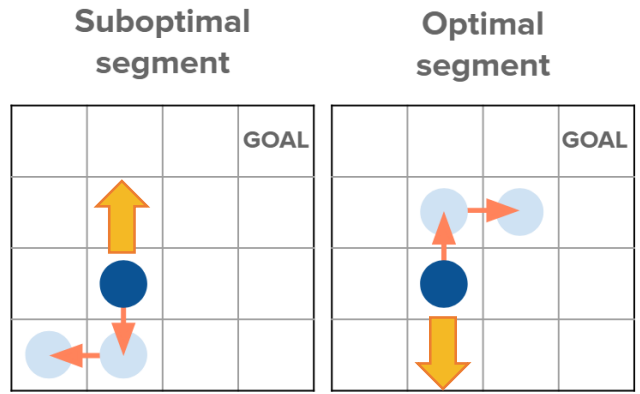
Each execution is not just an action.
It is a realization of both $\mathbb{P}$ and π!

I wanted to go up, but the dynamics made me go down!

Policy π_A

Bad result,
but right policy



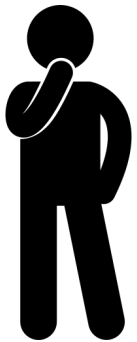
Equal partial return
Higher regret

Equal partial return
Lower regret

Policy π_B

Good result,
but wrong policy

I wanted to go down, but the dynamics made me go up!



Q. Which way should we prefer?:
the **result** or the **intended policy**?

A. Evaluate the **result** with **intended policy**

$$Q_*^{\pi}(s_t^{\sigma}, a_t^{\sigma})$$