Pairwise Elimination for Bandits with Cost Subsidy

Ishank Juneja, Carlee Joe-Wong, Osman Yağan

ICLR 2025

Arm 2

$\mu_2 = 0.6$

Arm 1

$\mu_1 = 0.4$

Arm 3

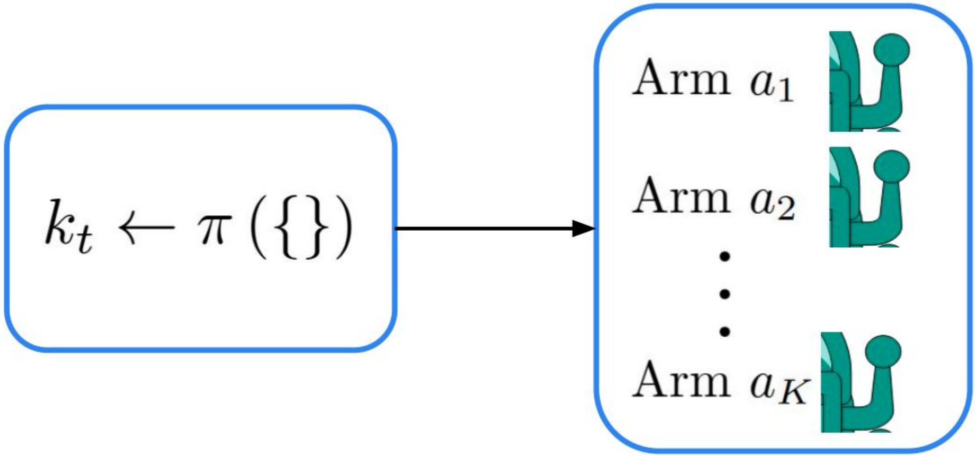
$\mu_3 = 0.5$

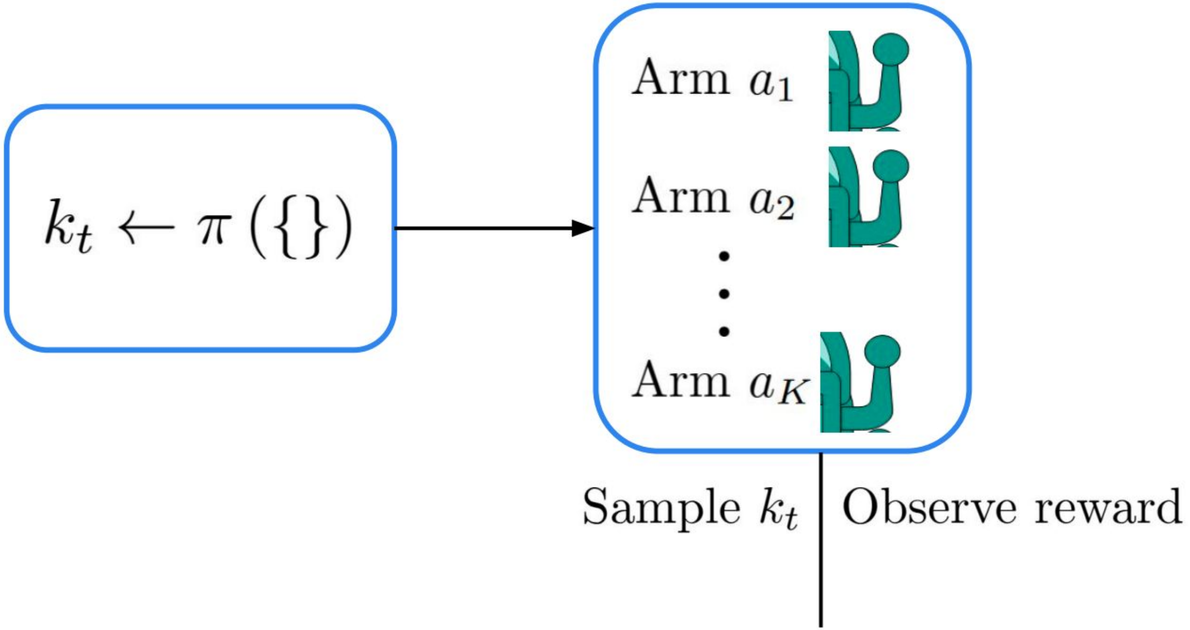
Arm 4

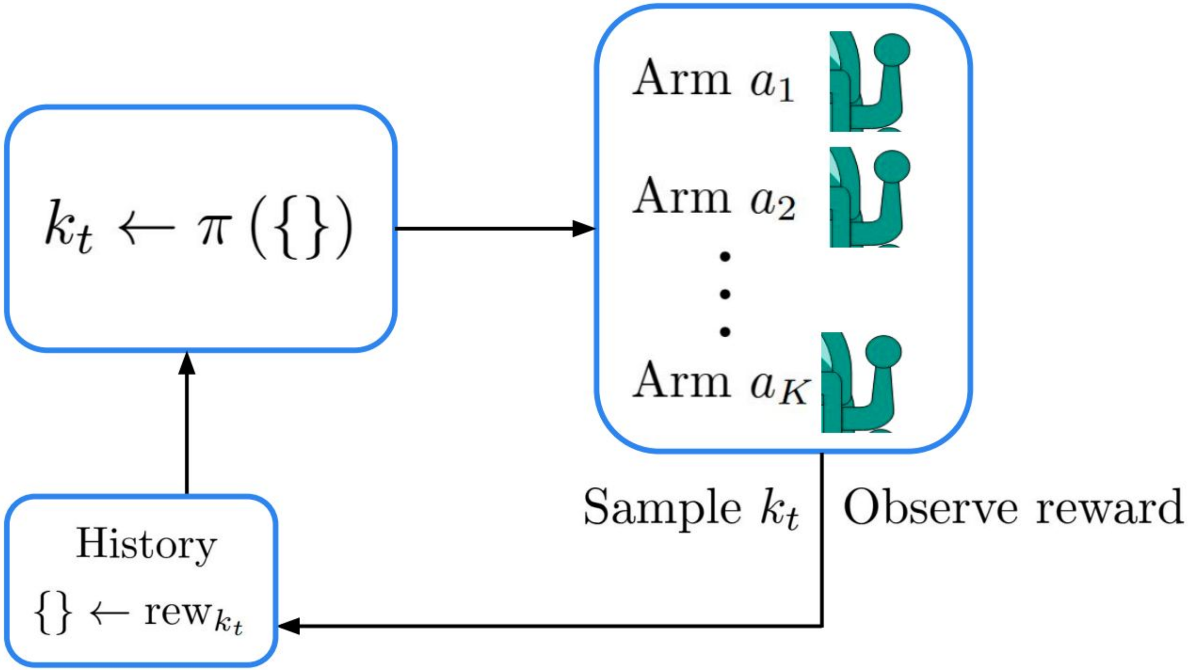
$\mu_4 = 0.9$

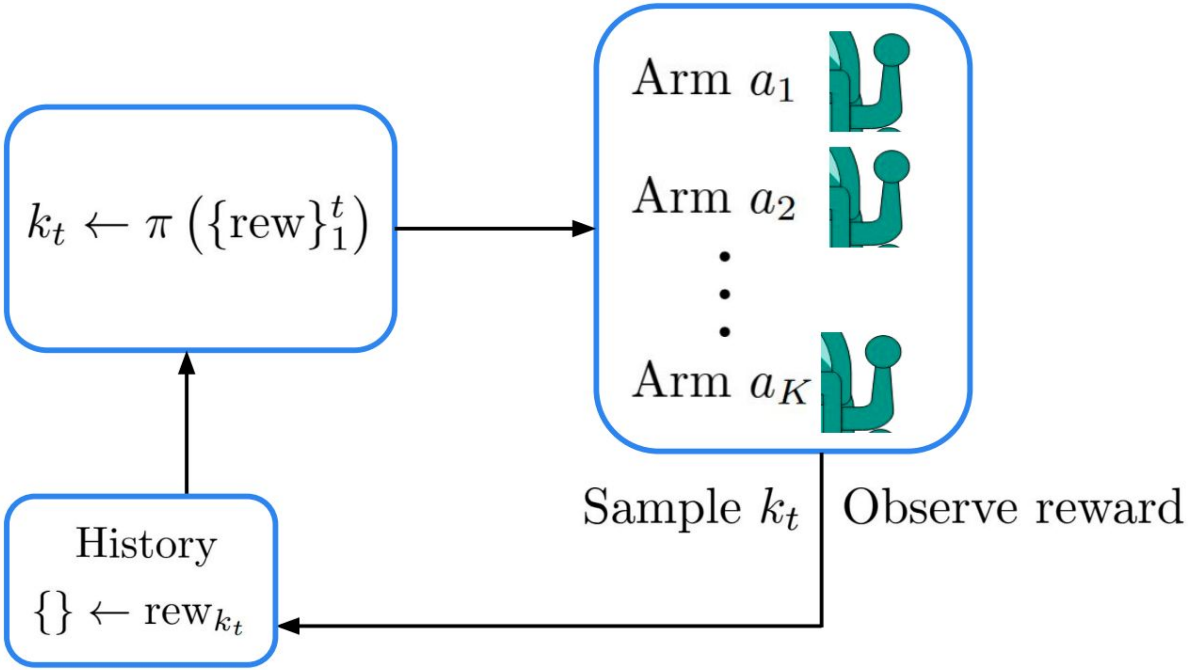
$$\boxed{\pi}$$

$$k_t \leftarrow \pi\left(\{\}\right)$$

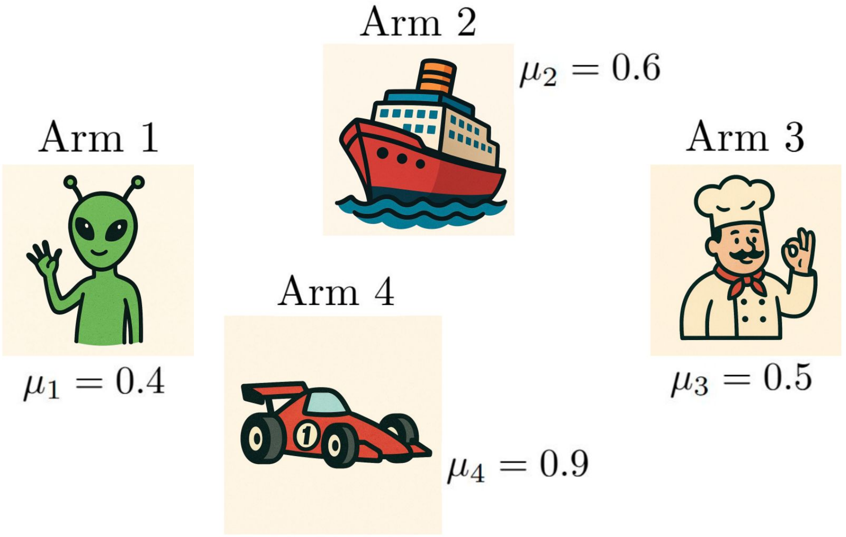

$k_t \leftarrow \pi(\{\})$
Arm a_1
Arm a_2
Arm a_K


$k_t \leftarrow \pi\left(\{\}\right)$
Arm a_1
Arm a_2
Arm a_K
Sample k_t
Observe reward


$k_t \leftarrow \pi(\{\})$
Arm a_1
Arm a_2
Arm a_K
Sample k_t
Observe reward
History
$\{\} \leftarrow \mathrm{rew}_{k_t}$


$k_t \leftarrow \pi\left(\{\text{rew}\}_1^t\right)$
Arm a_1
Arm a_2
Arm a_K
Sample k_t
Observe reward
History
$\{\} \leftarrow \text{rew}_{k_t}$

$$i^* = \arg\max_{i \in [K]} \mu_i$$

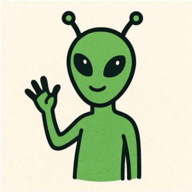


$$i^* = \arg\max_{i \in [K]} \mu_i$$

Arm 2

$\mu_2 = 0.6$

Arm 1



$\mu_1 = 0.4$

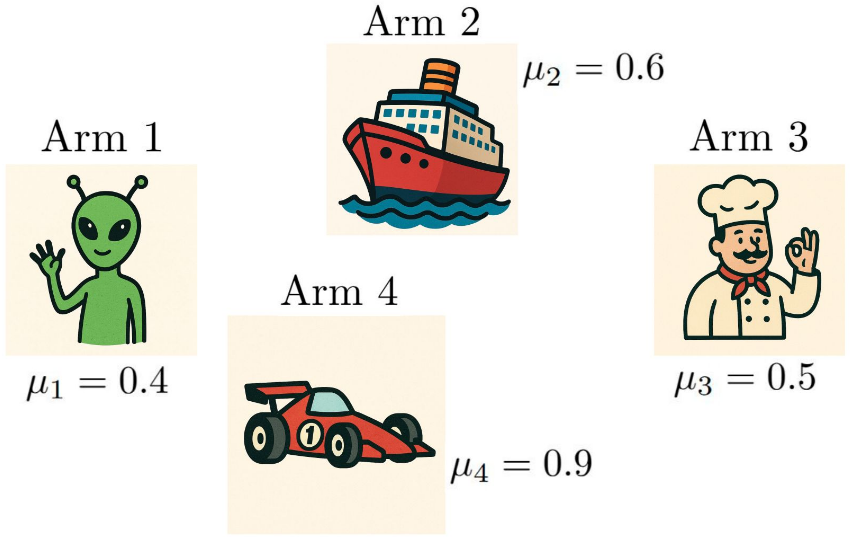
Arm 3

$\mu_3 = 0.5$

Arm 4

$\mu_4 = 0.9$

$$i^* = \arg\max_{i \in [K]} \mu_i$$

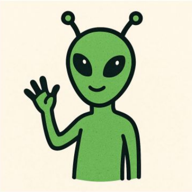


$$i^* = \arg\max_{i\in[K]} \mu_i \quad (i^* = 4)$$

Arm 2

$\mu_2 = 0.6$

Arm 1



$\mu_1 = 0.4$

Arm 3

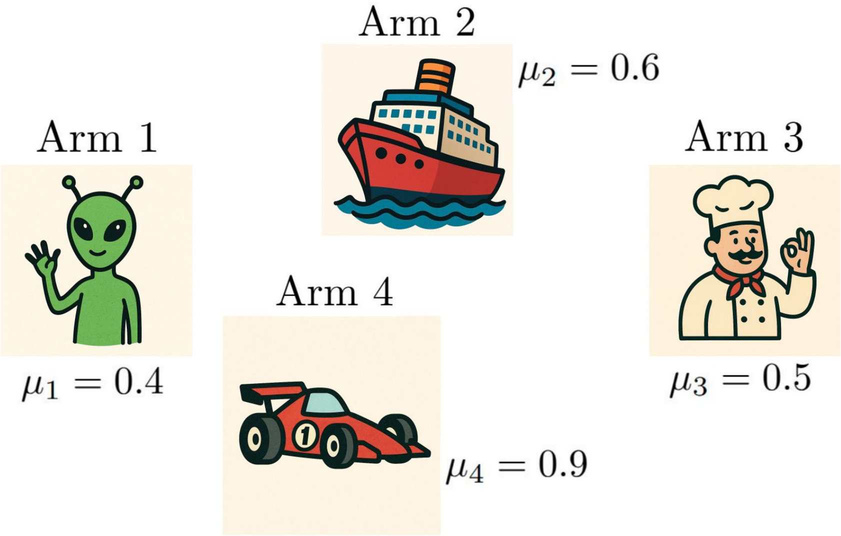
$\mu_3 = 0.5$

Arm 4

$\mu_4 = 0.9$

$$i^* = \arg\max_{i \in [K]} \mu_i \quad (i^* = 4)$$

$$\mu^* = \max_{i \in [K]} \mu_i \quad (\mu^* = 0.9)$$



$$i^* = \arg\max_{i \in [K]} \mu_i \qquad (i^* = 4)$$

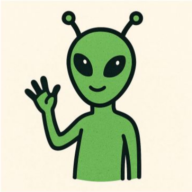
$$\mu^* = \max_{i \in [K]} \mu_i \qquad (\mu^* = 0.9)$$

$$\mathrm{Reg}(T, \nu) = \sum_{t=1}^{T} \mathbb{E}\left[\mu^* - \mu_{k_t}\right]$$

Arm 2

$\mu_2 = 0.6$

Arm 1



$\mu_1 = 0.4$

Arm 3

$\mu_3 = 0.5$

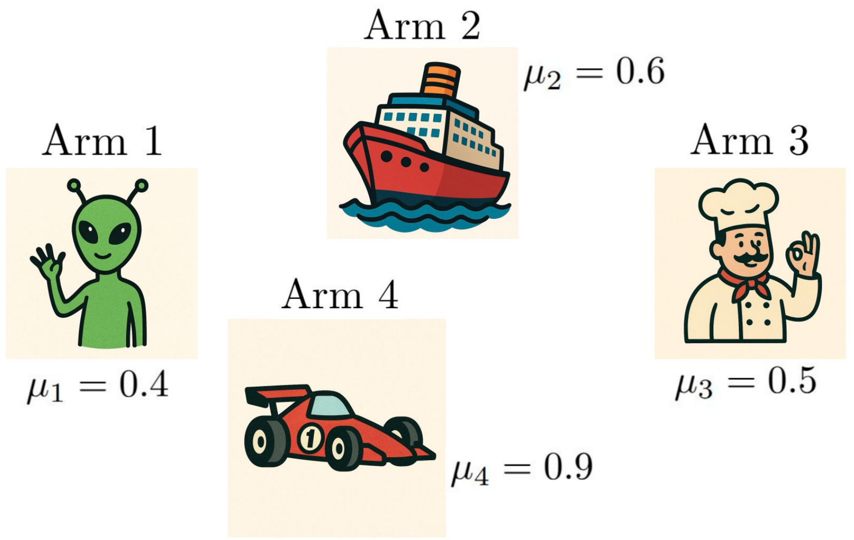
Arm 4

$\mu_4 = 0.9$

$$i^* = \arg\max_{i \in [K]} \mu_i \qquad (i^* = 4)$$

$$\mu^* = \max_{i \in [K]} \mu_i \qquad (\mu^* = 0.9)$$

$$\mathrm{Reg}(T, \nu) = \sum_{t=1}^{T} \mathbb{E}\left[\mu^* - \mu_{k_t}\right]$$

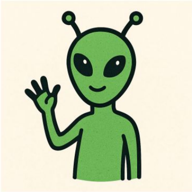


$$i^* = \arg\max_{i\in[K]} \mu_i \qquad (i^* = 4)$$

$$\mu^* = \max_{i\in[K]} \mu_i \qquad (\mu^* = 0.9)$$

$$\begin{aligned}\mathrm{Reg}(T,\nu) &= \sum_{t=1}^{T} \mathbb{E}\left[\mu^* - \mu_{k_t}\right] \\ &= \sum_{i\in[K]} \Delta_i \mathbb{E}\left[n_i(T)\right]\end{aligned}$$

Arm 1



$\mu_1 = 0.4$

Arm 2

$\mu_2 = 0.6$

Arm 3

$\mu_3 = 0.5$

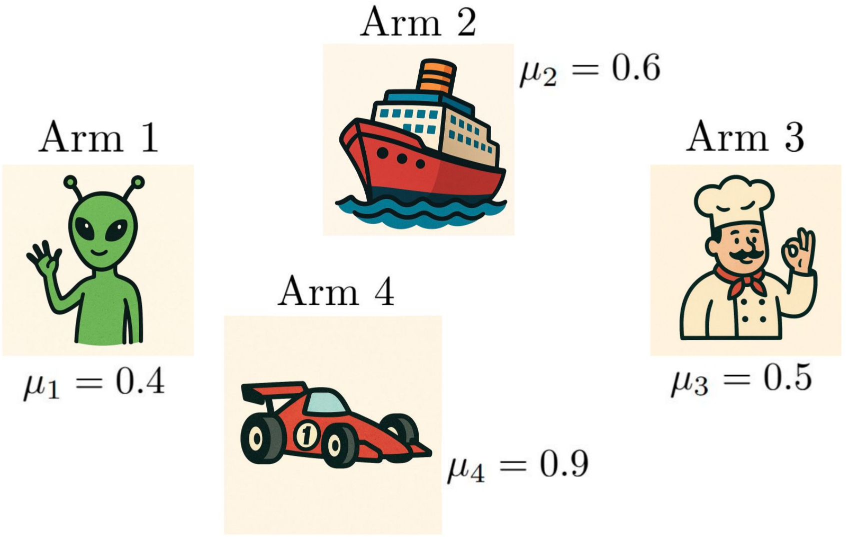
Arm 4

$\mu_4 = 0.9$

$$i^* = \arg\max_{i \in [K]} \mu_i \qquad (i^* = 4)$$

$$\mu^* = \max_{i \in [K]} \mu_i \qquad (\mu^* = 0.9)$$

$$\mathrm{Reg}(T, \nu) = \sum_{t=1}^{T} \mathbb{E}\left[\mu^* - \mu_{k_t}\right] = \sum_{i \in [K]} \Delta_i \mathbb{E}\left[n_i(T)\right]$$

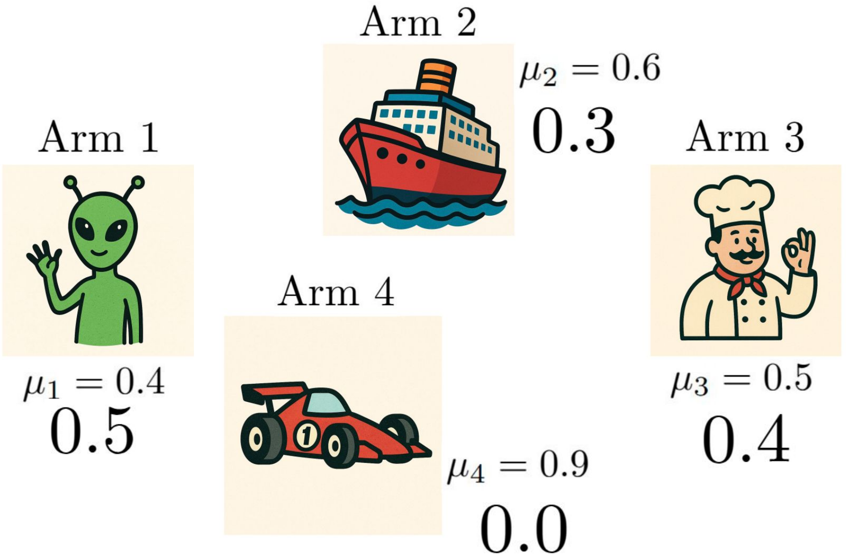


$$i^* = \arg\max_{i \in [K]} \mu_i \qquad (i^* = 4)$$

$$\mu^* = \max_{i \in [K]} \mu_i \qquad (\mu^* = 0.9)$$

$$\Delta_i = \mu^* - \mu_i$$

$$\begin{aligned} \mathrm{Reg}(T, \nu) &= \sum_{t=1}^{T} \mathbb{E}\left[\mu^* - \mu_{k_t}\right] \\ &= \sum_{i \in [K]} \Delta_i \mathbb{E}\left[n_i(T)\right] \end{aligned}$$



$$i^* = \arg\max_{i \in [K]} \mu_i \quad (i^* = 4)$$

$$\mu^* = \max_{i \in [K]} \mu_i \quad (\mu^* = 0.9)$$

$$\Delta_i = \mu^* - \mu_i$$

$$\mathrm{Reg}(T, \nu) = \sum_{t=1}^{T} \mathbb{E}\left[\mu^* - \mu_{k_t}\right] = \sum_{i \in [K]} \Delta_i \mathbb{E}\left[n_i(T)\right]$$

Arm 2

$\mu_2 = 0.6$

Arm 1

$\mu_1 = 0.4$

Arm 3

$\mu_3 = 0.5$

Arm 4

$\mu_4 = 0.9$

Arm 2

$\mu_2 = 0.6$
$c_2 = \$2$

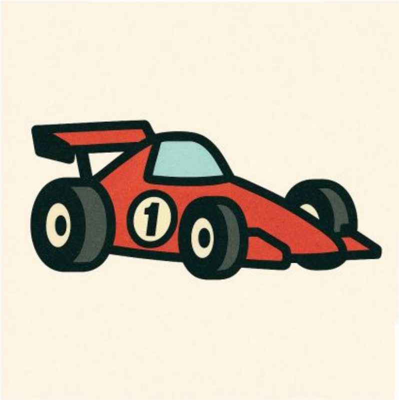
Arm 1

$\mu_1 = 0.4$
$c_1 = \$1$

Arm 3

$\mu_3 = 0.5$
$c_3 = \$5$

Arm 4



$\mu_4 = 0.9$
$c_4 = \$10$

$\mu_{k_t} \geq \mu_{\text{CS}}$ (In Aggregate)

Remove the reference to \mu_{\text{CS}} here, too early for it

Arm 2

$\mu_2 = 0.6$
$c_2 = \$2$

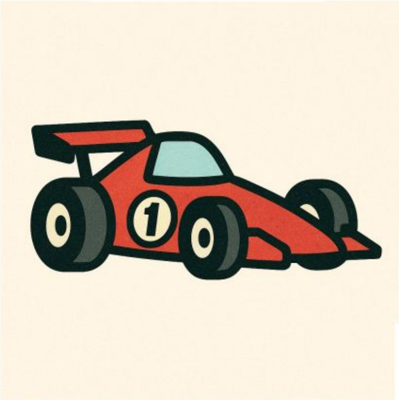
Arm 1

$\mu_1 = 0.4$
$c_1 = \$1$

Arm 3

$\mu_3 = 0.5$
$c_3 = \$5$

Arm 4



$\mu_4 = 0.9$
$c_4 = \$10$

Outline

- Contributions and Related Work
- MAB-CS objective
- PE and PE-CS Algorithms for MAB-CS
- Experiments
- Theoretical Results
    - Upper Bounds
    - Lower Bounds

Outline

- **Contributions and Related Work**
- MAB-CS objective
- PE and PE-CS Algorithms for MAB-CS
- Experiments
- Theoretical Results
  - Upper Bounds
  - Lower Bounds

Conservative Bandits

*Wu et al. Conservative Bandits. ICML 2016.

Conservative Bandits

$$\min\ \mathrm{Reg}(T, \nu)$$

*Wu et al. Conservative Bandits.
ICML 2016.

Conservative Bandits

$$\min \ \mathrm{Reg}(T, \nu)$$

$$\text{s.t.} \sum_{t=1}^{t_0} \mathrm{rew}_{k_t} \geq (1-\alpha) \sum_{t=1}^{t_0} \mathrm{rew}_{\ell}$$

$$\forall \ t_0 \leq T$$

*Wu et al. Conservative Bandits.
ICML 2016.

Conservative Bandits

$$\min \ \mathrm{Reg}(T, \nu)$$

$$\text{s.t.} \sum_{t=1}^{t_0} \mathrm{rew}_{k_t} \geq (1-\alpha) \sum_{t=1}^{t_0} \mathrm{rew}_{\ell}$$

$$\forall \ t_0 \leq T$$

✓ Quality Constraints

- Anytime constraint

✗ Sampling Costs

*Wu et al. Conservative Bandits. ICML 2016.

Conservative Bandits

$$\min \ \mathrm{Reg}(T, \nu)$$

$$\text{s.t.} \sum_{t=1}^{t_0} \mathrm{rew}_{k_t} \geq (1-\alpha) \sum_{t=1}^{t_0} \mathrm{rew}_{\ell}$$

$$\forall \ t_0 \leq T$$

✓ Quality Constraints

- Anytime constraint

✗ Sampling Costs

*Wu et al. Conservative Bandits. ICML 2016.

Bandits w/ Knapsacks

*Badanidiyuru et al. Bandits with knapsacks. JACM (2018).

Conservative Bandits

$$\min \ \mathrm{Reg}(T,\nu)$$
$$\text{s.t.} \sum_{t=1}^{t_0} \mathrm{rew}_{k_t} \geq (1-\alpha) \sum_{t=1}^{t_0} \mathrm{rew}_{\ell}$$
$$\forall \ t_0 \leq T$$

✓ Quality Constraints

- Anytime constraint

✗ Sampling Costs

*Wu et al. Conservative Bandits. ICML 2016.

Bandits w/ Knapsacks

$$\min \ \mathrm{Reg}(T,\nu)$$

*Badanidiyuru et al. Bandits with knapsacks. JACM (2018).

Conservative Bandits

$$\min \ \mathrm{Reg}(T, \nu)$$

$$\text{s.t.} \sum_{t=1}^{t_0} \mathrm{rew}_{k_t} \geq (1-\alpha) \sum_{t=1}^{t_0} \mathrm{rew}_{\ell}$$

$$\forall \ t_0 \leq T$$

✓ Quality Constraints

- Anytime constraint

✗ Sampling Costs

*Wu et al. Conservative Bandits. ICML 2016.

Bandits w/ Knapsacks

$$\min \ \mathrm{Reg}(T, \nu)$$

$$\text{s.t.} \sum_{t=1}^{T} c_{k_t} \leq B$$

*Badanidiyuru et al. Bandits with knapsacks. JACM (2018).

Conservative Bandits

$$\min \ \mathrm{Reg}(T,\nu)$$

$$\text{s.t.} \sum_{t=1}^{t_0} \mathrm{rew}_{k_t} \geq (1-\alpha) \sum_{t=1}^{t_0} \mathrm{rew}_{\ell}$$

$$\forall \ t_0 \leq T$$

✓ Quality Constraints

– Anytime constraint

✗ Sampling Costs

*Wu et al. Conservative Bandits. ICML 2016.

Bandits w/ Knapsacks

$$\min \ \mathrm{Reg}(T,\nu)$$

$$\text{s.t.} \sum_{t=1}^{T} c_{k_t} \leq B$$

✗ Quality Constraints

✓ Sampling Costs

– Fixed Budget

*Badanidiyuru et al. Bandits with knapsacks. JACM (2018).

Conservative Bandits

$$\min \ \mathrm{Reg}(T,\nu)$$
$$\text{s.t.} \sum_{t=1}^{t_0} \mathrm{rew}_{k_t} \geq (1-\alpha) \sum_{t=1}^{t_0} \mathrm{rew}_\ell$$
$$\forall \ t_0 \leq T$$

✓ Quality Constraints

– Anytime constraint

✗ Sampling Costs

*Wu et al. Conservative Bandits. ICML 2016.

Bandits w/ Knapsacks

$$\min \ \mathrm{Reg}(T,\nu)$$
$$\text{s.t.} \sum_{t=1}^{T} c_{k_t} \leq B$$

✗ Quality Constraints

✓ Sampling Costs

– Fixed Budget

*Badanidiyuru et al. Bandits with knapsacks. JACM (2018).

MAB w/ Cost-Subsidy

*Sinha et al. Multi-armed bandits with cost subsidy. AISTATS 2021.

Conservative Bandits

$$\min \ \mathrm{Reg}(T,\nu)$$

$$\text{s.t.} \sum_{t=1}^{t_0} \mathrm{rew}_{k_t} \geq (1-\alpha)\sum_{t=1}^{t_0} \mathrm{rew}_\ell$$

$$\forall \ t_0 \leq T$$

✓ Quality Constraints

– Anytime constraint

✗ Sampling Costs

*Wu et al. Conservative Bandits. ICML 2016.

Bandits w/ Knapsacks

$$\min \ \mathrm{Reg}(T,\nu)$$

$$\text{s.t.} \sum_{t=1}^{T} c_{k_t} \leq B$$

✗ Quality Constraints

✓ Sampling Costs

– Fixed Budget

*Badanidiyuru et al. Bandits with knapsacks. JACM (2018).

MAB w/ Cost-Subsidy

$$\min \ \mathrm{Cost\_Reg}(T,\nu),$$

$$\min \ \mathrm{Qual\_Reg}(T,\nu)$$

*Sinha et al. Multi-armed bandits with cost subsidy. AISTATS 2021.

Conservative Bandits

$$\min \ \mathrm{Reg}(T,\nu)$$

$$\text{s.t.} \sum_{t=1}^{t_0} \mathrm{rew}_{k_t} \geq (1-\alpha) \sum_{t=1}^{t_0} \mathrm{rew}_\ell$$

$$\forall \ t_0 \leq T$$

✓ Quality Constraints

- Anytime constraint

✗ Sampling Costs

*Wu et al. Conservative Bandits. ICML 2016.

Bandits w/ Knapsacks

$$\min \ \mathrm{Reg}(T,\nu)$$

$$\text{s.t.} \sum_{t=1}^{T} c_{k_t} \leq B$$

✗ Quality Constraints

✓ Sampling Costs

- Fixed Budget

*Badanidiyuru et al. Bandits with knapsacks. JACM (2018).

MAB w/ Cost-Subsidy

$$\min \ \mathrm{Cost\_Reg}(T,\nu),$$
$$\min \ \mathrm{Qual\_Reg}(T,\nu)$$

✓ Quality Constraints

✓ Sampling Costs

*Sinha et al. Multi-armed bandits with cost subsidy. AISTATS 2021.

Conservative Bandits

$$\min \ \mathrm{Reg}(T,\nu)$$

$$\text{s.t.} \sum_{t=1}^{t_0} \mathrm{rew}_{k_t} \geq (1-\alpha) \sum_{t=1}^{t_0} \mathrm{rew}_\ell$$

$$\forall \ t_0 \leq T$$

✓ Quality Constraints

- Anytime constraint

✗ Sampling Costs

*Wu et al. Conservative Bandits. ICML 2016.

Bandits w/ Knapsacks

$$\min \ \mathrm{Reg}(T,\nu)$$

$$\text{s.t.} \sum_{t=1}^{T} c_{k_t} \leq B$$

✗ Quality Constraints

✓ Sampling Costs

- Fixed Budget

*Badanidiyuru et al. Bandits with knapsacks. JACM (2018).

MAB w/ Cost-Subsidy

$$\min \ \mathrm{Cost\_Reg}(T,\nu),$$

$$\min \ \mathrm{Qual\_Reg}(T,\nu)$$

✓ Quality Constraints

✓ Sampling Costs

✗ Variation on μ_{CS} (thresh)

✗ Instance dependent analysis

*Sinha et al. Multi-armed bandits with cost subsidy. AISTATS 2021.

Related Work: Sinha et al

*Sinha et al. Multi-armed bandits with cost subsidy. AISTATS 2021.

Related Work: Sinha et al

- Introduced objectives cost regret and quality regret

*Sinha et al. Multi-armed bandits with cost subsidy. AISTATS 2021.

Related Work: Sinha et al

- Introduced objectives cost regret and quality regret
- Works with $\mu_{\text{CS}} = (1 - \alpha)\mu^*$

*Sinha et al. Multi-armed bandits with cost subsidy. AISTATS 2021.

Related Work: Sinha et al

- Introduced objectives cost regret and quality regret
- Works with $\mu_{\text{CS}} = (1 - \alpha)\mu^*$
- Introduces the ETC-CS, UCB-CS, and TS-CS algorithms

*Sinha et al. Multi-armed bandits with cost subsidy. AISTATS 2021.

Related Work: Sinha et al

- Introduced objectives cost regret and quality regret
- Works with $\mu_{\text{CS}} = (1-\alpha)\mu^*$
- Introduces the ETC-CS, UCB-CS, and TS-CS algorithms
- ETC-CS: $O(T^{2/3})$ guarantee
  $\text{Cost\_Reg}(T, \nu) + \text{Qual\_Reg}(T, \nu)$

*Sinha et al. Multi-armed bandits with cost subsidy. AISTATS 2021.

*Juneja et al. Pairwise Elimination for bandits with cost subsidy. ICLR 2025.

Setting Name	μ_{CS}
Fixed Threshold	$\mu_0 \in \mathbb{R}$
Known Reference Arm	$(1-\alpha)\mu_\ell$
Subsidized Best Reward	$(1-\alpha)\mu^\star$

*Juneja et al. Pairwise Elimination for bandits with cost subsidy. ICLR 2025.

Setting Name	μ_{CS}
Fixed Threshold	$\mu_0 \in \mathbb{R}$
Known Reference Arm	$(1-\alpha)\mu_\ell$
Subsidized Best Reward	$(1-\alpha)\mu^\star$

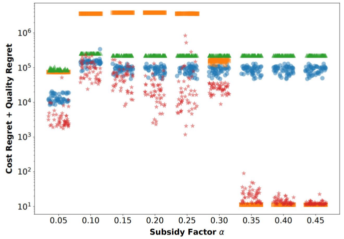
Known Reference Arm	PE
Subsidized Best Reward	PE-CS

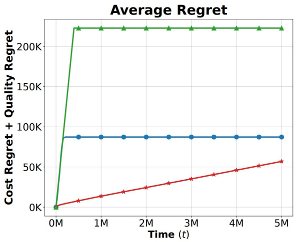
*Juneja et al. Pairwise Elimination for bandits with cost subsidy. ICLR 2025.

Setting Name	μ_{CS}
Fixed Threshold	$\mu_0 \in \mathbb{R}$
Known Reference Arm	$(1-\alpha)\mu_\ell$
Subsidized Best Reward	$(1-\alpha)\mu^\star$

Known Reference Arm	PE
Subsidized Best Reward	PE-CS

Experimental validation



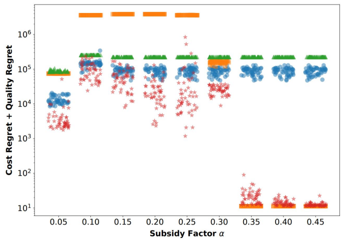


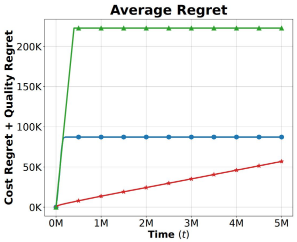
*Juneja et al. Pairwise Elimination for bandits with cost subsidy. ICLR 2025.

Setting Name	μ_{CS}
Fixed Threshold	$\mu_0 \in \mathbb{R}$
Known Reference Arm	$(1-\alpha)\mu_\ell$
Subsidized Best Reward	$(1-\alpha)\mu^\star$

Known Reference Arm	PE
Subsidized Best Reward	PE-CS

Experimental validation





*Juneja et al. Pairwise Elimination for bandits with cost subsidy. ICLR 2025.

Theory: Instance ν dependent upper bounds

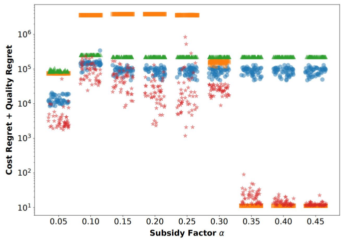
– PE

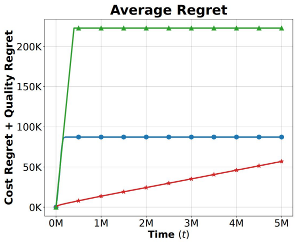
– PE-CS

Setting Name	μ_{CS}
Fixed Threshold	$\mu_0 \in \mathbb{R}$
Known Reference Arm	$(1-\alpha)\mu_\ell$
Subsidized Best Reward	$(1-\alpha)\mu^\star$

Known Reference Arm	PE
Subsidized Best Reward	PE-CS

Experimental validation





*Juneja et al. Pairwise Elimination for bandits with cost subsidy. ICLR 2025.

Theory: Instance ν dependent upper bounds

- PE
- PE-CS

Theory: Instance ν dependent lower bounds

Outline

- Contributions and Related Work
- **MAB-CS objective**
- PE and PE-CS Algorithms for MAB-CS
- Experiments
- Theoretical Results
    - Upper Bounds
    - Lower Bounds

$$a^* = \arg\min_{i \in S} c_i,$$

$$a^* = \arg\min_{i \in S} c_i,$$

$$S = \{i \in [K] : \mu_i \geq \mu_{\text{CS}}\}$$

$$a^* = \arg\min_{i \in S} c_i,$$

$$S = \{i \in [K] : \mu_i \geq \mu_{\text{CS}}\}$$

$\mu_1 = 0.4$
$c_1 = \$1$

$\mu_2 = 0.6$
$c_2 = \$2$

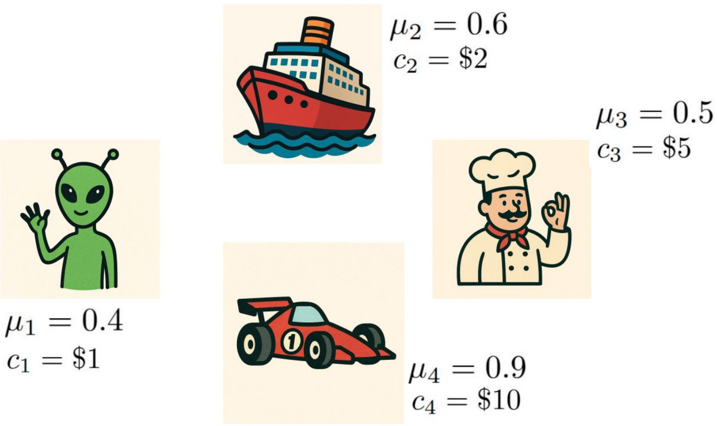
$\mu_3 = 0.5$
$c_3 = \$5$

$\mu_4 = 0.9$
$c_4 = \$10$

$$\alpha = 0.5$$

$$a^* = \arg\min_{i \in S} c_i,$$

$$S = \{i \in [K] : \mu_i \geq \mu_{\text{CS}}\}$$



$$\alpha = 0.5$$

$$\ell = 4 \implies \mu_{\text{CS}} = (1-\alpha)\mu_\ell = 0.45$$

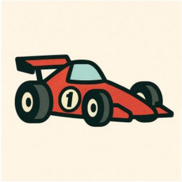
$$a^* = \arg\min_{i \in S} c_i,$$

$$S = \{i \in [K] : \mu_i \geq \mu_{\text{CS}}\}$$

$\mu_1 = 0.4$
$c_1 = \$1$

$\mu_2 = 0.6$
$c_2 = \$2$

$\mu_3 = 0.5$
$c_3 = \$5$



$\mu_4 = 0.9$
$c_4 = \$10$

$$\alpha = 0.5$$

$$\ell = 4 \implies \mu_{\mathrm{CS}} = (1-\alpha)\mu_\ell = 0.45$$

$$i^* = 4 \implies \mu_{\mathrm{CS}} = (1-\alpha)\mu^* = 0.45$$

$$a^* = \arg\min_{i \in S} c_i,$$

$$S = \{ i \in [K] : \mu_i \geq \mu_{\mathrm{CS}} \}$$

$\mu_2 = 0.6$
$c_2 = \$2$

$\mu_3 = 0.5$
$c_3 = \$5$

$\mu_1 = 0.4$
$c_1 = \$1$

$\mu_4 = 0.9$
$c_4 = \$10$

$$\alpha = 0.5$$

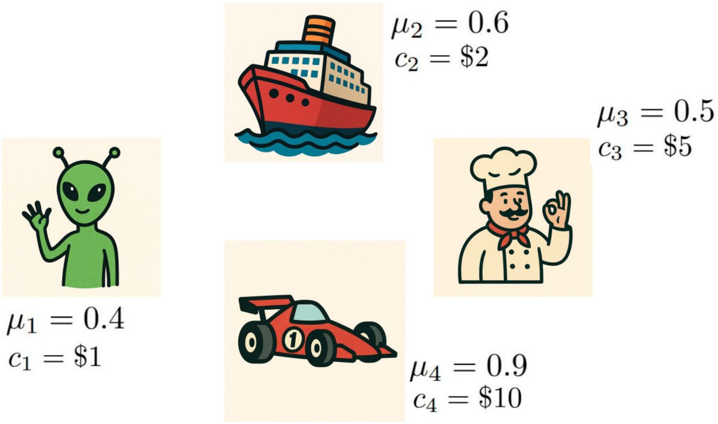
$$\ell = 4$$

$$i^* = 4$$

$$\mu_{\text{CS}} = 0.45$$

$$a^* = \arg\min_{i \in S} c_i,$$

$$S = \{i \in [K] : \mu_i \geq \mu_{\text{CS}}\}$$



$$\alpha = 0.5$$

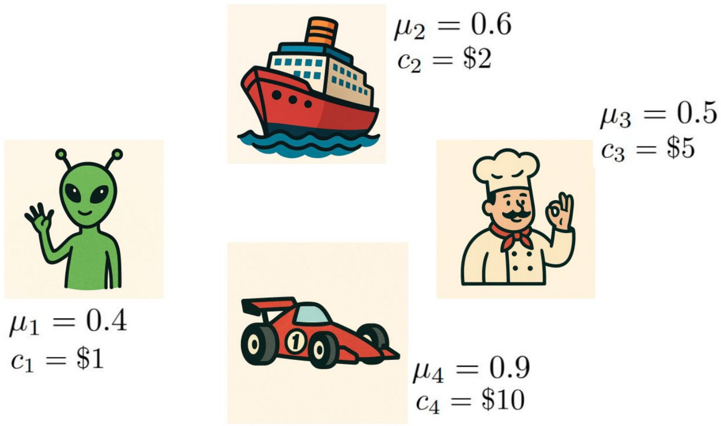
$$\ell = 4$$

$$i^* = 4$$

$$\mu_{\text{CS}} = 0.45$$

$$a^* = \arg\min_{i \in S} c_i,$$

$$S = \{i \in [K] : \mu_i \geq \mu_{\text{CS}}\} \quad (\{2, 3, 4\})$$



$$\alpha = 0.5$$

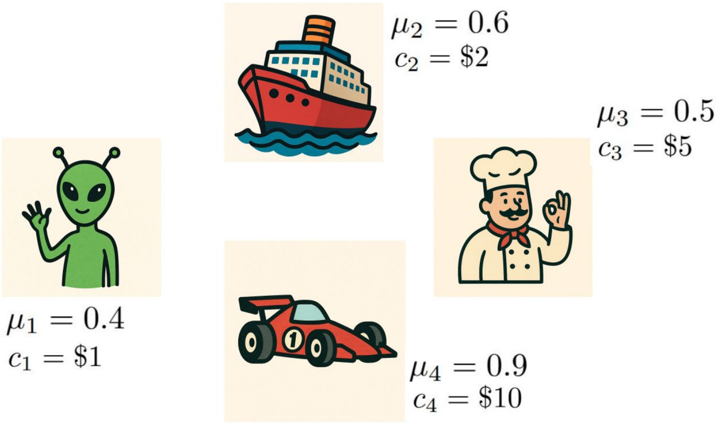
$$\ell = 4$$

$$i^* = 4$$

$$\mu_{\text{CS}} = 0.45$$

$$a^* = \arg\min_{i \in S} c_i, \quad \text{(Arm 2)}$$

$$S = \{i \in [K] : \mu_i \geq \mu_{\text{CS}}\} \quad (\{2, 3, 4\})$$



$$\alpha = 0.5$$

$$\ell = 4$$

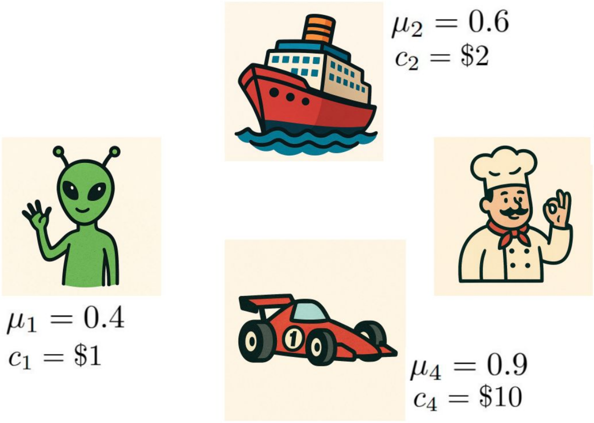
$$i^* = 4$$

$$\mu_{\text{CS}} = 0.45$$

$$a^* = \arg\min_{i \in S} c_i, \quad (\text{Arm } 2)$$

$$S = \{i \in [K] : \mu_i \geq \mu_{\text{CS}}\} \quad (\{2, 3, 4\})$$

$$\Delta_{Q,i} = \mu_{\text{CS}} - \mu_i$$



$$\alpha = 0.5$$

$$\ell = 4$$

$$i^* = 4$$

$$\mu_{\mathrm{CS}} = 0.45$$

$$a^* = \arg\min_{i \in S} c_i, \quad \text{(Arm 2)}$$

$$S = \{i \in [K] : \mu_i \geq \mu_{\mathrm{CS}}\} \quad (\{2, 3, 4\})$$

$$\Delta_{Q,i} = \mu_{\mathrm{CS}} - \mu_i$$

$$\Delta_{Q,i} = 0.05$$

$\mu_2 = 0.6$
$c_2 = \$2$

$\mu_3 = 0.5$
$c_3 = \$5$

$\mu_1 = 0.4$
$c_1 = \$1$

$\mu_4 = 0.9$
$c_4 = \$10$

$$\alpha = 0.5$$
$$\ell = 4$$
$$i^* = 4$$
$$\mu_{\mathrm{CS}} = 0.45$$

$$a^* = \arg\min_{i \in S} c_i, \quad \text{(Arm 2)}$$
$$S = \{i \in [K] : \mu_i \geq \mu_{\mathrm{CS}}\} \quad (\{2, 3, 4\})$$

$$\Delta_{Q,i} = \mu_{\mathrm{CS}} - \mu_i$$

$$\Delta_{Q,i} = 0.05$$

$\mu_2 = 0.6$
$c_2 = \$2$

$\mu_3 = 0.5$
$c_3 = \$5$

$\mu_1 = 0.4$
$c_1 = \$1$

$\Delta_{Q,i} < 0$

$\mu_4 = 0.9$
$c_4 = \$10$

$$\alpha = 0.5$$

$$\ell = 4$$

$$i^* = 4$$

$$\mu_{\text{CS}} = 0.45$$

$$a^* = \arg\min_{i \in S} c_i, \quad \text{(Arm 2)}$$

$$S = \{i \in [K] : \mu_i \geq \mu_{\text{CS}}\} \quad (\{2, 3, 4\})$$

$$\Delta_{Q,i} = \mu_{\text{CS}} - \mu_i$$

$$\Delta_{Q,i} = 0.05$$

$\mu_1 = 0.4$
$c_1 = \$1$

$\mu_2 = 0.6$
$c_2 = \$2$

$\Delta_{Q,i} < 0$

$\mu_3 = 0.5$
$c_3 = \$5$

$\mu_4 = 0.9$
$c_4 = \$10$

$$\alpha = 0.5$$

$$\ell = 4$$

$$i^* = 4$$

$$\mu_{\mathrm{CS}} = 0.45$$

$$a^* = \arg\min_{i \in S} c_i, \quad \text{(Arm 2)}$$

$$S = \{i \in [K] : \mu_i \geq \mu_{\mathrm{CS}}\} \quad (\{2, 3, 4\})$$

$$\Delta_{Q,i} = \mu_{\mathrm{CS}} - \mu_i$$

$$\Delta_{C,i} = c_i - c_{a^*}$$

$\Delta_{Q,i} = 0.05$

$\mu_1 = 0.4$
$c_1 = \$1$

$\mu_2 = 0.6$
$c_2 = \$2$

$\mu_3 = 0.5$
$c_3 = \$5$

$\Delta_{Q,i} < 0$

$\mu_4 = 0.9$
$c_4 = \$10$

$$\alpha = 0.5$$

$$\ell = 4$$

$$i^* = 4$$

$$\mu_{\mathrm{CS}} = 0.45$$

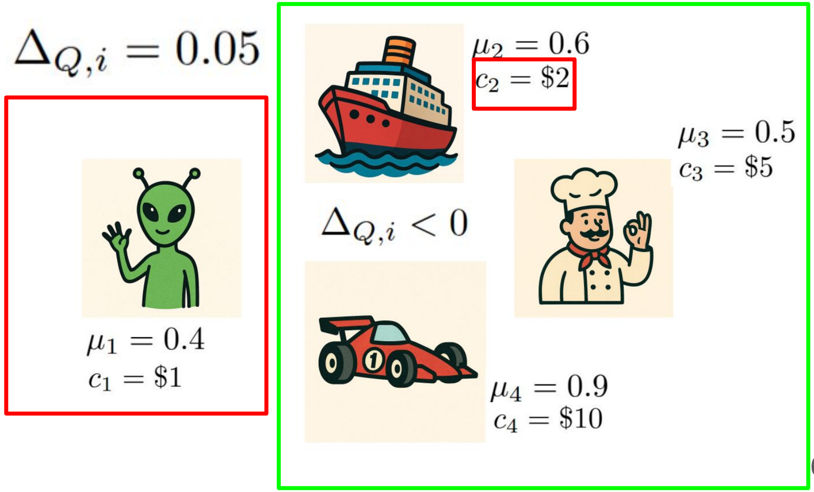
$$a^* = \arg\min_{i \in S} c_i, \quad \text{(Arm 2)}$$

$$S = \{i \in [K] : \mu_i \geq \mu_{\mathrm{CS}}\} \quad (\{2, 3, 4\})$$

$$\Delta_{Q,i} = \mu_{\mathrm{CS}} - \mu_i$$

$$\Delta_{C,i} = c_i - c_{a^*}$$

$$\Delta_{Q,i} = 0.05$$



$$\mathrm{Cost\_Reg}(T, \nu)$$

$$\text{Cost\_Reg}(T, \nu) = \sum_{t=1}^{T} \mathbb{E}\left[\Delta^{+}_{C,k_t}\right]$$

$$\text{Cost\_Reg}(T, \nu) = \sum_{t=1}^{T} \mathbb{E}\left[\Delta^{+}_{C,k_t}\right]$$

$$x^{+} = \max\{0, x\}$$

$$\mathrm{Cost\_Reg}(T,\nu) = \sum_{t=1}^{T} \mathbb{E}\left[\Delta^{+}_{C,k_t}\right] = \sum_{i=1}^{K} \Delta^{+}_{C,i}\mathbb{E}\left[n_i(T)\right]$$

$$x^{+} = \max\{0, x\}$$

$$\text{Cost\_Reg}(T,\nu) = \sum_{t=1}^{T} \mathbb{E}\left[\Delta^{+}_{C,k_t}\right] = \sum_{i=1}^{K} \Delta^{+}_{C,i}\mathbb{E}\left[n_i(T)\right]$$

$$\text{Qual\_Reg}(T,\nu) = \sum_{t=1}^{T} \mathbb{E}\left[\Delta^{+}_{Q,k_t}\right]$$

$$x^{+} = \max\{0, x\}$$

$$\mathrm{Cost\_Reg}(T,\nu) = \sum_{t=1}^{T} \mathbb{E}\left[\Delta^{+}_{C,k_t}\right] = \sum_{i=1}^{K} \Delta^{+}_{C,i}\mathbb{E}\left[n_i(T)\right]$$

$$\mathrm{Qual\_Reg}(T,\nu) = \sum_{t=1}^{T} \mathbb{E}\left[\Delta^{+}_{Q,k_t}\right] = \sum_{i=1}^{K} \Delta^{+}_{Q,i}\mathbb{E}\left[n_i(T)\right]$$

$$x^{+} = \max\{0, x\}$$

Outline

- Contributions and Related Work
- MAB-CS objective
- **PE and PE-CS Algorithms for MAB-CS**
- Experiments
- Theoretical Results
    - Upper Bounds
    - Lower Bounds

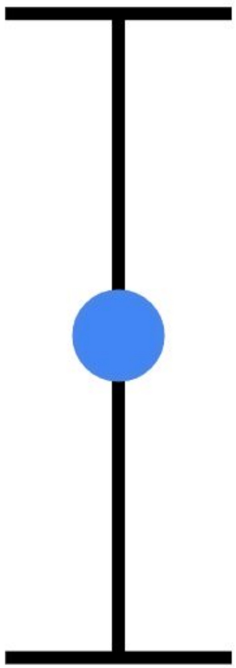
PE Algorithm

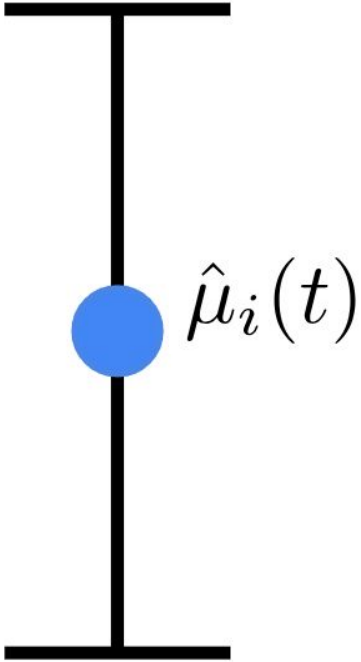
PE Algorithm

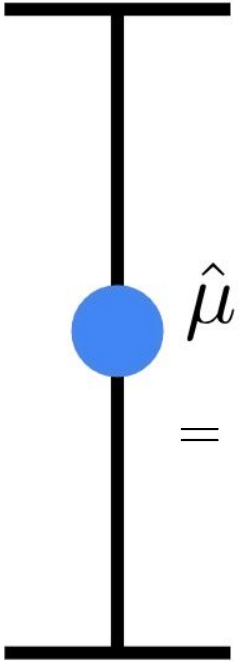
$$\mu_{\mathrm{CS}} = (1 - \alpha)\mu_\ell$$

PE Algorithm

$$\mu_{\mathrm{CS}} = (1 - \alpha)\mu_\ell$$

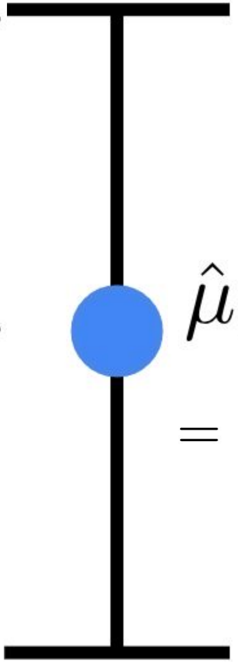



$\hat{\mu}_i(t)$



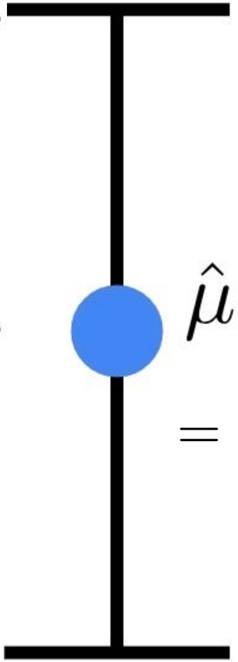
$$\hat{\mu}_i(t) = \frac{\sum_t \text{rewards sampled from arm } i}{n_i(t)}$$

$$\beta_{i,t} \propto^* \sqrt{\frac{\log T}{n_i(t)}}$$



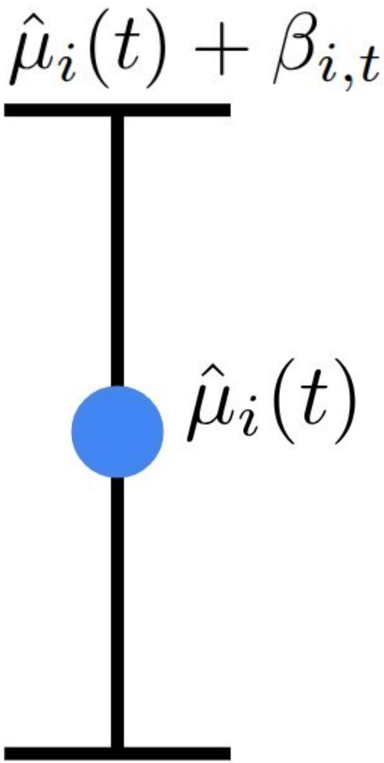
$$\hat{\mu}_i(t) = \frac{\sum_t \text{rewards sampled from arm } i}{n_i(t)}$$

$$\beta_{i,t} \propto^* \sqrt{\frac{\log T}{n_i(t)}}$$

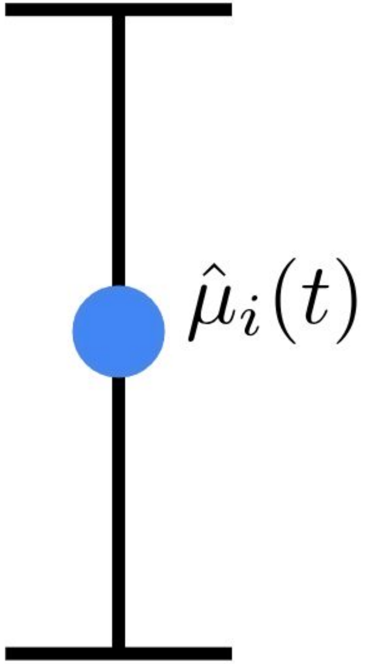


$$\hat{\mu}_i(t) = \frac{\sum_t \text{rewards sampled from arm } i}{n_i(t)}$$

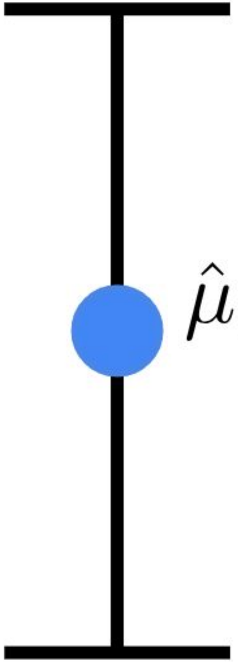
*Auer and Ortner. UCB revisited: Improved regret bounds for the stochastic multi-armed bandit problem. Periodica Mathematica Hungarica (2010).


$\hat{\mu}_i(t) + \beta_{i,t}$
$\hat{\mu}_i(t)$

$$\hat{\mu}_i(t) + \beta_{i,t} = \mathrm{UCB}_i$$



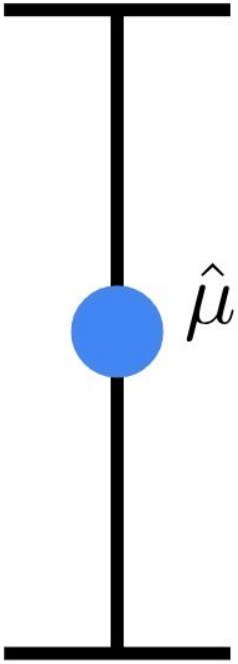
$$\hat{\mu}_i(t) + \beta_{i,t} = \mathrm{UCB}_i$$



$$\hat{\mu}_i(t)$$

$$\hat{\mu}_i(t) - \beta_{i,t} = \mathrm{LCB}_i$$

$$\hat{\mu}_i(t) + \beta_{i,t} = \mathrm{UCB}_i$$

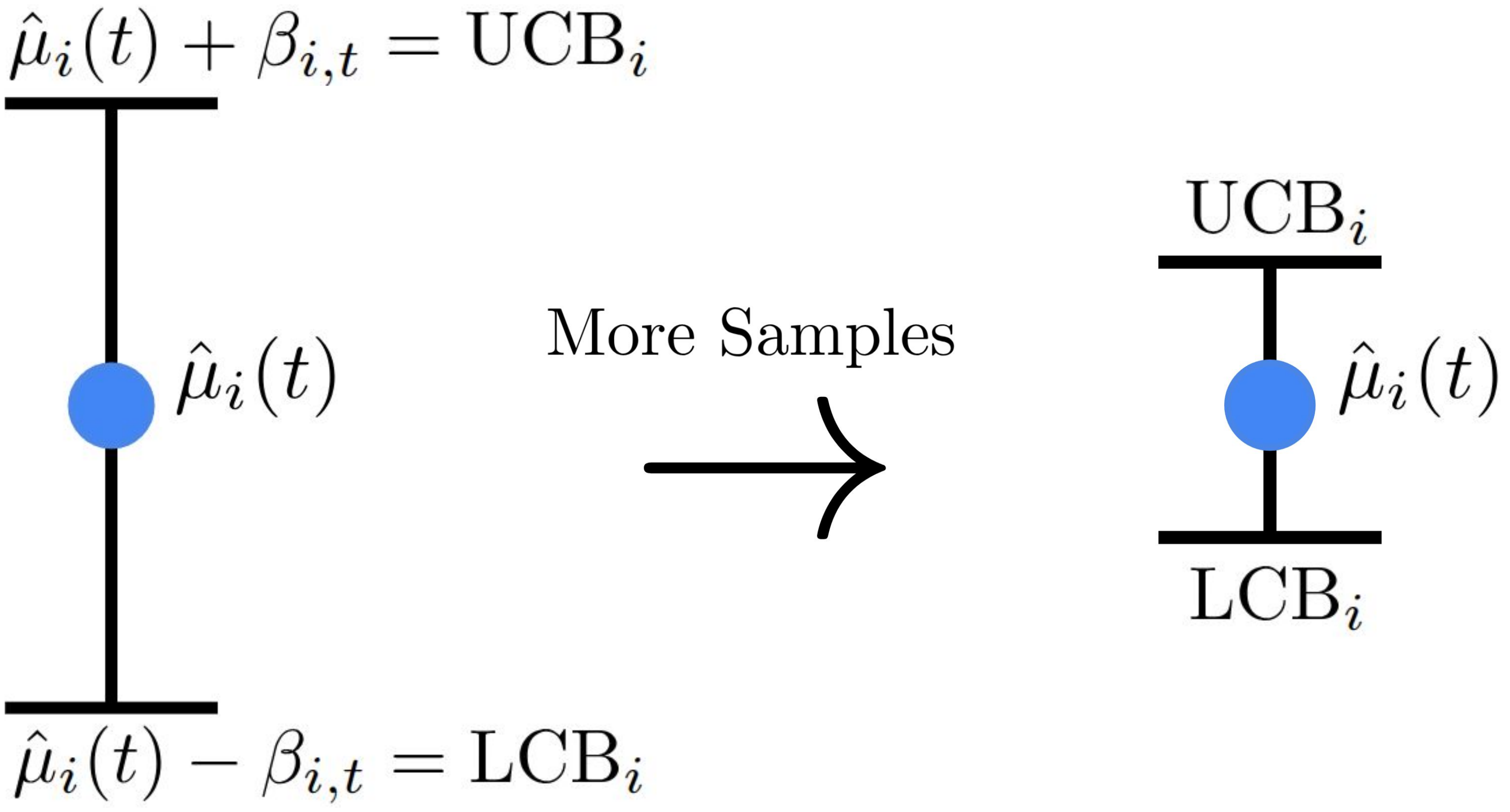


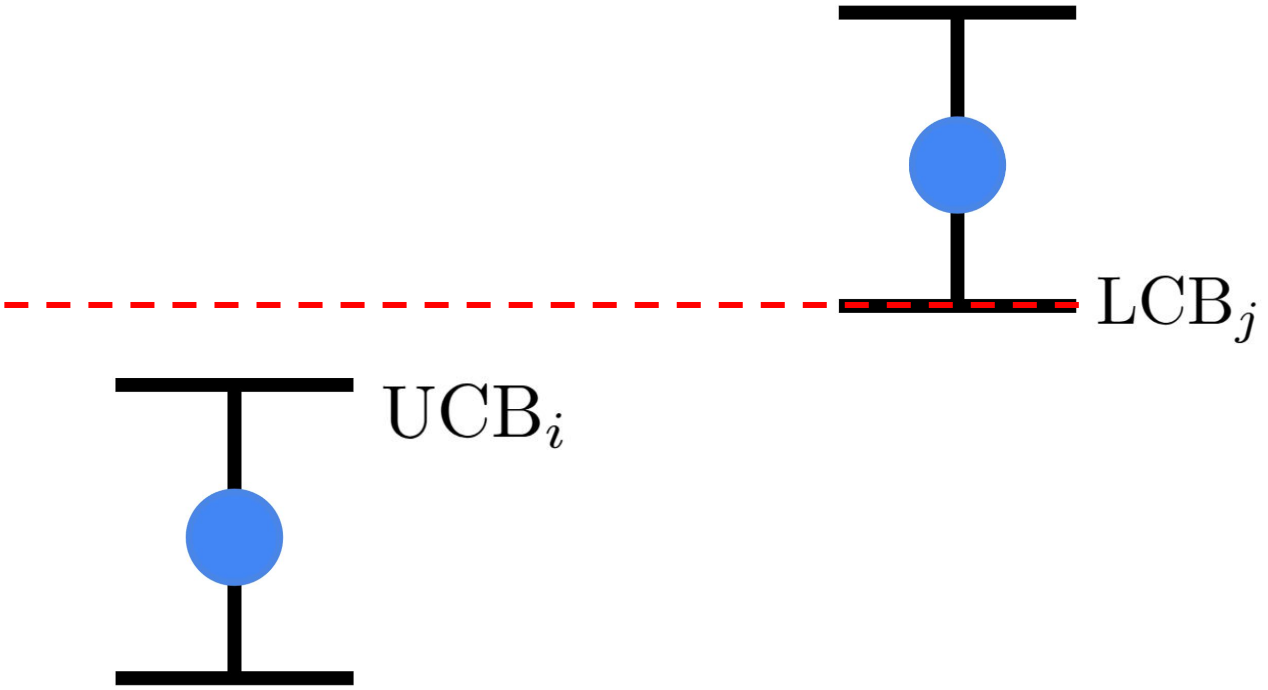
$$\hat{\mu}_i(t)$$

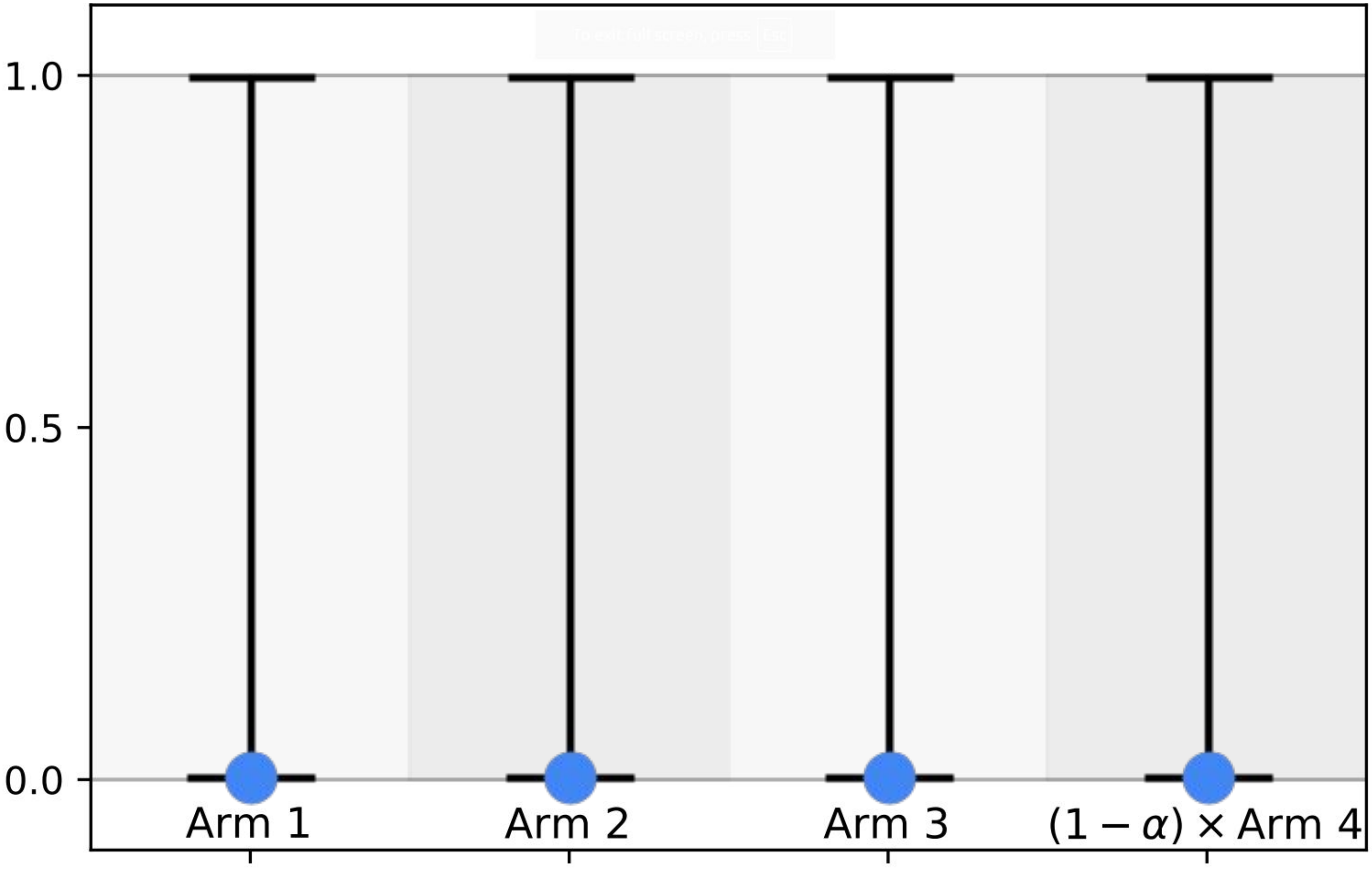
More Samples

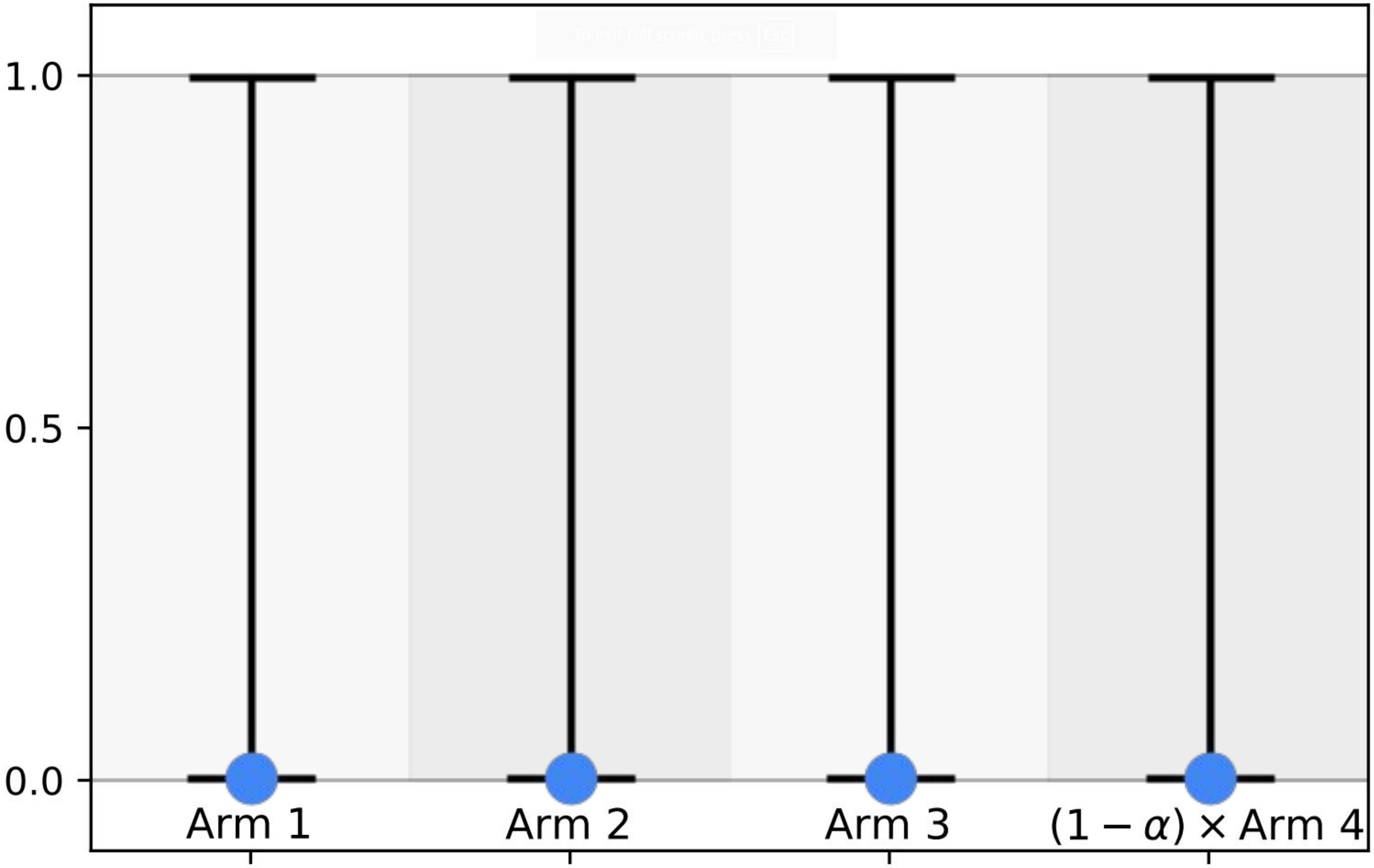
$$\longrightarrow$$

$$\hat{\mu}_i(t) - \beta_{i,t} = \mathrm{LCB}_i$$

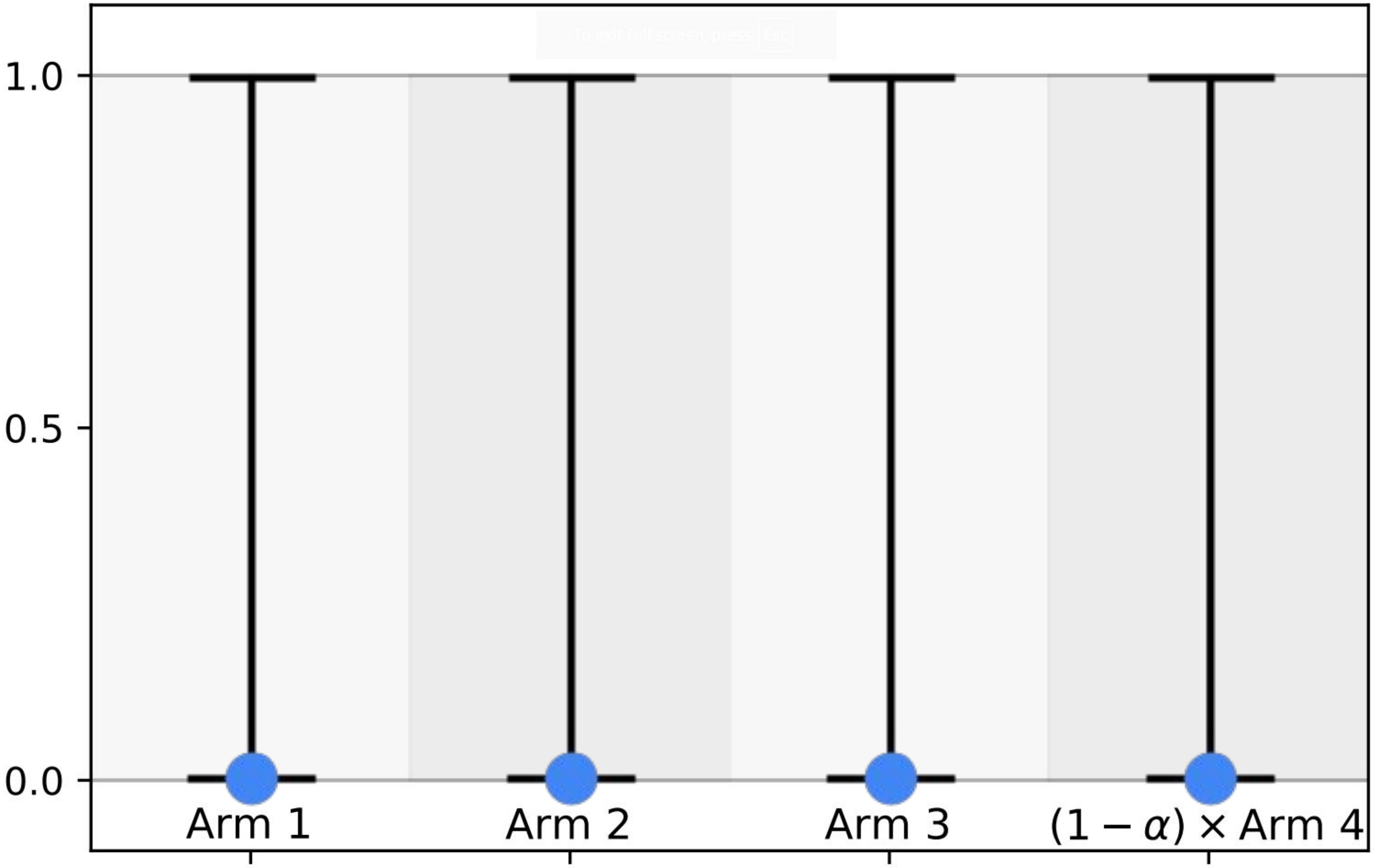

$\hat{\mu}_i(t) + \beta_{i,t} = \mathrm{UCB}_i$
$\hat{\mu}_i(t)$
$\hat{\mu}_i(t) - \beta_{i,t} = \mathrm{LCB}_i$
More Samples
UCB_i
$\hat{\mu}_i(t)$
LCB_i


LCB_j
UCB_i


1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4

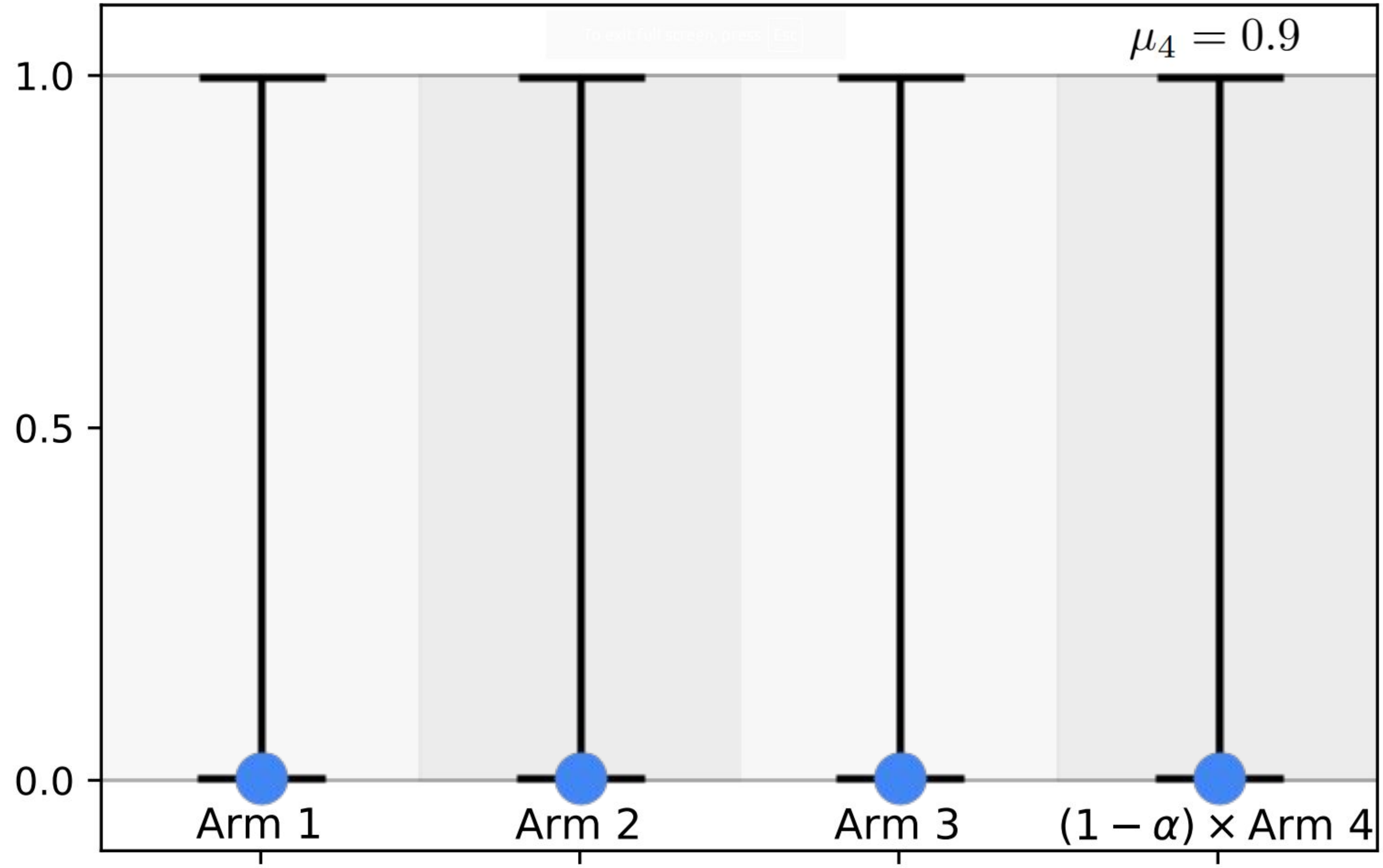


$$\hat{\mu}_i(t) = 0$$



$$\hat{\mu}_i(t) = 0$$

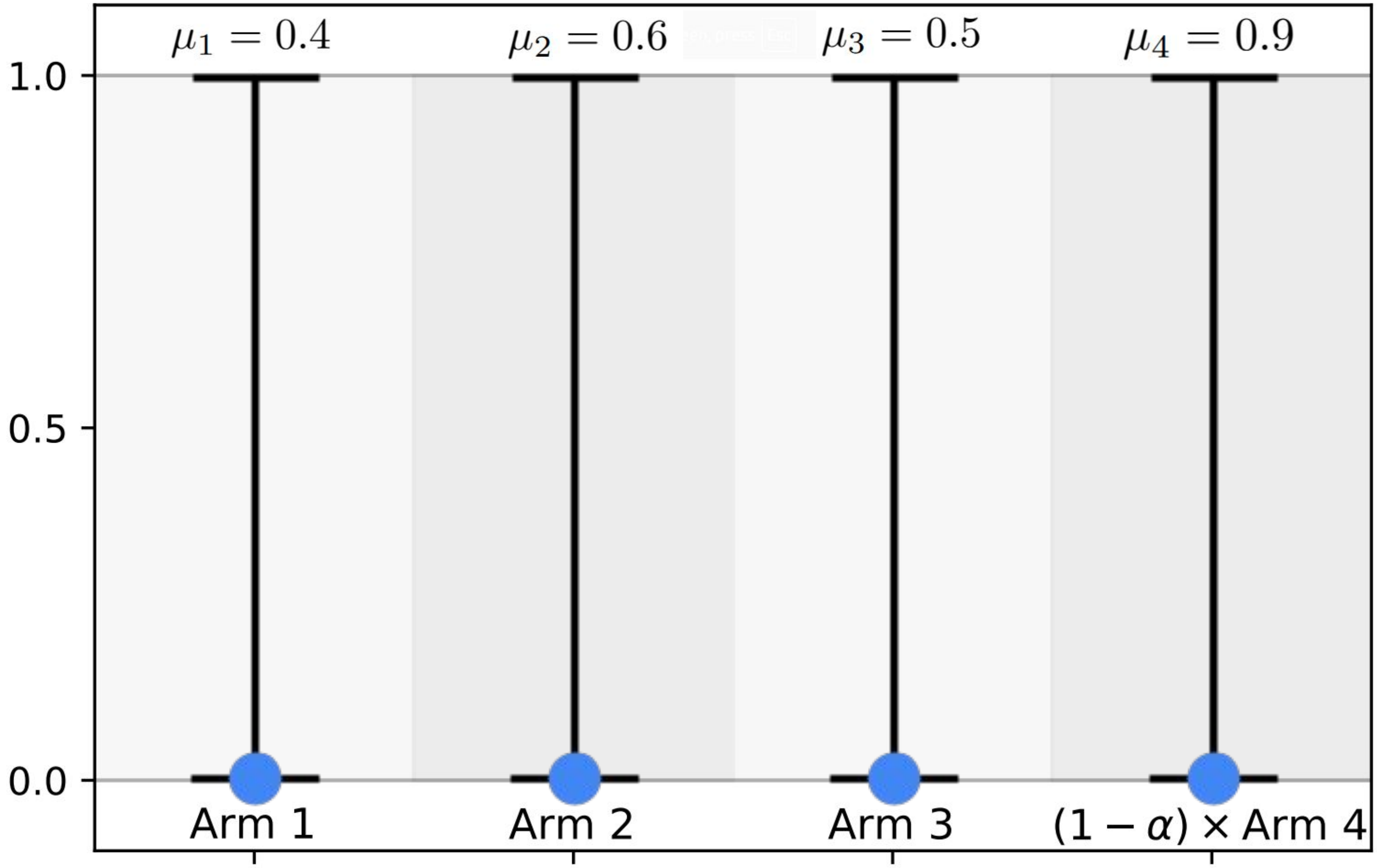
$$n_i(t) = 0$$



$$\begin{aligned} \mu_{\mathrm{CS}} &= \mathrm{thresh} \\ &= (1-\alpha)\mu_\ell \\ &= (1-0.5)0.9 \\ &= 0.45 \end{aligned}$$

$$\hat{\mu}_i(t) = 0$$

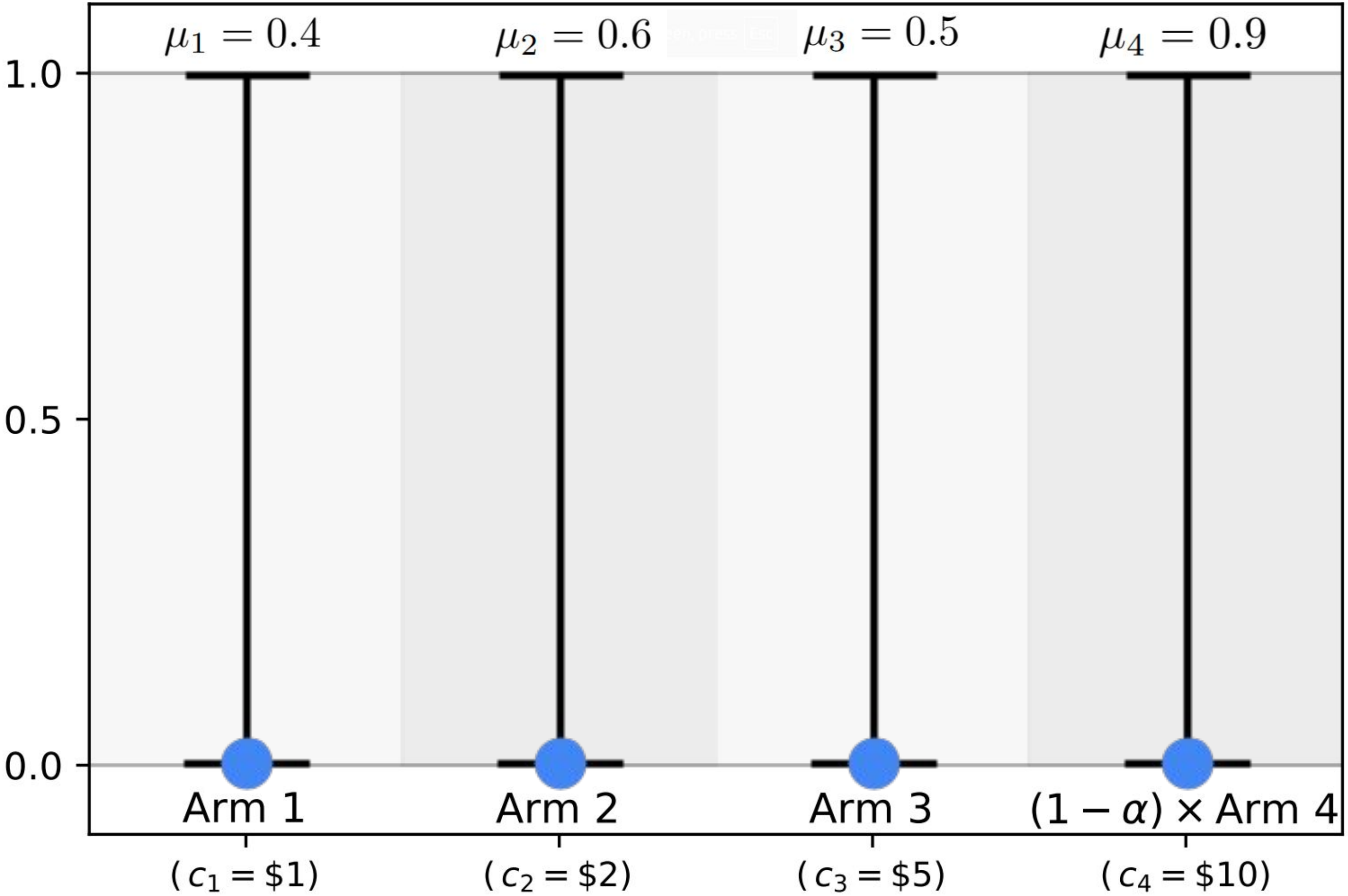
$$n_i(t) = 0$$



$$\begin{aligned}\mu_{\text{CS}} &= \text{thresh} \\ &= (1-\alpha)\mu_\ell \\ &= (1-0.5)0.9 \\ &= 0.45\end{aligned}$$

$$\hat{\mu}_i(t) = 0$$

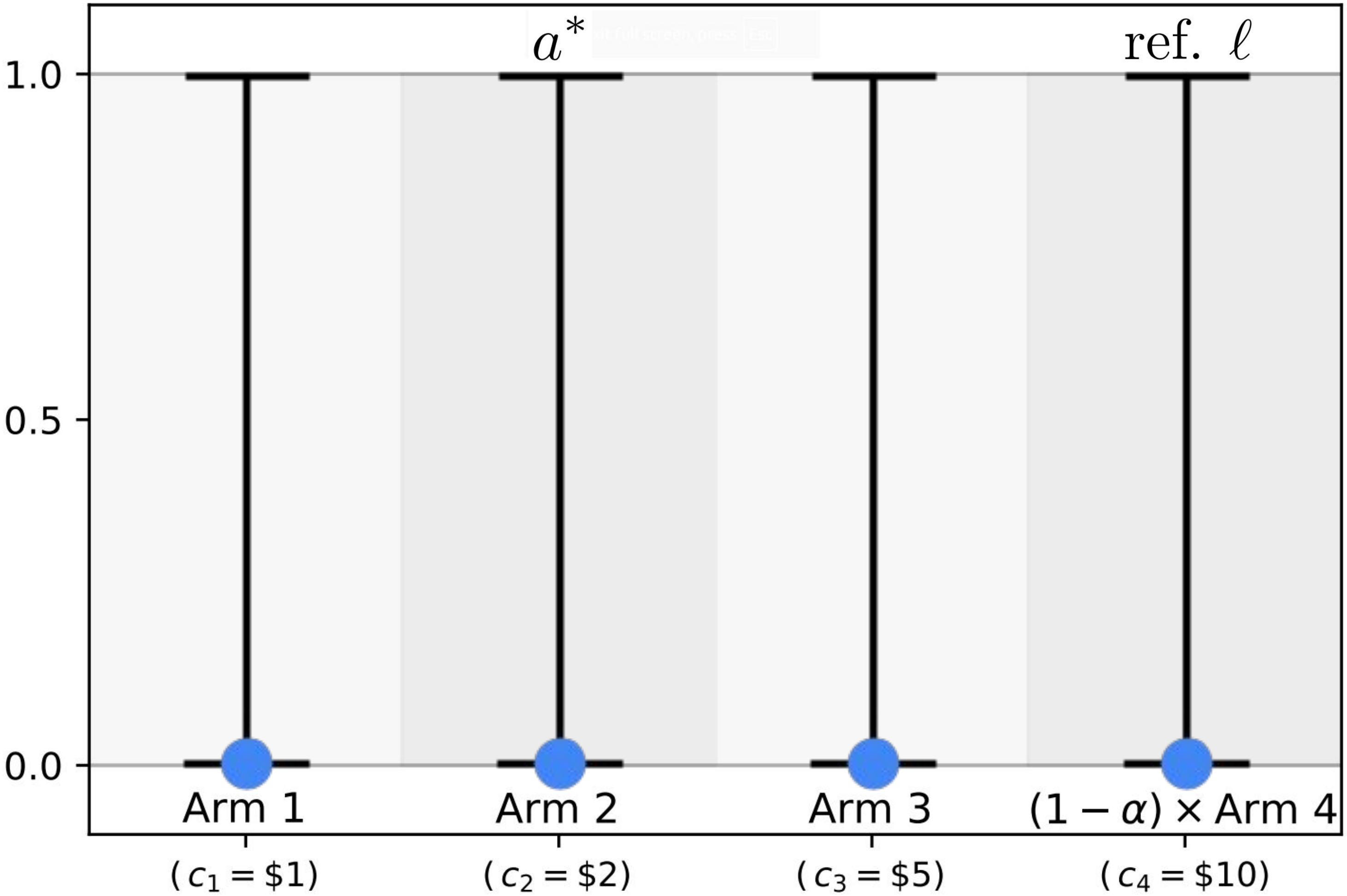
$$n_i(t) = 0$$



$$\begin{aligned}\mu_{\text{CS}} &= \text{thresh} \\ &= (1-\alpha)\mu_\ell \\ &= (1-0.5)0.9 \\ &= 0.45\end{aligned}$$

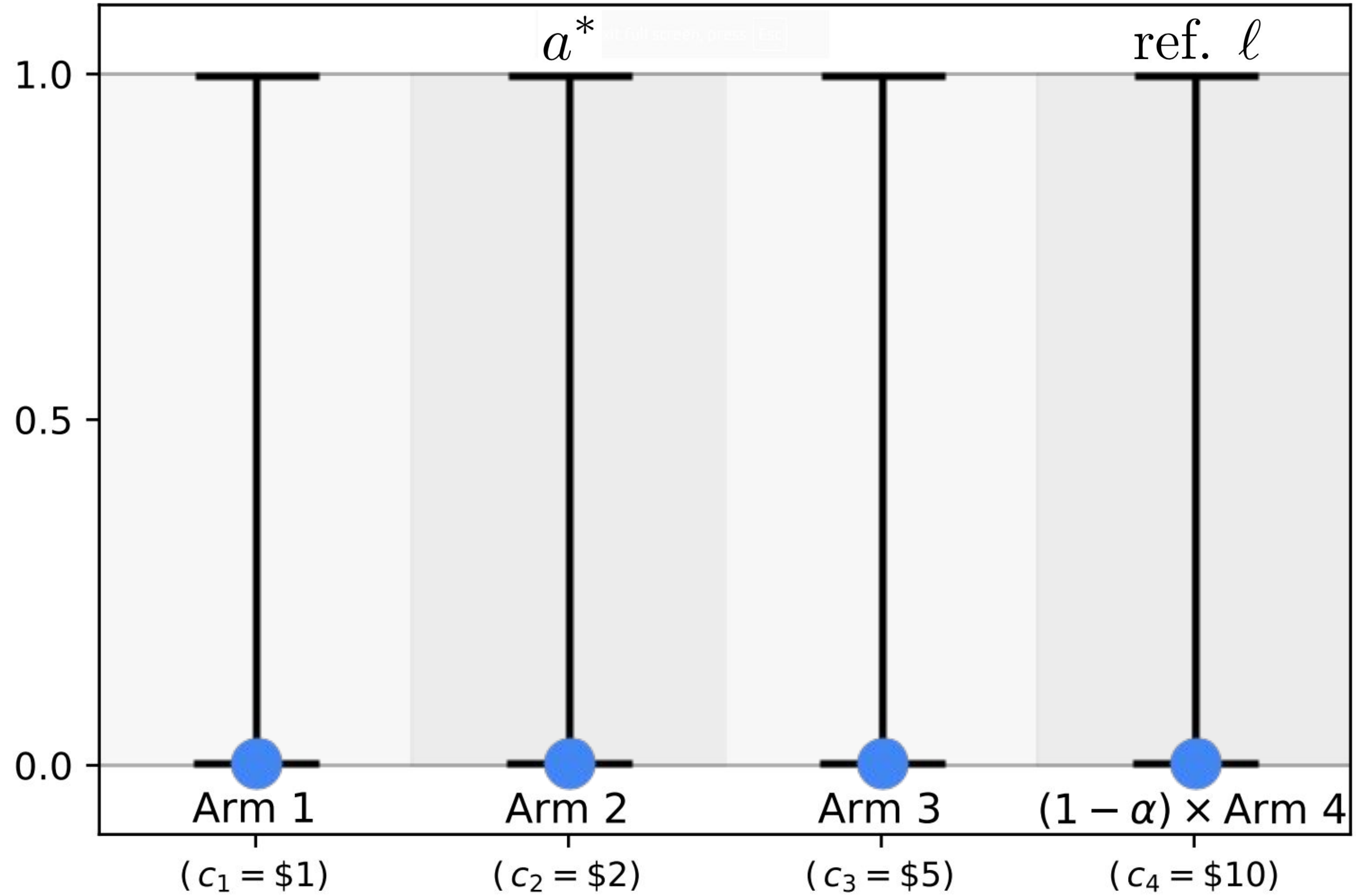
$$\hat{\mu}_i(t) = 0$$

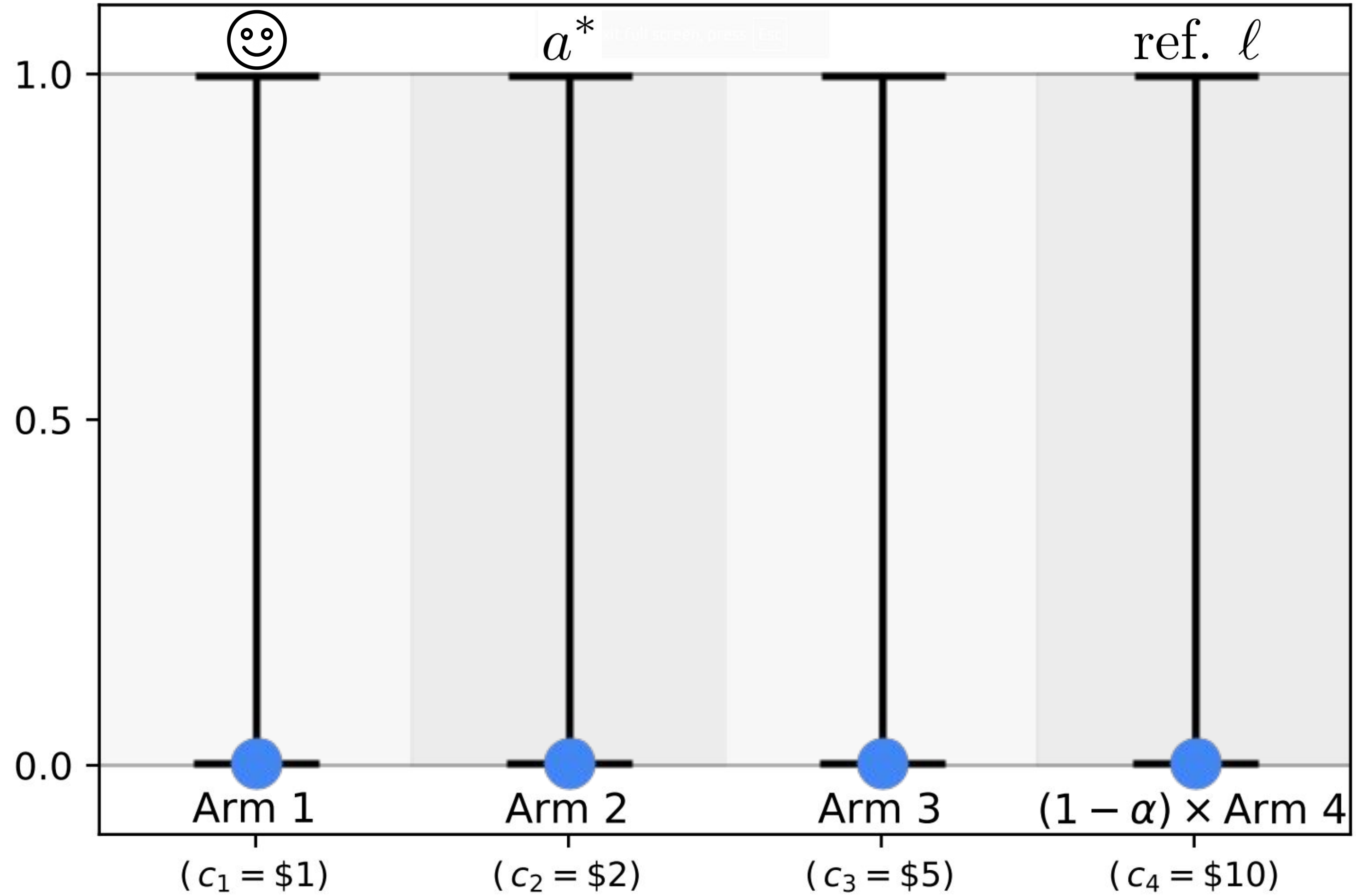
$$n_i(t) = 0$$



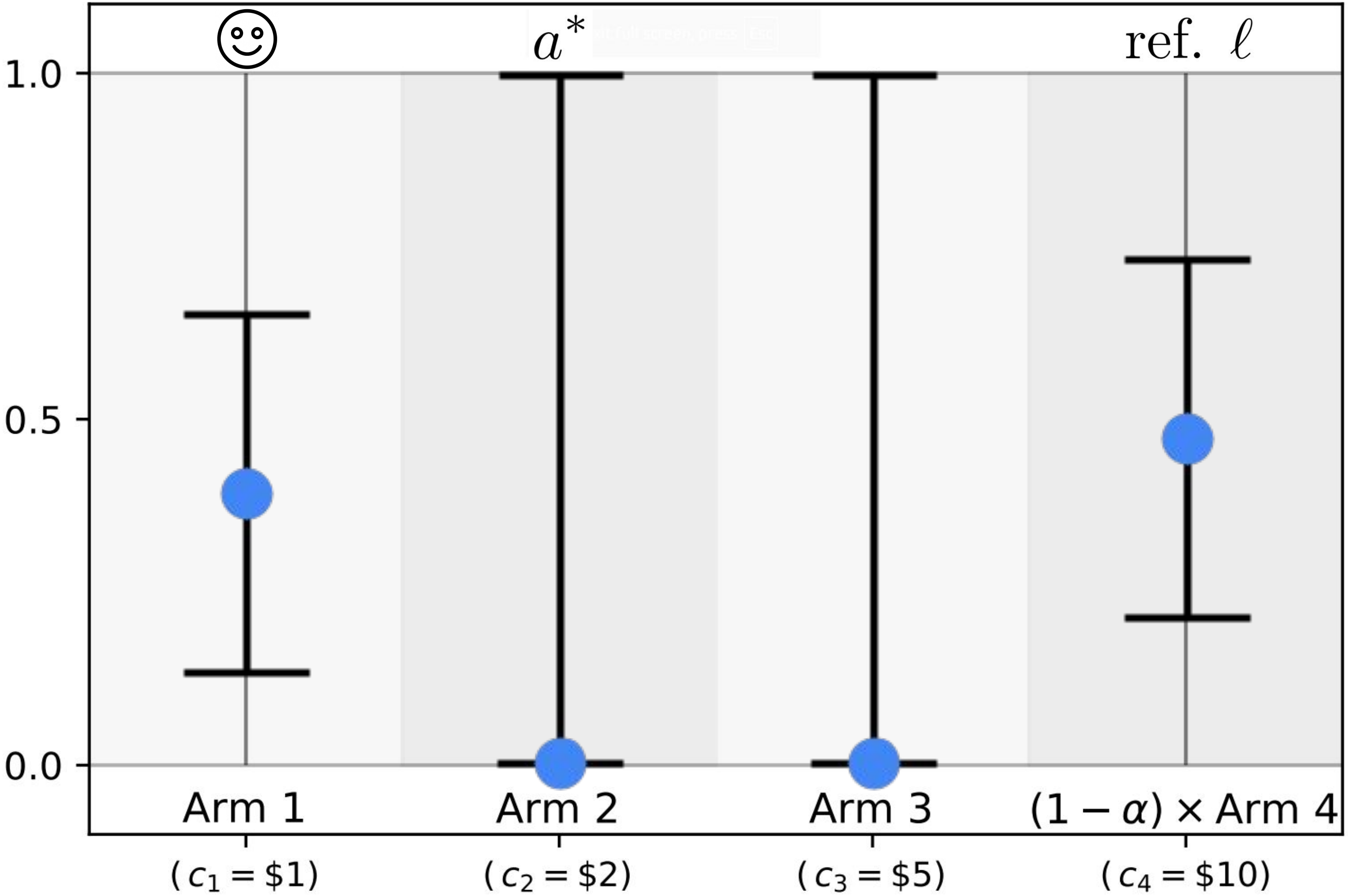
$$\hat{\mu}_i(t) = 0$$

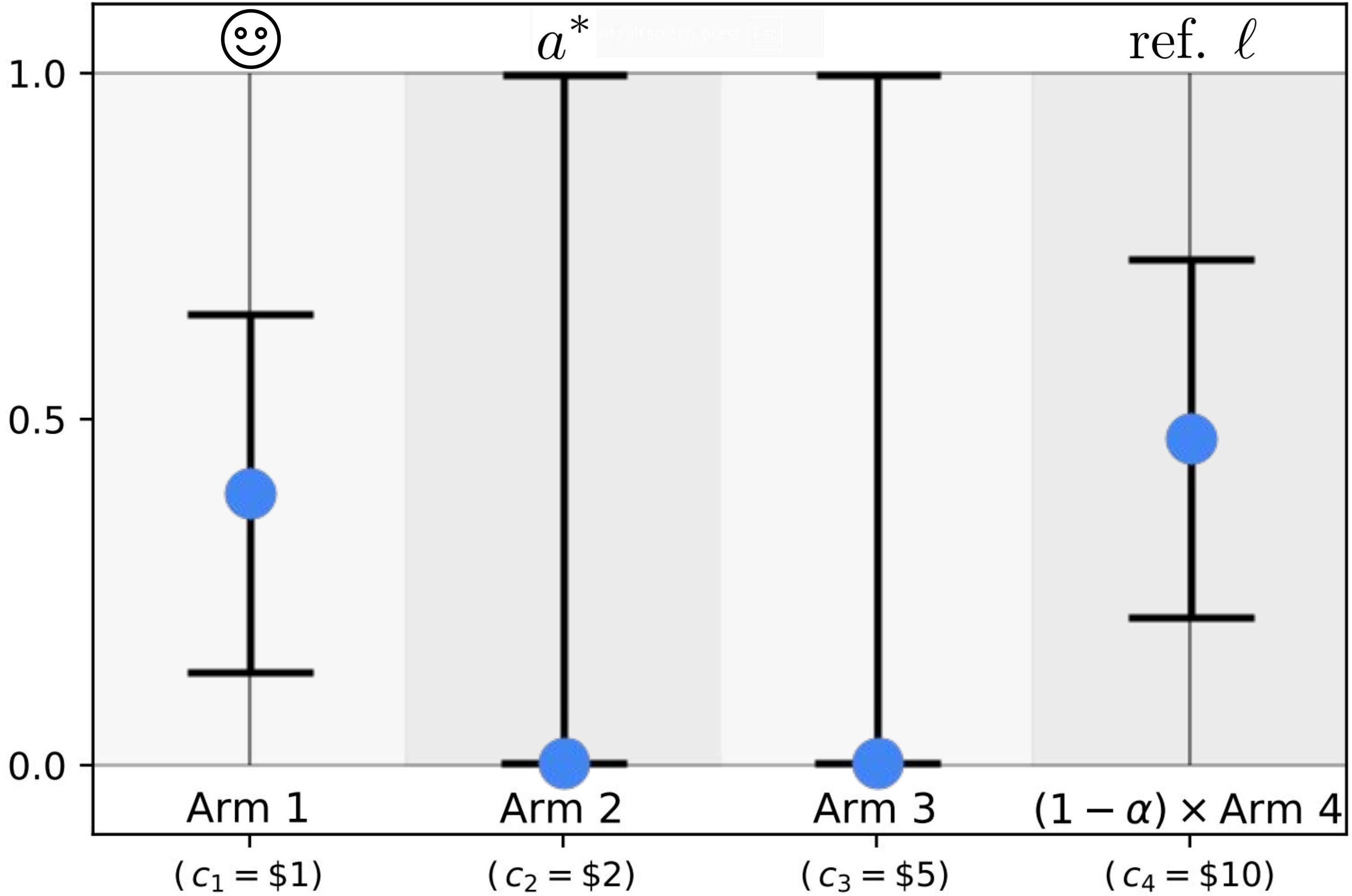
$$n_i(t) = 0$$


a^*
ref. ℓ
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$

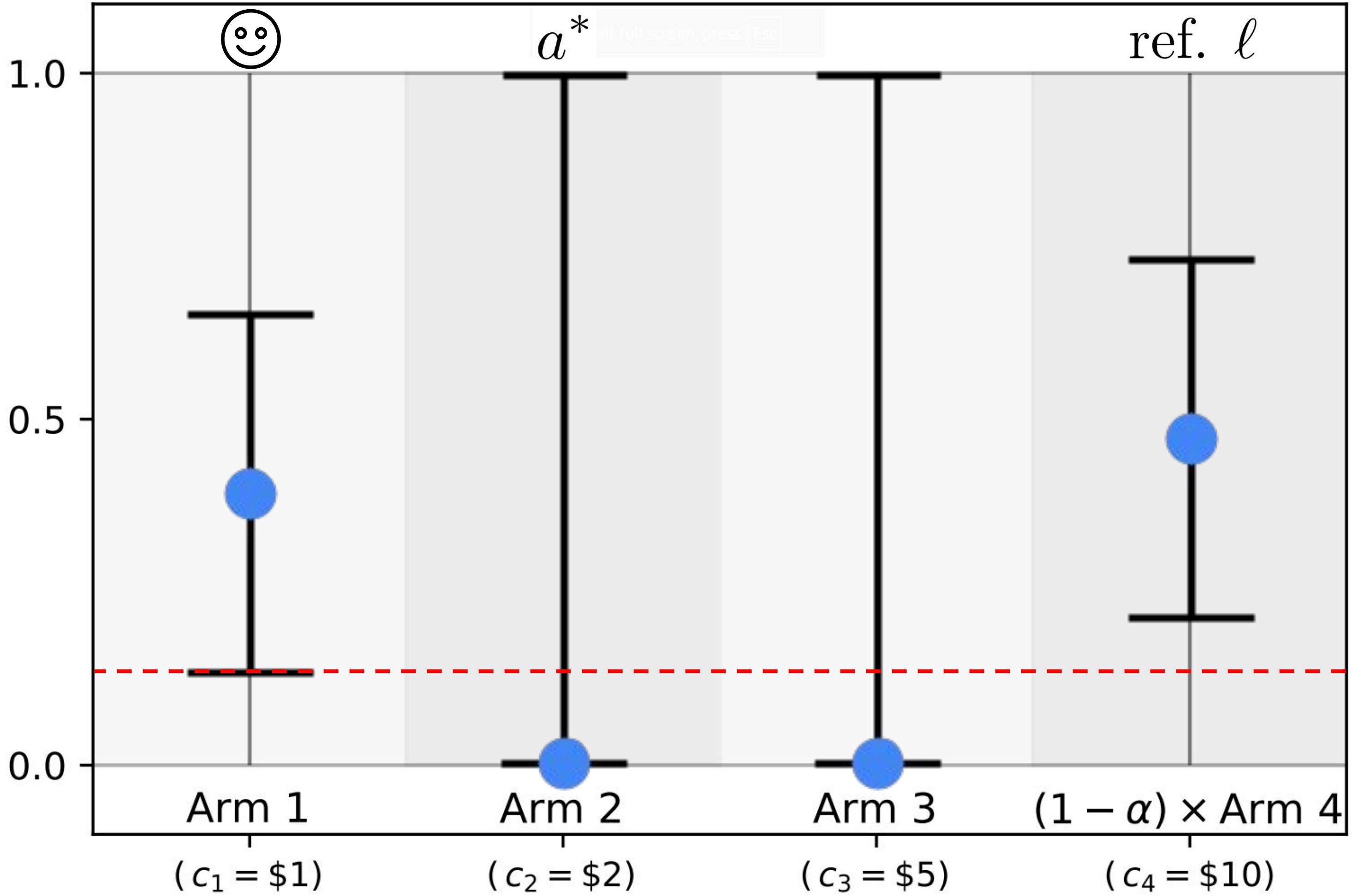


Sample arms $i = 1$ and ℓ to τ_1

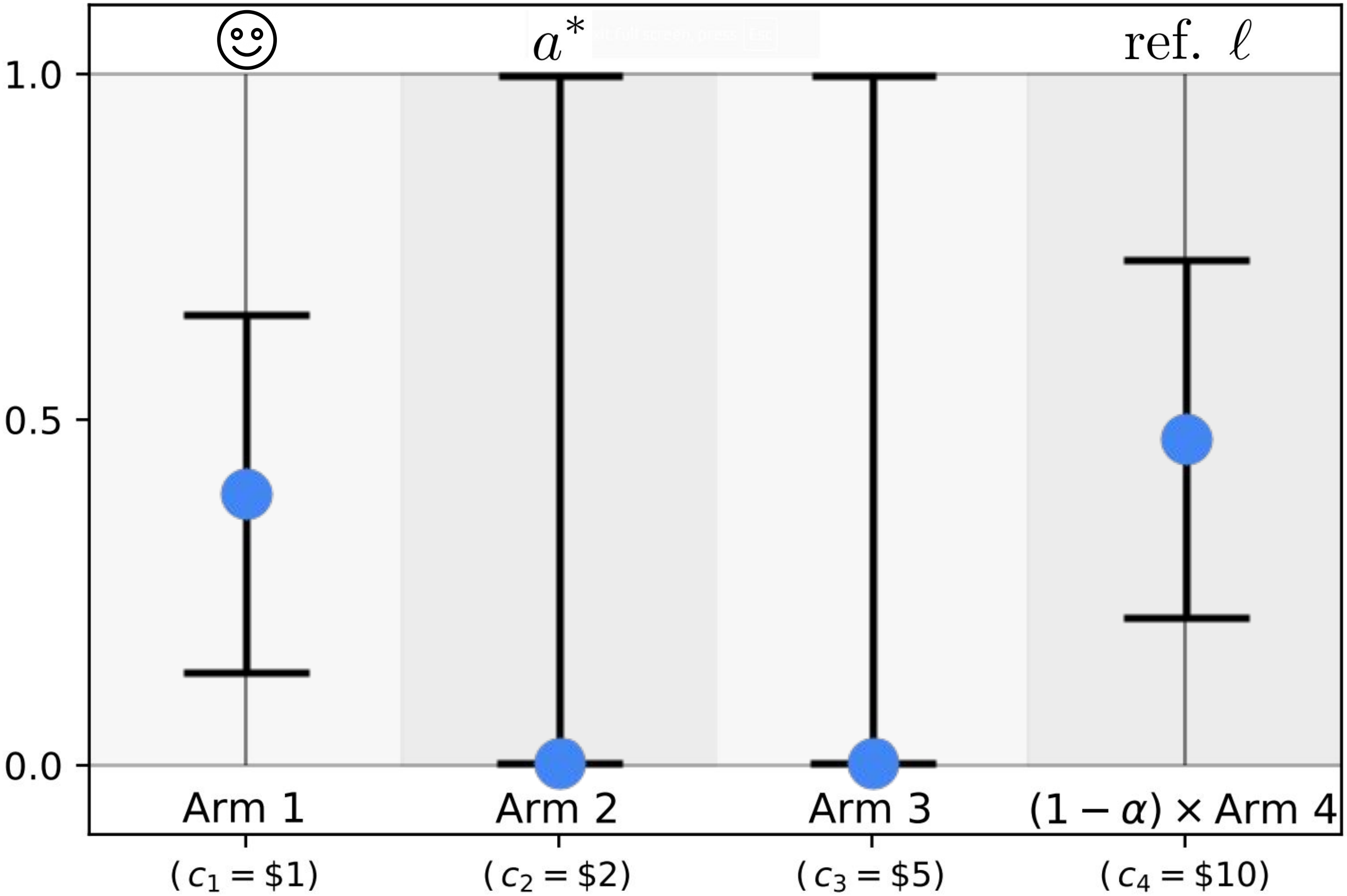

☺
a^*
ref. ℓ
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$



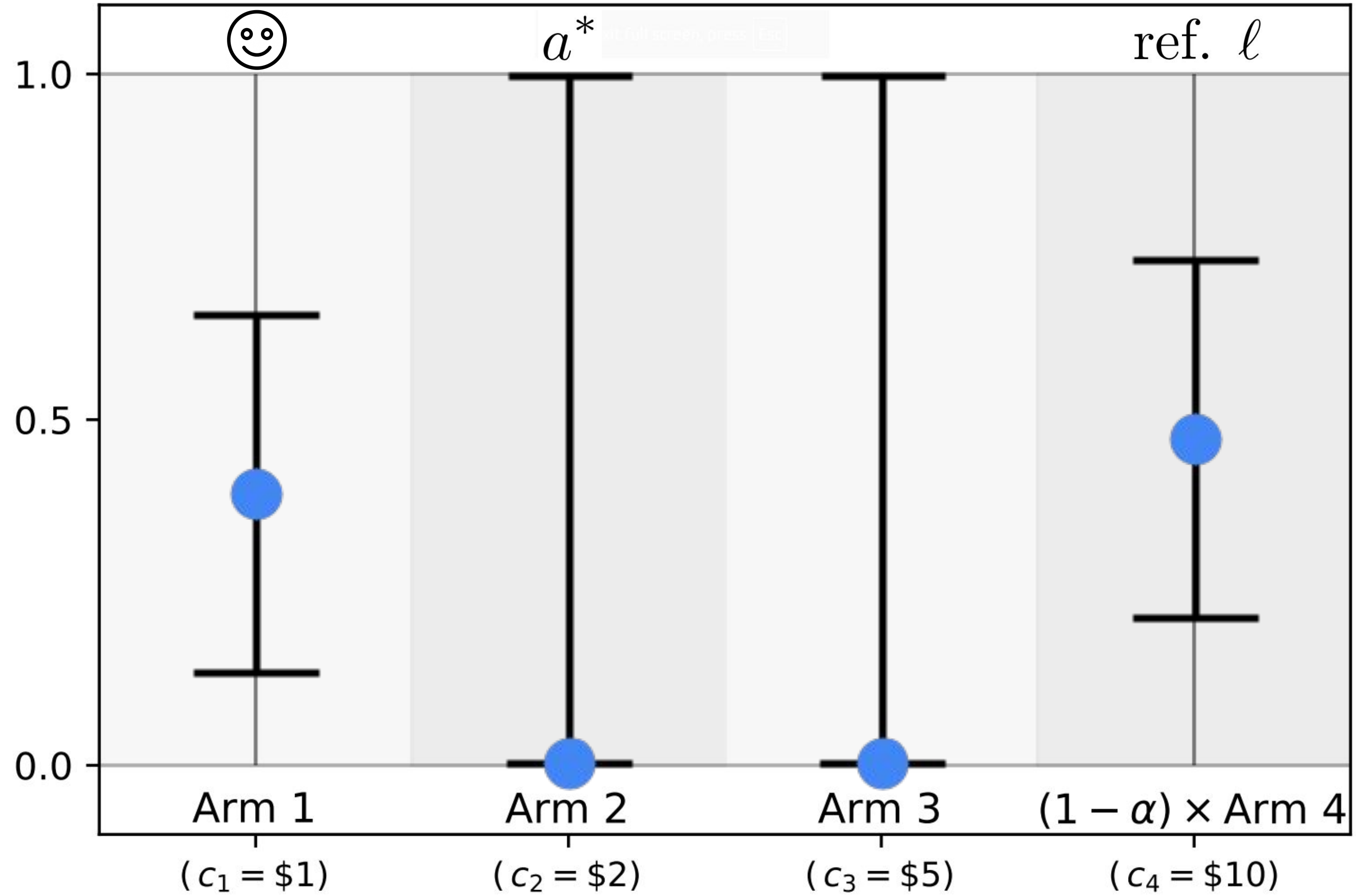
Stopping?
$\mathrm{LCB}_i \geq (1-\alpha)\mathrm{UCB}_\ell$



Stopping?
$\mathrm{LCB}_i \geq (1-\alpha)\mathrm{UCB}_\ell$

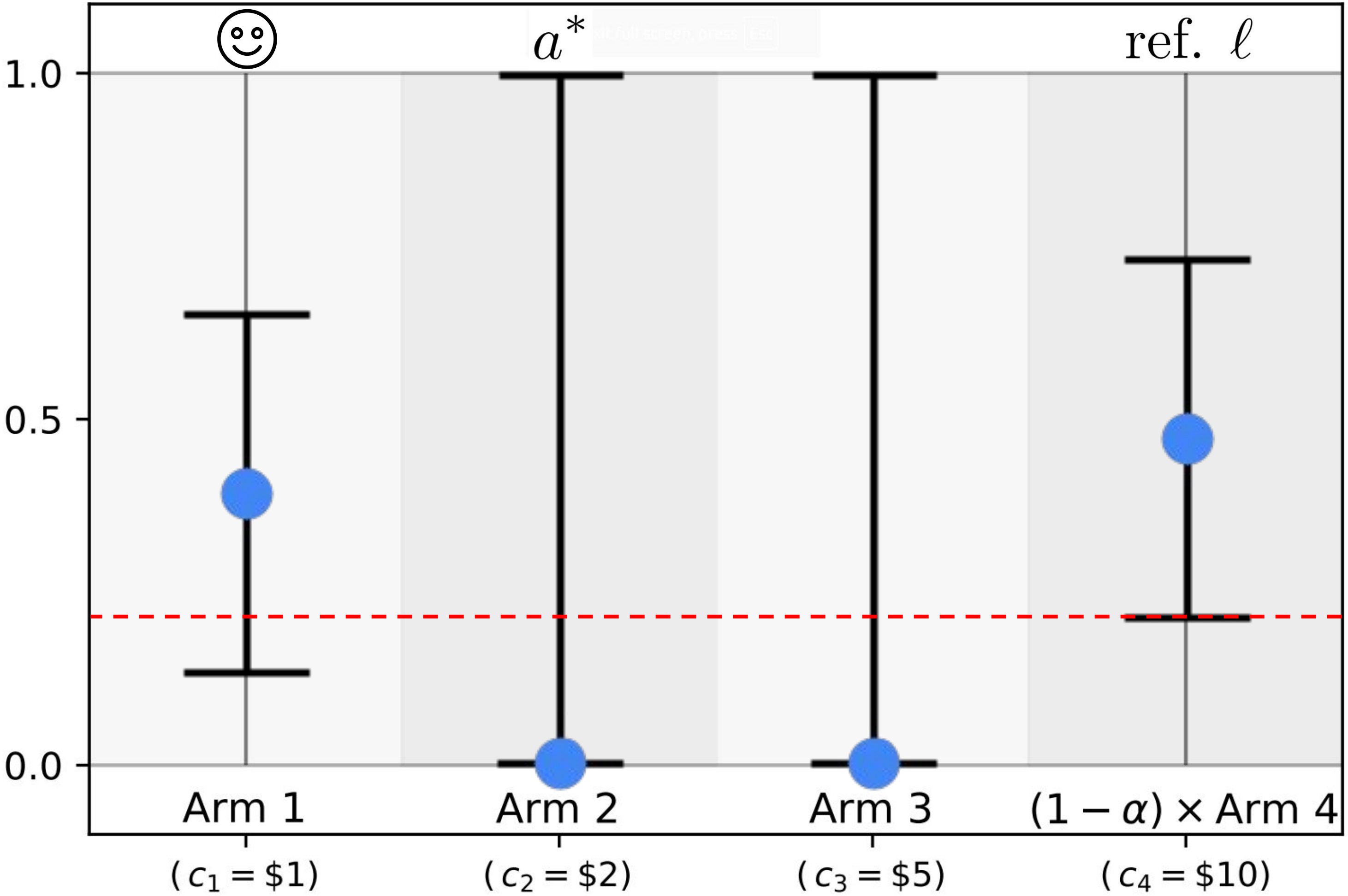


☺
a^*
ref. ℓ
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$



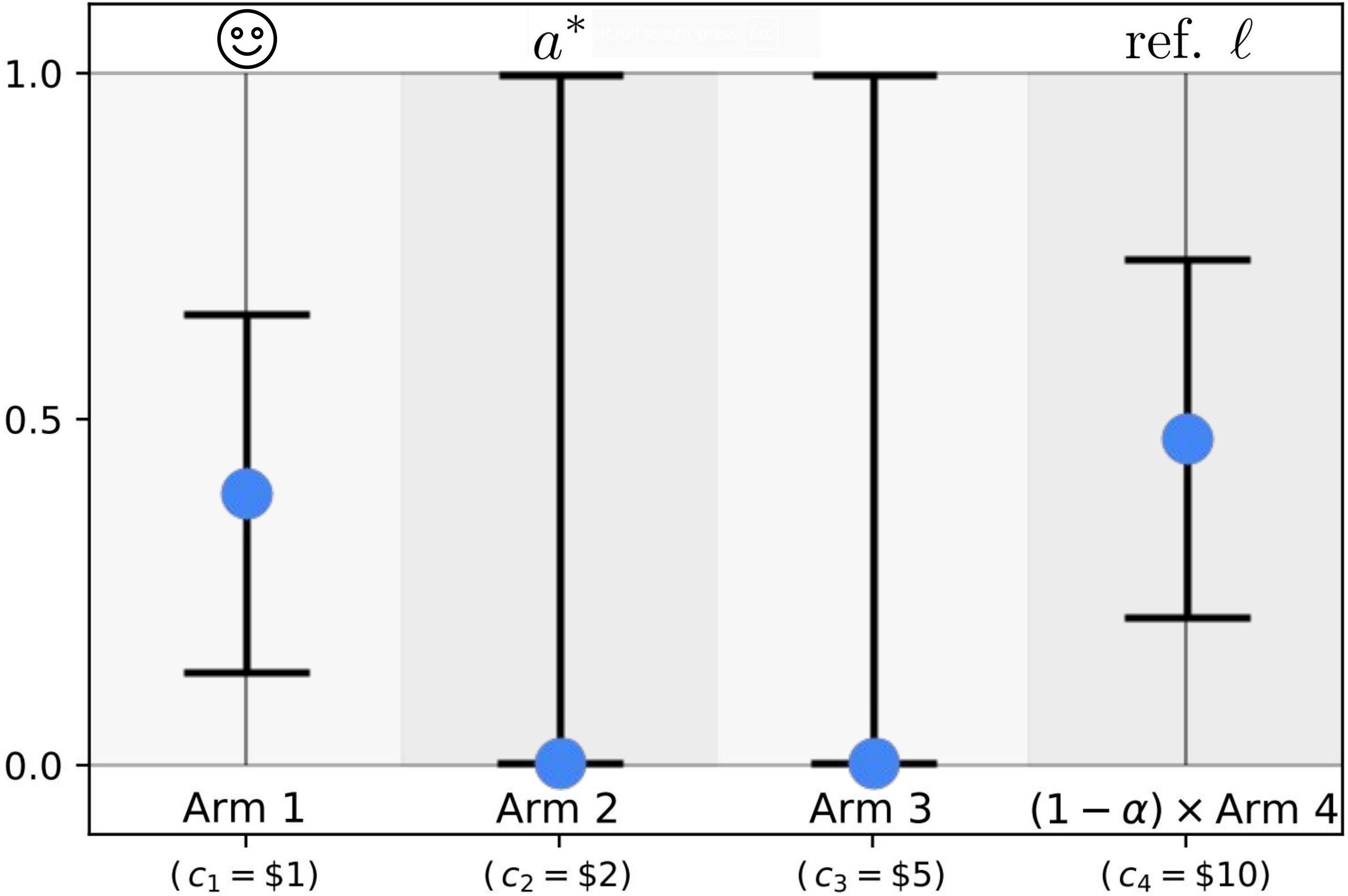
Elimination?
$(1-\alpha) \times \mathrm{LCB}_\ell \geq \mathrm{UCB}_i$

TODO: Drop the use of the times symbol here

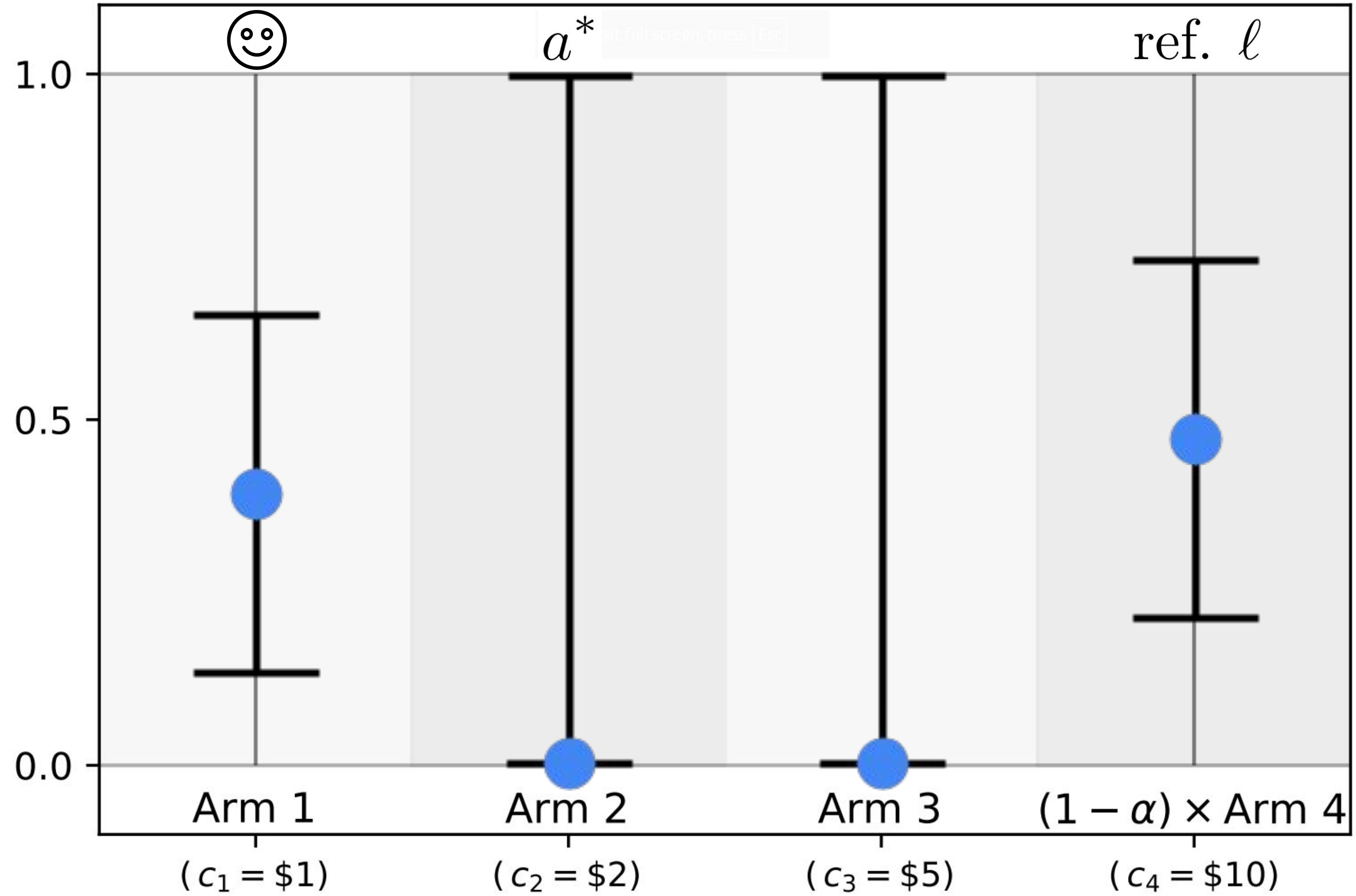


Elimination?

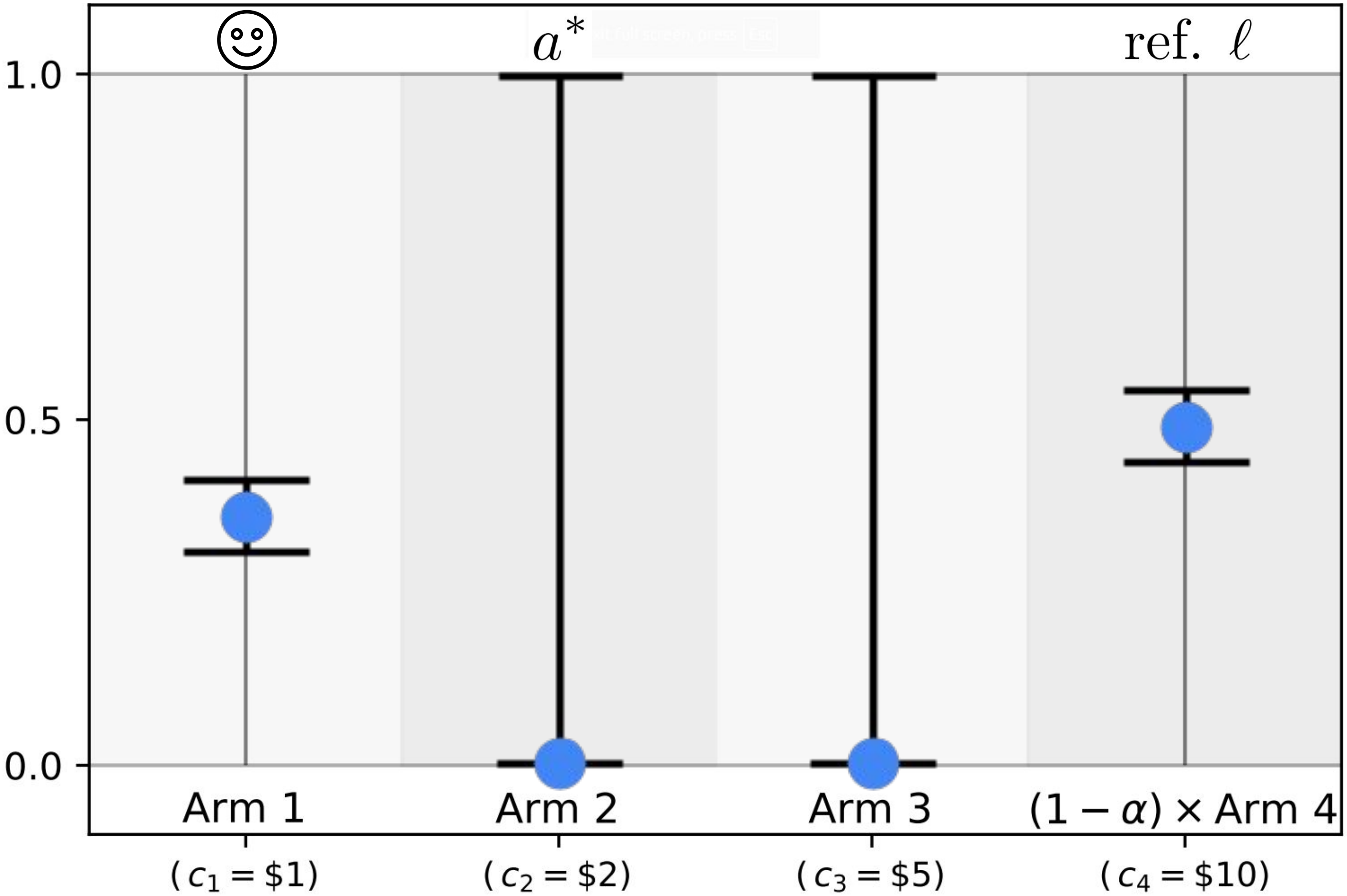
$(1-\alpha) \times \mathrm{LCB}_\ell \geq \mathrm{UCB}_i$

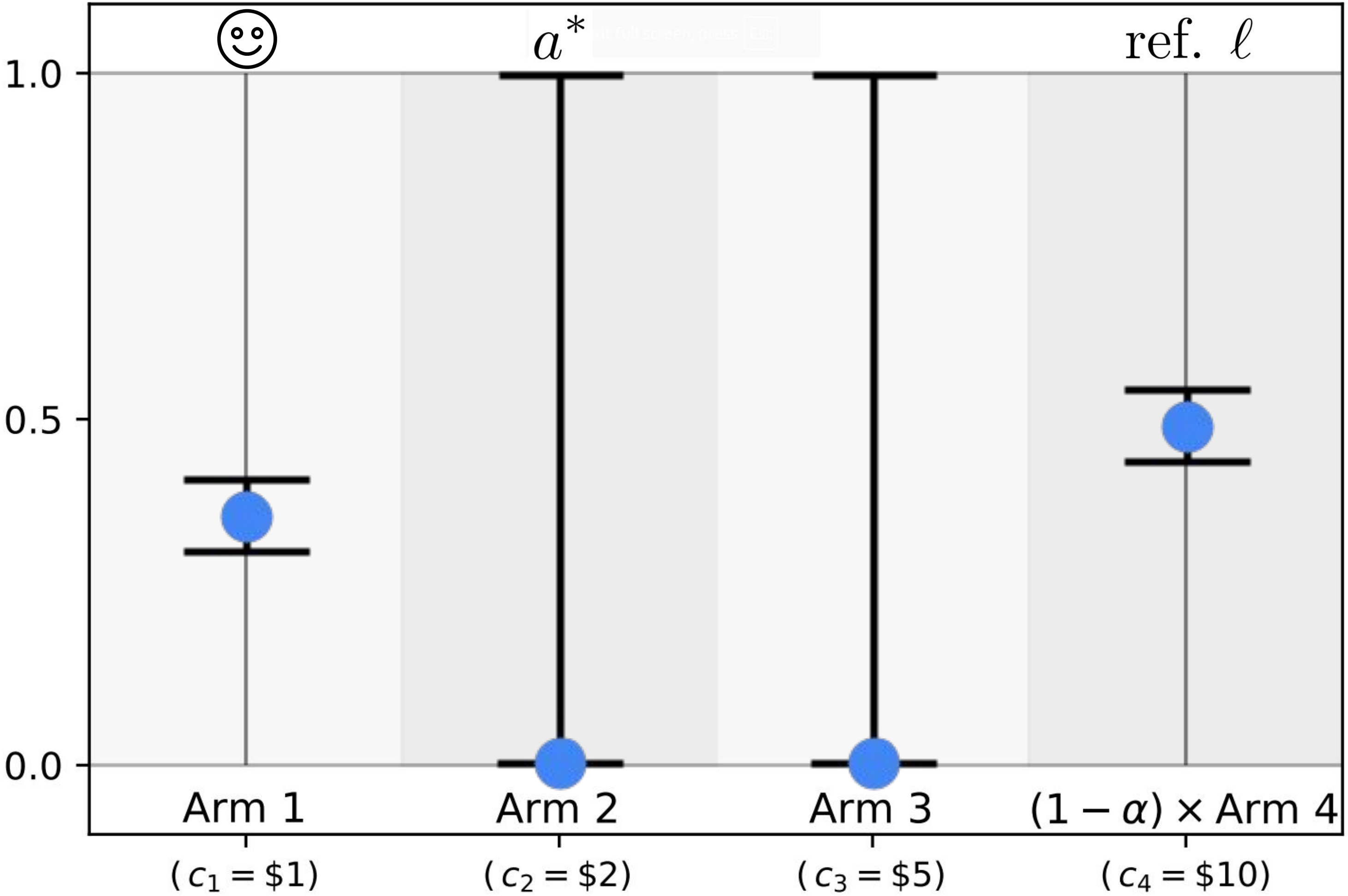


☺
a^*
ref. ℓ
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$

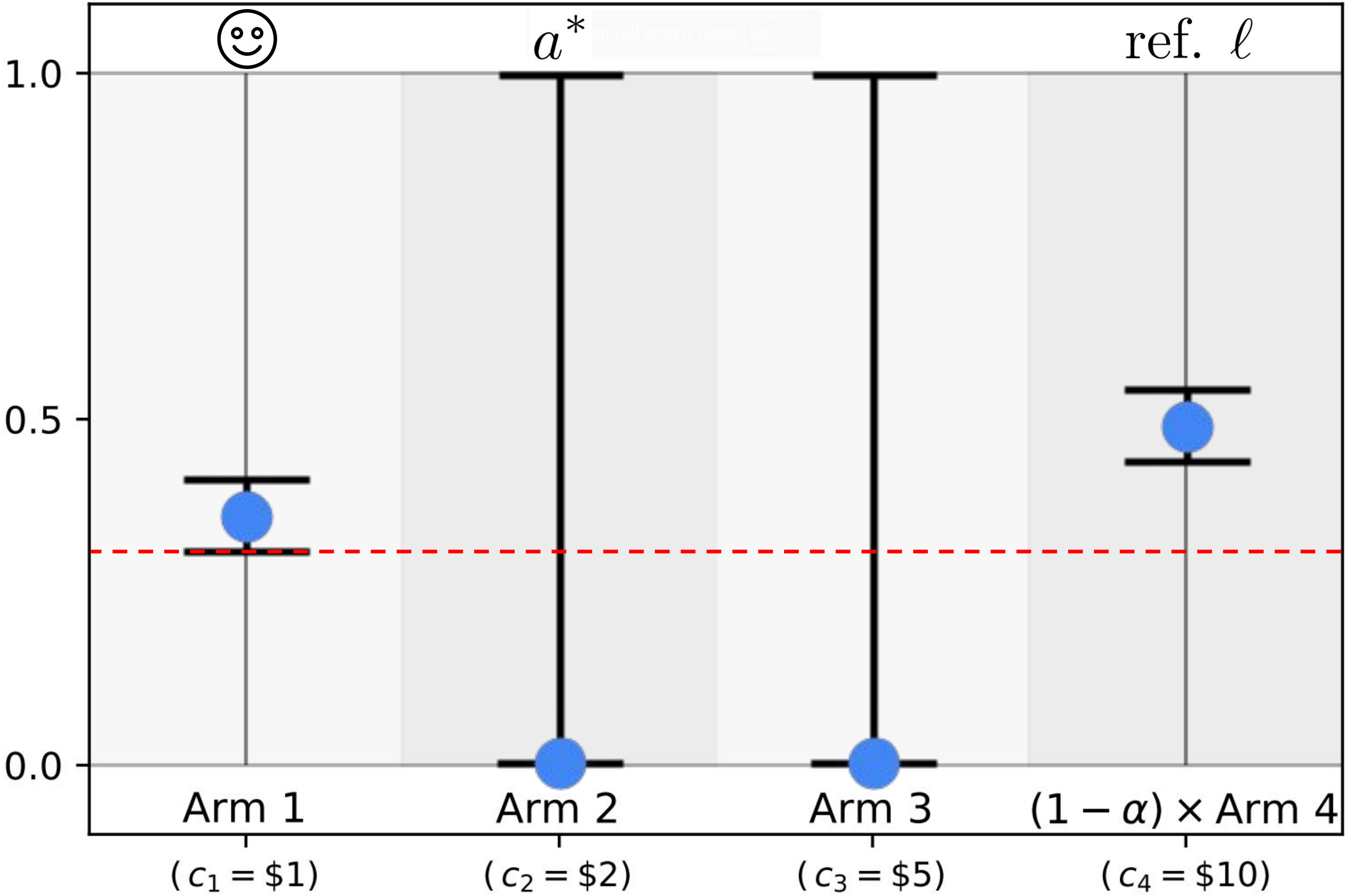


Sample arms $i = 1$ and ℓ to τ_2

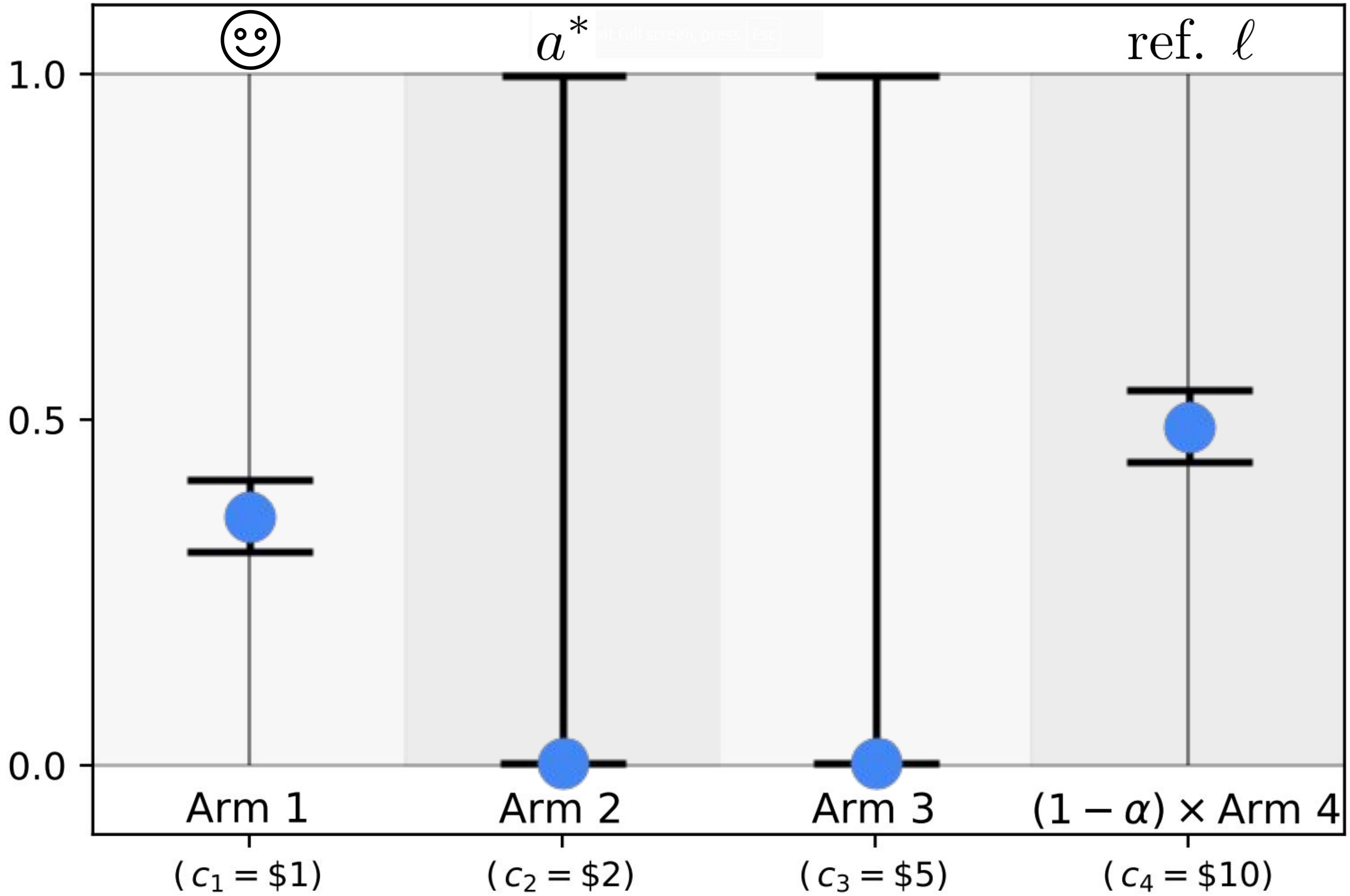

☺
a^*
ref. ℓ
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$

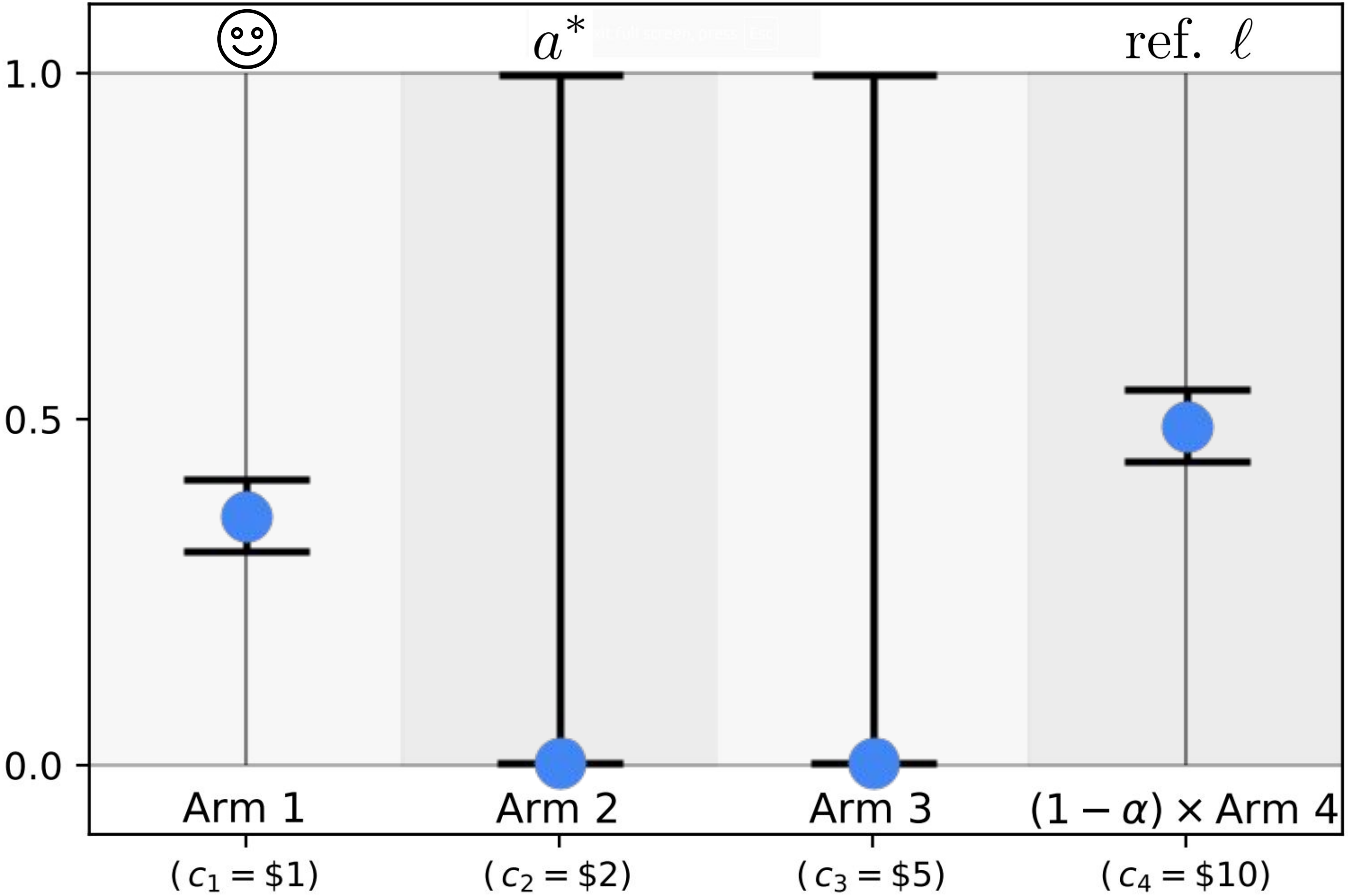


Stopping?
$\mathrm{LCB}_i \geq (1-\alpha)\mathrm{UCB}_\ell$

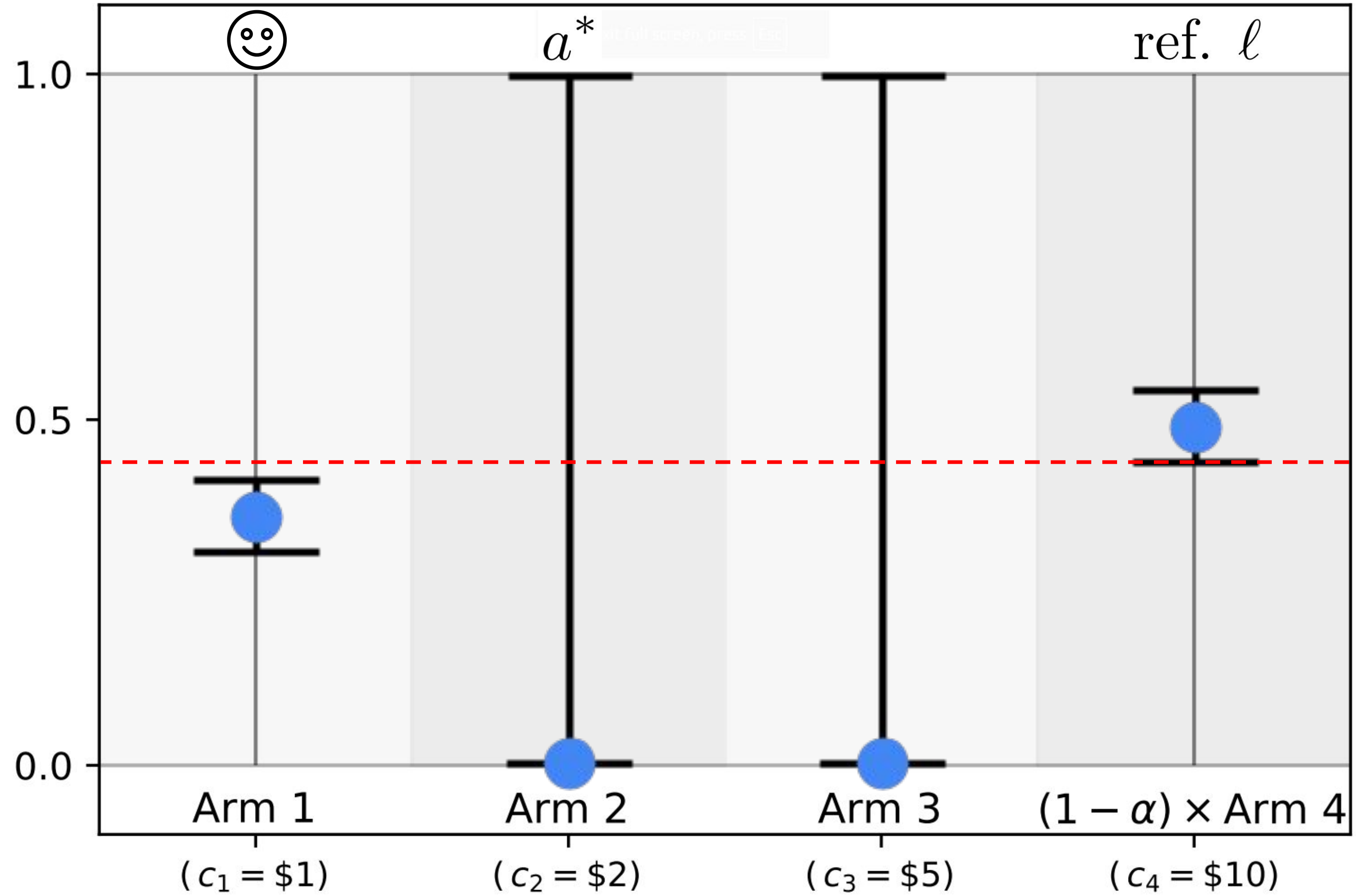


Stopping?
$\mathrm{LCB}_i \geq (1-\alpha)\mathrm{UCB}_\ell$

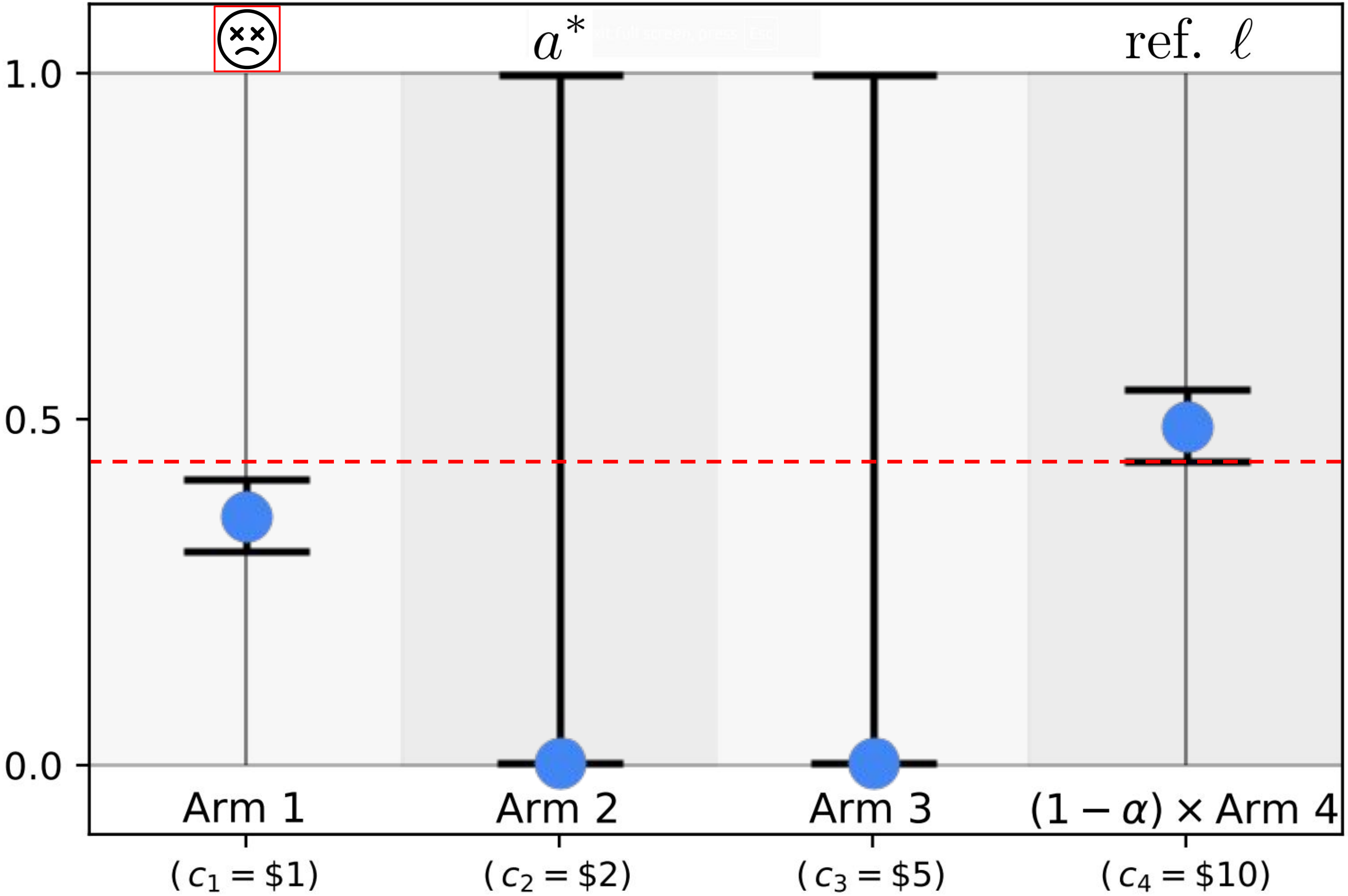

☺
a^*
ref. ℓ
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$



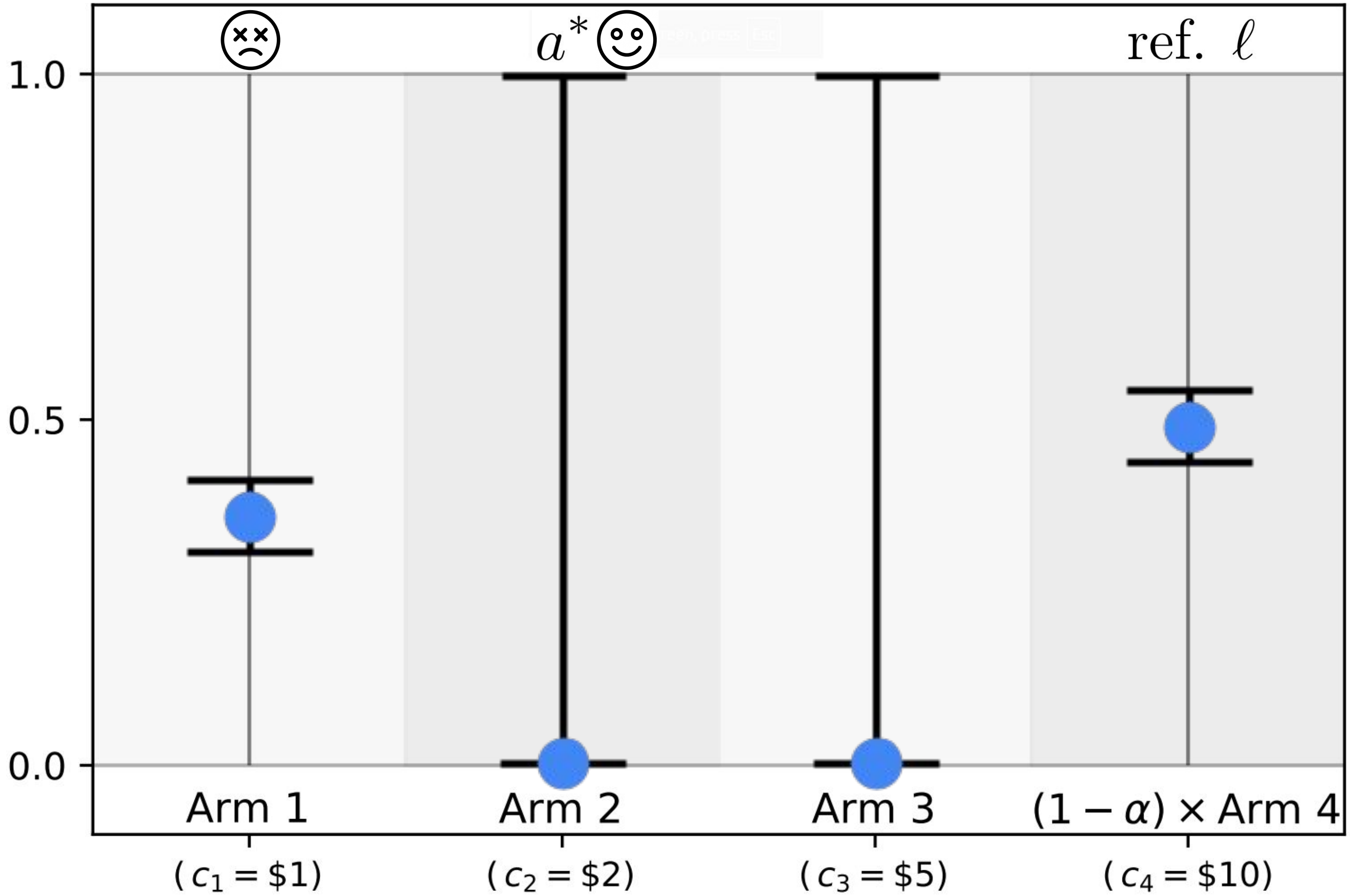
Elimination?
$(1-\alpha) \times \mathrm{LCB}_\ell \geq \mathrm{UCB}_i$

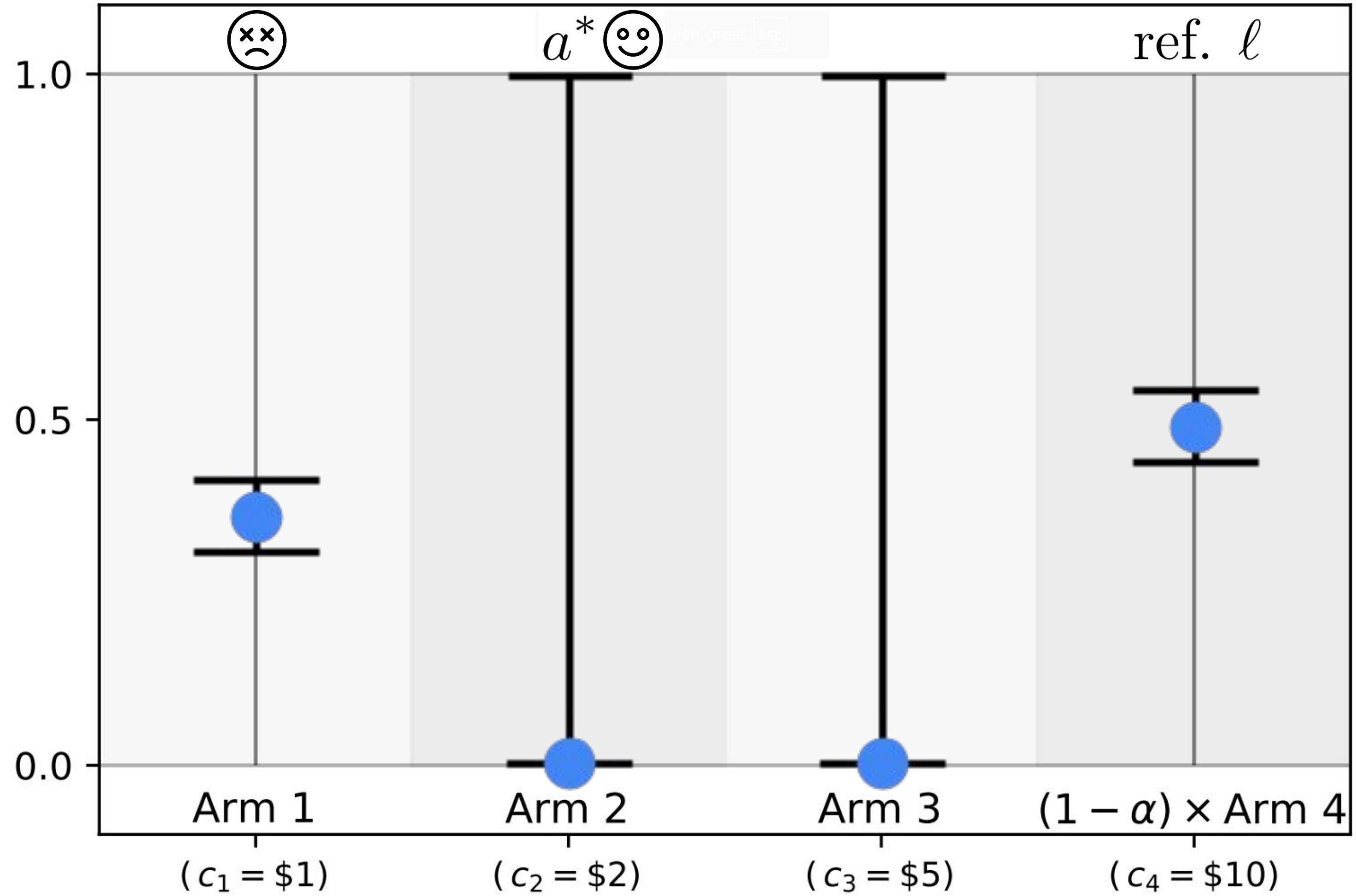


Elimination?
$(1-\alpha) \times \mathrm{LCB}_\ell \geq \mathrm{UCB}_i$

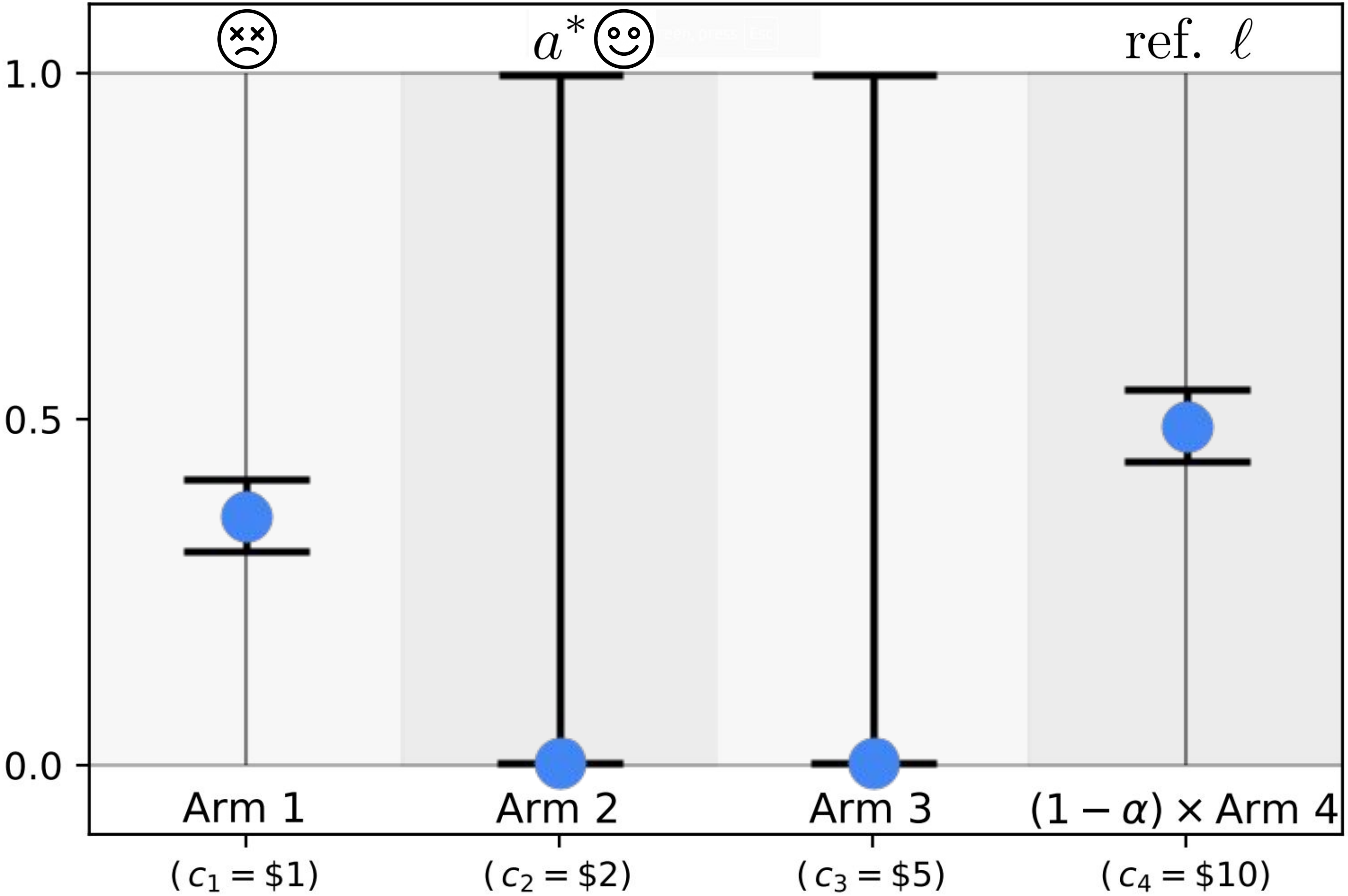


a^*
ref. ℓ
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$

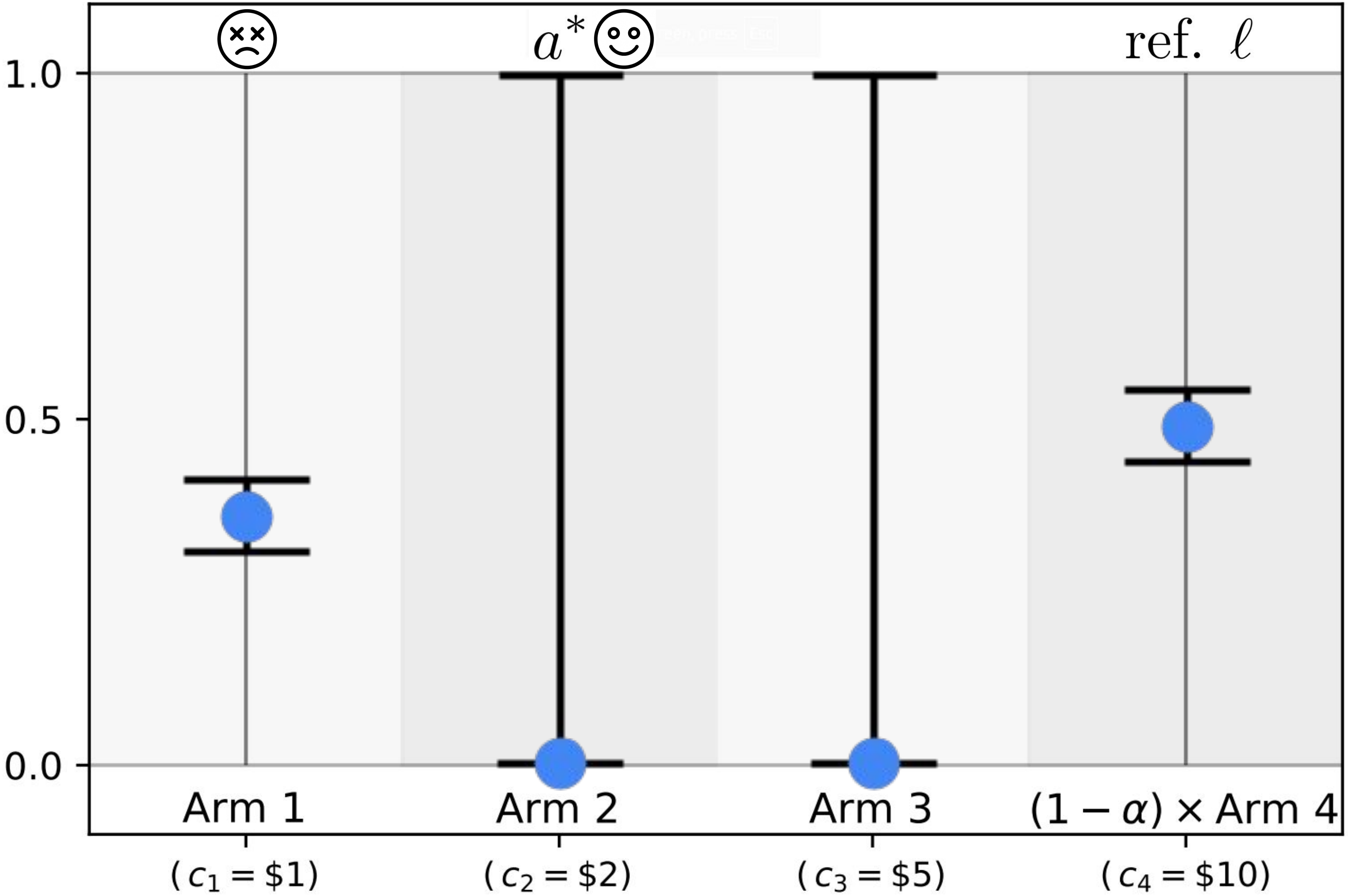

a^*
ref. ℓ
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$



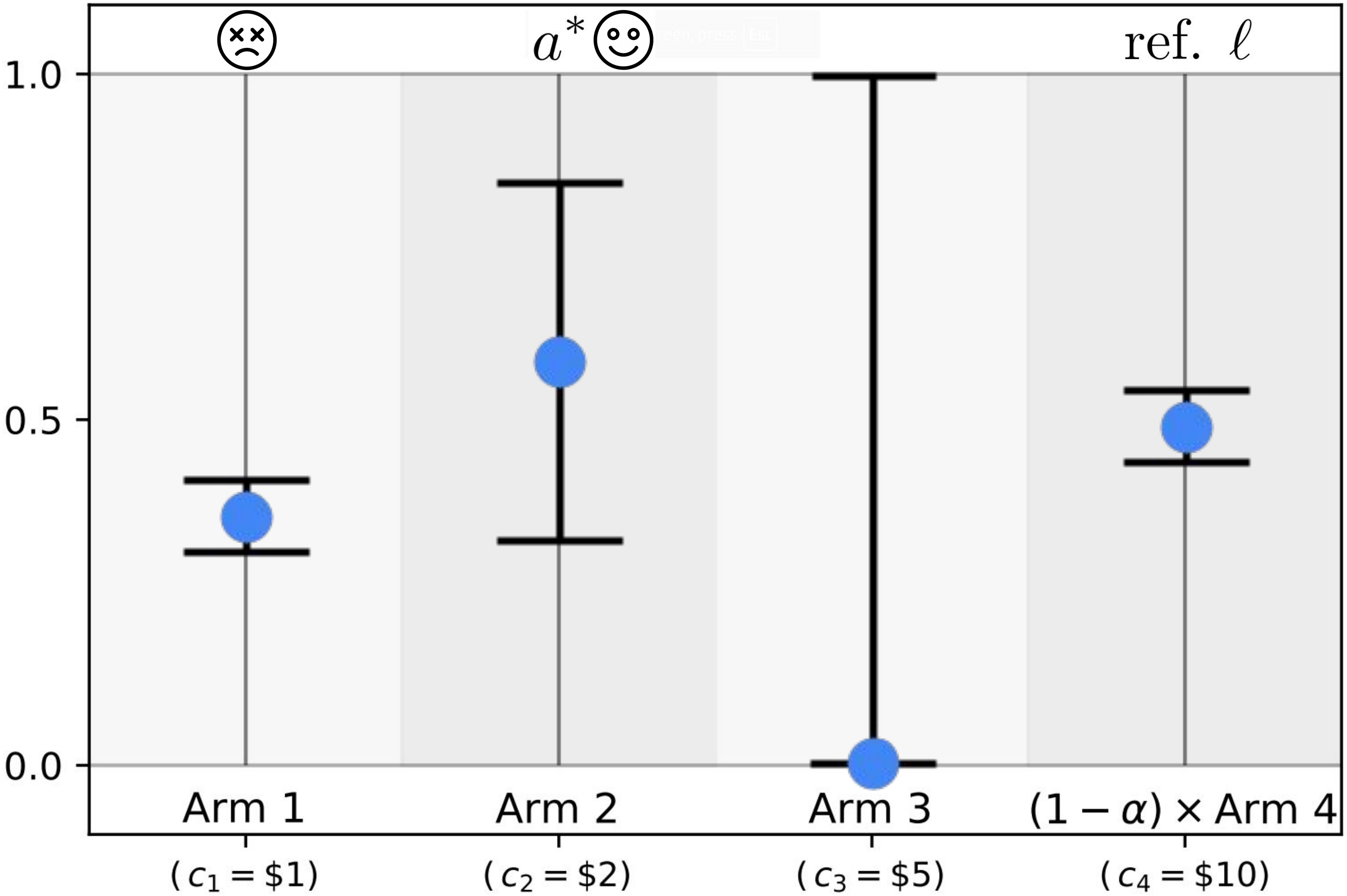
Evaluate feasibility of arm $i = 2$



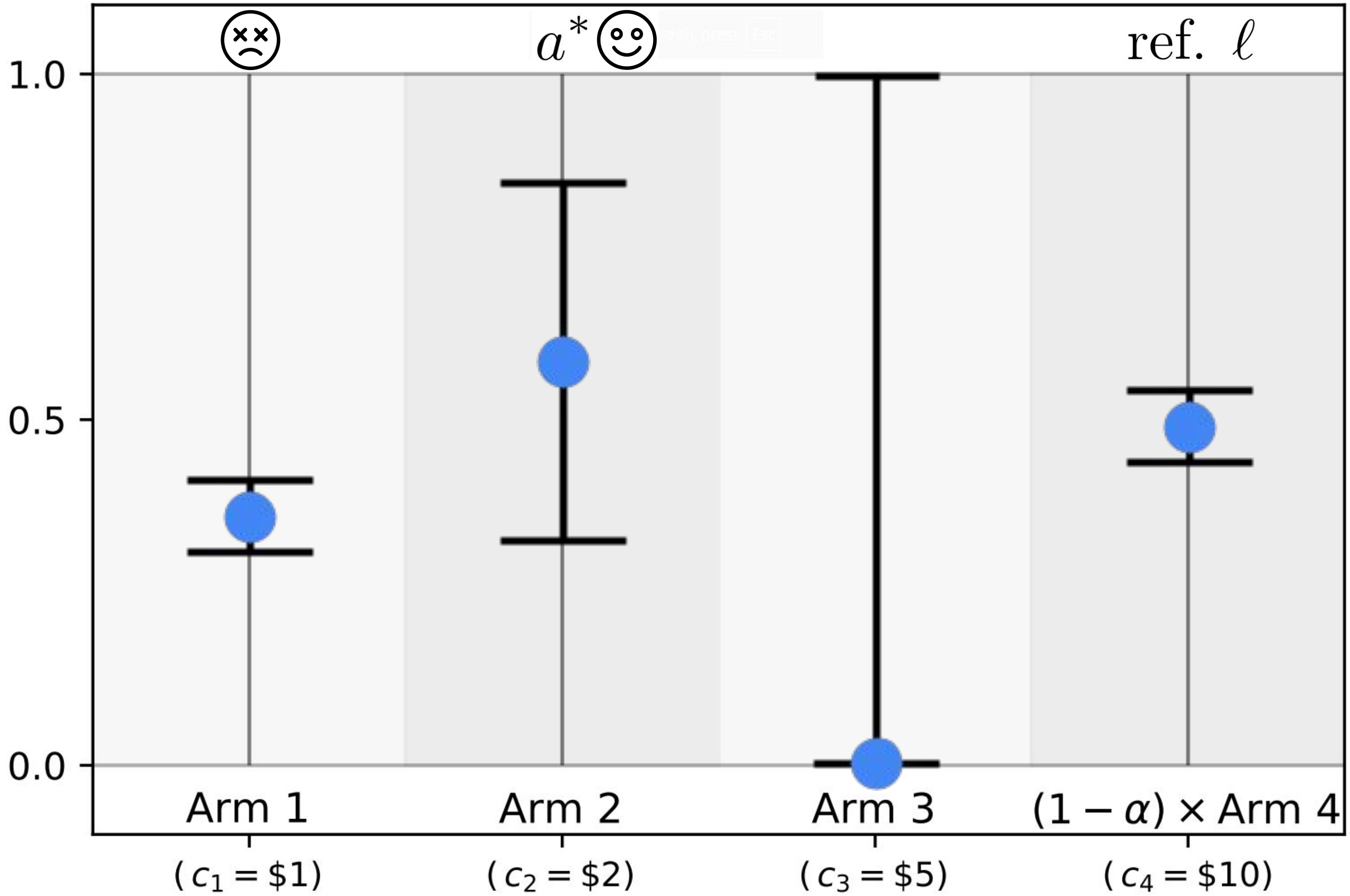
a^*
ref. ℓ
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$



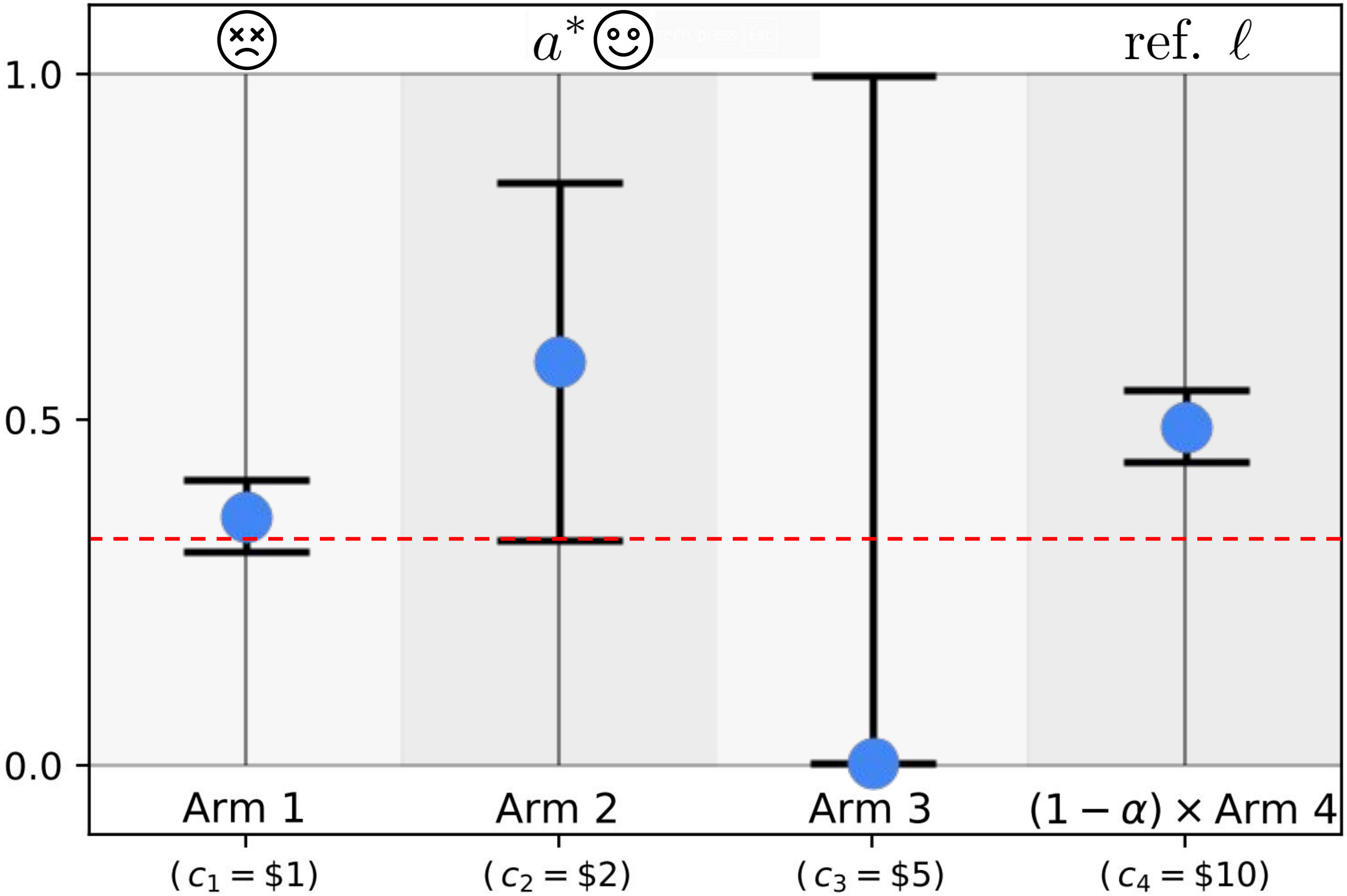
Sample arms $i = 2$ and ℓ to τ_1



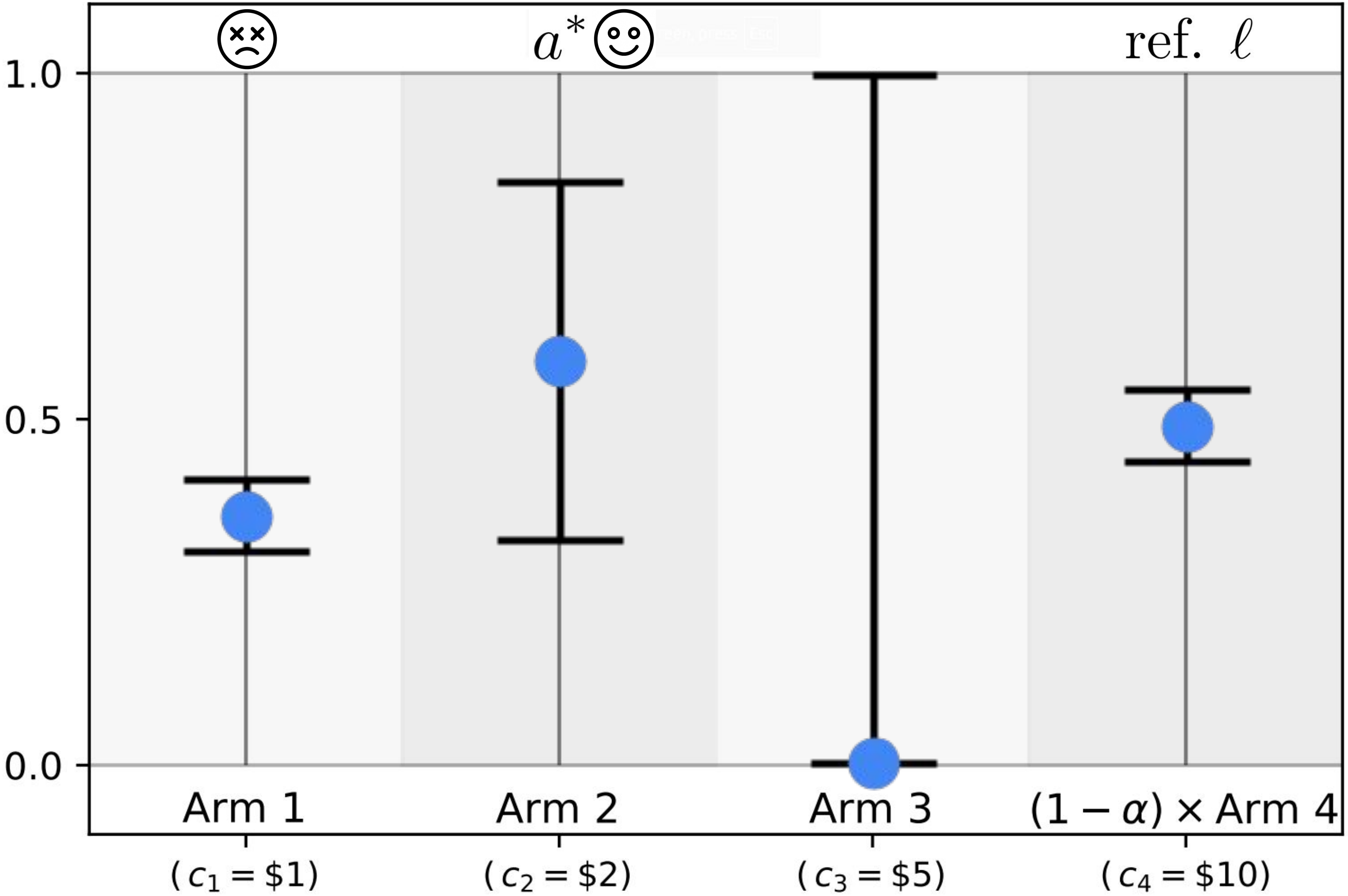
a^*
ref. ℓ
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$



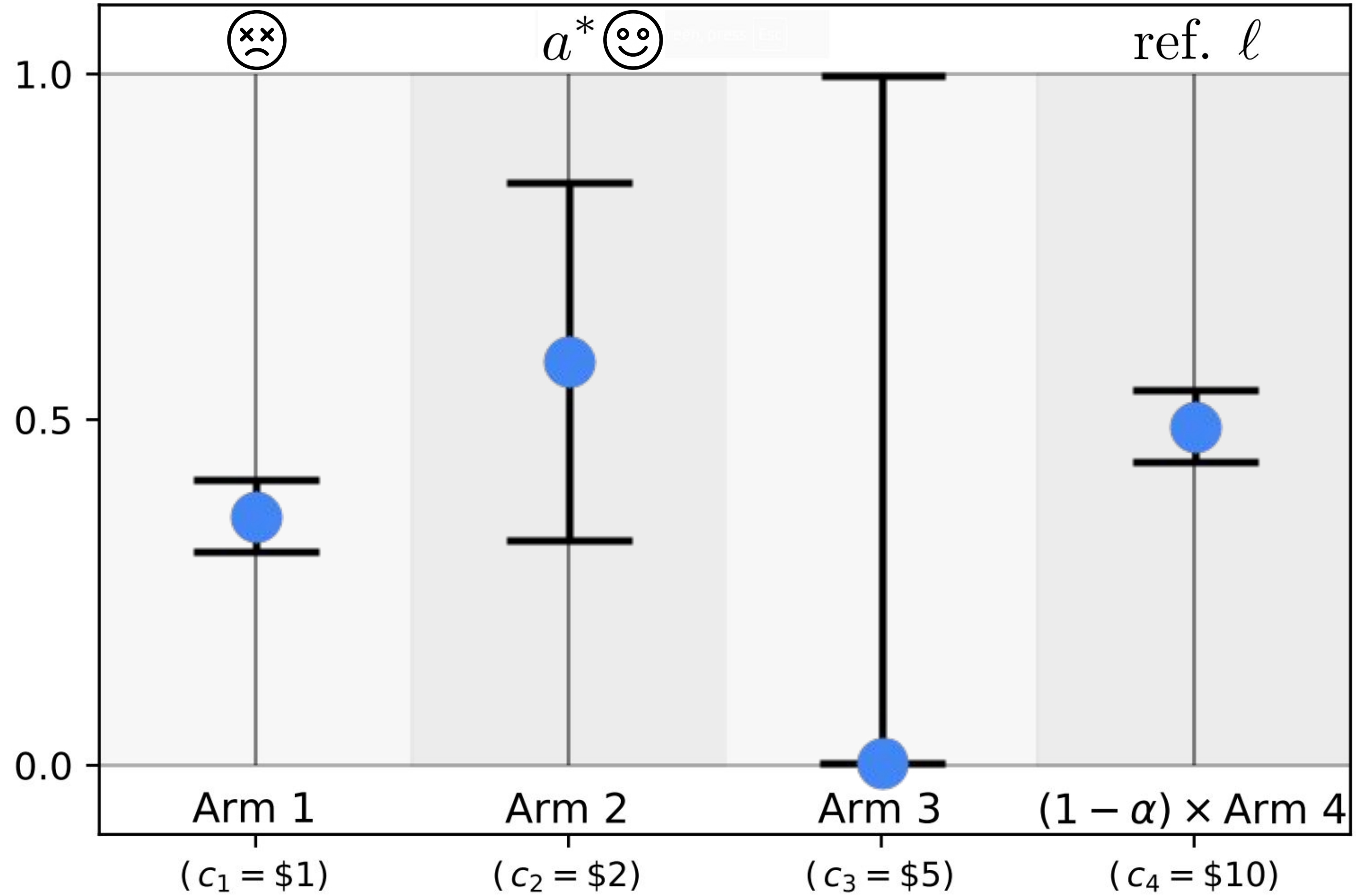
Stopping?
$\mathrm{LCB}_i \geq (1-\alpha)\mathrm{UCB}_\ell$



Stopping?
$\mathrm{LCB}_i \geq (1-\alpha)\mathrm{UCB}_\ell$

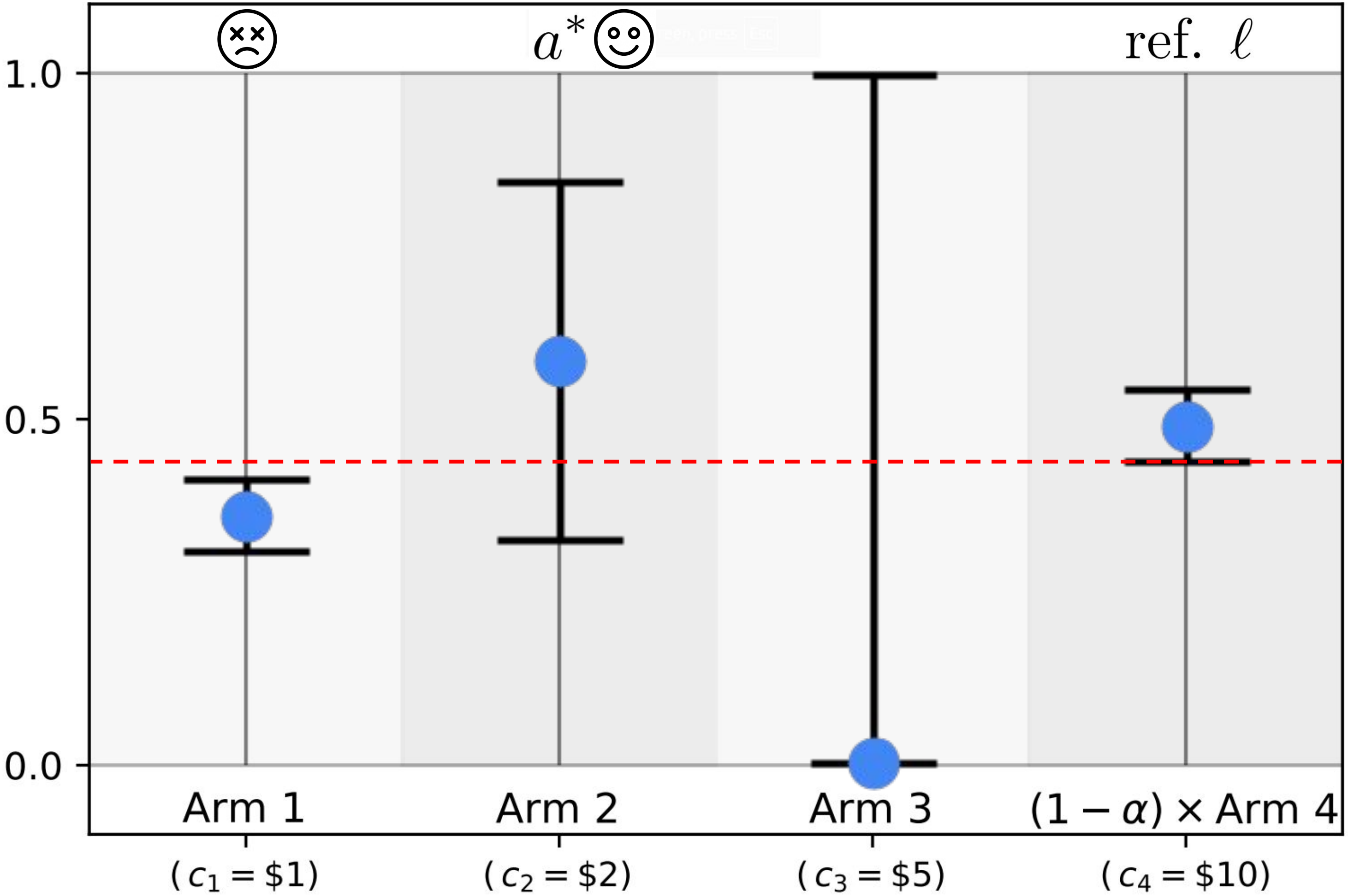


a^*
ref. ℓ
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$

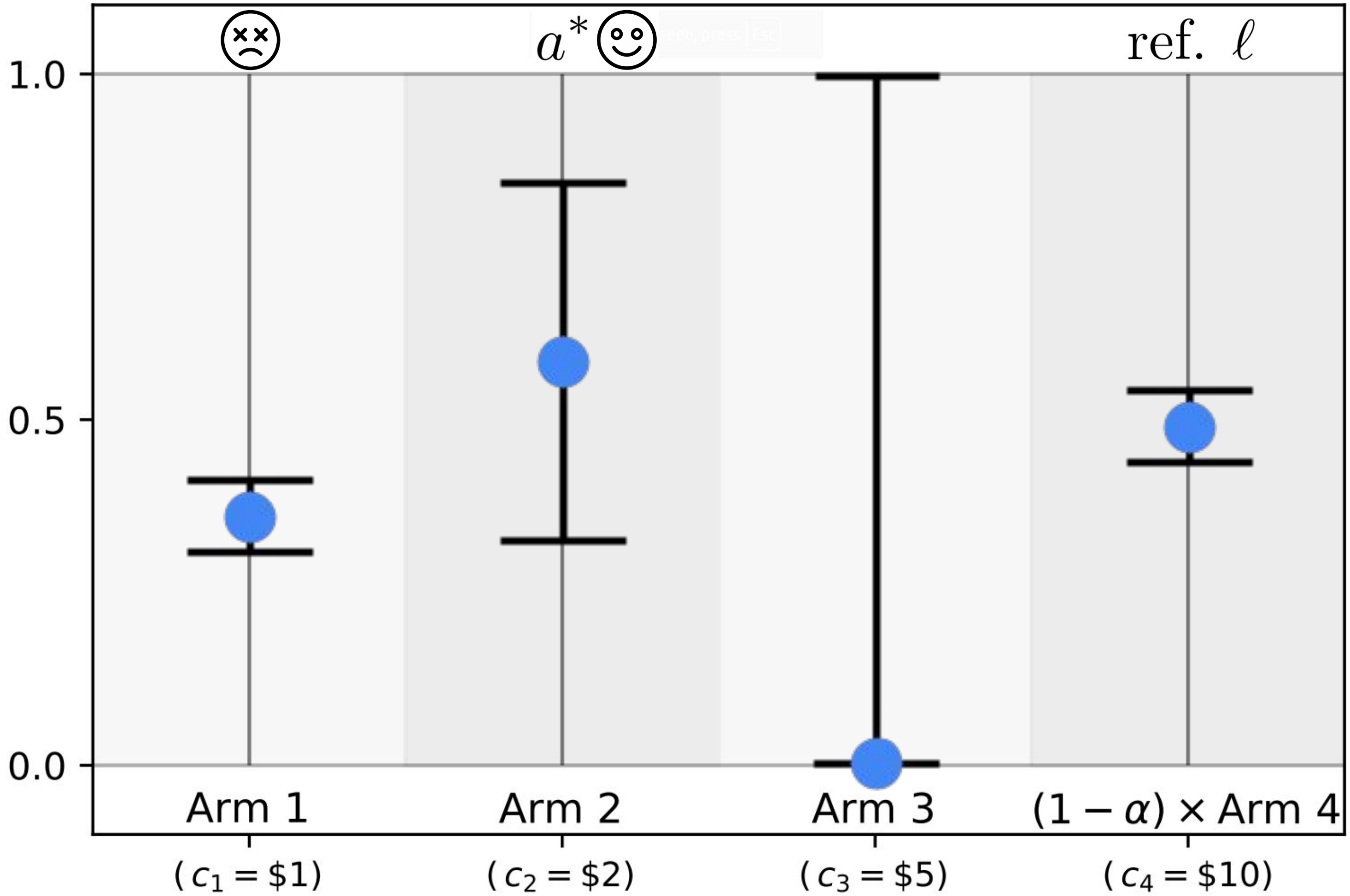


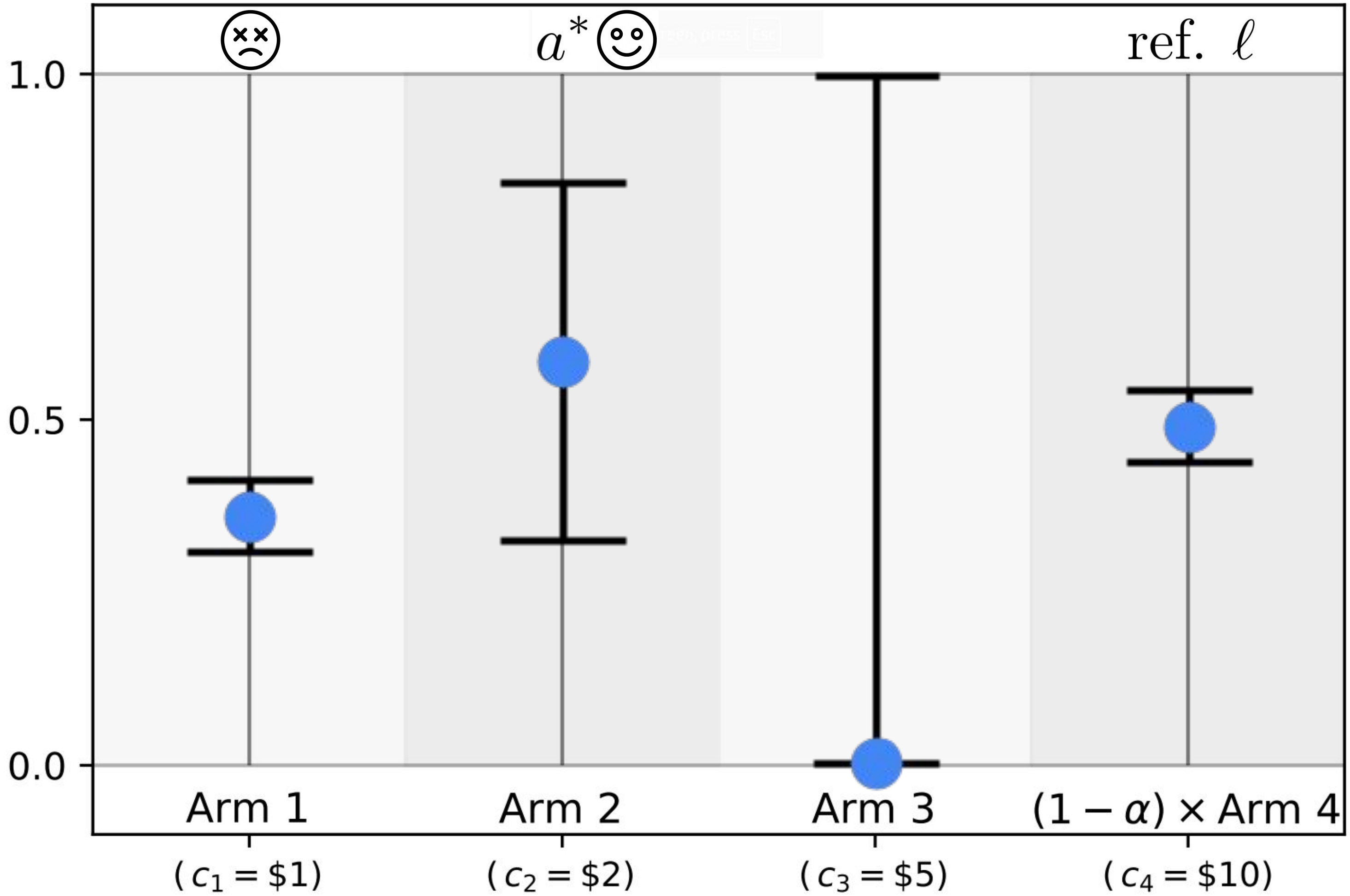
Elimination?

$(1-\alpha) \times \mathrm{LCB}_\ell \geq \mathrm{UCB}_i$

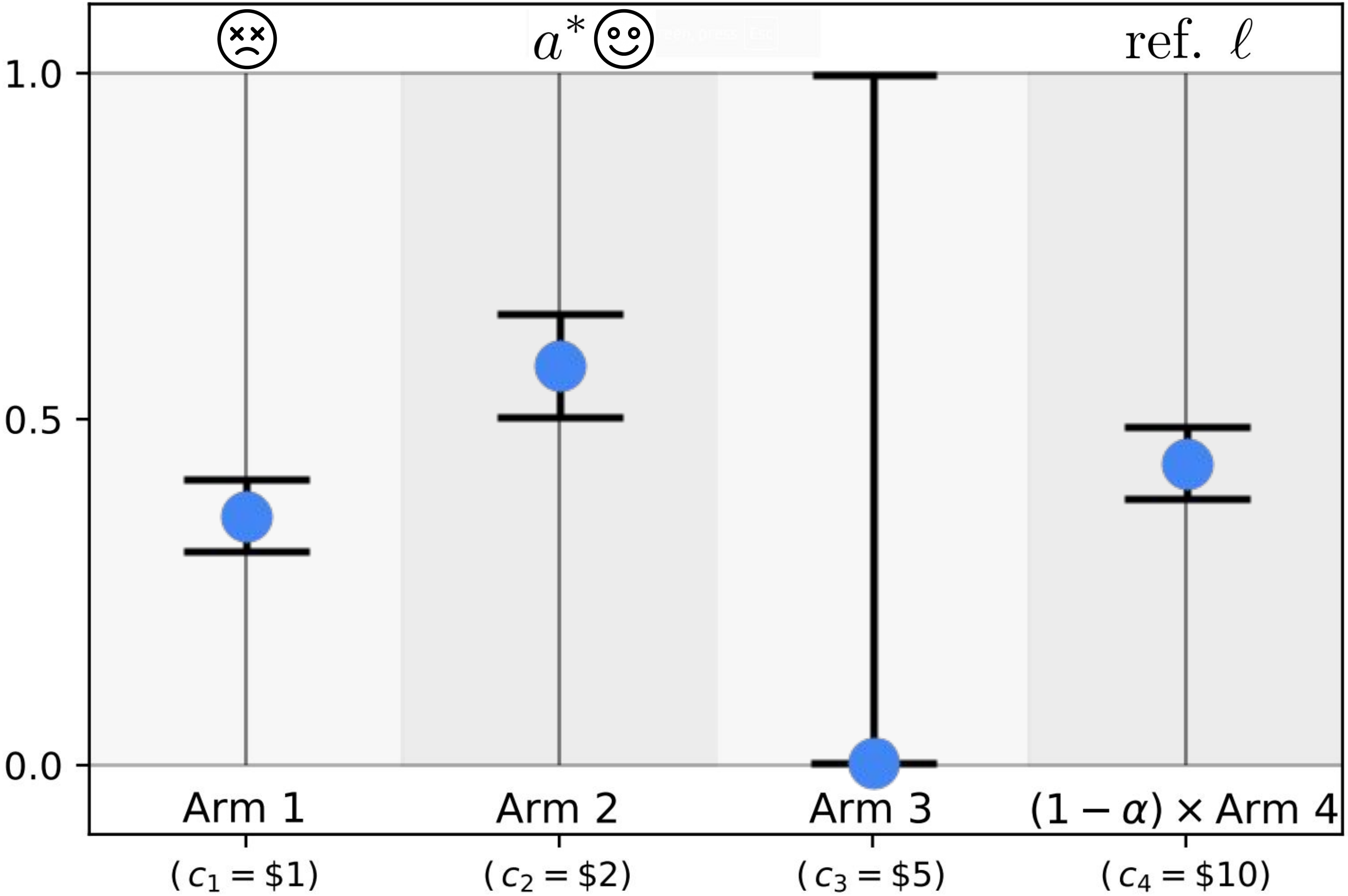


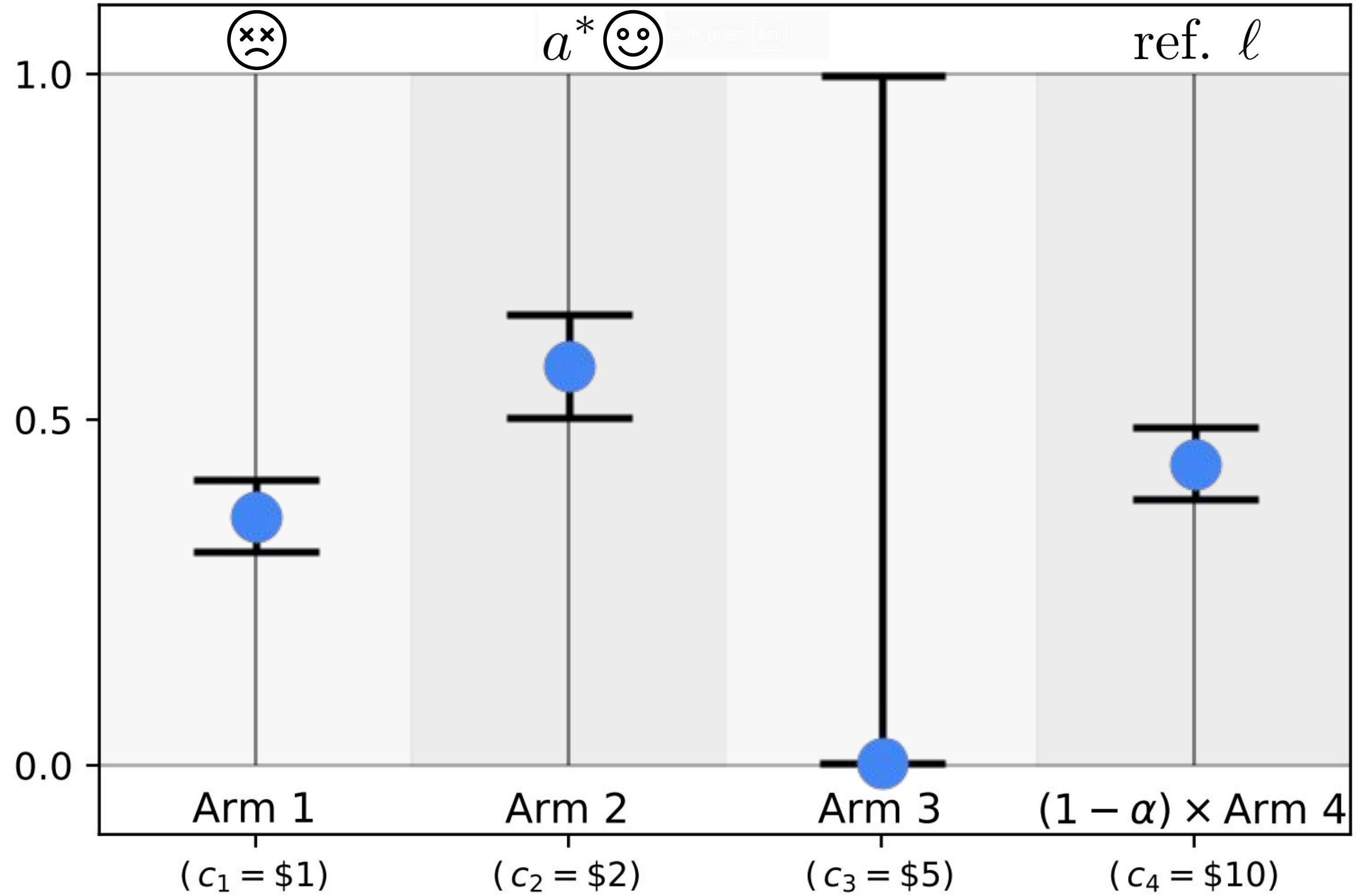
Elimination?
$(1-\alpha) \times \mathrm{LCB}_\ell \geq \mathrm{UCB}_i$


a^*
ref. ℓ
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$

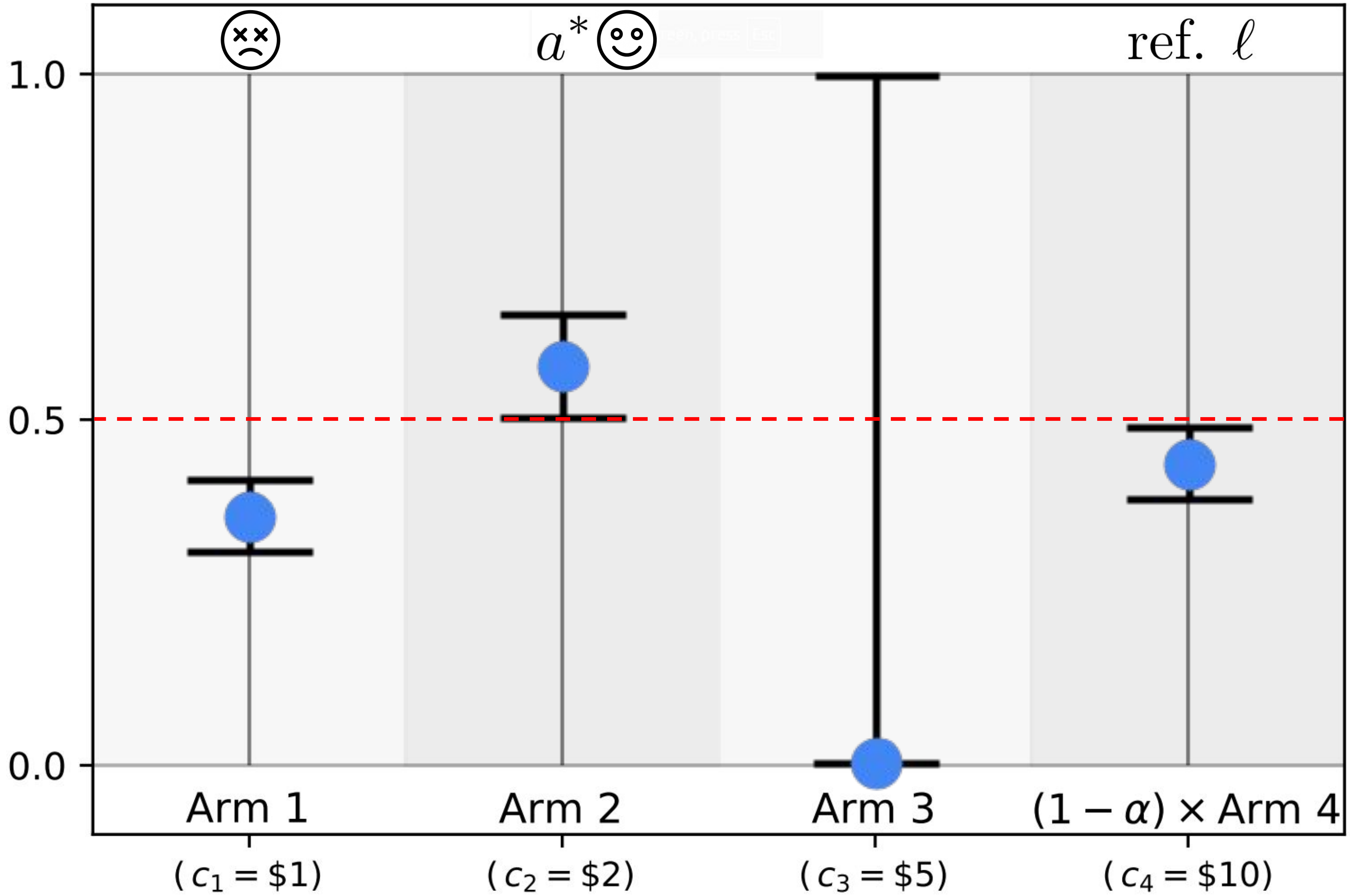


Sample arms $i = 2$ and ℓ to τ_2


a^*
ref. ℓ
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$

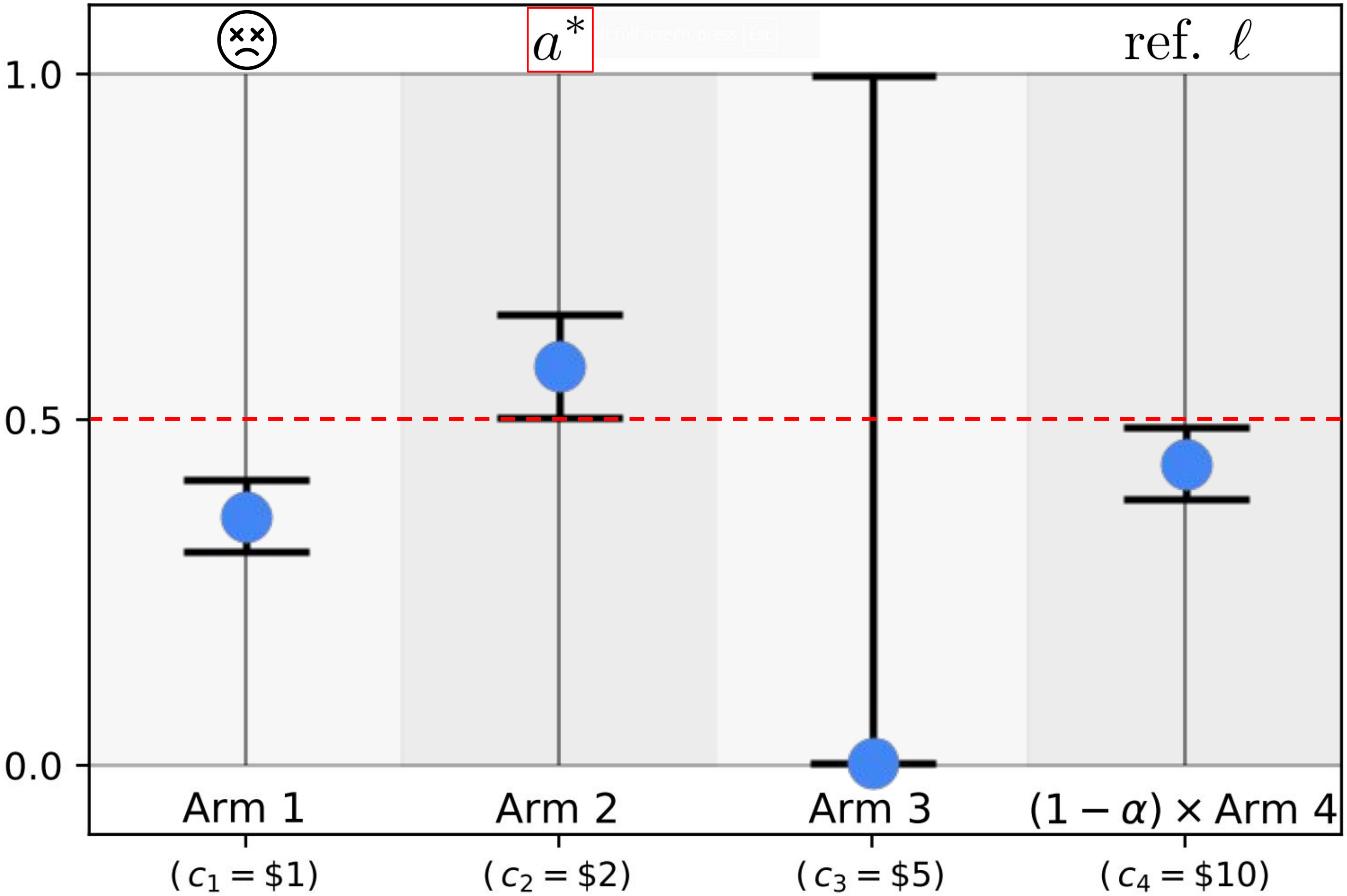


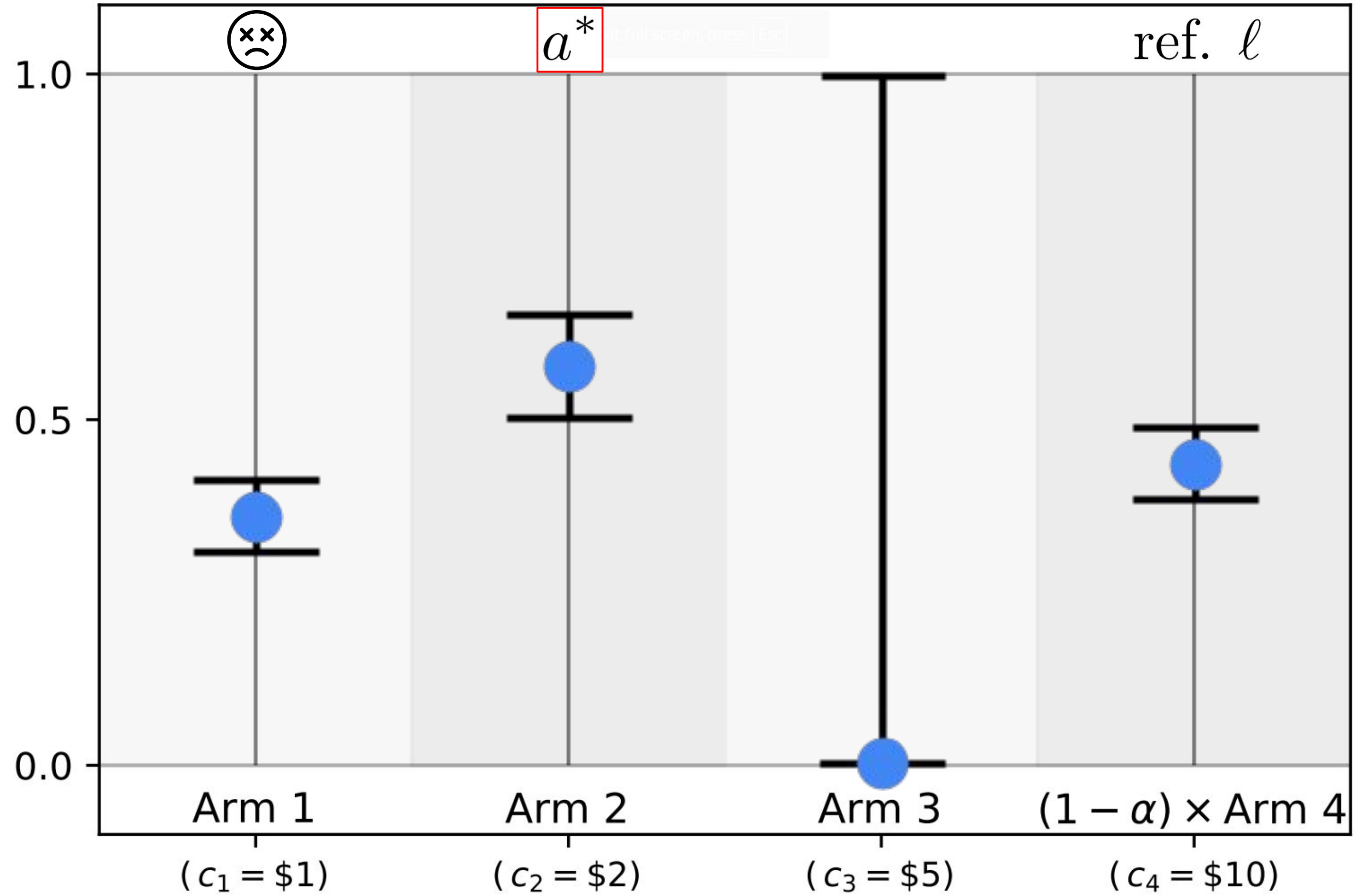
Stopping?
$\mathrm{LCB}_i \geq (1-\alpha)\mathrm{UCB}_\ell$



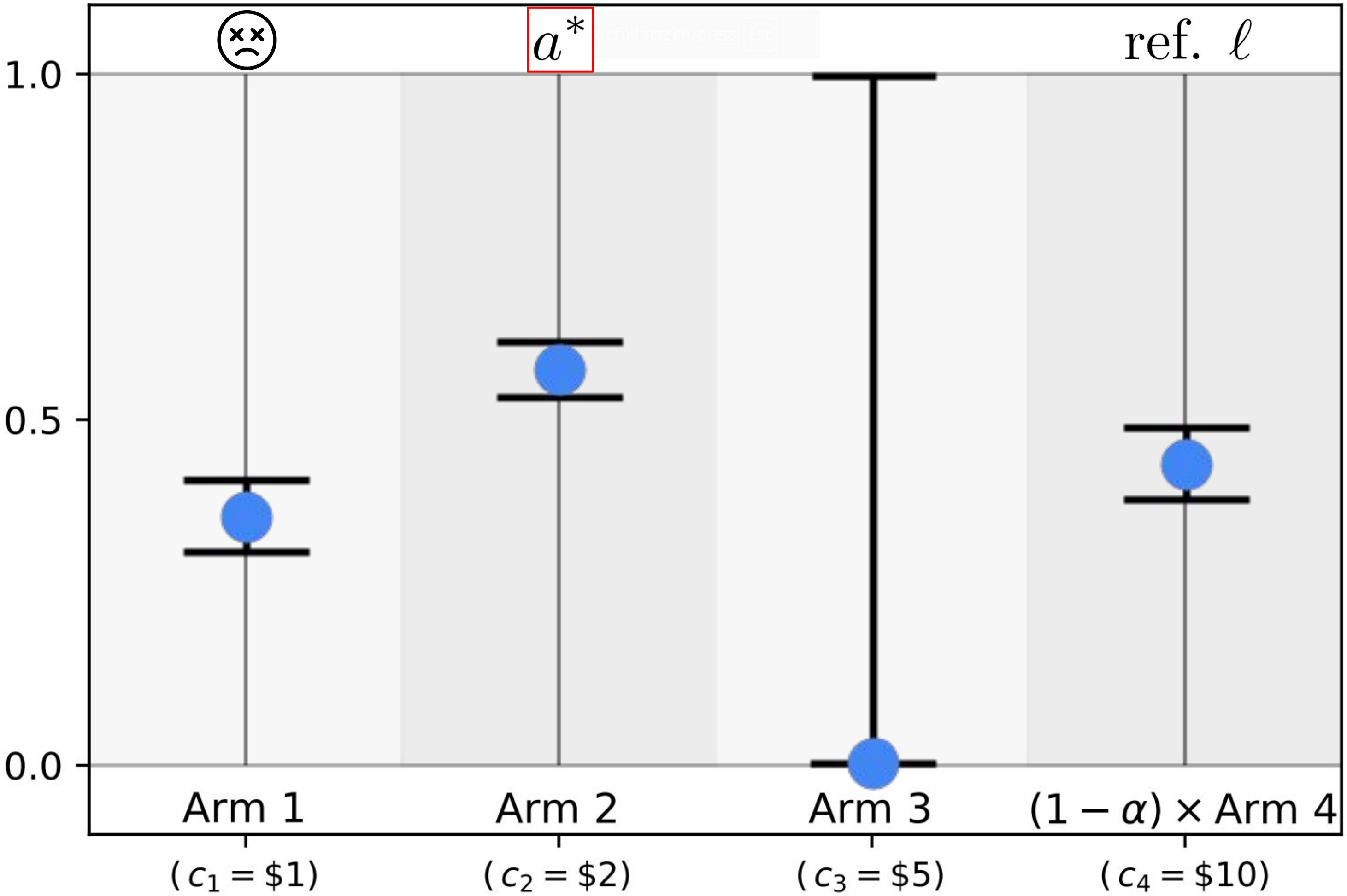
Stopping?

$\mathrm{LCB}_i \geq (1-\alpha)\mathrm{UCB}_\ell$


a^*
ref. ℓ
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$



Sample identified a^* for remaining samples


a*
ref. ℓ
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
(1 − α) × Arm 4
(c1 = $1)
(c2 = $2)
(c3 = $5)
(c4 = $10)

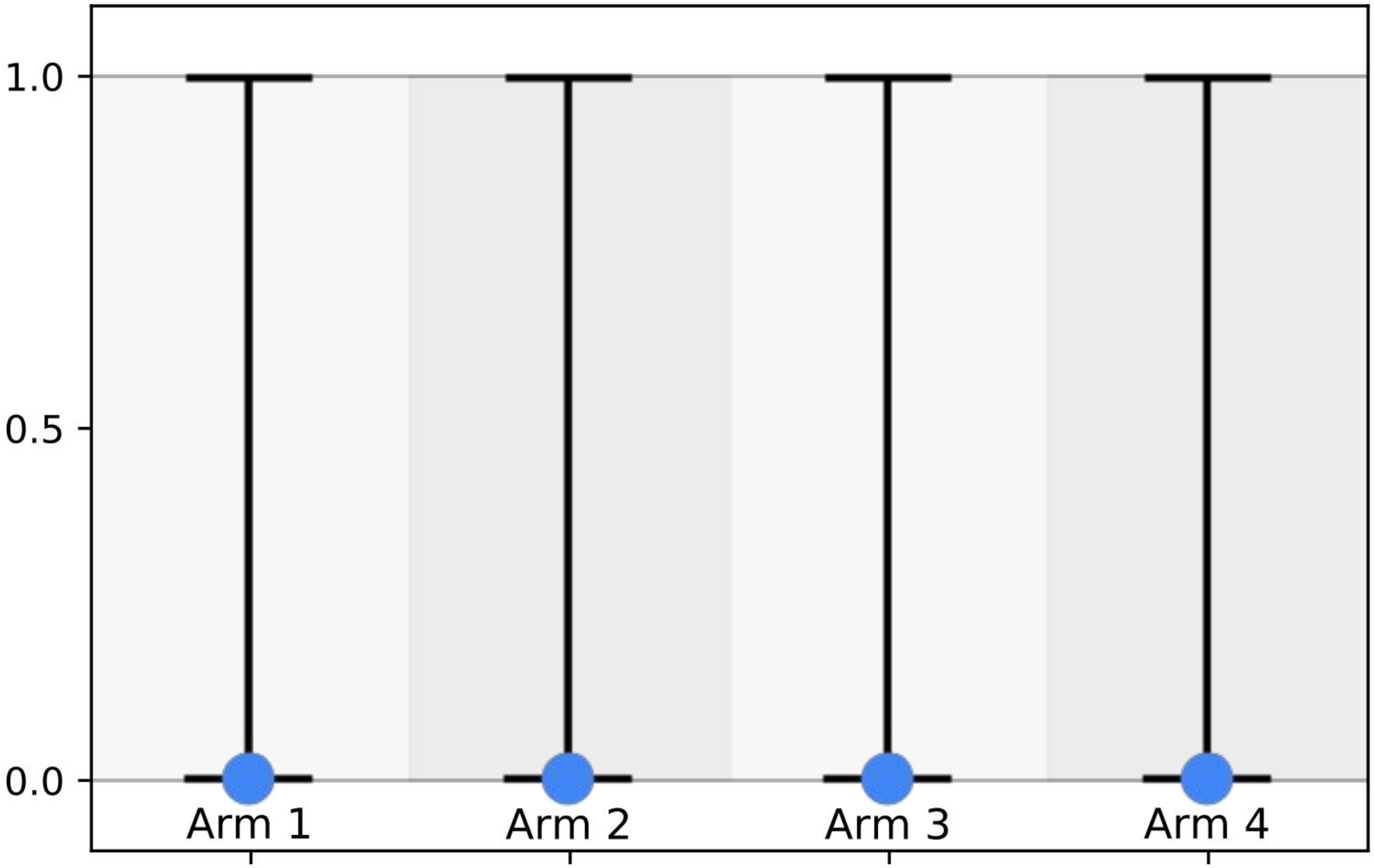
PE-CS Algorithm

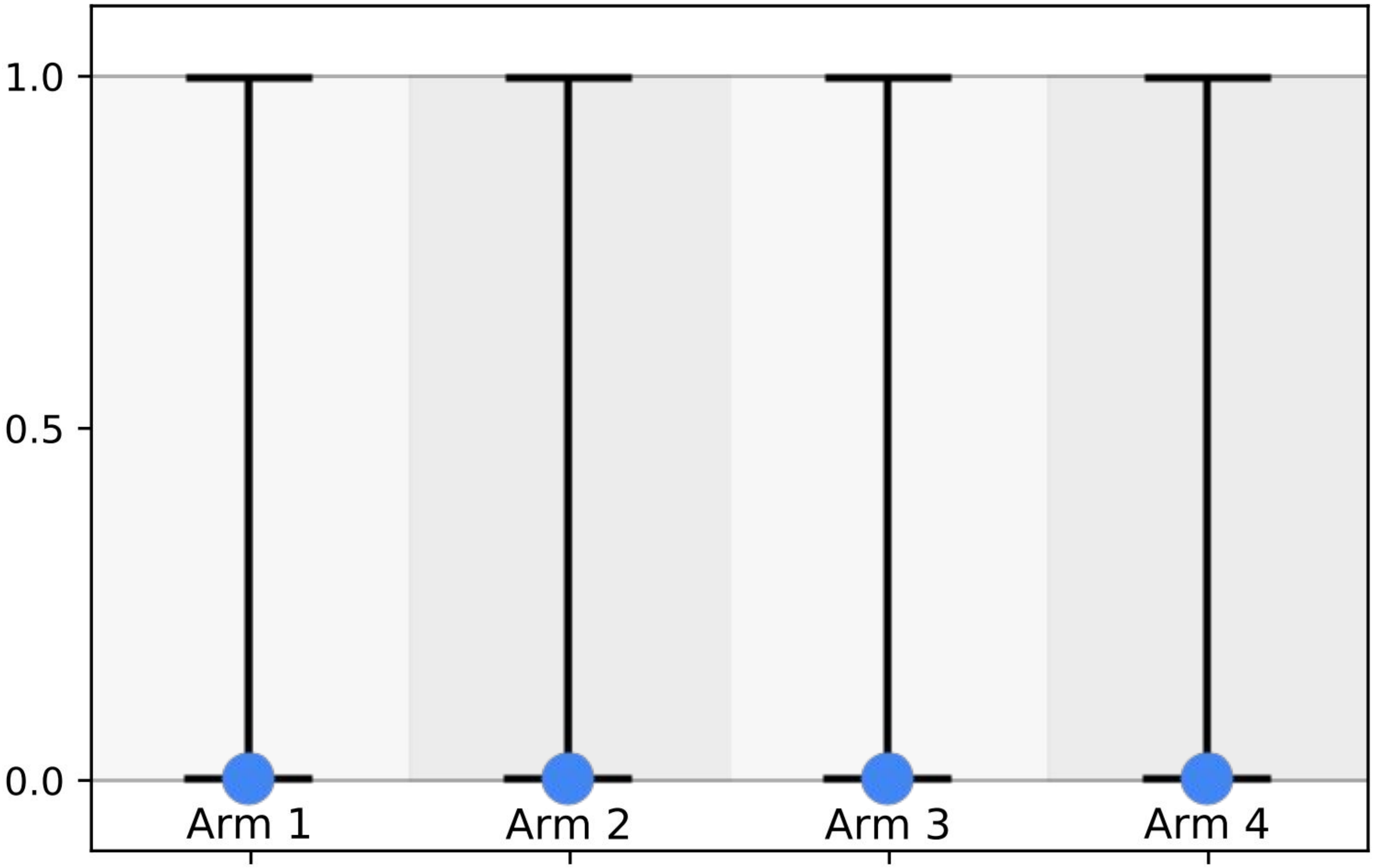
PE-CS Algorithm

$$\mu_{\mathrm{CS}} = (1 - \alpha)\mu^*$$

PE-CS Algorithm

$$\mu_{\mathrm{CS}} = (1 - \alpha)\mu^*$$


1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
Arm 4



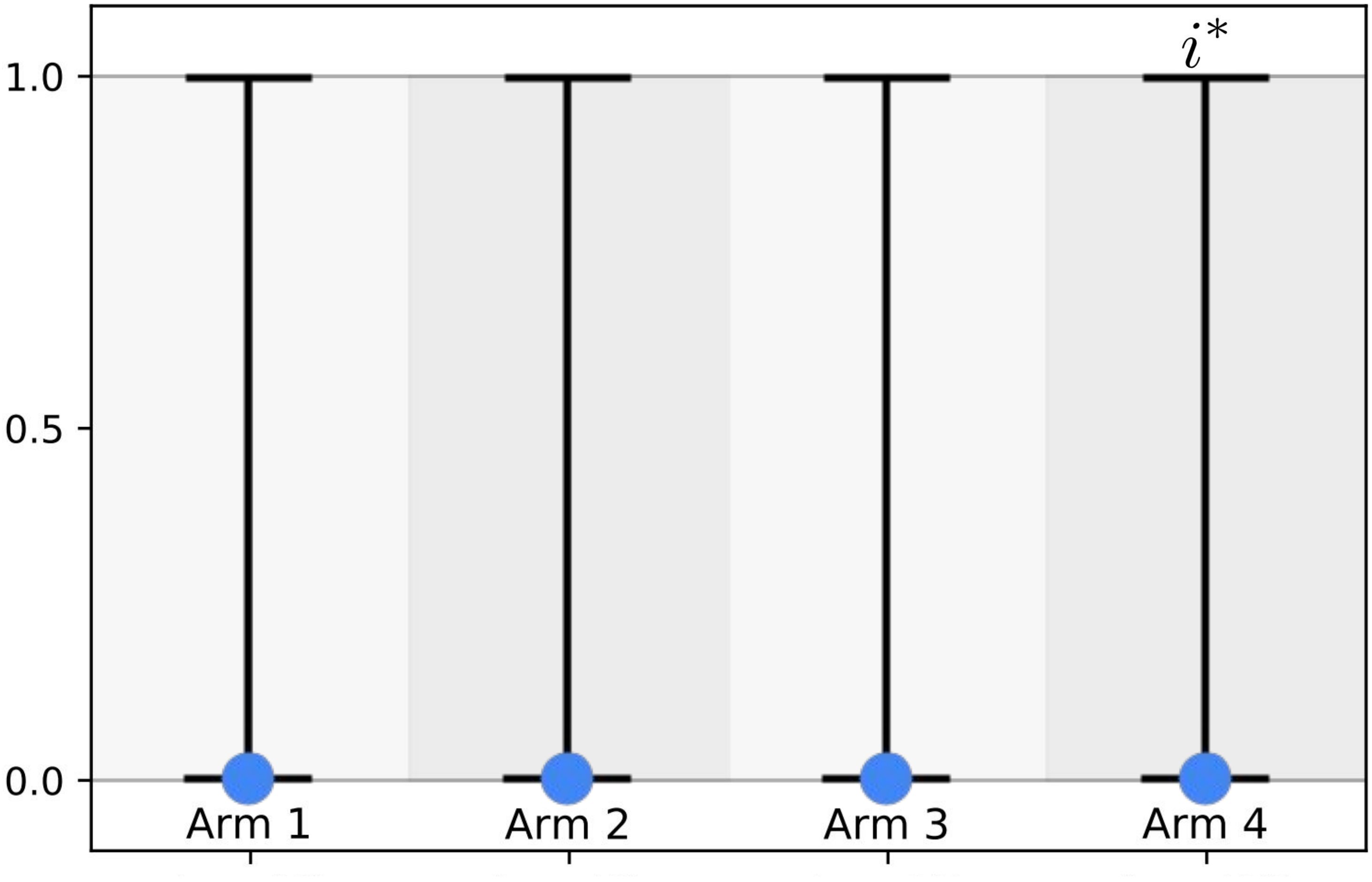
Identify i^* as
best reward arm

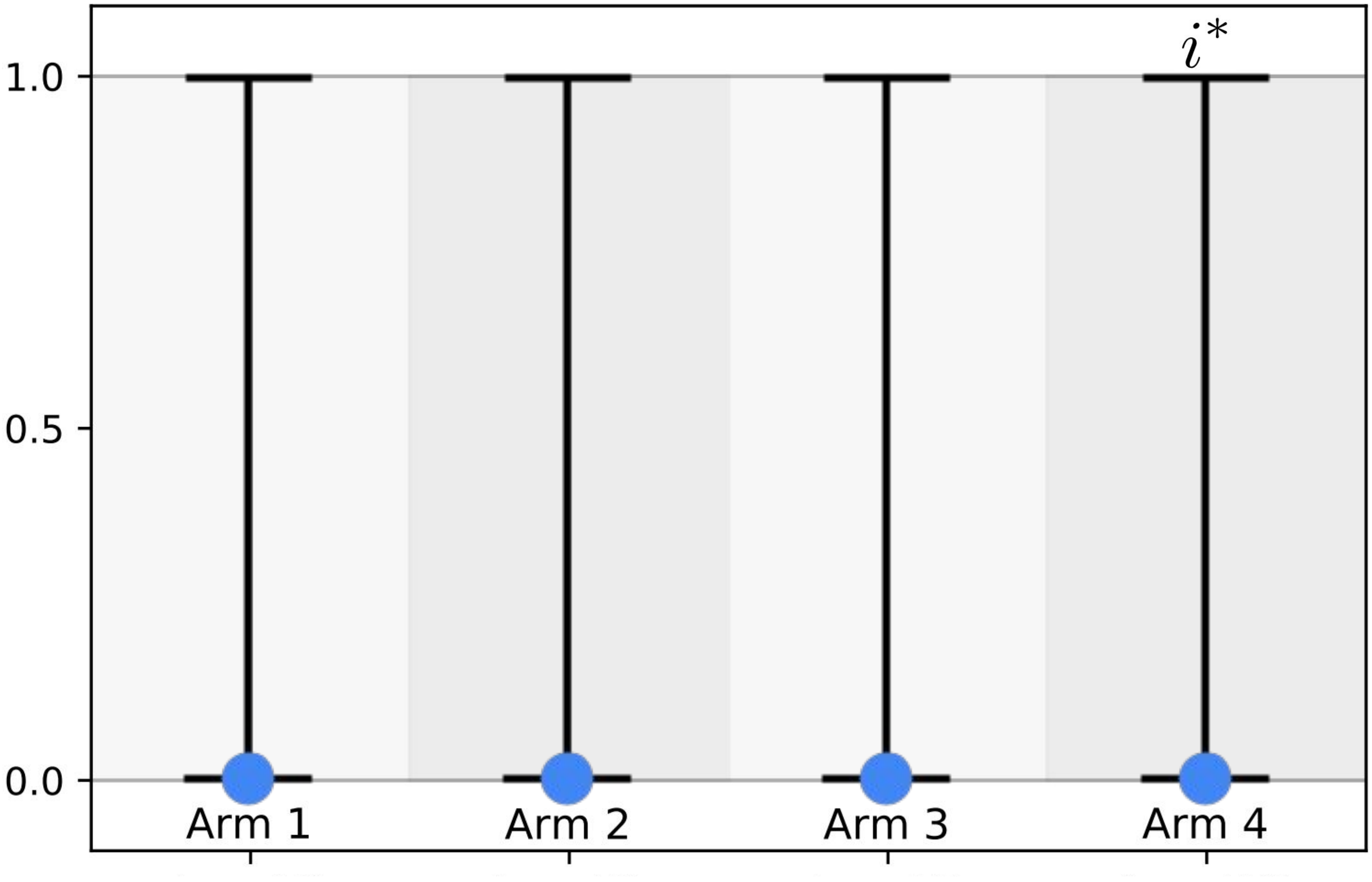
BAI Stage

(**B**est **A**rm **I**dentification)

BAI Stage

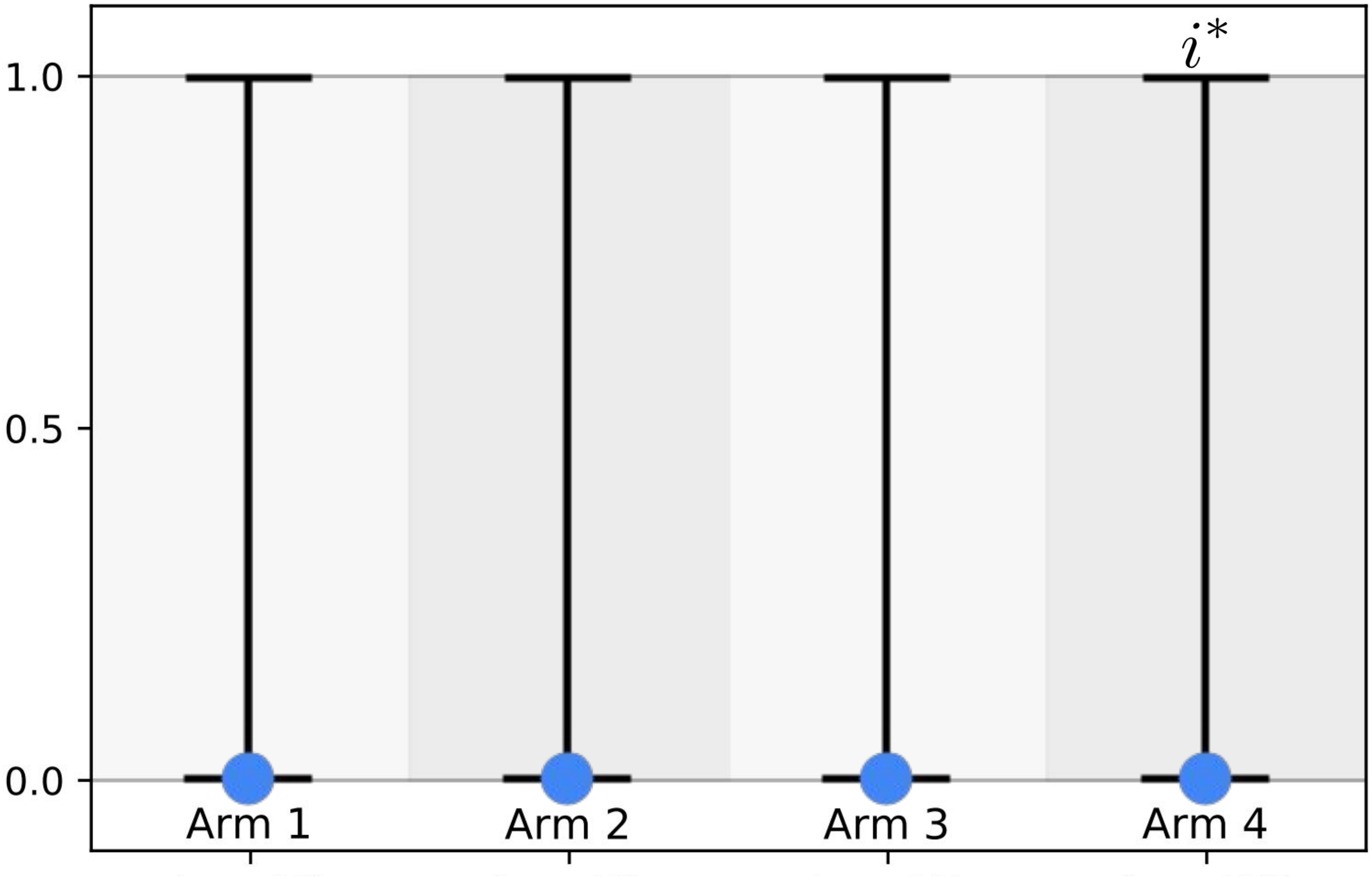
Identify best reward arm i^*

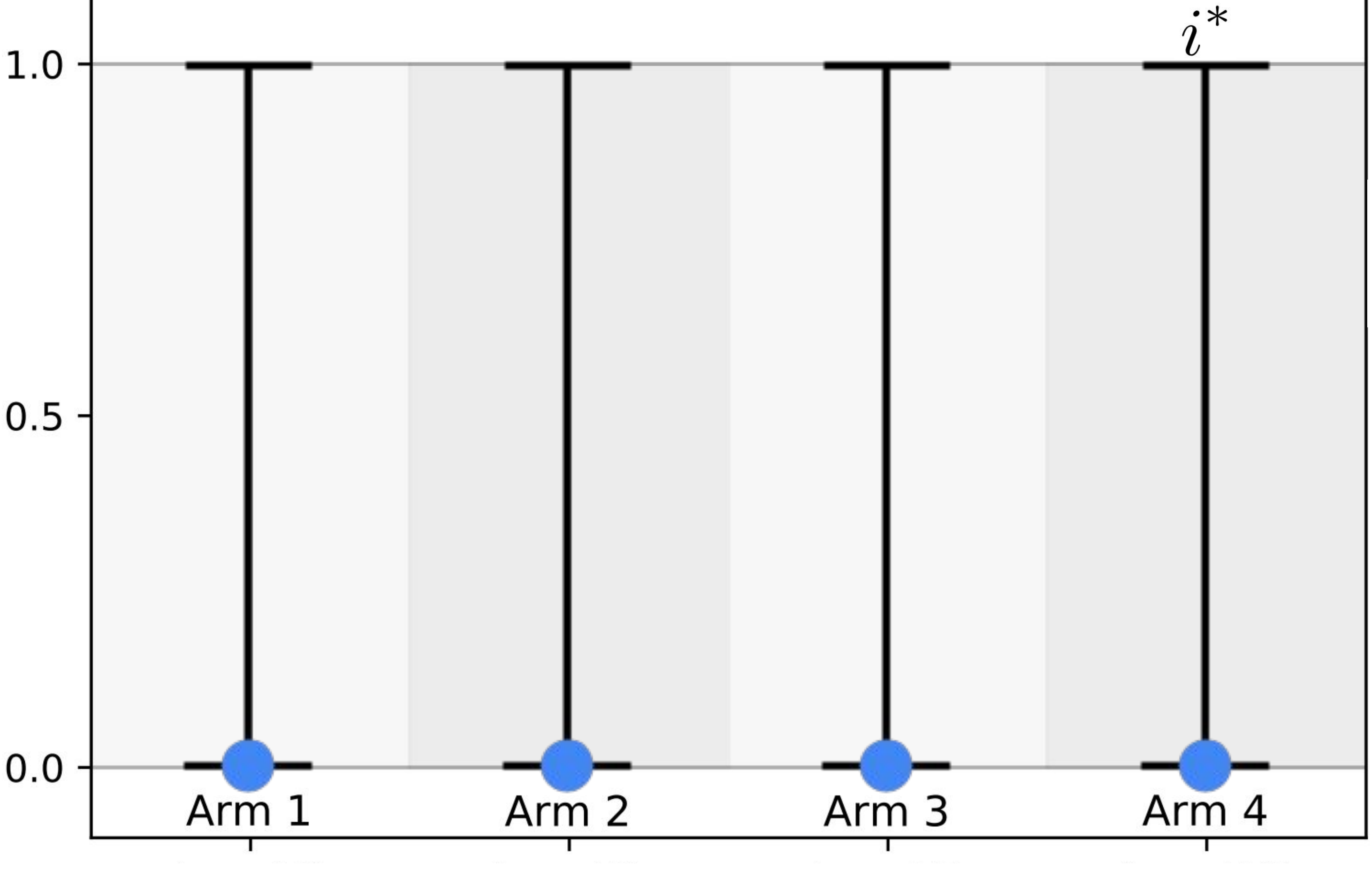

i^*
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
Arm 4



$$\hat{\mu}_i(t) = 0$$

$$n_i(t) = 0$$


i^*
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
Arm 4

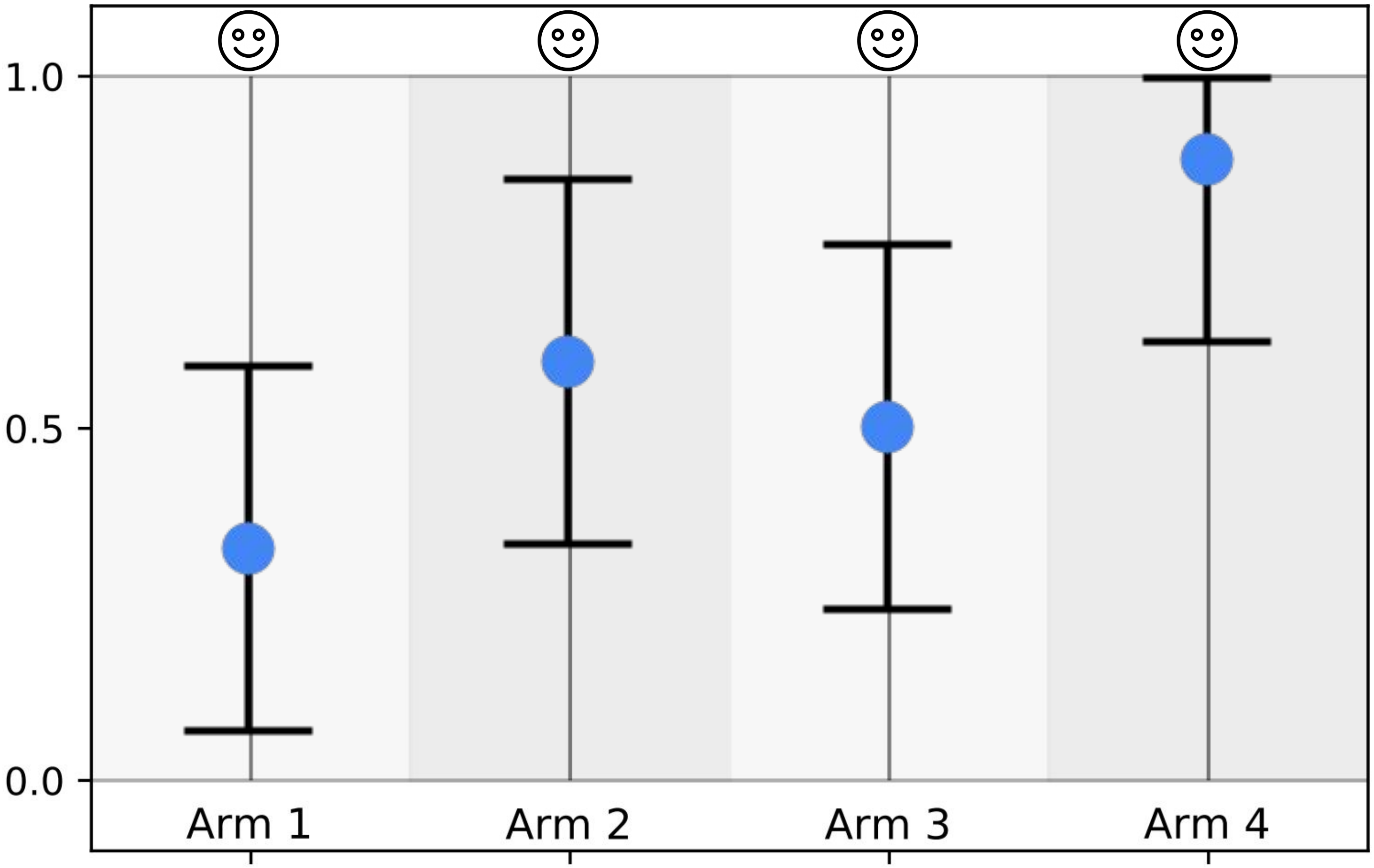


Mark all arms
as active

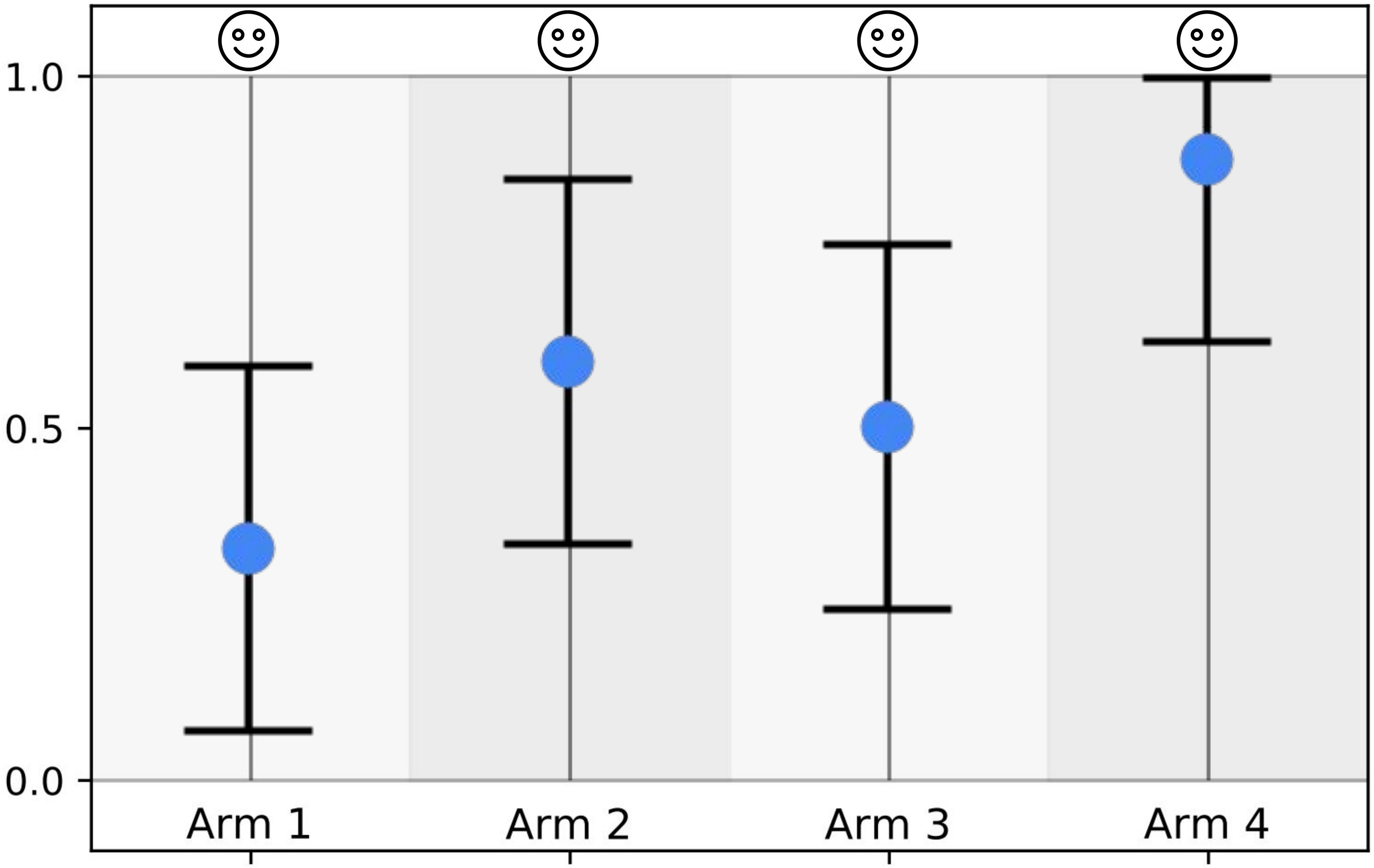
Mark all arms as active


1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
Arm 4

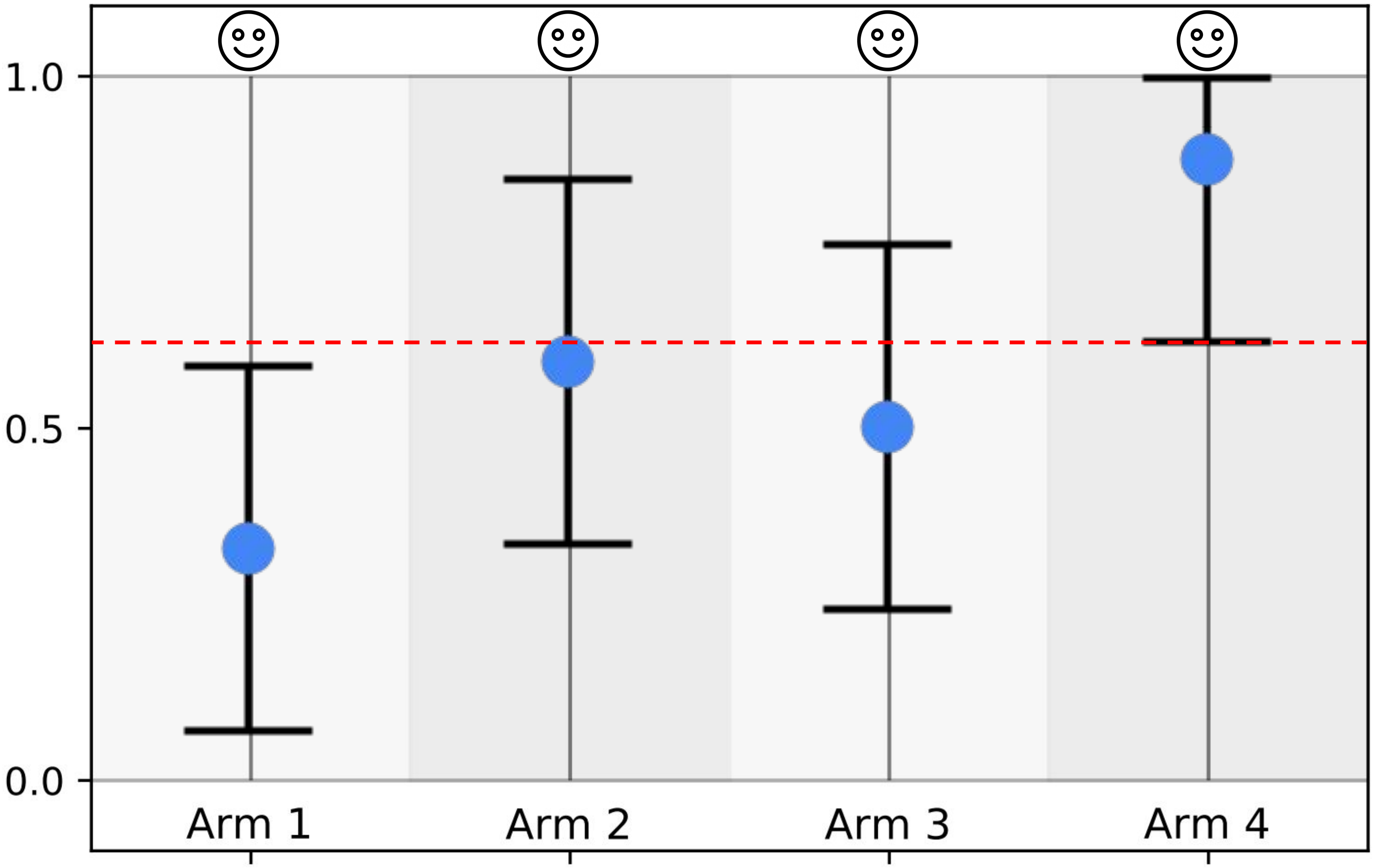
Sample active arms to τ_1



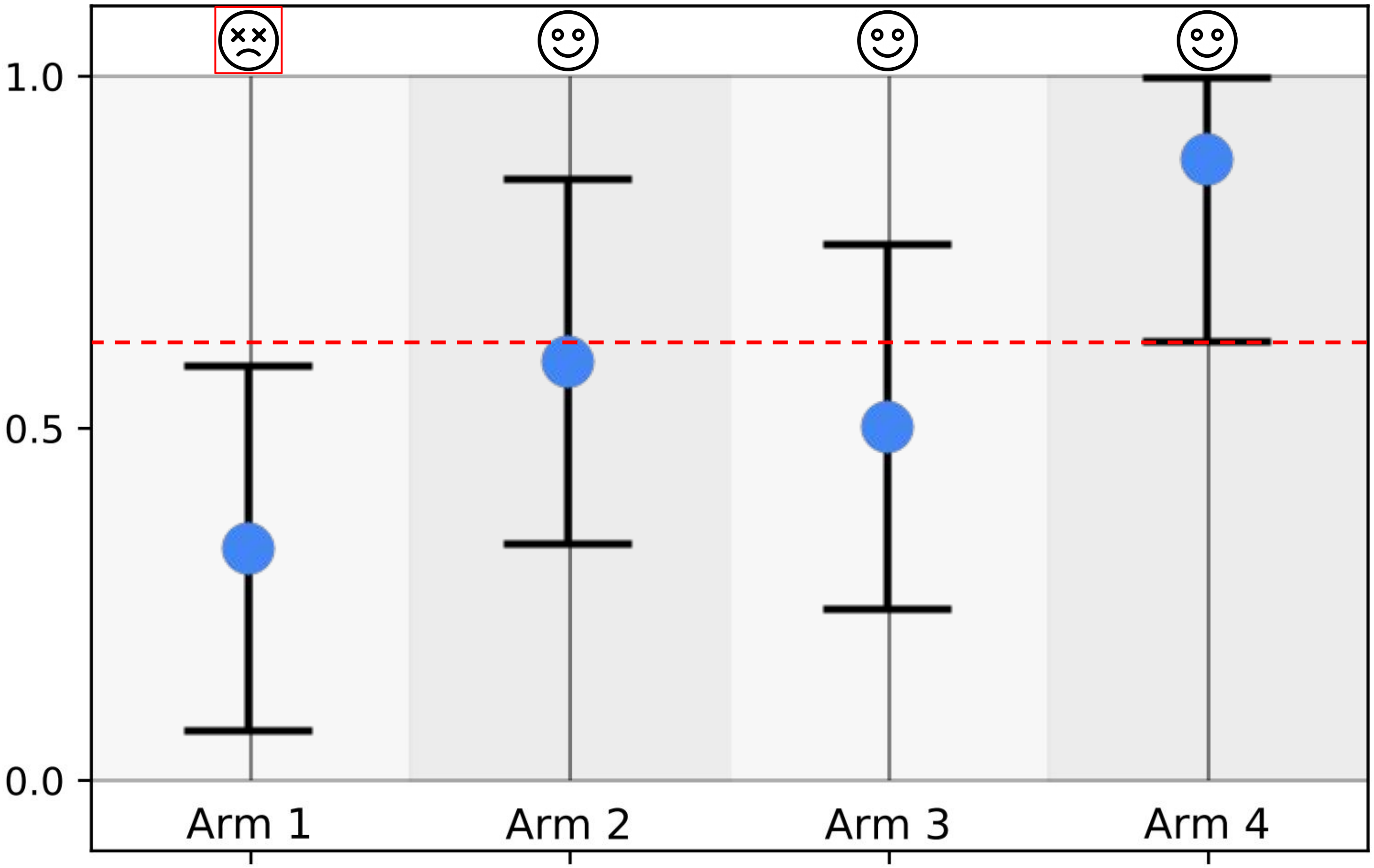
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
Arm 4



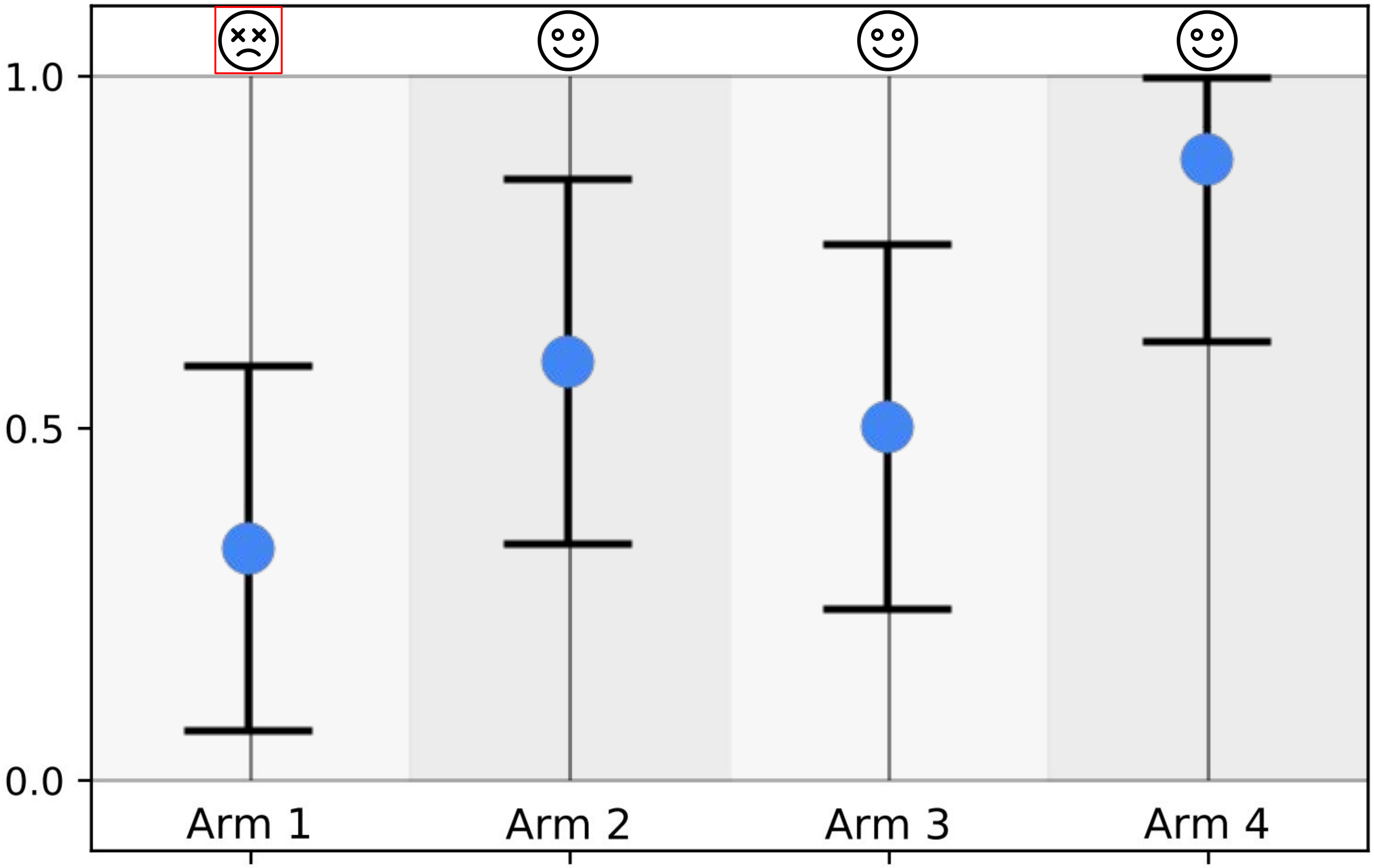
Delete i if
$\mathrm{UCB}_i \leq \max_j \mathrm{LCB}_j$

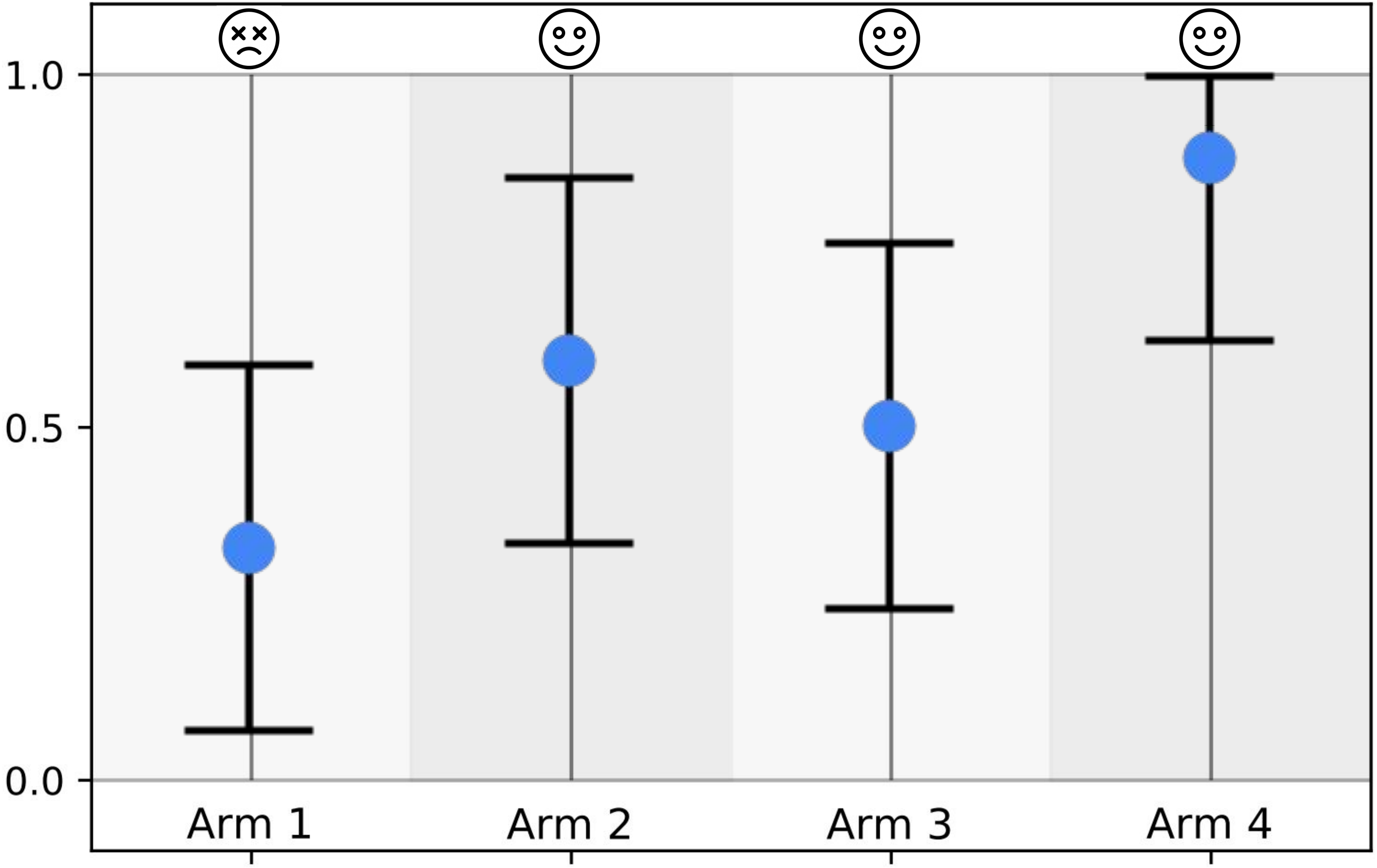


Delete i if
$\mathrm{UCB}_i \leq \max_j \mathrm{LCB}_j$

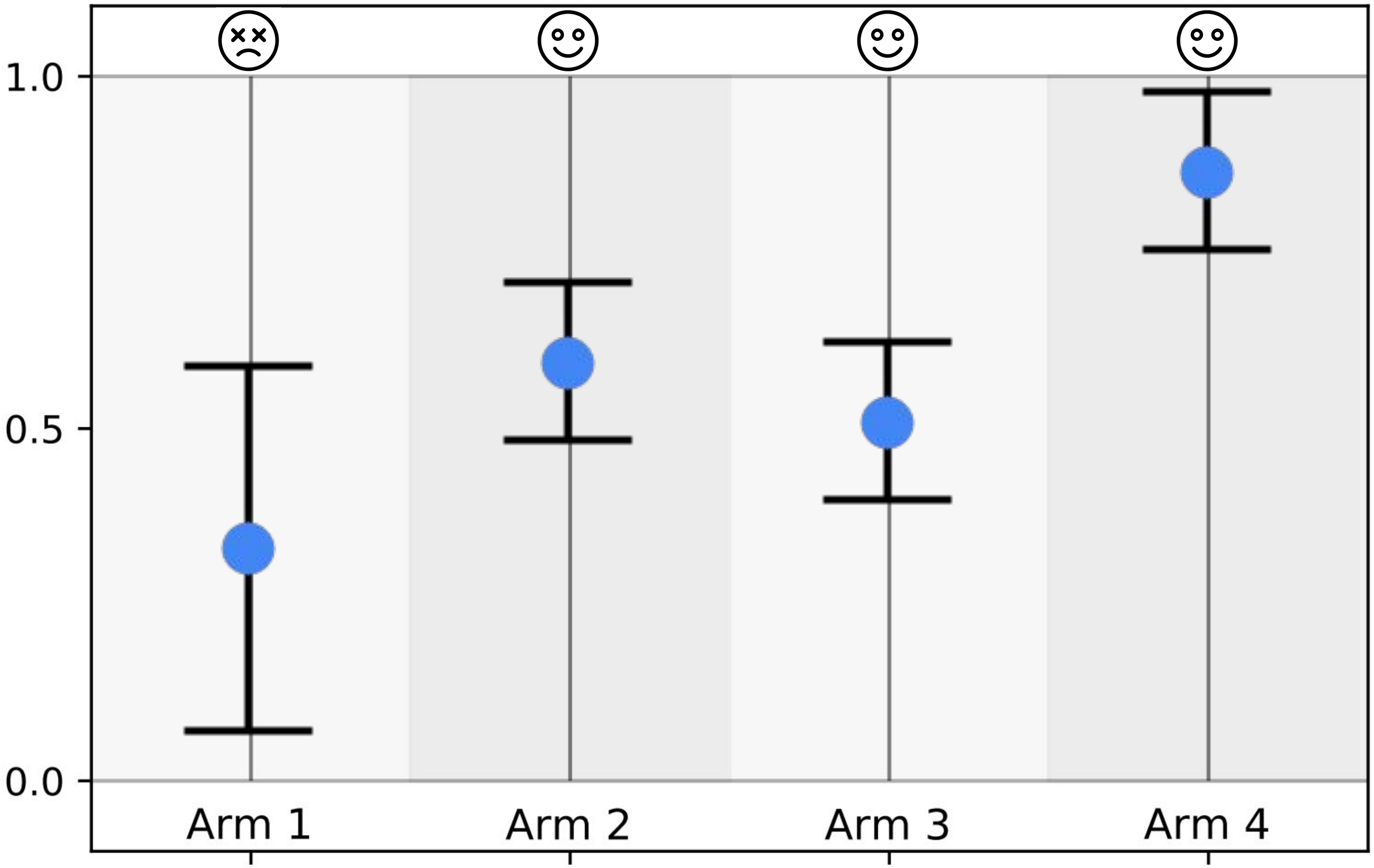


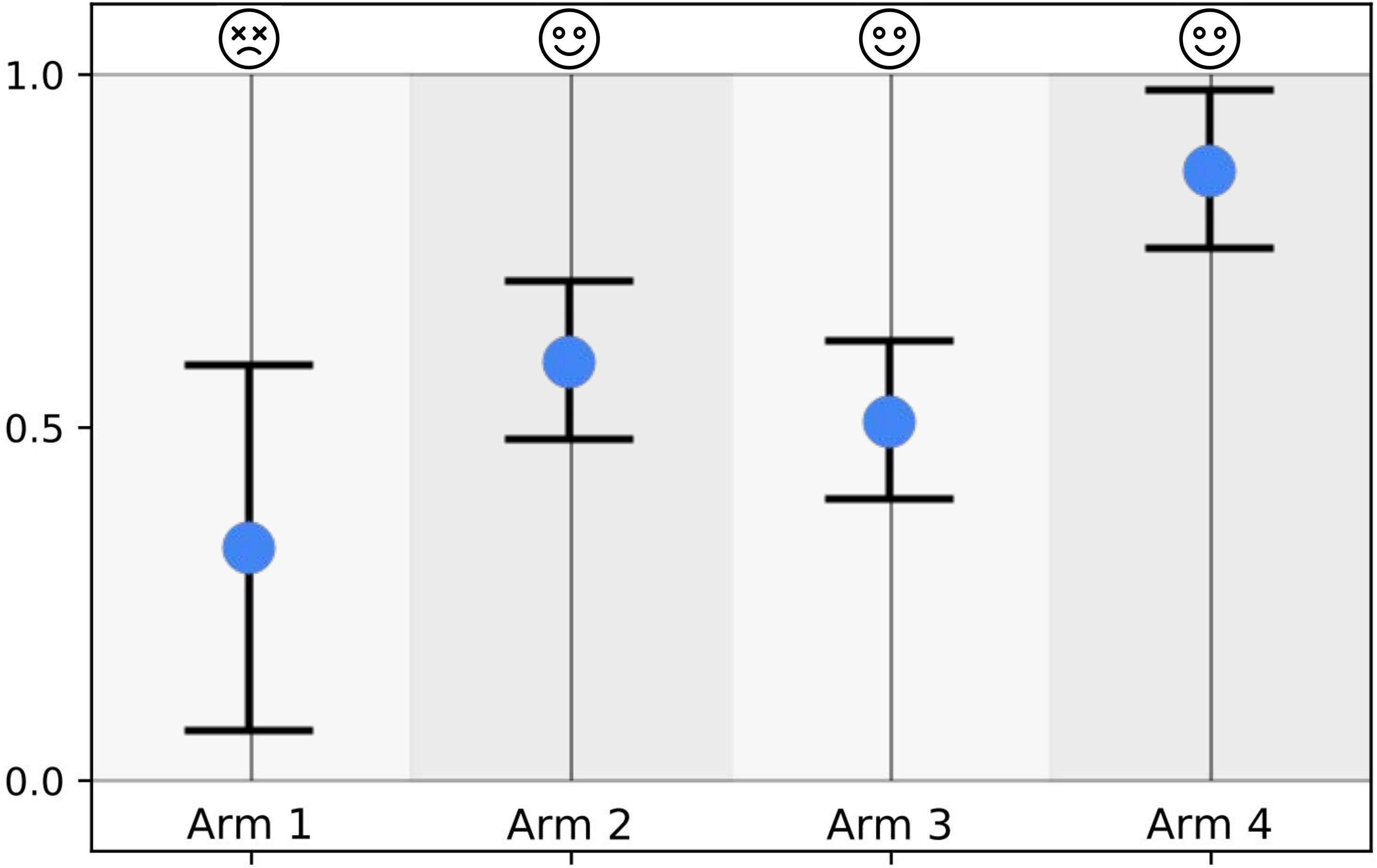
Delete i if
$\mathrm{UCB}_i \leq \max_j \mathrm{LCB}_j$


1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
Arm 4

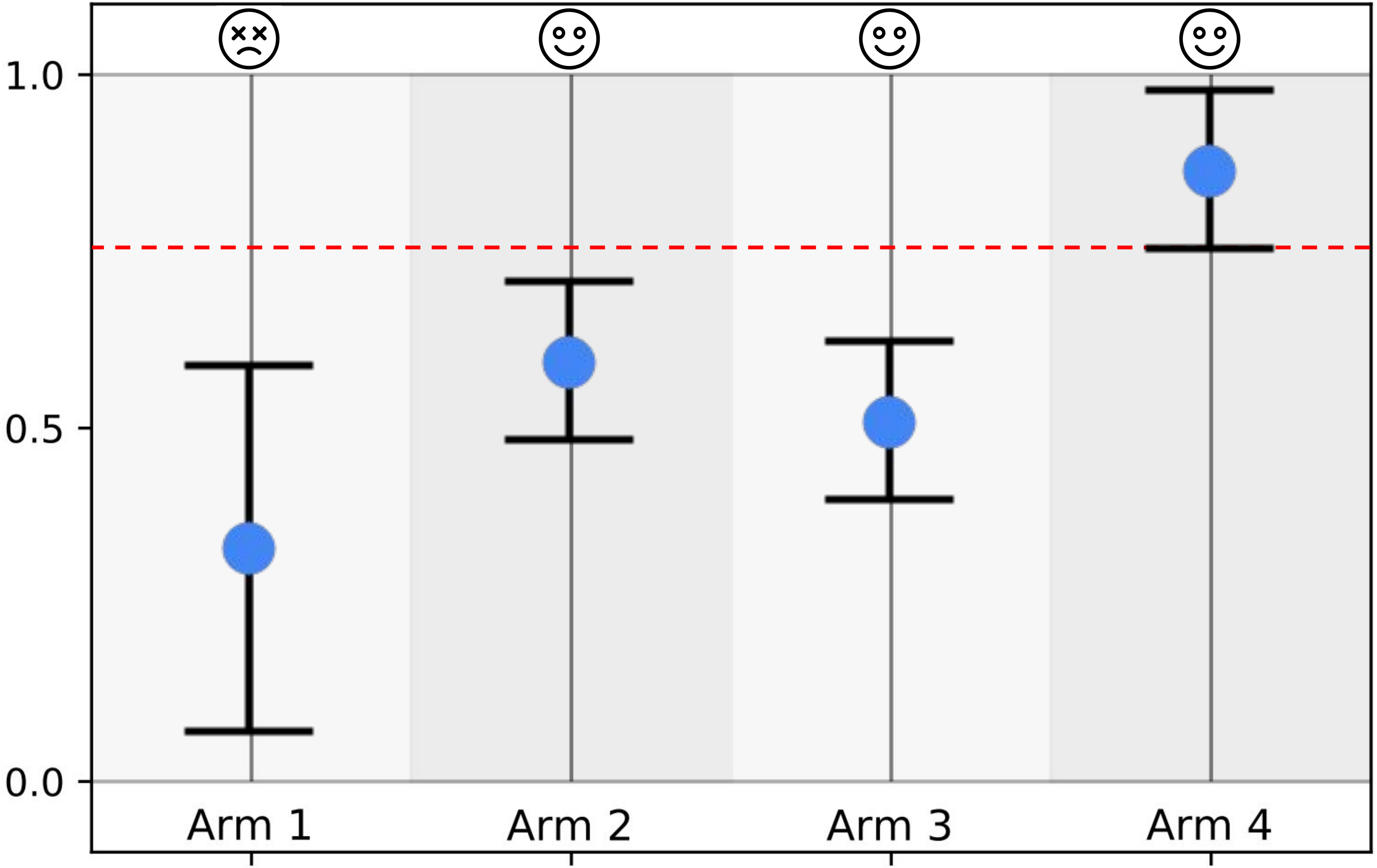


Sample active arms to τ_2

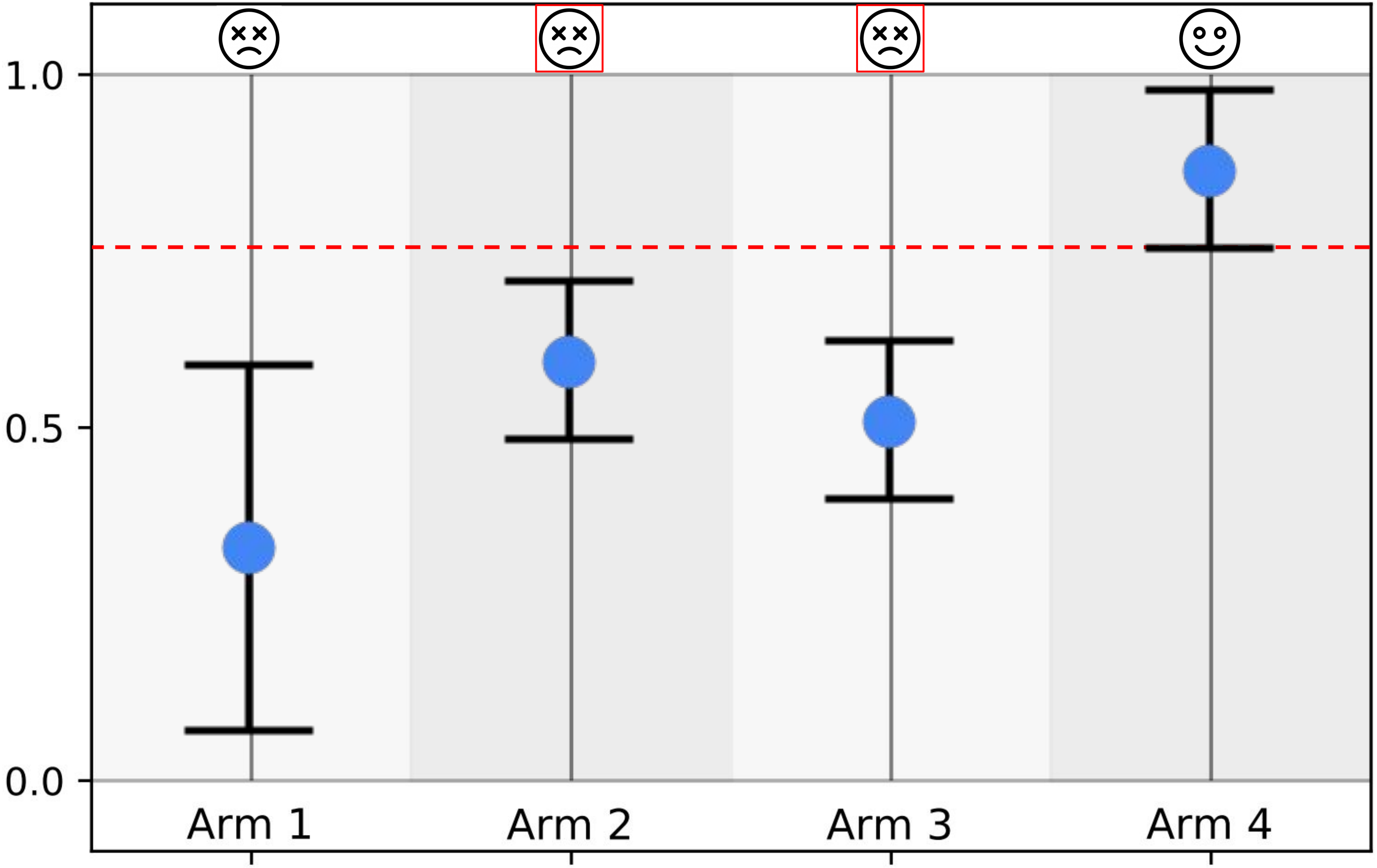

1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
Arm 4



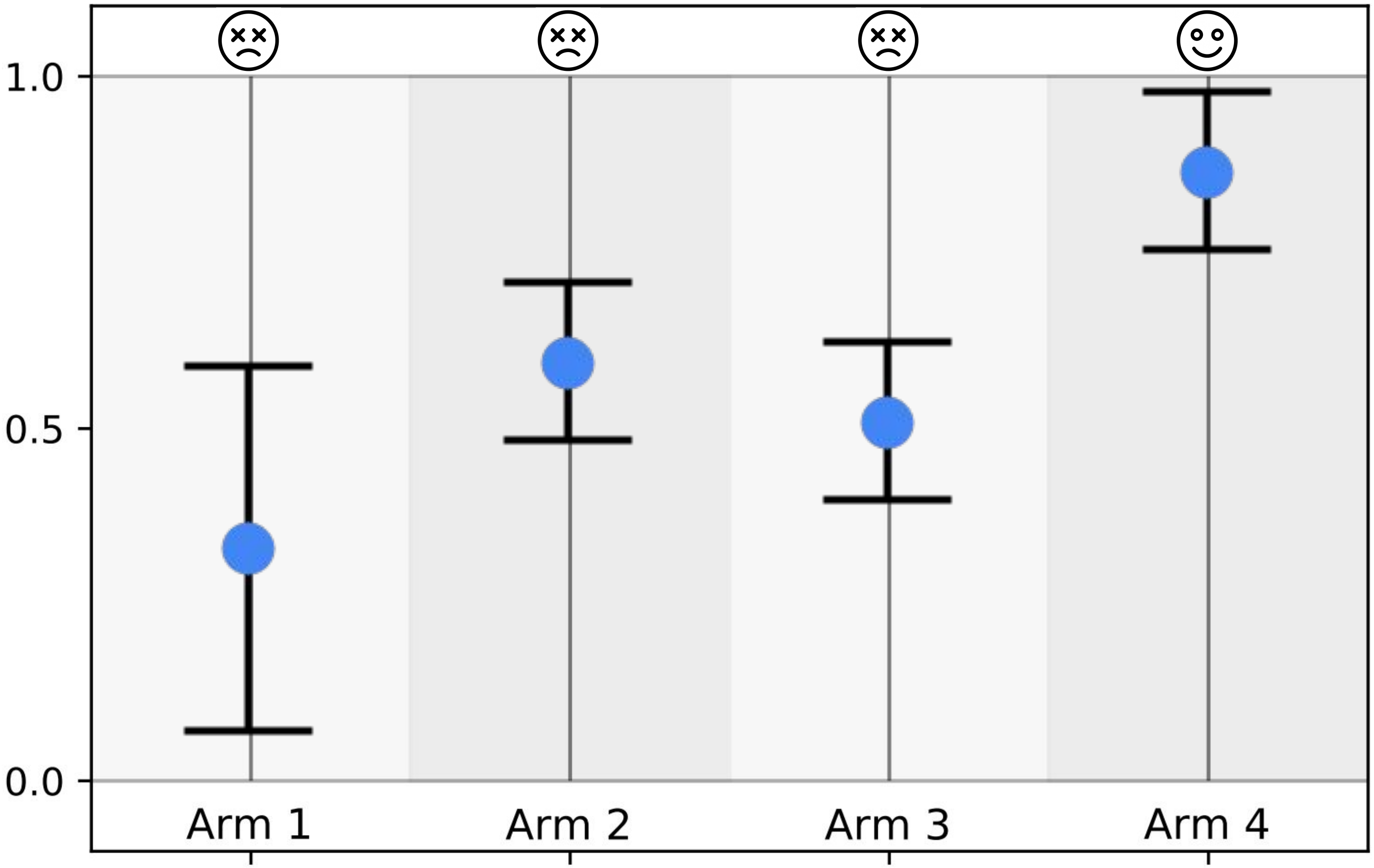
Delete i if
$\mathrm{UCB}_i \leq \max_j \mathrm{LCB}_j$

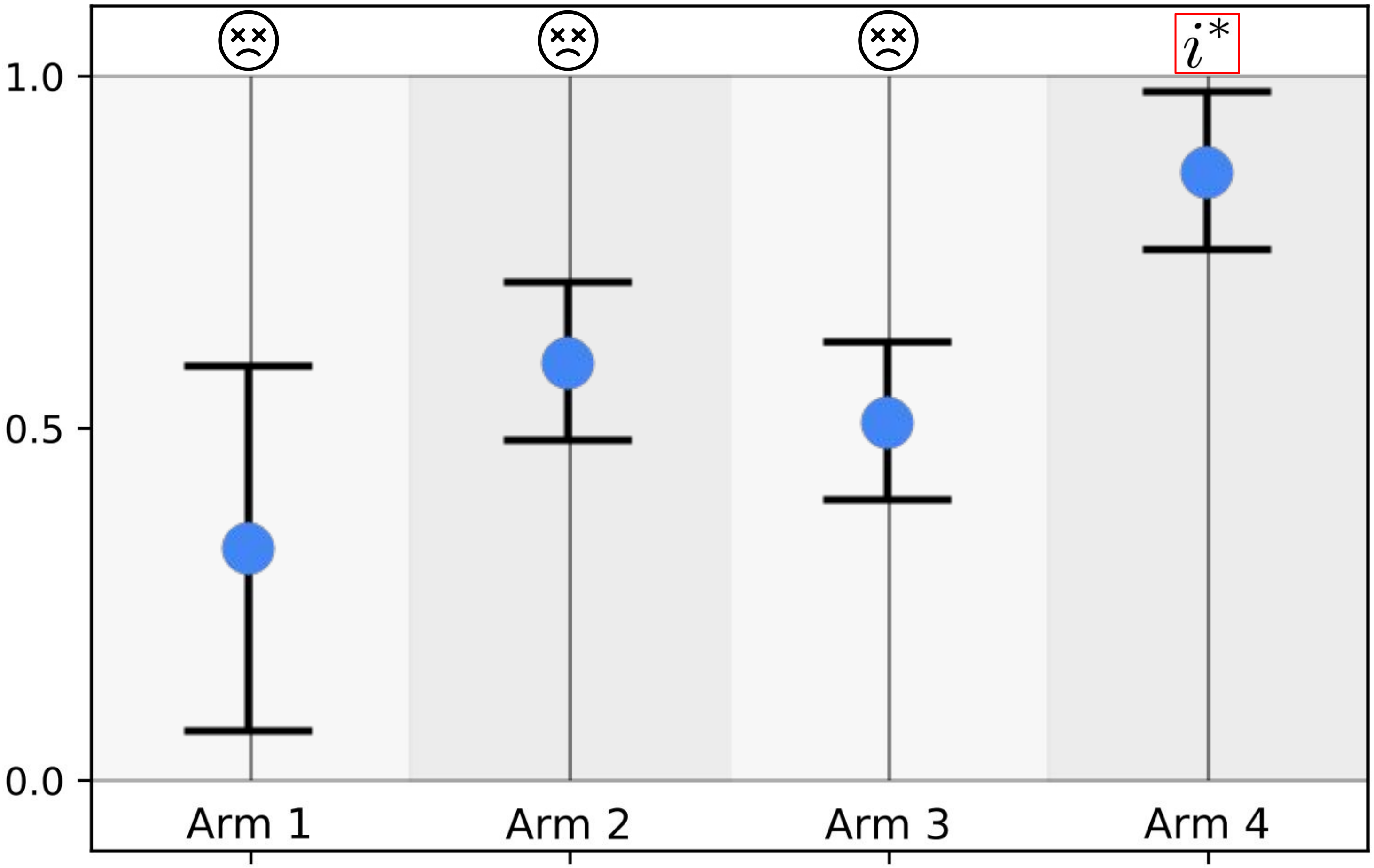


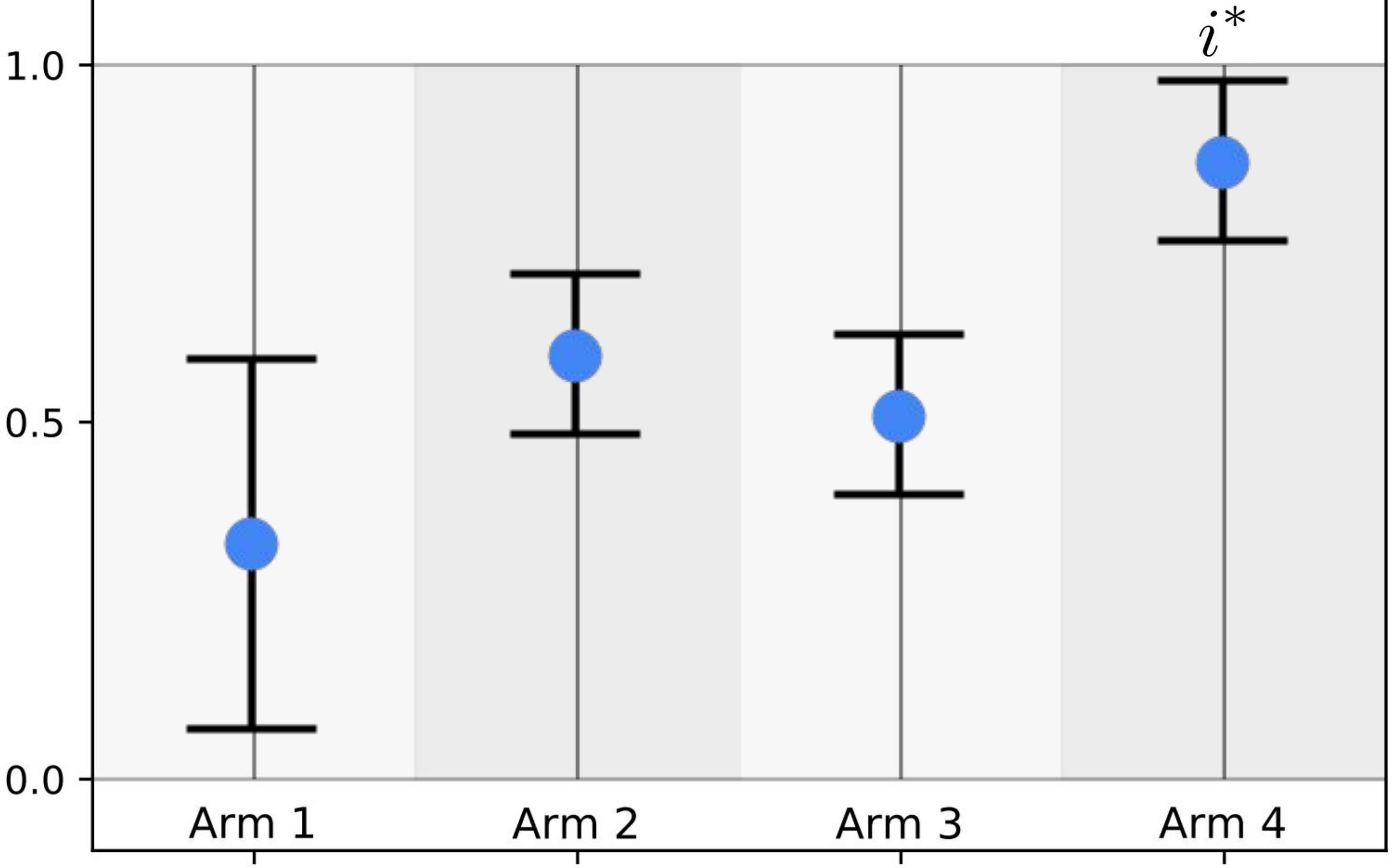
Delete i if
$\mathrm{UCB}_i \leq \max_j \mathrm{LCB}_j$

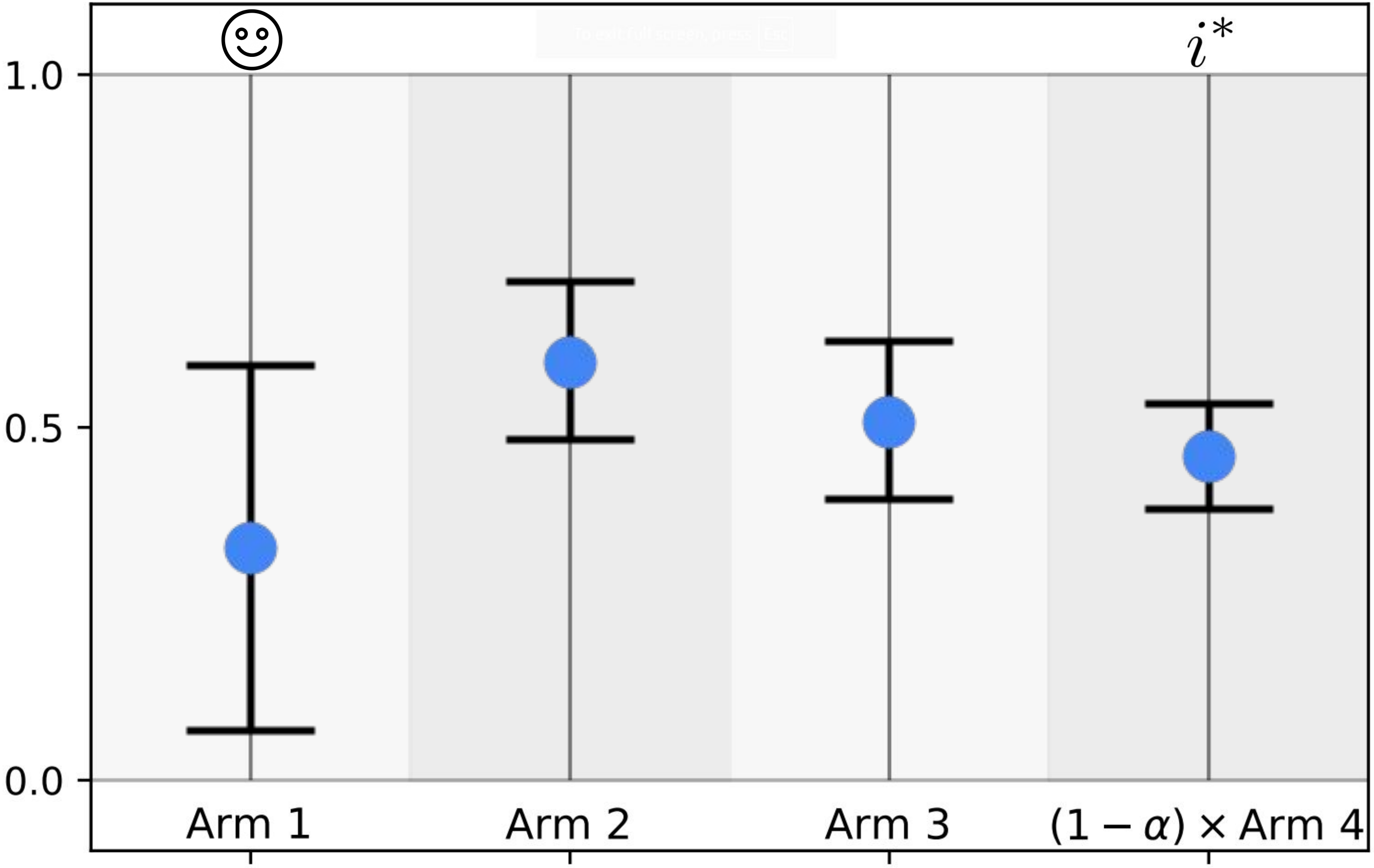


Delete i if
$\mathrm{UCB}_i \leq \max_j \mathrm{LCB}_j$


1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
Arm 4


i^*
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
Arm 4


i^*
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
Arm 4


i^*
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4

Outline

- Contributions and Related Work
- MAB-CS objective
- PE and PE-CS Algorithms for MAB-CS
- **Experiments**
- Theoretical Results
    - Upper Bounds
    - Lower Bounds

MovieLens 25M dataset

*Maxwell and Konstan. The movielens datasets: History and context. Acm transactions on interactive intelligent systems (tiis) (2015)

MovieLens 25M dataset

- 25 million ratings for $62,000$ movies rated by $162,000$ users

*Maxwell and Konstan. The movielens datasets: History and context. Acm transactions on interactive intelligent systems (tiis) (2015)

MovieLens 25M dataset

- 25 million ratings for $62,000$ movies rated by $162,000$ users
- Each movie tagged with one or more genres

*Maxwell and Konstan. The movielens datasets: History and context. Acm transactions on interactive intelligent systems (tiis) (2015)

MovieLens 25M dataset

- 25 million ratings for $62,000$ movies rated by $162,000$ users
- Each movie tagged with one or more genres
- 20 such genres $\rightarrow$ bandit arms

*Maxwell and Konstan. The movielens datasets: History and context. Acm transactions on interactive intelligent systems (tiis) (2015)

MovieLens 25M dataset

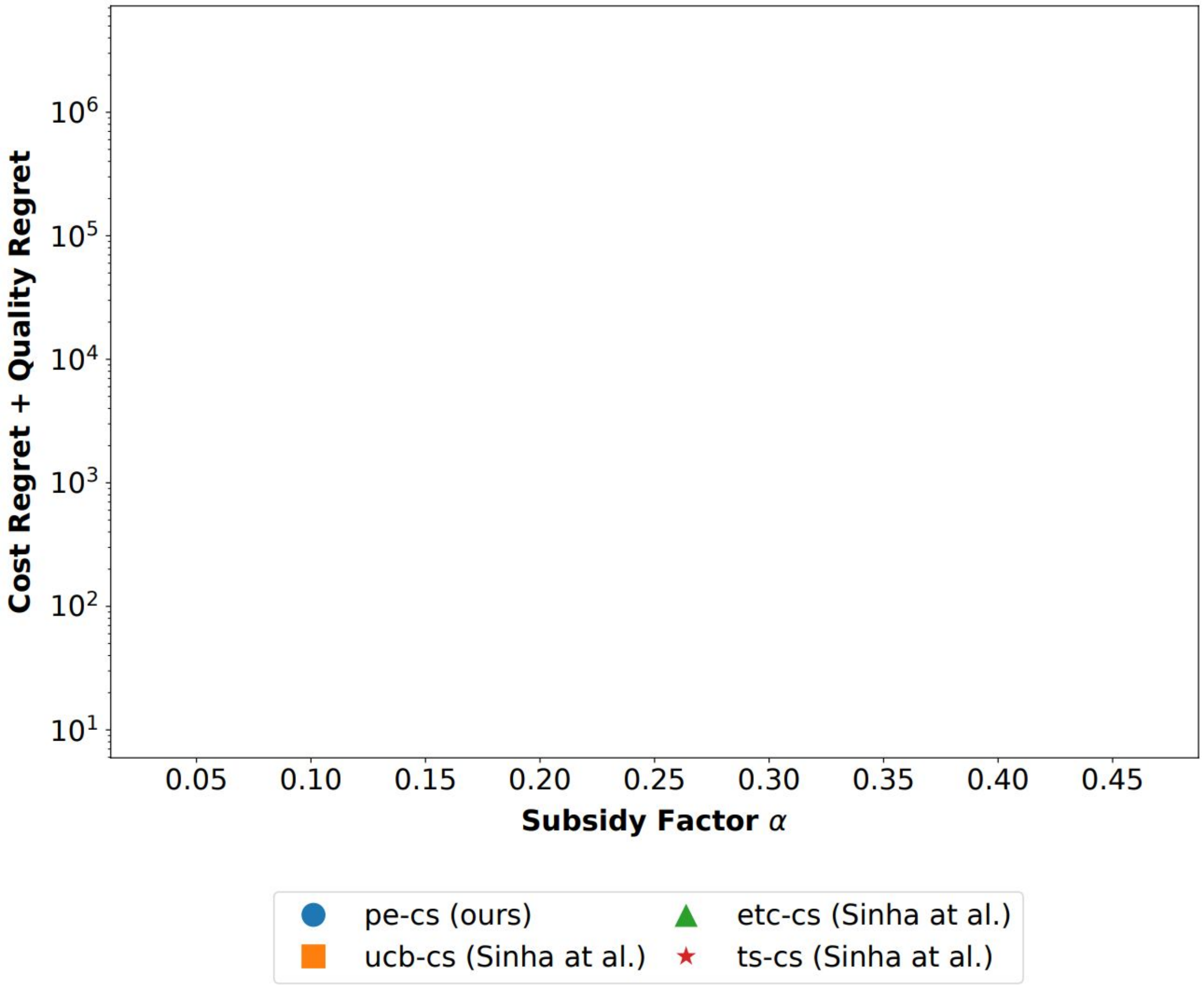
- 25 million ratings for $62,000$ movies rated by $162,000$ users
- Each movie tagged with one or more genres
- 20 such genres $\rightarrow$ bandit arms
- μ_i computed as average rating of genre i

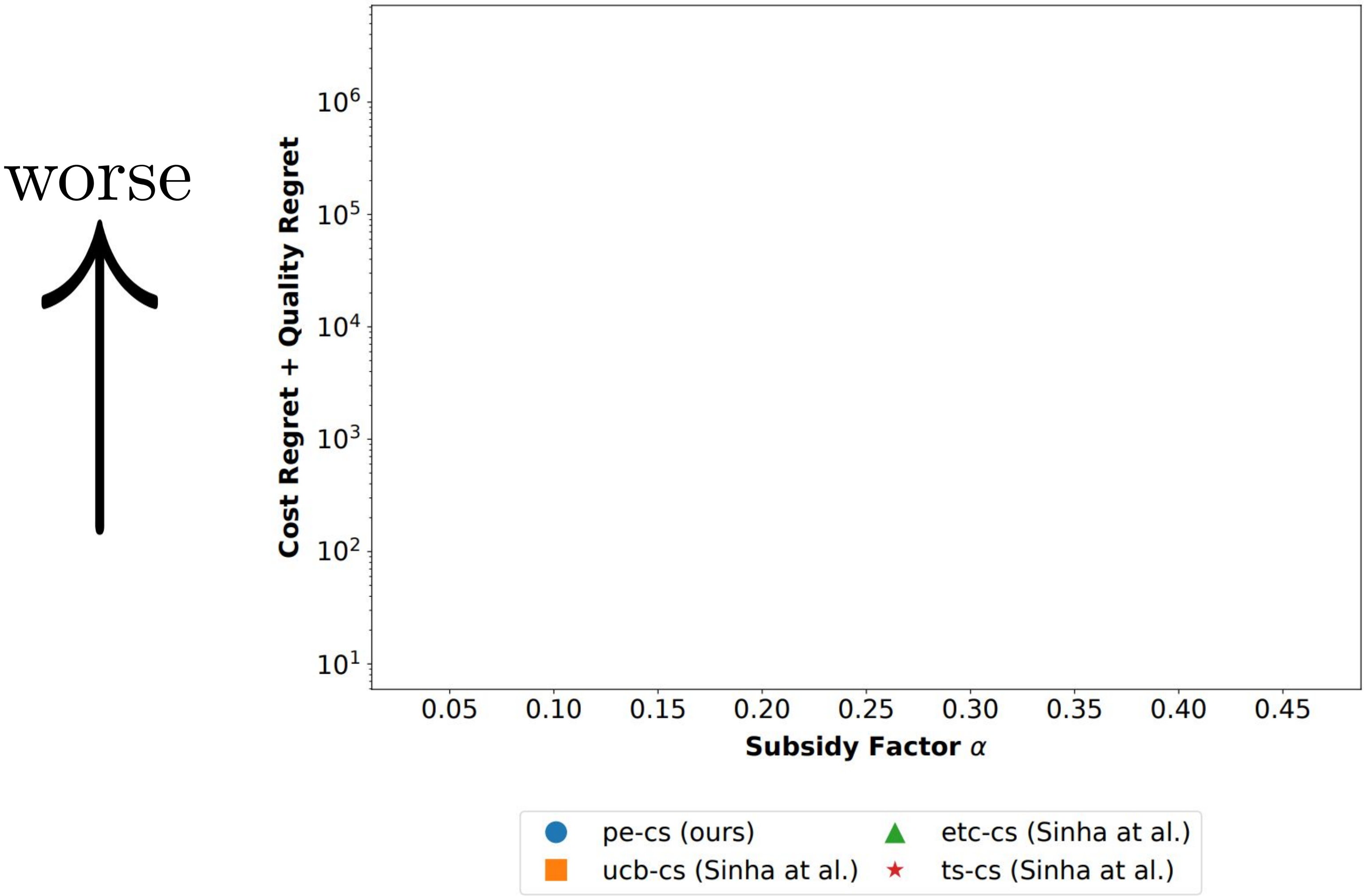
*Maxwell and Konstan. The movielens datasets: History and context. Acm transactions on interactive intelligent systems (tiis) (2015)

MovieLens 25M dataset

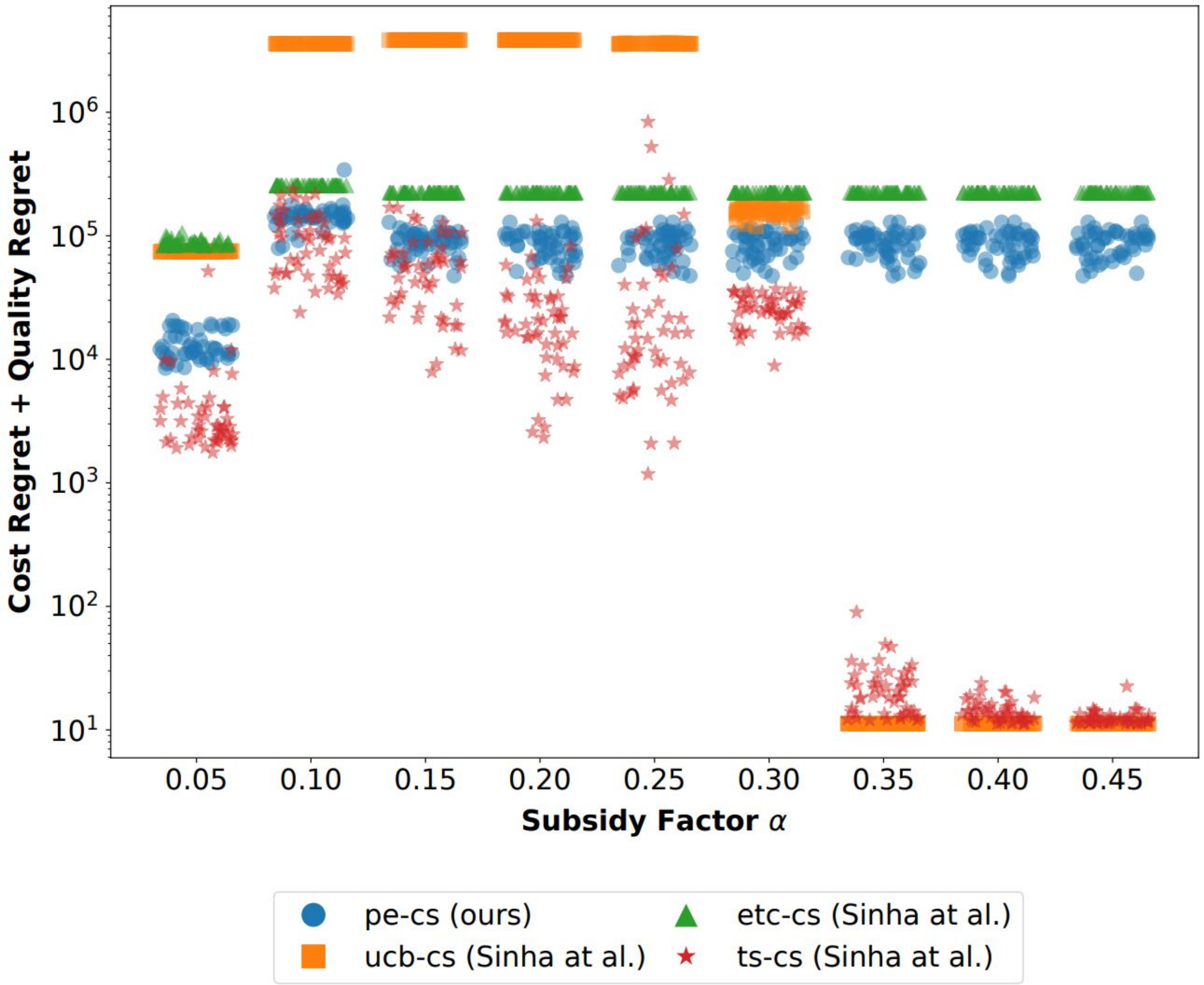
- 25 million ratings for $62,000$ movies rated by $162,000$ users
- Each movie tagged with one or more genres
- 20 such genres $\rightarrow$ bandit arms
- μ_i computed as average rating of genre i
- $c_i \sim \text{Unif}\,[0, 1]$ cost of sampling i

*Maxwell and Konstan. The movielens datasets: History and context. Acm transactions on interactive intelligent systems (tiis) (2015)

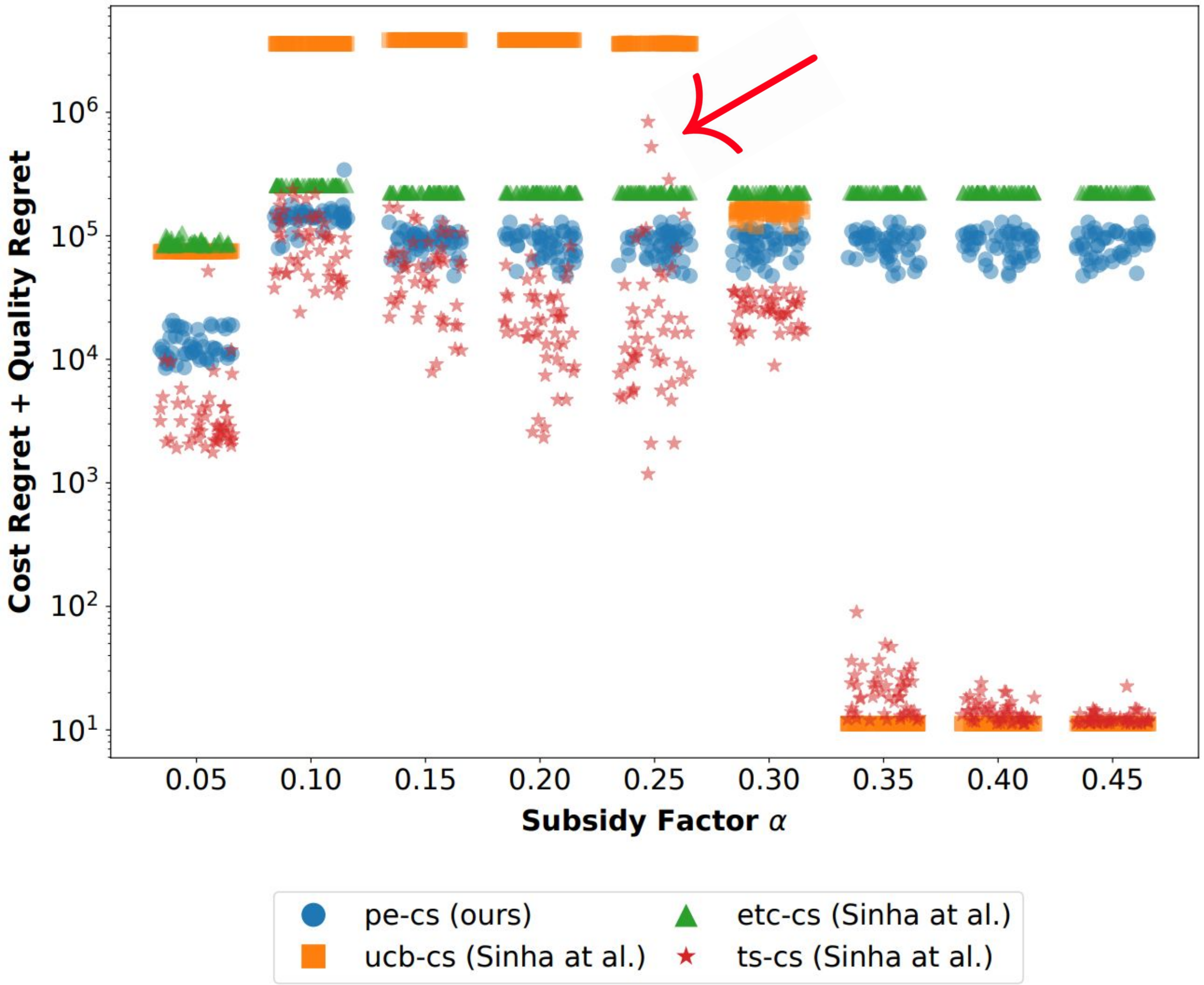

Cost Regret + Quality Regret
10^6
10^5
10^4
10^3
10^2
10^1
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
Subsidy Factor α
pe-cs (ours)
ucb-cs (Sinha at al.)
etc-cs (Sinha at al.)
ts-cs (Sinha at al.)



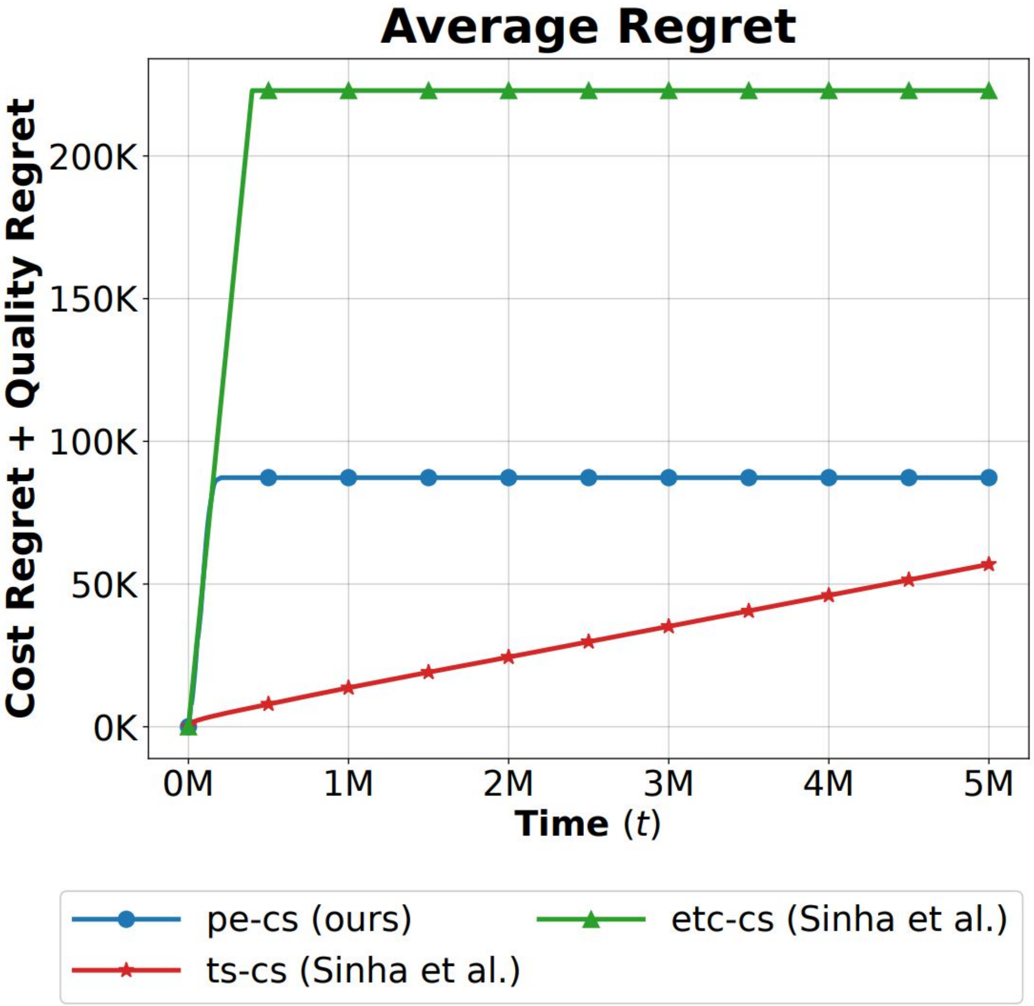
worse
Cost Regret + Quality Regret
10^6
10^5
10^4
10^3
10^2
10^1
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
Subsidy Factor α
pe-cs (ours)
etc-cs (Sinha at al.)
ucb-cs (Sinha at al.)
ts-cs (Sinha at al.)



Cost Regret + Quality Regret
10^1
10^2
10^3
10^4
10^5
10^6
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
Subsidy Factor α
pe-cs (ours)
ucb-cs (Sinha at al.)
etc-cs (Sinha at al.)
ts-cs (Sinha at al.)



Cost Regret + Quality Regret
10^1
10^2
10^3
10^4
10^5
10^6
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
Subsidy Factor α
pe-cs (ours)
ucb-cs (Sinha at al.)
etc-cs (Sinha at al.)
ts-cs (Sinha at al.)



Average Regret
Cost Regret + Quality Regret
200K
150K
100K
50K
0K
0M
1M
2M
3M
4M
5M
Time (t)
pe-cs (ours)
etc-cs (Sinha et al.)
ts-cs (Sinha et al.)

Outline

- Contributions and Related Work
- MAB-CS objective
- PE and PE-CS Algorithms for MAB-CS
- Experiments
- **Theoretical Results**
  - Upper Bounds
  - Lower Bounds

Outline

- Contributions and Related Work
- MAB-CS objective
- PE and PE-CS Algorithms for MAB-CS
- Experiments
- **Theoretical Results**
    - **Upper Bounds**
    - Lower Bounds

$$\text{Cost\_Reg}(T, \nu) = \sum_{i=1}^{K} \Delta_{C,i}^{+} \mathbb{E}\left[n_i(T)\right]$$

$$\text{Cost\_Reg}(T, \nu) = \sum_{i=1}^{K} \underbrace{\Delta_{C,i}^{+}}_{\substack{\text{Known} \\ \text{given } \nu}} \underbrace{\mathbb{E}[n_i(T)]}_{\substack{\text{Develop} \\ \text{Bound}}}$$

$$\text{Cost\_Reg}(T,\nu) = \sum_{i=1}^{K} \underbrace{\Delta^{+}_{C,i}}_{\substack{\text{Known} \\ \text{given } \nu}} \underbrace{\mathbb{E}[n_i(T)]}_{\substack{\text{Develop} \\ \text{Bound}}}$$

$$\text{Cost\_Reg}(T,\nu) = \sum_{i=1}^{K} \underbrace{\Delta_{C,i}^{+}}_{\substack{\text{Known} \\ \text{given } \nu}} \underbrace{\mathbb{E}[n_i(T)]}_{\substack{\text{Develop} \\ \text{Bound}}}$$

Arm category	Condition on event

Arm category	Condition on event
$i < a^*$	Arm i being eliminated by a specified $n_i(t), n_\ell(t)$

Arm category	Condition on event
$i < a^*$	Arm i being eliminated by a specified $n_i(t), n_\ell(t)$
$i > a^*, i \neq \ell$	PE terminating in episode a^*

Arm category	Condition on event
$i < a^*$	Arm i being eliminated by a specified $n_i(t), n_\ell(t)$
$i > a^*, i \neq \ell$	PE terminating in episode a^*
$i = \ell$	PE terminating correctly in episode a^*

$$\text{Cost\_Reg}(T, \nu) \leq$$

$$\mathrm{Cost\_Reg}(T,\nu) \leq \underbrace{\sum_{i=1}^{a^*-1}\left[C_1(\nu)\frac{\log T}{\Delta_{Q,i}^2}\right]\cancelto{0}{\Delta_{C,i}^{+}}}_{\substack{\text{Contribution from } i<a^* \text{ arms} \\ \text{under nominal termination}}} +$$

$$\mathrm{Cost\_Reg}(T,\nu) \le \underbrace{\sum_{i=1}^{a^*-1}\left[C_1(\nu)\frac{\log T}{\Delta_{Q,i}^2}\right]\cancelto{0}{\Delta_{C,i}^+}}_{\substack{\text{Contribution from } i < a^* \text{ arms} \\ \text{under nominal termination}}} + \underbrace{\sum_{i=1}^{a^*-1}\left[C_2(\nu)\right]\cancelto{0}{\Delta_{C,i}^+}}_{\substack{\text{Contribution from } i < a^* \text{ arms} \\ \text{under incorrect termination}}}$$

$$\mathrm{Cost\_Reg}(T,\nu) \leq \underbrace{\sum_{i=1}^{a^*-1} \left[C_1(\nu)\frac{\log T}{\Delta_{Q,i}^2}\right] \cancelto{0}{\Delta_{C,i}^+}}_{\substack{\text{Contribution from } i<a^* \text{ arms} \\ \text{under nominal termination}}} + \underbrace{\sum_{i=1}^{a^*-1} [C_2(\nu)] \cancelto{0}{\Delta_{C,i}^+}}_{\substack{\text{Contribution from } i<a^* \text{ arms} \\ \text{under incorrect termination}}}$$

$$+ \underbrace{C_3(\nu) \max_{i>a^*} \Delta_{C,i}^+}_{\substack{\text{Contribution from } i>a^* \text{ arms} \\ \text{under incorrect termination}}}$$

$$
\begin{aligned}
\mathrm{Cost\_Reg}(T,\nu) \leq\; & \underbrace{\sum_{i=1}^{a^*-1} \left[C_1(\nu) \frac{\log T}{\Delta_{Q,i}^2} \right] \cancelto{0}{\Delta_{C,i}^+}}_{\substack{\text{Contribution from } i < a^* \text{ arms} \\ \text{under nominal termination}}} + \underbrace{\sum_{i=1}^{a^*-1} \left[C_2(\nu) \right] \cancelto{0}{\Delta_{C,i}^+}}_{\substack{\text{Contribution from } i < a^* \text{ arms} \\ \text{under incorrect termination}}} \\
& + \underbrace{C_3(\nu) \max_{i > a^*} \Delta_{C,i}^+}_{\substack{\text{Contribution from } i > a^* \text{ arms} \\ \text{under incorrect termination}}} \\
& + \underbrace{\left[C_1(\nu) \max_{i \leq a^*} \frac{\log T}{\Delta_{Q,i}^2} \right] \Delta_{C,\ell}^+}_{\substack{\text{Contribution from arm } \ell \text{ under} \\ \text{nominal termination in PE episode } a^*}}
\end{aligned}
$$

$$\mathrm{Cost\_Reg}(T,\nu) \leq \underbrace{\sum_{i=1}^{a^*-1}\left[C_1(\nu)\frac{\log T}{\Delta_{Q,i}^2}\right]\cancelto{0}{\Delta_{C,i}^+}}_{\substack{\text{Contribution from } i<a^* \text{ arms}\\ \text{under nominal termination}}} + \underbrace{\sum_{i=1}^{a^*-1}\left[C_2(\nu)\right]\cancelto{0}{\Delta_{C,i}^+}}_{\substack{\text{Contribution from } i<a^* \text{ arms}\\ \text{under incorrect termination}}}$$

$$+ \quad \underbrace{C_3(\nu)\ \max_{i>a^*}\Delta_{C,i}^+}_{\substack{\text{Contribution from } i>a^* \text{ arms}\\ \text{under incorrect termination}}}$$

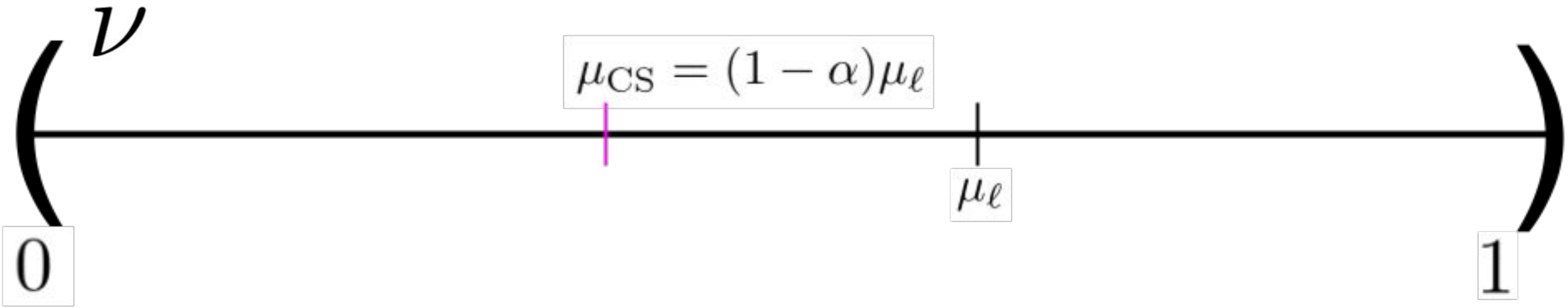
$$+ \quad \underbrace{\left[C_1(\nu)\ \max_{i\leq a^*}\frac{\log T}{\Delta_{Q,i}^2}\right]\Delta_{C,\ell}^+}_{\substack{\text{Contribution from arm } \ell \text{ under}\\ \text{nominal termination in PE episode } a^*}} \quad + \quad \underbrace{\sum_{i=1}^{a^*}\left[C_2(\nu)\right]\Delta_{C,\ell}^+}_{\substack{\text{Contribution from arm } \ell\\ \text{under incorrect termination}}}$$

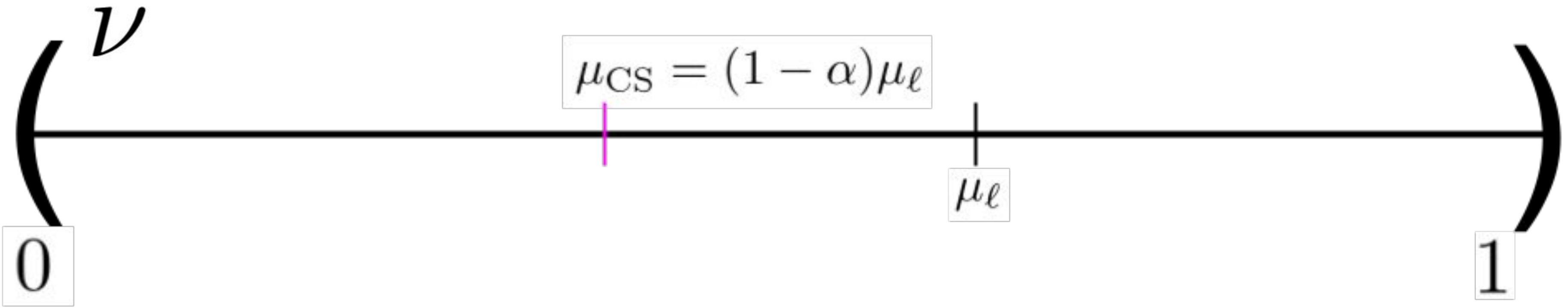
Outline

- Multi-Armed Bandits
  - Conventional, Cost Subsidy
- Related Work
- PE and and PE-CS Algorithms for MAB-CS
- Experiments
- **Theoretical Results**
  - Upper Bounds
  - **Lower Bounds**

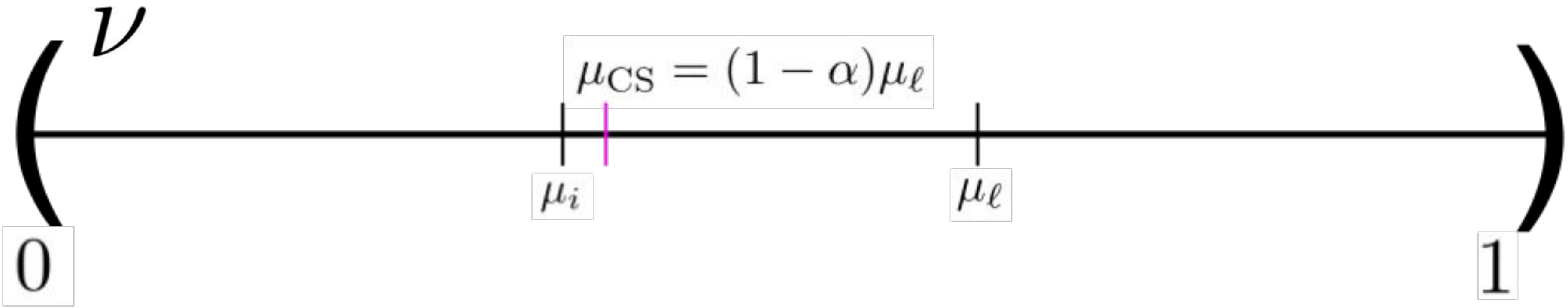

ν
μ_{CS}
0
1


ν
μ_{CS}
μ_ℓ
0
1

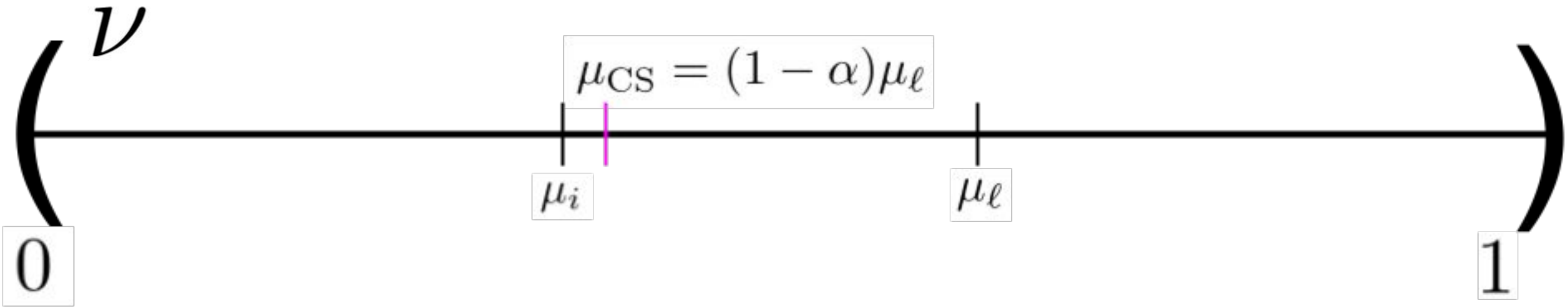

ν
$\mu_{\mathrm{CS}} = (1-\alpha)\mu_\ell$
μ_ℓ
0
1

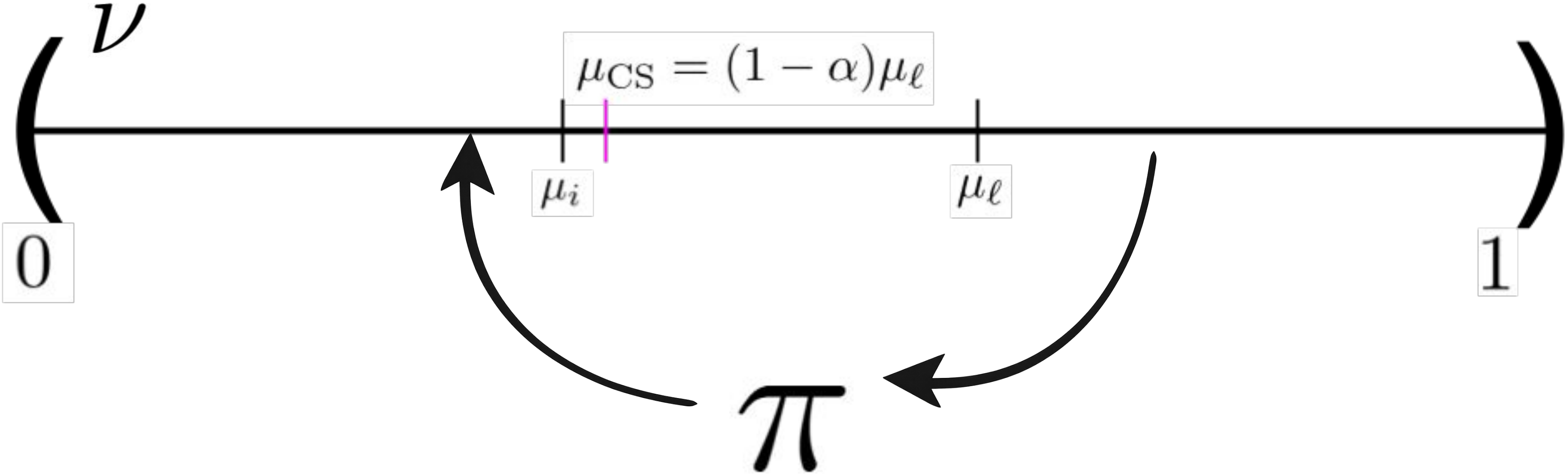


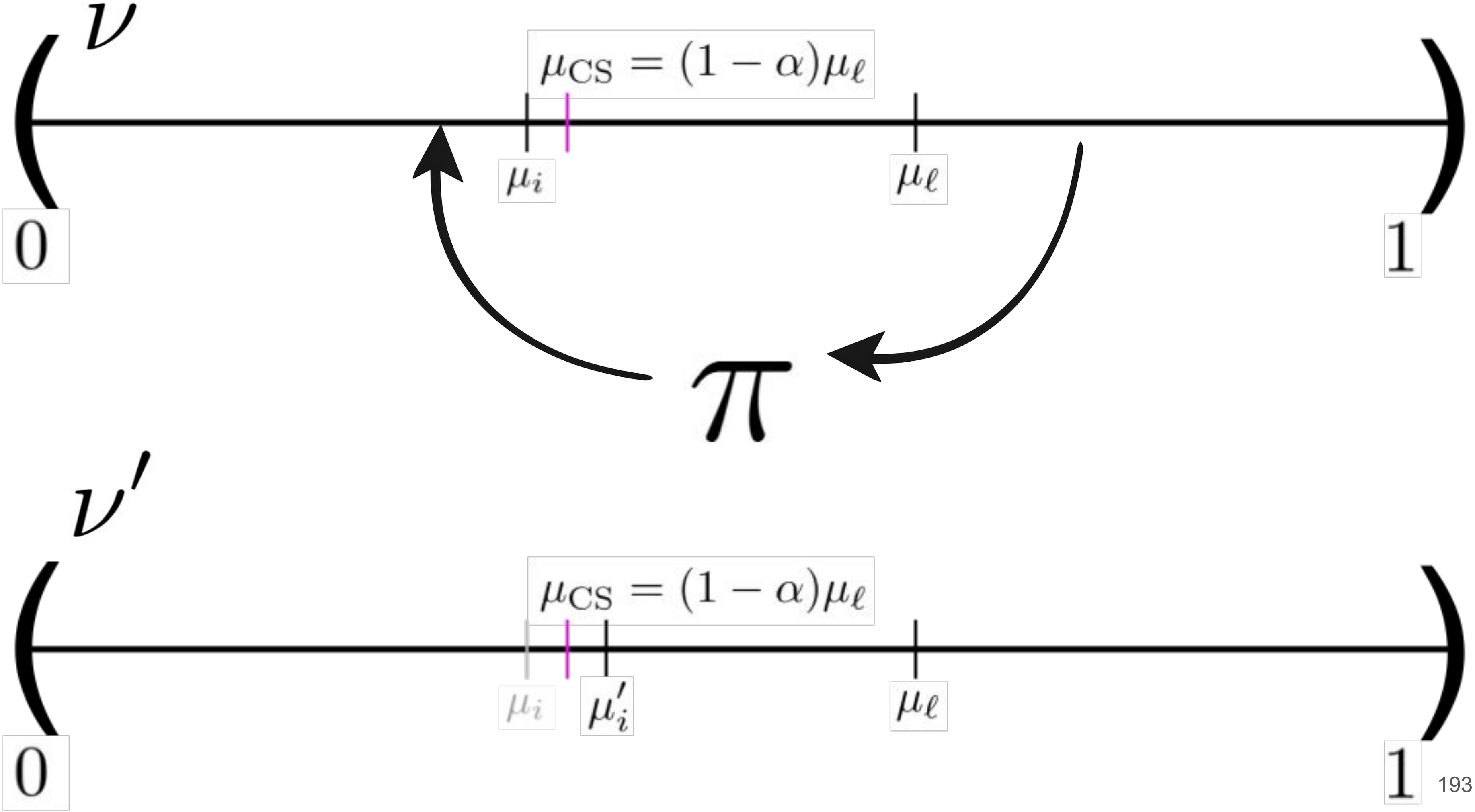
Consider arm i: $c_i < c_{a^*}$

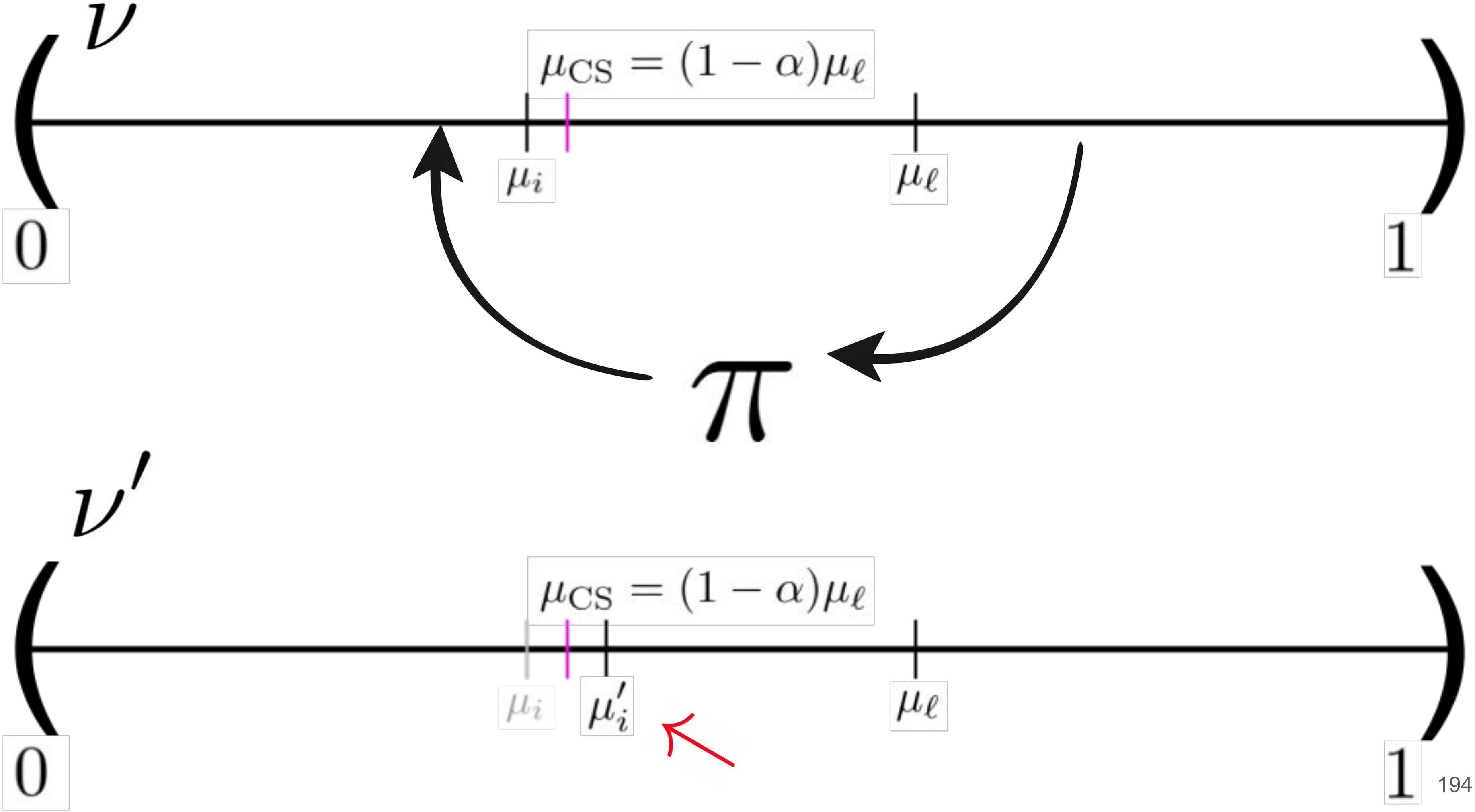


Consider arm i: $c_i < c_{a^*}$

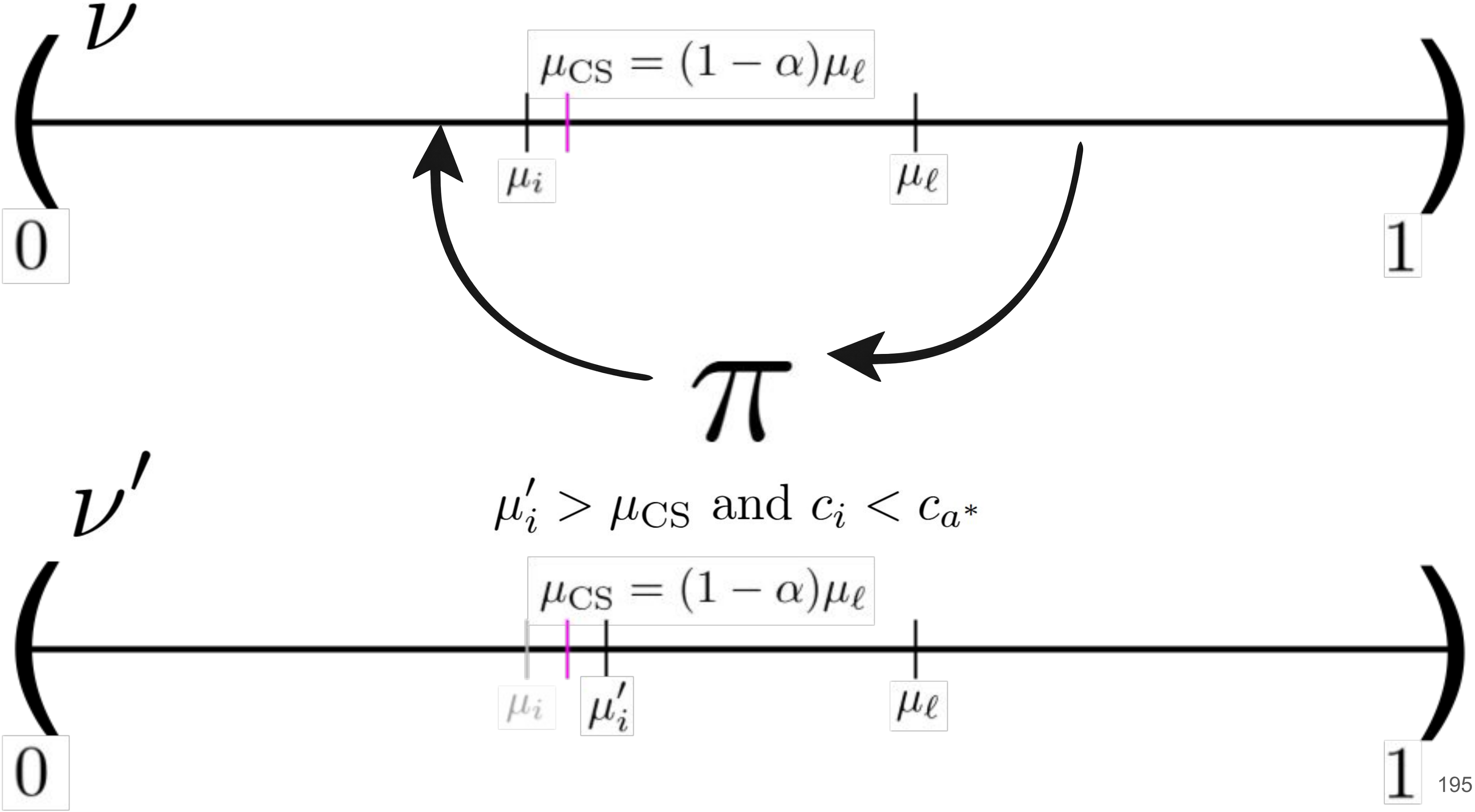

ν
$\mu_{\mathrm{CS}} = (1-\alpha)\mu_\ell$
μ_i
μ_ℓ
0
1

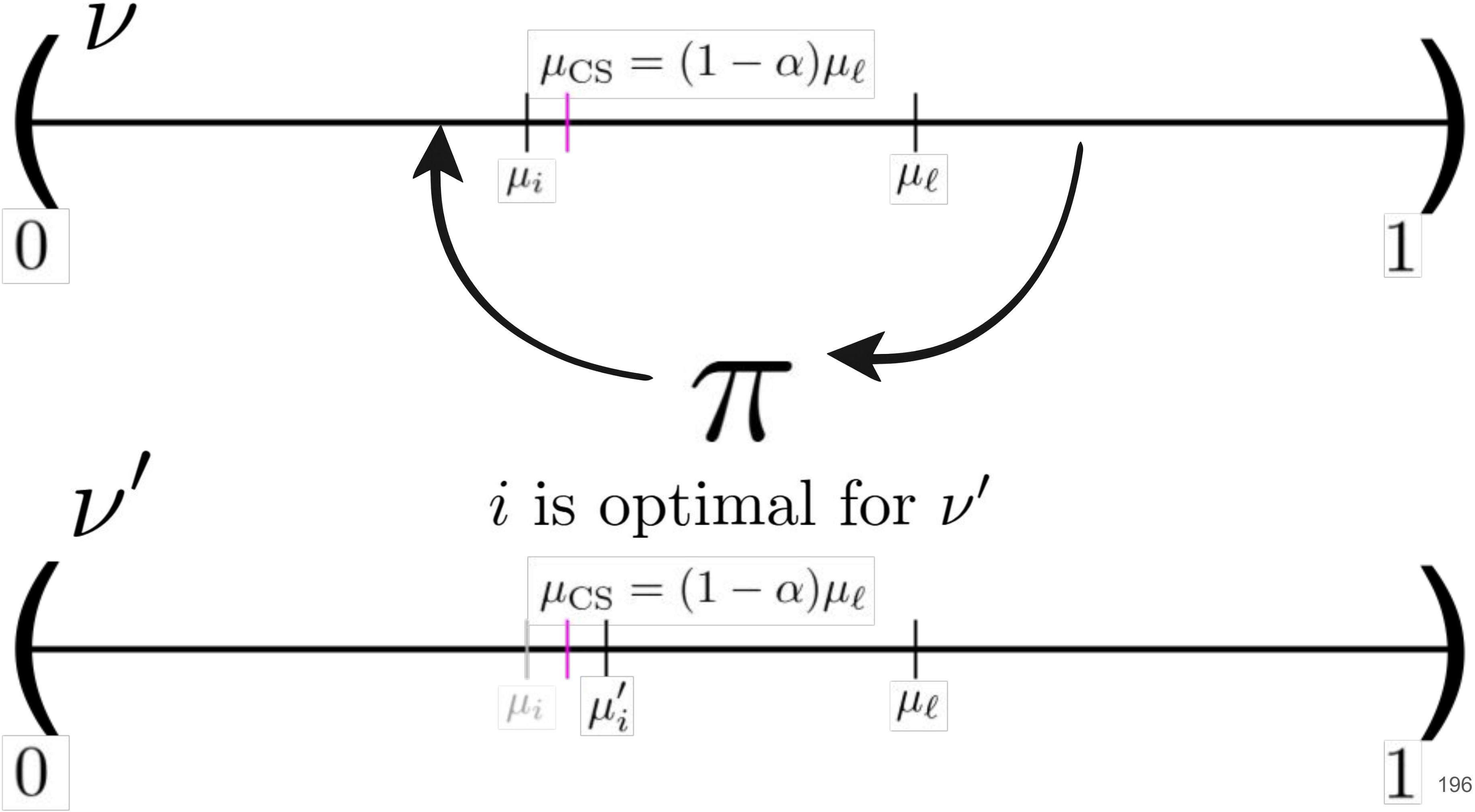

ν
$\mu_{\mathrm{CS}} = (1-\alpha)\mu_\ell$
μ_i
μ_ℓ
0
1
π


ν
$\mu_{\mathrm{CS}} = (1-\alpha)\mu_\ell$
μ_i
μ_ℓ
0
1
π
ν'
$\mu_{\mathrm{CS}} = (1-\alpha)\mu_\ell$
μ_i
μ_i'
μ_ℓ
0
1

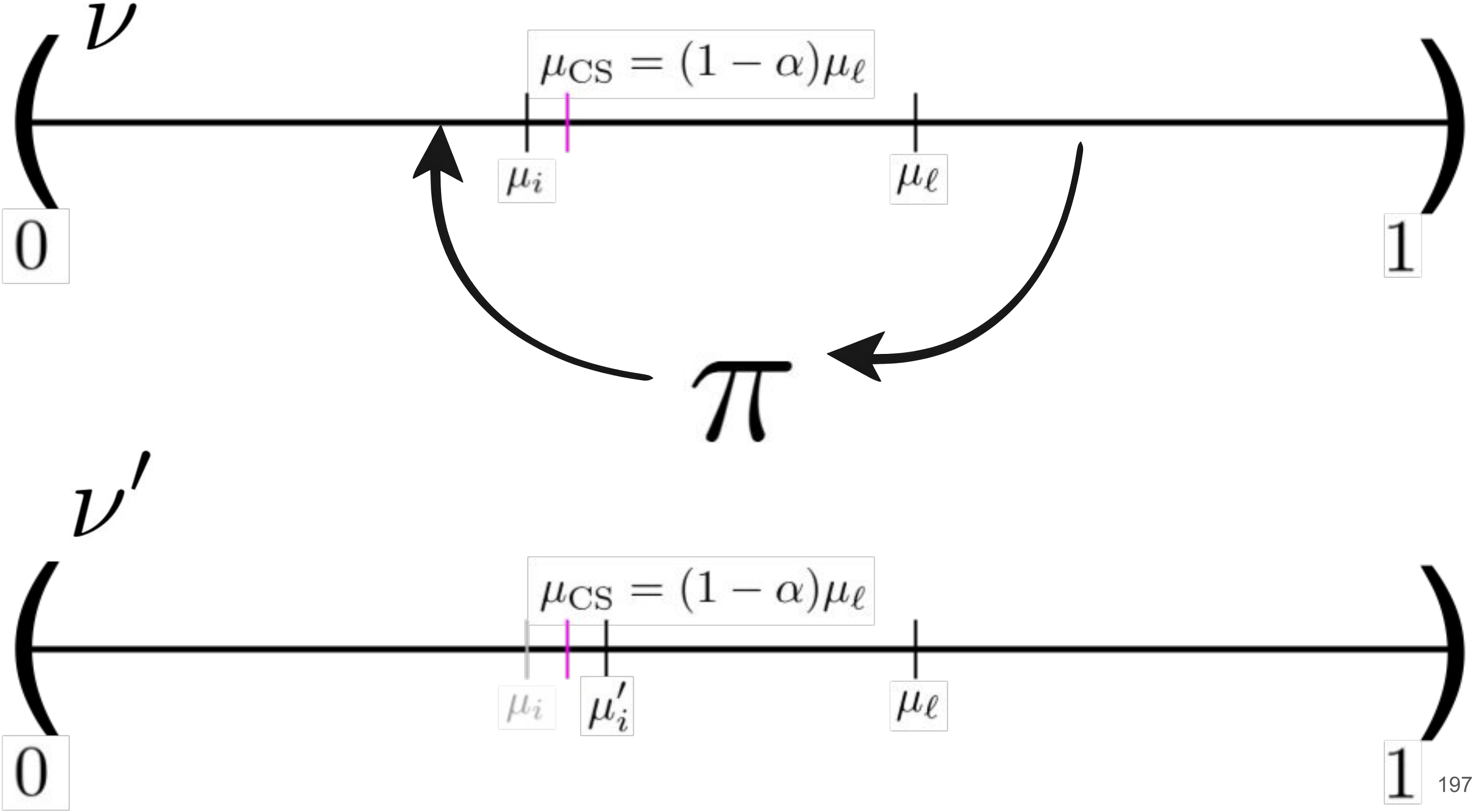


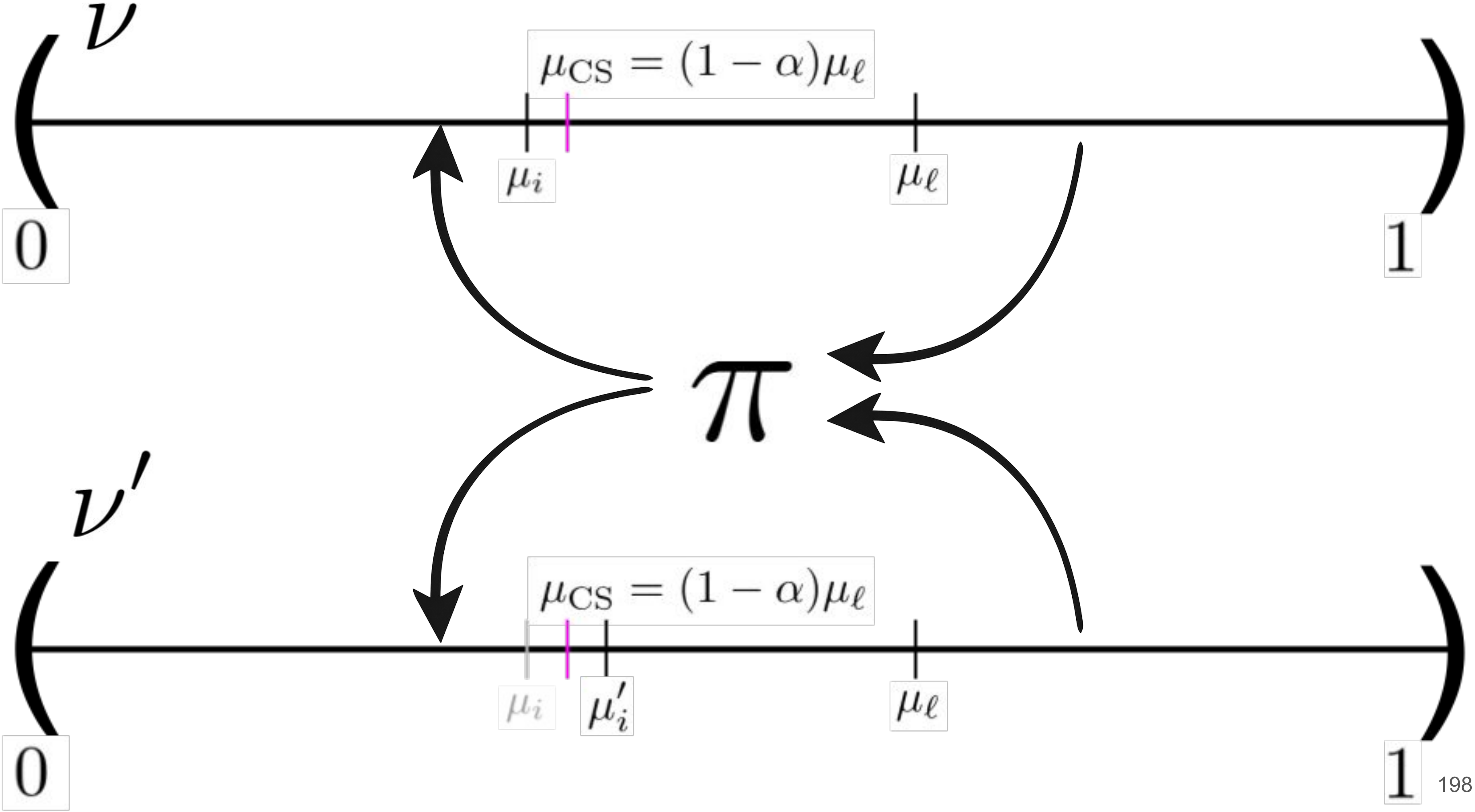
ν
$\mu_{\mathrm{CS}} = (1-\alpha)\mu_\ell$
μ_i
μ_ℓ
0
1
π
ν'
$\mu_{\mathrm{CS}} = (1-\alpha)\mu_\ell$
μ_i
μ_i'
μ_ℓ
0
1


ν
$\mu_{CS} = (1-\alpha)\mu_\ell$
μ_i
μ_ℓ
0
1
π
$\mu_i' > \mu_{CS}$ and $c_i < c_{a^*}$
ν'
$\mu_{CS} = (1-\alpha)\mu_\ell$
μ_i
μ_i'
μ_ℓ
0
1



ν
$\mu_{\mathrm{CS}} = (1-\alpha)\mu_\ell$
μ_i
μ_ℓ
0
1
π
i is optimal for ν'
ν'
$\mu_{\mathrm{CS}} = (1-\alpha)\mu_\ell$
μ_i
μ_i'
μ_ℓ
0
1


ν
$\mu_{\mathrm{CS}} = (1-\alpha)\mu_\ell$
μ_i
μ_ℓ
0
1
π
ν'
$\mu_{\mathrm{CS}} = (1-\alpha)\mu_\ell$
μ_i
μ_i'
μ_ℓ
0
1

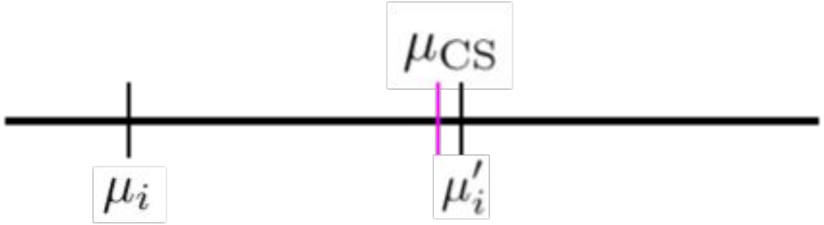

ν
$\mu_{\mathrm{CS}} = (1-\alpha)\mu_\ell$
μ_i
μ_ℓ
0
1
π
ν'
$\mu_{\mathrm{CS}} = (1-\alpha)\mu_\ell$
μ_i
μ_i'
μ_ℓ
0
1

$$\liminf_{T \to \infty} \frac{\mathbb{E}\left[n_i(T)\right]}{\log T} \geq \frac{1}{\mathrm{KL}(\nu_i, \nu_i')}$$

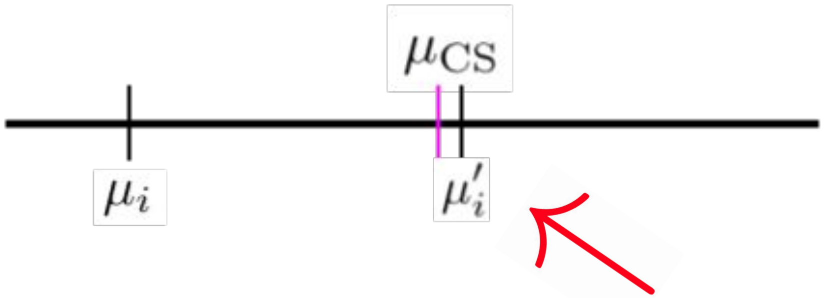
$$\liminf_{T\to\infty} \frac{\mathbb{E}\left[n_i(T)\right]}{\log T} \geq \frac{1}{\mathrm{KL}(\nu_i, \nu_i')}$$

$$\liminf_{T\to\infty} \frac{\mathbb{E}\left[n_i(T)\right]}{\log T} \geq \frac{1}{\mathrm{KL}(\nu_i, \nu_i')}$$

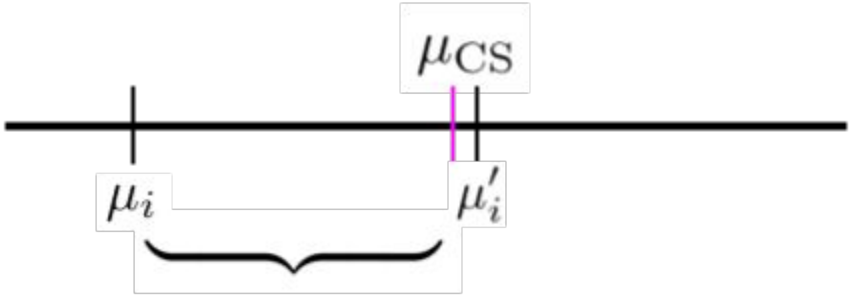
$$\liminf_{T\to\infty} \frac{\mathbb{E}\left[n_i(T)\right]}{\log T} \geq \frac{1}{\mathrm{KL}(\nu_i, \nu_i')}$$



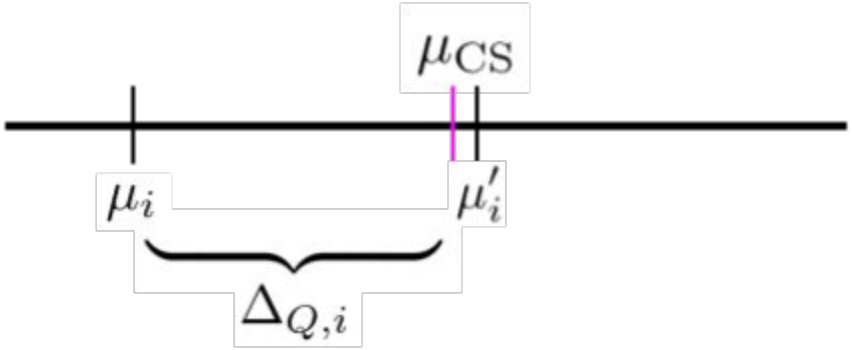
$$\liminf_{T\to\infty} \frac{\mathbb{E}\left[n_i(T)\right]}{\log T} \geq \frac{1}{\mathrm{KL}(\nu_i, \nu_i')}$$



$$\liminf_{T \to \infty} \frac{\mathbb{E}\left[n_i(T)\right]}{\log T} \geq \frac{1}{\mathrm{KL}(\nu_i, \nu_i')}$$



$$\liminf_{T\to\infty} \frac{\mathbb{E}\left[n_i(T)\right]}{\log T} \geq \frac{1}{\mathrm{KL}(\nu_i, \nu_i')}$$



$$\liminf_{T\to\infty} \frac{\mathbb{E}\left[n_i\left(T\right)\right]}{\log T} \geq \frac{2}{\Delta_{Q,i}^2}, \quad \textit{for arms} \;\; i < a^*$$

$$\liminf_{T\to\infty} \frac{\mathbb{E}\left[n_i\left(T\right)\right]}{\log T} \geq \frac{2}{\Delta_{Q,i}^2}, \quad \textit{for arms } \; i < a^*$$

Core Contributions and Conclusion

- Variants of MAB-CS
- Cost and quality regret minimization algorithms
  - PE
  - PE-CS
- Experimental Validation of PE and PE-CS
- Analysis of PE and PE-CS
- Instance specific lower bounds

QnA Segment

Extra Materials

Applications

Mention 3 applications including LLM Routing, Network Hardware allocation, and recommendation systems

Algorithm 1: Pairwise Elimination (PE) for a known reference arm ℓ

Inputs: Bandit Instance ν, Horizon T, Reference Arm ℓ, Subsidy Factor α .
Initialize: Samples $n_k = 0$, Empirical Means $\hat{\mu}_k = 0$, Current Rounds $\omega_k = 0$,
$\forall\, k \in [K] \cup \{\ell\}$, PE Episode $i = 1$, Time $t = 1$.

while $t \leq T$ **do**
  if $i \notin \{None, \ell\}$ **then**
    $k_t, \boldsymbol{\omega}, i \leftarrow \mathrm{PE}(\boldsymbol{n}, \hat{\boldsymbol{\mu}}, \boldsymbol{\omega}, T, i, \ell, \alpha)$ `// receive arm to be sampled, updated round numbers, and updated episode number`
  else
    $k_t \leftarrow k_{t-1}$ `// sample winning arm for remaining budget`
  $\hat{\boldsymbol{\mu}}(t+1), \boldsymbol{n}(t+1), t \leftarrow$ sample_and_update$(k_t, \hat{\boldsymbol{\mu}}(t), \boldsymbol{n}(t), t)$ `// in Appendix B`

Function 1: Pairwise Elimination Function PE()

Function *PE($\boldsymbol{n}$: Sample Vector, $\hat{\boldsymbol{\mu}}$: Empirical Means, $\boldsymbol{\omega}$: Round Numbers, T: Horizon, i: Episode, ℓ: Reference Arm, α: Subsidy Factor)*:

$\tilde{\Delta} \leftarrow 2^{-\omega_i}$

$\tau \leftarrow \left\lceil \frac{2\log(T\tilde{\Delta}^2)}{\tilde{\Delta}^2} \right\rceil$

for $k \in \{i, \ell\}$ **do**

  if $n_k < \tau$ **then**

    return $k, \boldsymbol{\omega}, i$ `//` $k_t = k$`, round numbers` ω `and ep.` i `unchanged`

$\beta \leftarrow \sqrt{\frac{\log(T\tilde{\Delta}^2)}{2\tau}}$

if $(1-\alpha)\left(\hat{\mu}_\ell + \beta\right) < \hat{\mu}_i - \beta$ **then**

  return $i, \boldsymbol{\omega}$, *None* `// Declare` i `as winner, set episode to None`

else if $\hat{\mu}_i + \beta < (1-\alpha)\left(\hat{\mu}_\ell - \beta\right)$ **then**

  return $i+1, \boldsymbol{\omega}, i+1$ `// Sample next candidate arm, Rounds` ω `unchanged, Update episode to that of next candidate arm`

else

  $\omega_i \leftarrow \omega_i + 1$ `// Increment round only for arm` i `being evaluated`

  return $i, \boldsymbol{\omega}, i$ `// Move to next round within same episode`

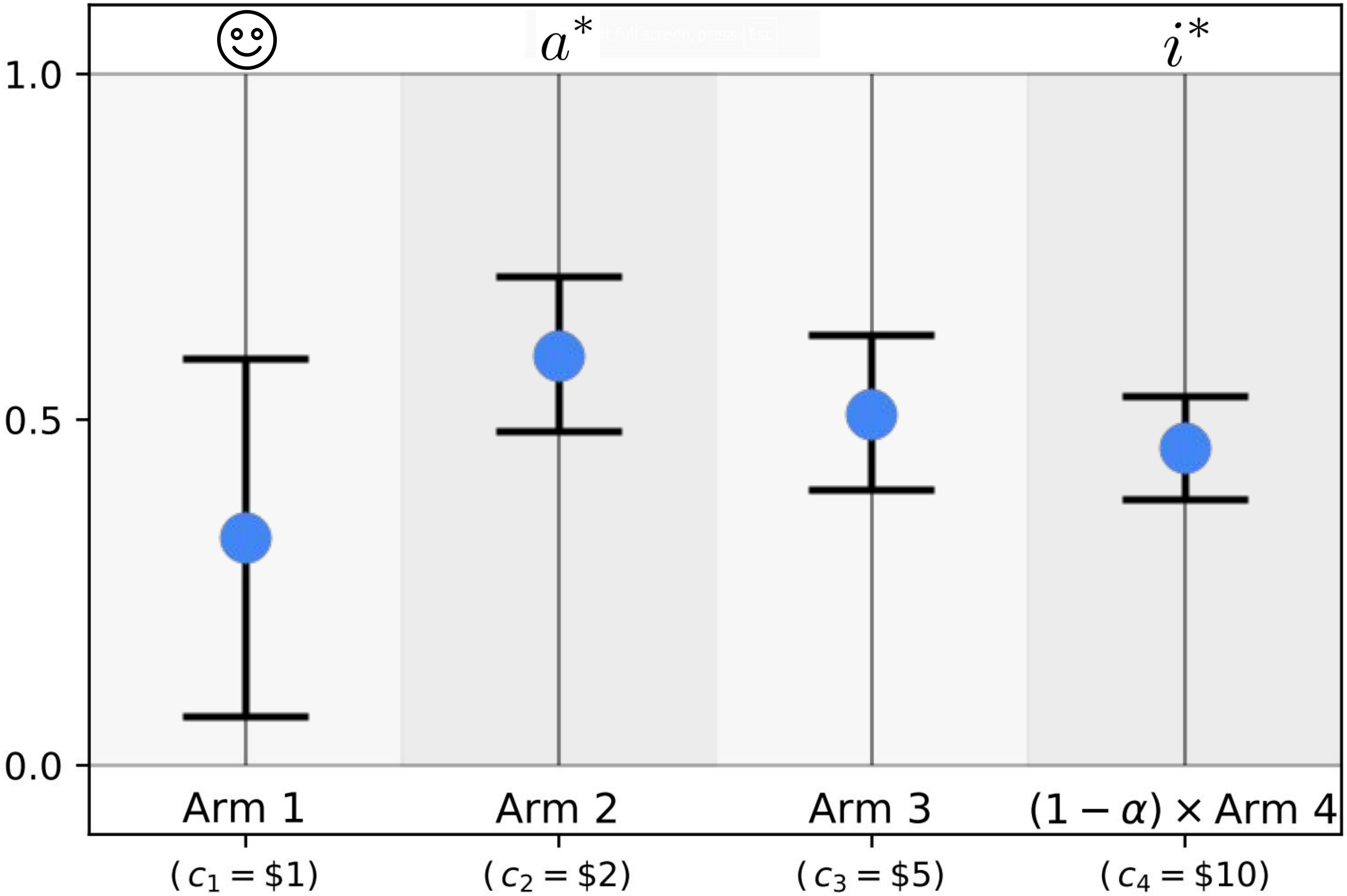
Algorithm 2: Pairwise Elimination for Cost Subsidy Problem (PE-CS).

Inputs: Bandit Instance ν, Horizon T, Subsidy Factor α.
Initialize: Samples $n_k = 0$, Empirical Means $\hat{\mu}_k = 0$, Current Rounds $\omega_k = 0 \; \forall k \in [K]$,
BAI Active Arms $\mathcal{A} = [K]$, Arm $\ell =$ None, PE Episode $i = 1$, Time $t = 1$.

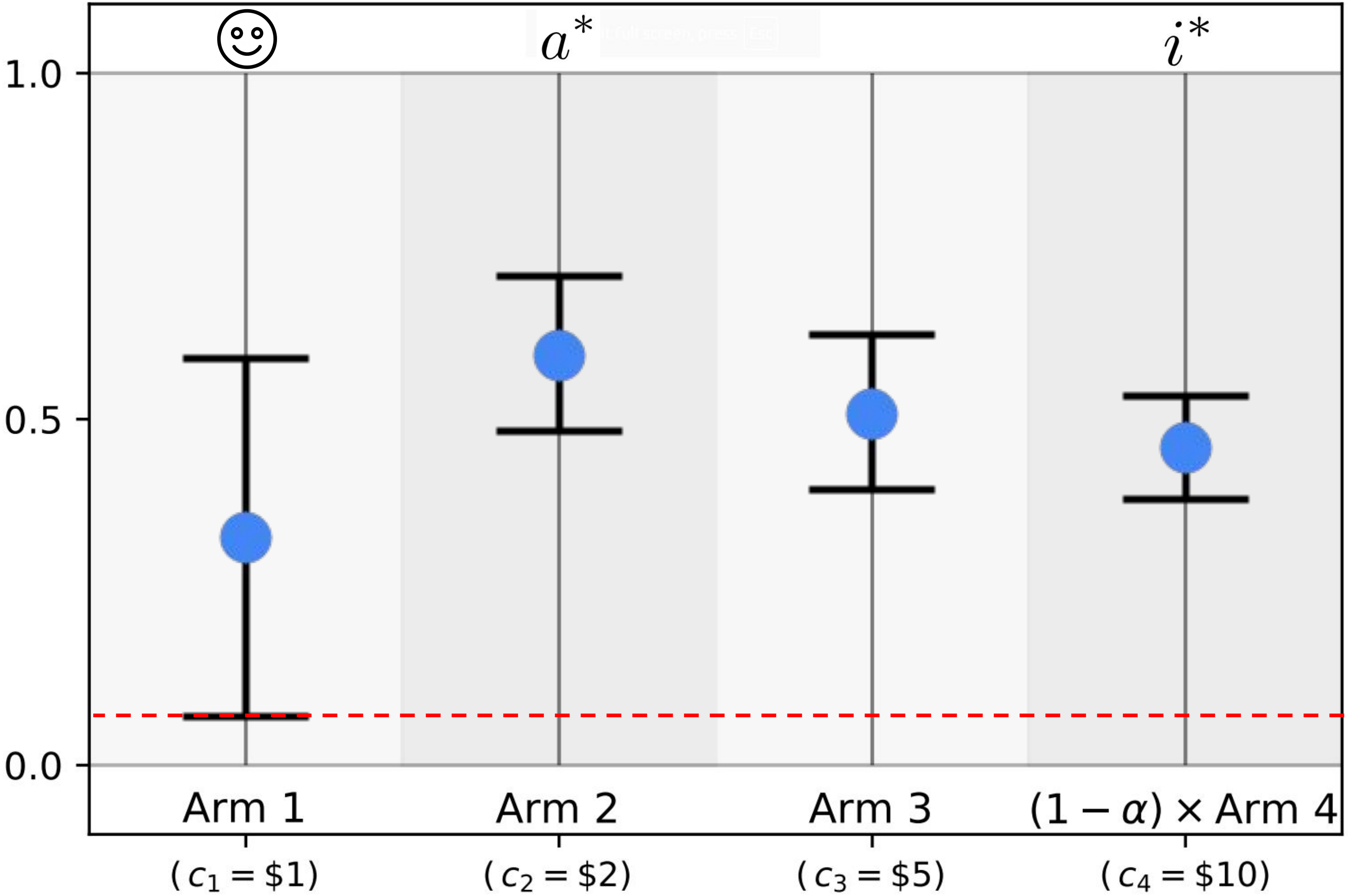
while $t \leq T$ **do**
  if $len(\mathcal{A}) > 1$ **then**
    $k_t, \boldsymbol{\omega}, \mathcal{A} \leftarrow \text{BAI}(\boldsymbol{n}, \hat{\boldsymbol{\mu}}, \boldsymbol{\omega}, T, \mathcal{A})$ `// receive arm to be sampled, updated round numbers, and updated active arms`
    if $len(\mathcal{A}) = 1$ **then**
      $\ell \leftarrow \mathcal{A}[0]$ `// set` ℓ `to be identified best arm`
      continue `// ignore sample recommendation` k_t
  else if $i \notin \{None, \ell\}$ **then**
    $k_t, \boldsymbol{\omega}, i \leftarrow \text{PE}(\boldsymbol{n}, \hat{\boldsymbol{\mu}}, \boldsymbol{\omega}, T, i, \ell, \alpha)$ `// receive arm to be sampled, updated round numbers, and updated episode number`
  else
    $k_t \leftarrow k_{t-1}$ `// sample winning arm for remaining budget`
  $\hat{\boldsymbol{\mu}}(t+1), \boldsymbol{n}(t+1), t \leftarrow \text{sample\_and\_update}(k_t, \hat{\boldsymbol{\mu}}(t), \boldsymbol{n}(t), t)$ `// in Appendix B`

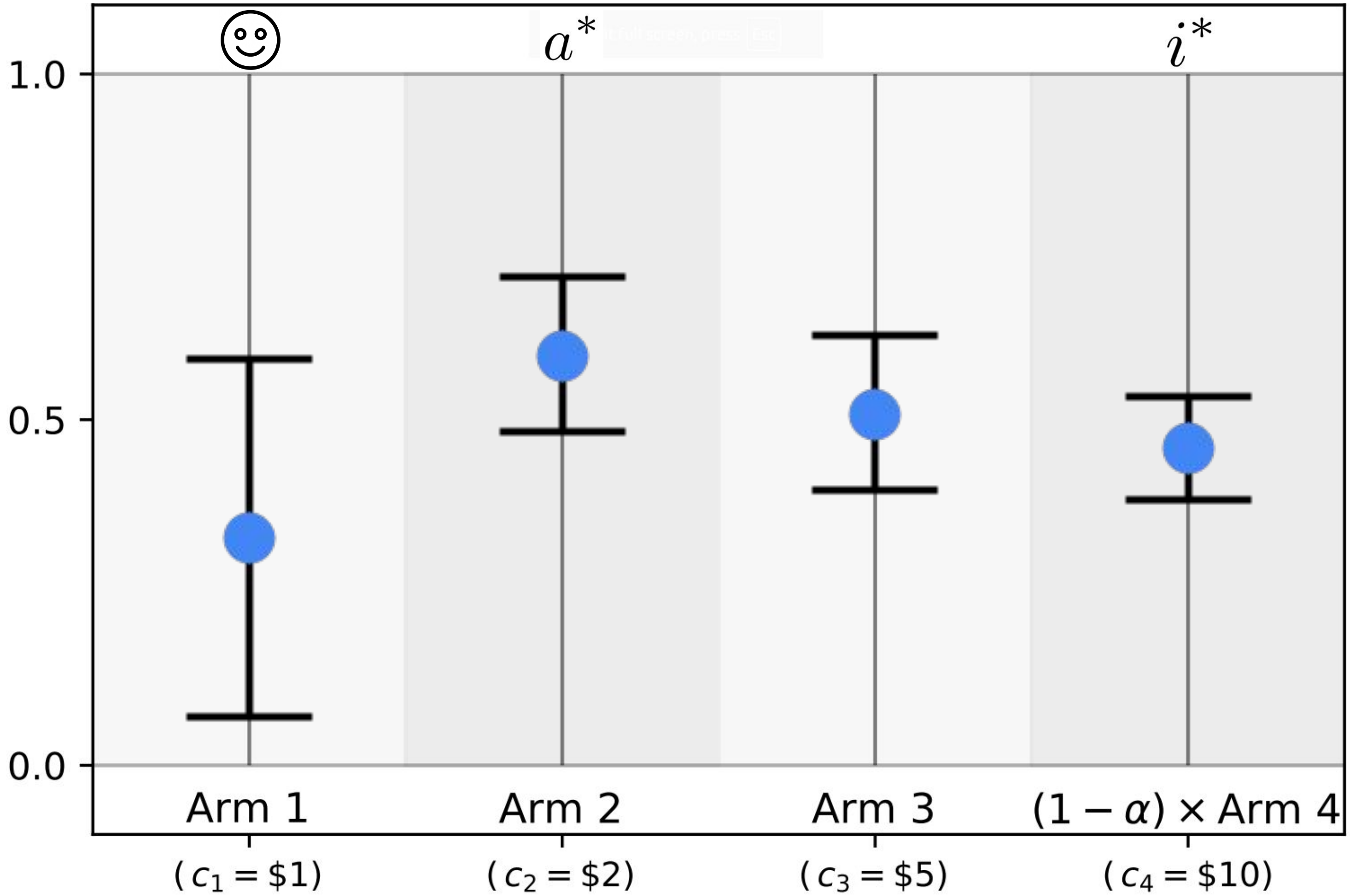
PE-CS Algorithm

PE Stage

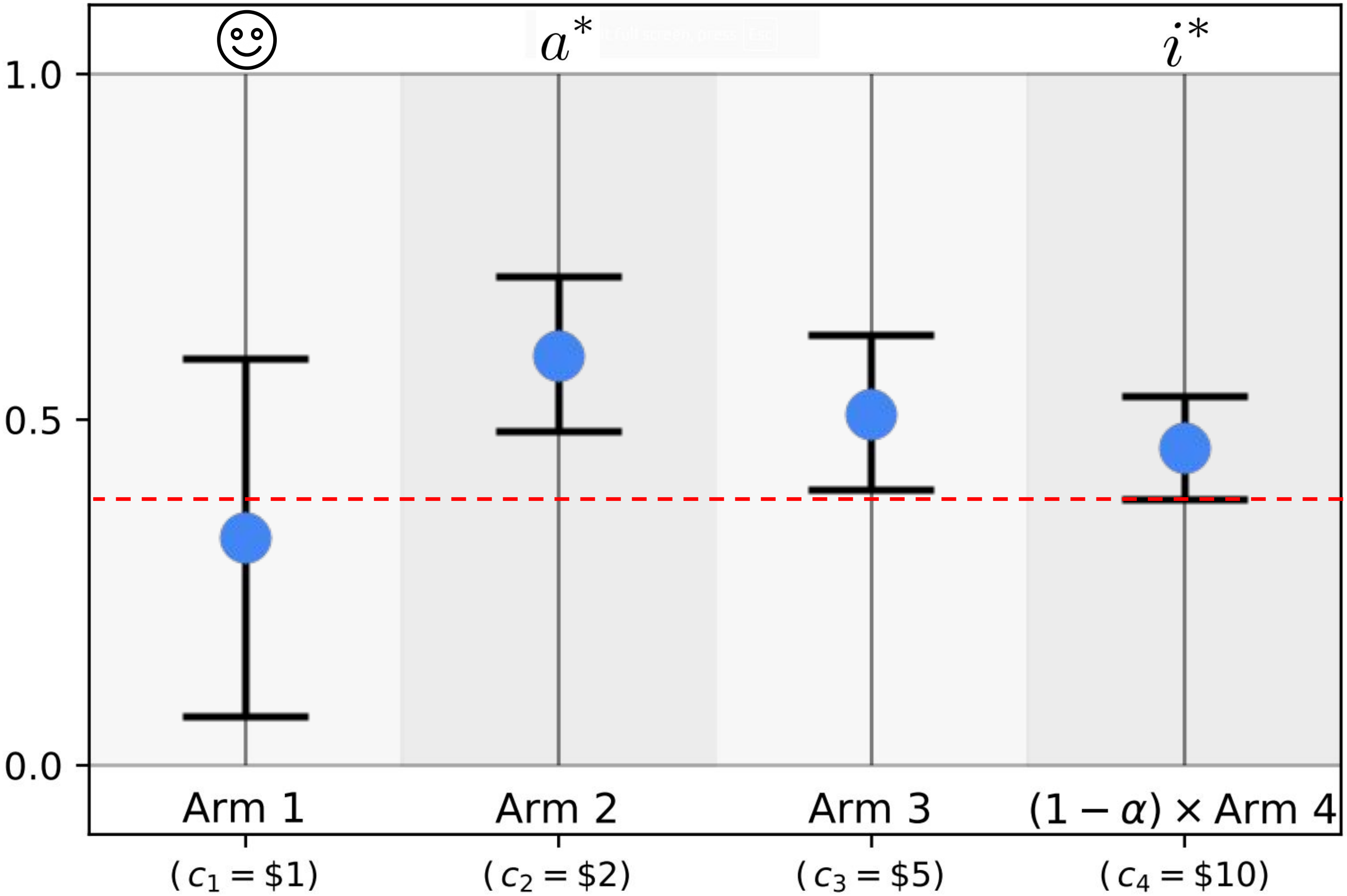


☺
a^*
i^*
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
($c_1 = \$1$)
($c_2 = \$2$)
($c_3 = \$5$)
($c_4 = \$10$)

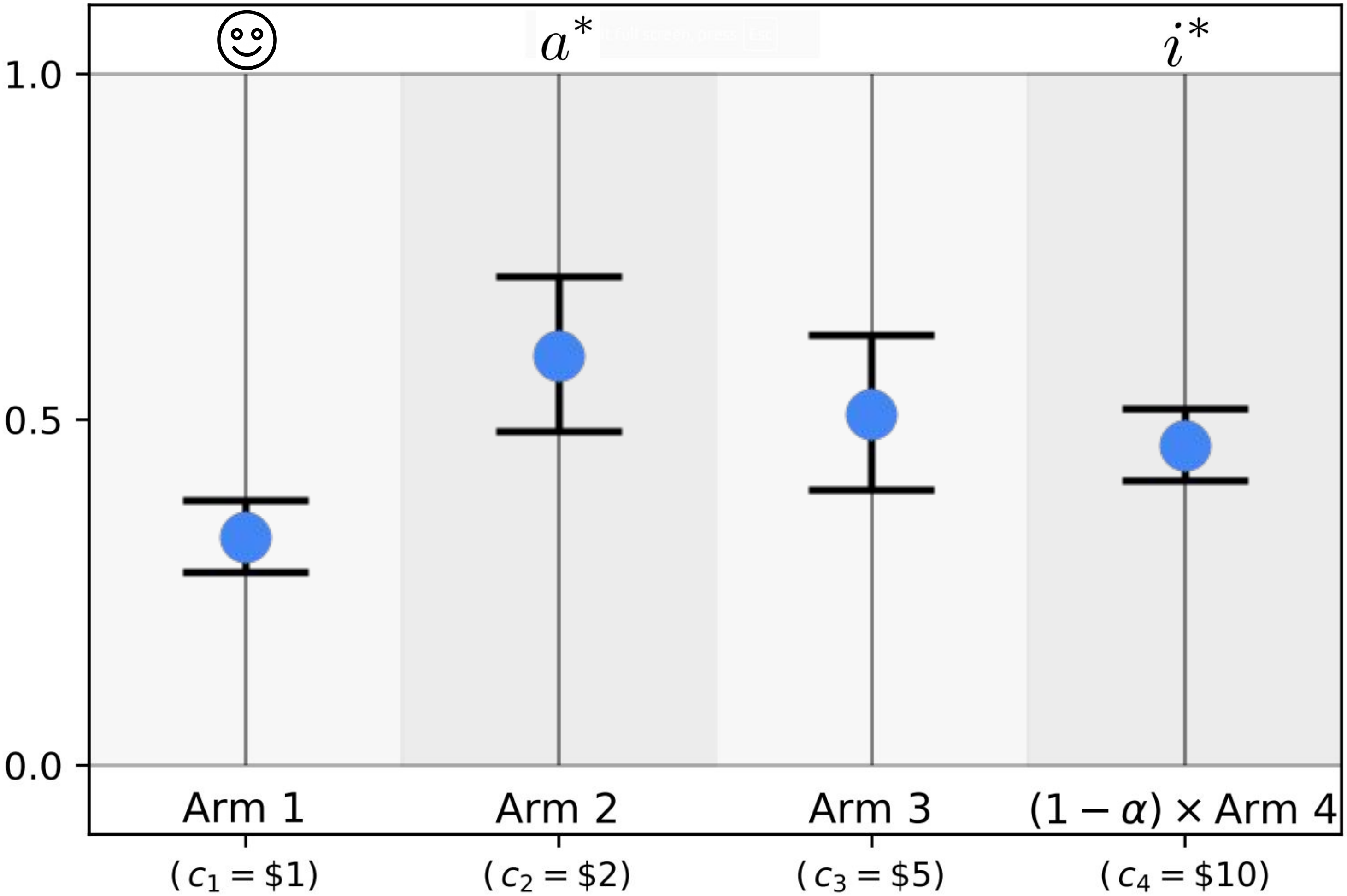

a^*
i^*
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$

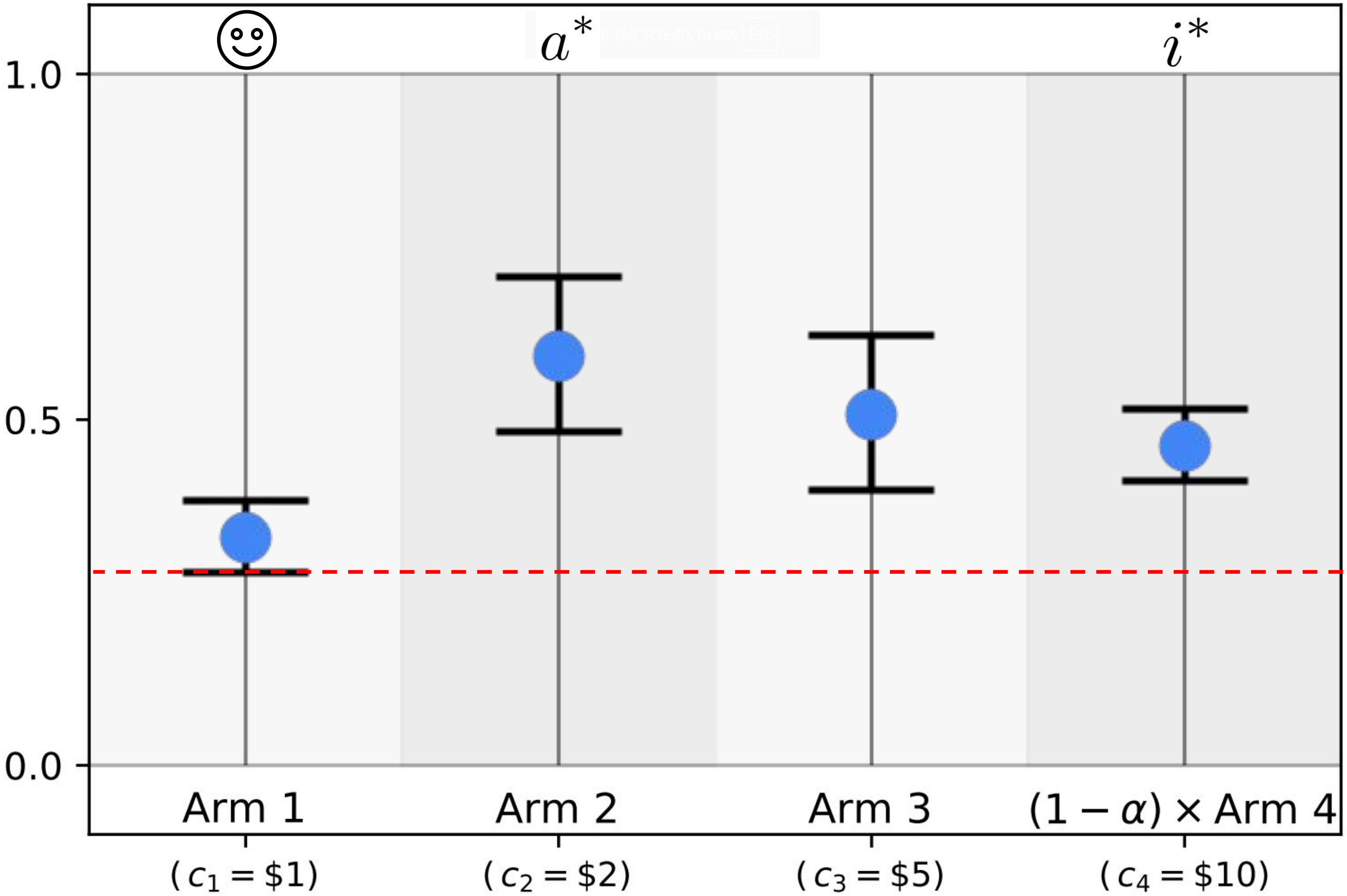


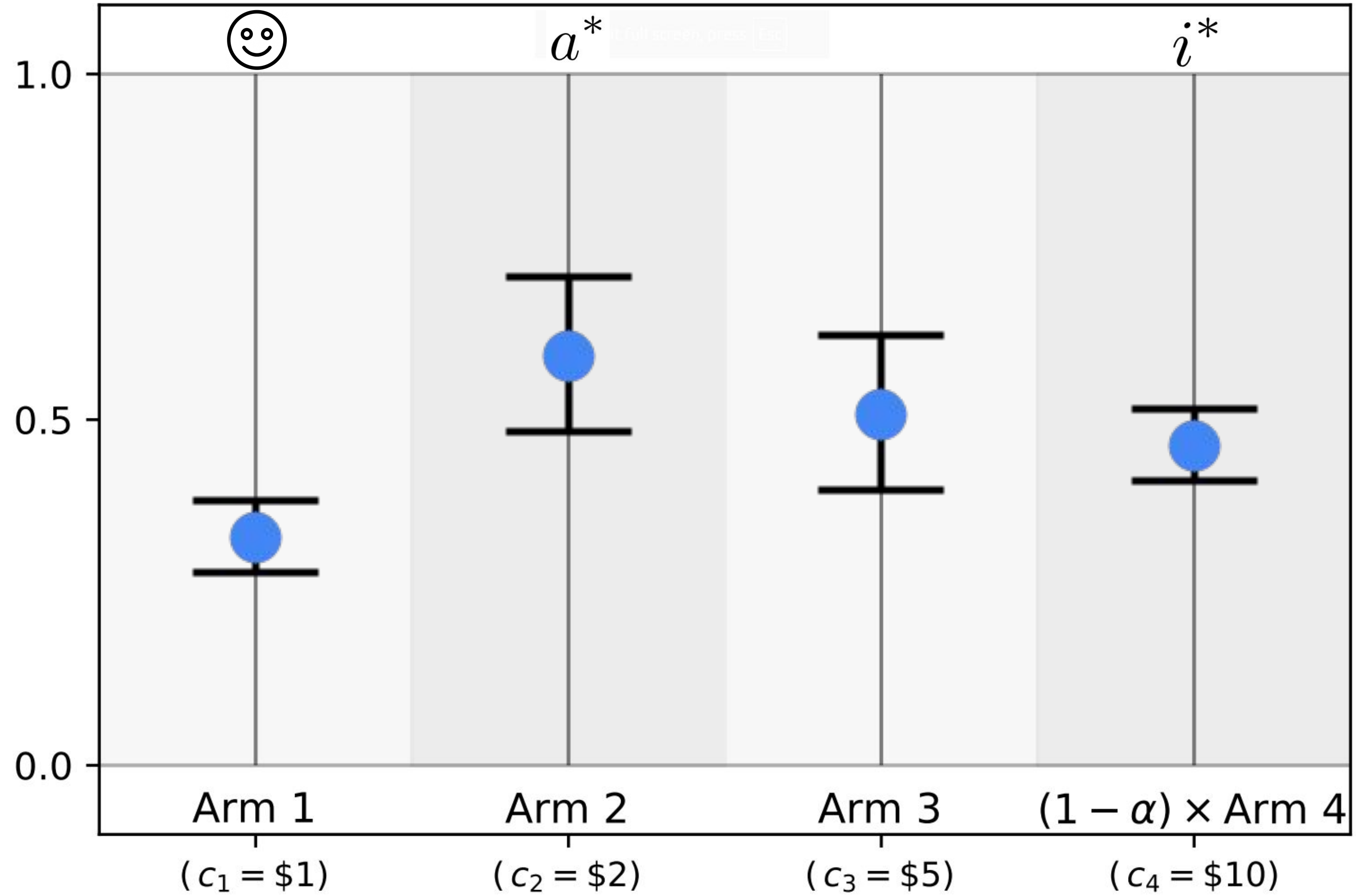
☺
a^*
i^*
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$

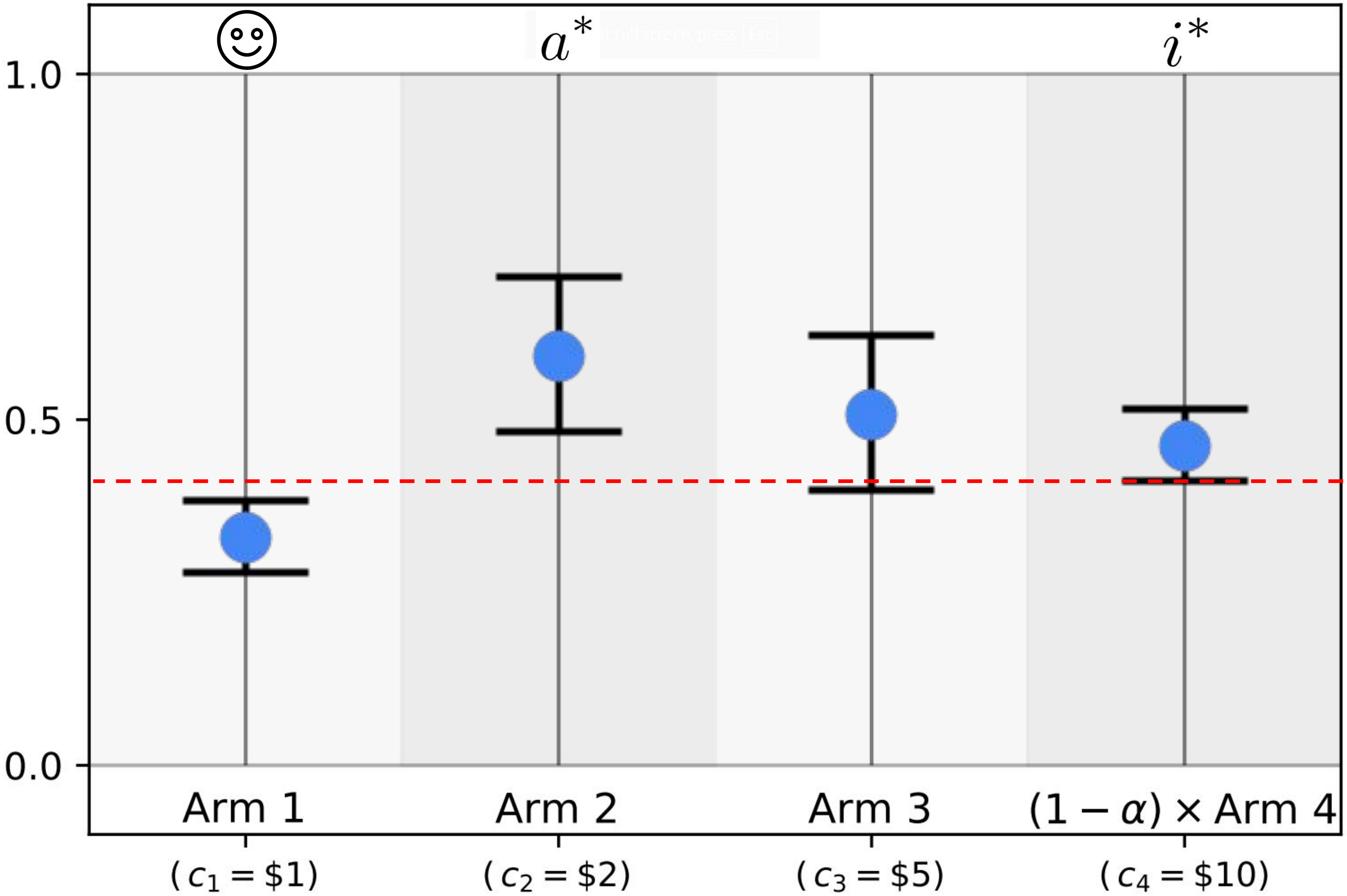


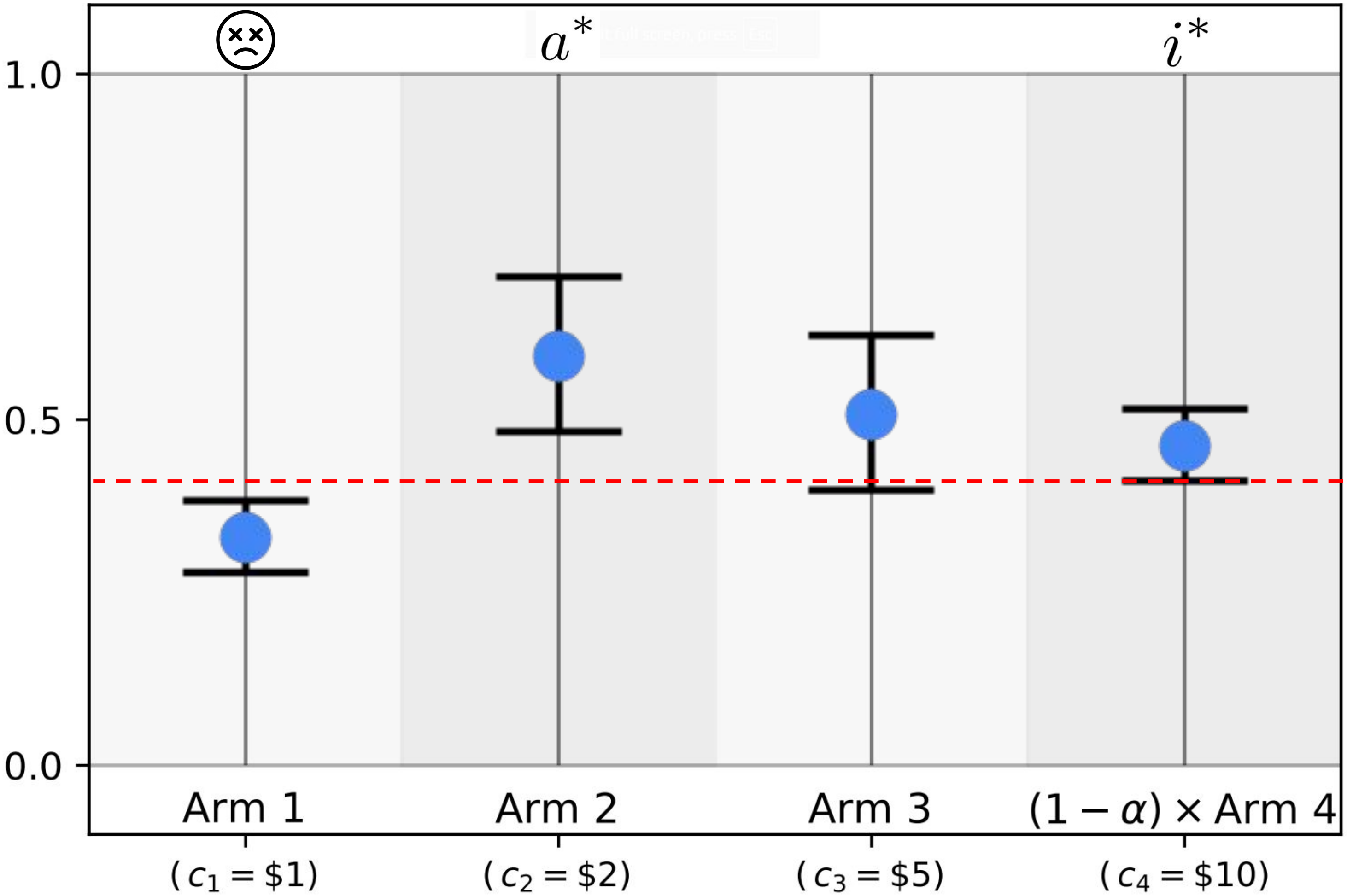
☺
a^*
i^*
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$

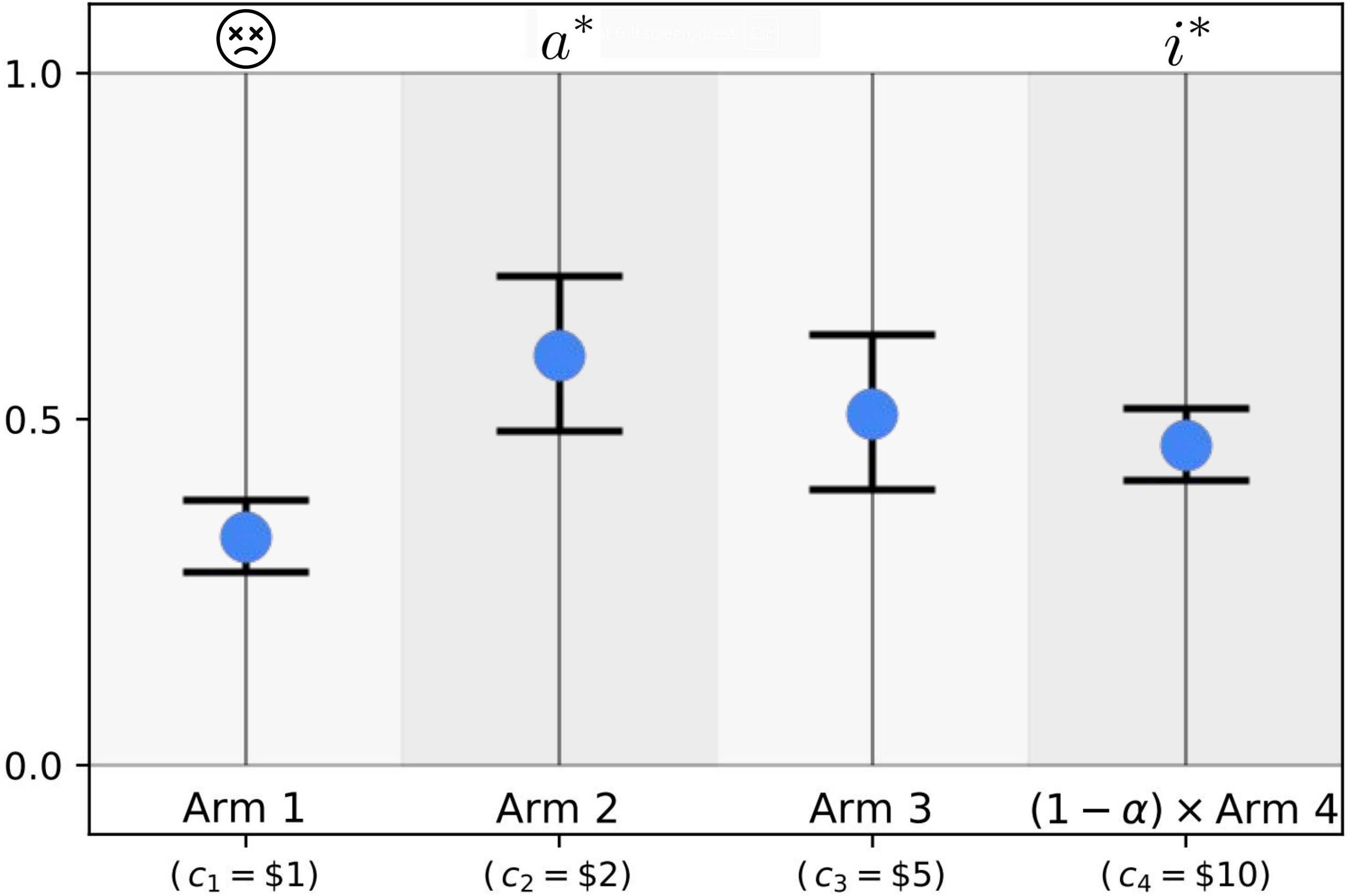

a^*
i^*
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$


☺
a^*
i^*
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$

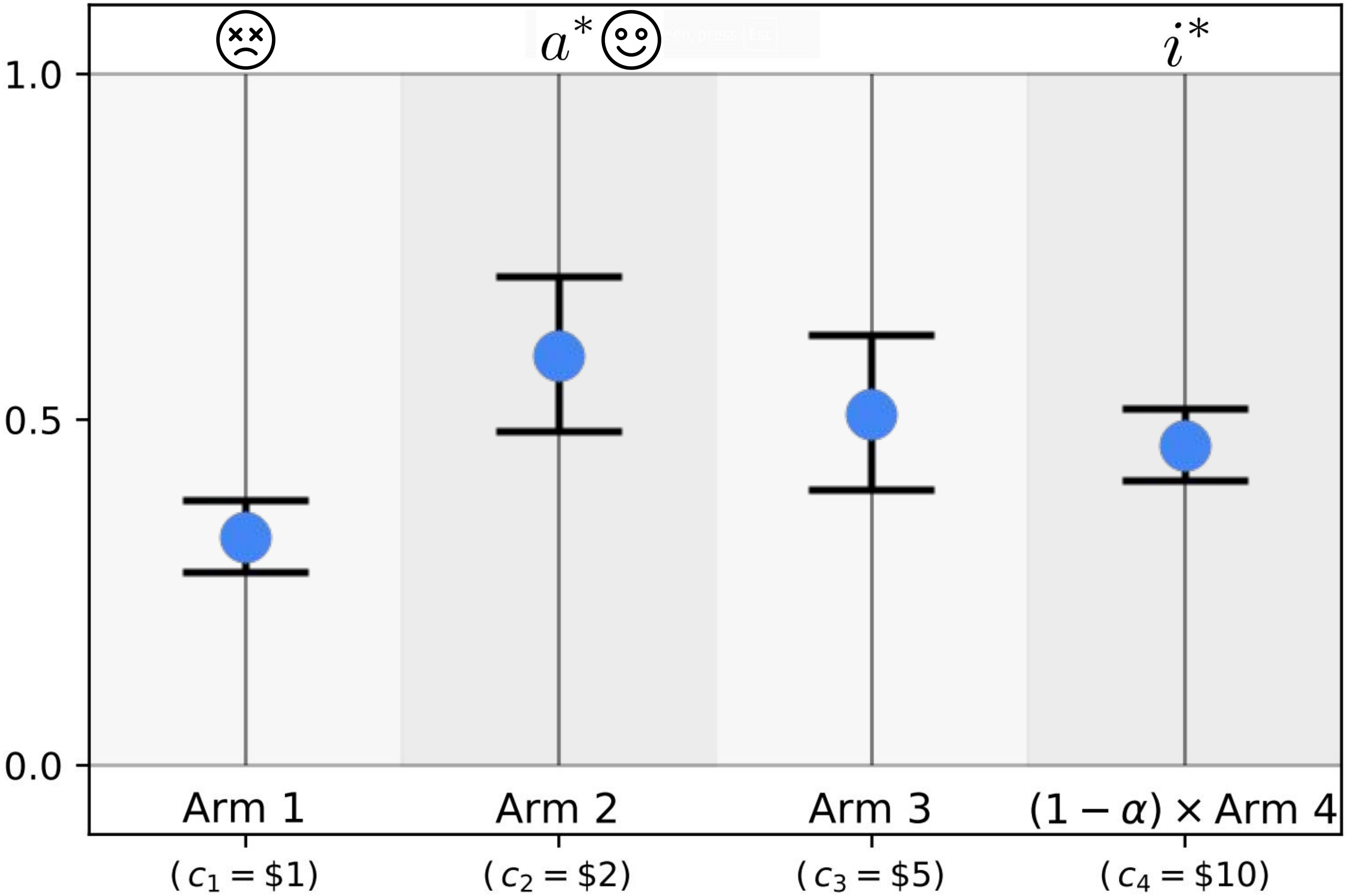

a^*
i^*
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$

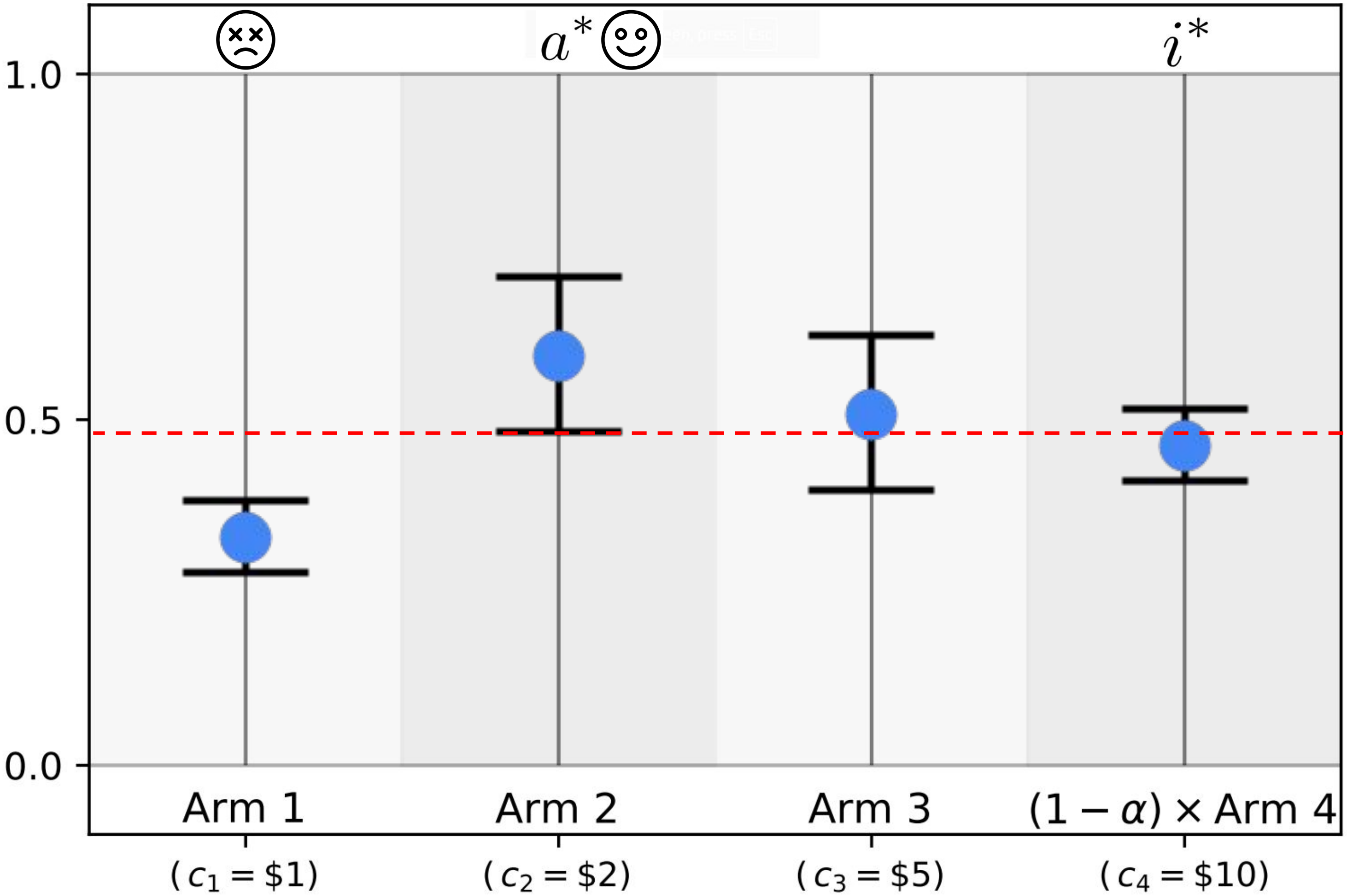

☺
a^*
i^*
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$


a^*
i^*
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$

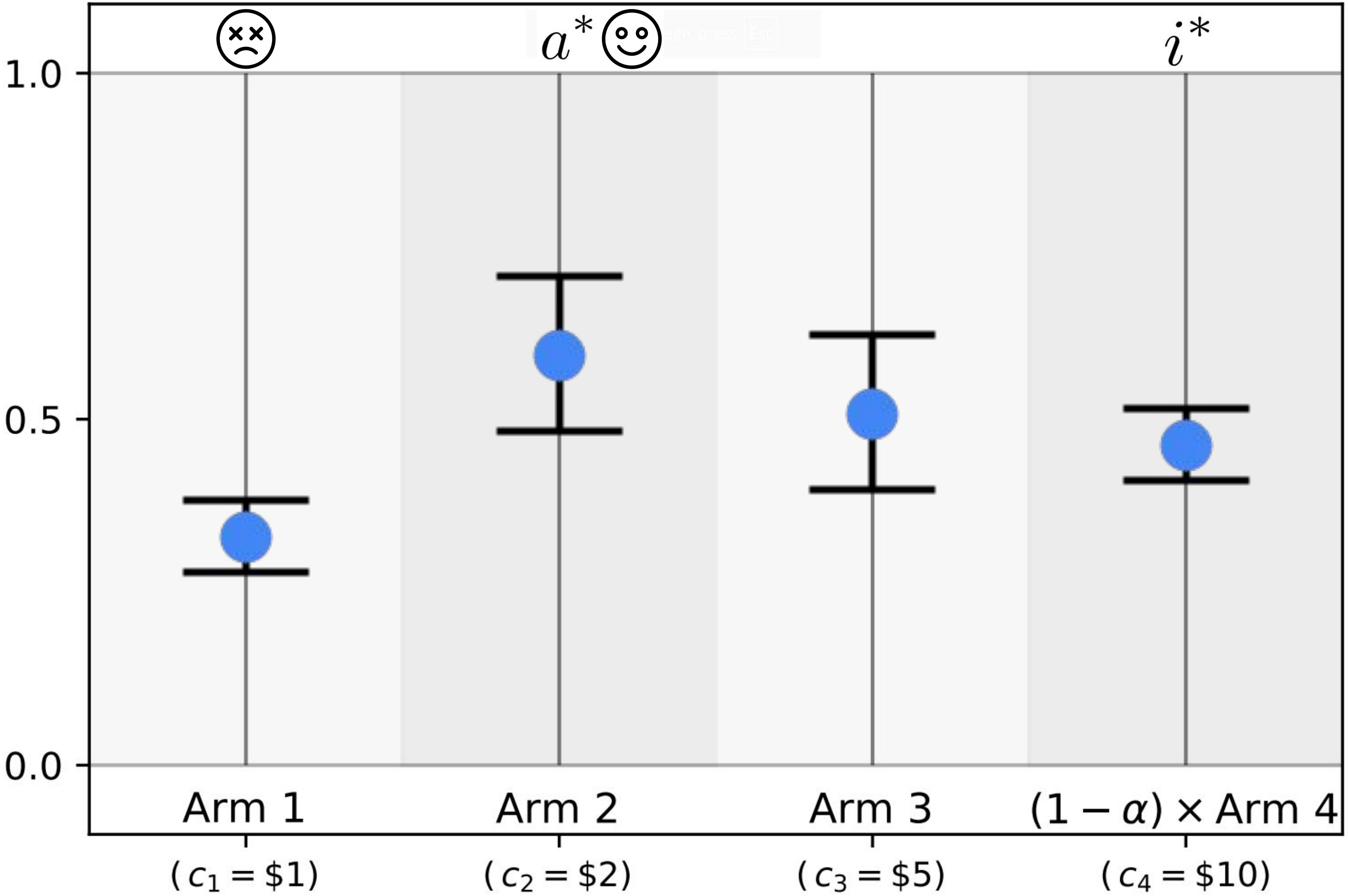


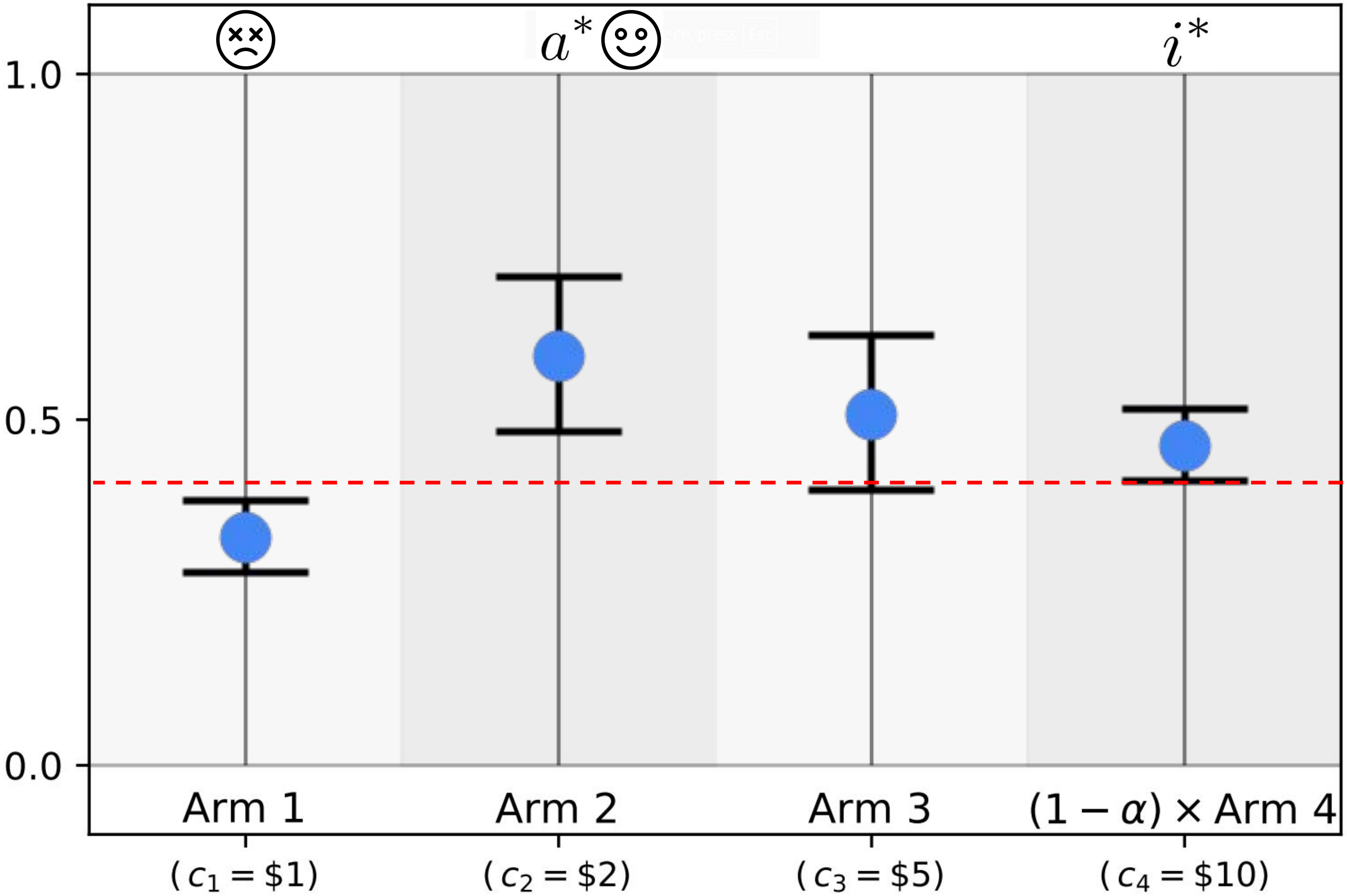
a^*
i^*
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$


a^*
i^*
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$

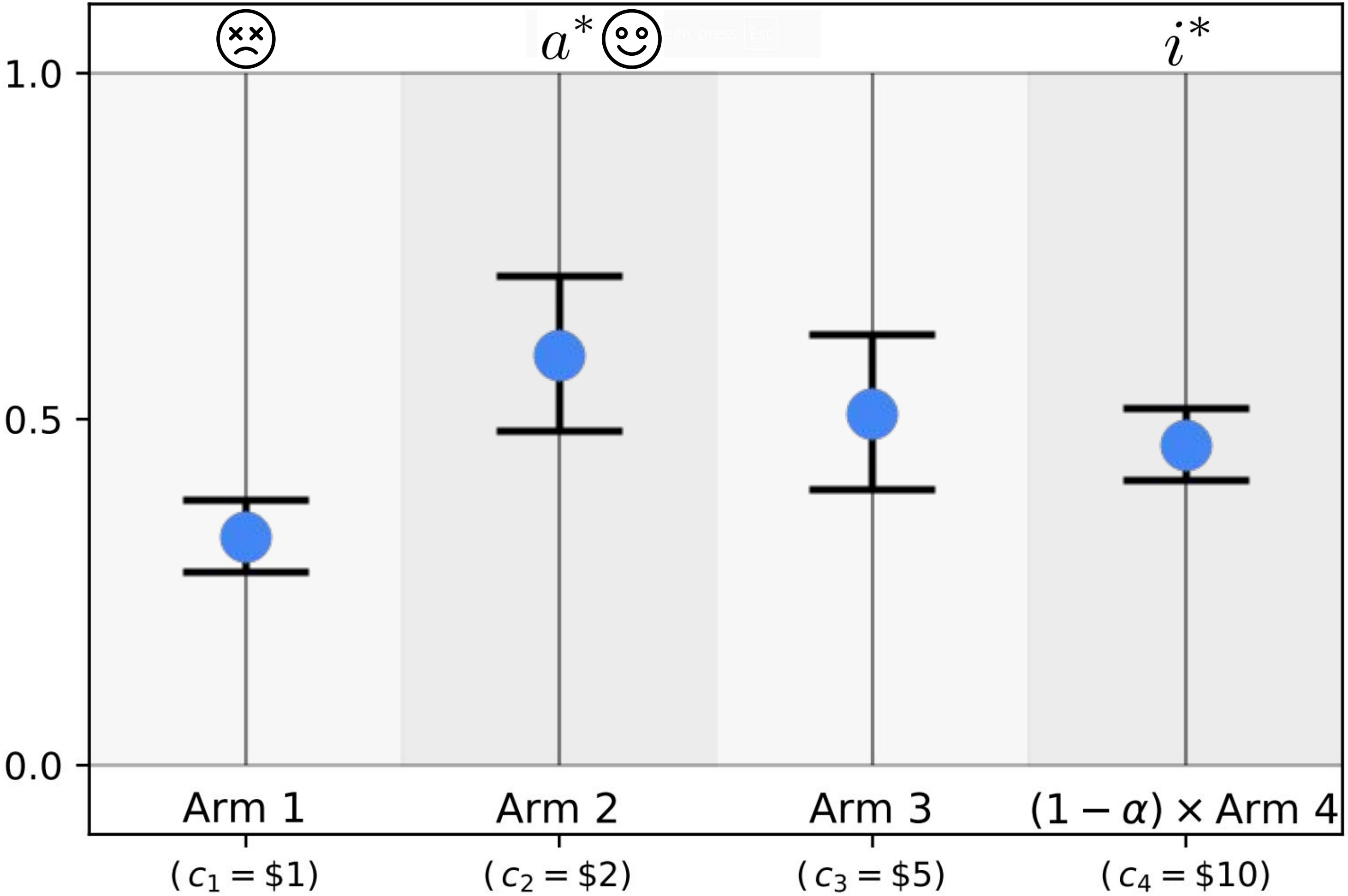


a^*
i^*
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$

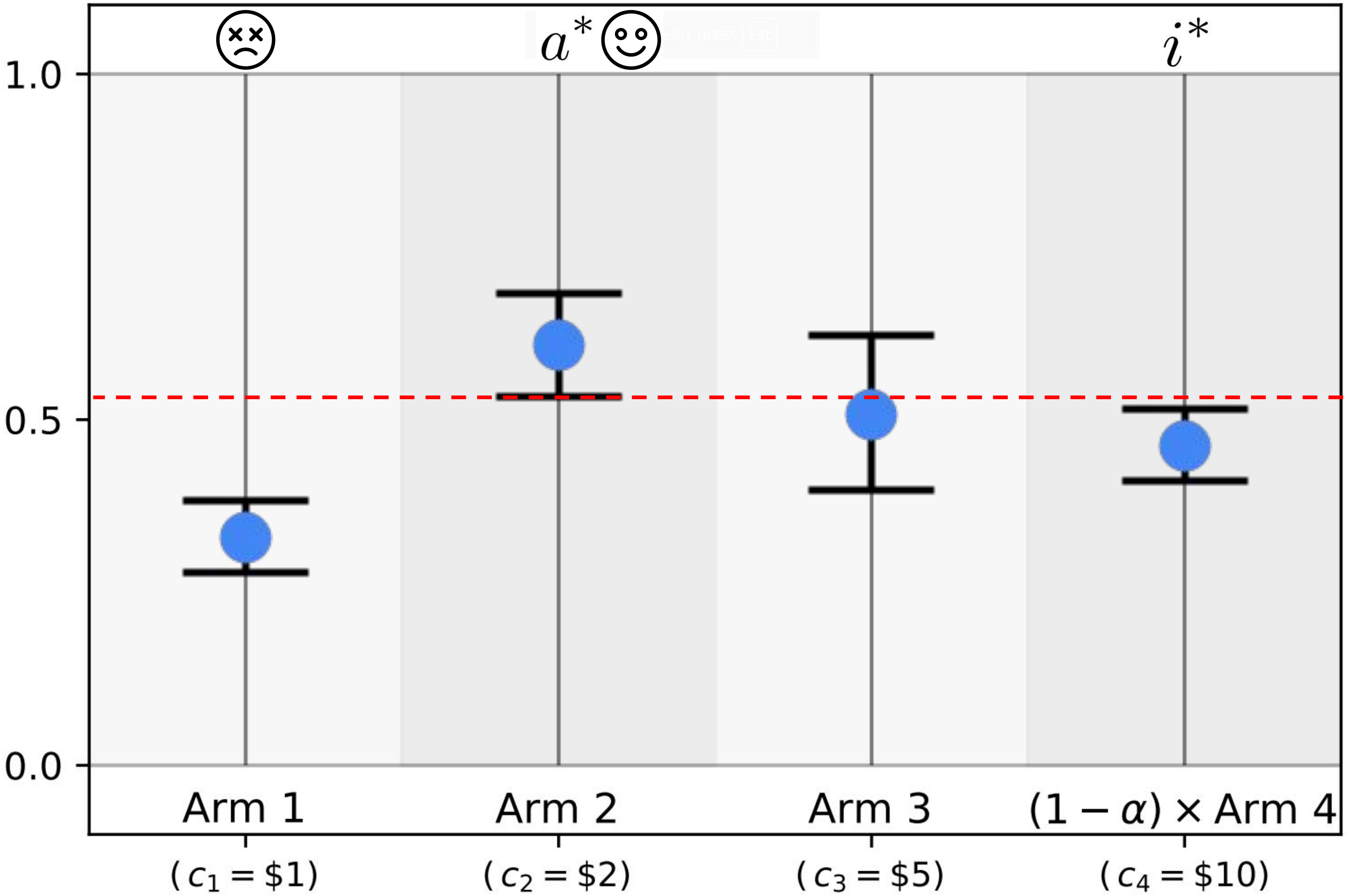

a^*
i^*
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$



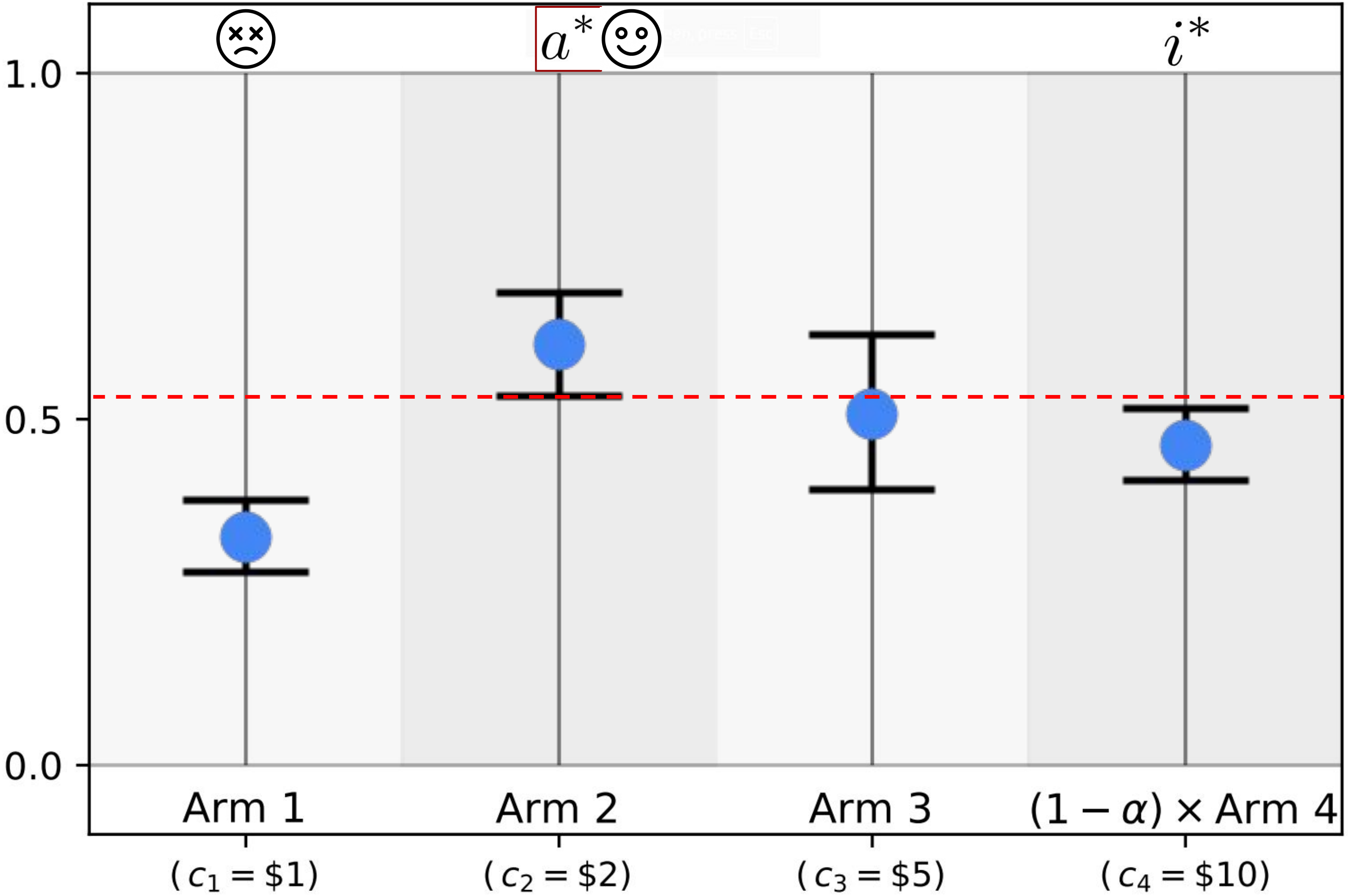
a^*
i^*
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$



a^*
i^*
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
($c_1 = \$1$)
($c_2 = \$2$)
($c_3 = \$5$)
($c_4 = \$10$)



a^*
i^*
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha)\times$ Arm 4
$(c_1=\$1)$
$(c_2=\$2)$
$(c_3=\$5)$
$(c_4=\$10)$



a^*
i^*
1.0
0.5
0.0
Arm 1
Arm 2
Arm 3
$(1-\alpha) \times$ Arm 4
$(c_1 = \$1)$
$(c_2 = \$2)$
$(c_3 = \$5)$
$(c_4 = \$10)$

Theorem 3.2 (Instance dependent upper bound on cumulative cost and quality regret for PE). *For bandit instance ν, over horizon T, the expected cumulative cost regret $\mathbb{E}\left[Cost\_Reg\left(T,\nu\right)\right]$ and quality regret $\mathbb{E}\left[Quality\_Reg\left(T,\nu\right)\right]$ of the PE algorithm are upper bounded respectively as,*

$$\underbrace{\left(1+\max_{i\leq a^*}\frac{32\log\left(T\Delta_{Q,i}^2\right)}{\Delta_{Q,i}^2}\right)\Delta_{C,\ell}^+}_{\substack{\text{Contribution from arm } \ell \text{ under}\\ \text{nominal termination in PE episode } a^*}}+\underbrace{\left(\sum_{i=1}^{a^*}\frac{43}{\Delta_{Q,i}^2}\right)\Delta_{C,\ell}^+}_{\substack{\text{Contribution from arm } \ell \text{ under}\\ \text{mis-termination in PE episode } \leq a^*}}+\underbrace{\frac{43}{\Delta_{Q,a^*}^2}\max_{a^*<i\leq\ell}\Delta_{C,i}^+}_{\substack{\text{Contribution from episodes } >a^*\\ \text{in case of mis-termination during ep } a^*}},$$

$$\underbrace{\sum_{i=1}^{a^*-1}\left(\Delta_{Q,i}+\frac{32\log\left(T\Delta_{Q,i}^2\right)}{\Delta_{Q,i}}\right)}_{\substack{\text{Contribution from arms } i<a^* \text{ under}\\ \text{nominal termination in PE episode } a^*}}+\underbrace{\sum_{i=1}^{a^*-1}\frac{43}{\Delta_{Q,i}}}_{\substack{\text{Contribution from arms } i<a^* \text{ under}\\ \text{mis-termination in PE episode } \leq a^*}}+\underbrace{\frac{43}{\Delta_{Q,a^*}^2}\max_{a^*<i<\ell}\Delta_{Q,i}^+}_{\substack{\text{Contribution from episodes } >a^*\\ \text{in case of mis-termination during ep } a^*}}.$$

Theorem 3.2 (Instance dependent upper bound on cumulative cost and quality regret for PE). *For bandit instance ν, over horizon T, the expected cumulative cost regret $\mathbb{E}\left[Cost\_Reg\left(T,\nu\right)\right]$ and quality regret $\mathbb{E}\left[Quality\_Reg\left(T,\nu\right)\right]$ of the PE algorithm are upper bounded respectively as,*

$$\underbrace{\left(1+\max_{i\le a^*}\frac{32\log\left(T\Delta_{Q,i}^2\right)}{\Delta_{Q,i}^2}\right)\Delta_{C,\ell}^+}_{\substack{\textit{Contribution from arm } \ell \textit{ under}\\ \textit{nominal termination in PE episode } a^*}} + \underbrace{\left(\sum_{i=1}^{a^*}\frac{43}{\Delta_{Q,i}^2}\right)\Delta_{C,\ell}^+}_{\substack{\textit{Contribution from arm } \ell \textit{ under}\\ \textit{mis-termination in PE episode } \le a^*}} + \underbrace{\frac{43}{\Delta_{Q,a^*}^2}\max_{a^*<i\le\ell}\Delta_{C,i}^+}_{\substack{\textit{Contribution from episodes } > a^*\\ \textit{in case of mis-termination during ep } a^*}},$$

$$\underbrace{\sum_{i=1}^{a^*-1}\left(\Delta_{Q,i}+\frac{32\log\left(T\Delta_{Q,i}^2\right)}{\Delta_{Q,i}}\right)}_{\substack{\textit{Contribution from arms } i<a^* \textit{ under}\\ \textit{nominal termination in PE episode } a^*}} + \underbrace{\sum_{i=1}^{a^*-1}\frac{43}{\Delta_{Q,i}}}_{\substack{\textit{Contribution from arms } i<a^* \textit{ under}\\ \textit{mis-termination in PE episode } \le a^*}} + \underbrace{\frac{43}{\Delta_{Q,a^*}^2}\max_{a^*<i<\ell}\Delta_{Q,i}^+}_{\substack{\textit{Contribution from episodes } > a^*\\ \textit{in case of mis-termination during ep } a^*}}.$$

$$
\underbrace{\Delta_{C,i^*}^{+}\left(1+\max_{\Delta\in\boldsymbol{\Delta}}\left\{\frac{32\log\left(T\Delta^2\right)}{\Delta^2}\right\}\right)}_{\substack{\text{Contribution from } i^* \text{ under nominal}\\ \text{termination in PE-stage episode } a^*}}+\underbrace{\Delta_{C,i^*}^{+}\left(\sum_{i=1}^{a^*-1}\frac{43}{\Delta_{Q,i}^2}+\left(\frac{32}{\Delta_{a^*}^2}+\frac{43}{\Delta_{Q,a^*}^2}\right)\right)}_{\substack{\text{Contribution from } i^* \text{ under}\\ \text{mis-termination in PE-stage episode } \le a^*}}
$$

$$
+\underbrace{\sum_{i>a^*,i\in[K]\setminus\{i^*\}}\Delta_{C,i}^{+}\left(1+\frac{32\log\left(T\Delta_i^2\right)}{\Delta_i^2}\right)}_{\substack{\text{Contribution from high-cost arms}\\ \text{with a proper end to the BAI-stage}}}+\underbrace{\max_{i>a^*,i\in[K]}\Delta_{C,i}^{+}\left(\frac{32}{\Delta_{a^*}^2}+\frac{43}{\Delta_{Q,a^*}^2}\right)}_{\substack{\text{Contribution from PE-stage episodes } >a^*\\ \text{in case of mis-termination during ep } a^*}}
$$

$$
+\underbrace{\max_{i\in[K]}\Delta_{C,i}^{+}\left(\frac{11}{\Delta_{\min}^2}+\sum_{j\neq i^*}\frac{32}{\Delta_j^2}\right)}_{\substack{\text{Contribution from improper}\\ \text{end to BAI stage}}}.
$$

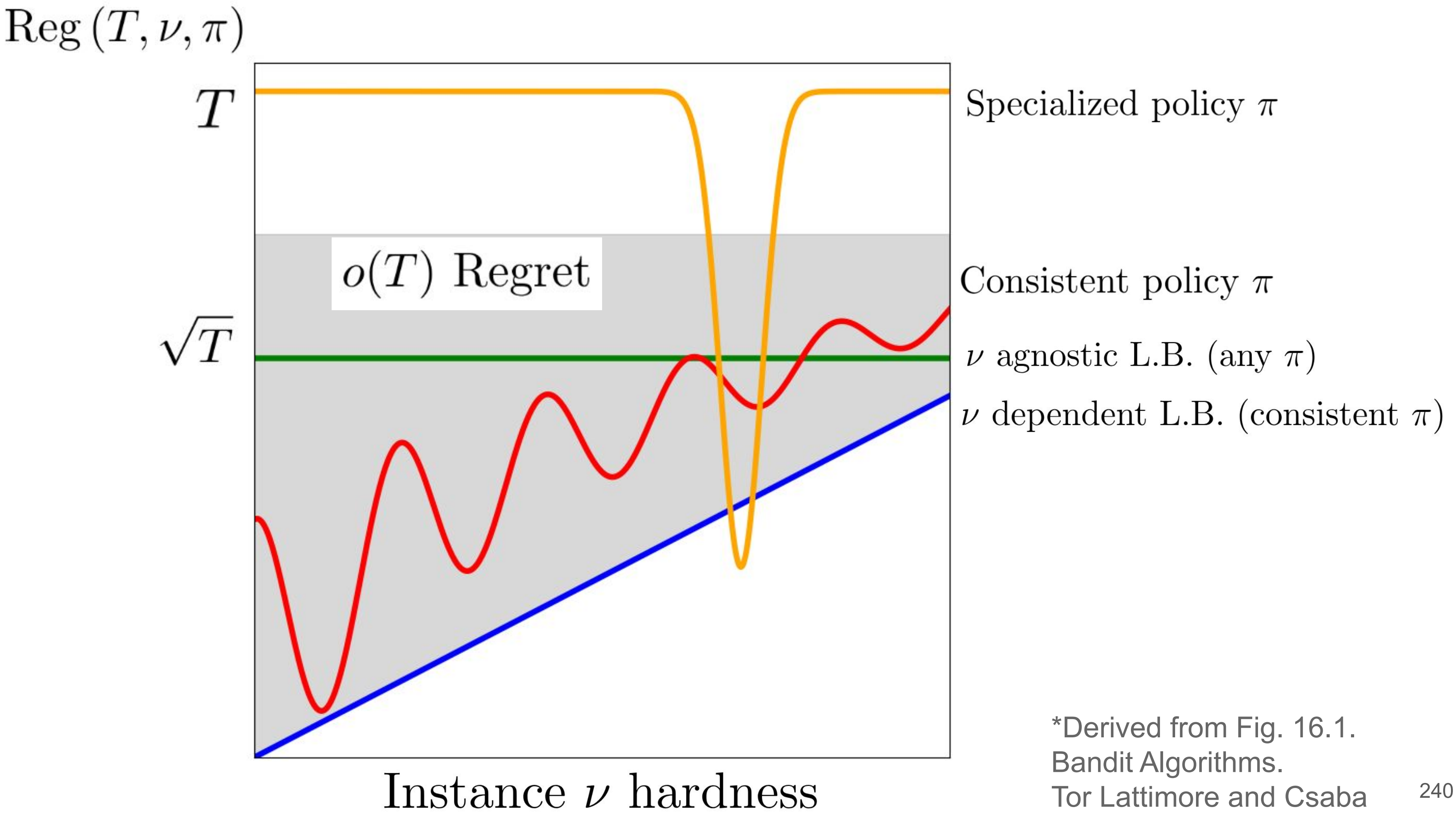
$$\underbrace{\sum_{i=1}^{a^*-1}\left(\Delta_{Q,i}+\frac{32\log\left(T\Delta_{Q,i}^2\right)}{\Delta_{Q,i}}\right)}_{\substack{\text{Contribution from } i<a^* \text{ under nominal}\\ \text{termination in PE-stage episode } a^*}}+\underbrace{\sum_{i=1}^{a^*-1}\frac{43}{\Delta_{Q,i}}}_{\substack{\text{Contribution from}\\ i<a^* \text{ under}\\ \text{mis-termination in}\\ \text{PE-stage episode } \leq a^*}}+\underbrace{\max_{i>a^*,i\in[K]}\Delta_{Q,i}^+\left(\frac{32}{\Delta_{a^*}^2}+\frac{43}{\Delta_{Q,a^*}^2}\right)}_{\substack{\text{Contribution from PE-stage episodes } >a^*\\ \text{in case of mis-termination during ep } a^*}}$$

$$+\underbrace{\sum_{i>a^*,i\in[K]\setminus\{i^*\}}\Delta_{Q,i}^+\left(1+\frac{32\log\left(T\Delta_i^2\right)}{\Delta_i^2}\right)}_{\substack{\text{Contribution from PE-stage episodes } >a^*\\ \text{in case of mis-termination during ep } a^*}}+\underbrace{\max_{i\in[K]}\Delta_{Q,i}^+\left(\frac{11}{\Delta_{\min}^2}+\sum_{j\neq i^*}\frac{32}{\Delta_j^2}\right)}_{\substack{\text{Contribution from improper}\\ \text{end to BAI stage}}}.$$

Consistent Policy π

π is consistent if for all bandit instances ν, in instance class $\mathcal{E}$, for all arms $i \neq a^*$,

$$\mathbb{E}\left[n_i(T)\right] = o\left(T^\gamma\right) \; \forall \gamma > 0$$



*Derived from Fig. 16.1.
Bandit Algorithms.
Tor Lattimore and Csaba Szepesvári

Theorem 3.1 (Lower bound for known reference arm setting). *Under any consistent policy π the expected number of samples of a low cost arm and of the reference arm ℓ are lower bounded as,*

$$\liminf_{T\to\infty} \frac{\mathbb{E}\left[n_i\left(T\right)\right]}{\log T} \geq \frac{2}{\Delta_{Q,i}^2}, \quad \textit{for arms} \;\; i < a^*, \quad \liminf_{T\to\infty} \frac{\mathbb{E}\left[n_\ell\left(T\right)\right]}{\log T} \geq \max_{i \leq a^*} \frac{2(1-\alpha)^2}{\Delta_{Q,i}^2}.$$

Where T denotes the problem horizon and the rewards of all K arms are Gaussian distributed with variance $\sigma^2 = 1$. Low cost is a term used relative to the cost of optimal arm a^*.

Theorem 3.3 (Lower bound for subsidized best reward setting). Under any consistent policy π, the expected number of samples of low cost arms, high cost arms, and the best reward arm i^* are lower bounded respectively as,

$$\liminf_{T\to\infty} \frac{\mathbb{E}\left[n_i\left(T\right)\right]}{\log T} \geq \frac{2}{\Delta_{Q,i}^2},\ i < a^*. \quad \liminf_{T\to\infty} \frac{\mathbb{E}\left[n_i\left(T\right)\right]}{\log T} \geq \frac{2}{(\frac{\mu_{a^*}}{1-\alpha} - \mu_i)^2}, i > a^*, i \neq i^*.$$

$$\liminf_{T\to\infty} \frac{\mathbb{E}\left[n_{i^*}\left(T\right)\right]}{\log T} \geq 2(1-\alpha)^2 \max\left\{\frac{1}{\Delta_{Q,a^*}^2}, \max_{i<a^*} \frac{1}{\Delta_{Q,i}^2}\mathbf{1}[\min_{i<a^*} \Delta_{Q,i} \leq (1-\alpha)\Delta_{min}]\right\}.$$

Where $\Delta_{min} = \mu^* - \max_{i\neq i^*} \mu_i$ is the smallest conventional gap, T denotes the horizon and the rewards of all K arms are Gaussian distributed with $\sigma^2 = 1$.