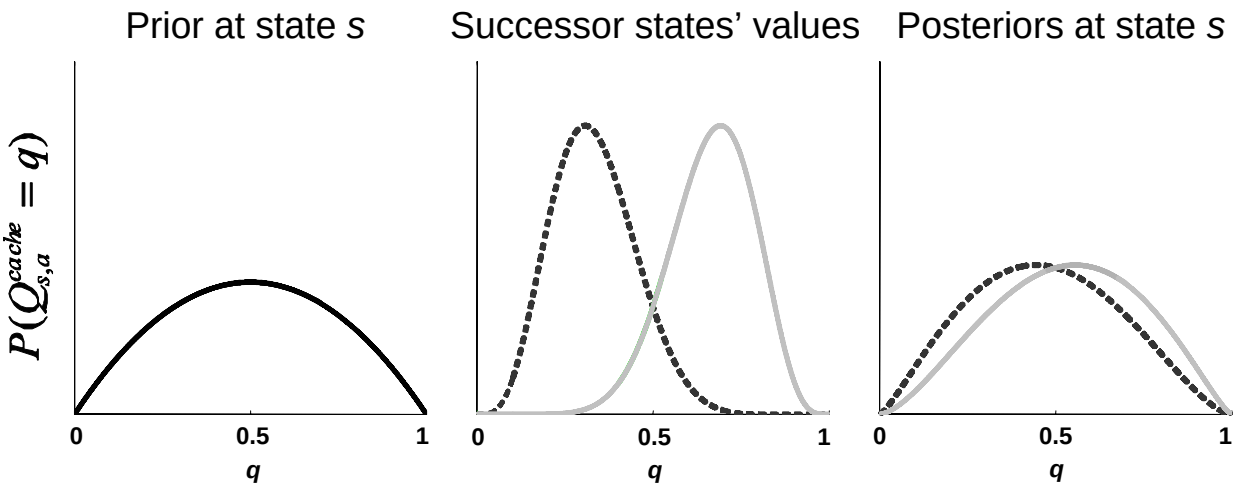

Supplementary Figure 2

Example of learning in the cache algorithm, following a single transition from state s to s' having taken action a. The leftmost panel shows the prior distribution $\mathbf{Q}^{\text{cache}}_{s,a}$. The middle panel plots the distribution over the value of the successor state s' (specifically, the distribution $\mathbf{Q}^{\text{cache}}_{s',a'}$ for the best action a' in s'). The curves illustrate two different successors, a more and less favorable state. For both of the successor states, the right panel plots the posterior distribution (whose moments are given by Equations 3 and 4 in **Supplementary Methods**) over the original state and action, $\mathbf{Q}^{\text{cache}}_{s,a}$, as updated following the visit to s'. The effect of learning was to nudge the predecessor state's value distribution in the direction of the value of the successor state.