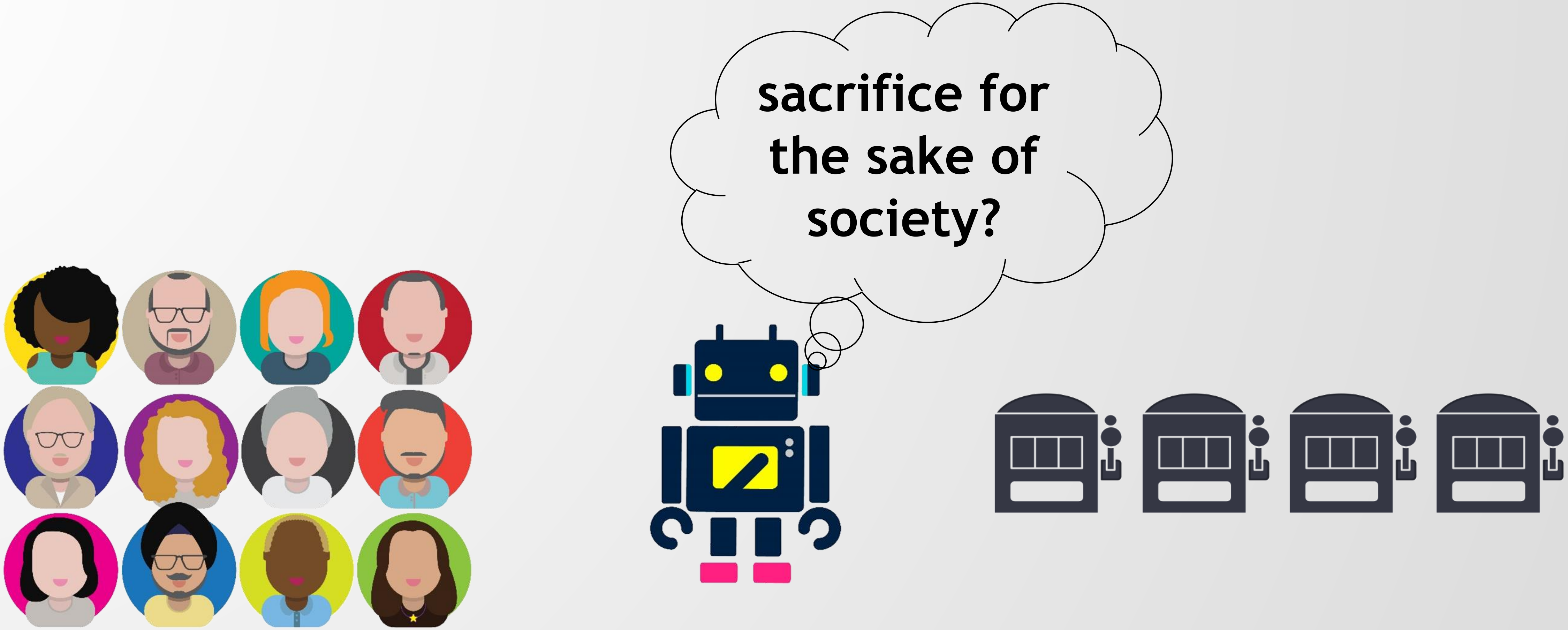
# Fiduciary Bandits


**Omer Ben–Porat**

**Technion – Israel Institute of Technology**

**Joint work with Gal Bahar, Kevin Leyton-Brown and Moshe Tennenholtz**



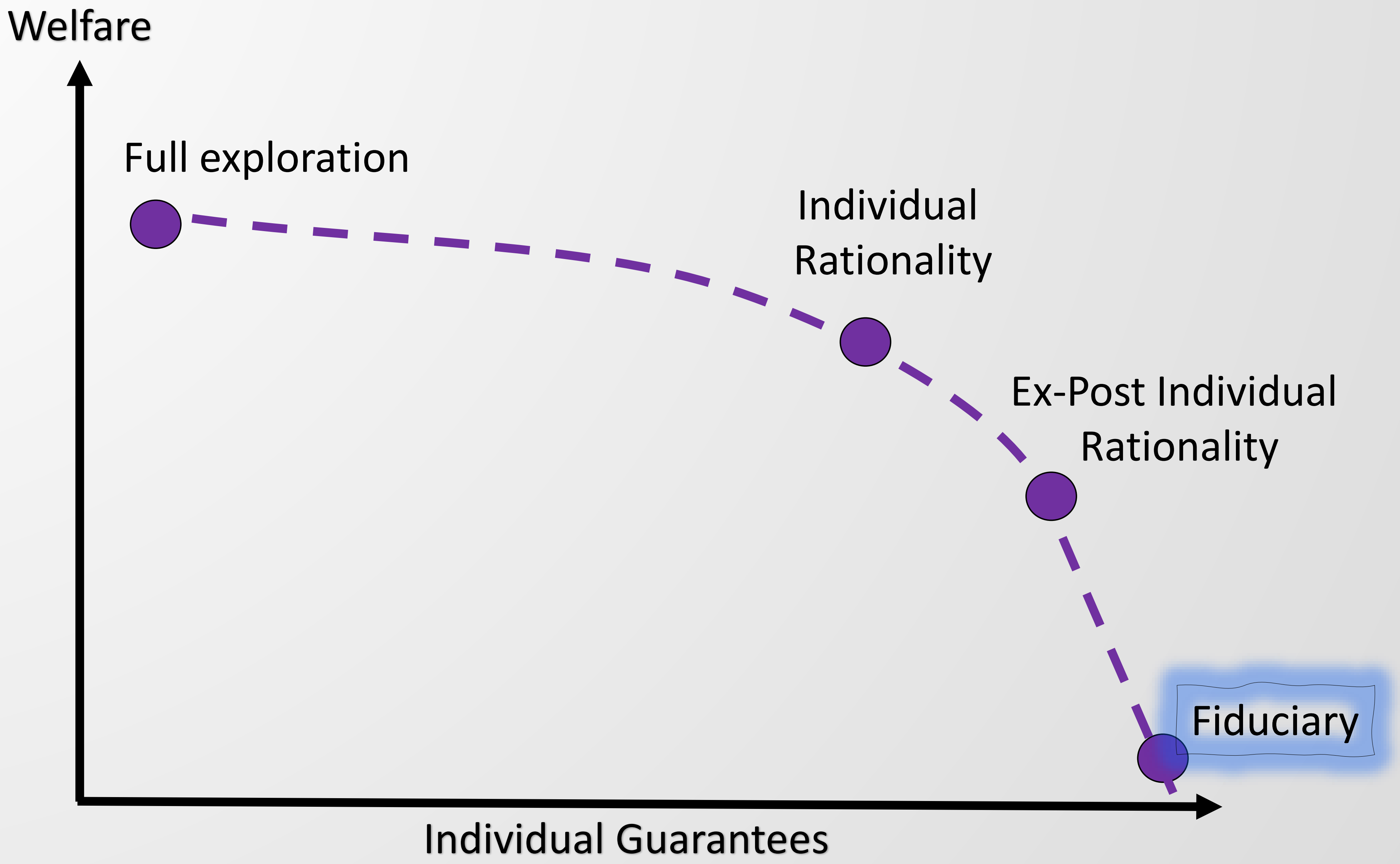

- What are meaningful individual guarantees?
- Optimal algorithms under individual guarantees?
- How much welfare deteriorates under these guarantees?

# Formal Model

- Arms $A = \{a_1, \dots, a_K\}$.
- The reward of $a_i$ is a random variable $X_i$, with $\mu_i = \mathbb{E}(X_i)$. (mutually ind.)
  - W.l.o.g. $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_K$.
- $X_i \in \{0,1, \dots, H\}$ almost surely. (Static)
- $n$ agents, agent $i$ arrives at time $i$. Agents follow the mechanism's action*.
- A mechanism maps histories to (possibly randomized) actions:

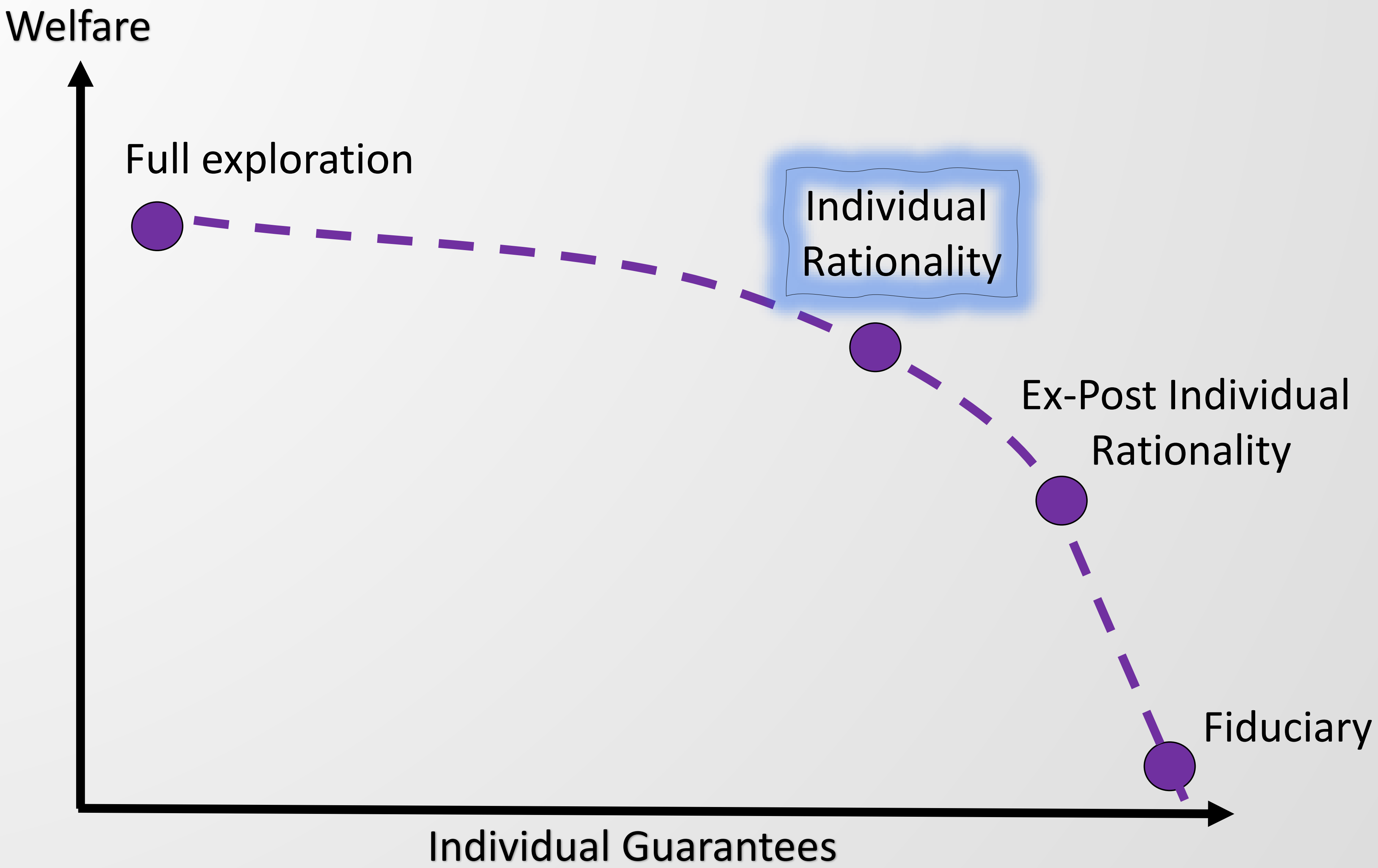
$$M: \bigcup_{l=1}^{n} (A \times \mathbb{R}_+)^{l-1} \to \Delta(A).$$

- Social welfare: $SW_n(M) = \mathbb{E}\left(\frac{1}{n}\sum_{l=1}^{n} X_{M(h_l)}\right)$.


Welfare
Full exploration
Individual Rationality
Ex-Post Individual Rationality
Fiduciary
Individual Guarantees

# Fiduciary

- Wikipedia: "A ***fiduciary*** is a person who holds a legal or ethical relationship of trust with one or more other parties."
- Intuitively: Operate in one's best interest.
- A mechanism $M$ is a fiduciary if for every $l \in \{1, \dots, n\}, h \in (A$


Welfare
Full exploration
Individual Rationality
Ex-Post Individual Rationality
Fiduciary
Individual Guarantees

# Agents Knowledge and Actions

- *Default arm*: the arm the agent would adopt if she doesn't use the mechanism.
- W.l.o.g. arm $a_1$ (Recall that $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_K$).
- A mechanism is *individually rational* if

  *every agent is, in expectation, better off using the mechanism.*
- Formally, $M$ is ***IR*** if for every $l$ and $h$ it holds that

$$\mathbb{E}(X_{M(h)}|h) \geq \mathbb{E}(X_1|h).$$

Using $M$, given the mechanism's knowledge

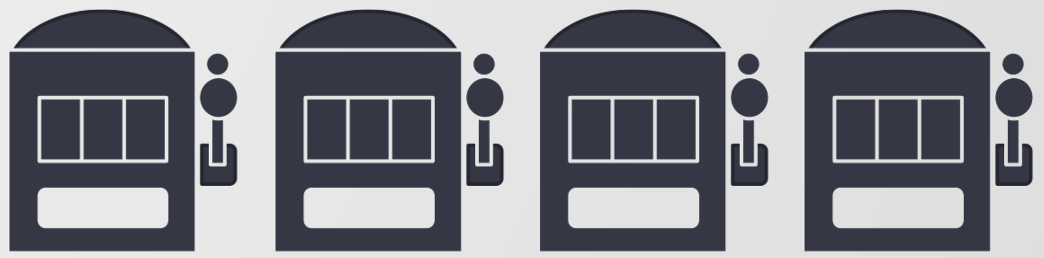
default arm, given the mechanism's knowledge

# Example

IR: $\mathbb{E}(X_{M(h)}|h) \geq \mathbb{E}(X_1|h)$.

- No exploration

$a_1$ $a_2$ $a_3$ $a_4$



~~$\mu_1 = 20$~~ $\mu_2 = 15$ $\mu_3 = 10$ $\mu_4 = 5$

$R_1 = 17$

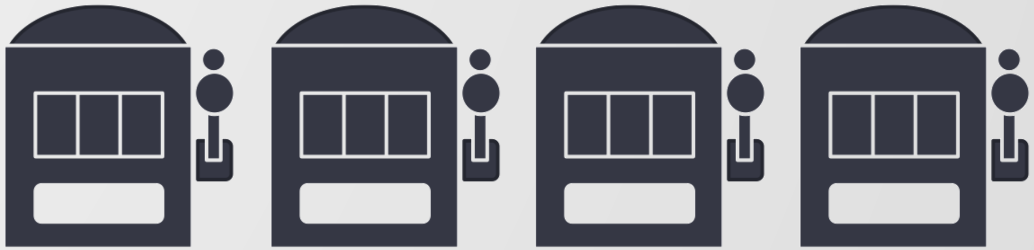
# Example

IR: $\mathbb{E}(X_{M(h)}|h) \geq \mathbb{E}(X_1|h)$.

- Explore $a_2$?
- A mixture of all remaining arms?
- $\frac{\mu_2}{2} + \frac{\mu_3}{2} = 12.5 > R_1$

- **Challenge**: Maximize welfare!
- Wrong exploration policy $\Rightarrow$ sub-optimal welfare.

$a_1$ $a_2$ $a_3$ $a_4$



~~$\mu_1 = 20$~~ $\mu_2 = 15$ $\mu_3 = 10$ $\mu_4 = 5$

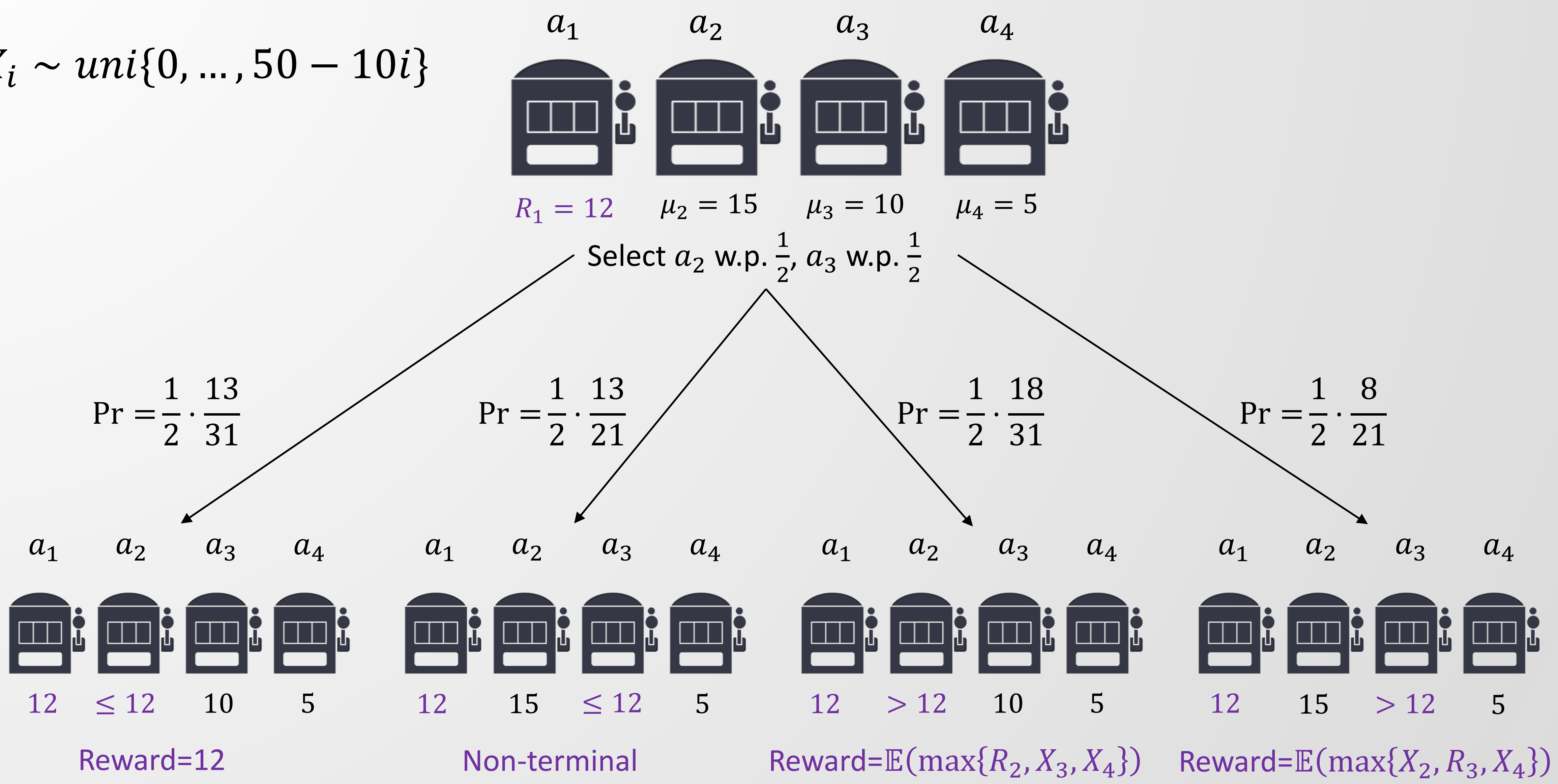
$R_1 = 12$

# Using Exploration Oracle

- Any exploration-seeking mechanism will hit one of these states:
  - An arm with a value $> R_1$ was found. (Jackpot)
  - All observed reward $\leq R_1$, all unobserved $\mu_i < R_1$. (Failure)
- Jackpot$\Rightarrow$ Explore all arms in *reasonable* time.
- Failure $\Rightarrow$ Select $a_1$.
- Everything boils down to the first $K$ agents
  - A Markov Decision Process (MDP) with continuum of actions.

Elaborate

$X_i \sim uni\{0, \ldots, 50 - 10i\}$



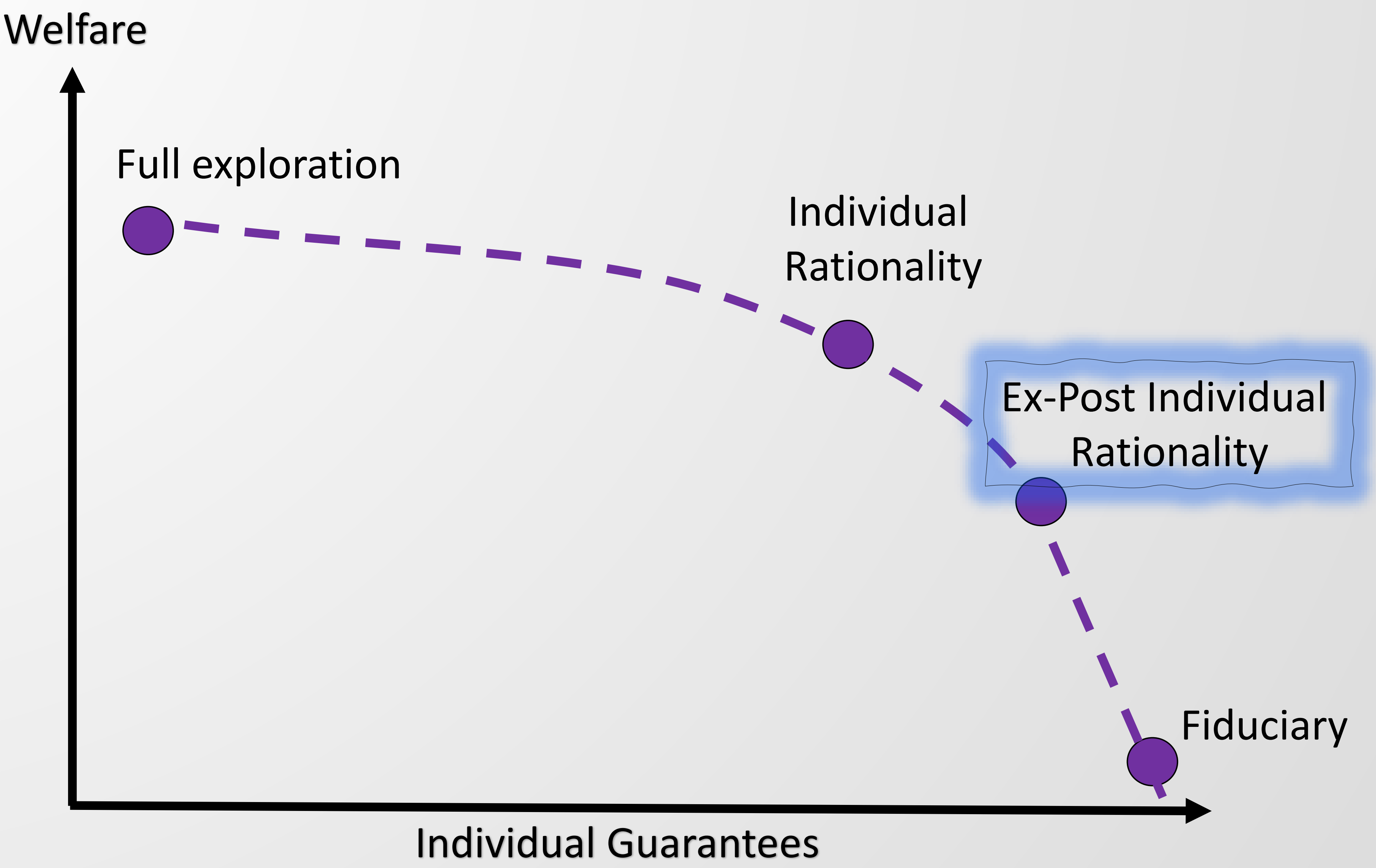
# Asymptotically Optimal IR Algorithm

1. Offline: Compute the optimal policy $\pi^*$ of the MDP.
2. While not hitting a terminal state: (Jackpot or Failure)
   - Select according to $\pi^*$.
3. If a superior arm is discovered: (Jackpot)
   - Mix that arm with an unobserved arm until all arms are observed.
   - From here on, exploit the best arm.
4. Else: (Failure)
   - From here on, select the default arm.

Computing $\pi^*$:
- Explores two arms at a time.
- Runtime: $O(2^K K \min\{K, H\})$.
- Pros/cons.

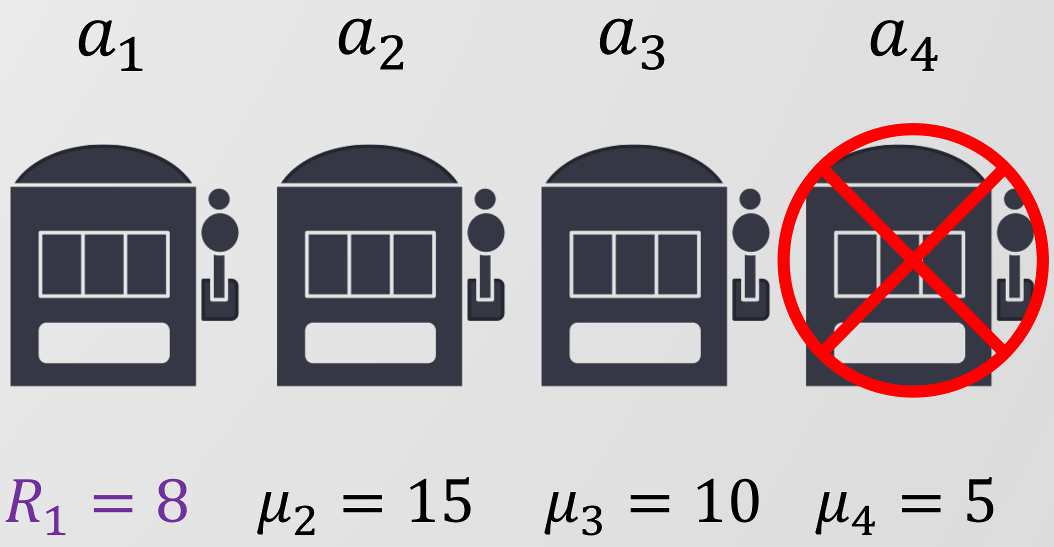
➢Theorem: For every $n$, it holds that

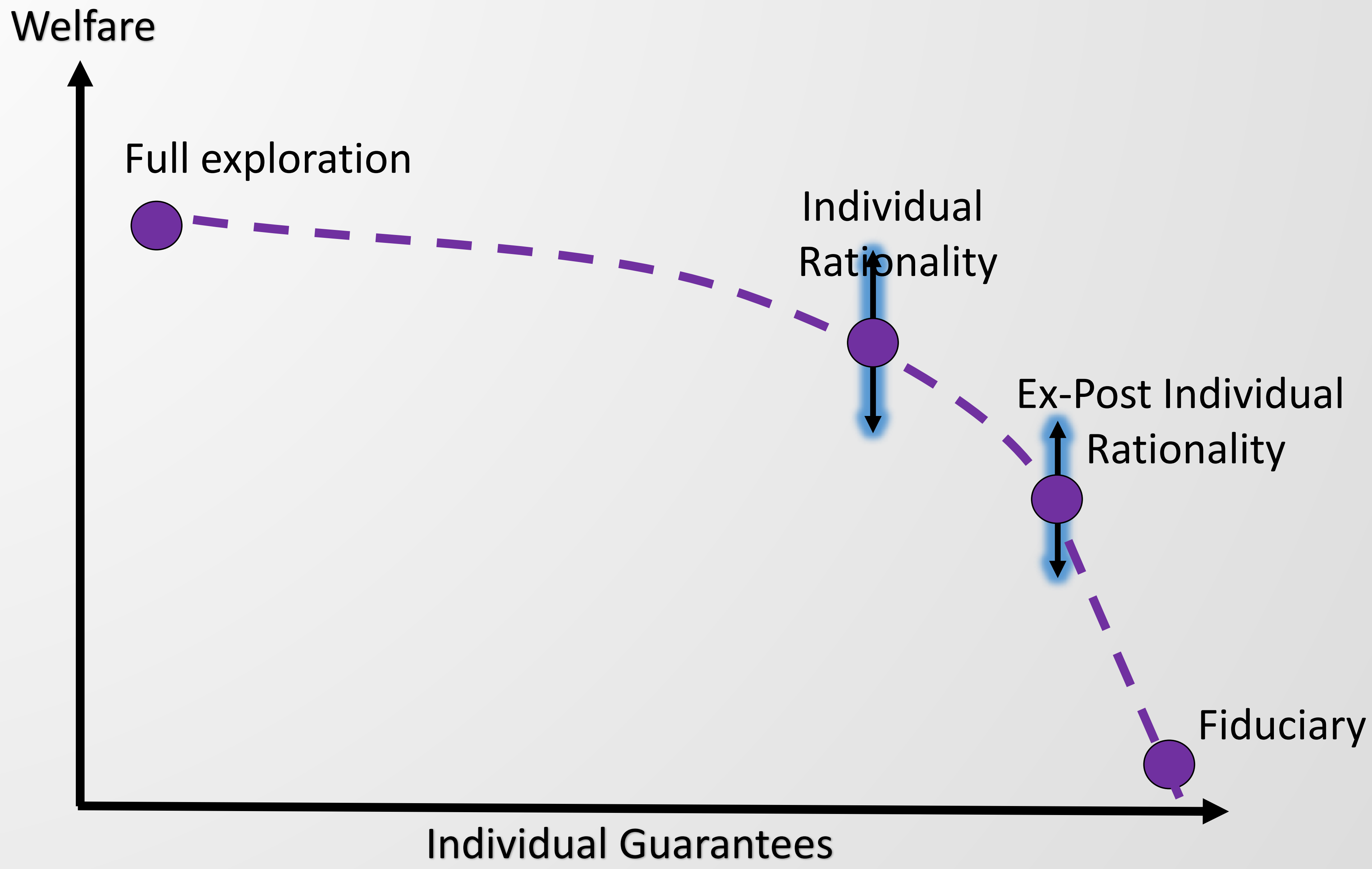
$$SW_n(ALG) \geq \sup_{M, M \text{ is IR}} SW_n(M)\left(1 - \frac{(K+1)H}{n}\right).$$


Welfare
Full exploration
Individual Rationality
Ex-Post Individual Rationality
Fiduciary
Individual Guarantees

# Ex-Post Individual Rationality

- A stronger individual guarantee: **IR** with forbidden lotteries.
- $M$ is Ex-Post Individually Rational if for every $l$ and $h$, if $\Pr_{M(h)}(a_i) > 0$, then $\mathbb{E}(X_i|h) \geq \mathbb{E}(X_1|h)$.




Welfare
Full exploration
Individual Rationality
Ex-Post Individual Rationality
Fiduciary
Individual Guarantees

# Social Welfare Analysis

➢ $OPT, OPT_{\mathrm{IR}}, OPT_{\mathrm{EPIR}}$

1. There is an instance such that
$$\frac{OPT}{OPT_{\mathrm{IR}}} \geq H\left(1 - e^{-\frac{K}{H}}\right).$$
2. If $X_1 \sim Uni[H] + \epsilon$ (for $\epsilon \rightarrow 0$) and $X_i \sim Uni[H]$ for every $i$, then
$$\frac{OPT}{OPT_{\mathrm{IR}}} \leq \frac{8}{7}.$$
3. There is an instance such that
$$\frac{OPT_{\mathrm{IR}}}{OPT_{\mathrm{EPIR}}} \geq \frac{H+2}{3}\left(1 - e^{-\frac{K}{H}}\right).$$

# Incentive Compatibility

- Assume that the mechanism only *recommends* which arm to use, but it is up to the agents to decide.
  - Kremer, Mansour and Perry (2014), but with $K \geq 2$ arms.
- A mechanism is *incentive compatible* if adopting the recommendation is a dominant strategy of every agent.
- **Theorem**: If agents' arrival is uniform, the proposed optimal IR mechanism is incentive compatible.

# Conclusions and Discussion

- Individual guarantees for the explore-exploit tradeoff.
- Optimal/asymptotically optimal algorithms.
- IC under uniform arrival.
- Open problem: For IR, could we compute $\pi^*$ in $poly(H, K)$?
- Future work: Extend these notions to stochastic arms/non-stationary rewards.

# Related work

- MAB with strategic agents
  - “Implementing the wisdom of the crowd”, Kremer, Mansour and Perry (2014).
  - Extensions to regret minimization, social networks, heterogeneous agents, monetary incentives, etc.
- Fair treatment of arms
  - “Calibrated fairness in bandits”, Liu et al. (2017).
  - “Fairness in learning: Classic and contextual bandits” Joseph et al. (2016).
- Fairness in ML, Safe Reinforcement Learning.
  - “Linear Stochastic Bandits Under Safety Constraints”, Amani et al. (2019).

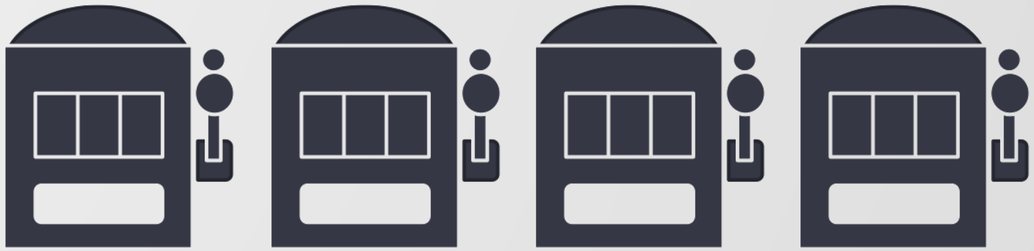
# Kremer, Mansour and Perry (2014)

- Two static arms
- Several agents will obtain

$$pR_1 + (1 - p)\mu_2 < R_1$$

- $\Rightarrow$ Not **IR**

$a_1$ $a_2$ $a_3$ $a_4$



~~$\mu_1 = 20$~~ $\mu_2 = 15$ $\mu_3 = 10$ $\mu_4 = 5$

$R_1 = 17$