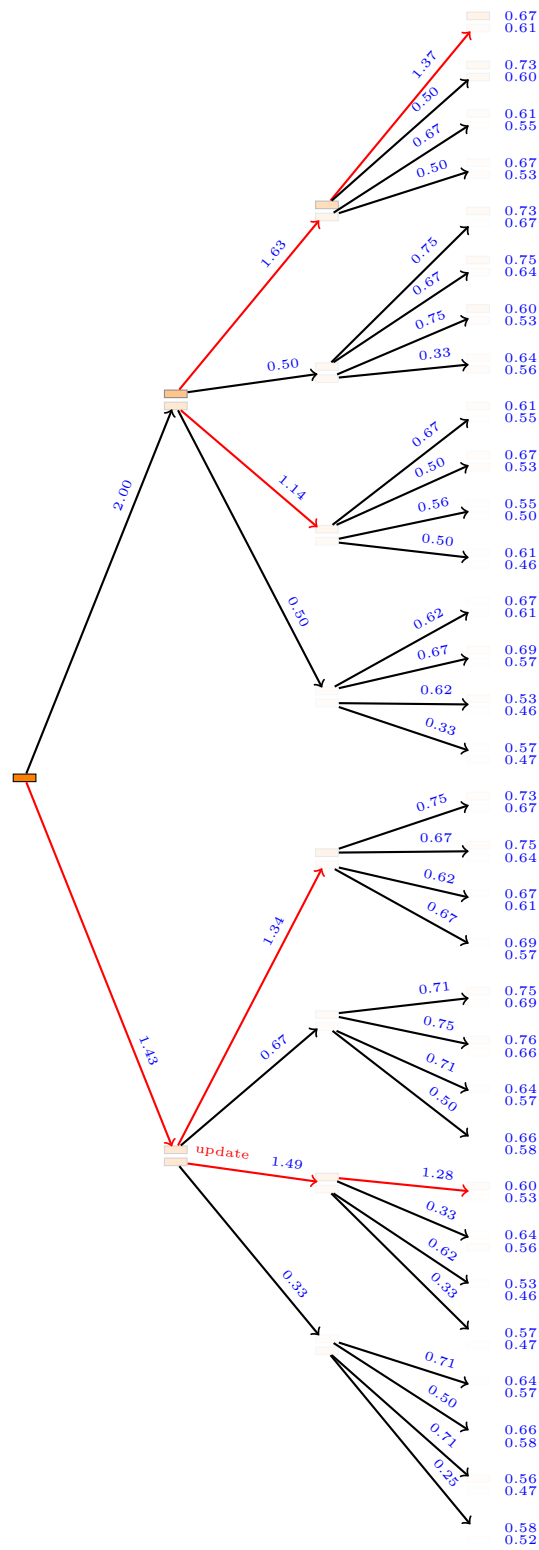


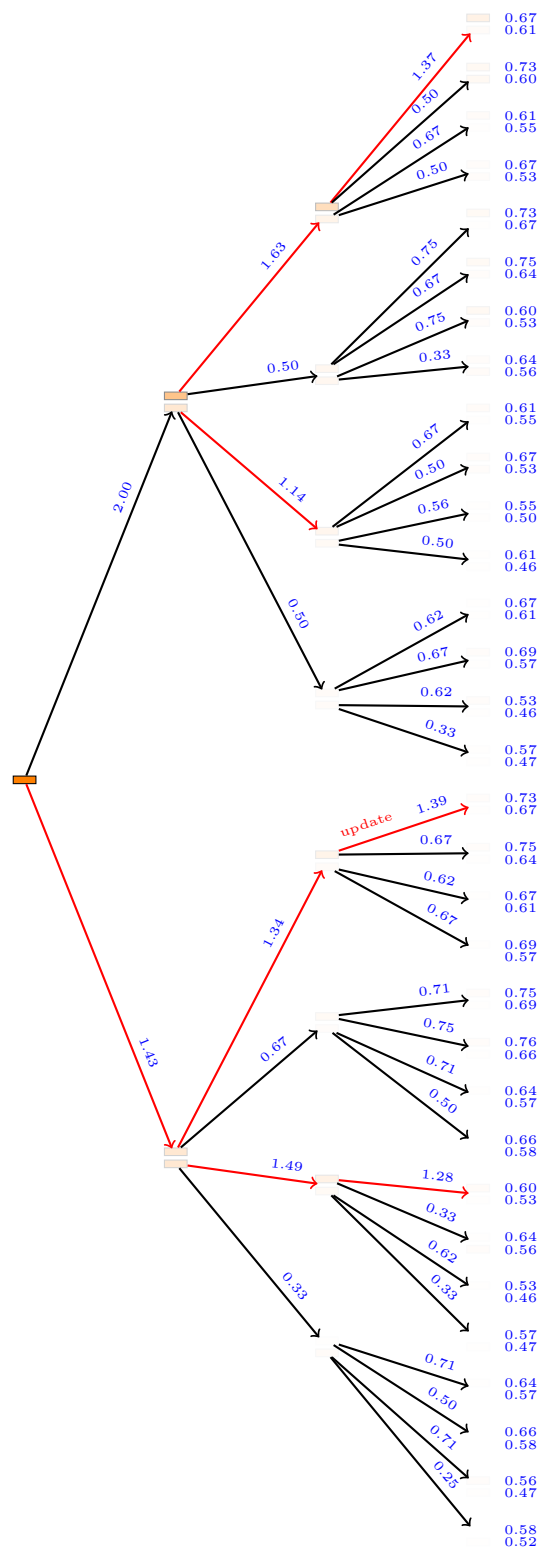


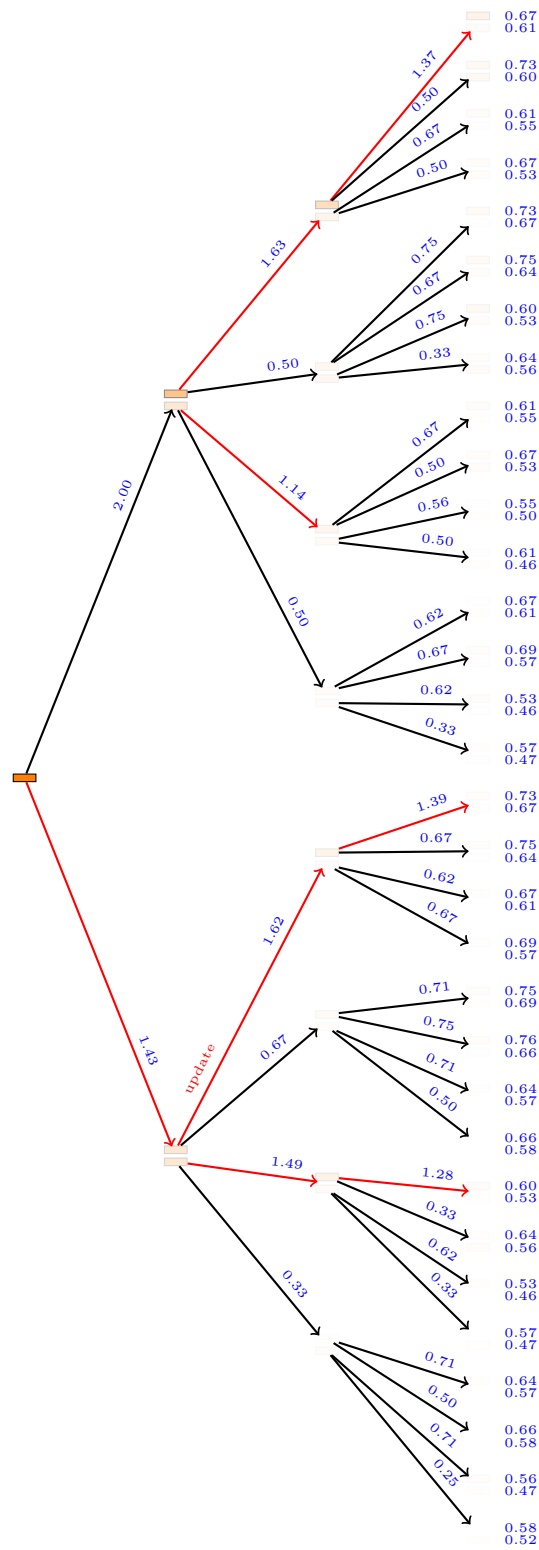




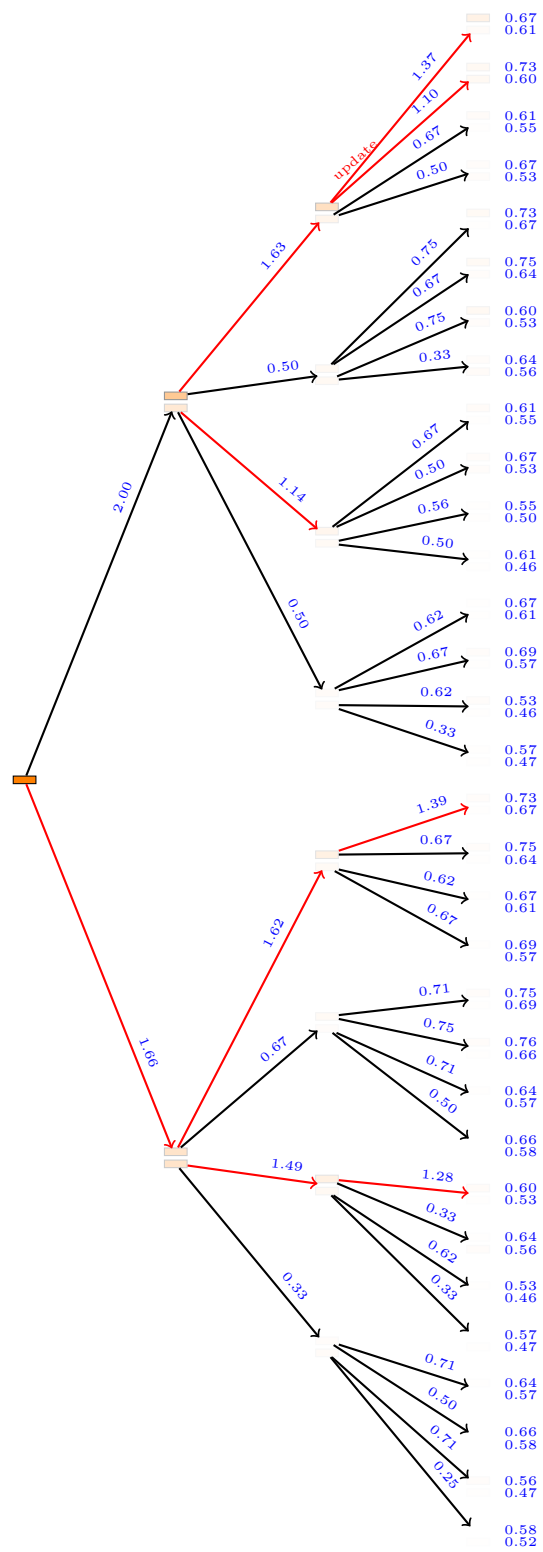


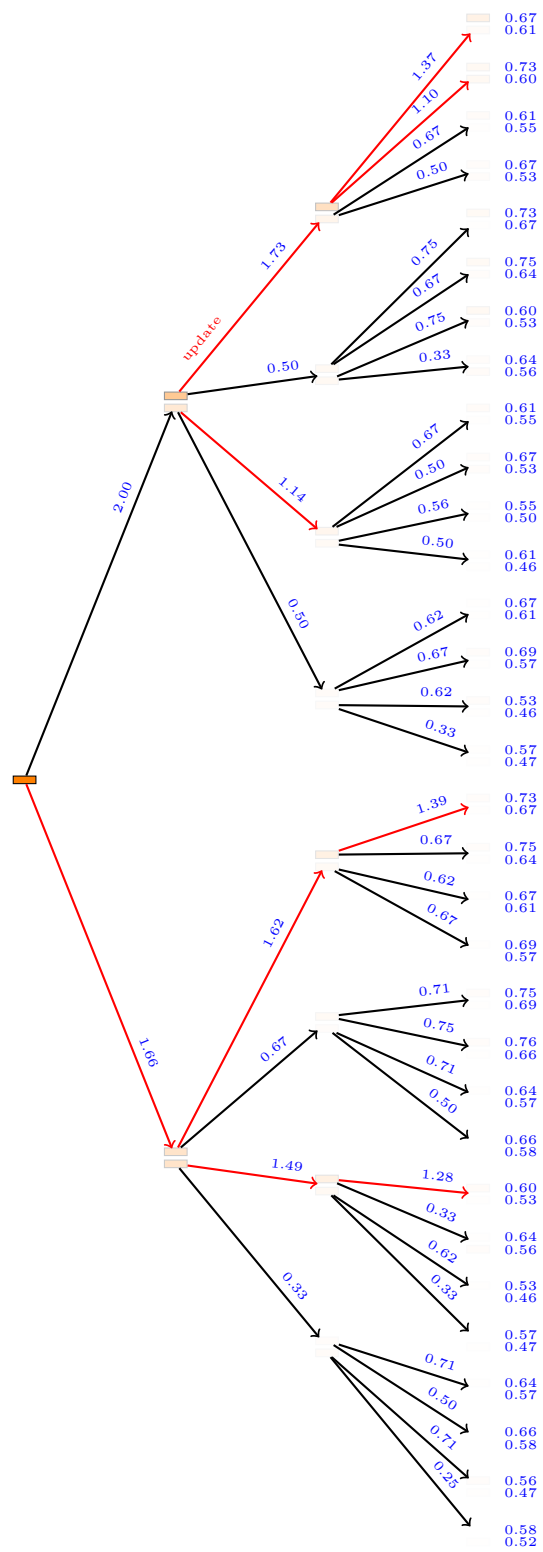


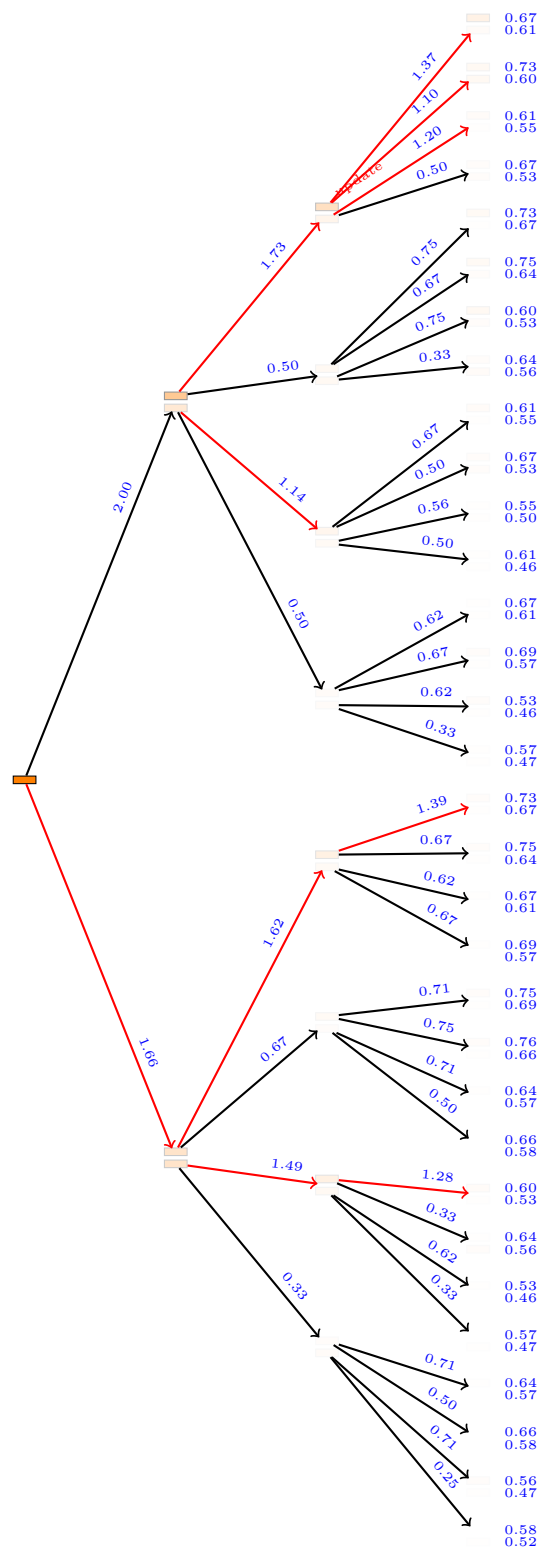


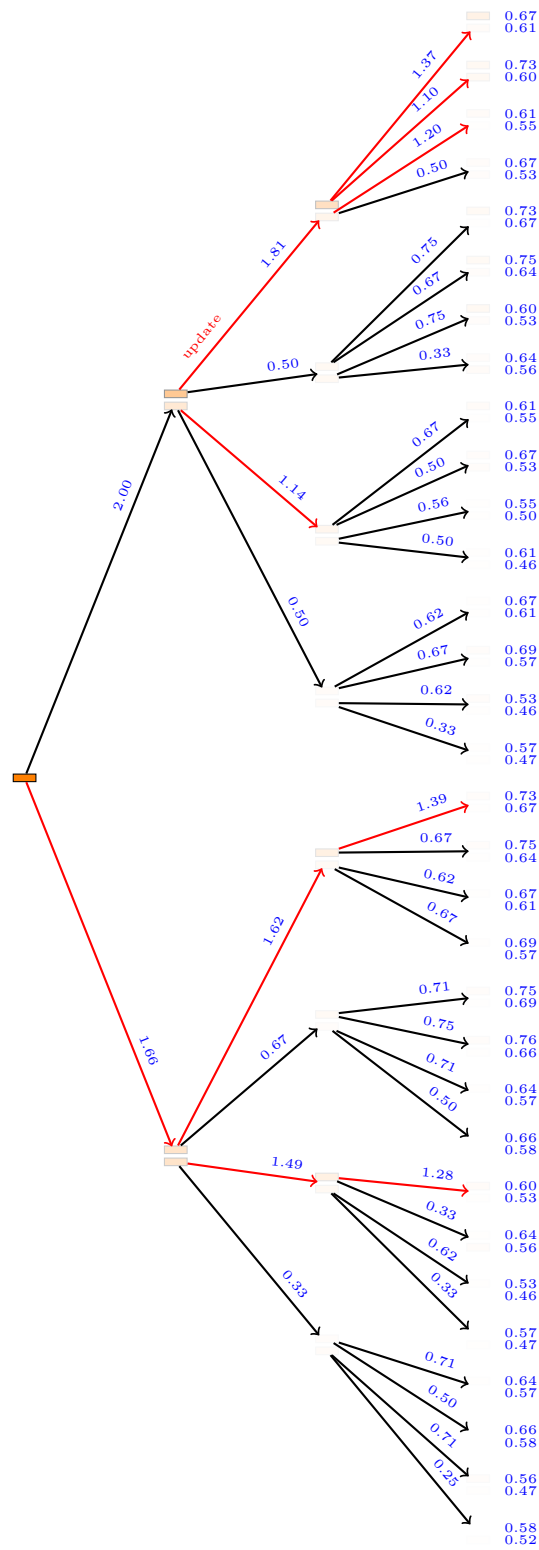








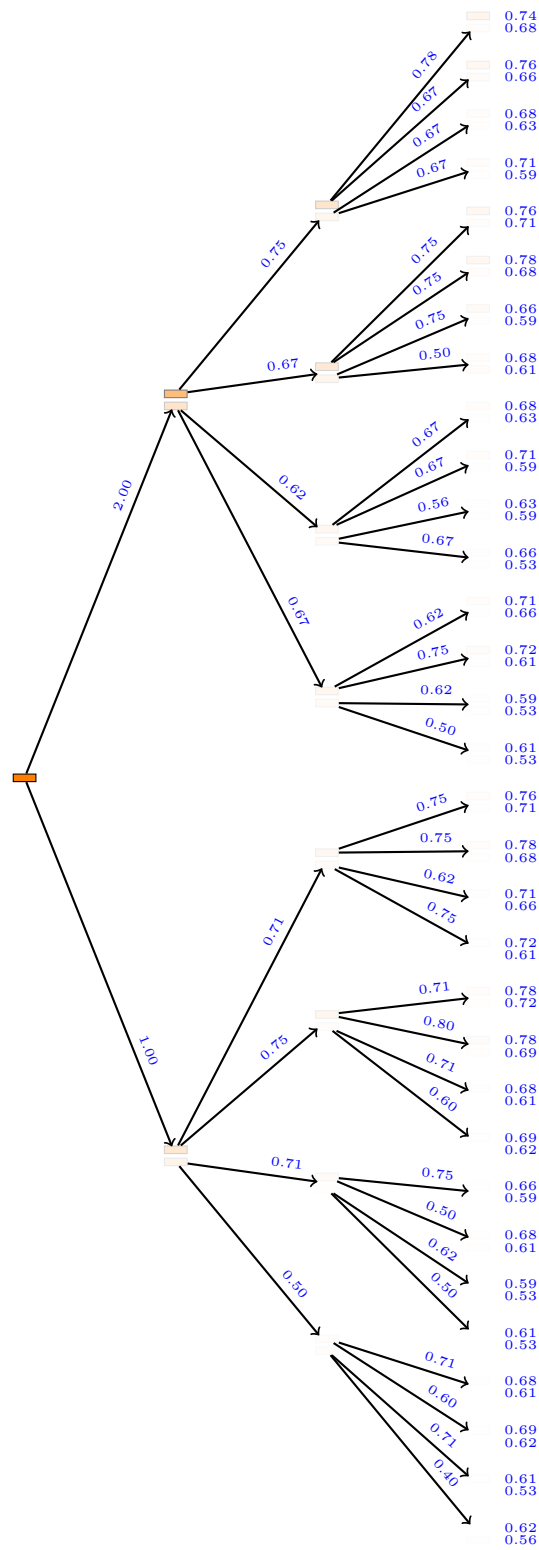


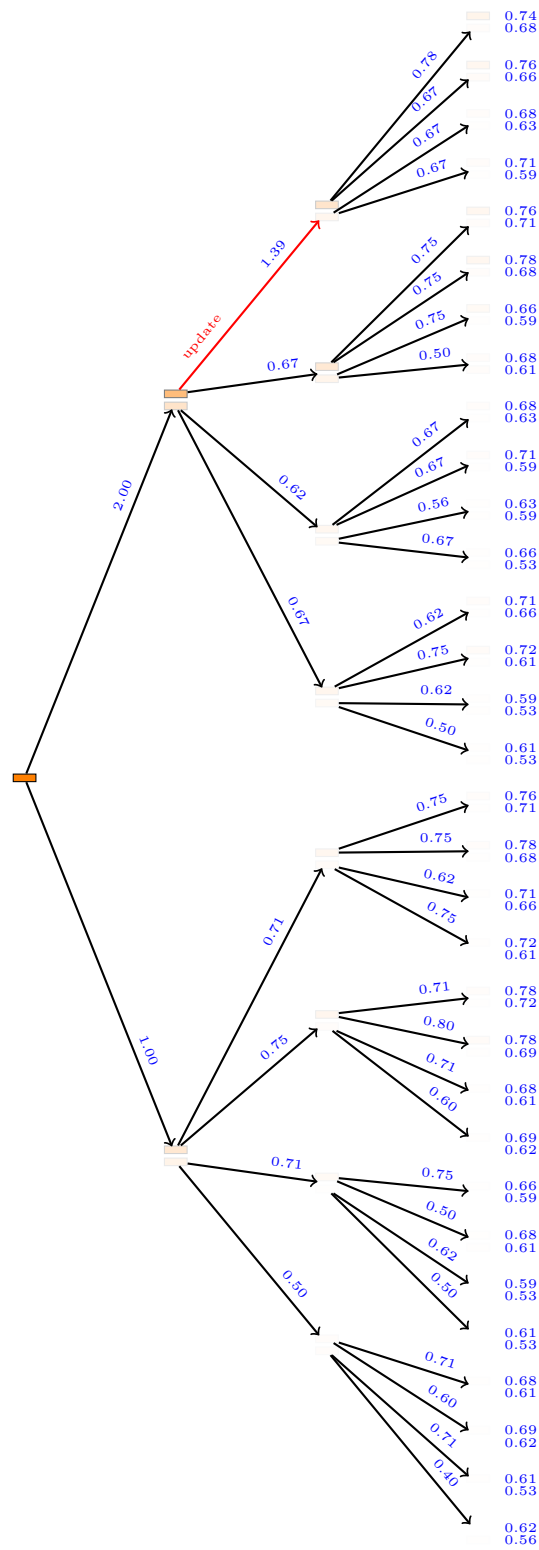


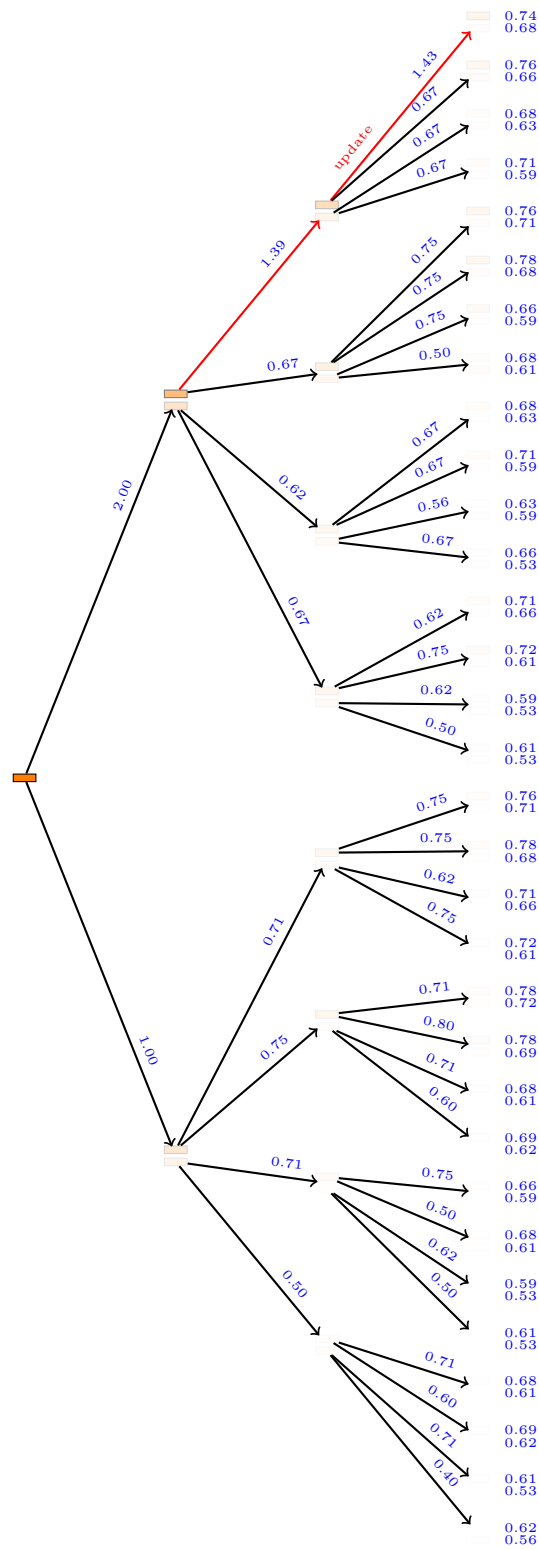
Example 2

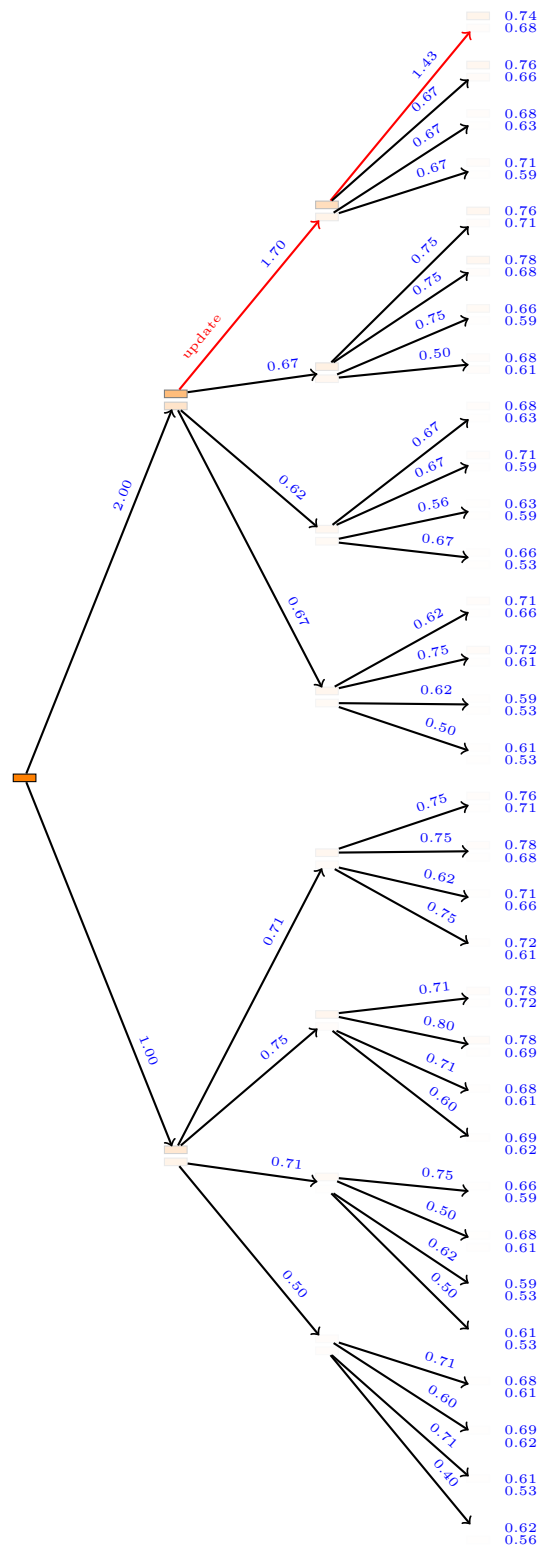
Belief at the root in this example is:

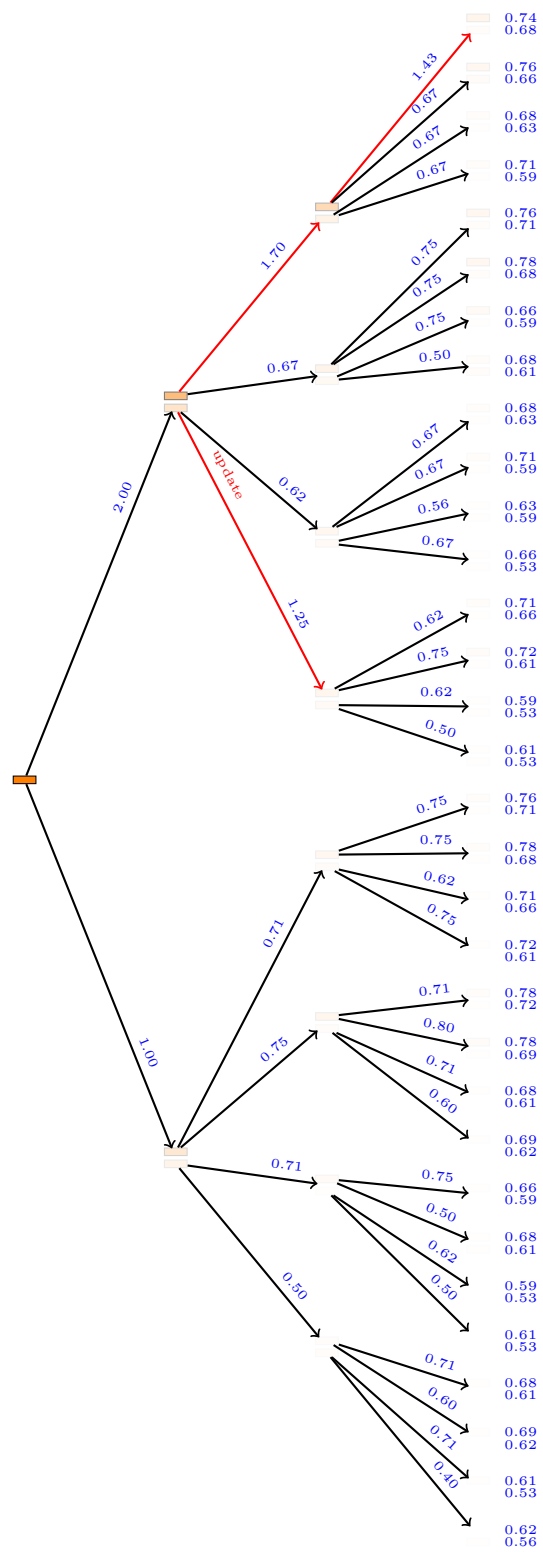
$$\begin{array}{ll} \alpha_0 = 5 & \beta_0 = 2 \\ \alpha_1 = 2 & \beta_1 = 1 \end{array}$$

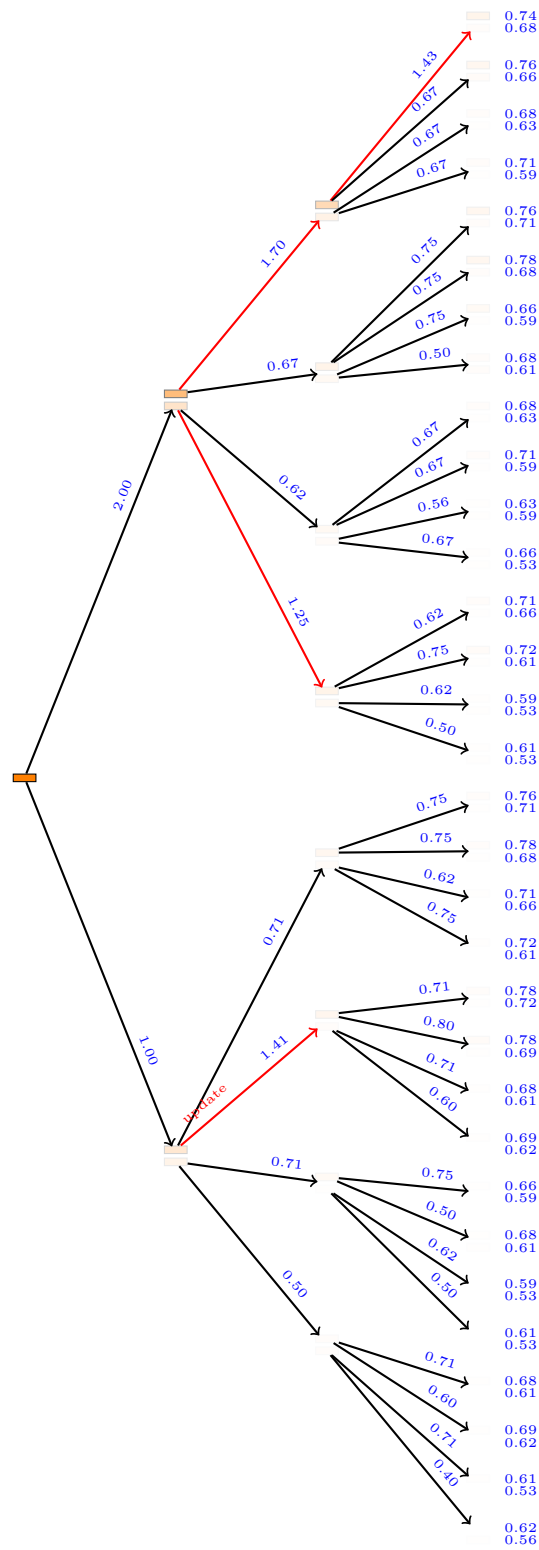


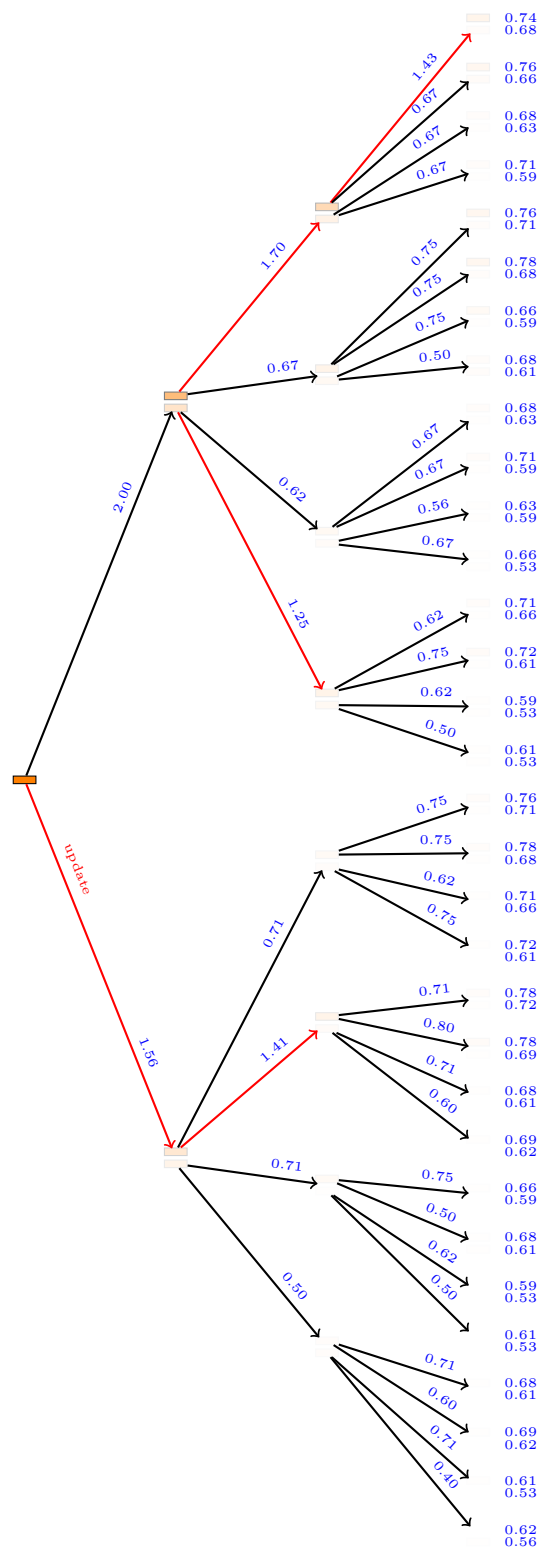




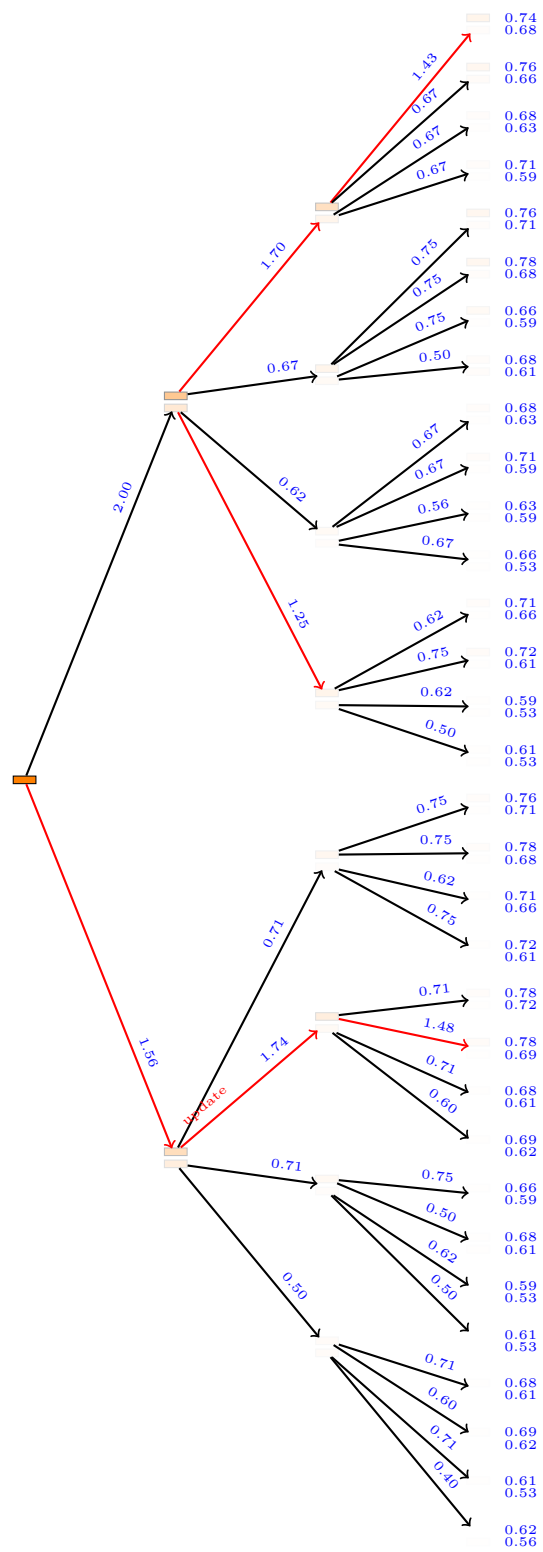


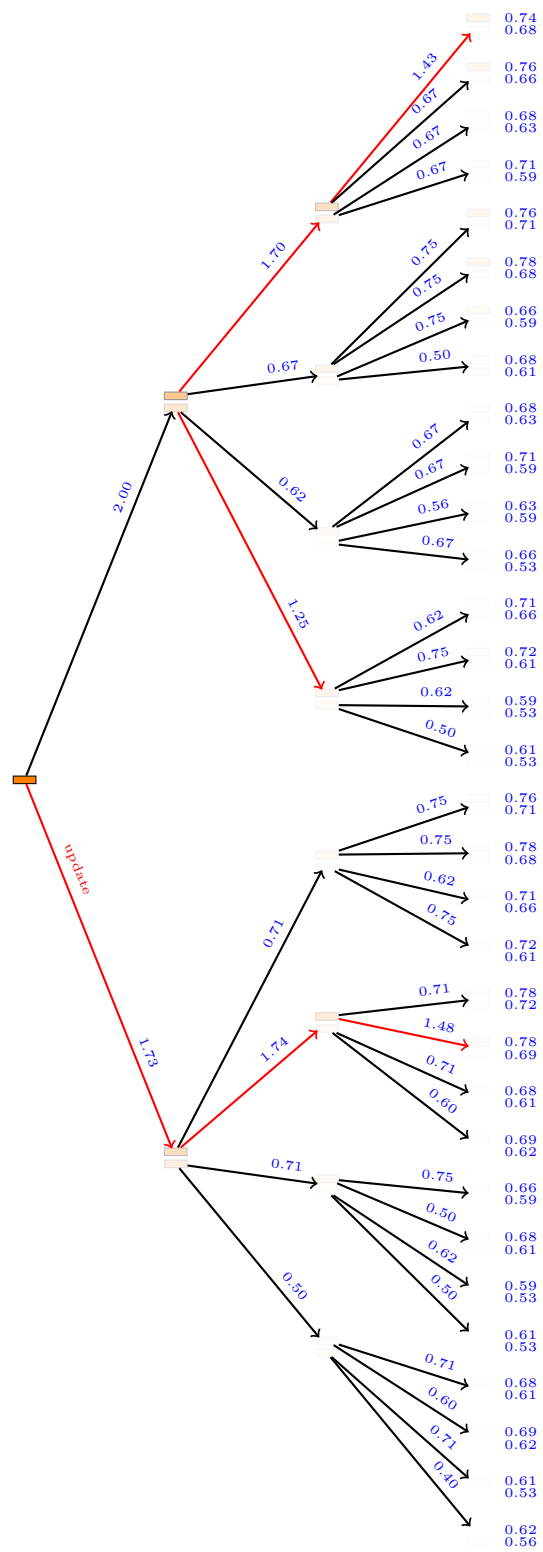


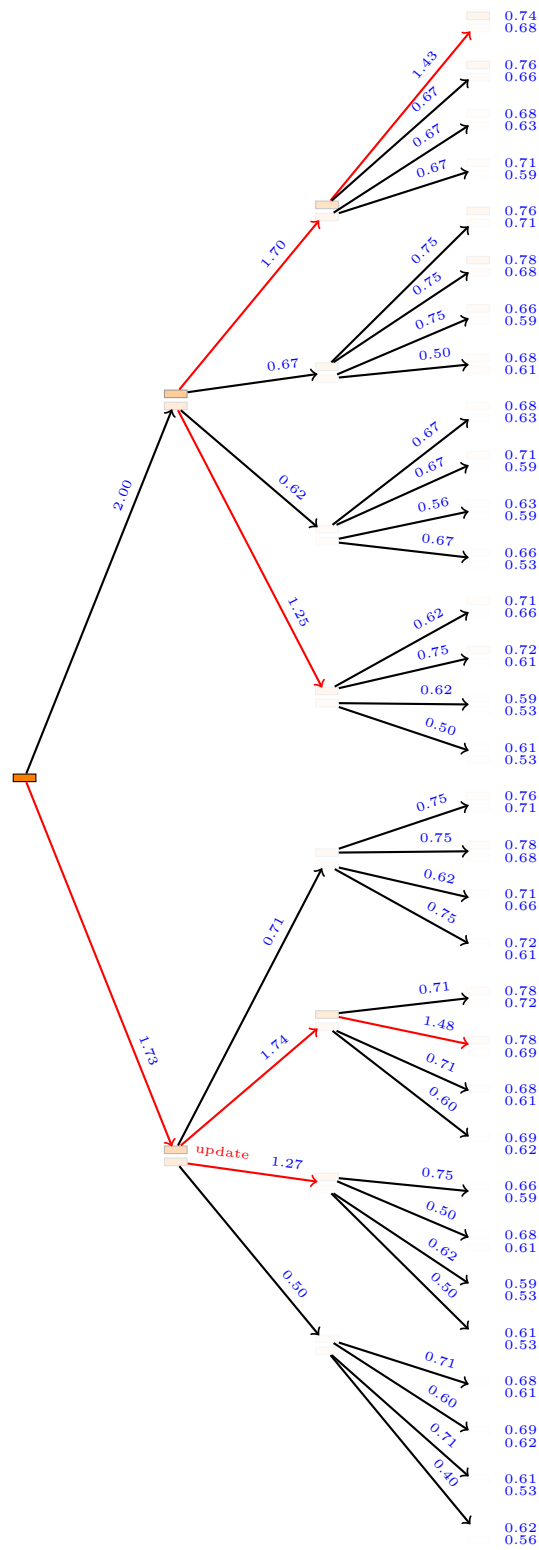












Example 3

Belief at the root in this example is:

$$\begin{array}{ll} \alpha_0 = 10 & \beta_0 = 4 \\ \alpha_1 = 2 & \beta_1 = 1 \end{array}$$

