**Agent
(Replay Buffer)
*TRANSIENT***

**Expert
(Demonstrations)
*FIXED***

points from
past policy
iterates