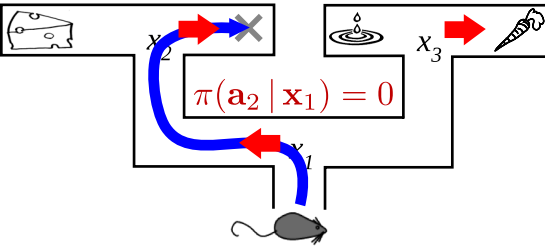


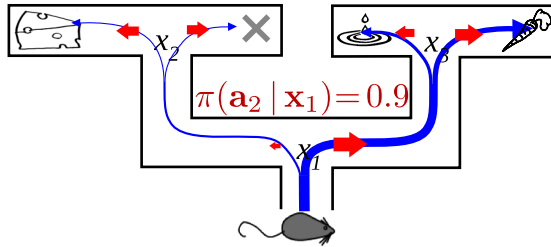
$$\pi(\mathbf{a}_2 \mid \mathbf{x}_2) = 1 \quad \pi(\mathbf{a}_2 \mid \mathbf{x}_3) = 1$$

**a bad policy**



$$\pi(\mathbf{a}_2 \mid \mathbf{x}_2) = 0.5 \quad \pi(\mathbf{a}_2 \mid \mathbf{x}_3) = 0.7$$

**a stochastic policy**



$$\pi(\mathbf{a}_2 \mid \mathbf{x}_2) = 0 \quad \pi(\mathbf{a}_2 \mid \mathbf{x}_3) = 1$$

**the best policy**

