Learning to Optimize

Exploration and Generalization

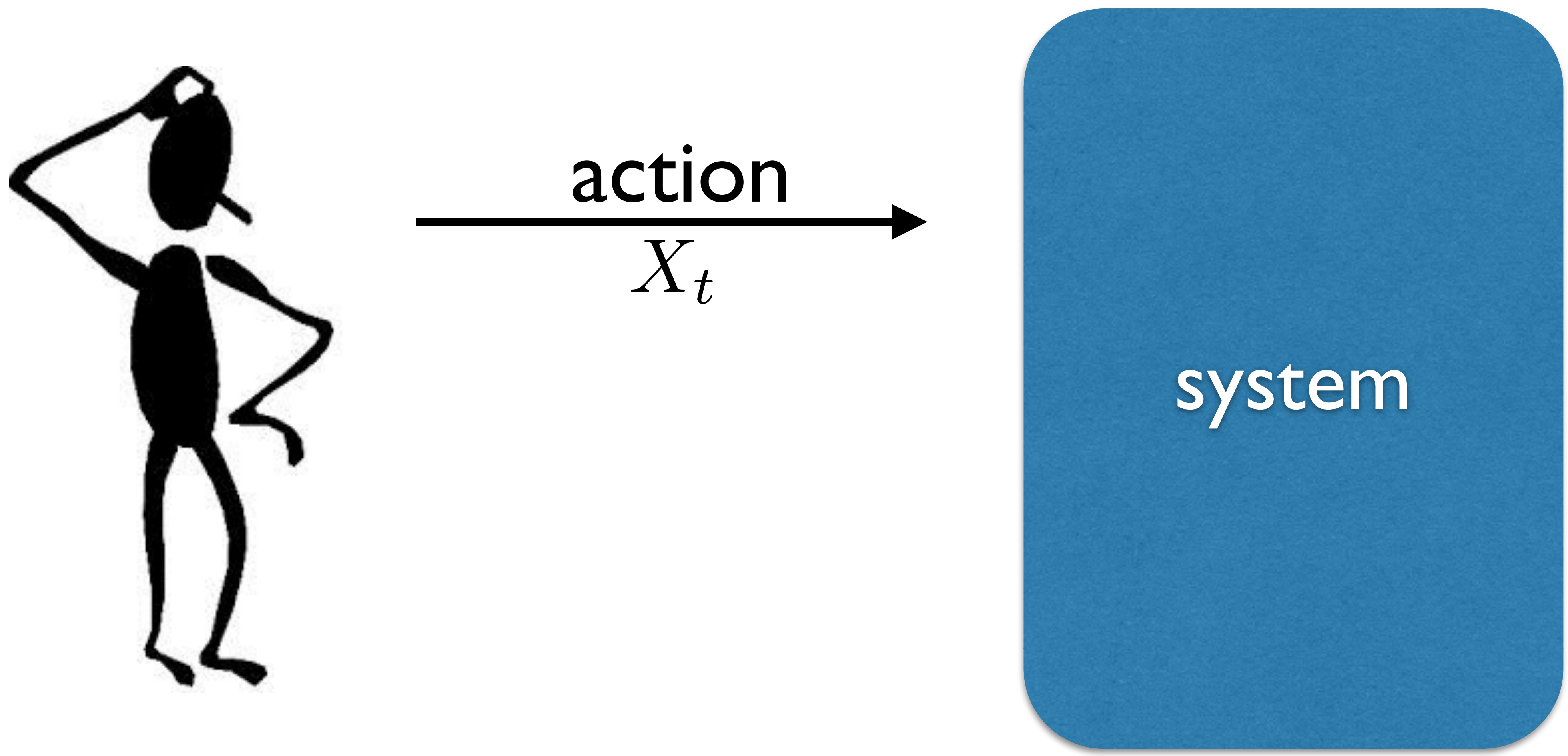
Benjamin Van Roy

work done with Dan Russo

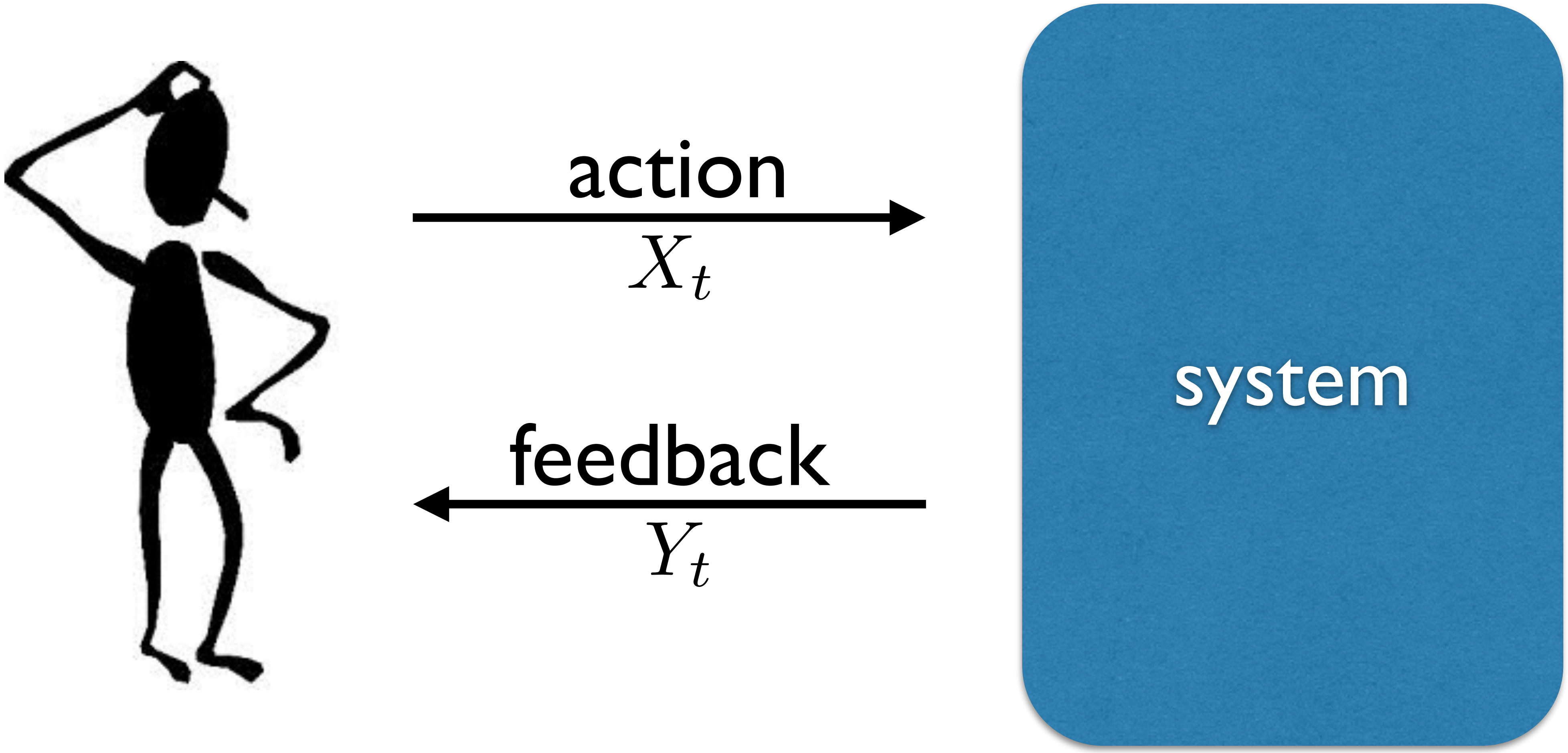
Learning to Optimize

system

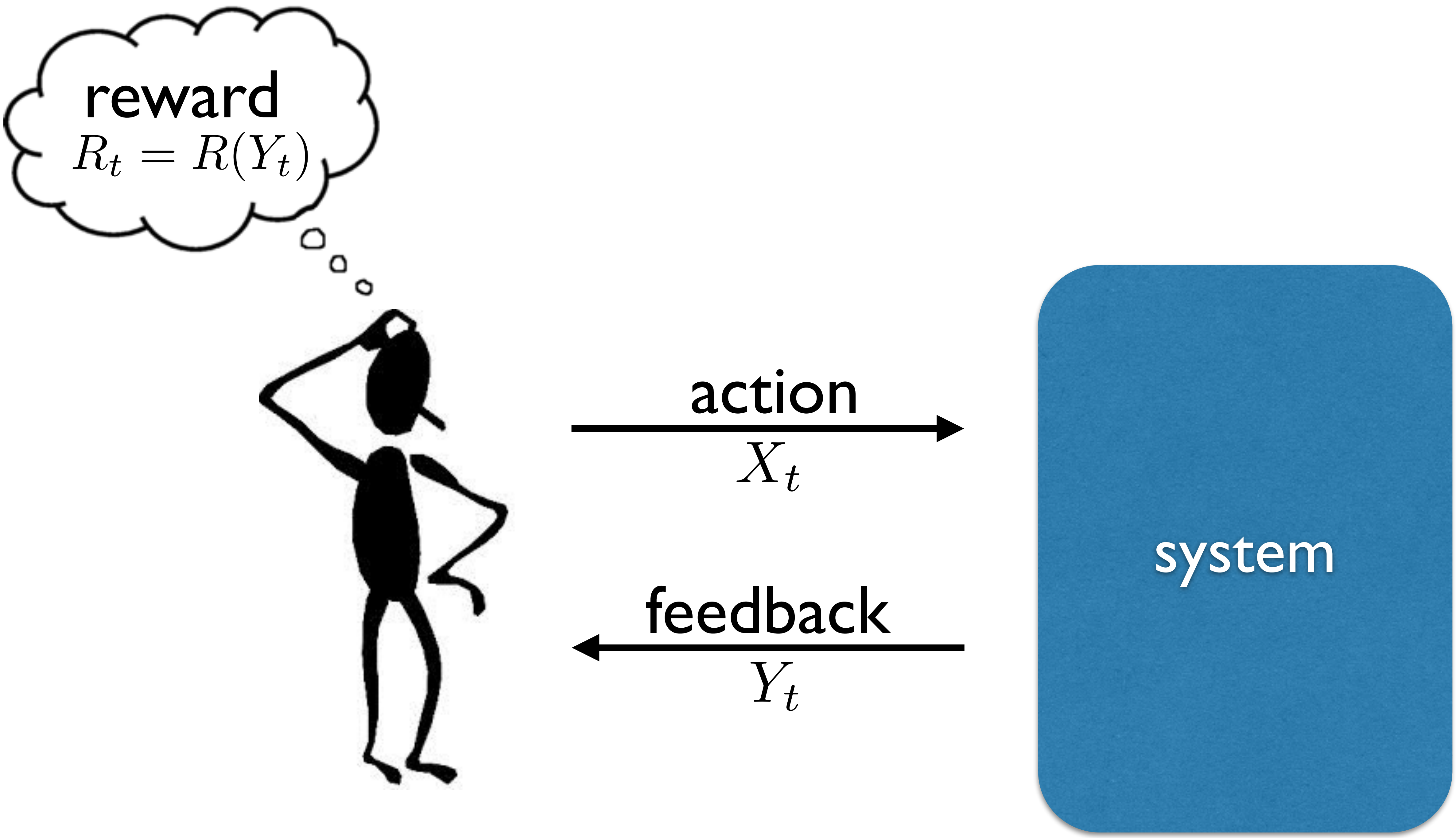
Learning to Optimize



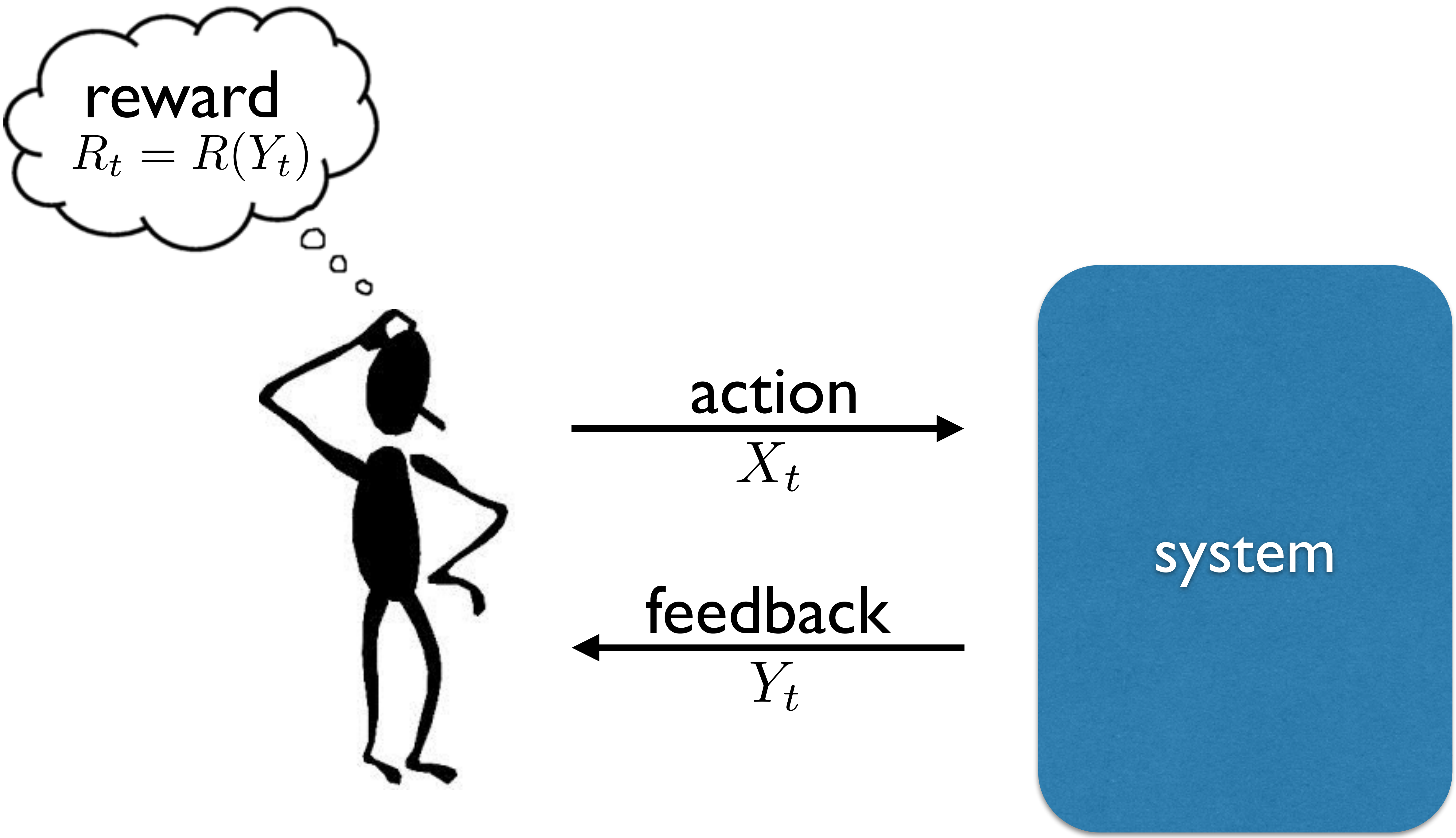
Learning to Optimize



Learning to Optimize



Learning to Optimize



exploration versus exploitation

A Generalization of Optimization

$$\max_{x \in \mathbb{X}} f(x)$$

A Generalization of Optimization

$$\max_{x \in \mathbb{X}} f_\theta(x) \qquad\qquad \theta \in \Theta$$

A Generalization of Optimization

$$\max_{x \in \mathbb{X}} f_\theta(x) \qquad \theta \in \Theta$$

- Expected reward $\qquad f_\theta(X_t) = E\left[R_t | X_t, \theta\right]$

A Generalization of Optimization

$$\max_{x \in \mathbb{X}} f_\theta(x) \qquad \theta \in \Theta$$

- Expected reward $\quad f_\theta(X_t) = E\left[R_t | X_t, \theta\right]$
- Represent knowledge about model via

A Generalization of Optimization

$$\max_{x \in \mathbb{X}} f_\theta(x) \qquad \theta \in \Theta$$

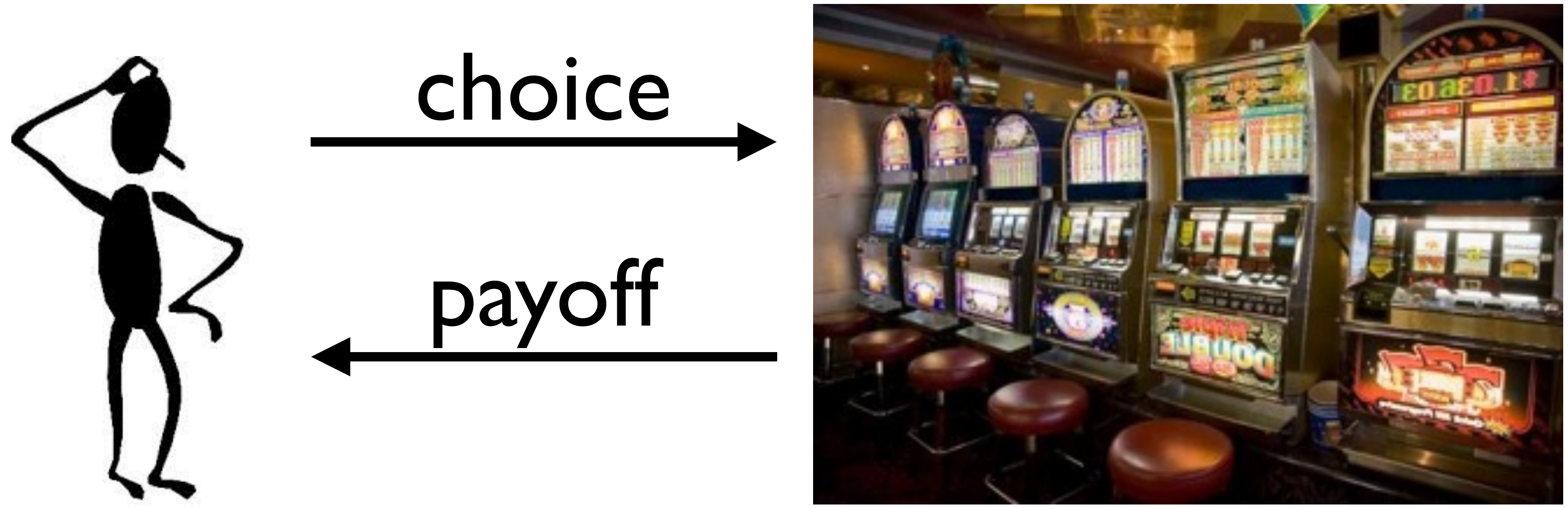
- Expected reward $\quad f_\theta(X_t) = E\left[R_t | X_t, \theta\right]$
- Represent knowledge about model via
    - Set membership $\quad \theta \in \Theta_t \subseteq \Theta$

A Generalization of Optimization

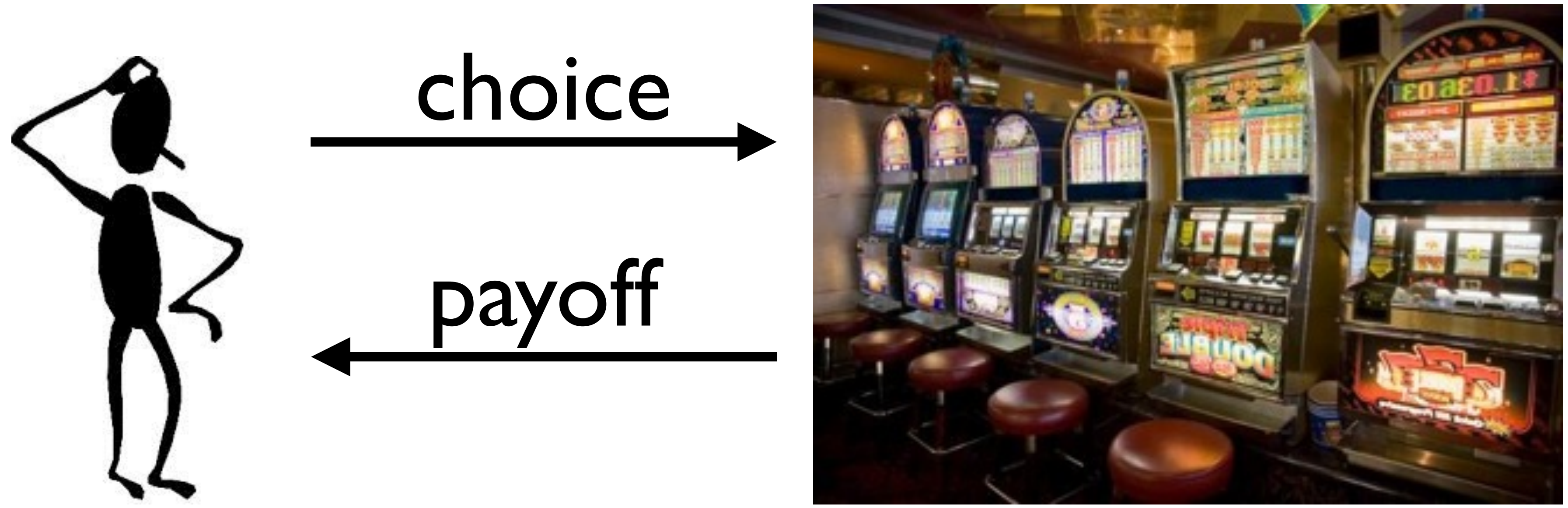
$$\max_{x \in \mathbb{X}} f_\theta(x) \qquad \theta \in \Theta$$

- Expected reward $\quad f_\theta(X_t) = E\left[R_t | X_t, \theta\right]$
- Represent knowledge about model via
  - Set membership $\quad \theta \in \Theta_t \subseteq \Theta$
  - Probability distribution $\quad \theta \sim p_t(\cdot)$

Example: Multi-Armed Bandit Problem

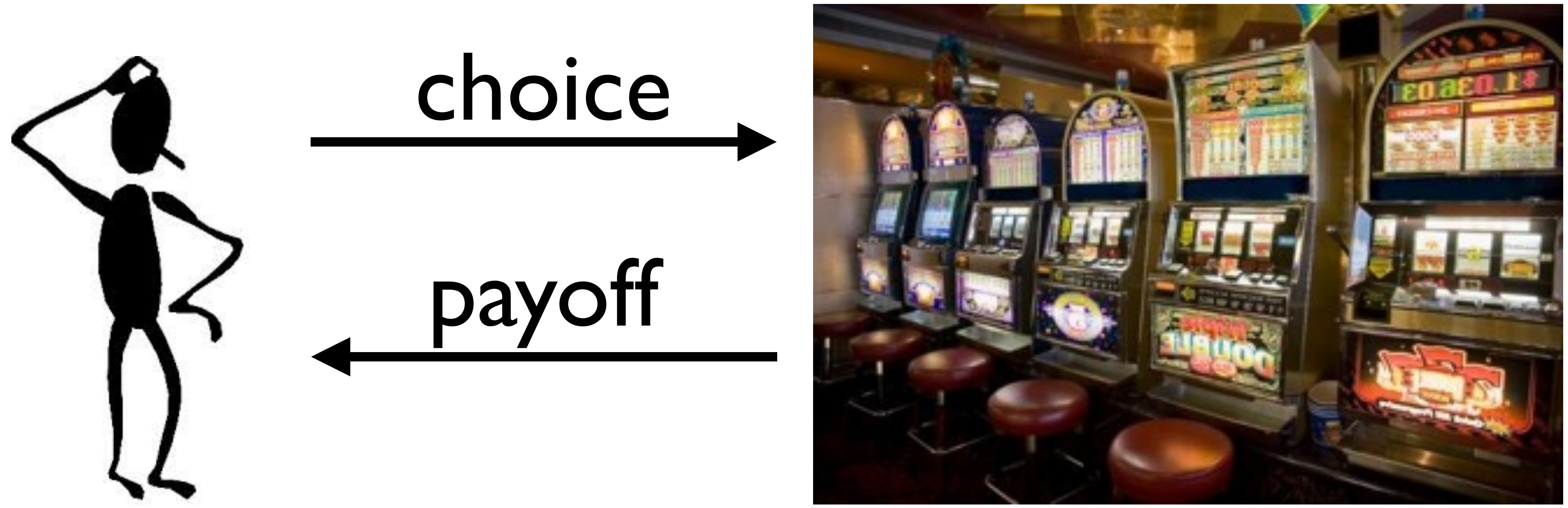


Example: Multi-Armed Bandit Problem



- Action/arm $\mathbb{X} = \{1, \ldots, n\}$

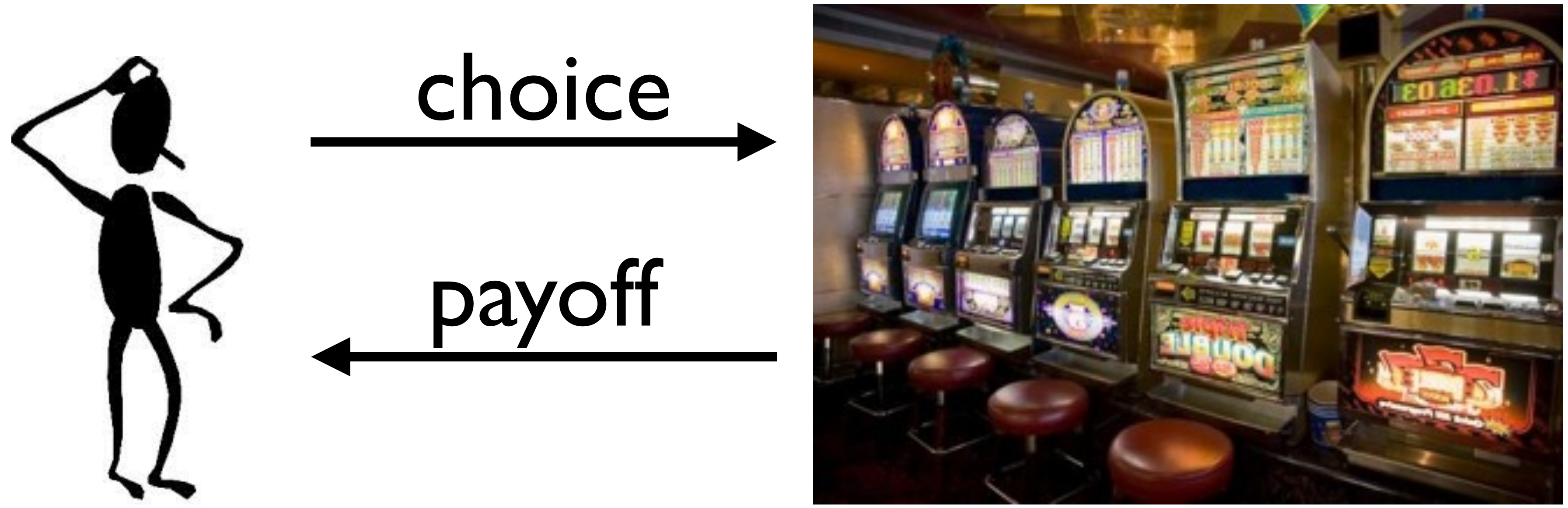
Example: Multi-Armed Bandit Problem



- Action/arm $\mathbb{X} = \{1, \ldots, n\}$
- Mean rewards with independent priors

$$f_\theta(x) = \theta_x \qquad\qquad \theta \sim p_0(\theta) = \prod_{x=1}^{N} p_0^x(\theta_x)$$

Example: Multi-Armed Bandit Problem

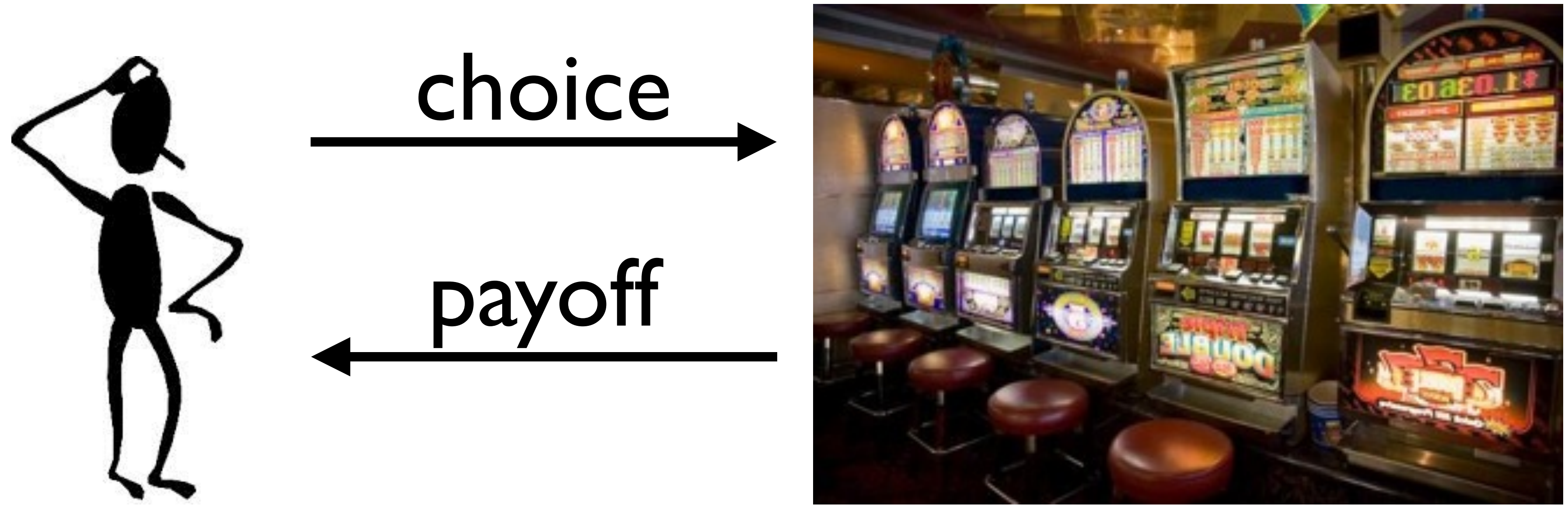


- Action/arm $\mathbb{X} = \{1, \ldots, n\}$
- Mean rewards with independent priors

$$f_\theta(x) = \theta_x \qquad\qquad \theta \sim p_0(\theta) = \prod_{x=1}^{N} p_0^x(\theta_x)$$

- Feedback/reward $R_t = Y_t = f_\theta(X_t) + W_t$

Example: Multi-Armed Bandit Problem



- Action/arm $\mathbb{X} = \{1, \ldots, n\}$
- Mean rewards with independent priors

$$f_\theta(x) = \theta_x \qquad\qquad \theta \sim p_0(\theta) = \prod_{x=1}^{N} p_0^x(\theta_x)$$

- Feedback/reward $R_t = Y_t = f_\theta(X_t) + W_t$
- Discounted objective addressed by Gittin's Index Theorem

Example: Linear Program

Example: Linear Program

- Linear program

$$f_\theta(x) = \theta^\top x$$

$$\mathbb{X} = \{x : Ax \leq b\}$$

Example: Linear Program

- Linear program $f_\theta(x) = \theta^\top x$

  $\mathbb{X} = \{x : Ax \leq b\}$

- Gaussian Prior $\theta \sim N(\mu, \Sigma)$

Example: Linear Program

- Linear program $f_\theta(x) = \theta^\top x$

  $\mathbb{X} = \{x : Ax \leq b\}$

- Gaussian Prior $\theta \sim N(\mu, \Sigma)$

- Noisy feedback / reward $R_t = Y_t = \theta^\top X_t + W_t$

  $W_t \sim N(0, \sigma^2)$

Example: Linear Program

- Linear program $f_\theta(x) = \theta^\top x$

  $\mathbb{X} = \{x : Ax \leq b\}$

- Gaussian Prior $\theta \sim N(\mu, \Sigma)$

- Noisy feedback / reward $R_t = Y_t = \theta^\top X_t + W_t$

  $W_t \sim N(0, \sigma^2)$

natural objectives are intractable

Heuristics

Heuristics

- Comparisons via

Heuristics

- Comparisons via
  - Simulations

Heuristics

- Comparisons via
  - Simulations
  - Theoretical objectives such as expected regret

$$E\left[\mathrm{Regret}(T)\right] = \sum_{t=1}^{T} E\left[\max_x f_\theta(x) - f_\theta(X_t)\right]$$

Heuristics

- Comparisons via
  - Simulations
  - Theoretical objectives such as expected regret

$$E\left[\text{Regret}(T)\right] = \sum_{t=1}^{T} E\left[\max_x \underbrace{f_\theta(x)}_{\text{for optimal action}} - f_\theta(X_t)\right]$$

Heuristics

- Comparisons via
  - Simulations
  - Theoretical objectives such as expected regret

$$E\left[\text{Regret}(T)\right] = \sum_{t=1}^{T} E\left[\max_x f_\theta(x) - f_\theta(X_t)\right]$$

for optimal action

for selected action

Heuristics

- Comparisons via
  - Simulations
  - Theoretical objectives such as expected regret

$$E\left[\mathrm{Regret}(T)\right] = \sum_{t=1}^{T} E\left[\max_x f_\theta(x) - f_\theta(X_t)\right]$$

expectation over models

for optimal action

for selected action

Heuristics

- Comparisons via
  - Simulations
  - Theoretical objectives such as expected regret

$$E\left[\text{Regret}(T)\right] = \sum_{t=1}^{T} E\left[\max_x f_\theta(x) - f_\theta(X_t)\right]$$

expectation over models

for optimal action

for selected action

minimizing expected regret maximizes expected reward

Heuristics

- Comparisons via
  - Simulations
  - Theoretical objectives such as expected regret

$$E\left[\text{Regret}(T)\right] = \sum_{t=1}^{T} E\left[\max_x f_\theta(x) - f_\theta(X_t)\right]$$

expectation over models

for optimal action

for selected action

minimizing expected regret maximizes expected reward

- Emphasis has been on "large" T

Heuristics

- Comparisons via
  - Simulations
  - Theoretical objectives such as expected regret

$$E\left[\text{Regret}(T)\right] = \sum_{t=1}^{T} E\left[\max_x f_\theta(x) - f_\theta(X_t)\right]$$

expectation over models

for optimal action

for selected action

minimizing expected regret maximizes expected reward

- Emphasis has been on "large" T
- Popular approaches to heuristic design

Heuristics

- Comparisons via
  - Simulations
  - Theoretical objectives such as expected regret

$$E\left[\mathrm{Regret}(T)\right] = \sum_{t=1}^{T} E\left[\max_x f_\theta(x) - f_\theta(X_t)\right]$$

expectation over models

for optimal action

for selected action

minimizing expected regret maximizes expected reward

- Emphasis has been on "large" T
- Popular approaches to heuristic design
  - Upper-confidence-bounds [Lai-Robbins, 1985; Dani-Hayes-Kakade, 2008; Rusmevichientong-Tsitsiklis, 2010; etc.]

Heuristics

- Comparisons via
  - Simulations
  - Theoretical objectives such as expected regret

$$E\left[\text{Regret}(T)\right] = \sum_{t=1}^{T} E\left[\max_x f_\theta(x) - f_\theta(X_t)\right]$$

expectation over models

for optimal action

for selected action

minimizing expected regret maximizes expected reward

- Emphasis has been on "large" T
- Popular approaches to heuristic design
  - Upper-confidence-bounds [Lai-Robbins, 1985; Dani-Hayes-Kakade, 2008; Rusmevichientong-Tsitsiklis, 2010; etc.]
  - Thompson sampling [Thompson, 1933]

Upper-Confidence-Bound Algorithms (UCB)

Upper-Confidence-Bound Algorithms (UCB)

- Confidence set Θ_t

Upper-Confidence-Bound Algorithms (UCB)

- Confidence set Θ_t
  - Set of "statistically plausible" models

Upper-Confidence-Bound Algorithms (UCB)

- Confidence set Θ_t
  - Set of “statistically plausible” models
  - Updated based on observations

Upper-Confidence-Bound Algorithms (UCB)

- Confidence set Θ_t
  - Set of “statistically plausible” models
  - Updated based on observations

- Upper confidence bounds $U_t(x) = \max_{\theta \in \Theta_t} f_\theta(x)$

Upper-Confidence-Bound Algorithms (UCB)

- Confidence set Θ_t
  - Set of “statistically plausible” models
  - Updated based on observations
- Upper confidence bounds $U_t(x) = \max_{\theta \in \Theta_t} f_\theta(x)$
- Optimistic optimization $X_t \in \arg\max_{x \in \mathbb{X}} U_t(x)$

Upper-Confidence-Bound Algorithms (UCB)

- Confidence set Θ_t
  - Set of "statistically plausible" models
  - Updated based on observations

- Upper confidence bounds $U_t(x) = \max_{\theta \in \Theta_t} f_\theta(x)$

- Optimistic optimization $X_t \in \arg\max_{x \in \mathbb{X}} U_t(x)$

- Bayes-UCB

Upper-Confidence-Bound Algorithms (UCB)

- Confidence set Θ_t
  - Set of “statistically plausible” models
  - Updated based on observations

- Upper confidence bounds $U_t(x) = \max_{\theta \in \Theta_t} f_\theta(x)$

- Optimistic optimization $X_t \in \arg\max_{x \in \mathbb{X}} U_t(x)$

- Bayes-UCB
  - Maintain probability distribution $p_t(d\theta) = \mathbb{P}\left[\theta \in d\theta | \mathbb{F}_{t-1}\right]$

Upper-Confidence-Bound Algorithms (UCB)

- Confidence set Θ_t
  - Set of "statistically plausible" models
  - Updated based on observations
- Upper confidence bounds $U_t(x) = \max_{\theta \in \Theta_t} f_\theta(x)$
- Optimistic optimization $X_t \in \arg\max_{x \in \mathbb{X}} U_t(x)$
- Bayes-UCB
  - Maintain probability distribution $p_t(d\theta) = \mathbb{P}\left[\theta \in d\theta | \mathbb{F}_{t-1}\right]$
  - Select level set as confidence set

Thompson Sampling (TS)

Thompson Sampling (TS)

- Maintain probability distribution $p_t(d\theta) = \mathbb{P}\left[\theta \in d\theta | \mathbb{F}_t\right]$

Thompson Sampling (TS)

- Maintain probability distribution $p_t(d\theta) = \mathbb{P}\left[\theta \in d\theta | \mathbb{F}_t\right]$
- Sample model $\theta_t \sim p_t$

Thompson Sampling (TS)

- Maintain probability distribution $p_t(d\theta) = \mathbb{P}\left[\theta \in d\theta | \mathbb{F}_t\right]$
- Sample model $\theta_t \sim p_t$
- Optimize sample $X_t \in \arg\max_{x \in \mathbb{X}} f_{\theta_t}(x)$

Thompson Sampling (TS)

- Maintain probability distribution $p_t(d\theta) = \mathbb{P}\left[\theta \in d\theta | \mathbb{F}_t\right]$
- Sample model $\theta_t \sim p_t$
- Optimize sample $X_t \in \arg\max_{x \in \mathbb{X}} f_{\theta_t}(x)$

sample each action with the probability that it is optimal

Regret Bounds

Regret Bounds

- UCB
  - Finite indep. $\mathbb{X}$ [Auer et al, 2002]

Regret Bounds

- UCB
  - Finite indep. $\mathbb{X}$ — [Auer et al, 2002]
  - Linear — [Dani-Hayes-Kakade, 2008; Rusmevichientong-Tsitsiklis, 2010; Abbasi-Yadkori et al, 2011]

Regret Bounds

- UCB
  - Finite indep. $\mathbb{X}$ [Auer et al, 2002]
  - Linear [Dani-Hayes-Kakade, 2008; Rusmevichientong-Tsitsiklis, 2010; Abbasi-Yadkori et al, 2011]
  - Generalized linear [Filippi et al, 2010]

Regret Bounds

- UCB
  - Finite indep. $\mathbb{X}$ [Auer et al, 2002]
  - Linear [Dani-Hayes-Kakade, 2008; Rusmevichientong-Tsitsiklis, 2010; Abbasi-Yadkori et al, 2011]
  - Generalized linear [Filippi et al, 2010]
- TS

Regret Bounds

- UCB
  - Finite indep. $\mathbb{X}$ [Auer et al, 2002]
  - Linear [Dani-Hayes-Kakade, 2008; Rusmevichientong-Tsitsiklis, 2010; Abbasi-Yadkori et al, 2011]
  - Generalized linear [Filippi et al, 2010]
- TS
  - Finite X, Bernoulli [Agrawal-Goyal, 2012]

Regret Bounds

- UCB
  - Finite indep. $\mathbb{X}$ — [Auer et al, 2002]
  - Linear — [Dani-Hayes-Kakade, 2008; Rusmevichientong-Tsitsiklis, 2010; Abbasi-Yadkori et al, 2011]
  - Generalized linear — [Filippi et al, 2010]
- TS
  - Finite X, Bernoulli — [Agrawal-Goyal, 2012]
  - Linear — [Agrawal-Goyal, 2012]

Regret Bounds

- UCB
  - Finite indep. $\mathbb{X}$ [Auer et al, 2002]
  - Linear [Dani-Hayes-Kakade, 2008; Rusmevichientong-Tsitsiklis, 2010; Abbasi-Yadkori et al, 2011]
  - Generalized linear [Filippi et al, 2010]
- TS
  - Finite X, Bernoulli [Agrawal-Goyal, 2012]
  - Linear [Agrawal-Goyal, 2012]

UCB Regret Bounds → TS E[Regret] Bounds

[Russo-Van Roy, 2013]

Regret Bounds

- UCB
  - Finite indep. $\mathbb{X}$ [Auer et al, 2002]
  - Linear [Dani-Hayes-Kakade, 2008; Rusmevichientong-Tsitsiklis, 2010; Abbasi-Yadkori et al, 2011]
  - Generalized linear [Filippi et al, 2010]
- TS
  - Finite X, Bernoulli [Agrawal-Goyal, 2012]
  - Linear [Agrawal-Goyal, 2012]

UCB Regret Bounds → TS E[Regret] Bounds

[Russo-Van Roy, 2013]

- The role of confidence sets

Regret Bounds

- UCB
  - Finite indep. $\mathbb{X}$ [Auer et al, 2002]
  - Linear [Dani-Hayes-Kakade, 2008; Rusmevichientong-Tsitsiklis, 2010; Abbasi-Yadkori et al, 2011]
  - Generalized linear [Filippi et al, 2010]
- TS
  - Finite X, Bernoulli [Agrawal-Goyal, 2012]
  - Linear [Agrawal-Goyal, 2012]

UCB Regret Bounds → TS E[Regret] Bounds

[Russo-Van Roy, 2013]

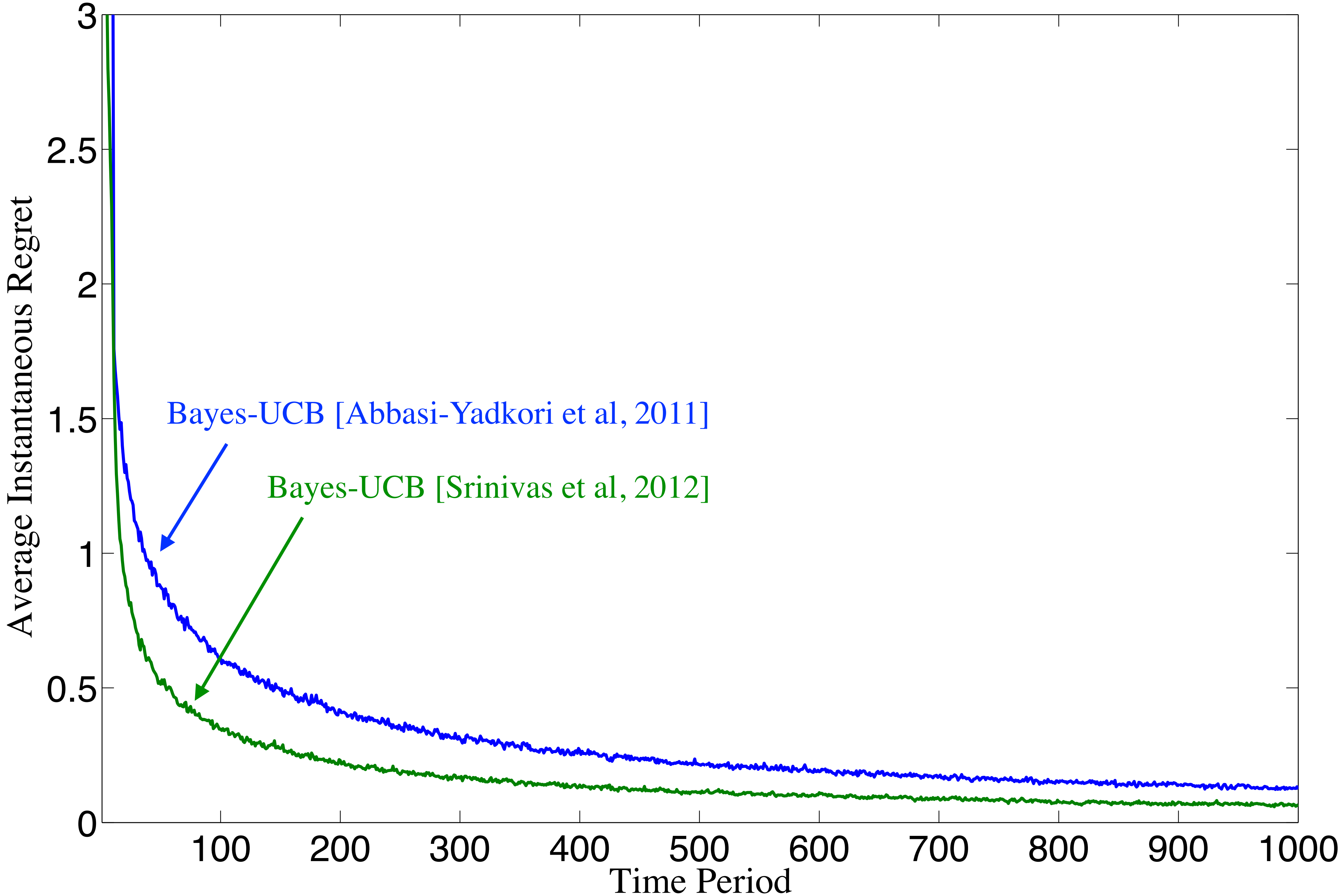
- The role of confidence sets
  - UCB: algorithm design and analysis
  - TS: analysis only

Linear Bandit Simulations

$$f_\theta(x) = \theta^\top x \qquad\qquad \mathbb{X} = \{z_1, \ldots, z_n\}$$

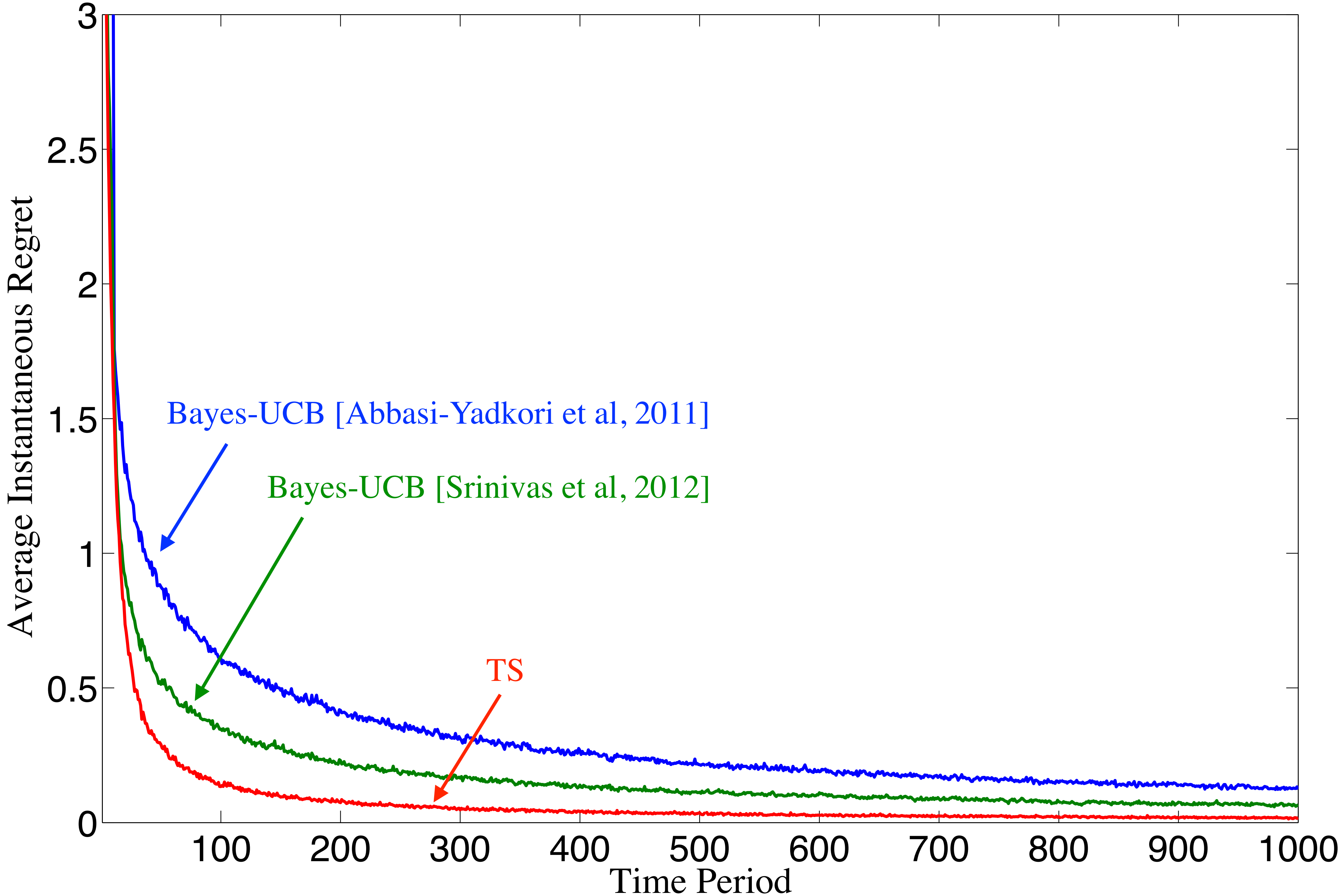
Linear Bandit Simulations

$$f_\theta(x) = \theta^\top x \qquad\qquad \mathbb{X} = \{z_1, \ldots, z_n\}$$



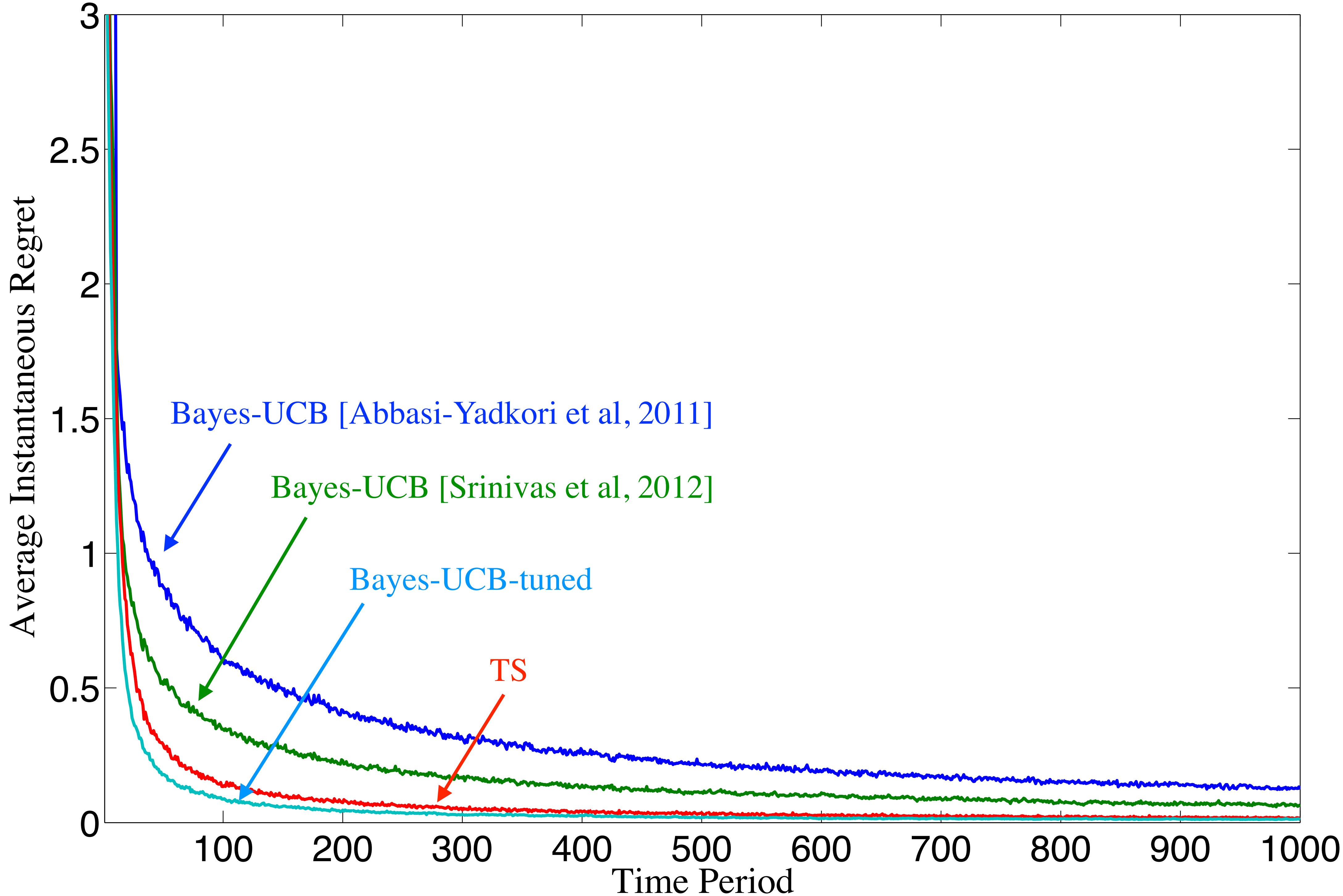
Linear Bandit Simulations

$$f_\theta(x) = \theta^\top x \qquad \mathbb{X} = \{z_1, \ldots, z_n\}$$



Linear Bandit Simulations

$$f_\theta(x) = \theta^\top x \qquad\qquad \mathbb{X} = \{z_1, \ldots, z_n\}$$



Computational Considerations

Computational Considerations

- TS is often tractable when Bayes-UCB is not

Computational Considerations

- TS is often tractable when Bayes-UCB is not
- Consider LP

$$f_\theta(x) = \theta^\top x \qquad \theta \sim N(\mu, \Sigma)$$

$$\mathbb{X} = \{x : Ax \leq b\} \qquad R_t = Y_t = \theta^\top X_t + W_t$$

$$W_t \sim N(0, \sigma^2)$$

Computational Considerations

- TS is often tractable when Bayes-UCB is not
- Consider LP

$$f_\theta(x) = \theta^\top x \qquad \theta \sim N(\mu, \Sigma)$$

$$\mathbb{X} = \{x : Ax \leq b\} \qquad R_t = Y_t = \theta^\top X_t + W_t$$

$$W_t \sim N(0, \sigma^2)$$

- TS is computationally efficient

Computational Considerations

- TS is often tractable when Bayes-UCB is not
- Consider LP

$$f_\theta(x) = \theta^\top x \qquad \theta \sim N(\mu, \Sigma)$$

$$\mathbb{X} = \{x : Ax \leq b\} \qquad R_t = Y_t = \theta^\top X_t + W_t$$

$$W_t \sim N(0, \sigma^2)$$

- TS is computationally efficient
- Bayes-UCB is computationally intractable

Computational Considerations

- TS is often tractable when Bayes-UCB is not
- Consider LP

$$f_\theta(x) = \theta^\top x \qquad \theta \sim N(\mu, \Sigma)$$

$$\mathbb{X} = \{x : Ax \leq b\} \qquad R_t = Y_t = \theta^\top X_t + W_t$$

$$W_t \sim N(0, \sigma^2)$$

- TS is computationally efficient
- Bayes-UCB is computationally intractable

- Computationally tractable version of UCB
  - Regret scaled by a factor of d [Dani-Hayes-Kakade, 2008]

Summary of TS versus UCB

Summary of TS versus UCB

- TS outperforms Bayes-UCB designed for analysis

Summary of TS versus UCB

- TS outperforms Bayes-UCB designed for analysis
- TS slightly underperforms well-tuned Bayes-UCB

Summary of TS versus UCB

- TS outperforms Bayes-UCB designed for analysis
- TS slightly underperforms well-tuned Bayes-UCB
- TS often tractable when Bayes-UCB not

Summary of TS versus UCB

- TS outperforms Bayes-UCB designed for analysis
- TS slightly underperforms well-tuned Bayes-UCB
- TS often tractable when Bayes-UCB not
- TS outperforms non-Bayes-UCB designed for tractability

General Bound

General Bound

- Bound via general notion of function class complexity

General Bound

- Bound via general notion of function class complexity

$$E\left[Regret(T)\right] \leq \tilde{O}\left(\sqrt{d_E(T)\log\left(N(T)\right)T}\right)$$ [Russo-Van Roy, 2013]

T^{-2}-scale eluder dimension of function class

T^{-2}-covering number of function class

General Bound

- Bound via general notion of function class complexity

$$E\left[Regret(T)\right] \leq \tilde{O}\left(\sqrt{d_E(T)\log\left(N(T)\right)T}\right)$$

[Russo-Van Roy, 2013]

T^{-2}-scale eluder dimension of function class

T^{-2}-covering number of function class

- CN is representative of supervised learning concepts

General Bound

- Bound via general notion of function class complexity

$$E\left[Regret(T)\right] \leq \tilde{O}\left(\sqrt{d_E(T)\log\left(N(T)\right)T}\right)$$

[Russo-Van Roy, 2013]

T^{-2}-scale eluder dimension of function class

T^{-2}-covering number of function class

- CN is representative of supervised learning concepts
- ED is new and necessary

General Bound

- Bound via general notion of function class complexity

$$E\left[Regret(T)\right] \leq \tilde{O}\left(\sqrt{d_E(T) \log\left(N(T)\right) T}\right)$$

[Russo-Van Roy, 2013]

T^{-2}-scale eluder dimension of function class

T^{-2}-covering number of function class

- CN is representative of supervised learning concepts
- ED is new and necessary

- Specializes to various model classes

General Bound

- Bound via general notion of function class complexity

$$E\left[Regret(T)\right] \leq \tilde{O}\left(\sqrt{d_E(T)\log\left(N(T)\right)T}\right)$$

[Russo-Van Roy, 2013]

T^{-2}-scale eluder dimension of function class

T^{-2}-covering number of function class

- CN is representative of supervised learning concepts
- ED is new and necessary

- Specializes to various model classes
  - Linear bandits: recovers best previous bounds

General Bound

- Bound via general notion of function class complexity

$$E\left[Regret(T)\right] \leq \tilde{O}\left(\sqrt{d_E(T)\log\left(N(T)\right)T}\right)$$

[Russo-Van Roy, 2013]

T^{-2}-scale eluder dimension of function class

T^{-2}-covering number of function class

- CN is representative of supervised learning concepts
- ED is new and necessary

- Specializes to various model classes
  - Linear bandits: recovers best previous bounds
  - Generalized linear bandits: slight improvement

Troubling Example: Sparse Linear Bandit

Troubling Example: Sparse Linear Bandit

- A 1-sparse case

Troubling Example: Sparse Linear Bandit

- A 1-sparse case

$$f_\theta(x) = \theta^\top x$$

Troubling Example: Sparse Linear Bandit

- A 1-sparse case

$$f_\theta(x) = \theta^\top x \qquad \mathbb{X} = \left\{ x \in \left\{ 0, \frac{1}{m} \right\}^d : \|x\|_0 = m \right\}$$

Troubling Example: Sparse Linear Bandit

- A 1-sparse case

$$f_\theta(x) = \theta^\top x \qquad \mathbb{X} = \left\{ x \in \left\{0, \frac{1}{m}\right\}^d : \|x\|_0 = m \right\}$$

$$R_t = Y_t = f_\theta(X_t)$$

Troubling Example: Sparse Linear Bandit

- A 1-sparse case

$$f_\theta(x) = \theta^\top x \qquad \mathbb{X} = \left\{ x \in \left\{0, \frac{1}{m}\right\}^d : \|x\|_0 = m \right\}$$

$$R_t = Y_t = f_\theta(X_t)$$

$$\theta \sim \mathrm{unif}\left(\left\{\theta \in \{0,1\}^d : \|\theta\|_0 = 1\right\}\right)$$

Troubling Example: Sparse Linear Bandit

- A 1-sparse case

$$f_\theta(x) = \theta^\top x \qquad \mathbb{X} = \left\{ x \in \left\{ 0, \frac{1}{m} \right\}^d : \|x\|_0 = m \right\}$$

$$R_t = Y_t = f_\theta(X_t)$$

$$\theta \sim \mathrm{unif}\left(\left\{ \theta \in \{0,1\}^d : \|\theta\|_0 = 1 \right\} \right)$$

- UCB/TS require $\Omega(d)$ samples to identify

Troubling Example: Sparse Linear Bandit

- A 1-sparse case

$$f_\theta(x) = \theta^\top x \qquad \mathbb{X} = \left\{ x \in \left\{0, \frac{1}{m}\right\}^d : \|x\|_0 = m \right\}$$

$$R_t = Y_t = f_\theta(X_t)$$

$$\theta \sim \mathrm{unif}\left(\left\{\theta \in \{0,1\}^d : \|\theta\|_0 = 1\right\}\right)$$

- UCB/TS require $\Omega(d)$ samples to identify
  - UCB/TS rule out one action per period

Troubling Example: Sparse Linear Bandit

- A 1-sparse case

$$f_\theta(x) = \theta^\top x \qquad \mathbb{X} = \left\{ x \in \left\{0, \frac{1}{m}\right\}^d : \|x\|_0 = m \right\}$$

$$R_t = Y_t = f_\theta(X_t)$$

$$\theta \sim \mathrm{unif}\left(\left\{\theta \in \{0,1\}^d : \|\theta\|_0 = 1\right\}\right)$$

- UCB/TS require $\Omega(d)$ samples to identify
  - UCB/TS rule out one action per period

- Easy to design algorithms for which $\log_2(d)$ suffice

Troubling Example: Assortment Optimization

Troubling Example: Assortment Optimization

- A simple context

Troubling Example: Assortment Optimization

- A simple context
  - N customer types

Troubling Example: Assortment Optimization

- A simple context
  - N customer types
  - Many products, each geared for a particular type

Troubling Example: Assortment Optimization

- A simple context
  - N customer types
  - Many products, each geared for a particular type
  - Action: recommend assortment of size M

Troubling Example: Assortment Optimization

- A simple context
  - N customer types
  - Many products, each geared for a particular type
  - Action: recommend assortment of size M
  - Customer purchases at most one product per period

Troubling Example: Assortment Optimization

- A simple context
  - N customer types
  - Many products, each geared for a particular type
  - Action: recommend assortment of size M
  - Customer purchases at most one product per period
  - Learn about customer through repeated interactions

Troubling Example: Assortment Optimization

- A simple context
  - N customer types
  - Many products, each geared for a particular type
  - Action: recommend assortment of size M
  - Customer purchases at most one product per period
  - Learn about customer through repeated interactions

- UCB/TS focus on a single customer type

Troubling Example: Assortment Optimization

- A simple context
  - N customer types
  - Many products, each geared for a particular type
  - Action: recommend assortment of size M
  - Customer purchases at most one product per period
  - Learn about customer through repeated interactions
- UCB/TS focus on a single customer type
- Diversifying can reduce regret by a factor of M

Information-Directed Sampling (IDS)

Information-Directed Sampling (IDS)

- Information ratio (IR)

$$\Psi_t = \frac{\left(E\left[f_\theta(X^*) - f_\theta(X_t) | \mathbb{F}_{t-1}\right]\right)^2}{I_t(X^*, Y_t)} = \frac{(\text{expected regret})^2}{\text{mutual information}}$$

Information-Directed Sampling (IDS)

- Information ratio (IR)

$$\Psi_t = \frac{\left(E\left[f_\theta(X^*) - f_\theta(X_t) | \mathbb{F}_{t-1}\right]\right)^2}{I_t(X^*, Y_t)} = \frac{(\text{expected regret})^2}{\text{mutual information}}$$

- Mutual information measures information gain

$$I_t(X^*, Y_t) = E\left[H_t(X^*) - H_{t+1}(X^*) | \mathbb{F}_{t-1}\right]$$

Information-Directed Sampling (IDS)

- Information ratio (IR)

$$\Psi_t = \frac{\left(E\left[f_\theta(X^*) - f_\theta(X_t) | \mathbb{F}_{t-1}\right]\right)^2}{I_t(X^*, Y_t)} = \frac{(\text{expected regret})^2}{\text{mutual information}}$$

- Mutual information measures information gain

$$I_t(X^*, Y_t) = E\left[H_t(X^*) - H_{t+1}(X^*) | \mathbb{F}_{t-1}\right]$$

- Entropy $H_t(X^*)$ measures degree of uncertainty

Information-Directed Sampling (IDS)

- Information ratio (IR)

$$\Psi_t = \frac{\left(E\left[f_\theta(X^*) - f_\theta(X_t)|\mathbb{F}_{t-1}\right]\right)^2}{I_t(X^*, Y_t)} = \frac{(\text{expected regret})^2}{\text{mutual information}}$$

- Mutual information measures information gain

$$I_t(X^*, Y_t) = E\left[H_t(X^*) - H_{t+1}(X^*)|\mathbb{F}_{t-1}\right]$$

- Entropy $H_t(X^*)$ measures degree of uncertainty

- IDS: select action distribution that minimizes Ψ_t

Information-Directed Sampling (IDS)

- Information ratio (IR)

$$\Psi_t = \frac{\left(E\left[f_\theta(X^*) - f_\theta(X_t) | \mathbb{F}_{t-1}\right]\right)^2}{I_t(X^*, Y_t)} = \frac{(\text{expected regret})^2}{\text{mutual information}}$$

- Mutual information measures information gain

$$I_t(X^*, Y_t) = E\left[H_t(X^*) - H_{t+1}(X^*) | \mathbb{F}_{t-1}\right]$$

- Entropy $H_t(X^*)$ measures degree of uncertainty

- IDS: select action distribution that minimizes Ψ_t
  - Trades off between expected regret and information gain

Information-Directed Sampling (IDS)

- Information ratio (IR)

$$\Psi_t = \frac{\left(E\left[f_\theta(X^*) - f_\theta(X_t)|\mathbb{F}_{t-1}\right]\right)^2}{I_t(X^*, Y_t)} = \frac{(\text{expected regret})^2}{\text{mutual information}}$$

- Mutual information measures information gain

$$I_t(X^*, Y_t) = E\left[H_t(X^*) - H_{t+1}(X^*)|\mathbb{F}_{t-1}\right]$$

- Entropy $H_t(X^*)$ measures degree of uncertainty

- IDS: select action distribution that minimizes Ψ_t
  - Trades off between expected regret and information gain
  - Support is of cardinality at most 2

Relation to TS and Regret Bound

Relation to TS and Regret Bound

- A regret bound that applies to all algorithms

$$E[\text{Regret}(T)] \leq \sqrt{\overline{\Psi}_T H(X^*) T}$$

[Russo-Van Roy, 2014]

$$\overline{\Psi}_T = \frac{1}{T} \sum_{t=1}^{T} E[\Psi_t]$$

Relation to TS and Regret Bound

- A regret bound that applies to all algorithms

$$E[\text{Regret}(T)] \leq \sqrt{\overline{\Psi}_T H(X^*) T}$$

[Russo-Van Roy, 2014]

$$\overline{\Psi}_T = \frac{1}{T} \sum_{t=1}^{T} E[\Psi_t]$$

- For IDS:

Relation to TS and Regret Bound

- A regret bound that applies to all algorithms

$$E[\text{Regret}(T)] \leq \sqrt{\overline{\Psi}_T H(X^*) T}$$

[Russo-Van Roy, 2014]

$$\overline{\Psi}_T = \frac{1}{T} \sum_{t=1}^{T} E[\Psi_t]$$

- For IDS:
  - $\overline{\Psi}_t \leq |\mathbb{X}|/2$ always

Relation to TS and Regret Bound

- A regret bound that applies to all algorithms

$$E[\text{Regret}(T)] \leq \sqrt{\overline{\Psi}_T H(X^*) T}$$

[Russo-Van Roy, 2014]

$$\overline{\Psi}_T = \frac{1}{T} \sum_{t=1}^{T} E[\Psi_t]$$

- For IDS:
  - $\overline{\Psi}_t \leq |\mathbb{X}|/2$ always
  - $\overline{\Psi}_t \leq d/2$ for *d*-dimensional linear bandit

Relation to TS and Regret Bound

- A regret bound that applies to all algorithms

$$E[\text{Regret}(T)] \leq \sqrt{\overline{\Psi}_T H(X^*) T}$$ [Russo-Van Roy, 2014]

$$\overline{\Psi}_T = \frac{1}{T} \sum_{t=1}^{T} E[\Psi_t]$$

- For IDS:
  - $\overline{\Psi}_t \leq |\mathbb{X}|/2$ always
  - $\overline{\Psi}_t \leq d/2$ for d-dimensional linear bandit
  - $1/2$ with full feedback

Relation to TS and Regret Bound

- A regret bound that applies to all algorithms

$$E[\text{Regret}(T)] \leq \sqrt{\overline{\Psi}_T H(X^*) T}$$

[Russo-Van Roy, 2014]

$$\overline{\Psi}_T = \frac{1}{T} \sum_{t=1}^{T} E[\Psi_t]$$

- For IDS:
  - $\overline{\Psi}_t \leq |\mathbb{X}|/2$ always
  - $\overline{\Psi}_t \leq d/2$ for d-dimensional linear bandit
  - $1/2$ with full feedback

- Grew out of information-theoretic analysis of TS

[Russo-Van Roy, 2014]

Computational Considerations

Computational Considerations

- Tractable implementations for several cases

Computational Considerations

- Tractable implementations for several cases
  - Beta-Bernouli bandit (independent arms)

Computational Considerations

- Tractable implementations for several cases
  - Beta-Bernouli bandit (independent arms)
  - Gaussian bandit (independent arms)

Computational Considerations

- Tractable implementations for several cases
  - Beta-Bernouli bandit (independent arms)
  - Gaussian bandit (independent arms)
  - Linear bandit (mean-based IDS)

Computational Considerations

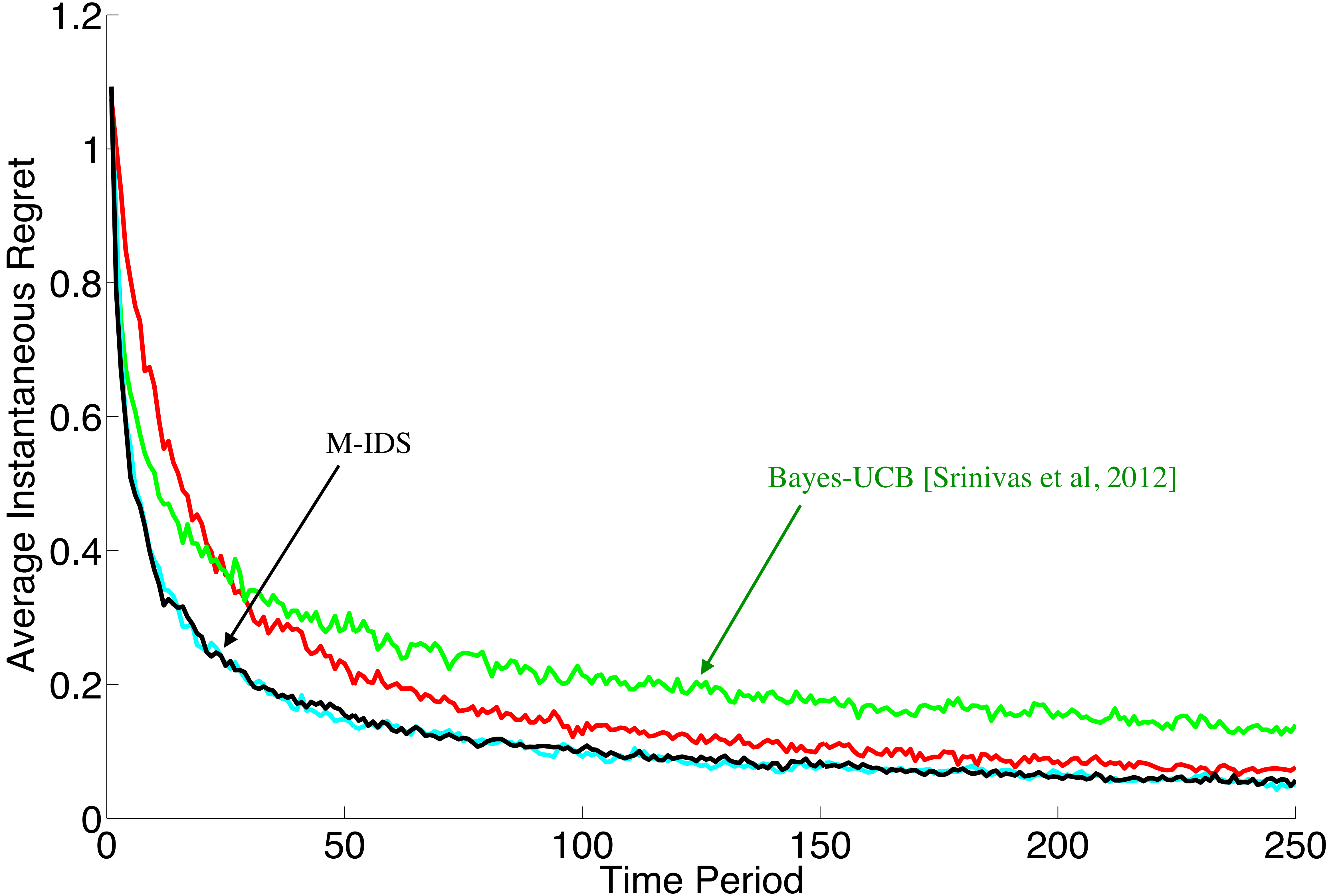
- Tractable implementations for several cases
  - Beta-Bernouli bandit (independent arms)
  - Gaussian bandit (independent arms)
  - Linear bandit (mean-based IDS)

- UCB/TS do well in these cases

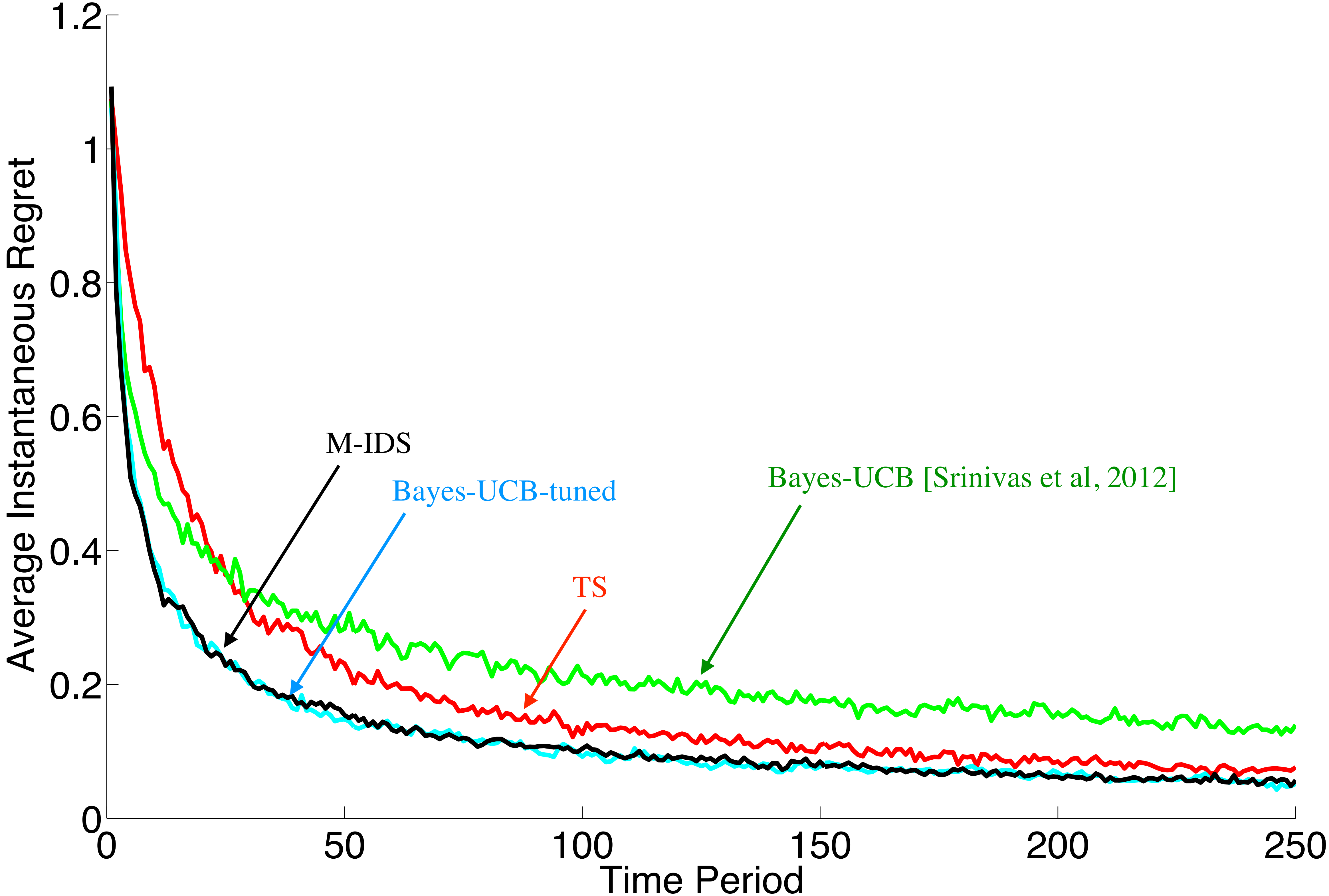
Computational Considerations

- Tractable implementations for several cases
  - Beta-Bernouli bandit (independent arms)
  - Gaussian bandit (independent arms)
  - Linear bandit (mean-based IDS)
- UCB/TS do well in these cases
- New algorithms needed for other cases

Linear Bandit Simulation



Linear Bandit Simulation



Summary on IDS

Summary on IDS

- IDS addresses cases where UCB/TS miserably fails

Summary on IDS

- IDS addresses cases where UCB/TS miserably fails
- IDS accomplishes this by measuring information gain

Summary on IDS

- IDS addresses cases where UCB/TS miserably fails
- IDS accomplishes this by measuring information gain
- IDS performs as well or better than UCB/TS in several cases where all are tractable

Summary on IDS

- IDS addresses cases where UCB/TS miserably fails
- IDS accomplishes this by measuring information gain
- IDS performs as well or better than UCB/TS in several cases where all are tractable
- New algorithms are needed to implement IDS in other cases, especially those in which UCB/TS miserably fail