Optimism in the Face of Uncertainty

- Despite our lack of knowledge in what actions are best we will construct an optimistic guess as to how good the expected reward of each action is, and pick the action with the highest guess.
- If our guess is wrong, then our optimistic guess will quickly decrease and we'll be compelled to switch to a different action.
- But if we pick well, we'll be able to exploit that action and incur little regret. In this way we balance exploration and exploitation.

Upper confidence bound

- At time t, sample from the arm with maximum UCB$_a(t)$.
- Recall that, UCB$_a(t) = \hat{\mu}_a(t) + \sqrt{\frac{2\log(T)}{N_a(t)}}$.
- **Thus, UCB prioritises arms having:**
  - Higher empirical mean $\hat{\mu}_a(t)$ – (Exploitation)
  - Lower no. of samples $N_a(t)$ – (Exploration)