



$$\begin{aligned}
 J_\theta &= E_{\tau \sim p(\tau; \theta)}[r(\tau)] = \int r(\tau) p(\tau; \theta) d\tau = \int d\tau r(\tau) \int \nabla_\theta p(\tau; \theta) d\theta \\
 &= \int d\tau r(\tau) \int p(\tau; \theta) \nabla_\theta \log p(\tau; \theta) d\theta = \int d\theta \int p(\tau; \theta) r(\tau) \nabla_\theta \log p(\tau; \theta) d\tau \\
 &= \int d\theta E_{\tau \sim p(\tau; \theta)}[r(\tau) \nabla_\theta \log p(\tau; \theta)] = E_{\tau \sim p(\tau; \theta)}[r(\tau) \log p(\tau; \theta)] \\
 &\approx \frac{1}{|batch|} \sum_{batch} r(\tau) \log p(\tau; \theta) \triangleq \frac{1}{|batch|} \sum_{batch} \sum_{t \geq 0} [r(\tau) - b_t] \log p(\tau_t; \theta)
 \end{aligned}$$