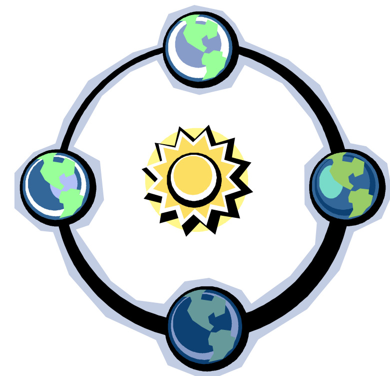
Better algorithms for sleeping experts & bandits

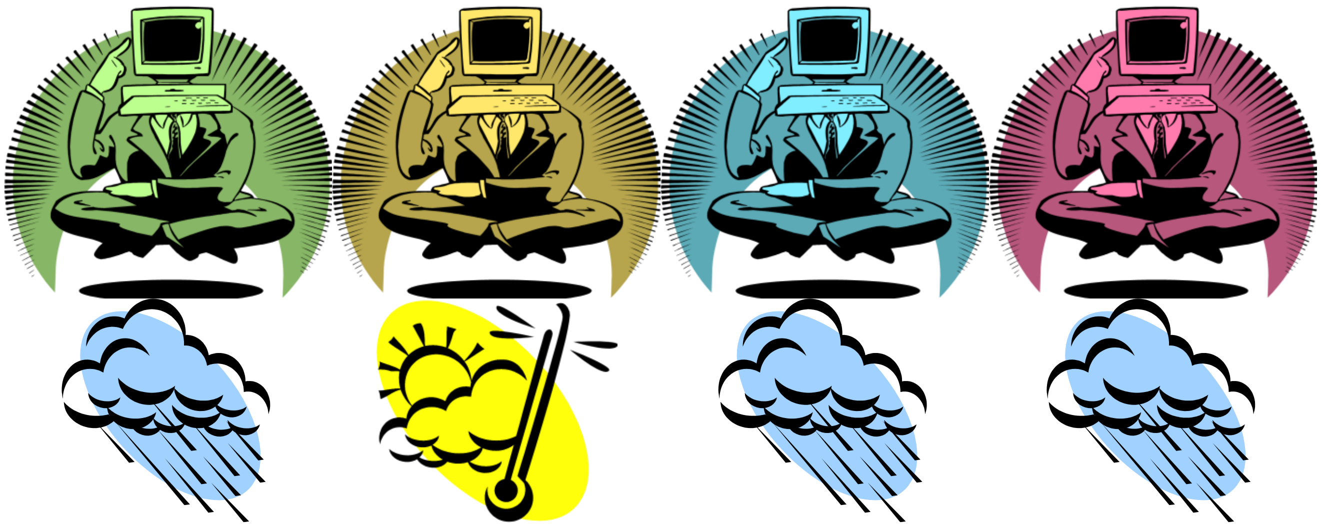
Gergely Neu
INRIA, SequeL team

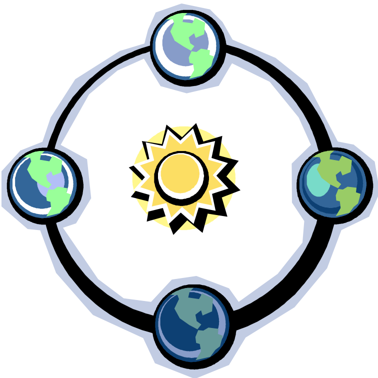
joint work with Michal Valko, to appear at NIPS 2014

Prediction with expert advice

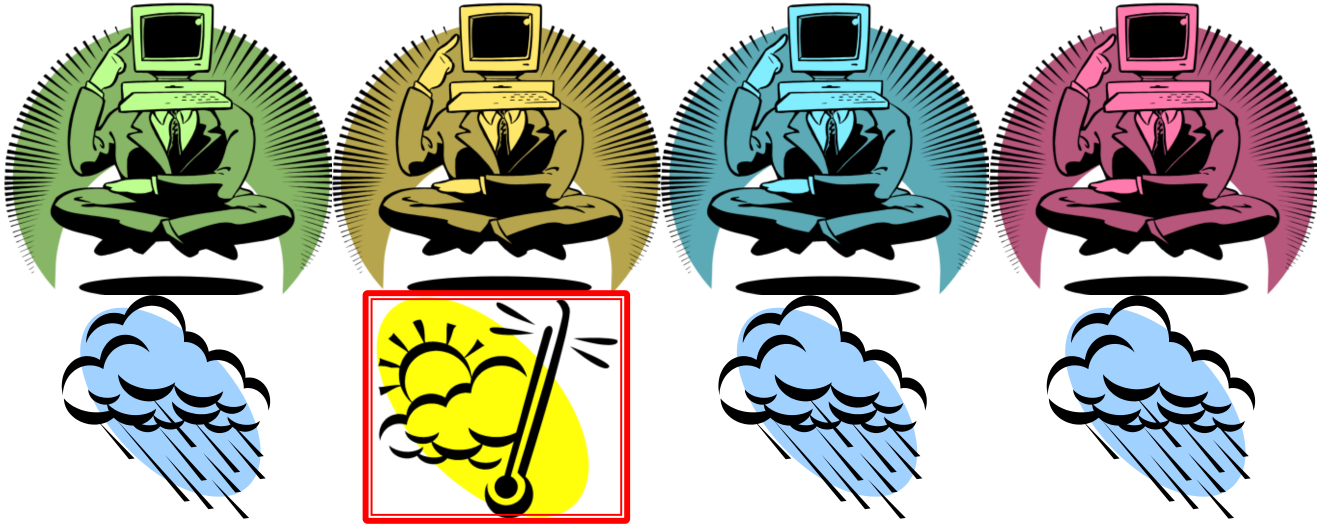


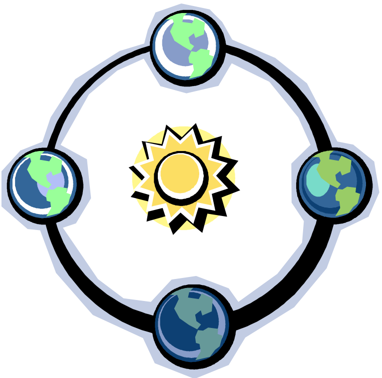
Prediction with expert advice



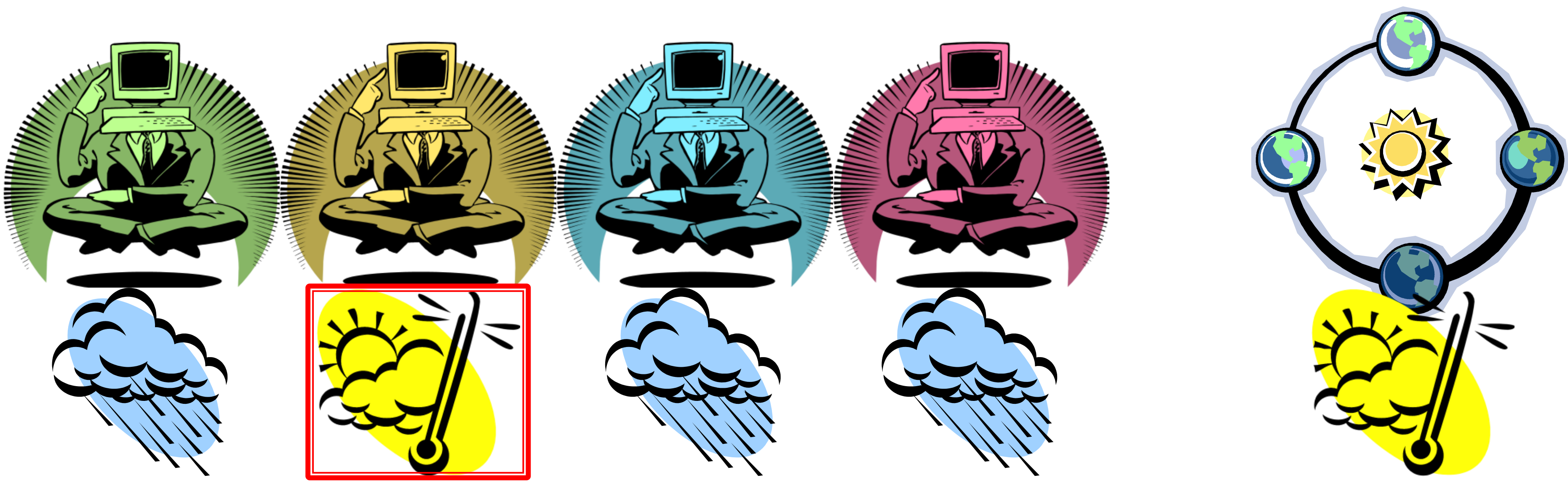


Prediction with expert advice

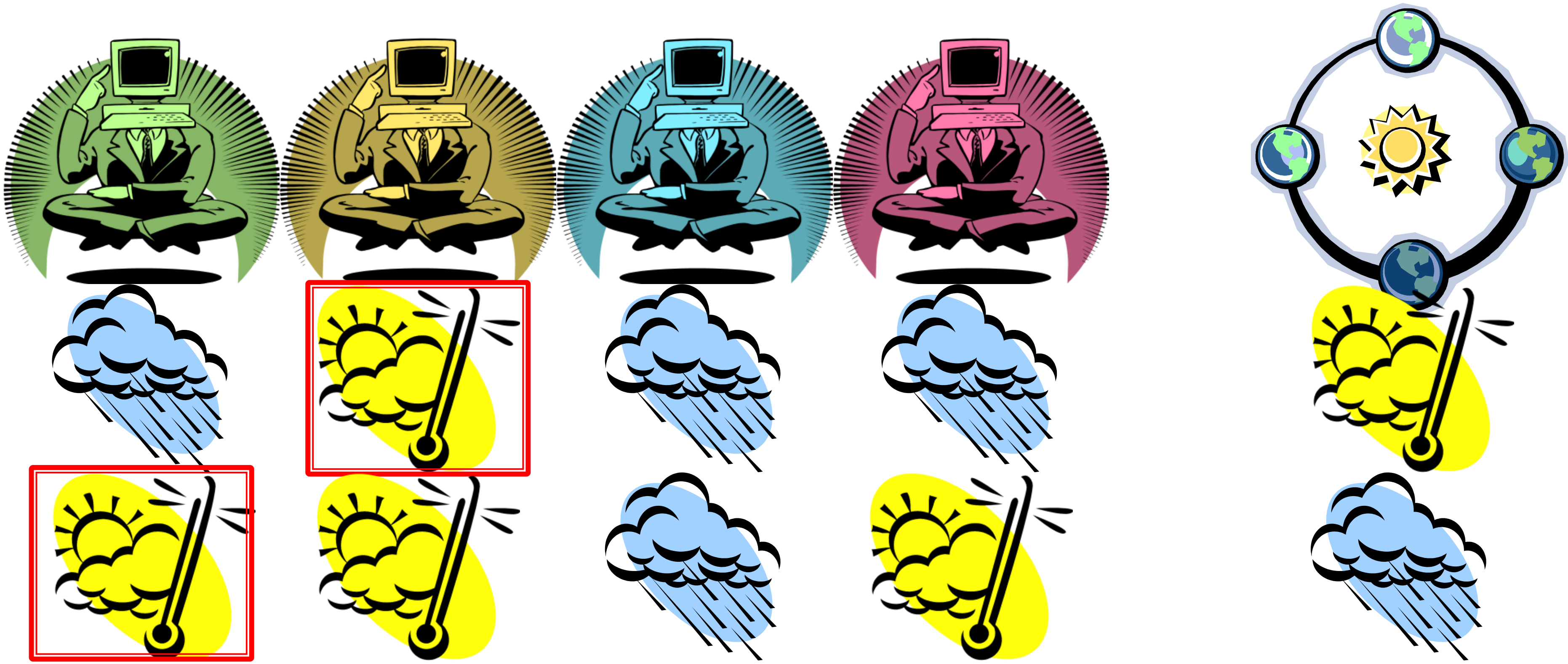




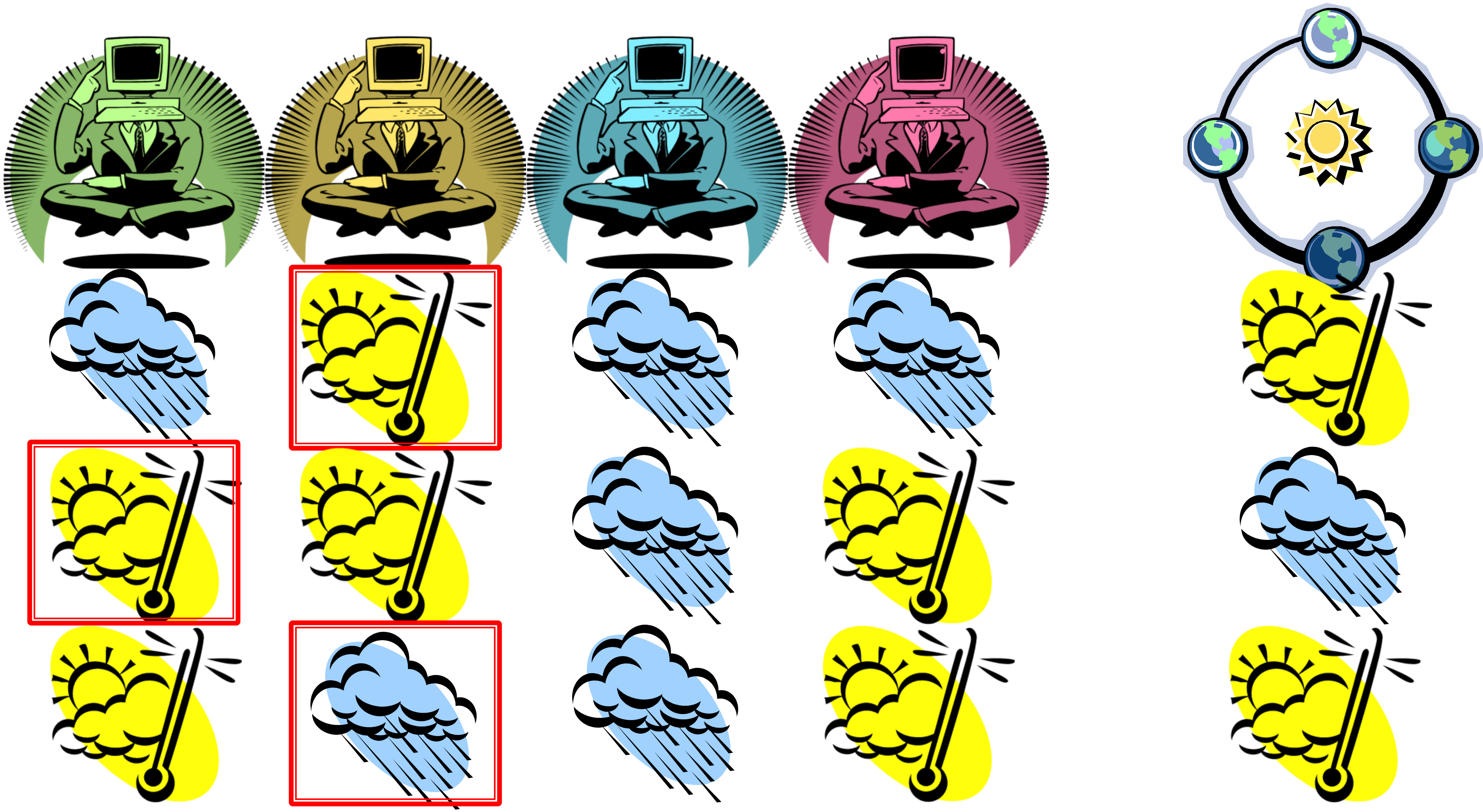
Prediction with expert advice



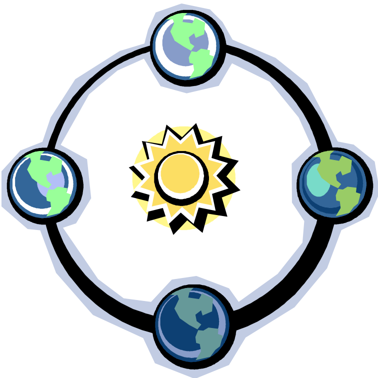
Prediction with expert advice



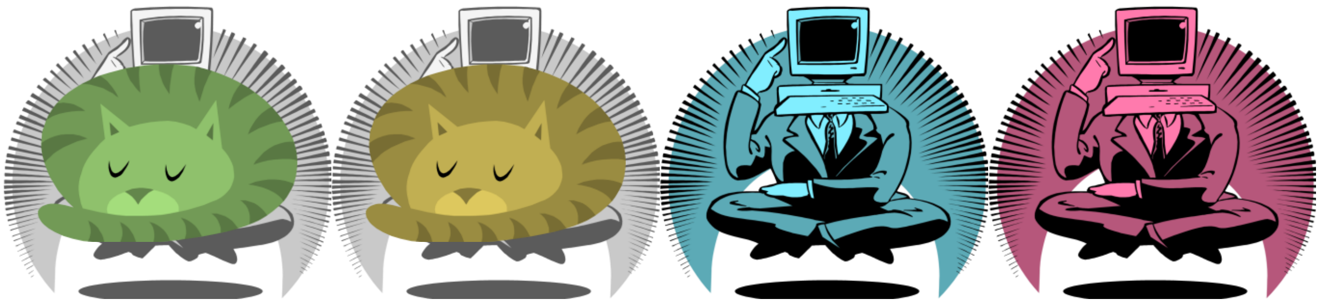
Prediction with expert advice

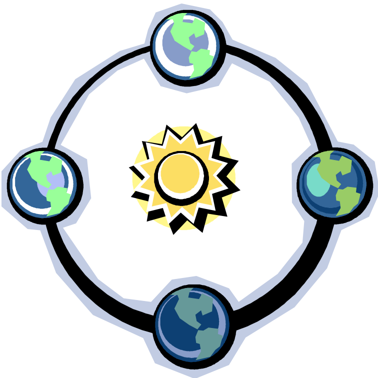


Prediction with sleeping experts

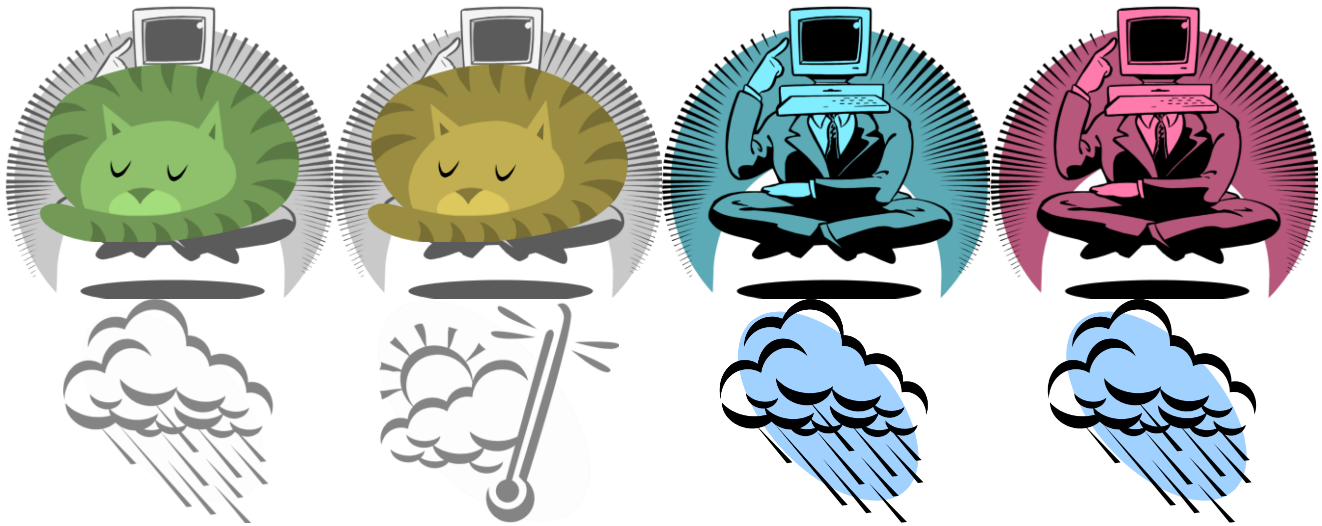


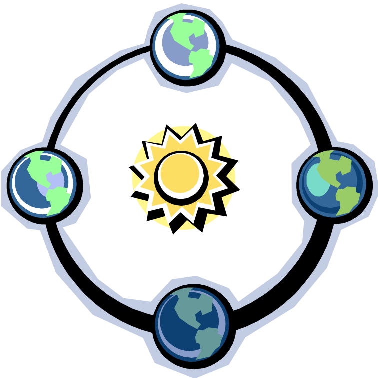
Prediction with sleeping experts



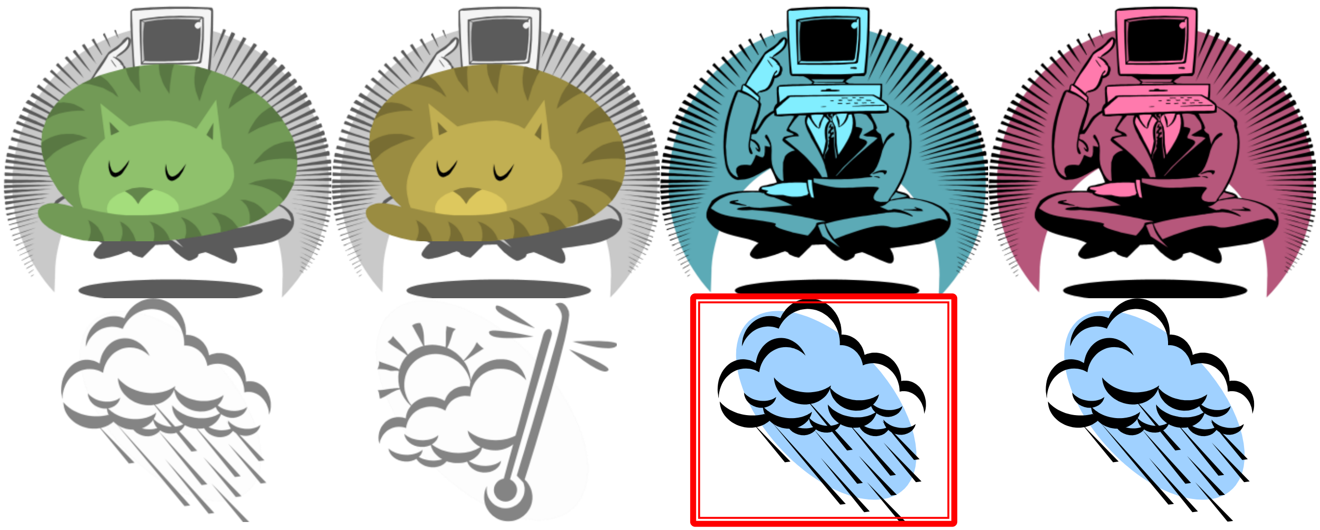


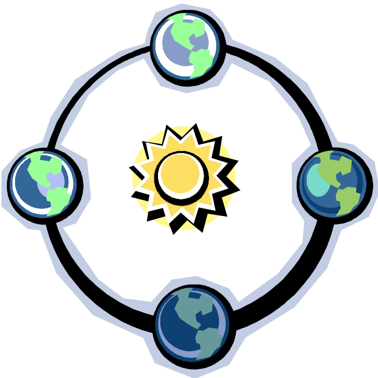
Prediction with sleeping experts



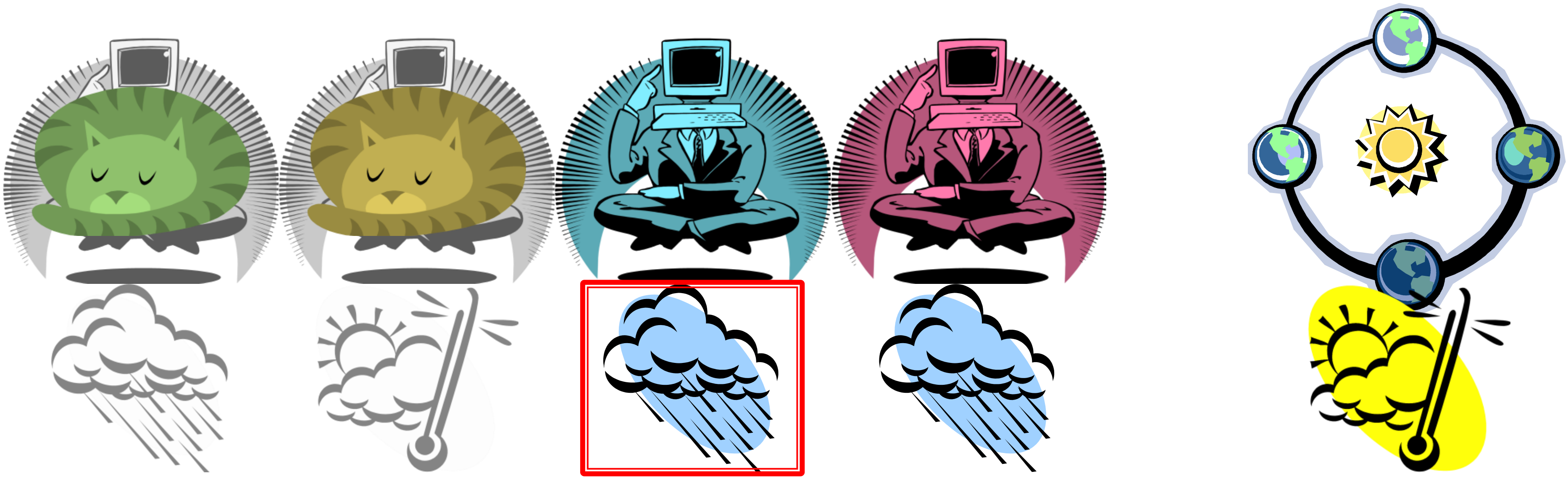


Prediction with sleeping experts

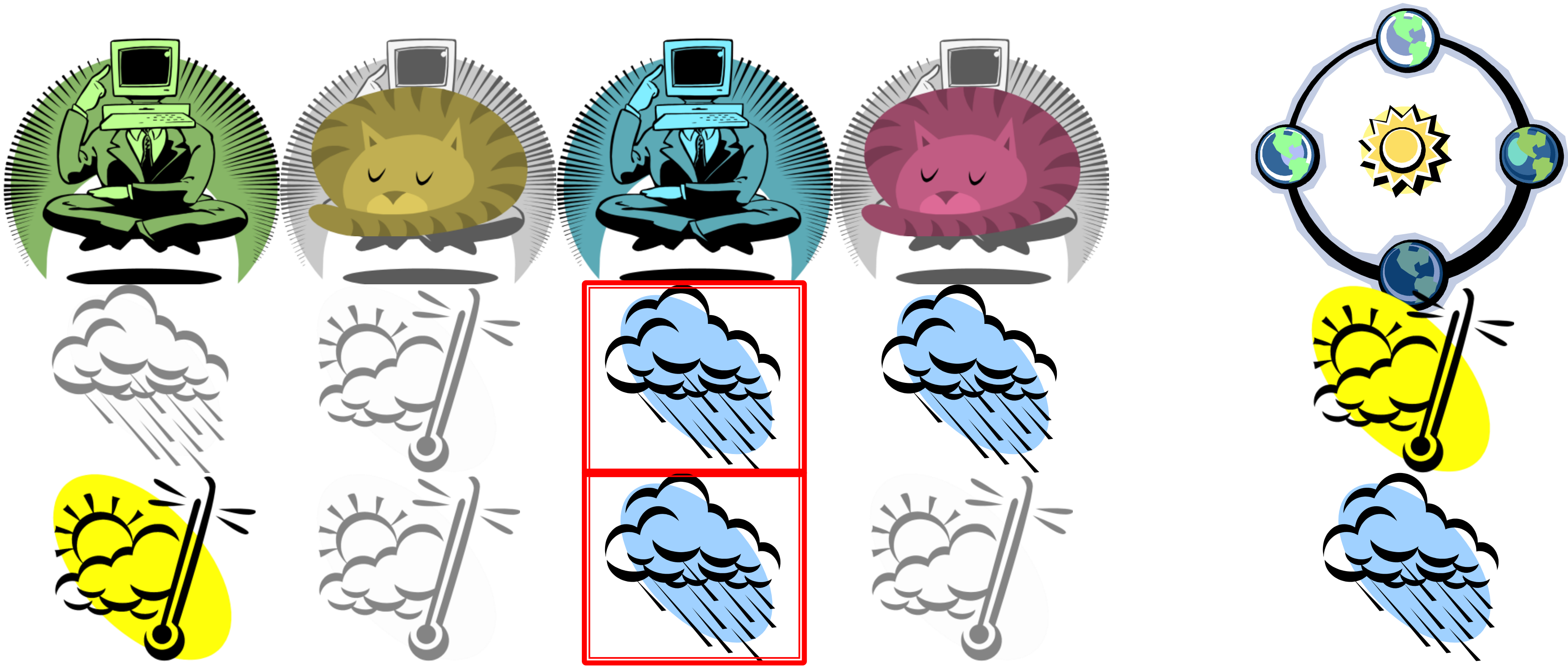




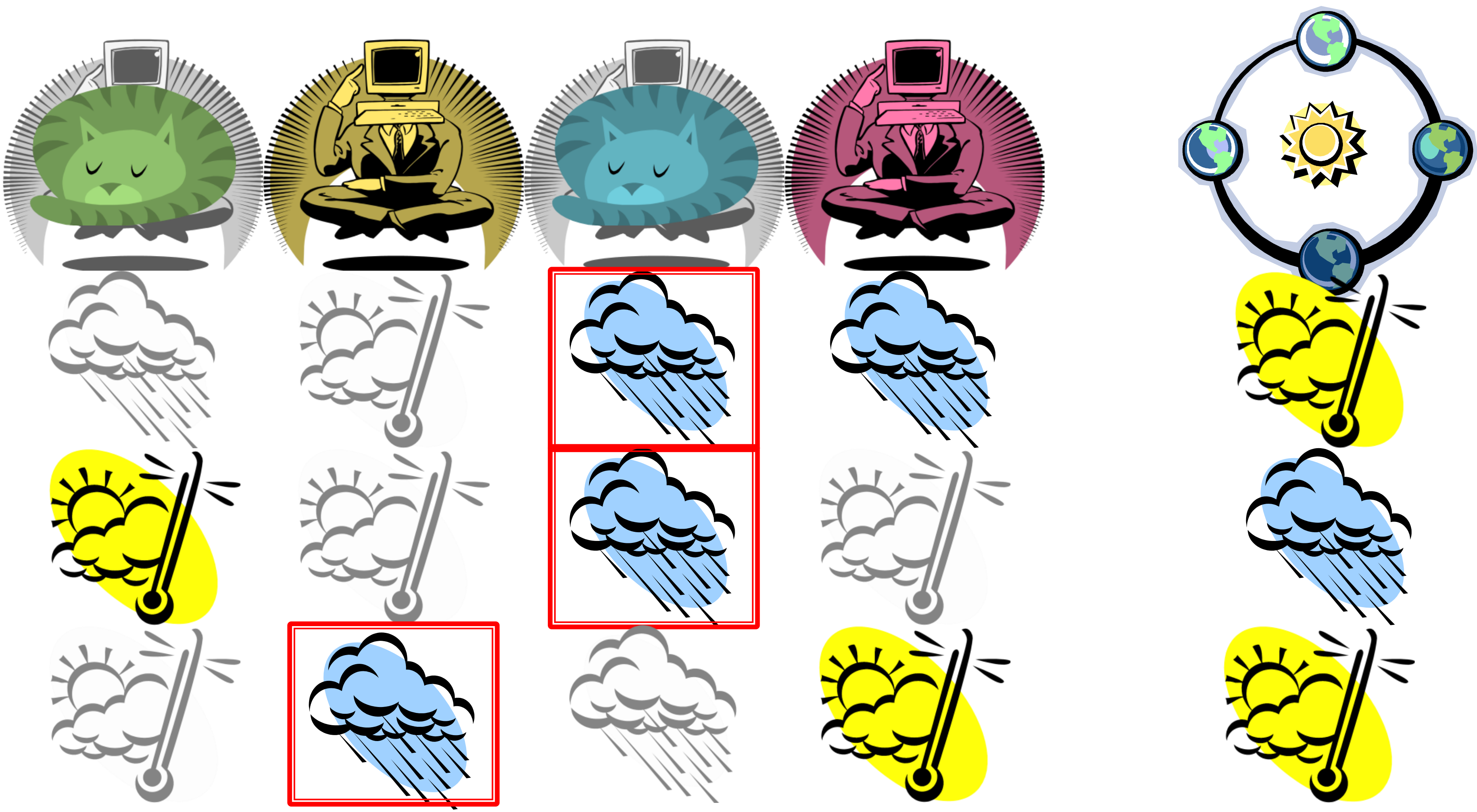
Prediction with sleeping experts



Prediction with sleeping experts



Prediction with sleeping experts



More formally…

Parameters: set of N experts

In each round $t = 1,2, \ldots, T$

- Environment chooses losses $\ell_{t,i} \in [0,1]$ for all experts
- Environment chooses the set of available experts $S_t \in \{1,2, \ldots, N\}$
- Learner picks distribution $\mathbf{p}_t$ on available experts
- Learner suffers loss $\mathbf{p}_t^\top \mathbf{l}_t$

Regret definition

- Usual notion of regret:

$$R_T = \sum_{t=1}^{T} \mathbf{p}_t^\top \mathbf{l}_t - \min_{i \in \{1,\ldots,N\}} \sum_{t=1}^{T} \ell_{t,i}$$

Regret definition

- Usual notion of regret:

$$R_T = \sum_{t=1}^{T} \mathbf{p}_t^\top \mathbf{l}_t - \min_{i \in \{1,\ldots,N\}} \sum_{t=1}^{T} \ell_{t,i}$$

This comparator is pointless!

Regret definition

- Usual notion of regret:

$$R_T = \sum_{t=1}^{T} \mathbf{p}_t^\top \mathbf{l}_t - \min_{i \in \{1,\ldots,N\}} \sum_{t=1}^{T} \ell_{t,i}$$

This comparator is pointless!

- We should actually compete with policies of the form $\pi: 2^{[N]} \to N$ such that $\pi(S) \in S$!

Regret definition

- Regret against policy class Π:

$$R_T = \sum_{t=1}^{T} \mathbf{p}_t^\top \mathbf{l}_t - \min_{\pi \in \Pi} \sum_{t=1}^{T} \ell_{t,\pi(s_t)}$$

Previous results

	IID availability	Adversarial availability
IID losses	(that's kind of trivial)	Kleinberg et al. (2008): $R_T = \Theta\left(\sqrt{TN \log N}\right)$
Adversarial losses	Kanade et al. (2009): $R_T = O\left(\sqrt{T \log N}\right)$	Kleinberg et al. (2008): $R_T = \Theta\left(N\sqrt{T \log N}\right)$

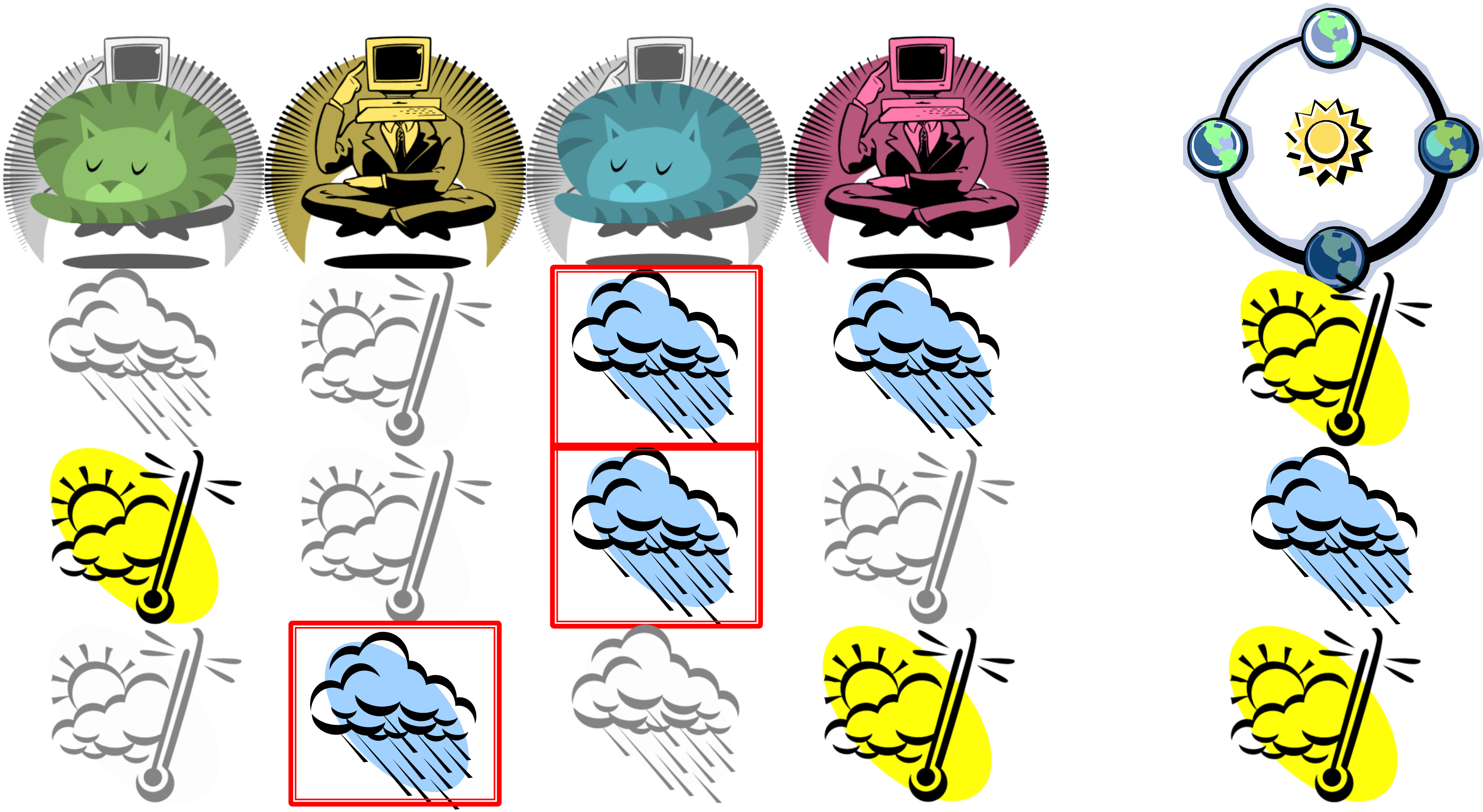
Previous results

	IID availability	Adversarial availability
IID losses	(that's kind of trivial)	Kleinberg et al. (2008): $R_T = \Theta\left(\sqrt{TN \log N}\right)$
Adversarial losses	Kanade et al. (2009): $R_T = O\left(\sqrt{T \log N}\right)$	Kleinberg et al. (2008): $\Omega\left(N\sqrt{T \log N}\right)$ HARD

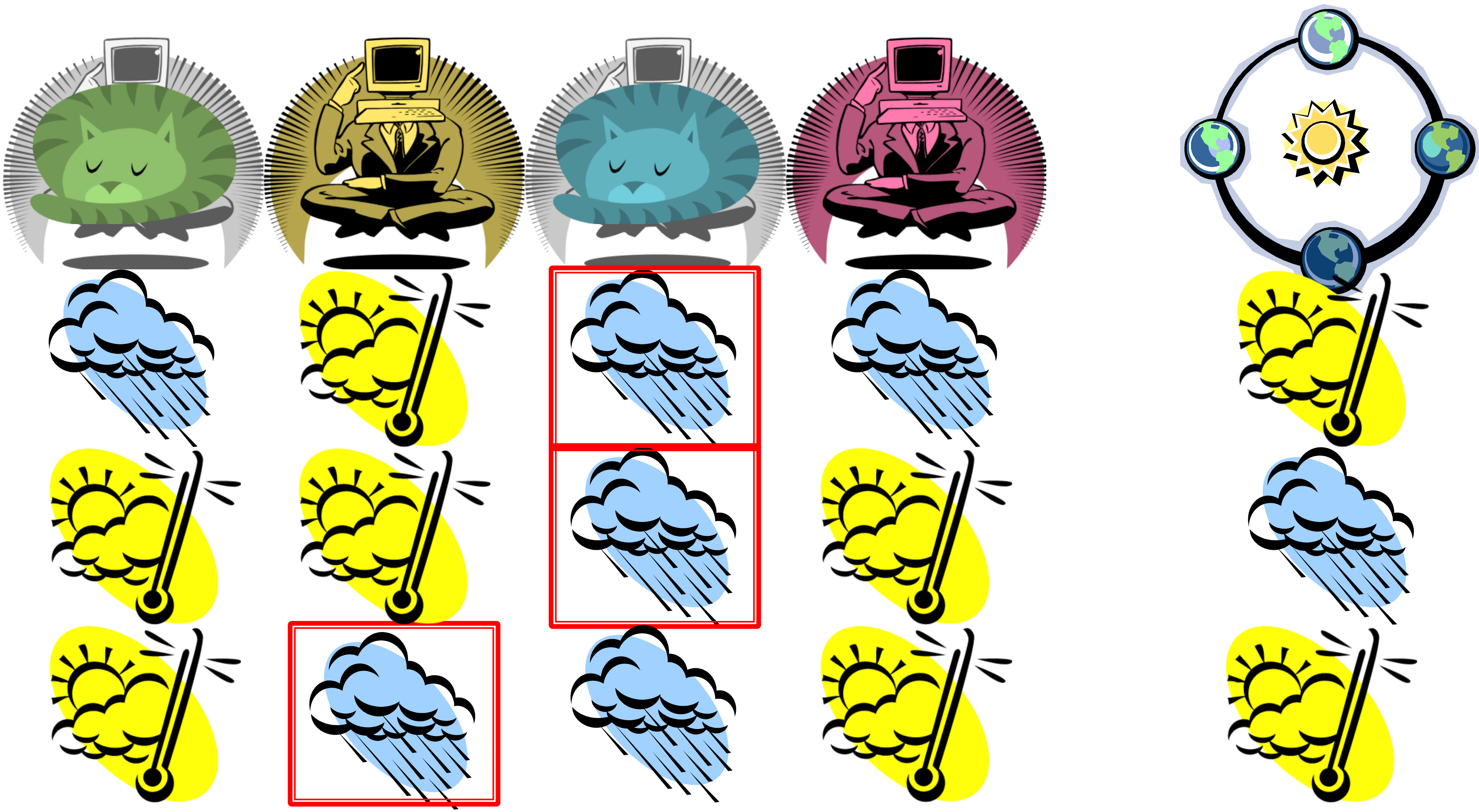
Previous results

	IID availability	Adversarial availability
IID losses	(that's kind of trivial)	Kleinberg et al. (2008): $R_T = \Theta\left(\sqrt{TN \log N}\right)$
Adversarial losses	CHEAT Kanade et al. (2009): $R_T = O\left(\sqrt{T \log N}\right)$	HARD Kleinberg et al. (2008): $\Omega\left(N\sqrt{T}\log N\right)$

Where's the cheat?



Where's the cheat?



The cheat – now formally

- Kanade et al. assume that $\ell_{t,i}$ is observed for all $i \in [N]$!

The cheat – now formally

- Kanade et al. assume that $\ell_{t,i}$ is observed for all $i \in [N]$!
- A more realistic assumption:

Observe $\ell_{t,i}$ only for $i \in S_t \subseteq [N]$

Algorithm: Follow the Perturbed Leader

Initialization: let $L_{t,i} = 0$ for all $i \in [N]$

For all rounds $t = 1,2, \dots, T$:

- Observe $S_t \subseteq [N]$
- Draw perturbations $Z_{t,i} \sim \text{Exp}(\eta)$ for all $i \in S_t$
- Play expert

$$I_t = \arg\min_{i \in S_t} \left(L_{t-1,i} - Z_{t,i} \right)$$

- Observe feedback and set for all $i \in N$

$$L_{t,i} = L_{t-1,i} + \ell_{t,i}$$

Algorithm: Follow the Perturbed Leader

Initialization: let $\hat{L}_{t,i} = 0$ for all $i \in [N]$

For all rounds $t = 1,2, \dots, T$:

- Observe $S_t \subseteq [N]$
- Draw perturbations $Z_{t,i} \sim \mathrm{Exp}(\eta)$ for all $i \in S_t$
- Play expert

$$I_t = \arg\min_{i \in S_t} \left(\hat{L}_{t-1,i} - Z_{t,i} \right)$$

- Observe feedback and set for all $i \in N$

$$\hat{L}_{t,i} = \hat{L}_{t-1,i} + \hat{\ell}_{t,i}$$

Loss estimation

- Assume IID availability:

$$S_t \sim Q \quad \forall t = 1,2,\dots,T$$

- Then we can set $q_i = \mathbf{P}[i \in S_t]$ for all $i \in [N]$
- Losses can be estimated as

$$\hat{\ell}_{t,i} = \begin{cases} \dfrac{\ell_{t,i}}{q_i}, & \text{if } i \text{ is observed} \\ 0, & \text{otherwise} \end{cases}$$

Loss estimation

- Assume IID availability:
$$S_t \sim Q \quad \forall t = 1,2,\dots,T$$
- Then we can set $q_i = \mathbf{P}[i \in S_t]$ for all $i \in [N]$
- Losses can be estimated as
$$\hat{\ell}_{t,i} = \begin{cases} \dfrac{\ell_{t,i}}{q_i}, & \text{if } i \text{ is observed} \\ 0, & \text{otherwise} \end{cases}$$

Unbiased:
$\mathbf{E}\left[\hat{\ell}_{t,i}\right] = \ell_{t,i}$

Loss estimation

- Assume IID availability:

$$S_t \sim Q \quad \forall t = 1,2,\ldots,T$$

- Then we can set $q_i = \mathbf{P}[i \in S_t]$ for all $i \in [N]$
- Losses can be esti

$$\widehat{\ell}_{t,i} = \begin{cases} \dfrac{\ell_{t,i}}{q_i}, & \text{if } i \text{ is observed} \\ 0, & \text{otherwise} \end{cases}$$

But the q_i's are unknown!!

Unbiased:
$\mathbf{E}\left[\widehat{\ell}_{t,i}\right] = \ell_{t,i}$

Loss estimation – the bad way

Idea:

- Use K samples to estimate Q!
- Compute estimates of q_i!
- Obtain low-bias reward estimates!

Loss estimation – the bad way

Idea:

- Use K samples to estimate Q!
- Compute estimates of q_i!
- Obtain low-bias reward estimates!

Bad news:

- Regret becomes $O\left(T^{3/4}\right)$
- Can fail horribly for large action sets

Loss estimation – the right way

- Observe that the downtime is a geometric RV!

Loss estimation – the right way

- Observe that the downtime is a geometric RV!

t

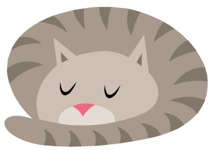
Arm i

Loss estimation – the right way

- Observe that the downtime is a geometric RV!

t $\quad$ $t+1$

Arm i

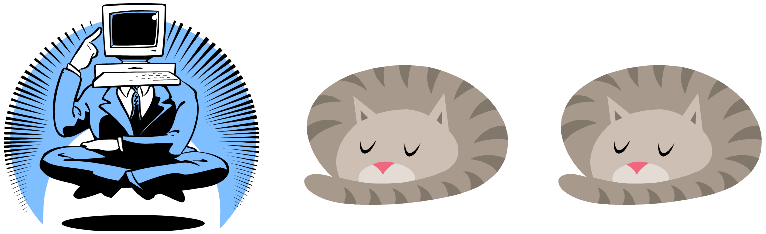


Loss estimation – the right way

- Observe that the downtime is a geometric RV!

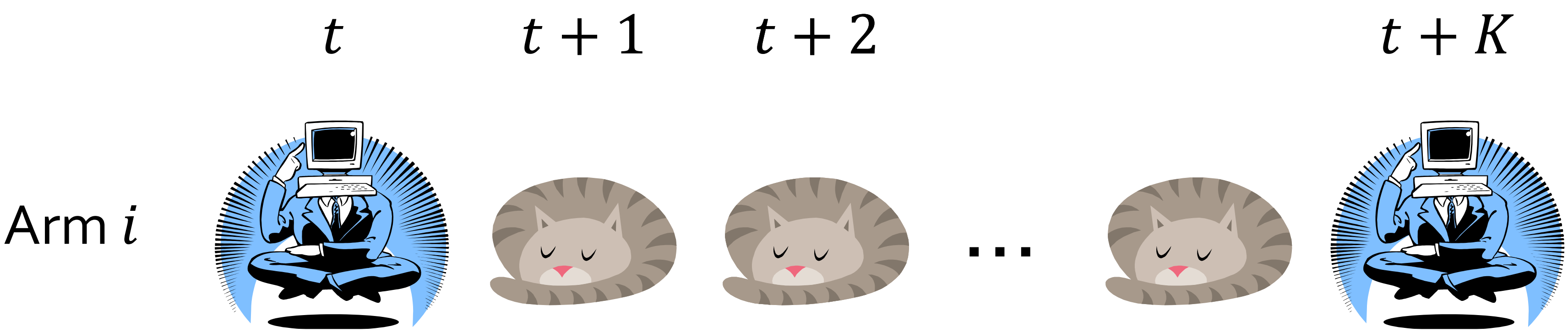
t $\quad$ $t+1$ $\quad$ $t+2$

Arm i



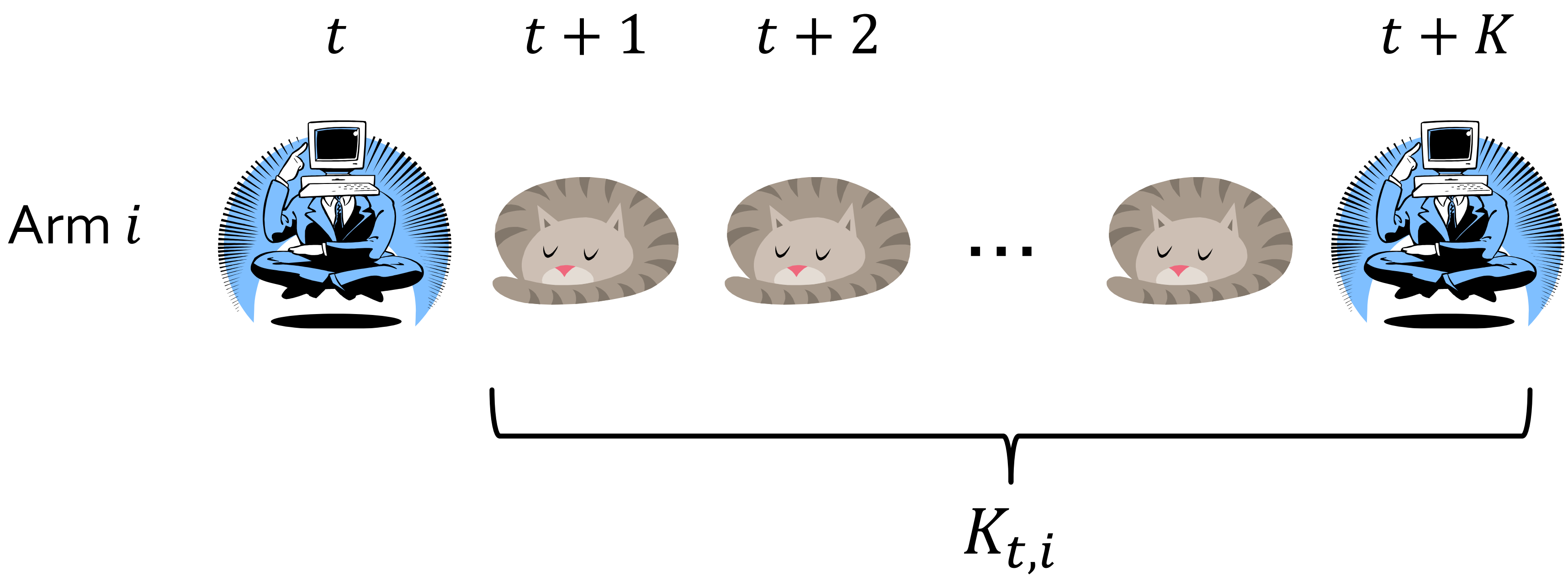
Loss estimation – the right way

- Observe that the downtime is a geometric RV!



Loss estimation – the right way

- Observe that the downtime is a geometric RV!

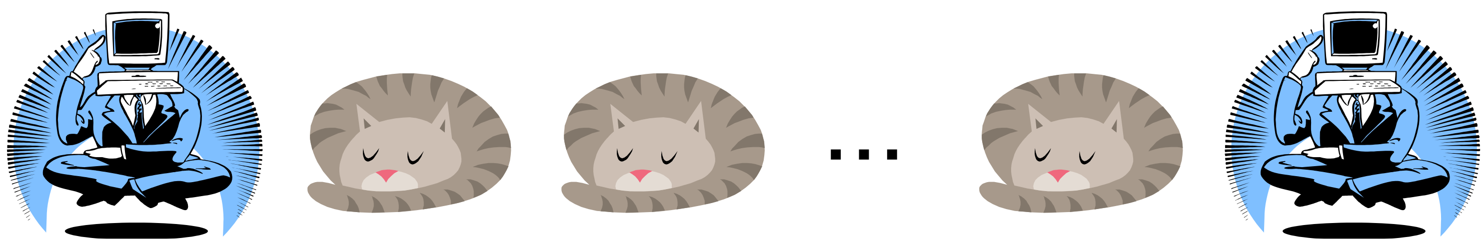


Loss estimation – the right way

- Observe that the downtime is a geometric RV!

t $\quad t+1 \quad t+2 \quad \dots \quad t+K$

Arm i



$K_{t,i}$

$$\mathbf{E}_t\left[K_{t,i}\right] = \frac{1}{q_i}$$

Loss estimation – the right way

- Observe that the … c RV!

K

Arm i

Estimate losses as

$$\hat{\ell}_{t,i} = \begin{cases} \ell_{t,i} K_{t,i}, & \text{if } i \text{ is observed} \\ 0, & \text{otherwise} \end{cases}$$

$K_{t,i}$

$$\mathbf{E}_t\left[K_{t,i}\right] = \frac{1}{q_i}$$

Main result

Theorem 1
Assuming IID expert availability, the expected regret of FPL fed with loss estimates $\{\hat{\ell}_{t,i}\}$ satisfies

$$R_T = O\left(\sqrt{TN \log N}\right)$$

Main result

Theorem 1
Assuming IID expert availability, the expected regret of FPL fed with loss estimates $\{\hat{\ell}_{t,i}\}$ satisfies

$$R_T = O\left(\sqrt{TN \log N}\right)$$

- This is worse by a factor of $\sqrt{N}$ than the bound of Kanade et al. (2009)…

Main result

Theorem 1
Assuming IID expert availability, the expected regret of FPL fed with loss estimates $\{\hat{\ell}_{t,i}\}$ satisfies

$$R_T = O\left(\sqrt{TN \log N}\right)$$

- This is worse by a factor of $\sqrt{N}$ than the bound of Kanade et al. (2009)…
- …but we didn't cheat!

Lower bound

Theorem 1: $R_T = O\left(\sqrt{TN \log N}\right)$

Theorem 2
Assuming IID expert availability, no algorithm can achieve better regret than

$$R_T = \Omega\left(\sqrt{TN}\right)$$

Extensions: large action spaces

- Assume that
  - each expert $i \in [N]$ is associated with a binary vector $\mathbf{v}(i) \in \{0,1\}^d$
  - losses are described by a loss vector $\mathbf{l}_t \in [0,1]^d$
  - loss of expert i in round t is given as $\mathbf{v}(i)^\top \mathbf{l}_t \leq m$

Extensions: large action spaces

- Assume that
  - each expert $i \in [N]$ is associated with a binary vector $\mathbf{v}(i) \in \{0,1\}^d$
  - losses are described by a loss vector $\mathbf{l}_t \in [0,1]^d$
  - loss of expert i in round t is given as $\mathbf{v}(i)^\top \mathbf{l}_t \leq m$

Theorem 3
Assuming IID expert availability, the expected regret of the combinatorial extension of FPL is

$$R_T = O\left(m\sqrt{dT \log d}\right)$$

Extensions: bandit feedback

- So far: assume we observe $\ell_{t,i}$ for all $i \in S_t$
- Now: assume we only observe the loss ℓ_{t,I_t}
- Using a simple extension of FPL, we prove

Extensions: bandit feedback

- So far: assume we observe $\ell_{t,i}$ for all $i \in S_t$
- Now: assume we only observe the loss ℓ_{t,I_t}
- Using a simple extension of FPL, we prove

Theorem 4
Assuming IID expert availability, the expected regret of the bandit extension of FPL satisfies

$$R_T = O\left(T^{2/3}\right)$$

Extensions: bandit feedback

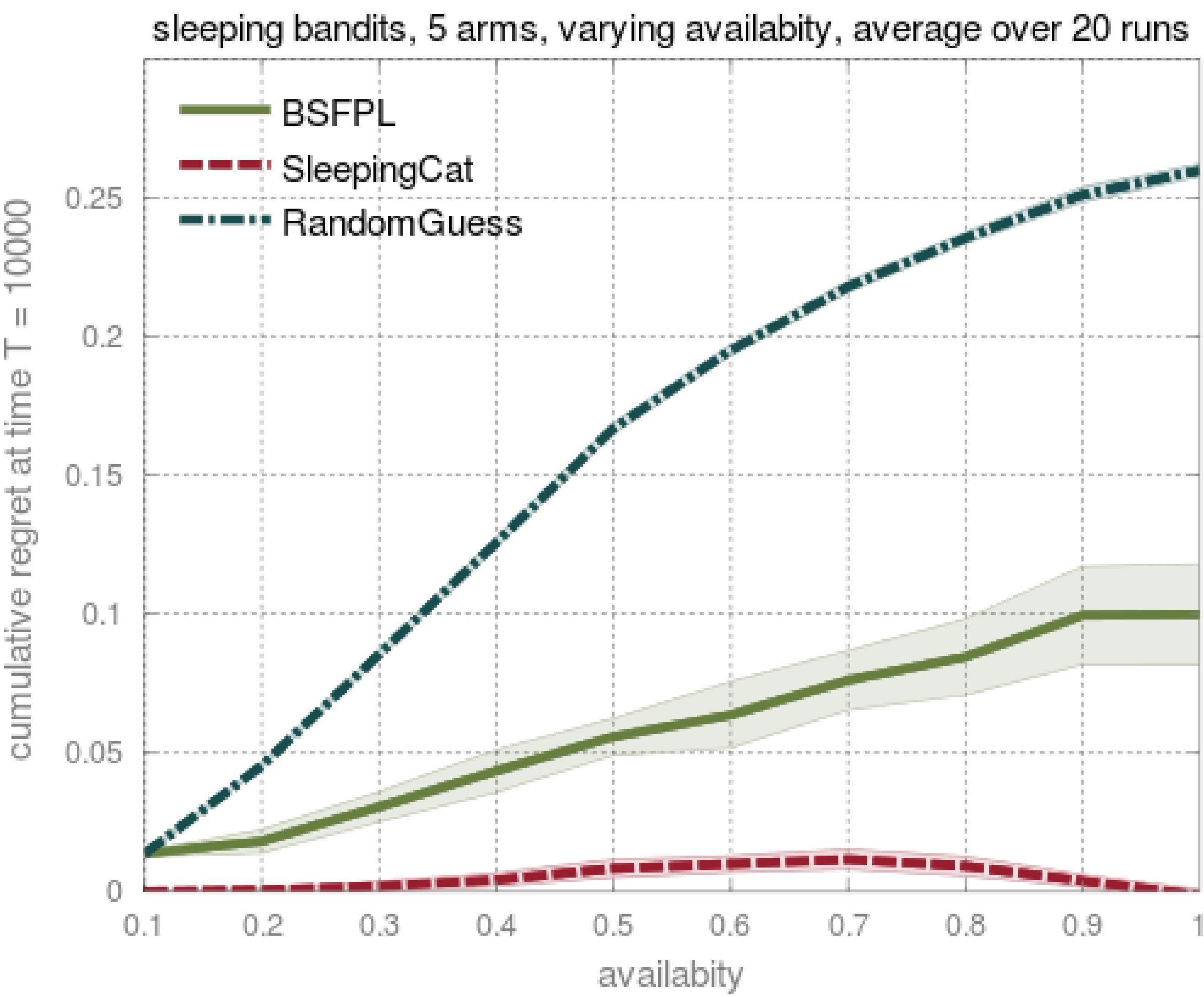
- So far: assume we observe $\ell_{t,i}$ for all $i \in S_t$
- Now: assume we only observe the loss ℓ_{t,I_t}
- Using a simple extension of FPL, we prove

Theorem 4
Assuming IID expert availability, the expected regret of the bandit extension of FPL satisfies

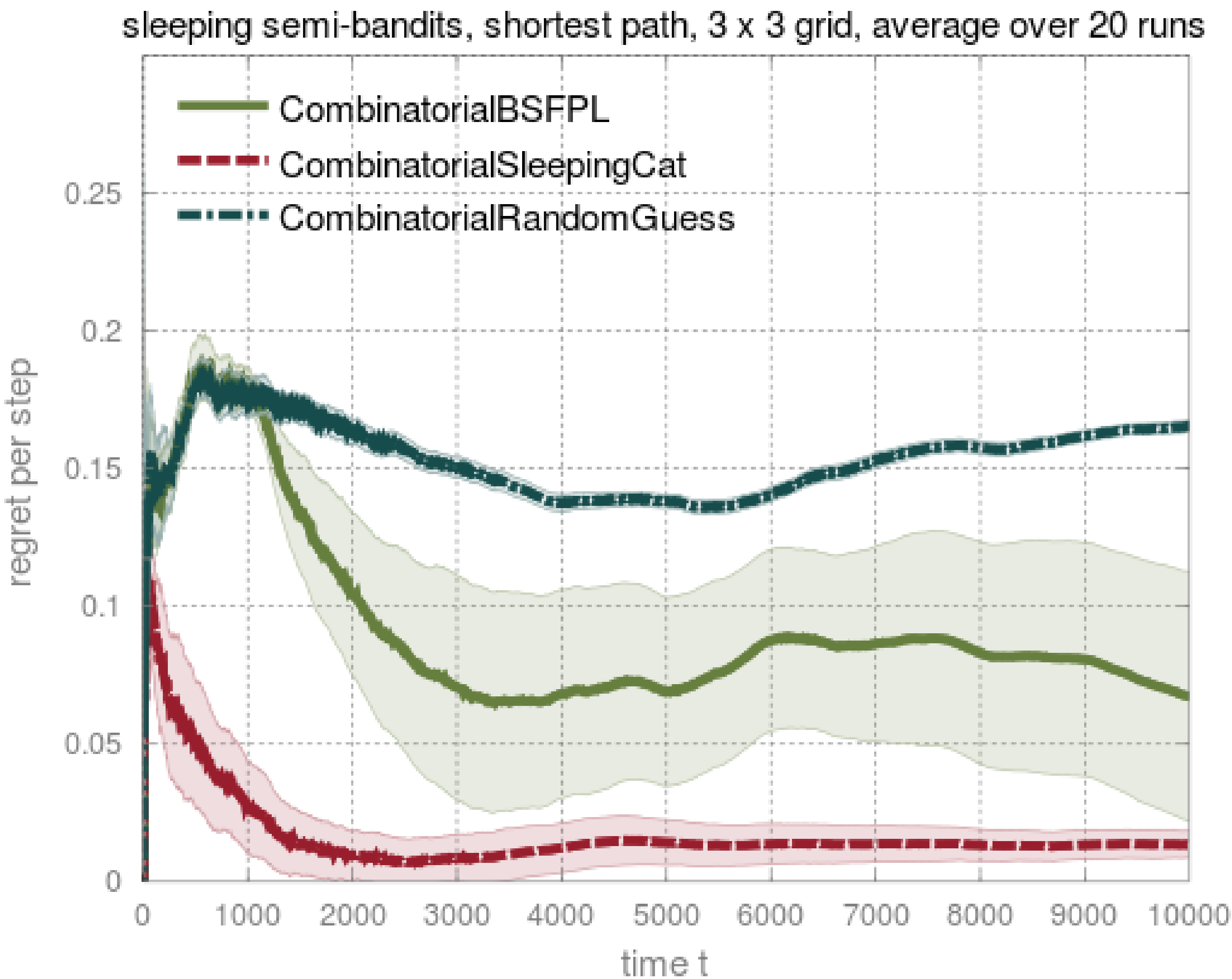
$$R_T = O\left(T^{2/3}\right)$$

Best previous result was $O\left(T^{4/5}\right)$

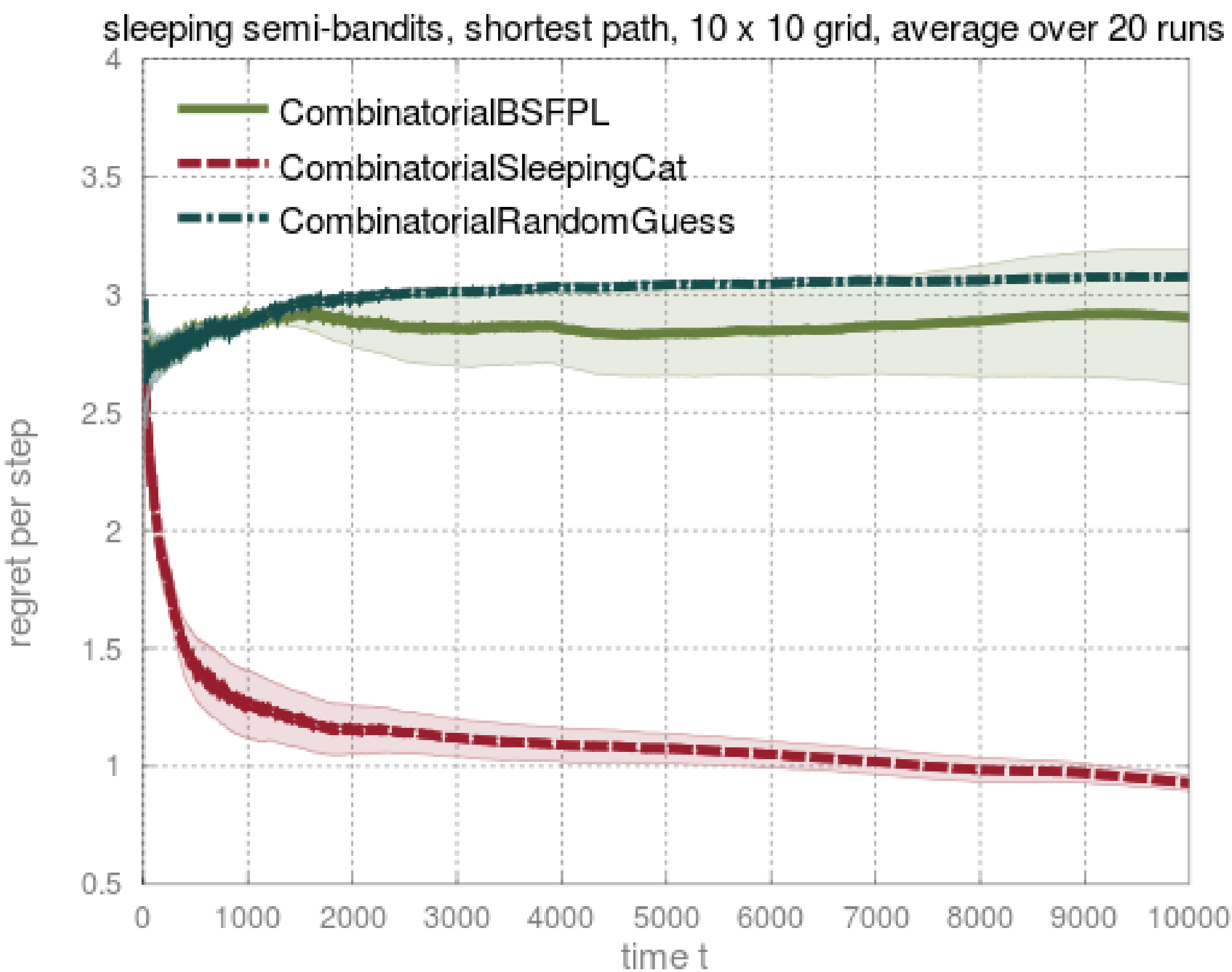
Experiments



Experiments



Experiments



Future work

- Prove $R_T = O\left(\sqrt{T}\right)$ for sleeping bandits?
  - Problem: knowing the q_i's is not enough
- Extend results to more complicated availability assumptions:
  - Markovian arms
  - Mortal arms

Thanks!