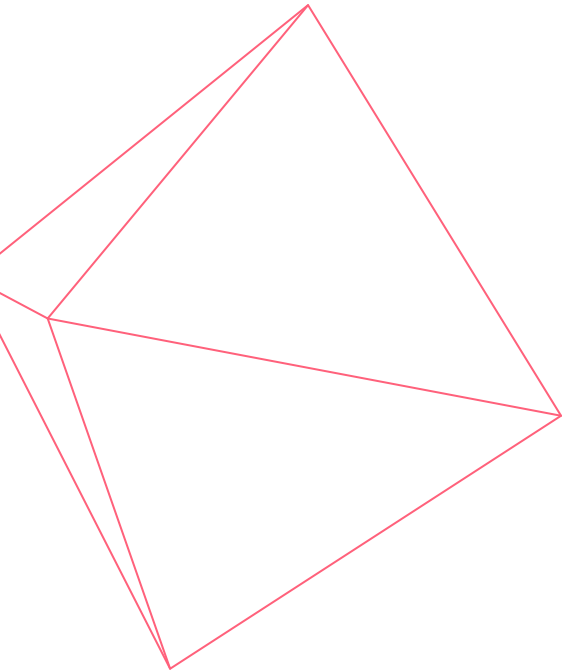
DeepMind

Hindsight Credit Assignment

Anna Harutyunyan, Will Dabney, Thomas Mesnard, Mo Azar, Bilal Piot, Nicolas Heess, Hado van Hasselt, Greg Wayne, Satinder Singh, Doina Precup, Remi Munos

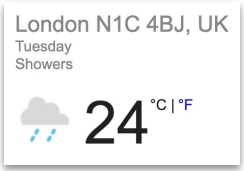
NeurIPS 2019



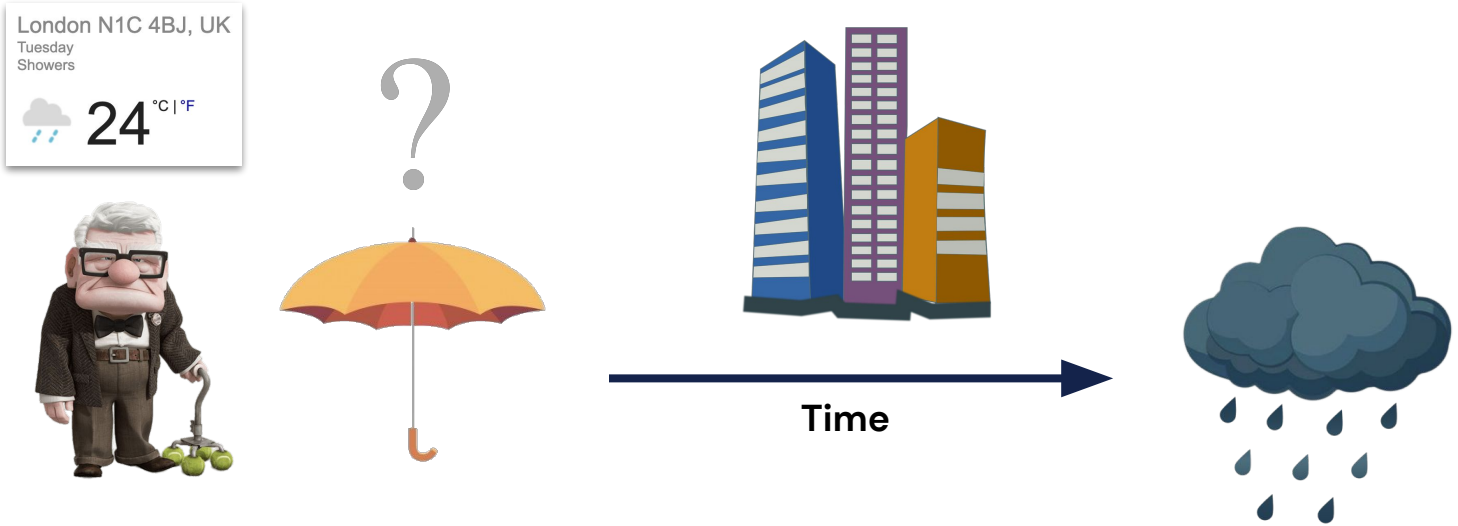
How did past actions influence future outcomes?

RL relies on MDP structure, and takes **time** as main proxy for credit relevance

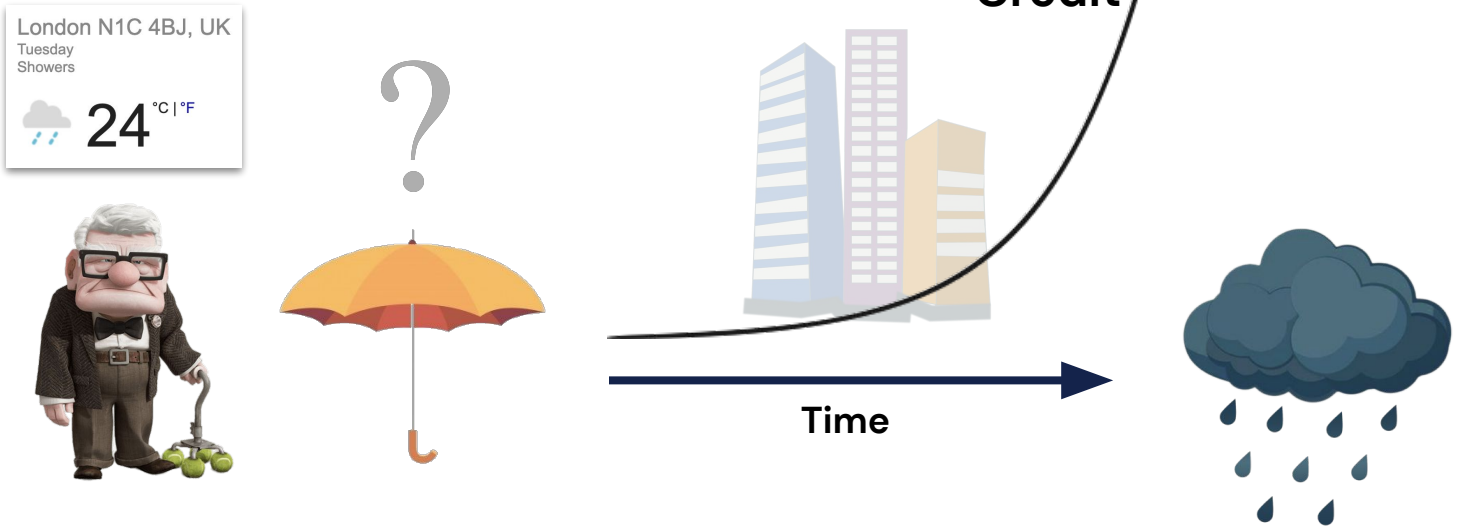
RL relies on MDP structure, and takes **time** as main proxy for credit relevance



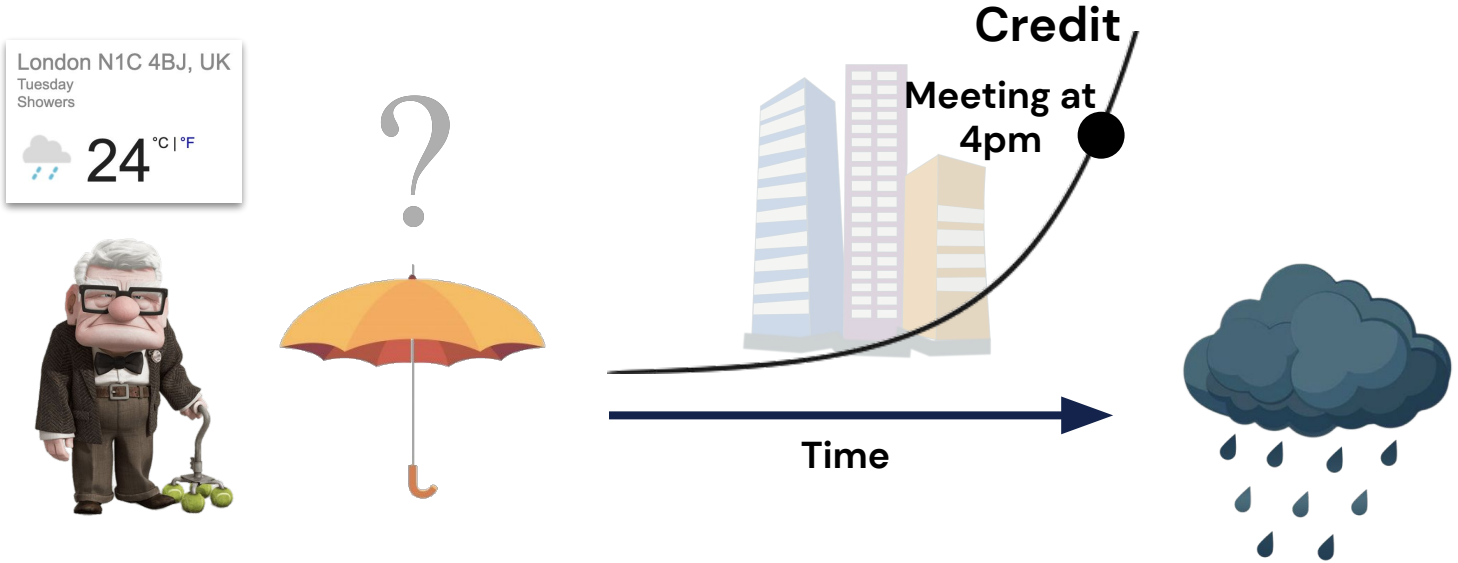
RL relies on MDP structure, and takes **time** as main proxy for credit relevance



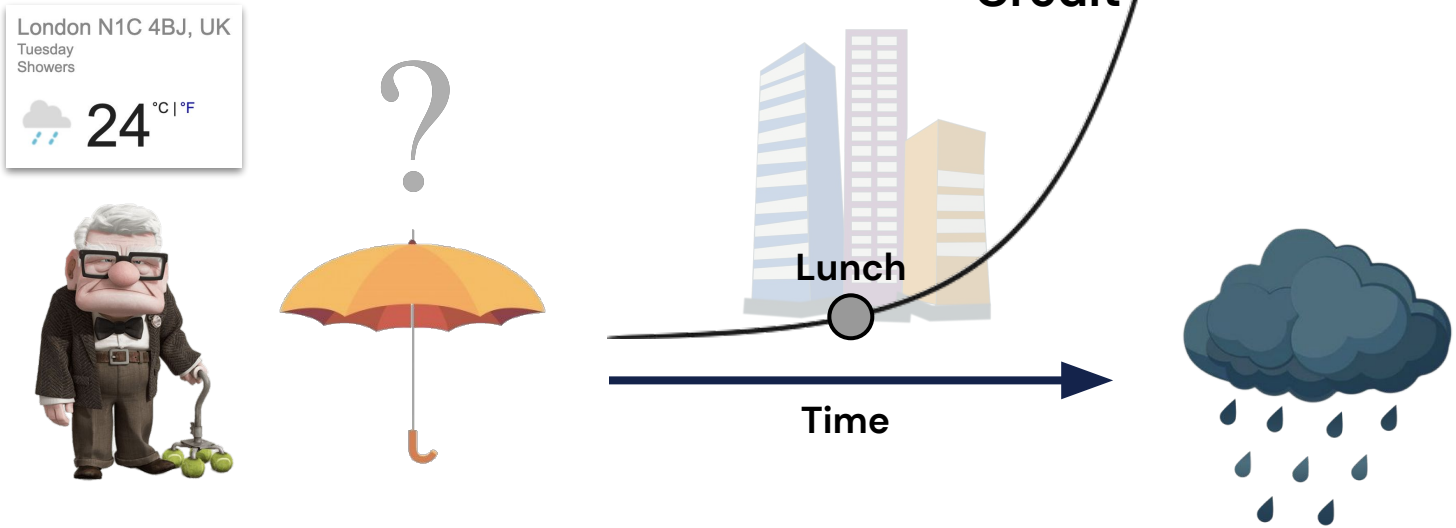
RL relies on MDP structure, and takes **time** as main proxy for credit relevance



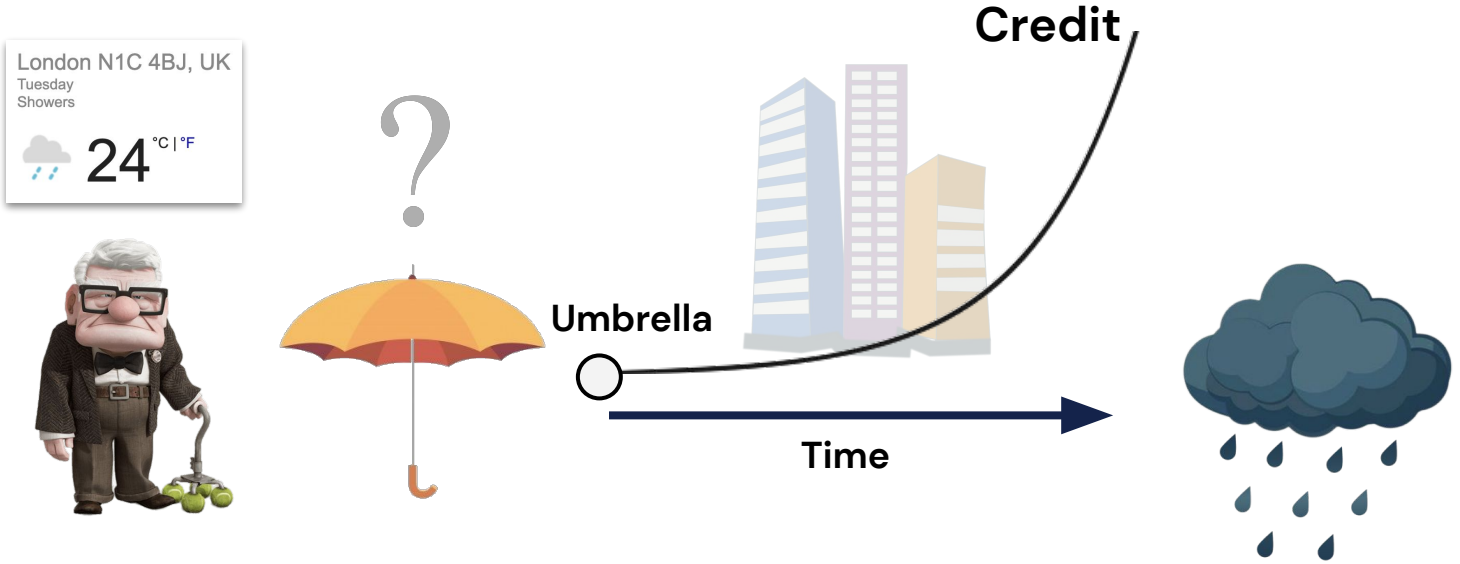
RL relies on MDP structure, and takes **time** as main proxy for credit relevance



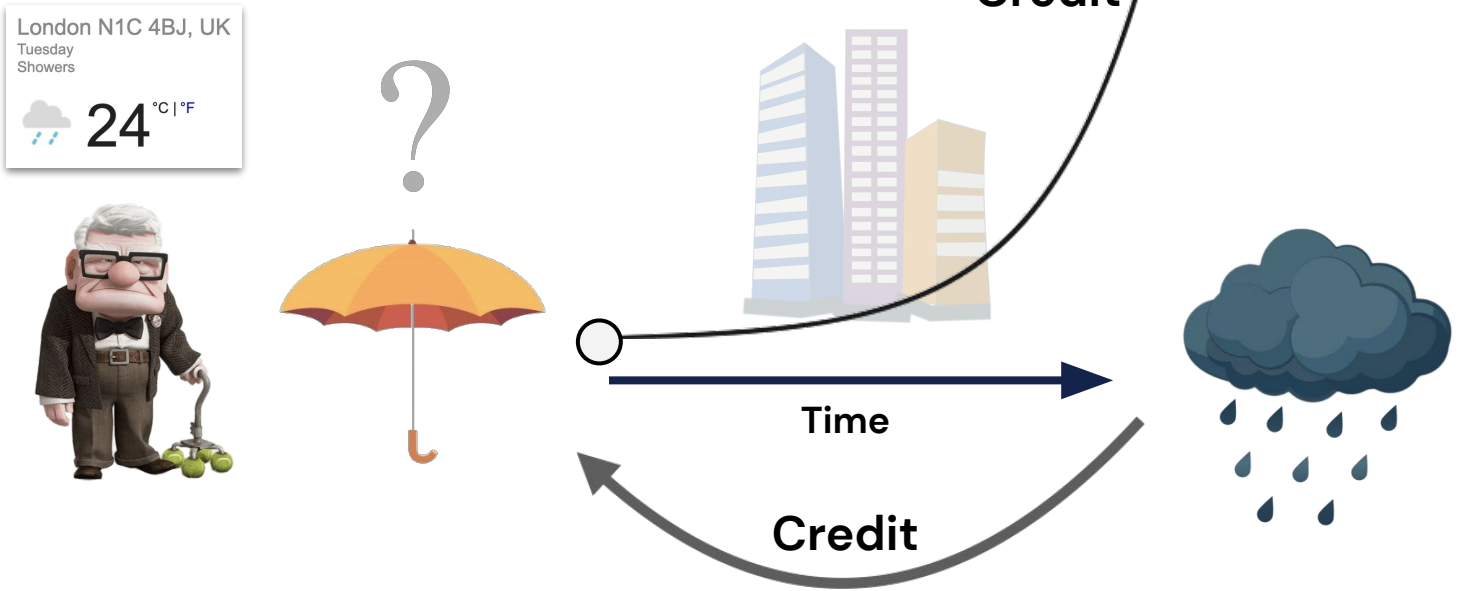
RL relies on MDP structure, and takes **time** as main proxy for credit relevance



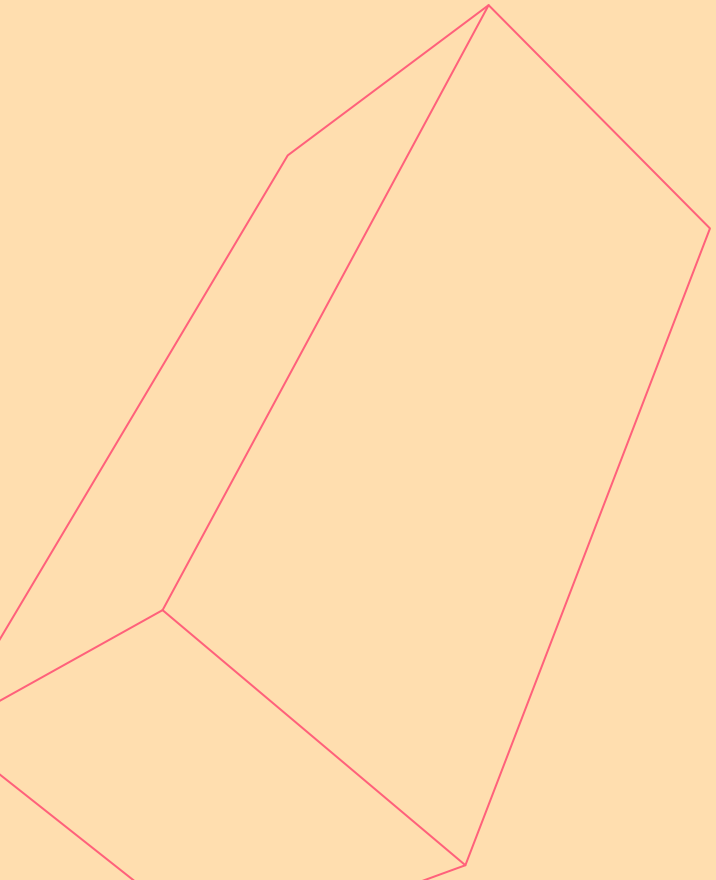
RL relies on MDP structure, and takes **time** as main proxy for credit relevance



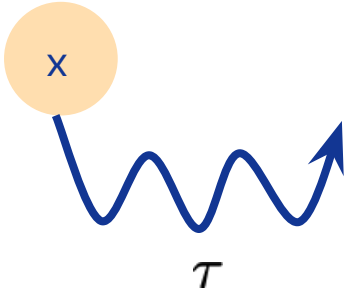
RL relies on MDP structure, and takes **time** as main proxy for credit relevance



DeepMind



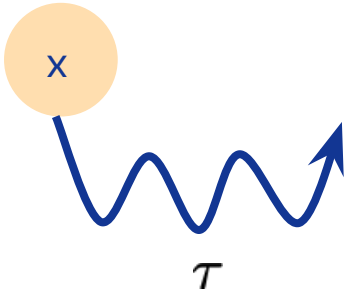
Instead of only relying on MDP assumptions, let's learn credit relevance explicitly!


X
τ

$$\mathbb{P}(a|x, f(\tau))$$

past action

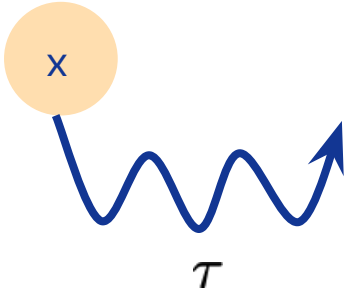
future outcome



$$\mathbb{P}(a|x, f(\tau))$$

past action

future outcome

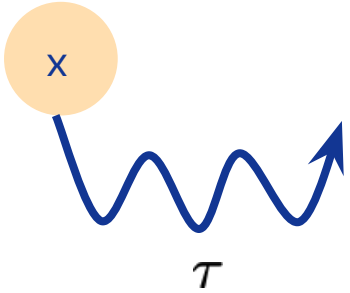


How did past actions influence future outcomes?

$$\mathbb{P}(a|x, f(\tau))$$

past action

future outcome

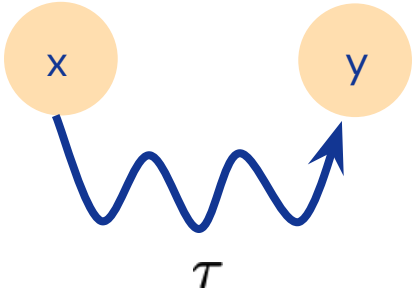


How did past actions influence future outcomes?

$$\mathbb{P}(a|x, f(\tau))$$

past action

future outcome

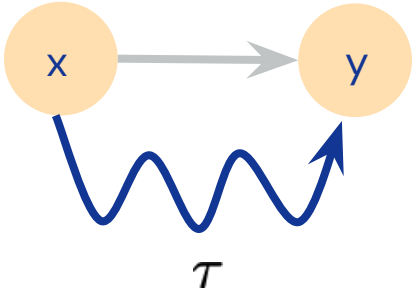


How did past actions influence future outcomes?

$$\mathbb{P}(a|x, f(\tau))$$

past action

future outcome

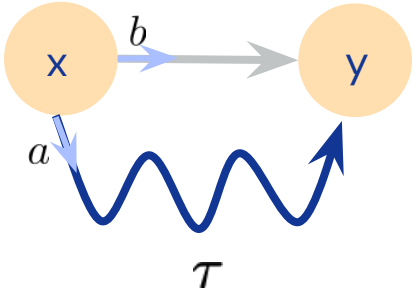


How did past actions influence future outcomes?

$$\mathbb{P}(a|x, f(\tau))$$

past action

future outcome

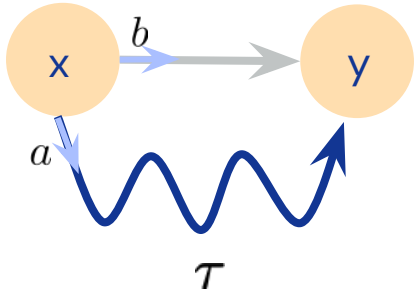


How did past actions influence future outcomes?

$$\mathbb{P}(a|x, f(\tau))$$

past action

future outcome



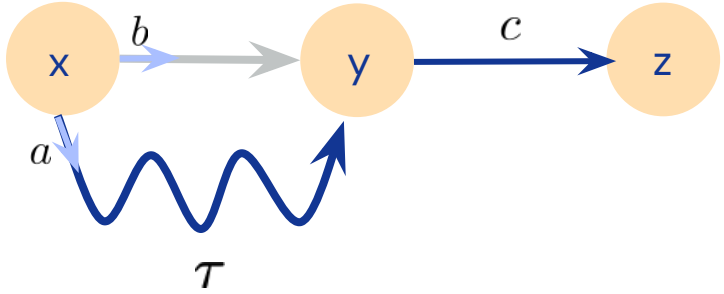
$$\mathbb{P}(b|x, y) > \mathbb{P}(a|x, y)$$

How did past actions influence future outcomes?

$$\mathbb{P}(a|x, f(\tau))$$

past action

future outcome



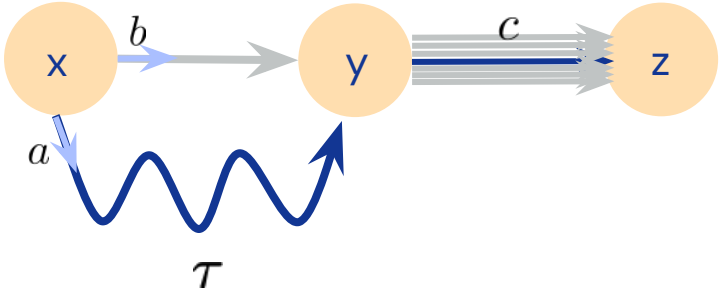
$$\mathbb{P}(b|x, y) > \mathbb{P}(a|x, y)$$

How did past actions influence future outcomes?

$$\mathbb{P}(a|x, f(\tau))$$

past action

future outcome



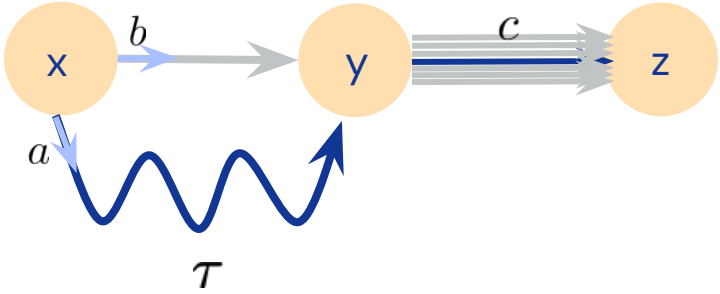
$$\mathbb{P}(b|x, y) > \mathbb{P}(a|x, y)$$

How did past actions influence future outcomes?

$$\mathbb{P}(a|x, f(\tau))$$

past action

future outcome



$$\mathbb{P}(b|x, y) > \mathbb{P}(a|x, y)$$
$$\mathbb{P}(c|y, z) = \pi(c|y)$$

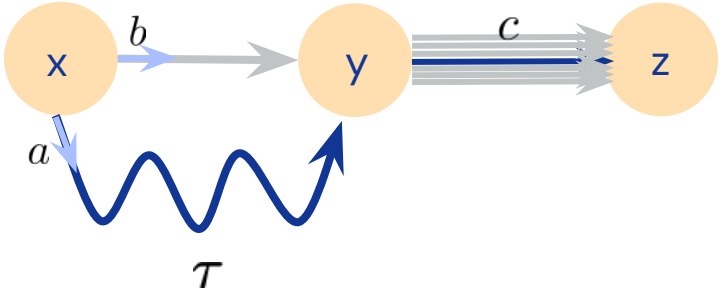
How did past actions influence future outcomes?

$$\mathbb{P}(a|x, f(\tau))$$

past action

future outcome

State $f(\tau) = X_k$



$$\mathbb{P}(b|x, y) > \mathbb{P}(a|x, y)$$

$$\mathbb{P}(c|y, z) = \pi(c|y)$$

How did past actions influence future outcomes?

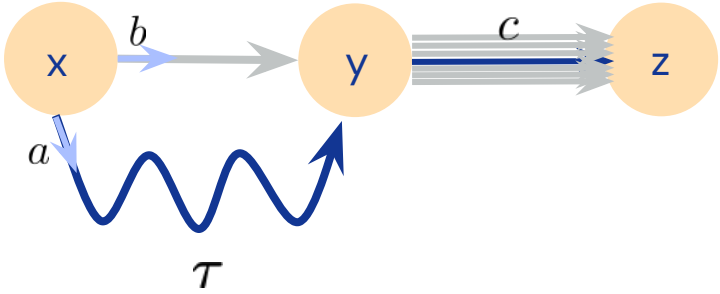
$$\mathbb{P}(a|x, f(\tau))$$

past action

future outcome

State $f(\tau) = X_k$

Return $f(\tau) = Z(\tau) = \sum_{k \geq 0} \gamma^k R_k$



$$\mathbb{P}(b|x, y) > \mathbb{P}(a|x, y)$$

$$\mathbb{P}(c|y, z) = \pi(c|y)$$

Hindsight Credit Assignment

$$Q^{\pi}(x,a) = r(x,a) + \mathbb{E}_{\tau \sim \mathcal{T}(x,\pi)}\left[\sum_{k \geq 1} \gamma^{k} \frac{\mathbb{P}(a|x, X_k)}{\pi(a|x)} R_k\right]$$

How relevant was ***a*** to get to a state X_k?

Hindsight Credit Assignment

$$Q^{\pi}(x,a) = r(x,a) + \mathbb{E}_{\tau \sim \mathcal{T}(x,\pi)} \left[\sum_{k \geq 1} \gamma^k \frac{\mathbb{P}(a|x, X_k)}{\pi(a|x)} R_k \right]$$

How relevant was ***a*** to get to a state X_k?

$$Q^{\pi}(x,a) = \mathbb{E}_{\tau \sim \mathcal{T}(x,\pi)} \left[\frac{\mathbb{P}(a|x, Z(\tau))}{\pi(a|x)} Z(\tau) \right]$$

How relevant was ***a*** to achieve the return ***Z***?

Hindsight Credit Assignment

$$Q^{\pi}(x,a) = r(x,a) + \mathbb{E}_{\tau \sim \mathcal{T}(x,\pi)}\left[\sum_{k \geq 1} \gamma^{k} \frac{\mathbb{P}(a|x, X_k)}{\pi(a|x)} R_k\right]$$

How relevant was ***a*** to get to a state X_k?

$$Q^{\pi}(x,a) = \mathbb{E}_{\tau \sim \mathcal{T}(x,\pi)}\left[\frac{\mathbb{P}(a|x, Z(\tau))}{\pi(a|x)} Z(\tau)\right]$$

How relevant was ***a*** to achieve the return ***Z***?

Hindsight Credit Assignment

HCA Algorithms: ***Learn the hindsight distribution P, and use it to better estimate value functions or policy gradients***

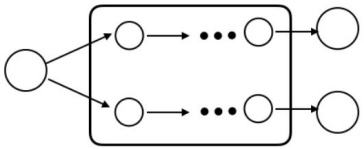
$$Q^\pi(x,a) = r(x,a) + \mathbb{E}_{\tau\sim\mathcal{T}(x,\pi)}\left[\sum_{k\geq 1}\gamma^k \frac{\mathbb{P}(a|x,X_k)}{\pi(a|x)} R_k\right]$$

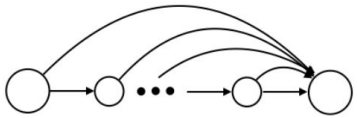
How relevant was ***a*** to get to a state X_k?

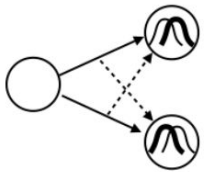
$$Q^\pi(x,a) = \mathbb{E}_{\tau\sim\mathcal{T}(x,\pi)}\left[\frac{\mathbb{P}(a|x,Z(\tau))}{\pi(a|x)} Z(\tau)\right]$$

How relevant was ***a*** to achieve the return ***Z***?

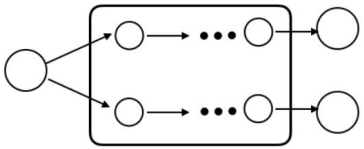
Experiments

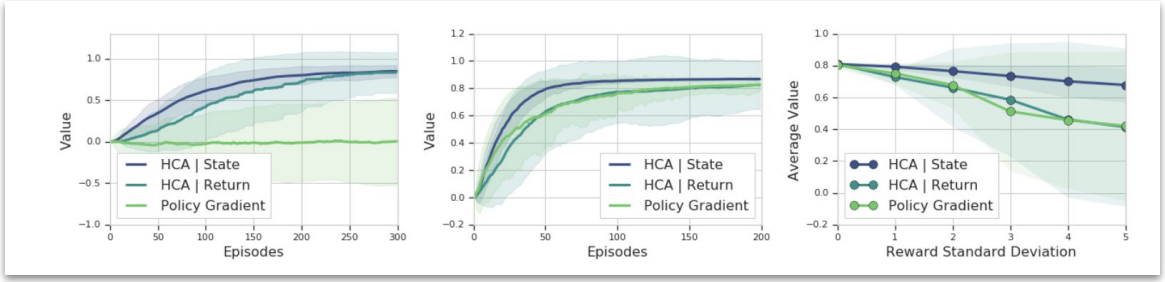


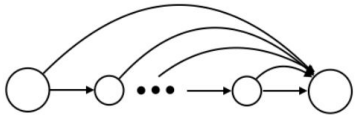


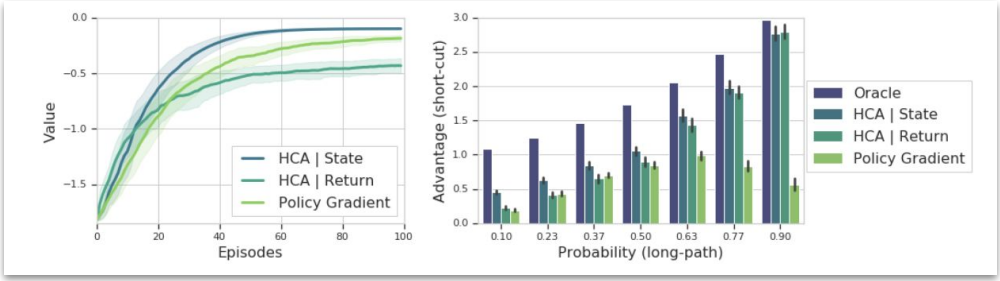


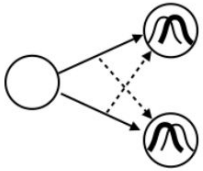
Experiments

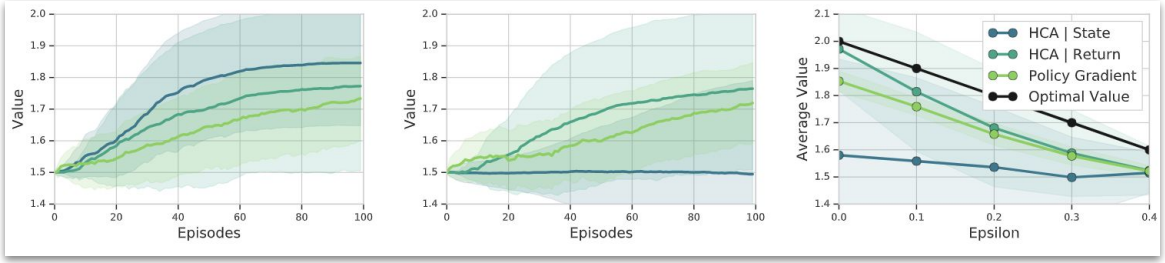












DeepMind

Thank you for your attention!

Poster #204 :)