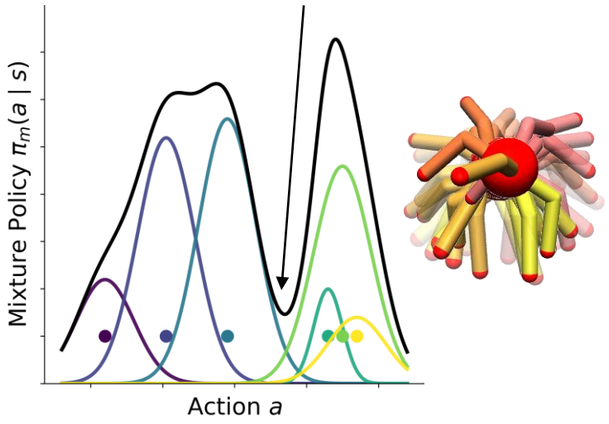
Achieving maximum action path occupancy with low entropy components

Actions leading to terminal states (e.g., falling) have low probabilities



$$\pi_m(a|s) = \sum_{k=1}^{K} \omega_k(s)\, \pi_k(a|s)$$

$$\underset{\pi_m}{\text{argmax}} \sum_{t}^{\infty} \gamma^t \mathbb{E}\left[H(\pi_m(\cdot\,|s_t)) - \alpha \sum_{k} \omega_k(s_t) H(\pi_k(\cdot\,|s_t)) \right]$$