Input: Set of arms A
Initialization: Set prior distribution parameters for each arm $a \in A$;
for $t = 1$ *to* T **do**
    for $a \in A$ **do**
        Sample a reward from the posterior distribution: $r(a) \leftarrow$ sample a reward from $P(a)$ based on the current posterior parameters;
    end
    Choose action $a_t = \arg\max_{a \in A} r(a)$;
    Perform action a_t on the chosen arm;
    Observe reward r_t obtained from the chosen action;
    Update the posterior distribution parameters for the selected arm based on the observed reward;
end