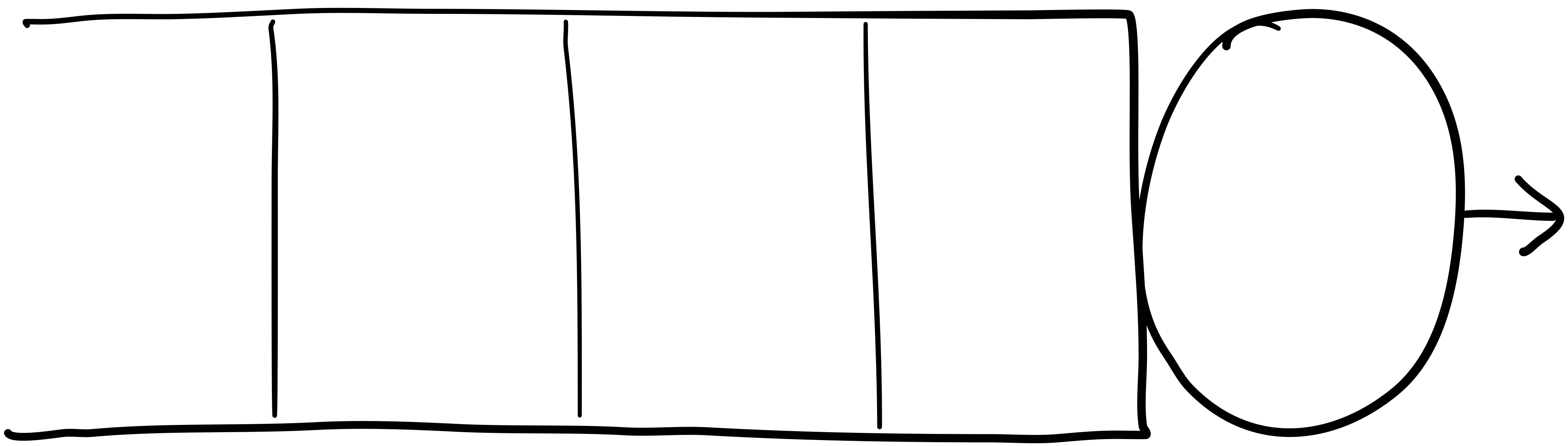
Simple Near-Optimal Scheduling

for the M/G/1

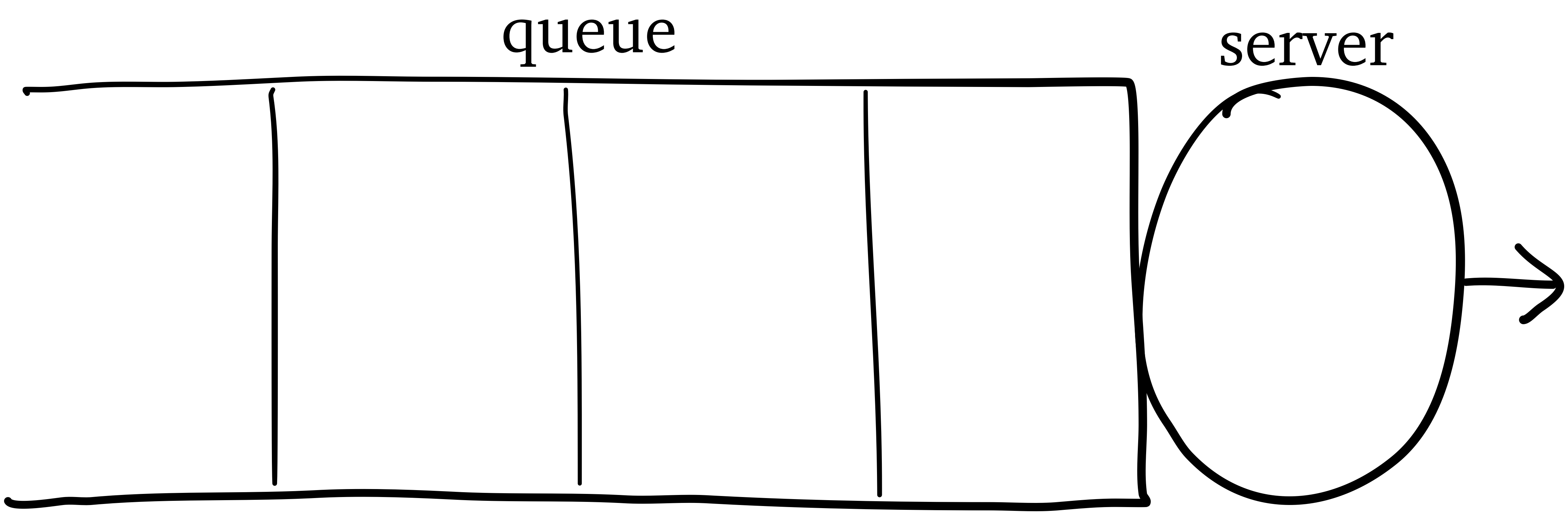
Ziv Scully
Mor Harchol-Balter
Alan Scheller-Wolf

Carnegie Mellon University

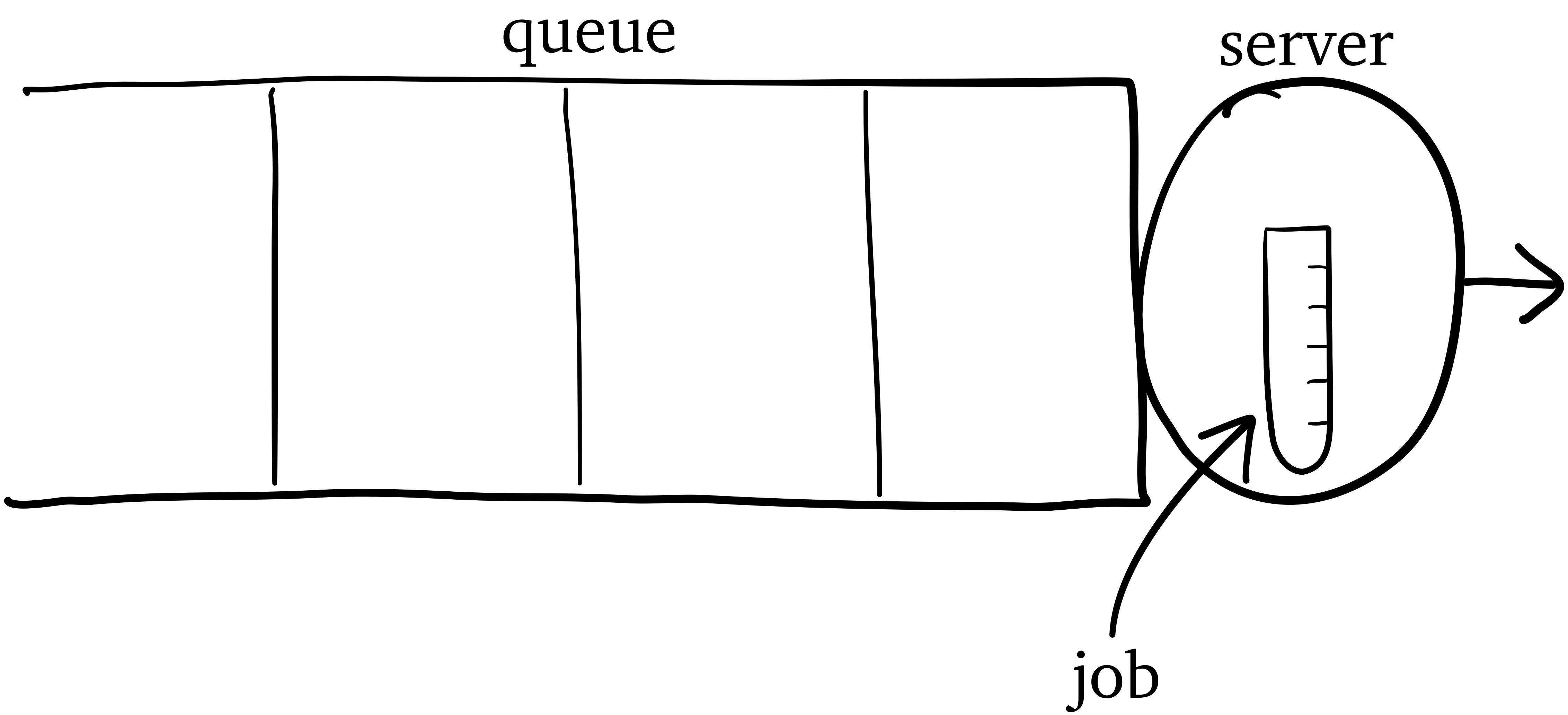
M/G/1 Queue



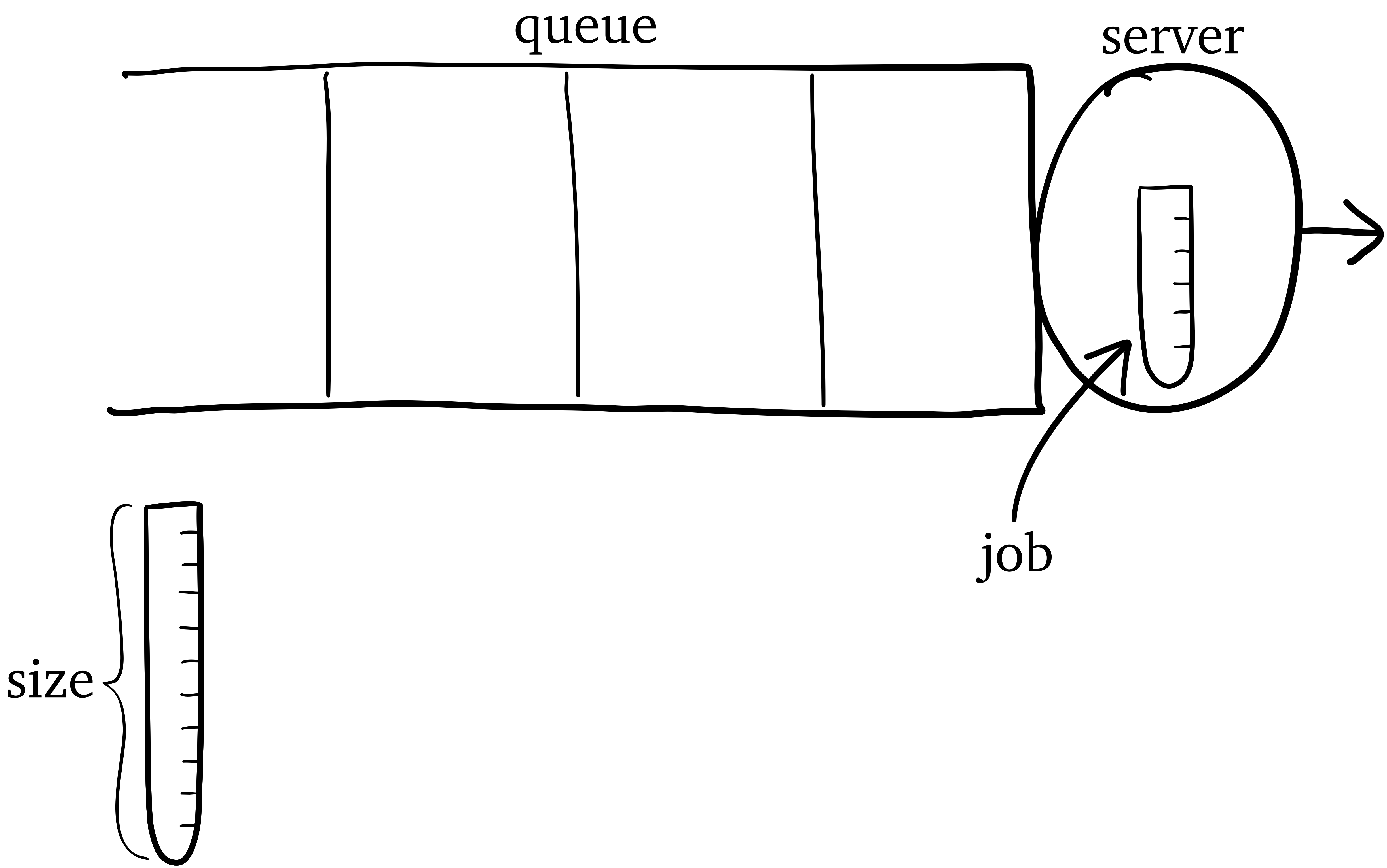
M/G/1 Queue



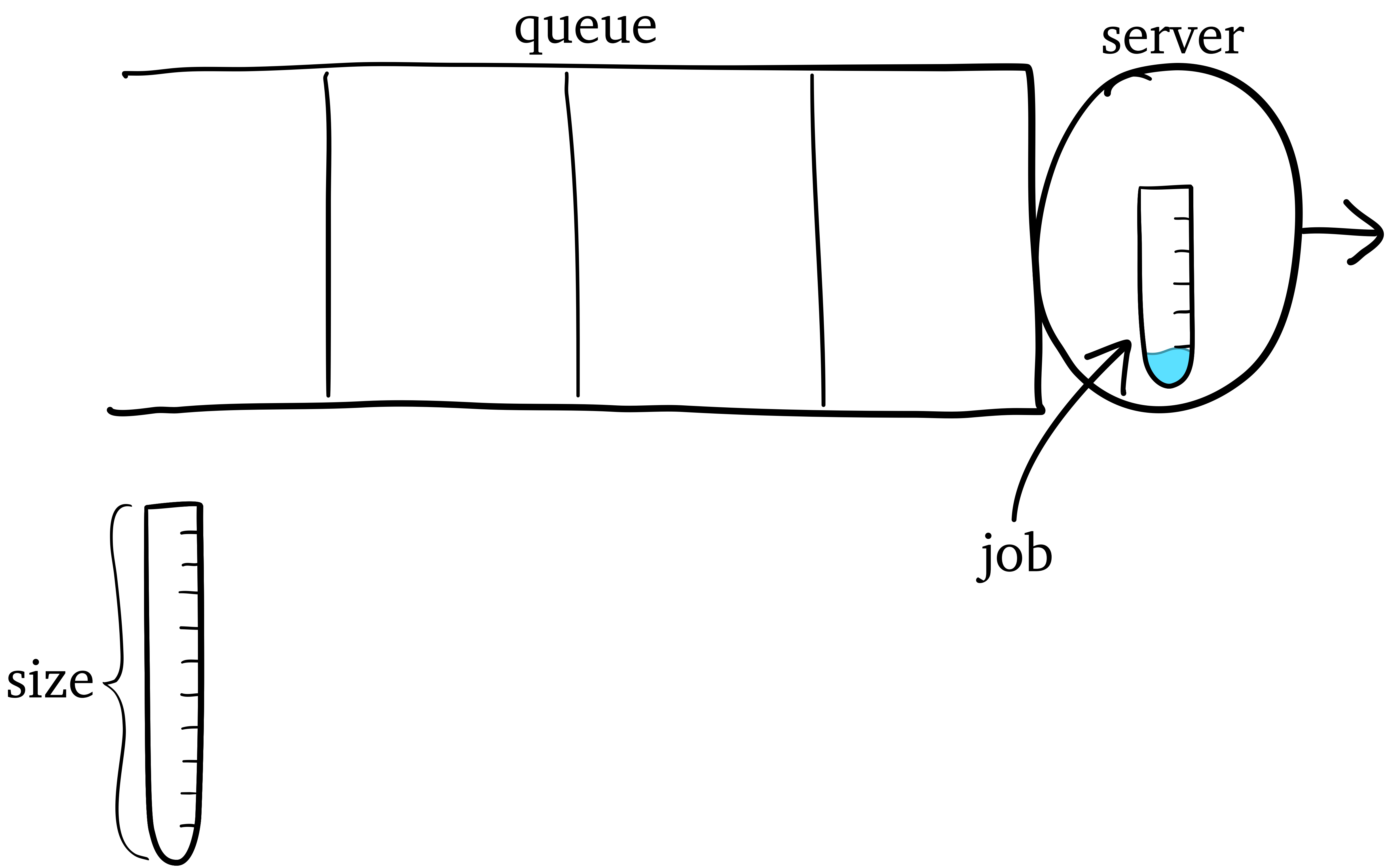
M/G/1 Queue



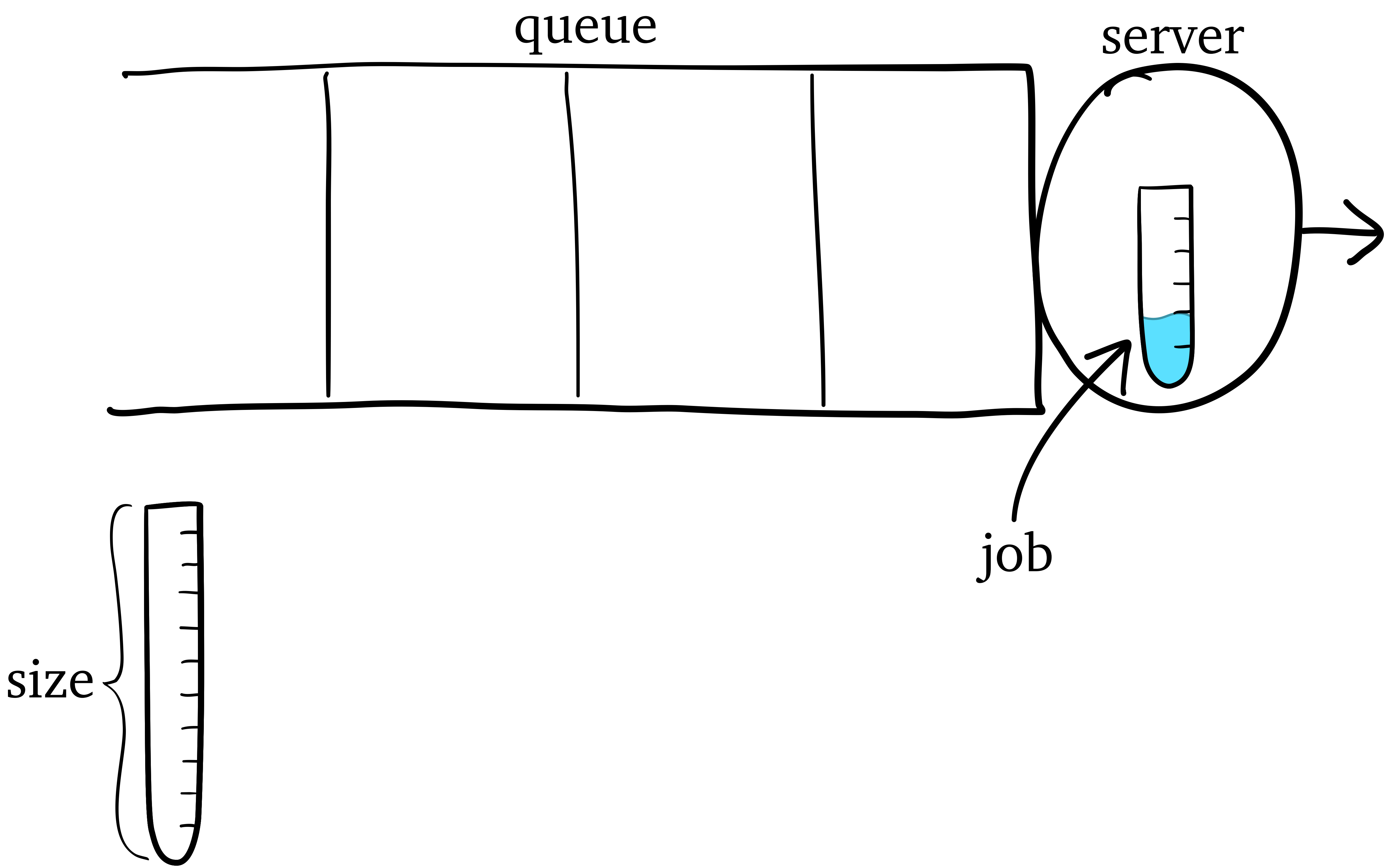
M/G/1 Queue



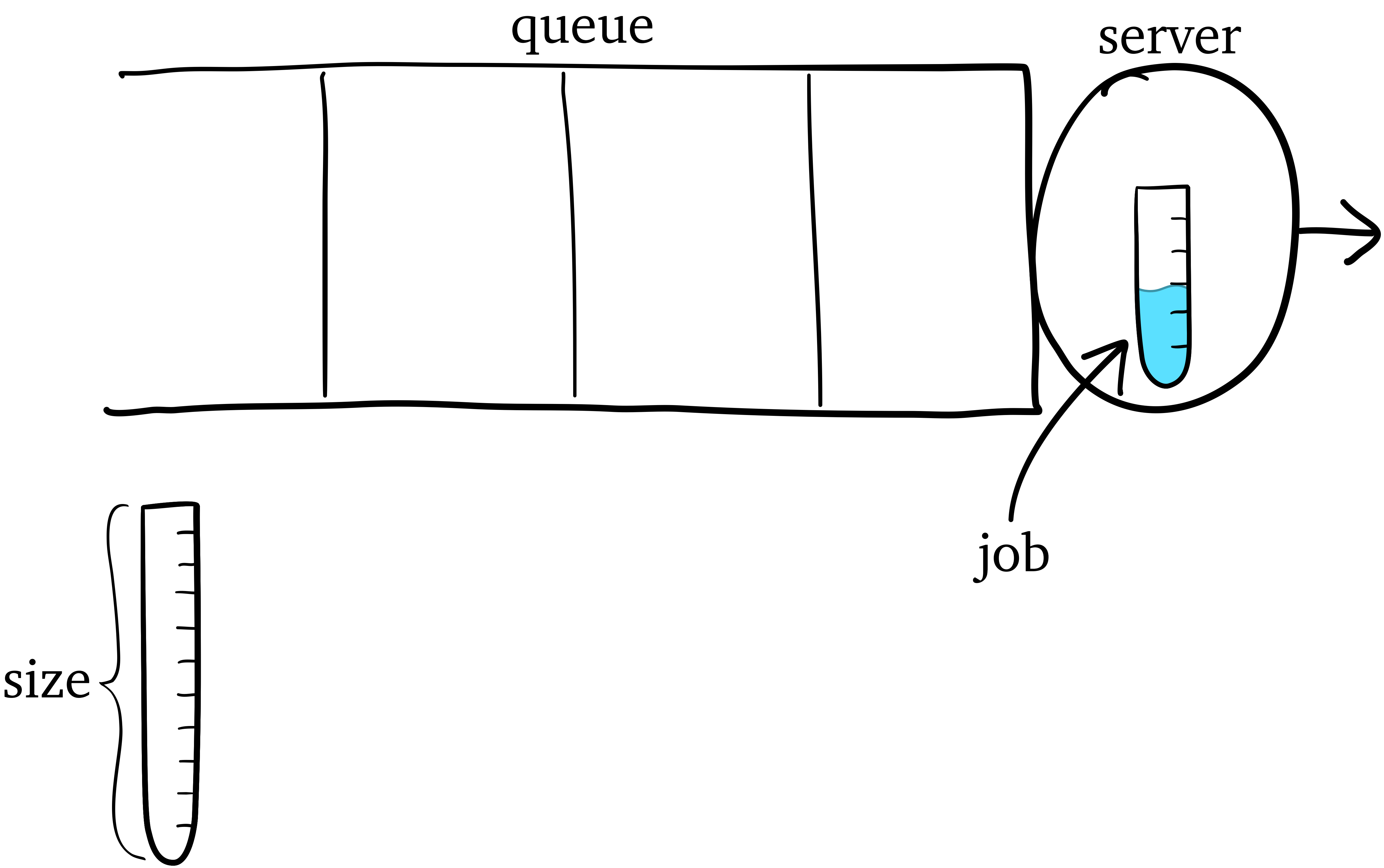
M/G/1 Queue



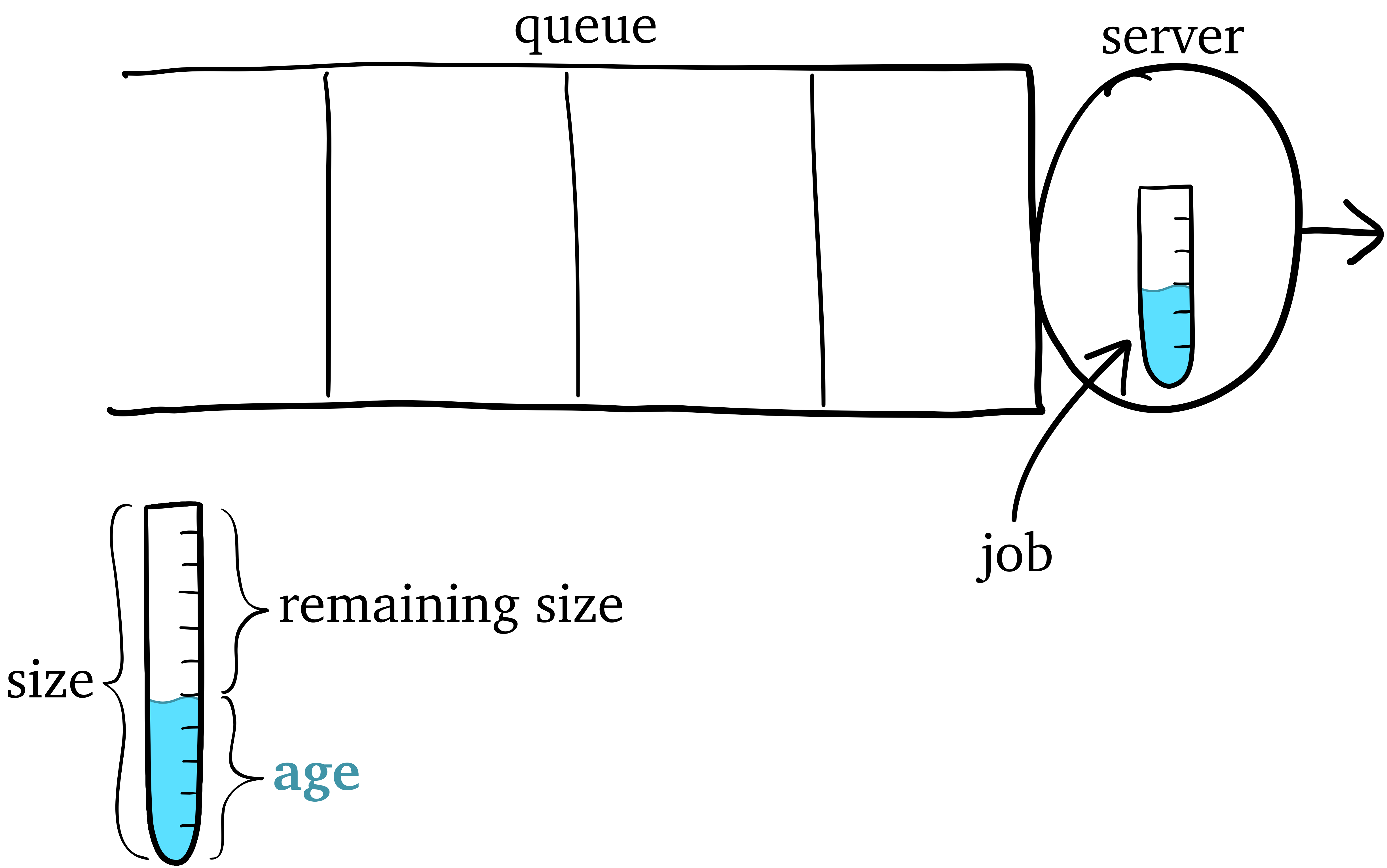
M/G/1 Queue



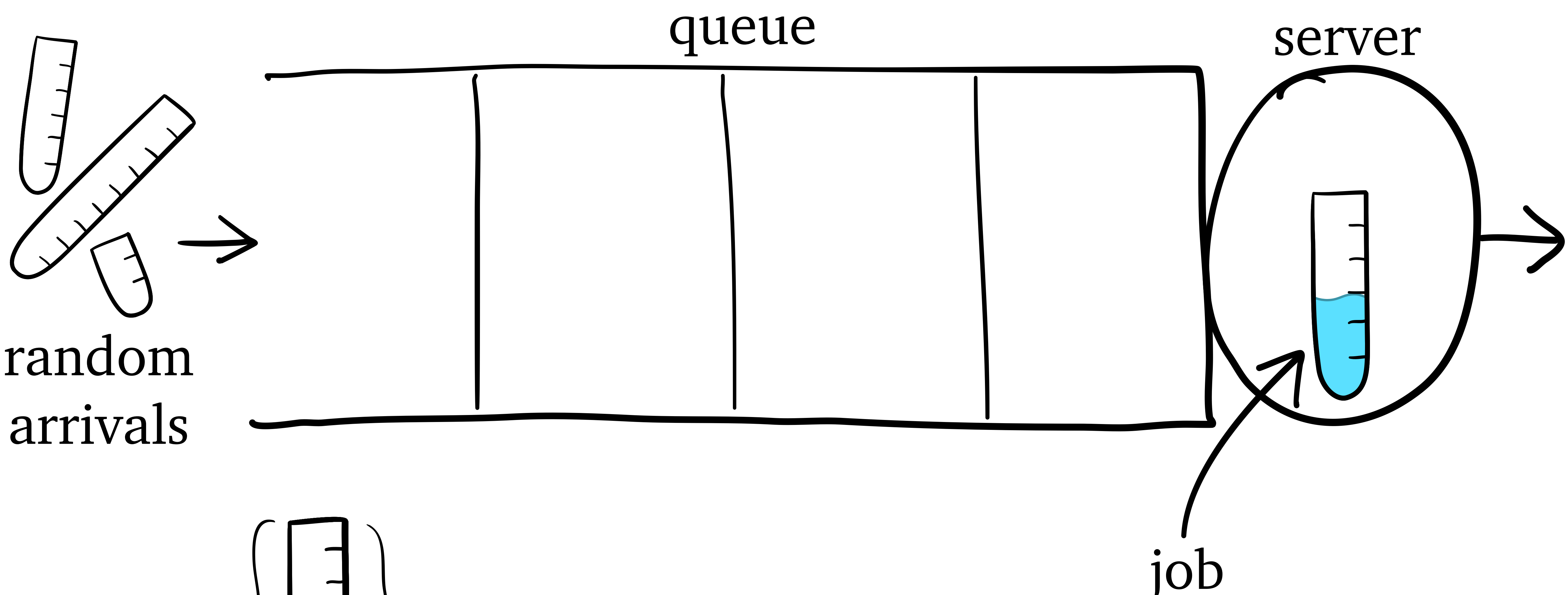
M/G/1 Queue

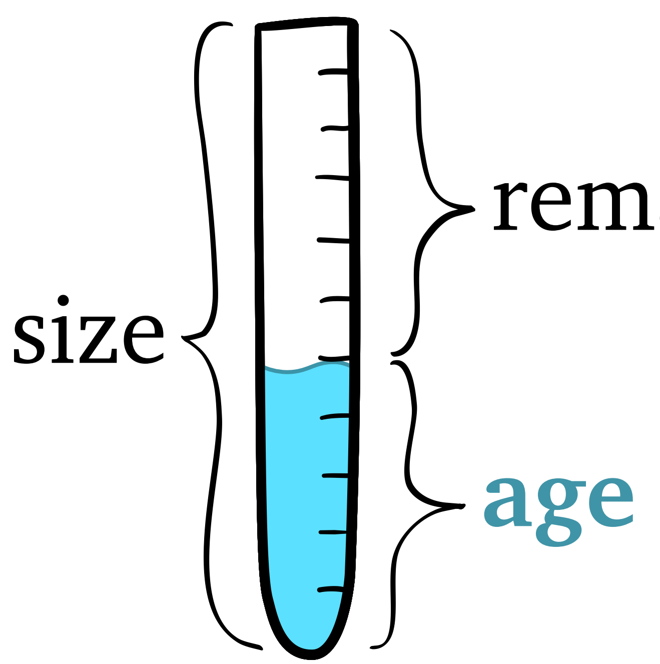


M/G/1 Queue

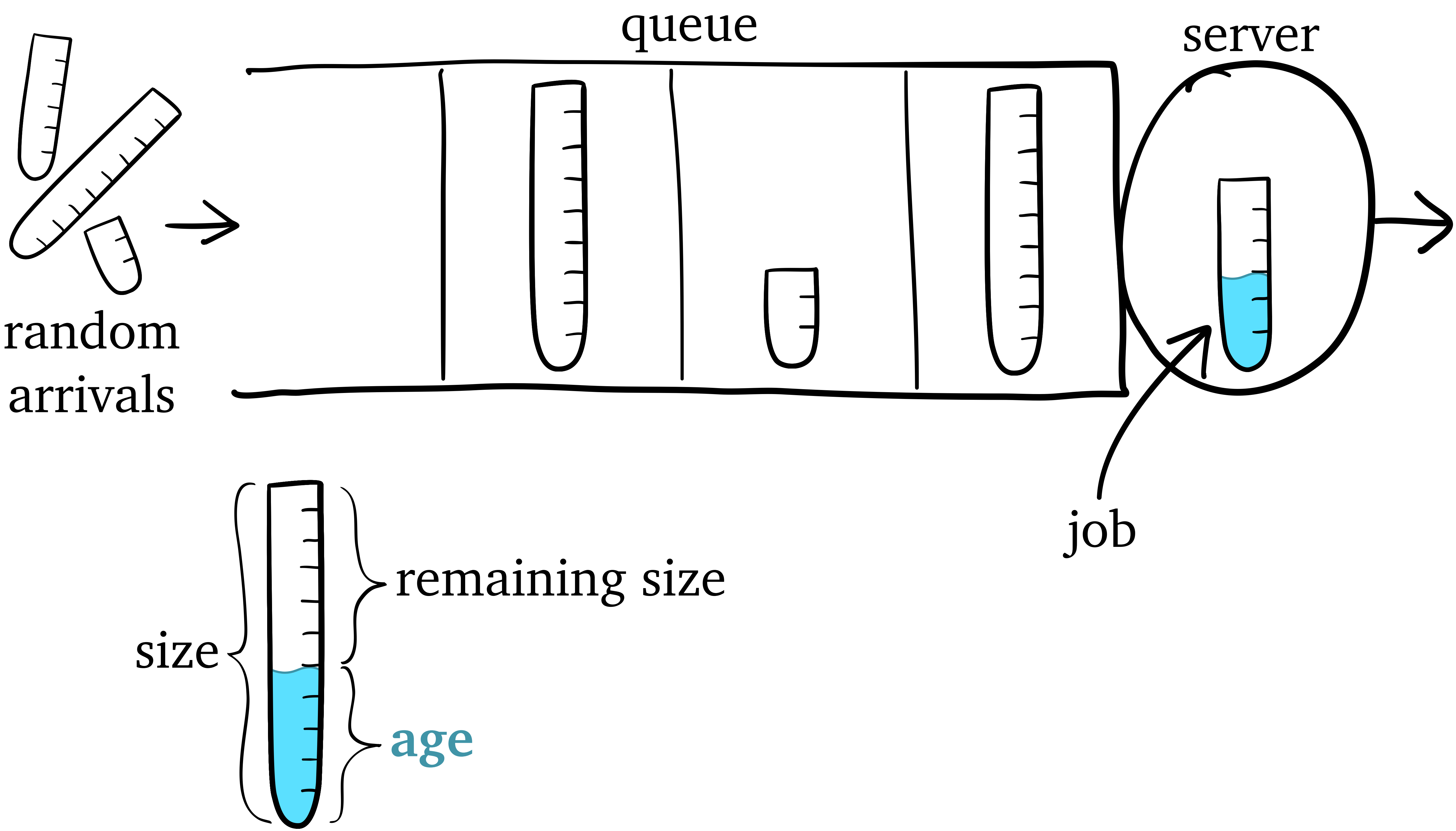


M/G/1 Queue

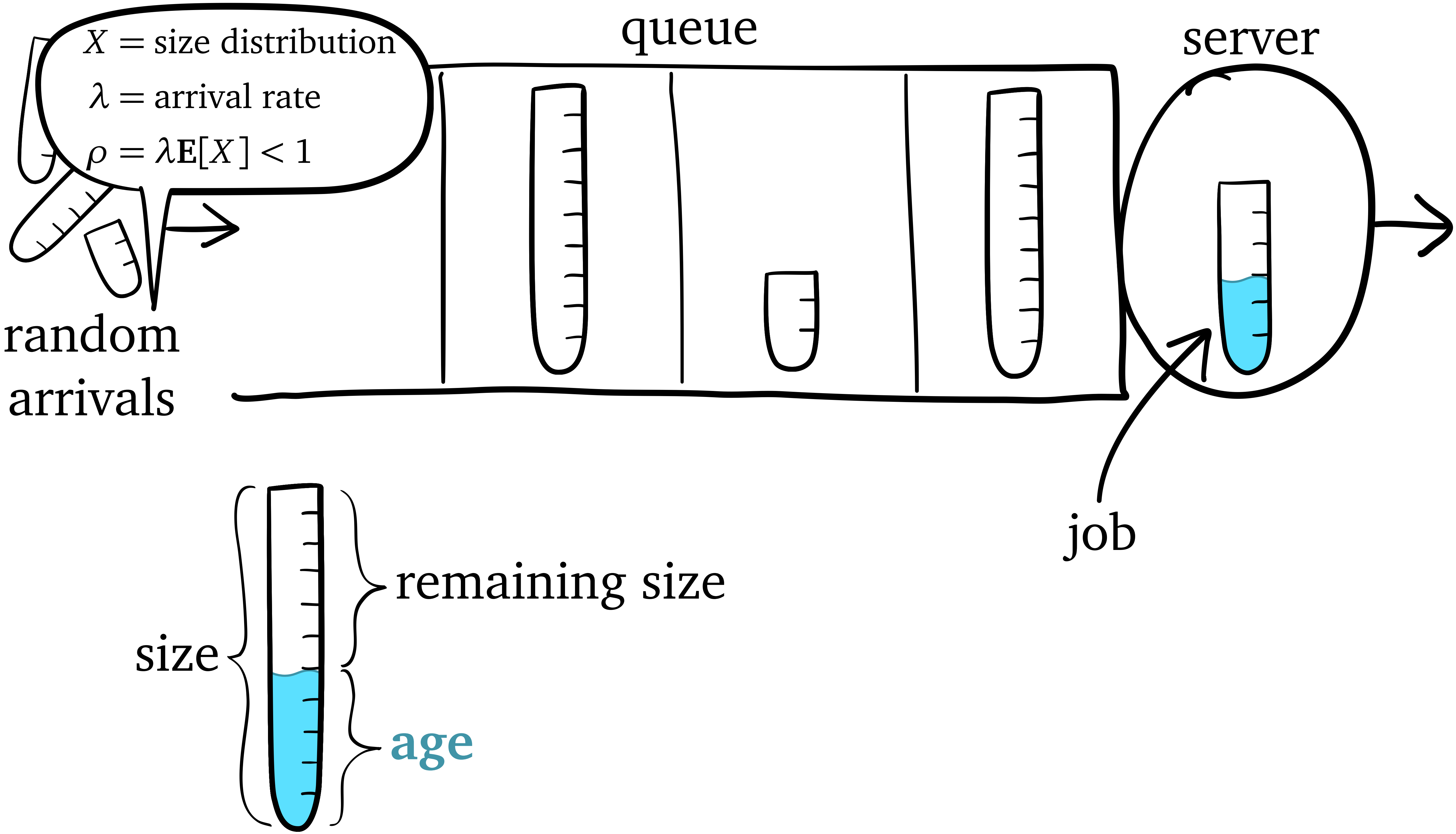




M/G/1 Queue

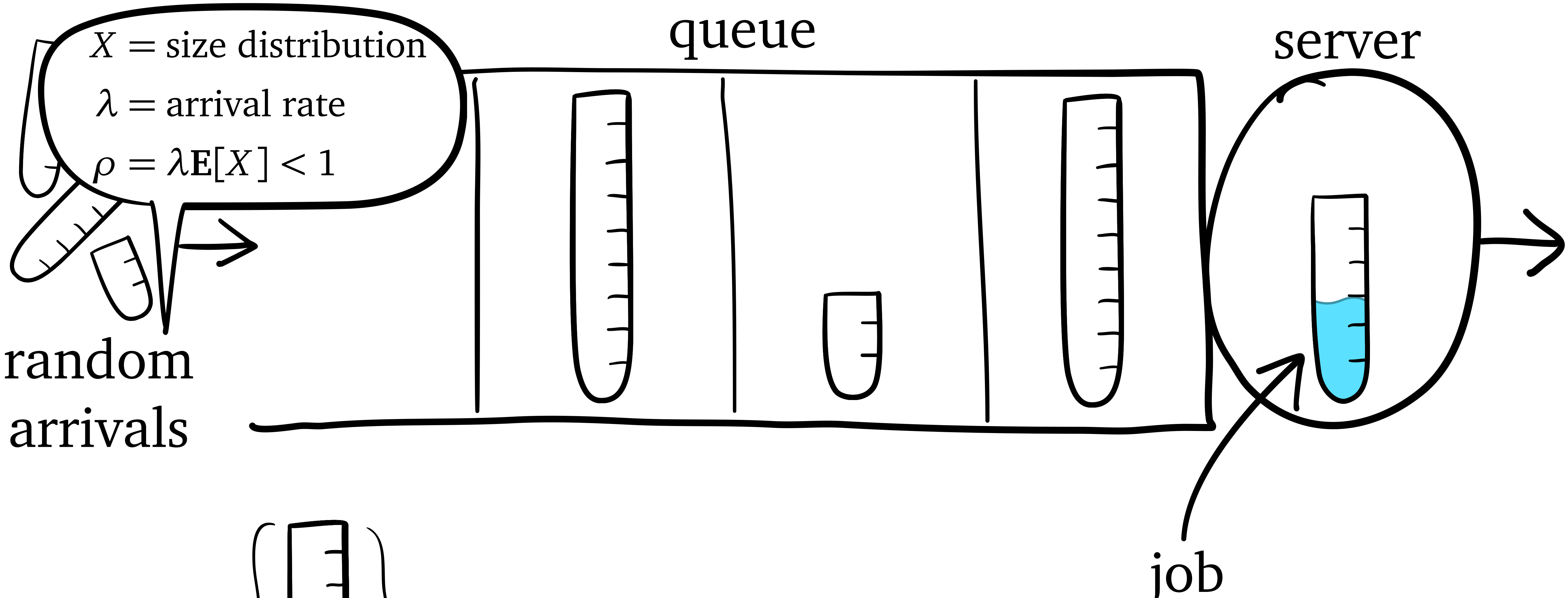


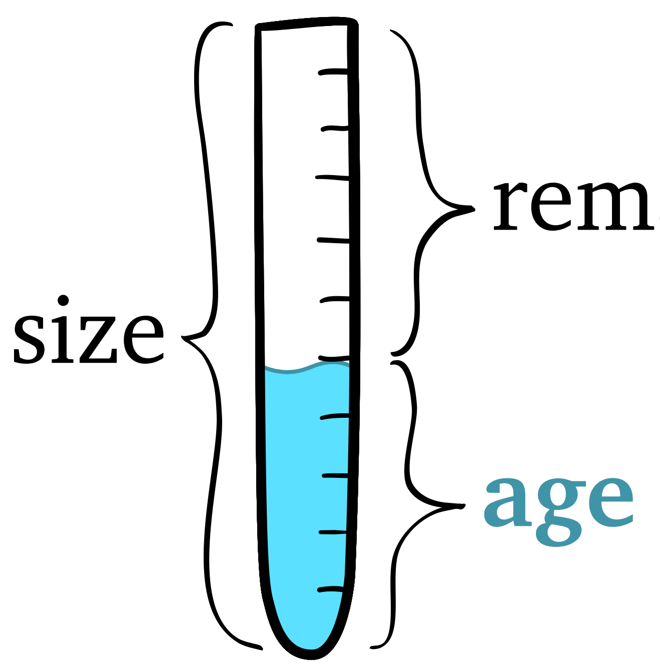
M/G/1 Queue

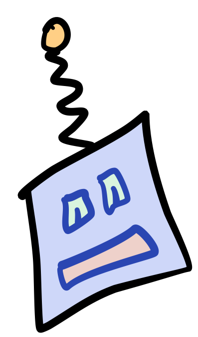


X = size distribution
λ = arrival rate
$\rho = \lambda \mathbf{E}[X] < 1$
queue
server
random arrivals
job
size
remaining size
age

M/G/1 Queue

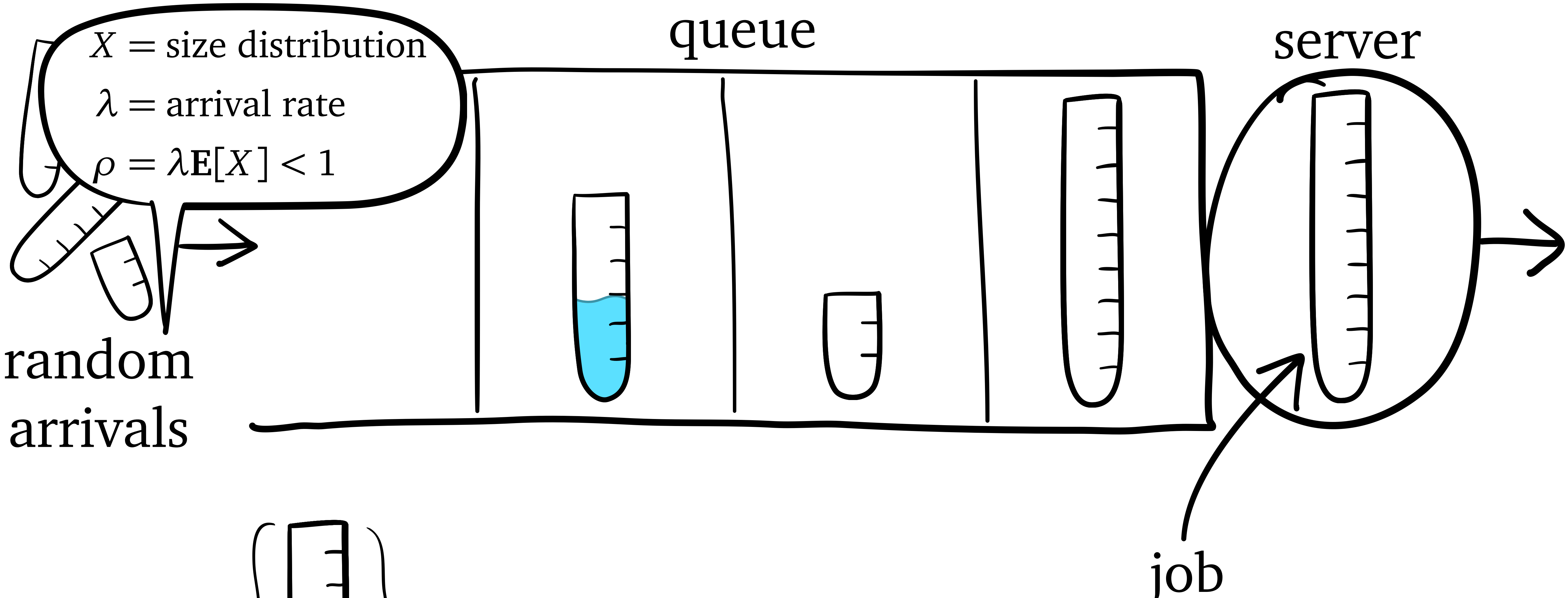


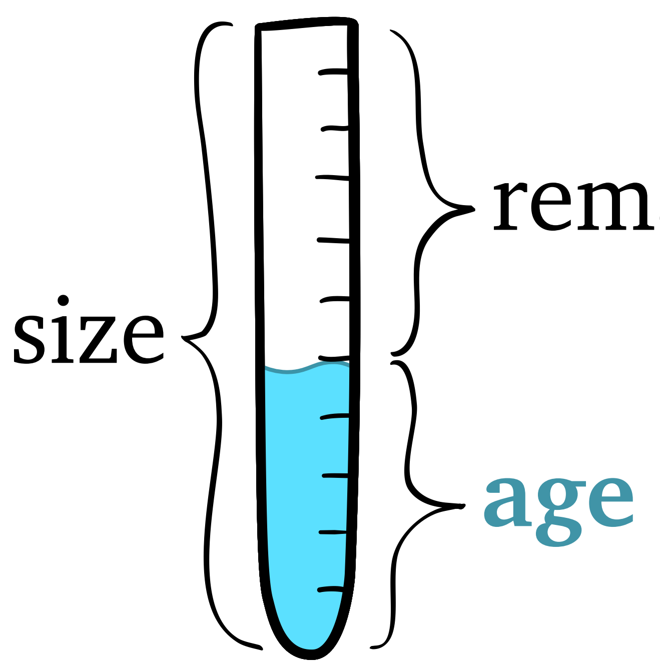


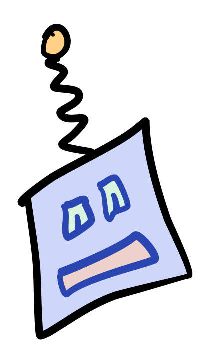


Scheduling policy:
picks which job to serve

M/G/1 Queue

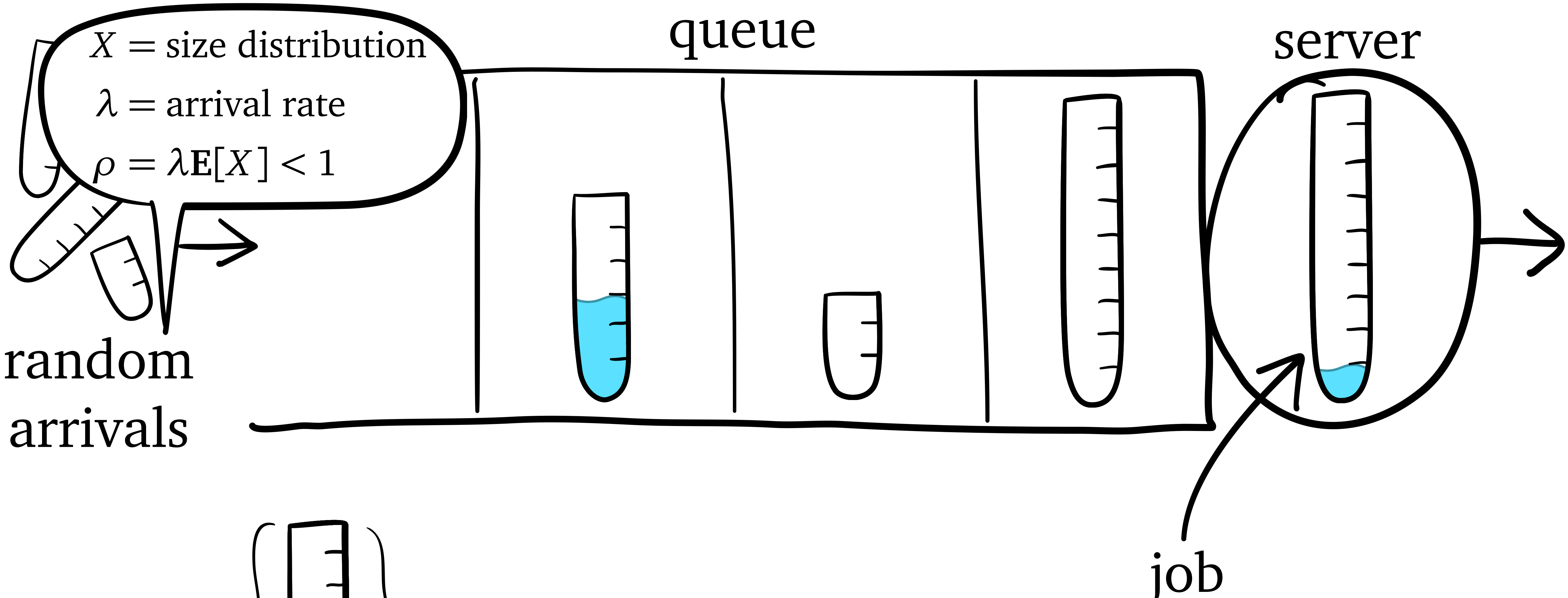


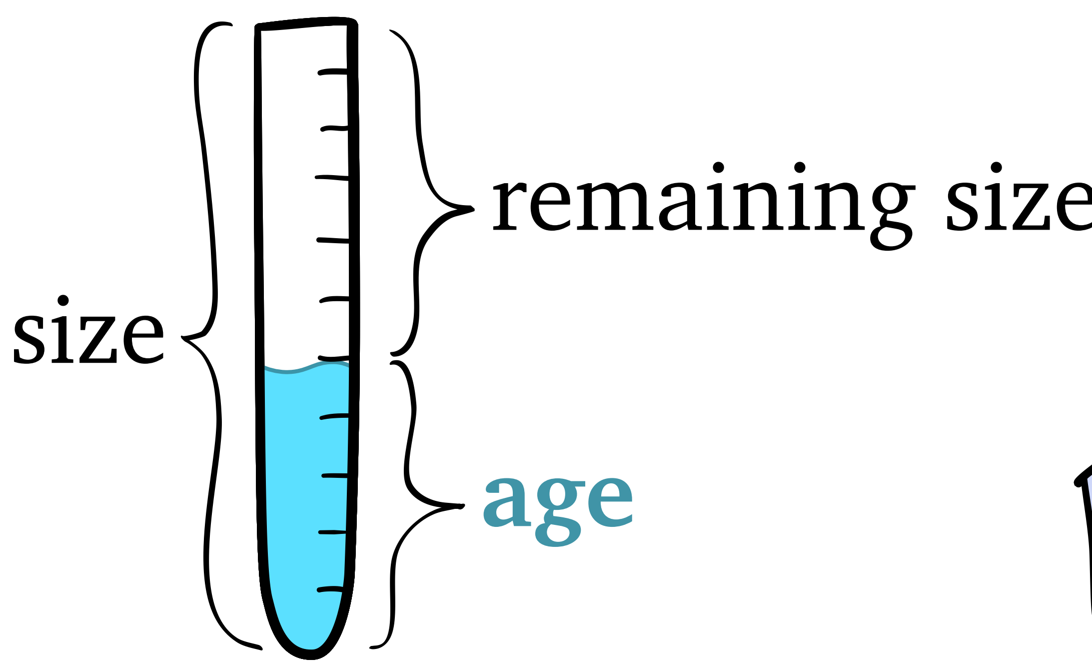


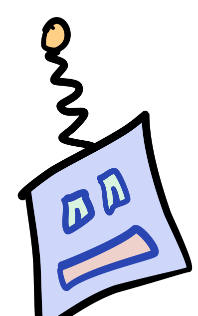


Scheduling policy:
picks which job to serve

M/G/1 Queue

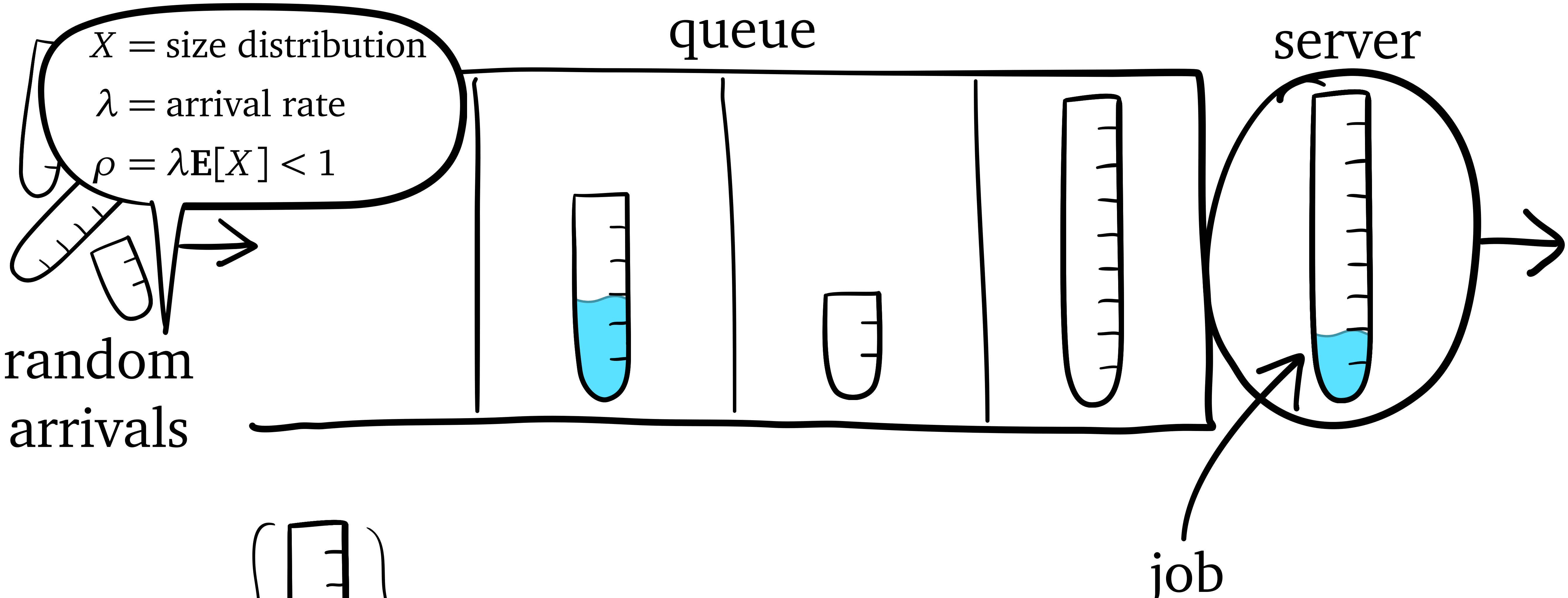


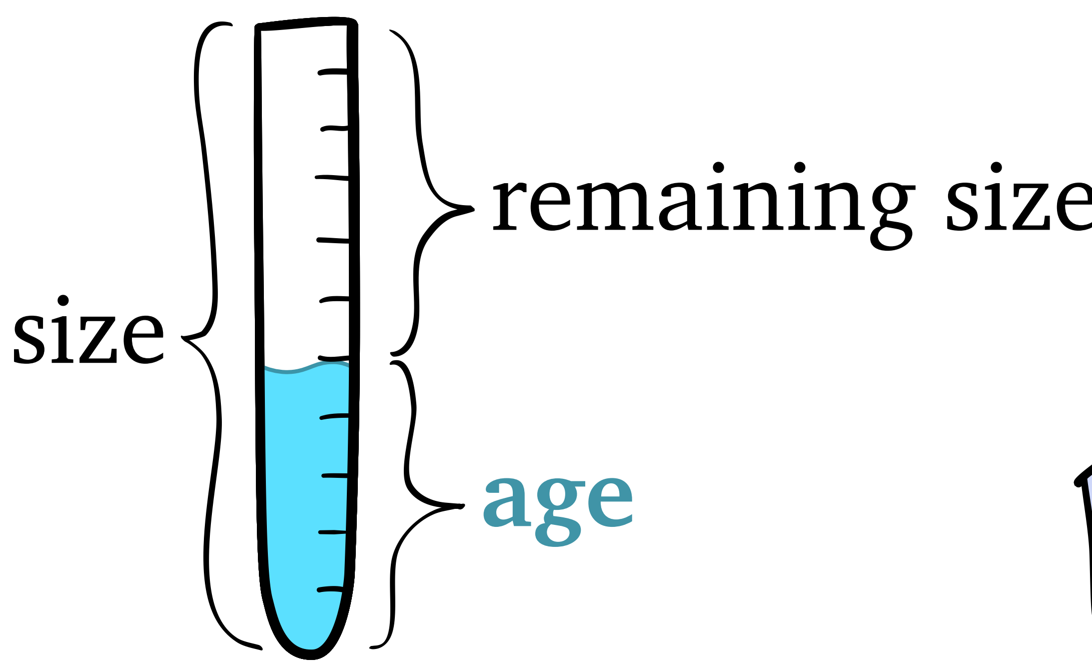


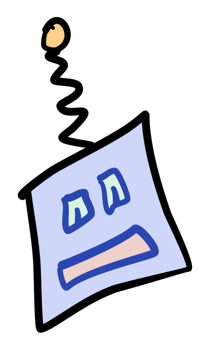


Scheduling policy:
picks which job to serve

M/G/1 Queue

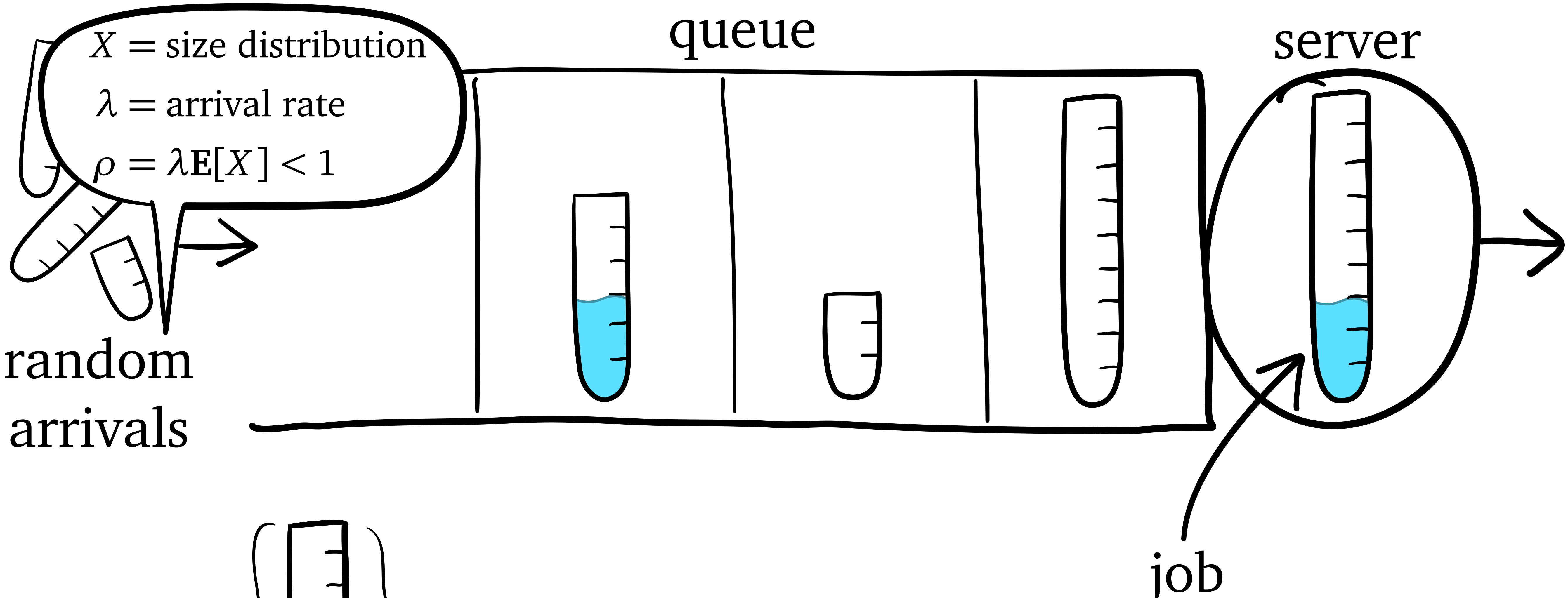


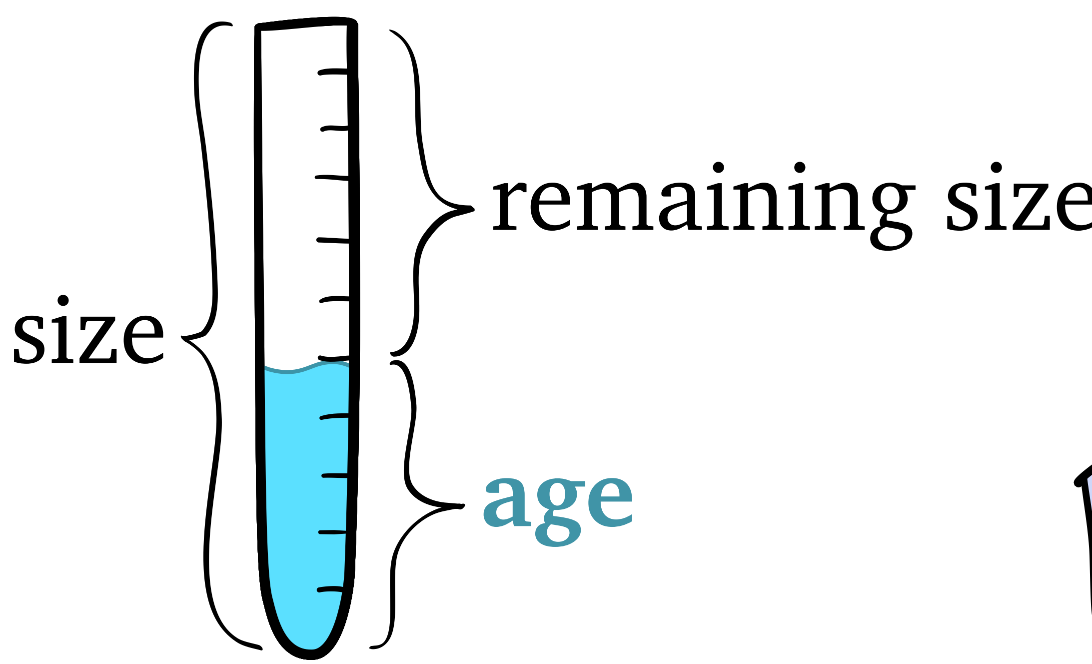


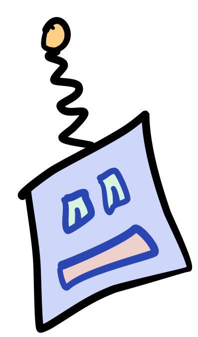


Scheduling policy:
picks which job to serve

M/G/1 Queue

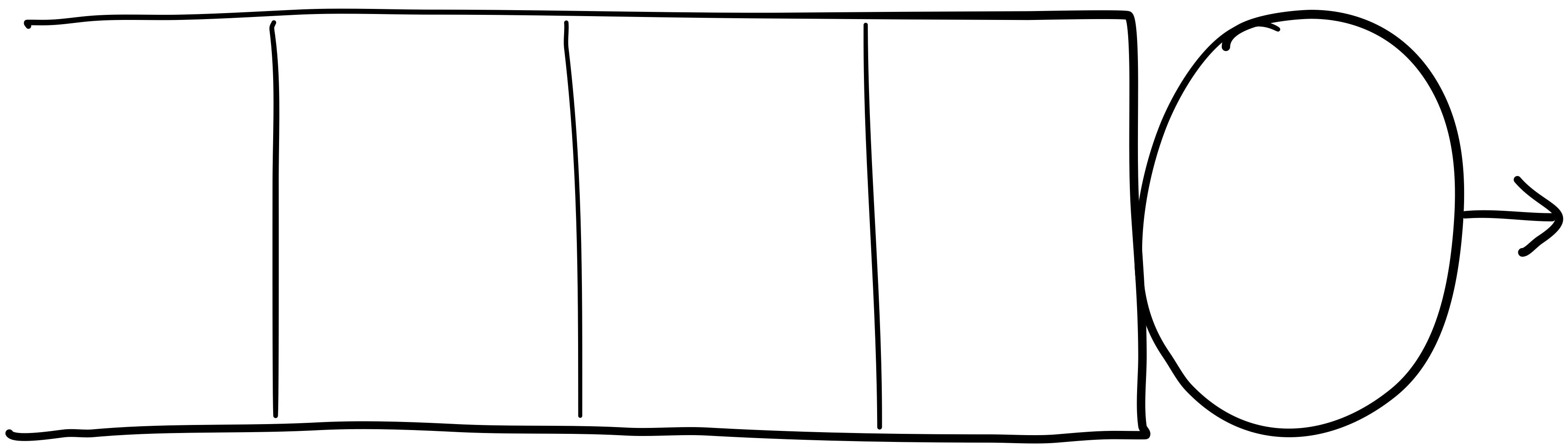




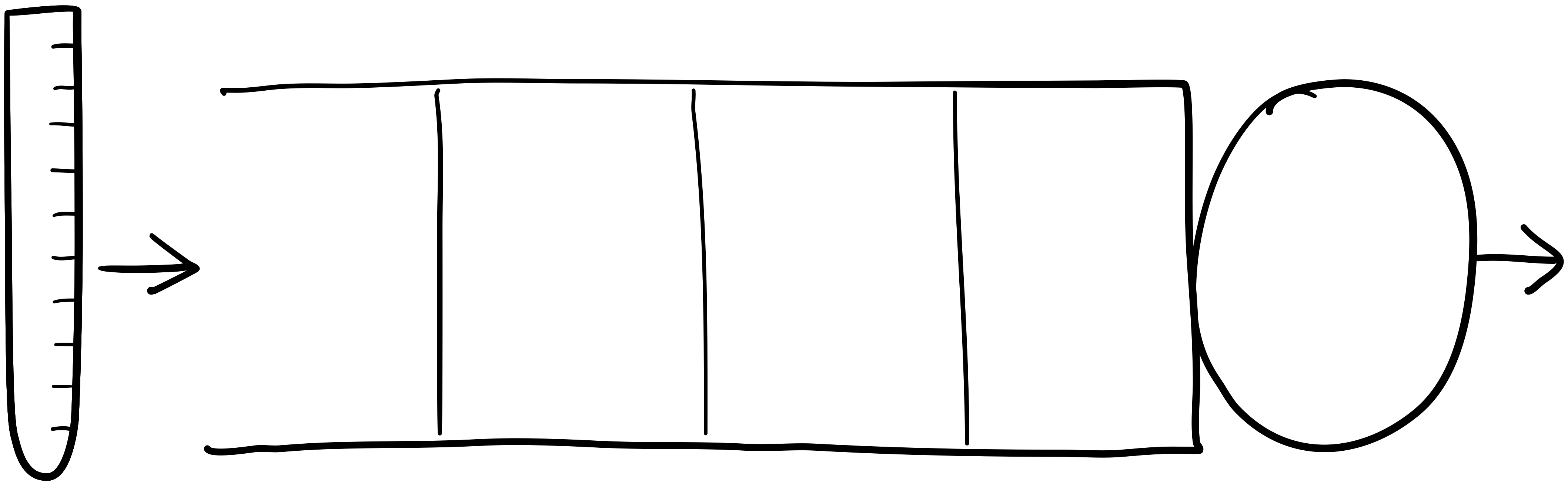


Scheduling policy:
picks which job to serve

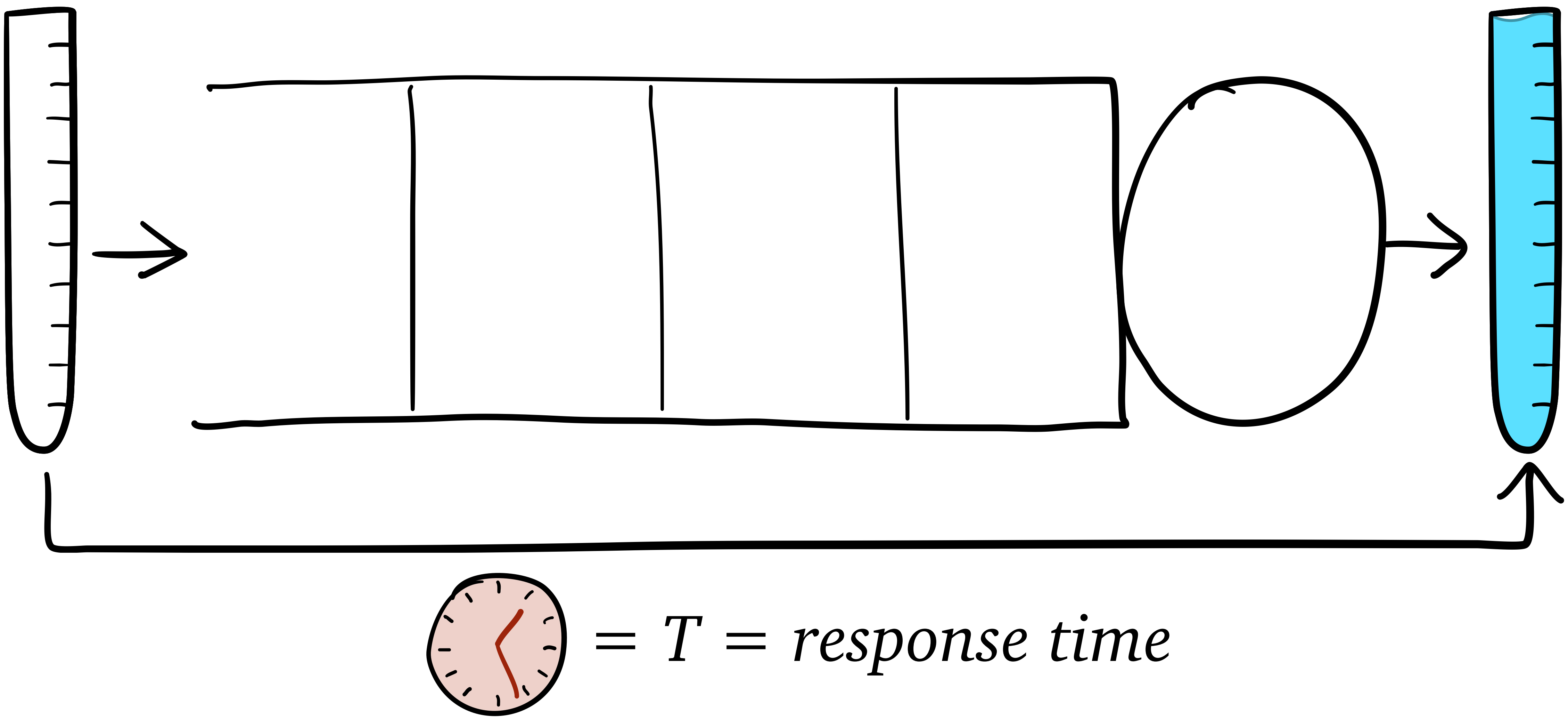
Response Time



Response Time

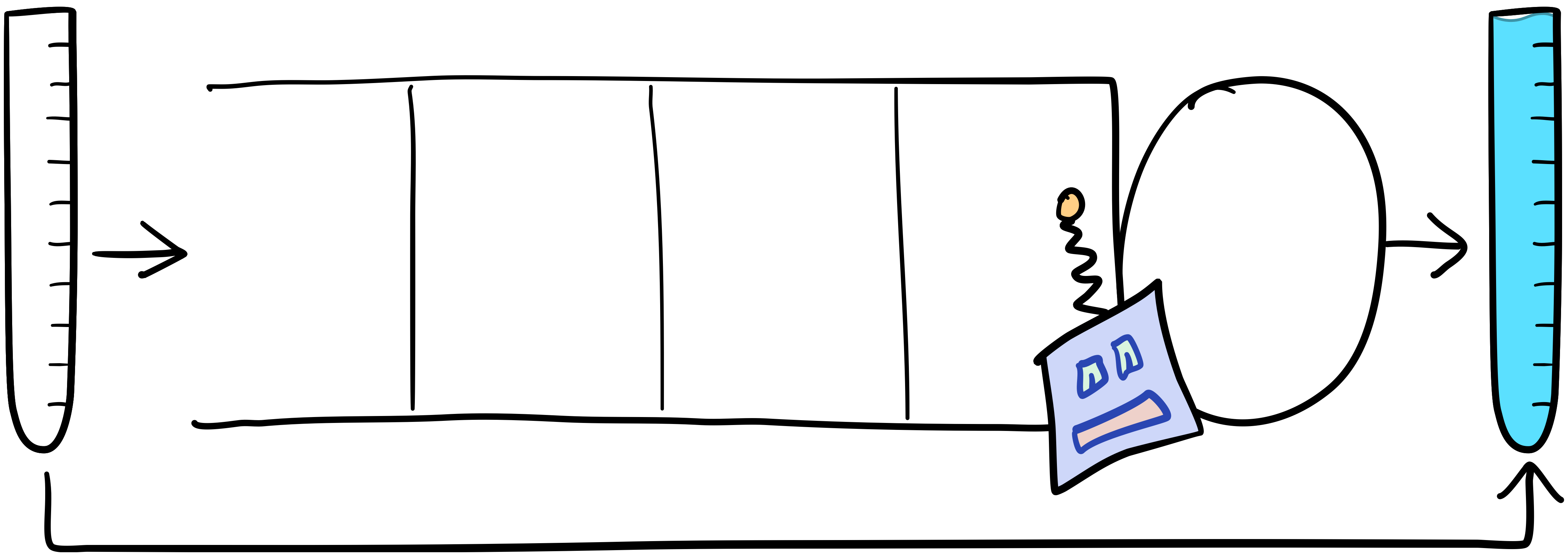


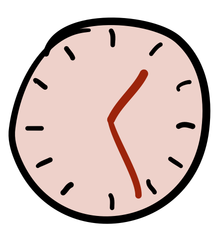
Response Time



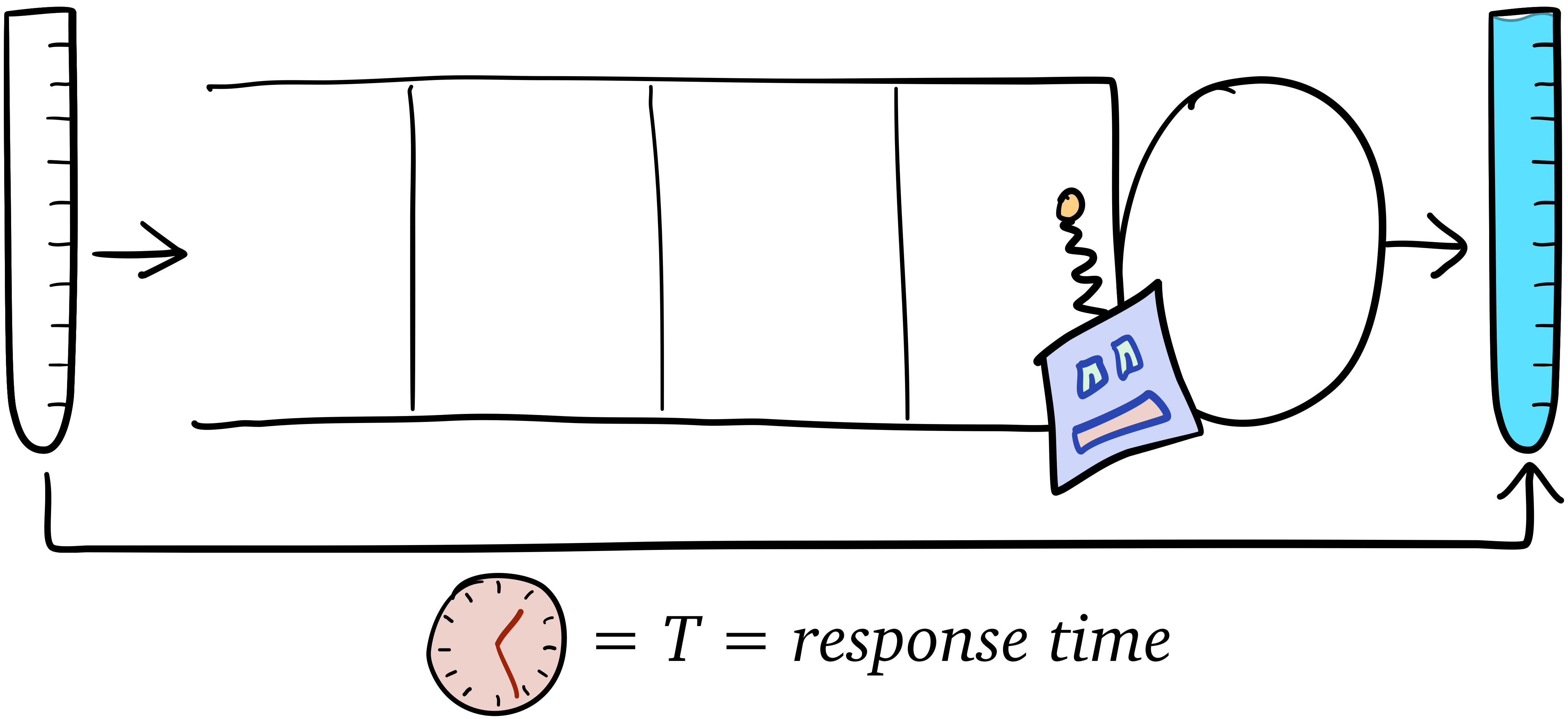
$= T =$ *response time*

Response Time



 $= T =$ *response time*

Response Time



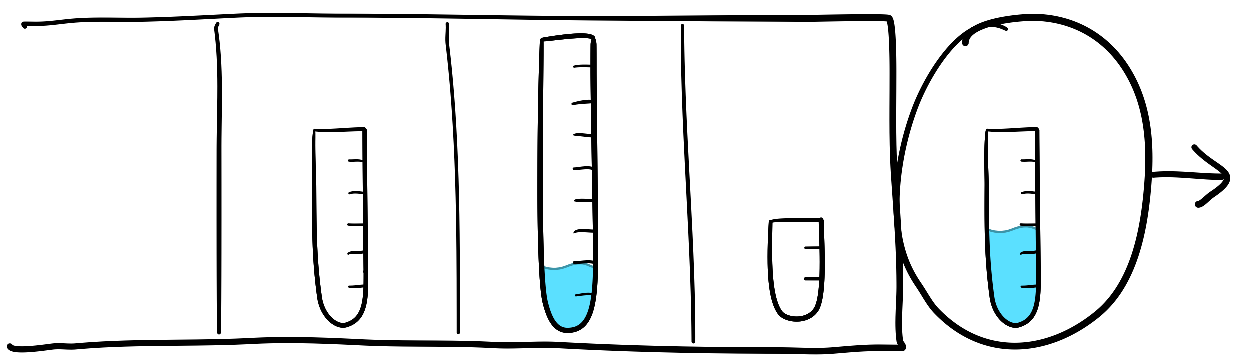
$= T =$ *response time*

Goal: schedule to minimize
mean response time $\mathbf{E}[T]$

State of the Art

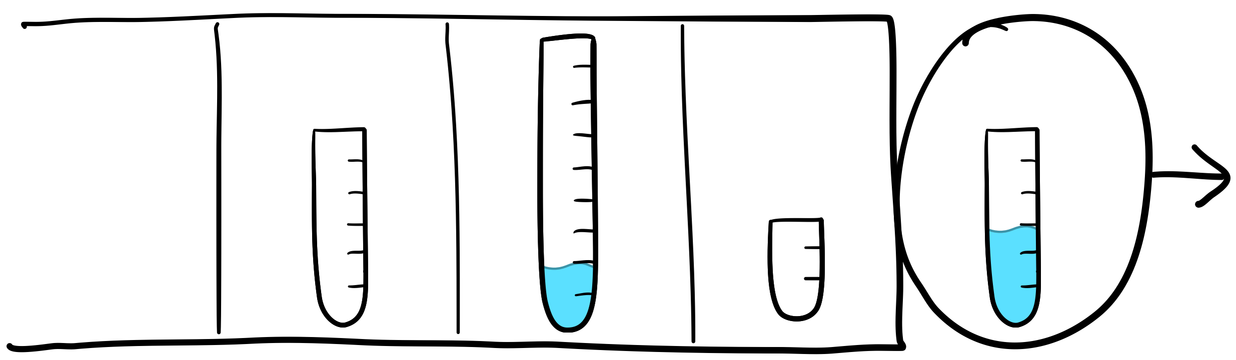
State of the Art

Known Job Sizes



State of the Art

Known Job Sizes

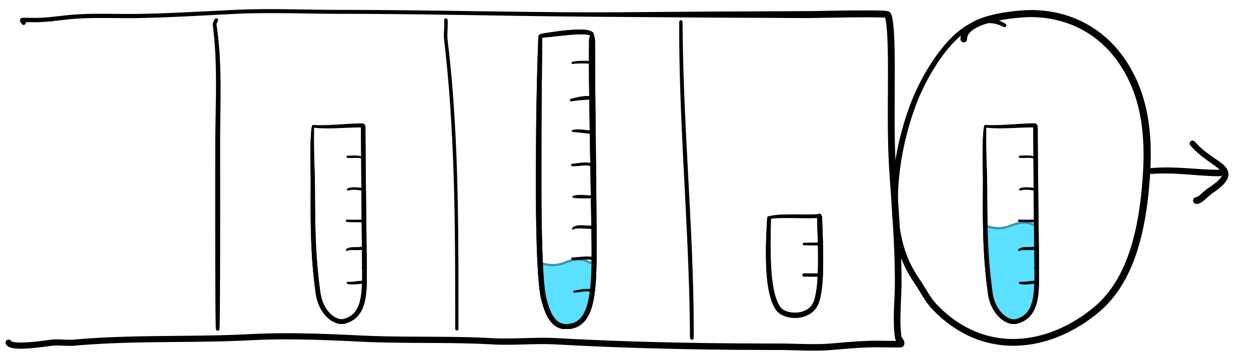


Optimal: **SRPT**

(serve job of least *remaining size*)

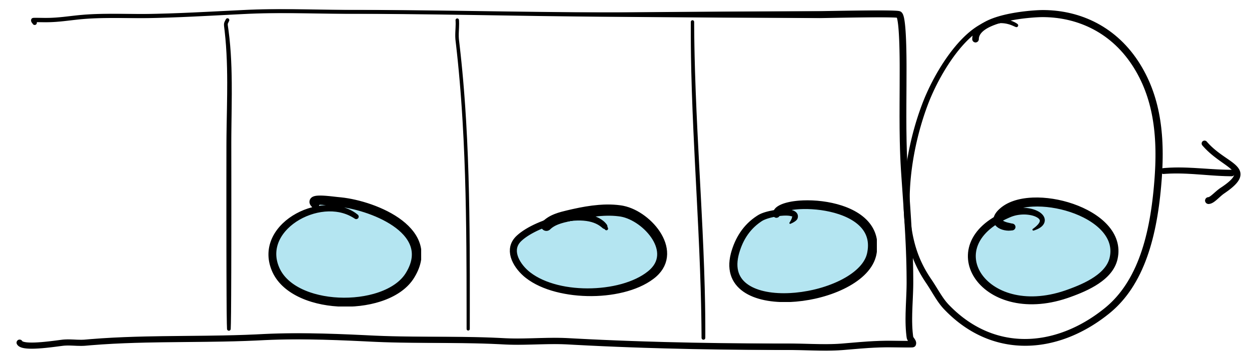
State of the Art

Known Job Sizes



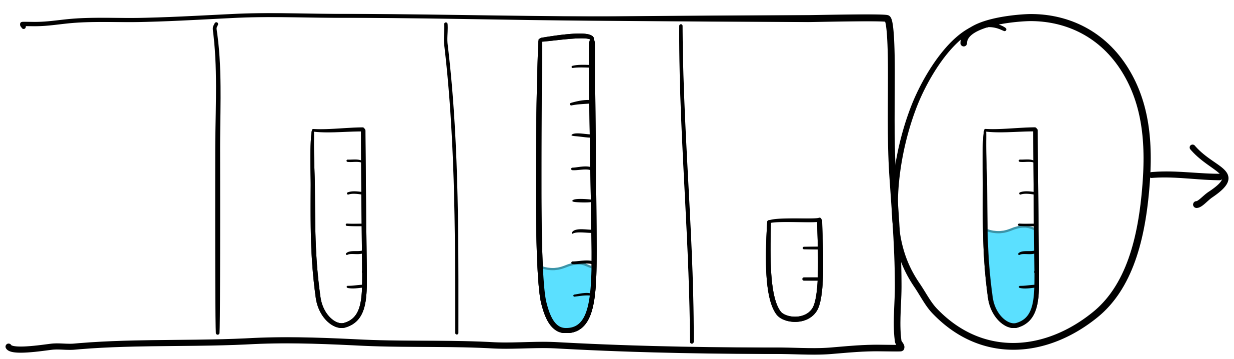
Optimal: **SRPT**
(serve job of least *remaining size*)

Unknown Job Sizes



State of the Art

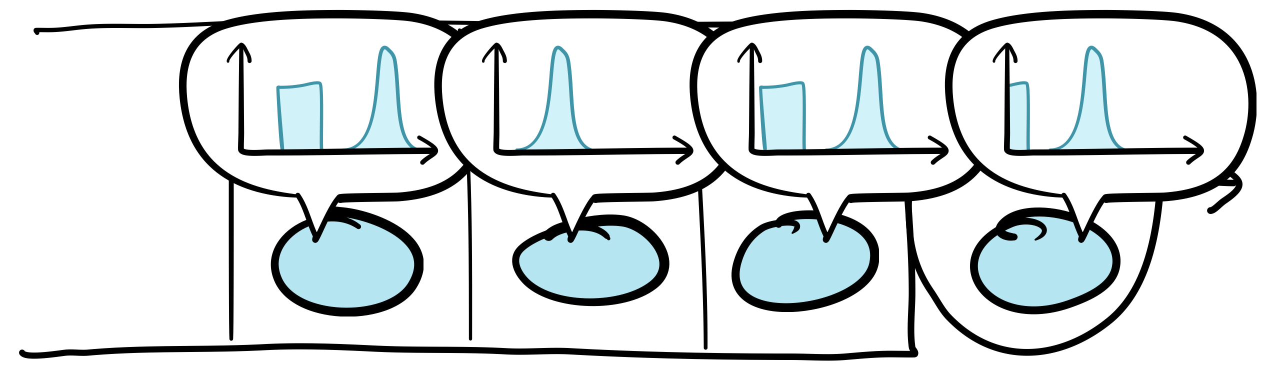
Known Job Sizes



Optimal: **SRPT**

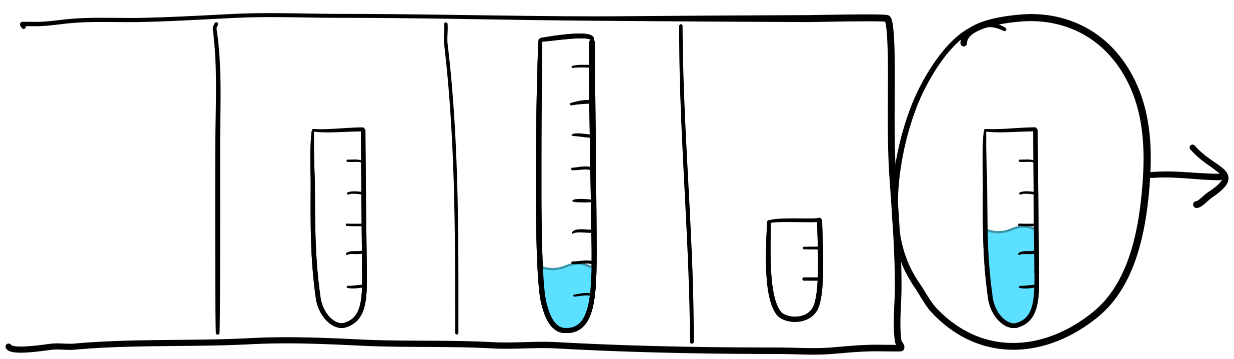
(serve job of least *remaining size*)

Unknown Job Sizes



State of the Art

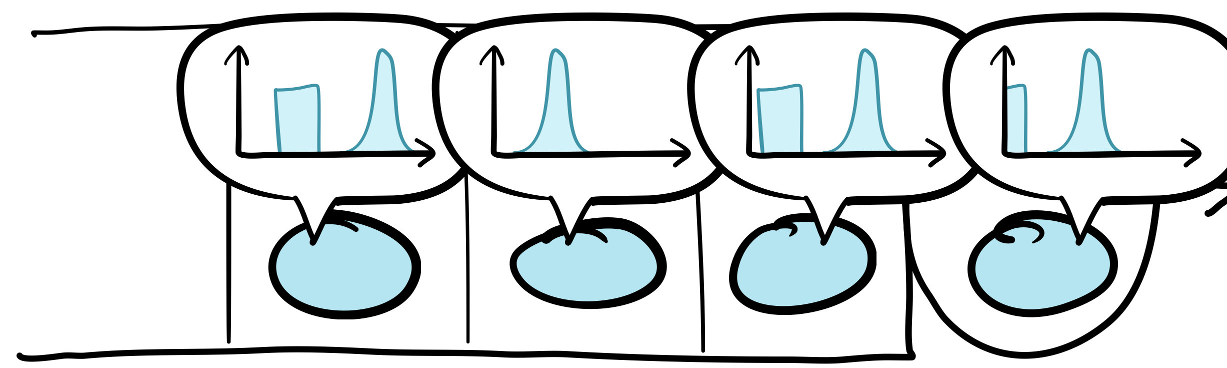
Known Job Sizes



Optimal: **SRPT**

(serve job of least *remaining size*)

Unknown Job Sizes

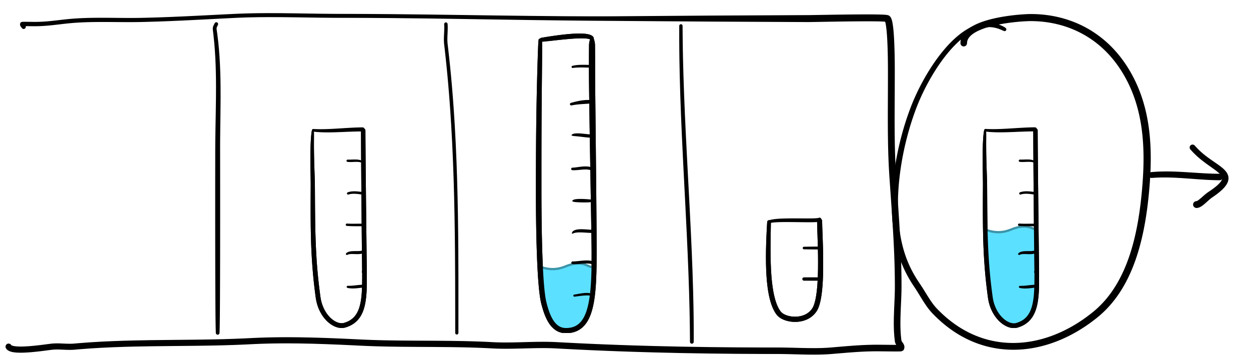


Seems good: **SERPT**

(serve job of least ***expected*** *remaining size*)

State of the Art

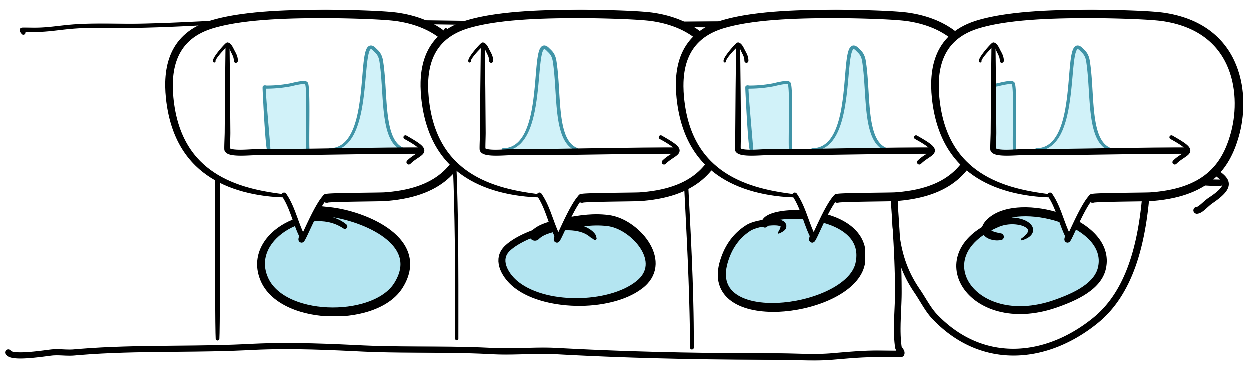
Known Job Sizes



Optimal: **SRPT**

(serve job of least *remaining size*)

Unknown Job Sizes



Seems good: **SERPT**

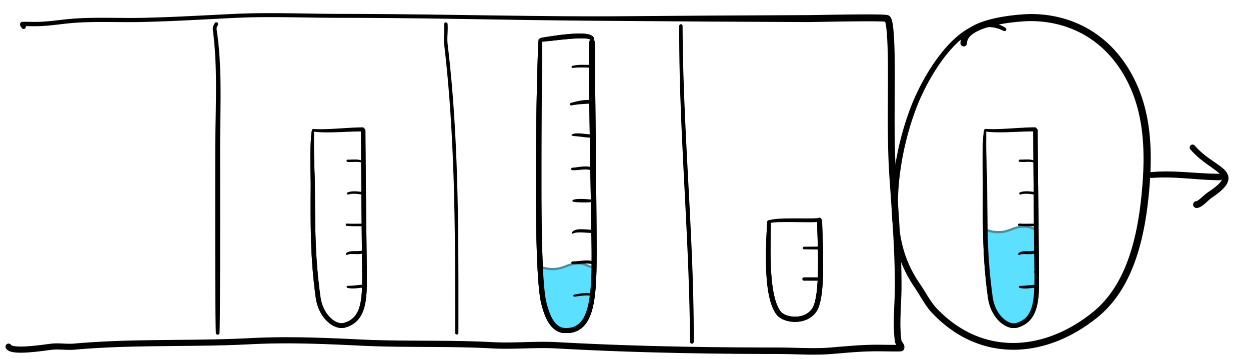
(serve job of least ***expected*** *remaining size*)

Optimal: **Gittins**

(serve job of least ***Gittins*** *rank*)

State of the Art

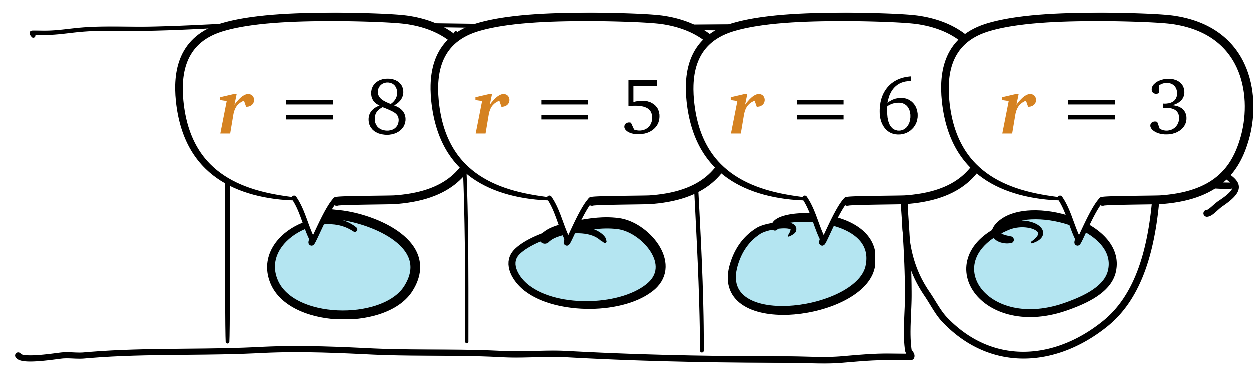
Known Job Sizes



Optimal: **SRPT**

(serve job of least *remaining size*)

Unknown Job Sizes



Seems good: **SERPT**

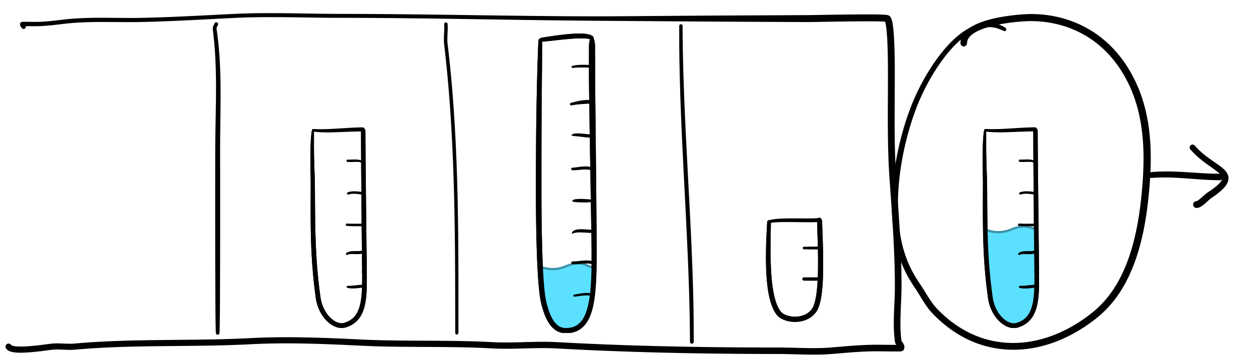
(serve job of least ***expected*** *remaining size*)

Optimal: **Gittins**

(serve job of least ***Gittins*** *rank*)

State of the Art

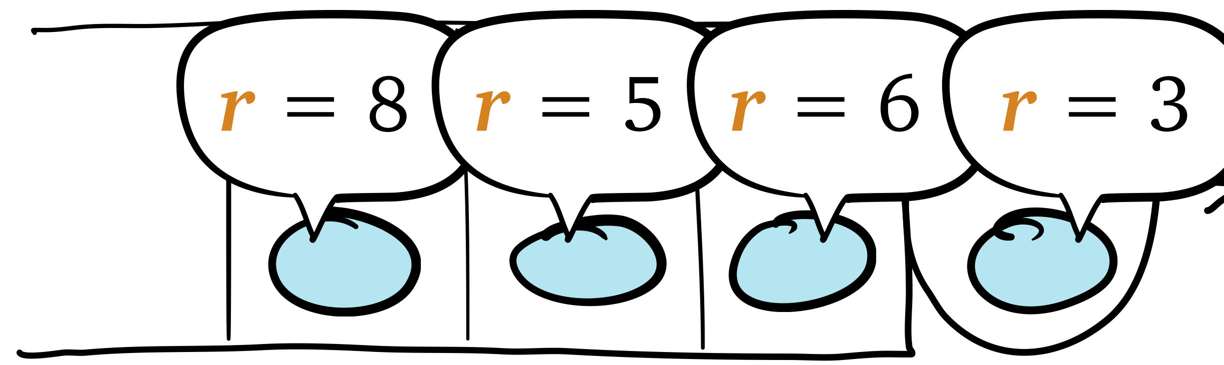
Known Job Sizes



Optimal: **SRPT**

(serve job of least *remaining size*)

Unknown Job Sizes



Seems good: **SERPT**

(serve job of least ***expected*** *remaining size*)

Optimal: **Gittins**

(serve job of least ***Gittins*** *rank*)

Warning!

- **Gittins** is hard to compute
- **SERPT** has no $\mathbf{E}[T]$ guarantee

I wish my policy had…

- *simple* definition like **SERPT**
- *provable* guarantee on $\mathbf{E}[T]$ like **Gittins**

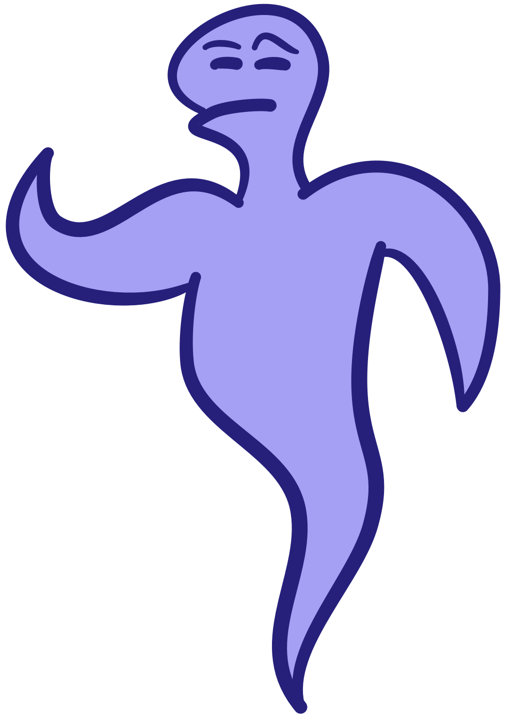
I wish my policy had…

- *simple* definition like **SERPT**
- *provable* guarantee on $\mathbf{E}[T]$ like **Gittins**

Open problem:
can we bound $\mathbf{E}[T]$
of **SERPT**?

I wish my policy had…

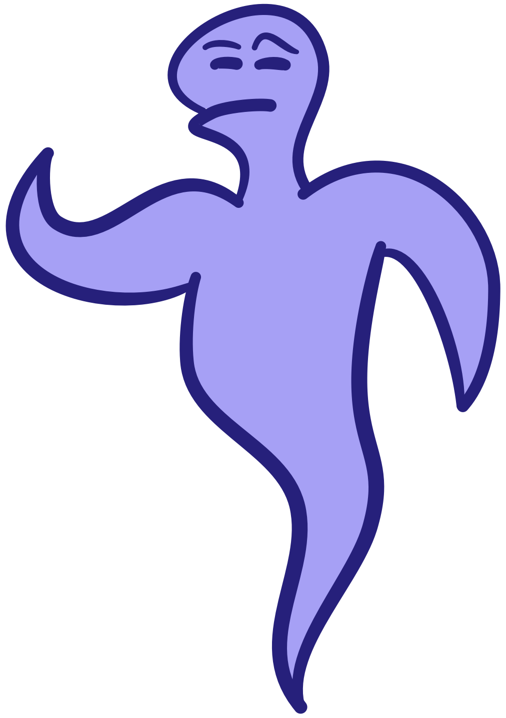
- *simple* definition like **SERPT**
- *provable* guarantee on $\mathbf{E}[T]$ like **Gittins**



M-SERPT

A new policy with both!

- *simple* definition like **SERPT**
- *provable* guarantee on $\mathbf{E}[T]$ like **Gittins**



M-SERPT

A new policy with both!

- *simple* definition like **SERPT**
- *provable* guarantee on $\mathbf{E}[T]$ like **Gittins**

First step: background on **SERPT** and **Gittins**

Background: **SERPT**

Background: **SERPT**

Job size distribution:

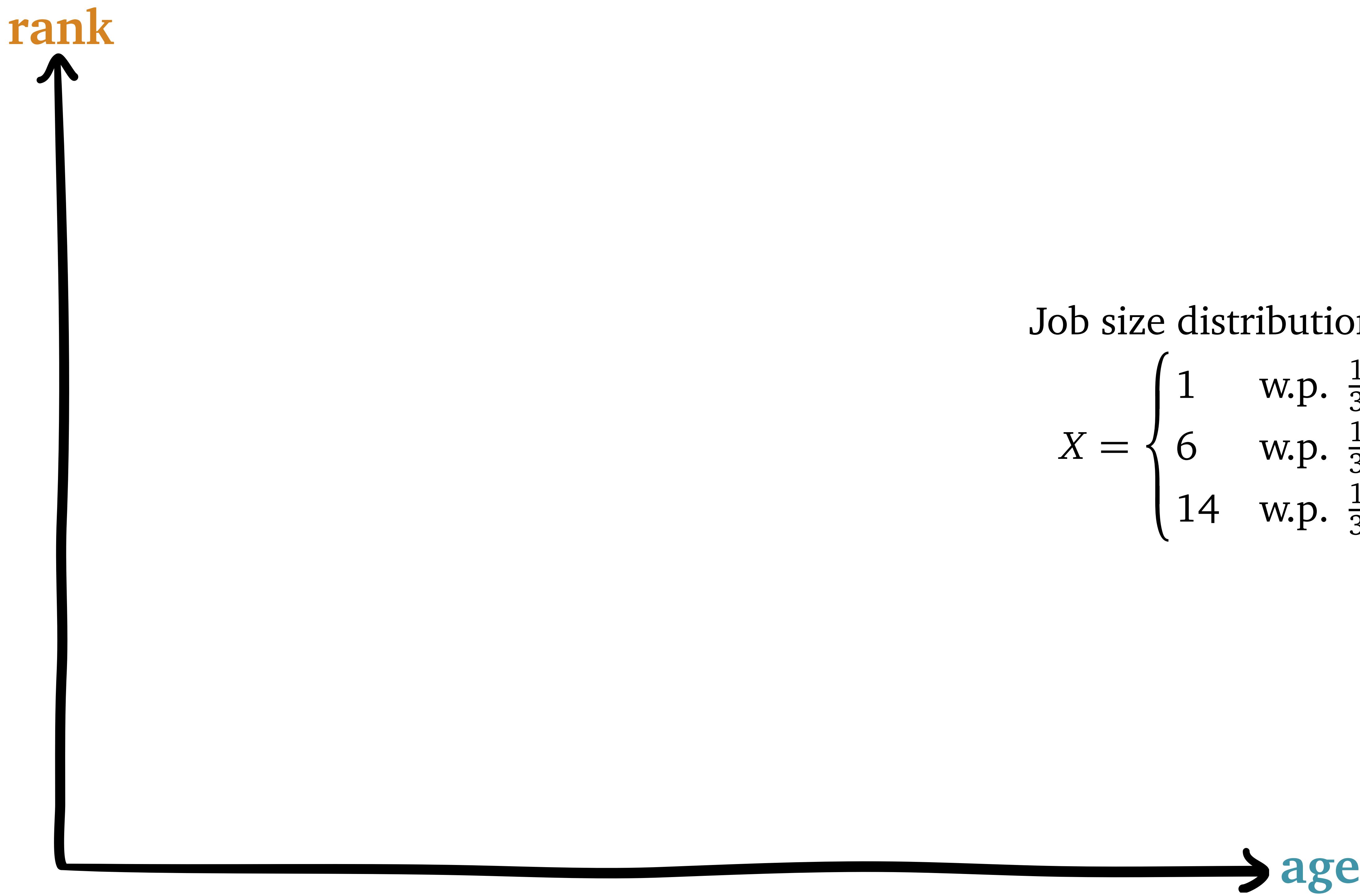
$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

Background: SERPT

Job size distribution:

$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

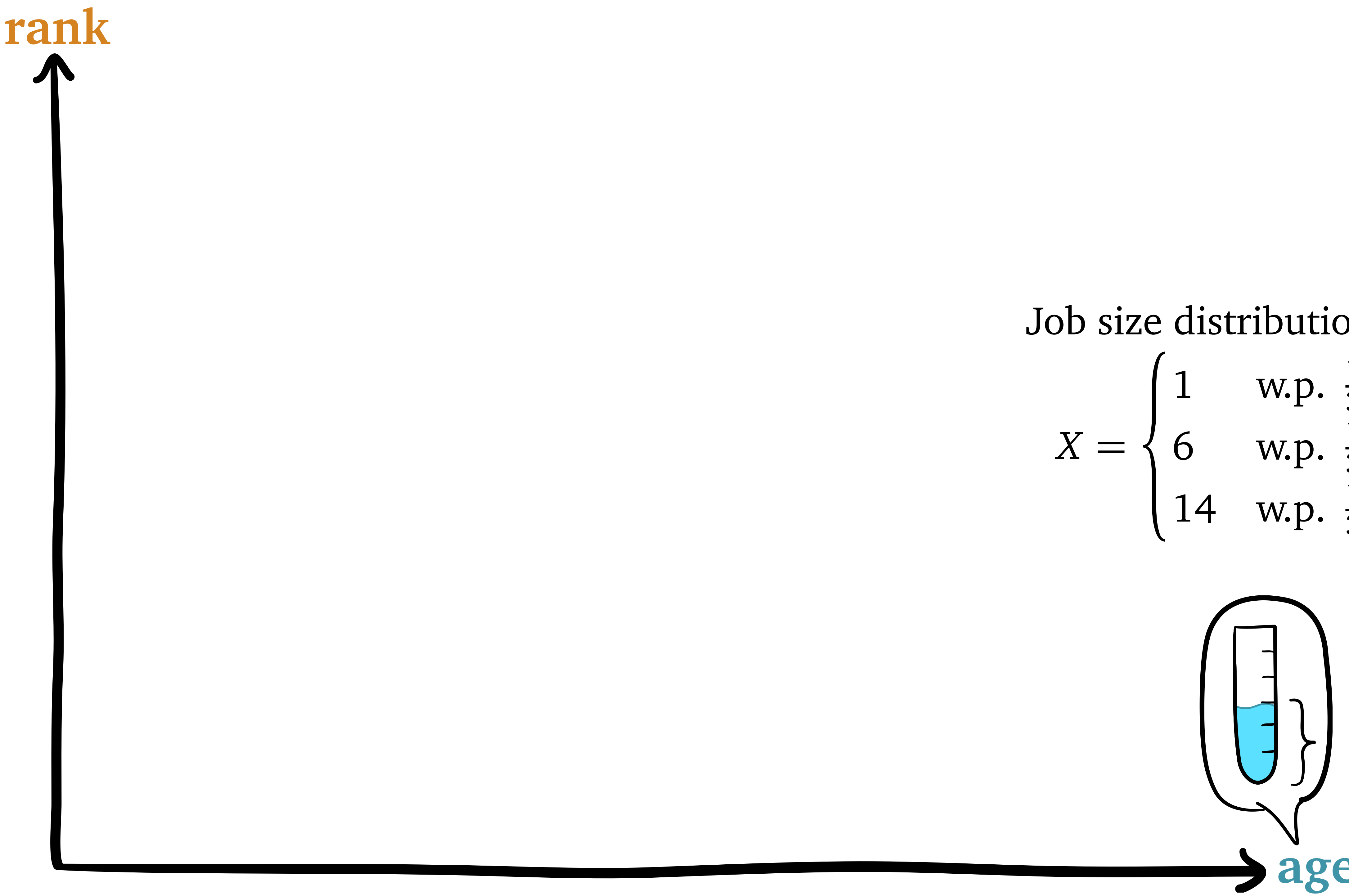
Background: SERPT



Job size distribution:

$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

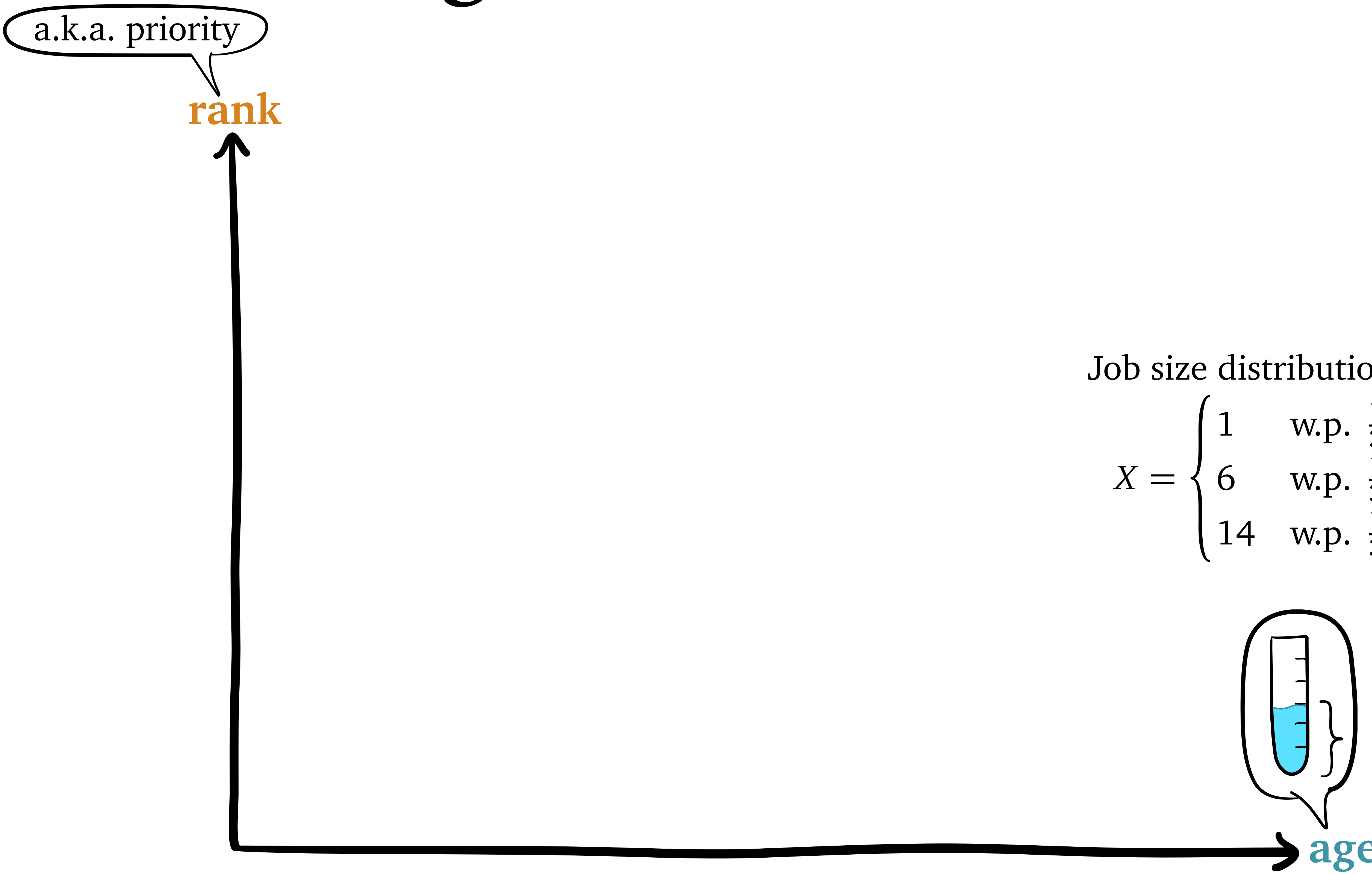
Background: SERPT



Job size distribution:

$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

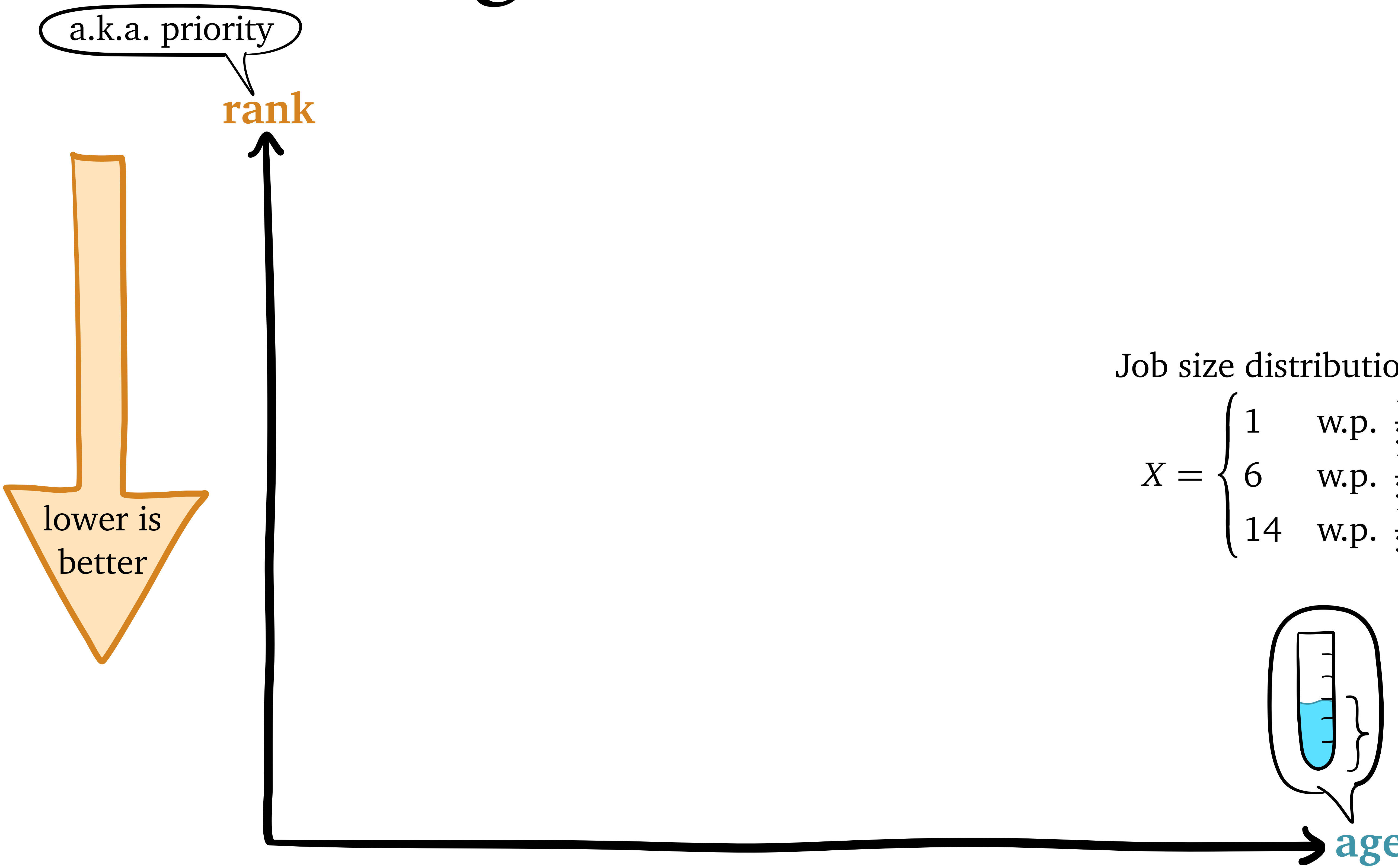
Background: SERPT



Job size distribution:

$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

Background: SERPT

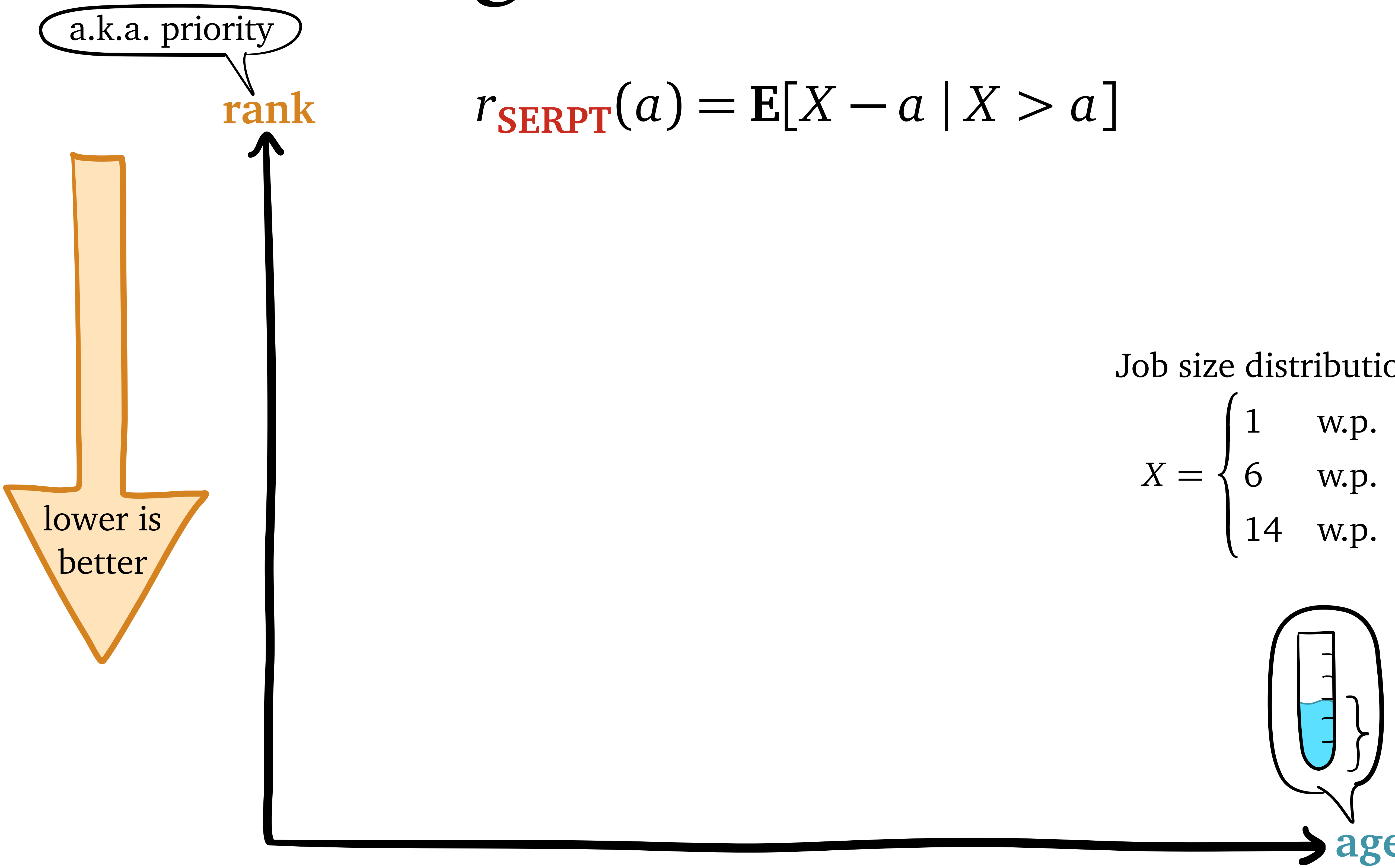


Job size distribution:

$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

Background: SERPT

$$r_{\mathbf{SERPT}}(a) = \mathbf{E}[X - a \mid X > a]$$

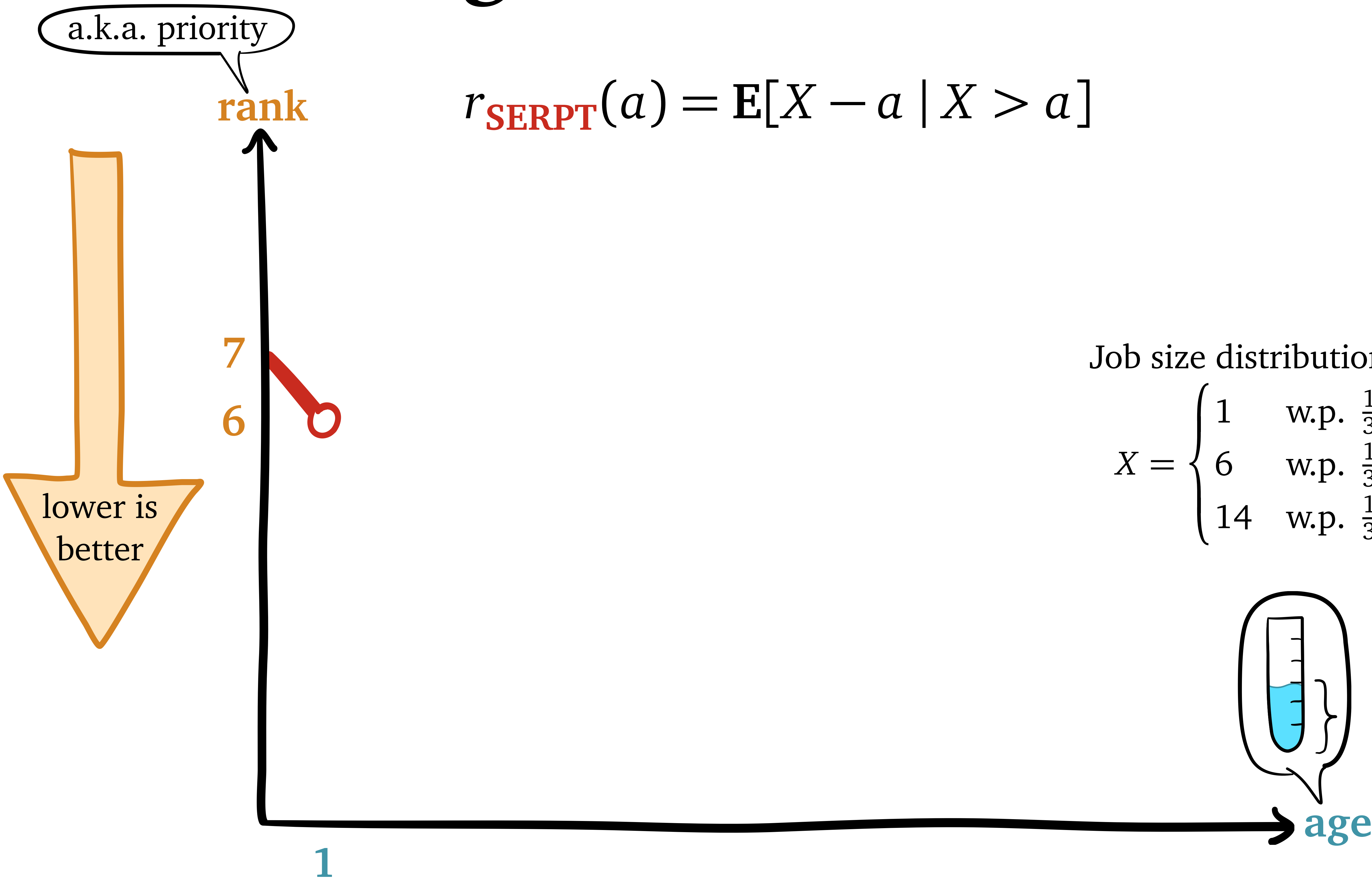


Job size distribution:

$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

Background: SERPT

$$r_{\text{SERPT}}(a) = \mathbf{E}[X - a \mid X > a]$$

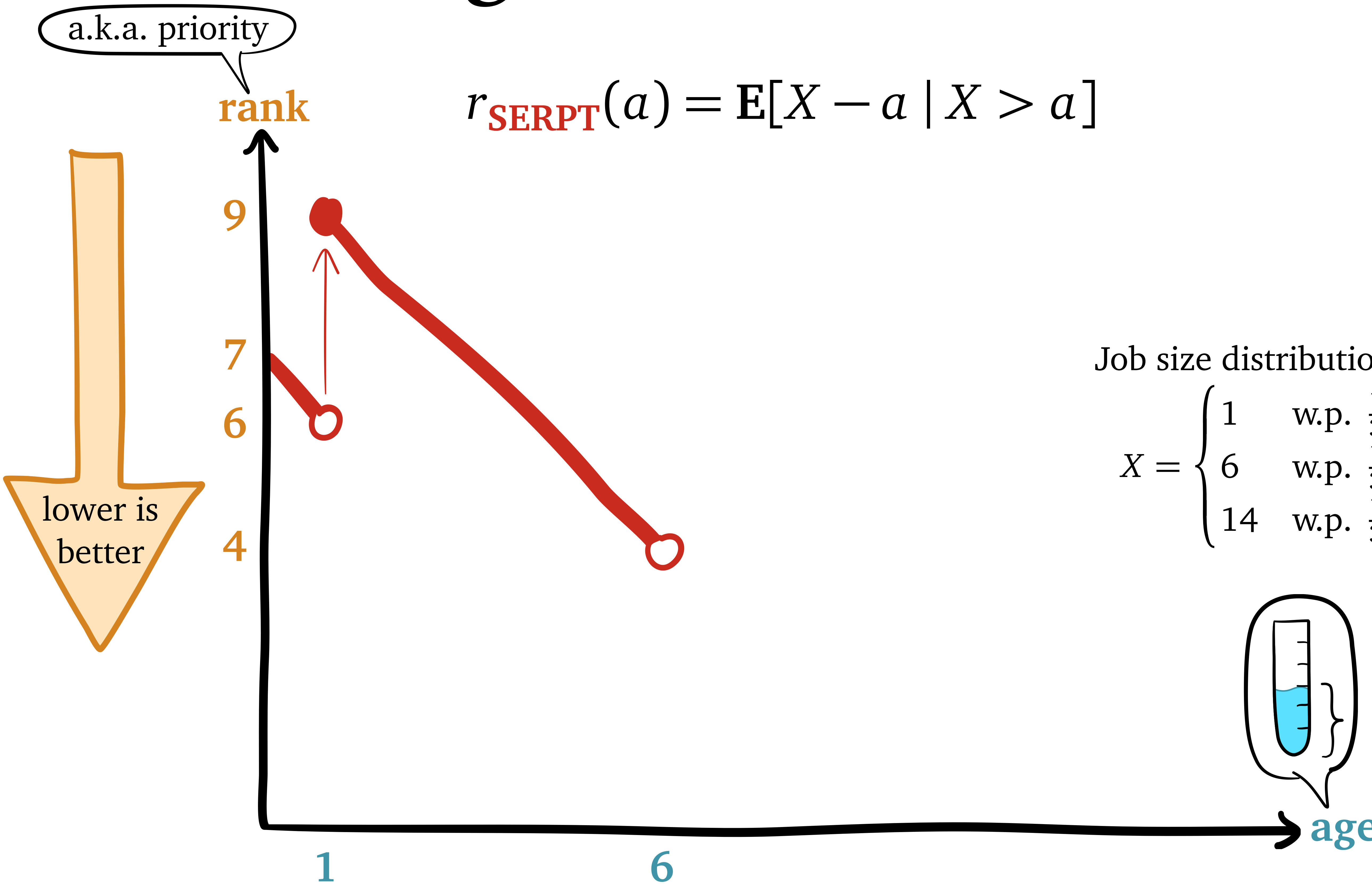


Job size distribution:

$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

Background: SERPT

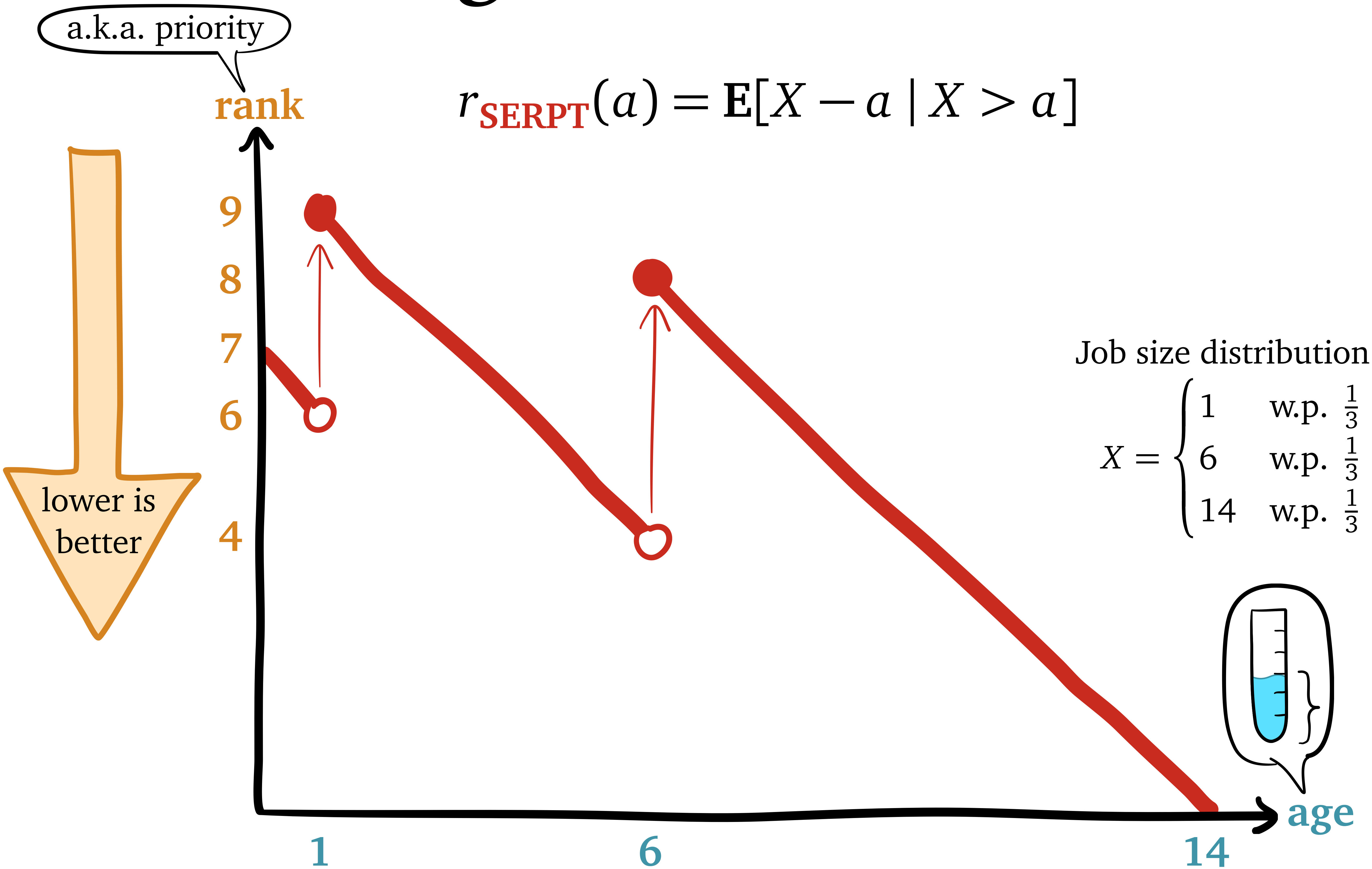
$$r_{\text{SERPT}}(a) = \mathbf{E}[X - a \mid X > a]$$



Job size distribution:

$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

Background: SERPT

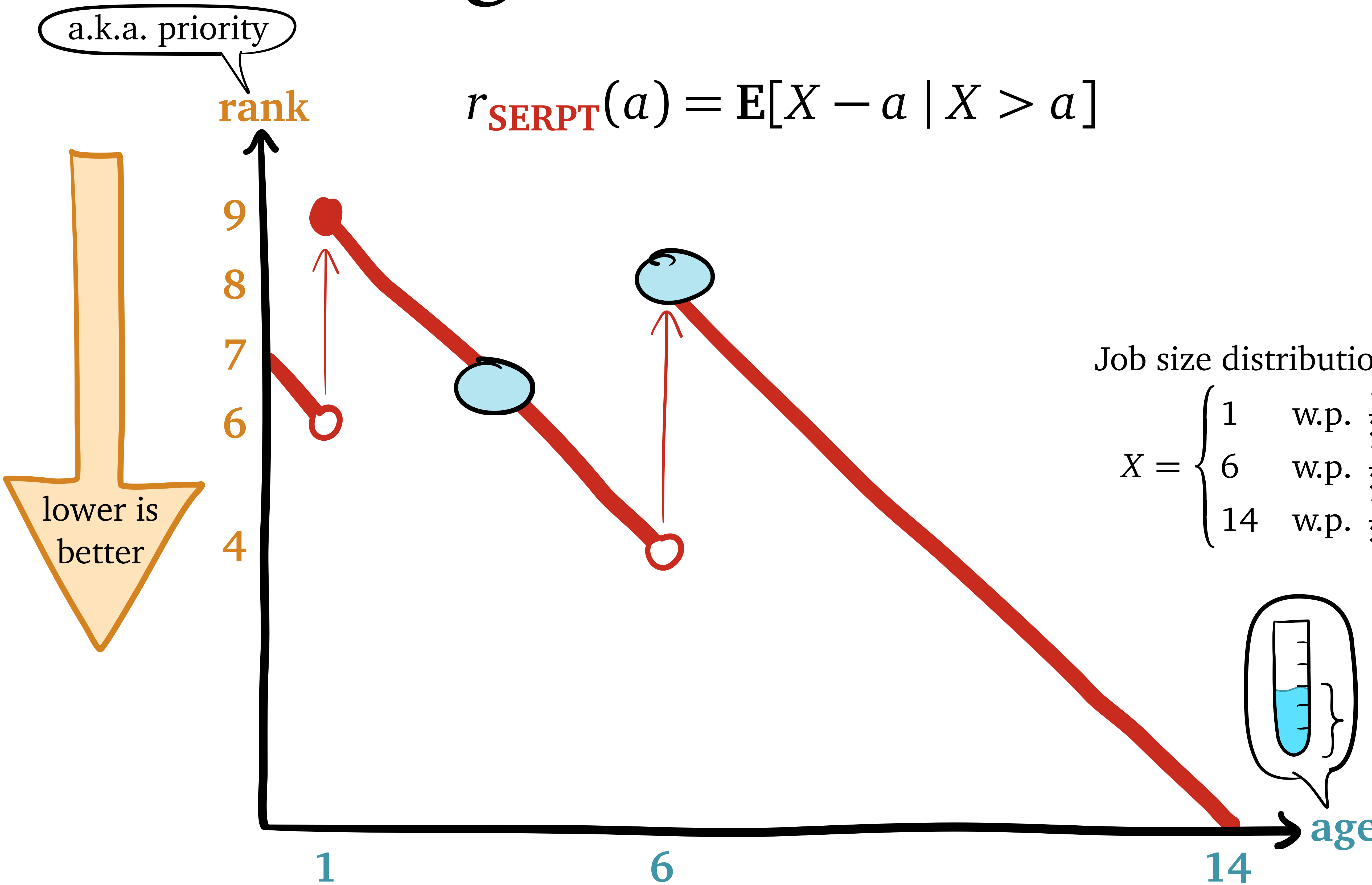


$$r_{\text{SERPT}}(a) = \mathbf{E}[X - a \mid X > a]$$

$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

Background: SERPT

$$r_{\text{SERPT}}(a) = \mathbf{E}[X - a \mid X > a]$$

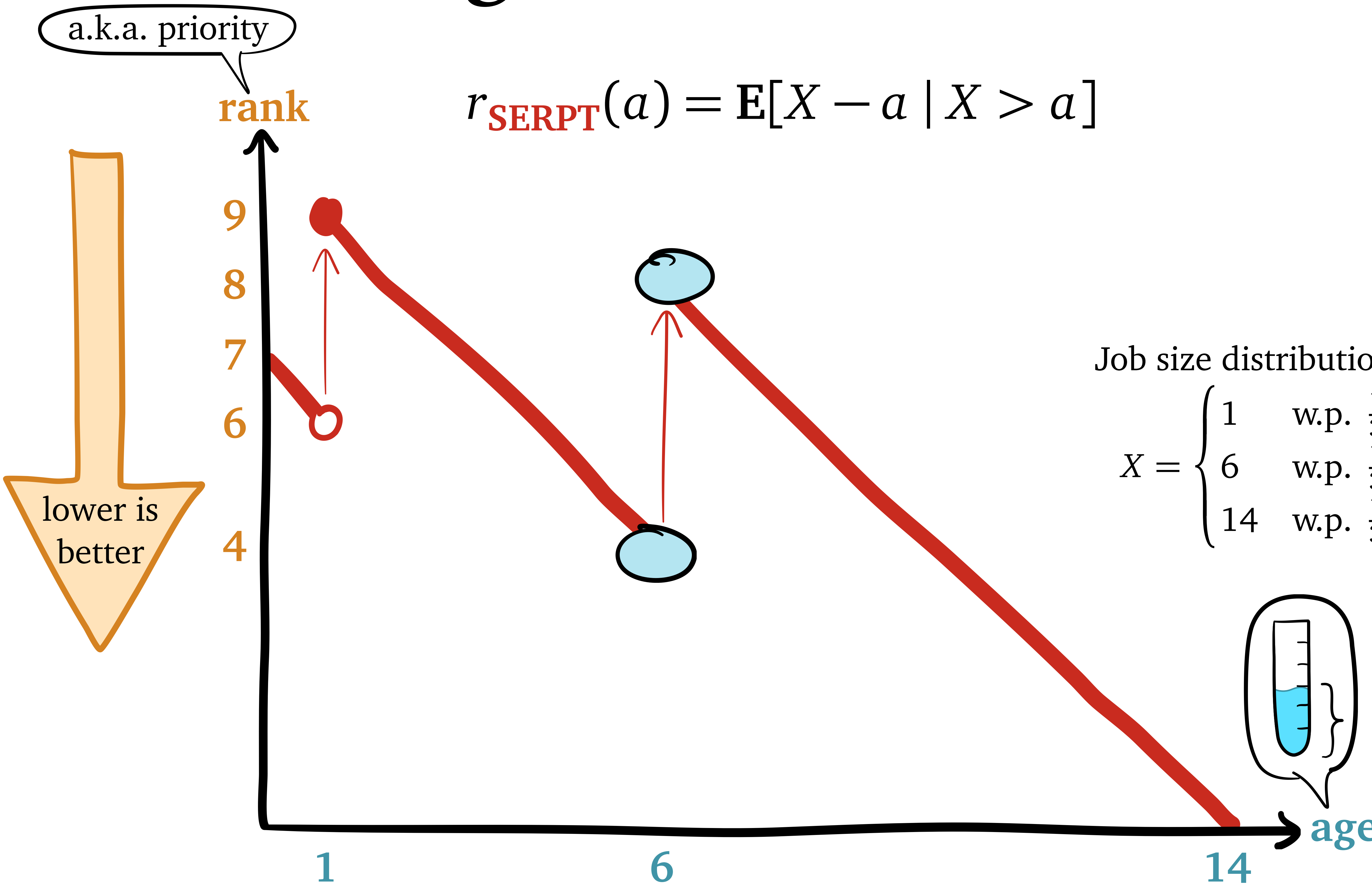


Job size distribution:

$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

Background: SERPT

$$r_{\text{SERPT}}(a) = \mathbf{E}[X - a \mid X > a]$$

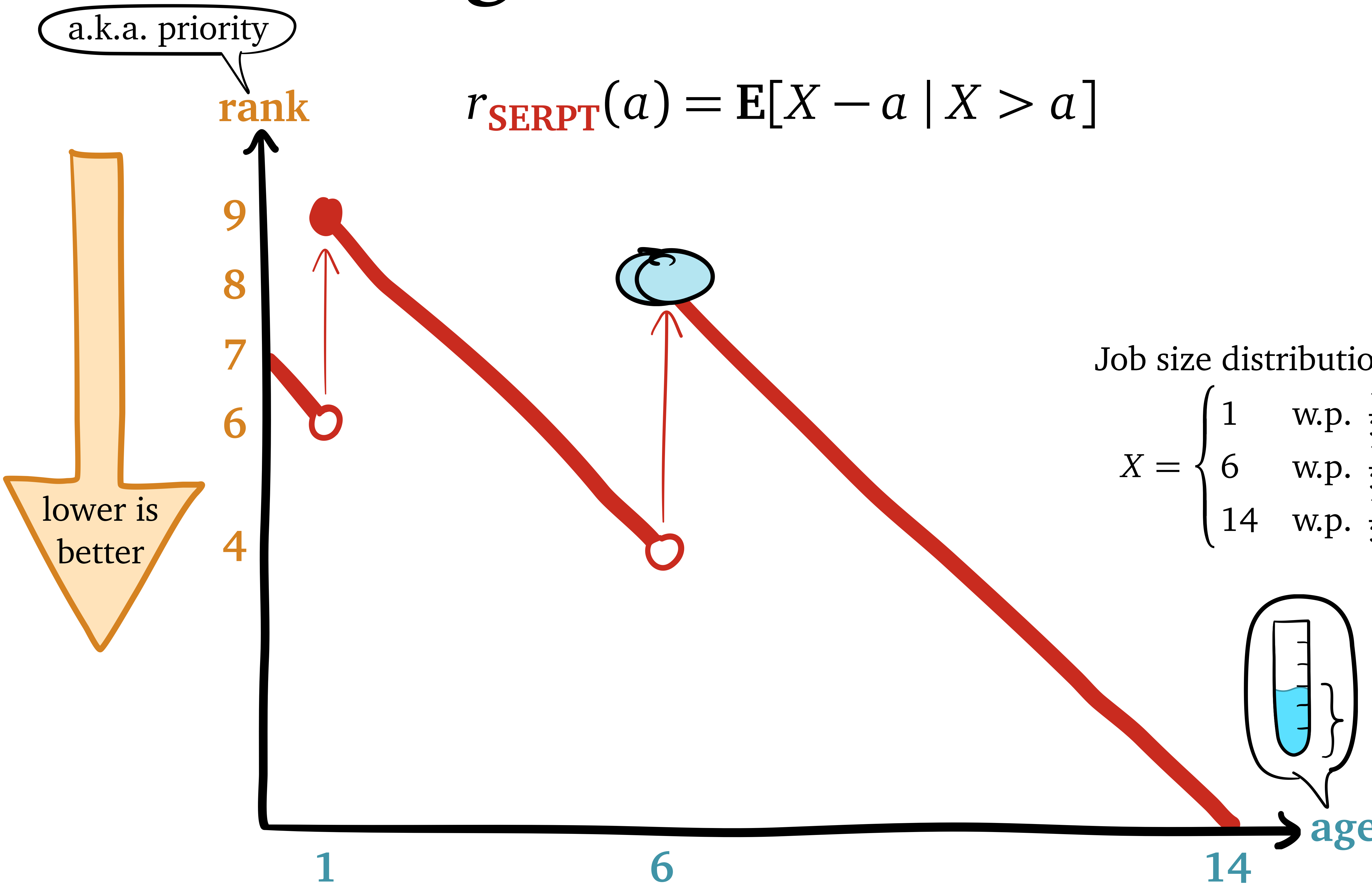


Job size distribution:

$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

Background: SERPT

$$r_{\text{SERPT}}(a) = \mathbf{E}[X - a \mid X > a]$$

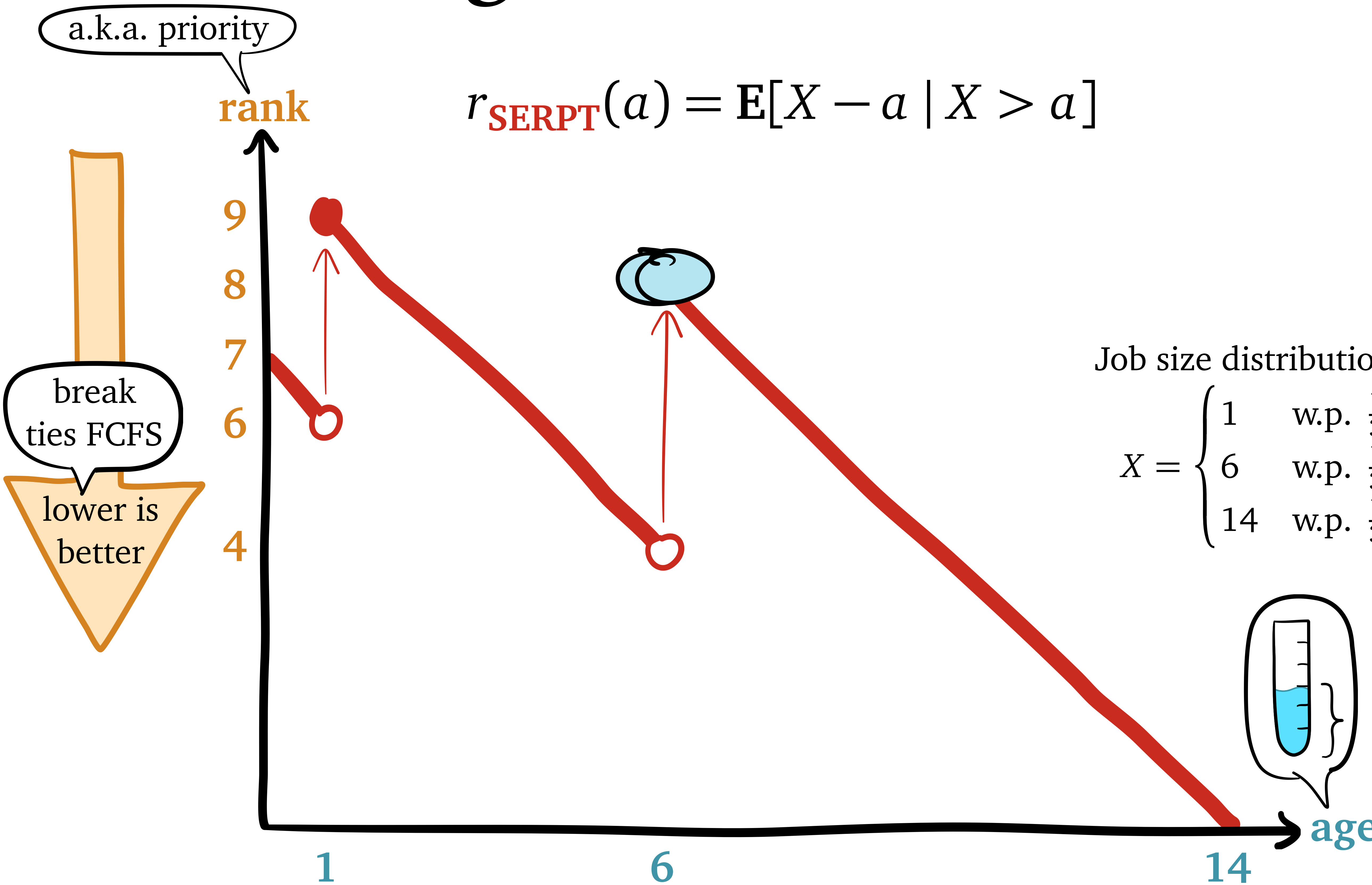


Job size distribution:

$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

Background: SERPT

$$r_{\text{SERPT}}(a) = \mathbf{E}[X - a \mid X > a]$$

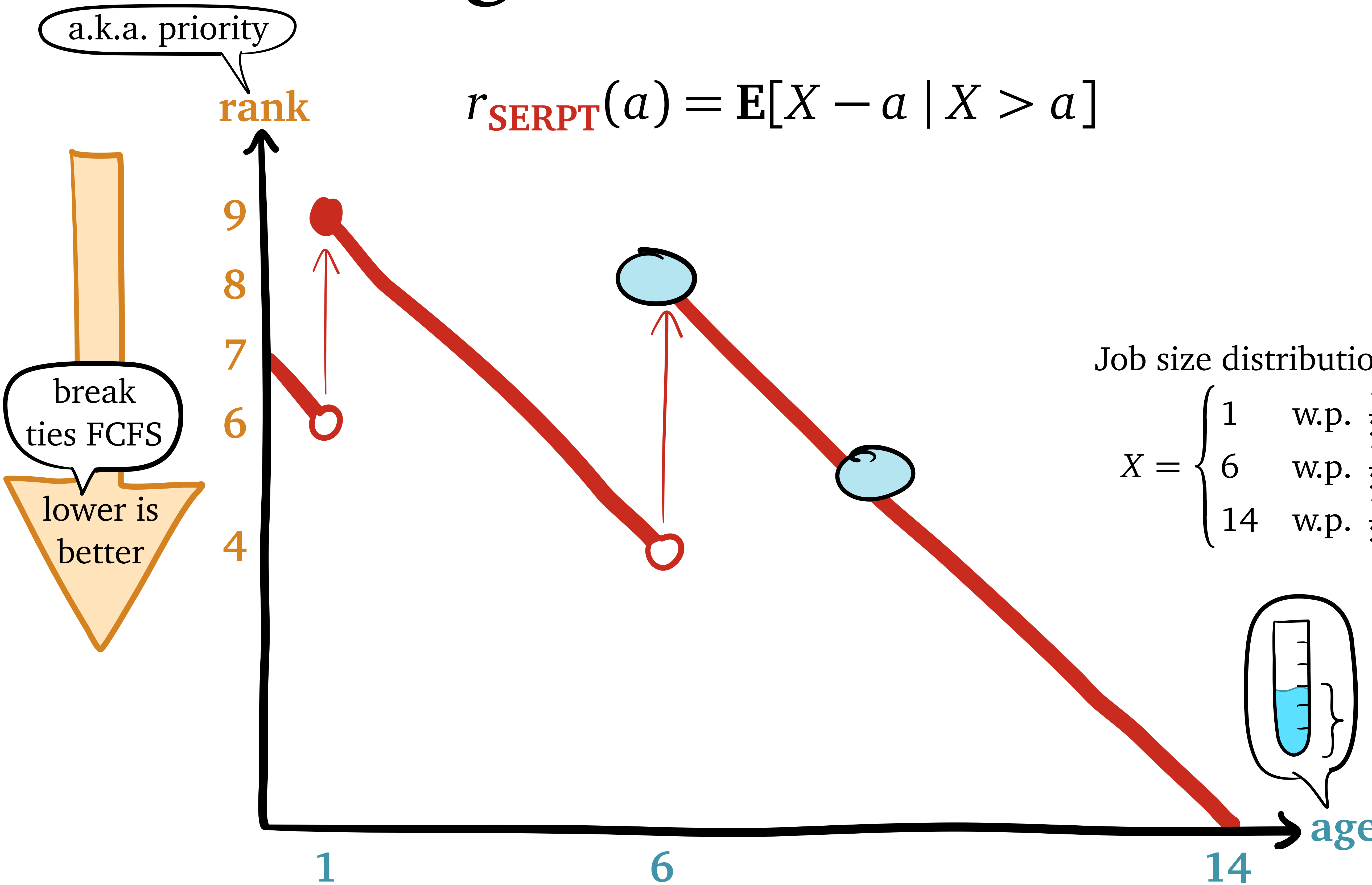


Job size distribution:

$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

Background: SERPT

$$r_{\text{SERPT}}(a) = \mathbf{E}[X - a \mid X > a]$$



Job size distribution:

$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

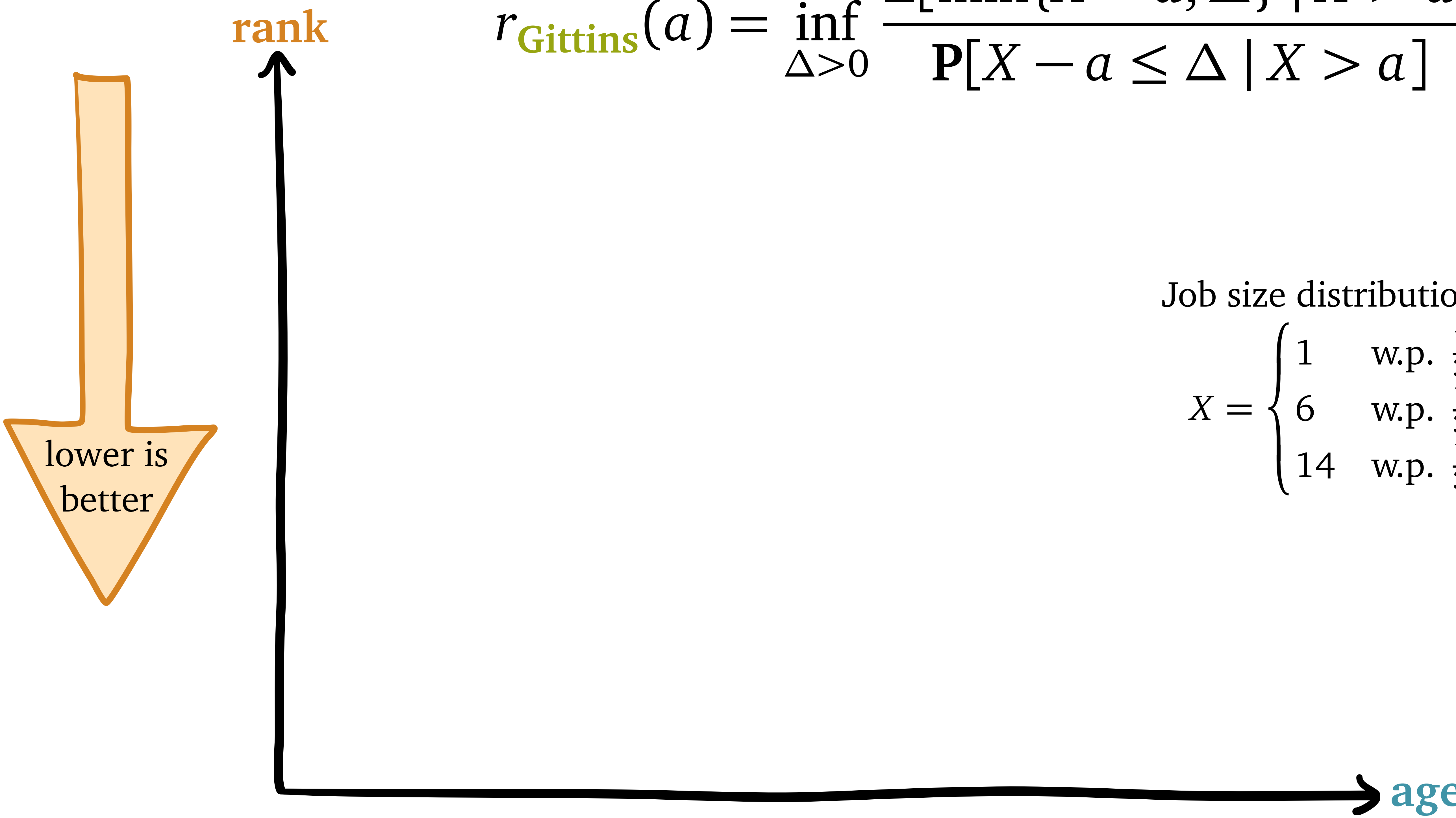
Background: **Gittins**

Job size distribution:

$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

Background: Gittins

$$r_{\text{Gittins}}(a) = \inf_{\Delta>0} \frac{\mathbf{E}[\min\{X-a,\Delta\} \mid X>a]}{\mathbf{P}[X-a \le \Delta \mid X>a]}$$

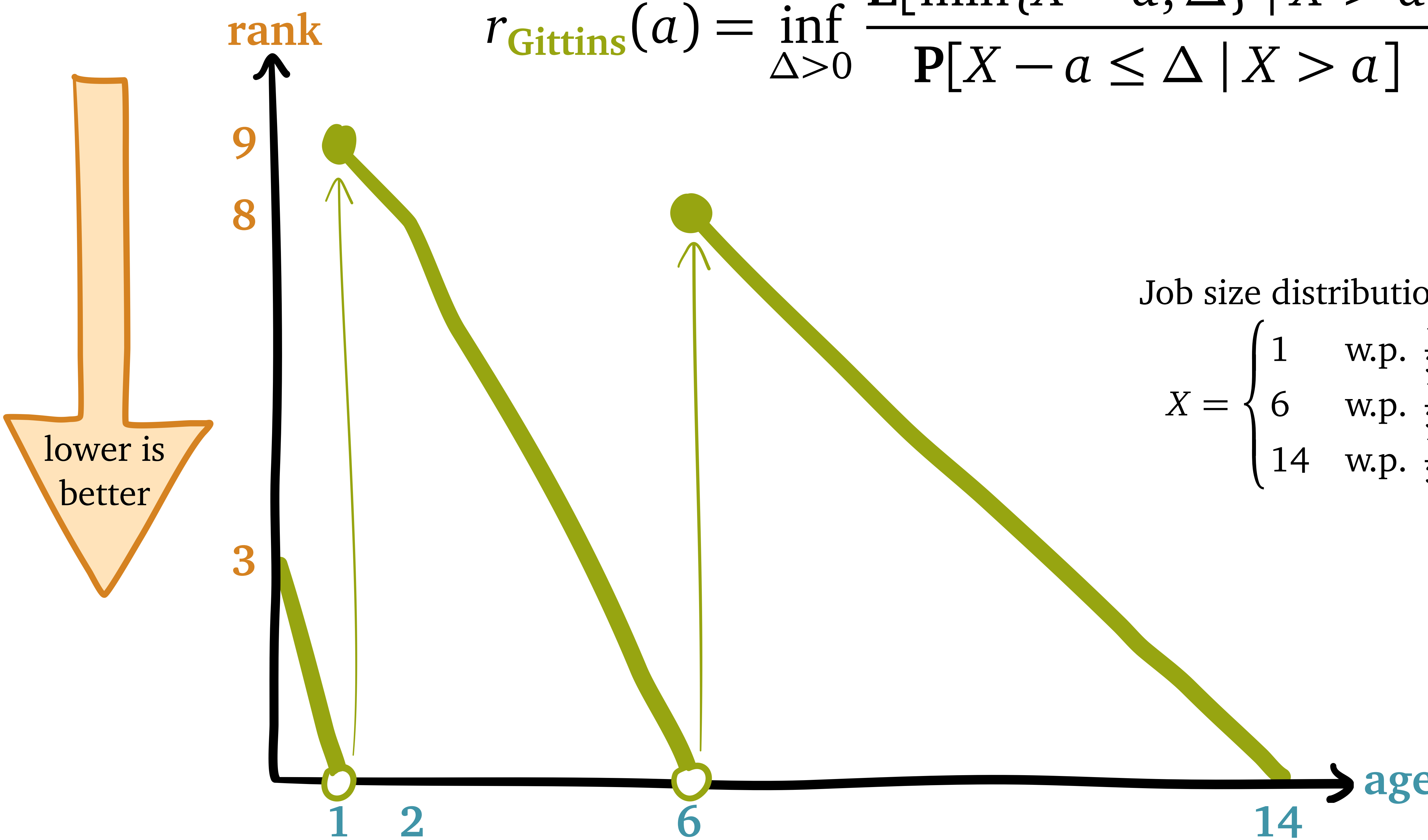


Job size distribution:

$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

Background: Gittins

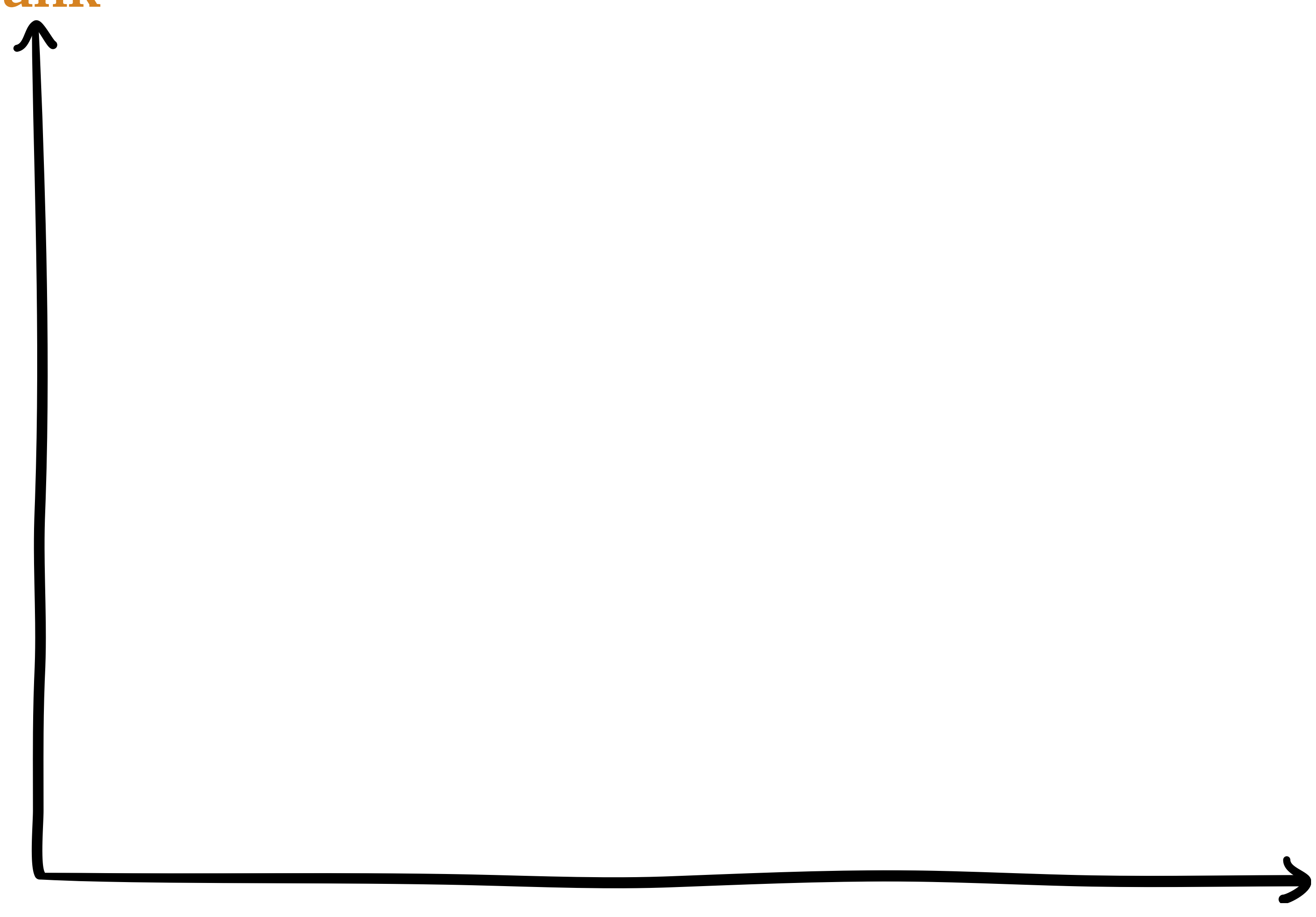
$$r_{\text{Gittins}}(a) = \inf_{\Delta>0} \frac{\mathbf{E}[\min\{X-a,\Delta\} \mid X>a]}{\mathbf{P}[X-a \le \Delta \mid X>a]}$$



Job size distribution:

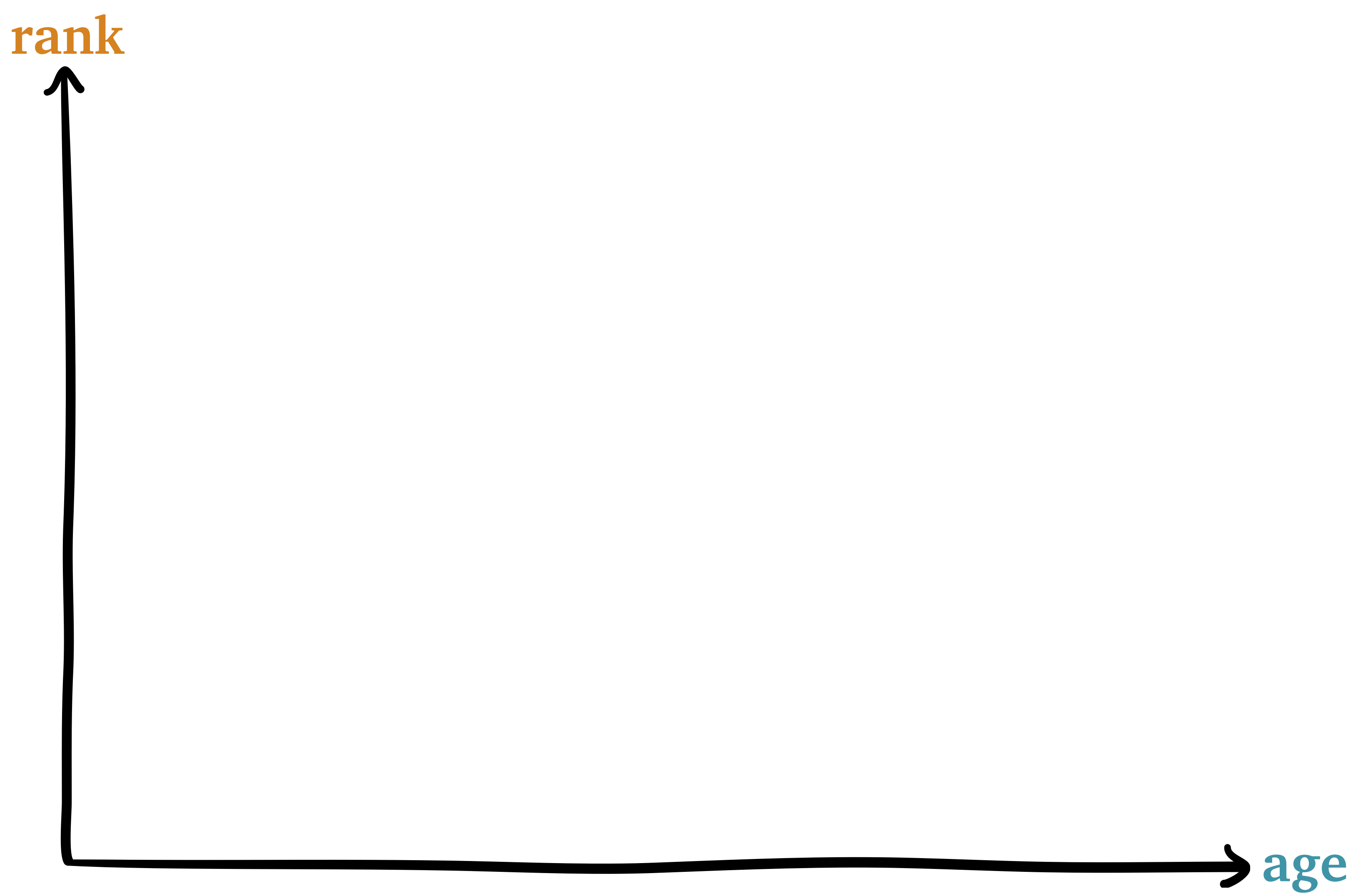
$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

Introducing **M-SERPT**



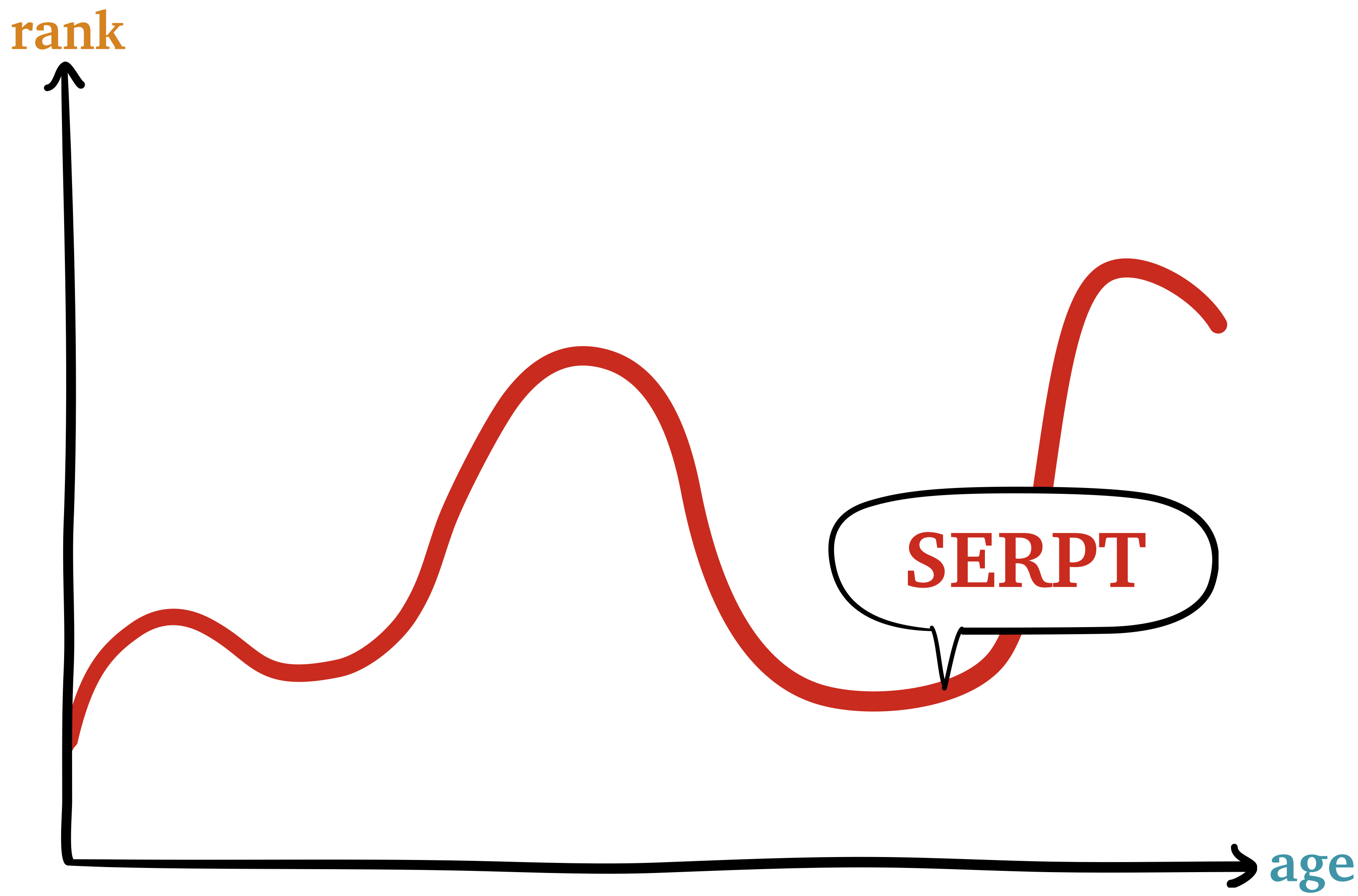
Introducing M-SERPT

$$r_{\text{M-SERPT}}(a) = \max_{0 \le b \le a} r_{\text{SERPT}}(b)$$



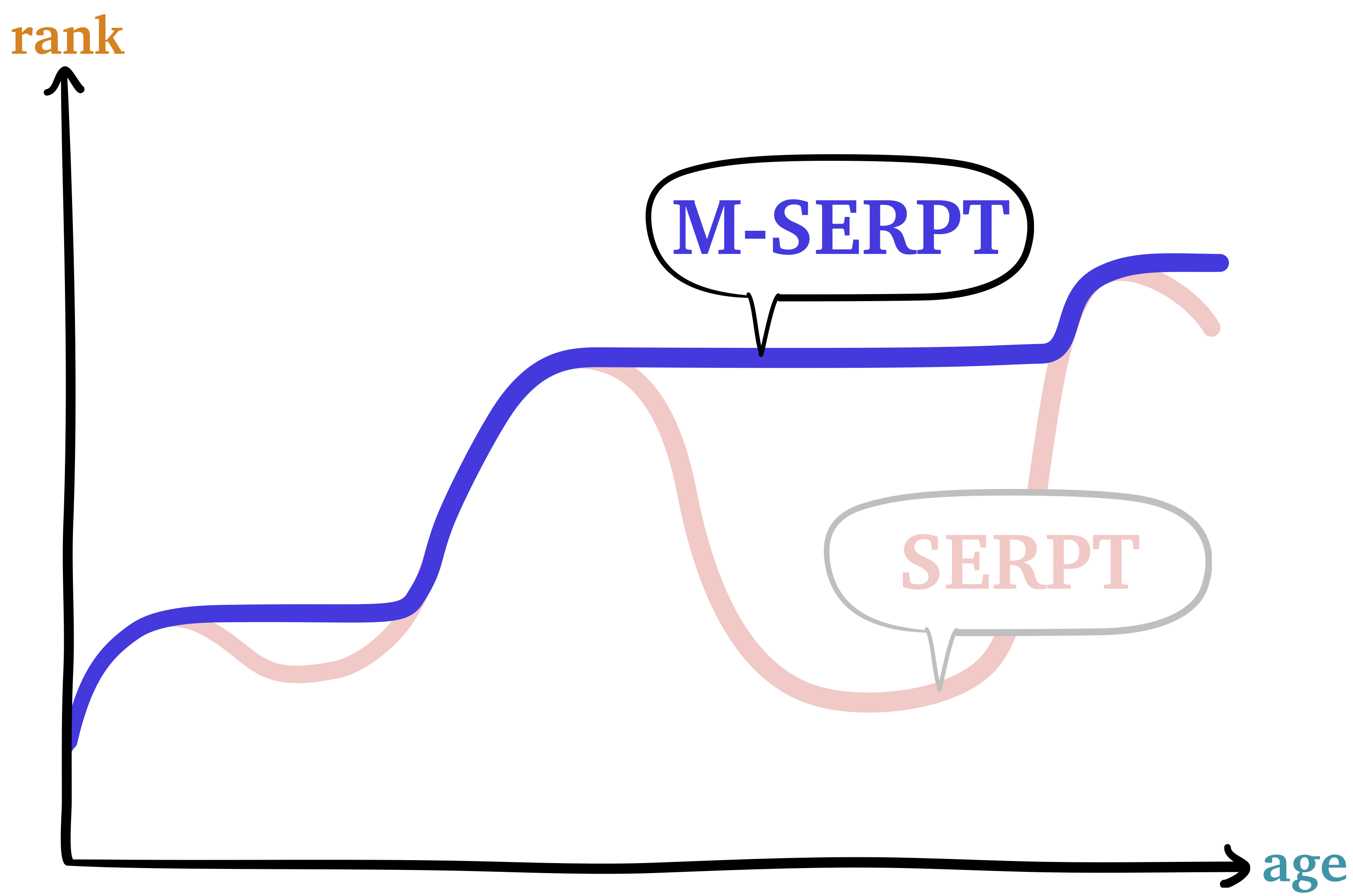
Introducing M-SERPT

$$r_{\text{M-SERPT}}(a) = \max_{0 \le b \le a} r_{\text{SERPT}}(b)$$



Introducing M-SERPT

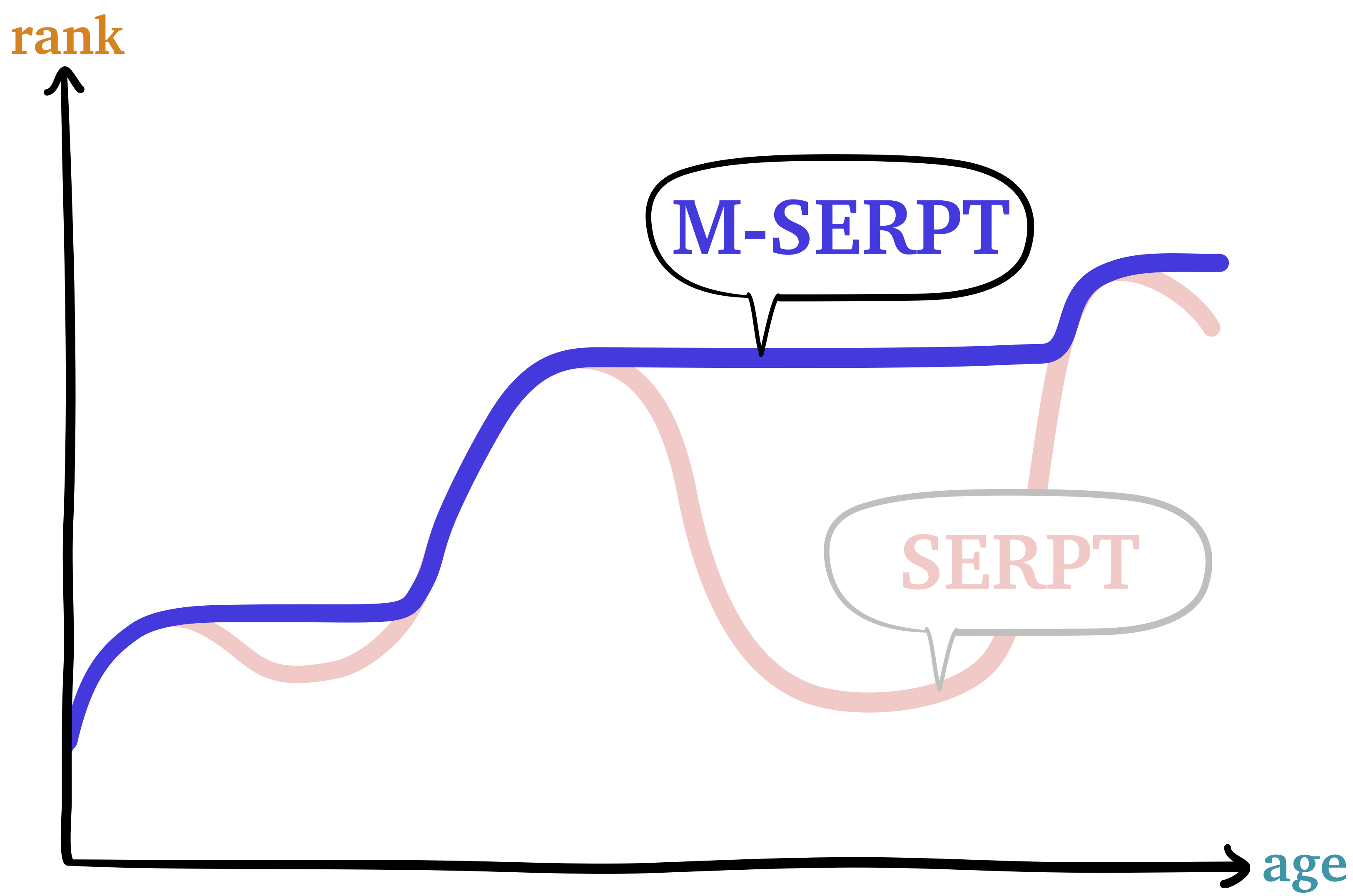
$$r_{\text{M-SERPT}}(a) = \max_{0 \le b \le a} r_{\text{SERPT}}(b)$$



Introducing M-SERPT

monotonic

$$r_{\text{M-SERPT}}(a) = \max_{0 \le b \le a} r_{\text{SERPT}}(b)$$



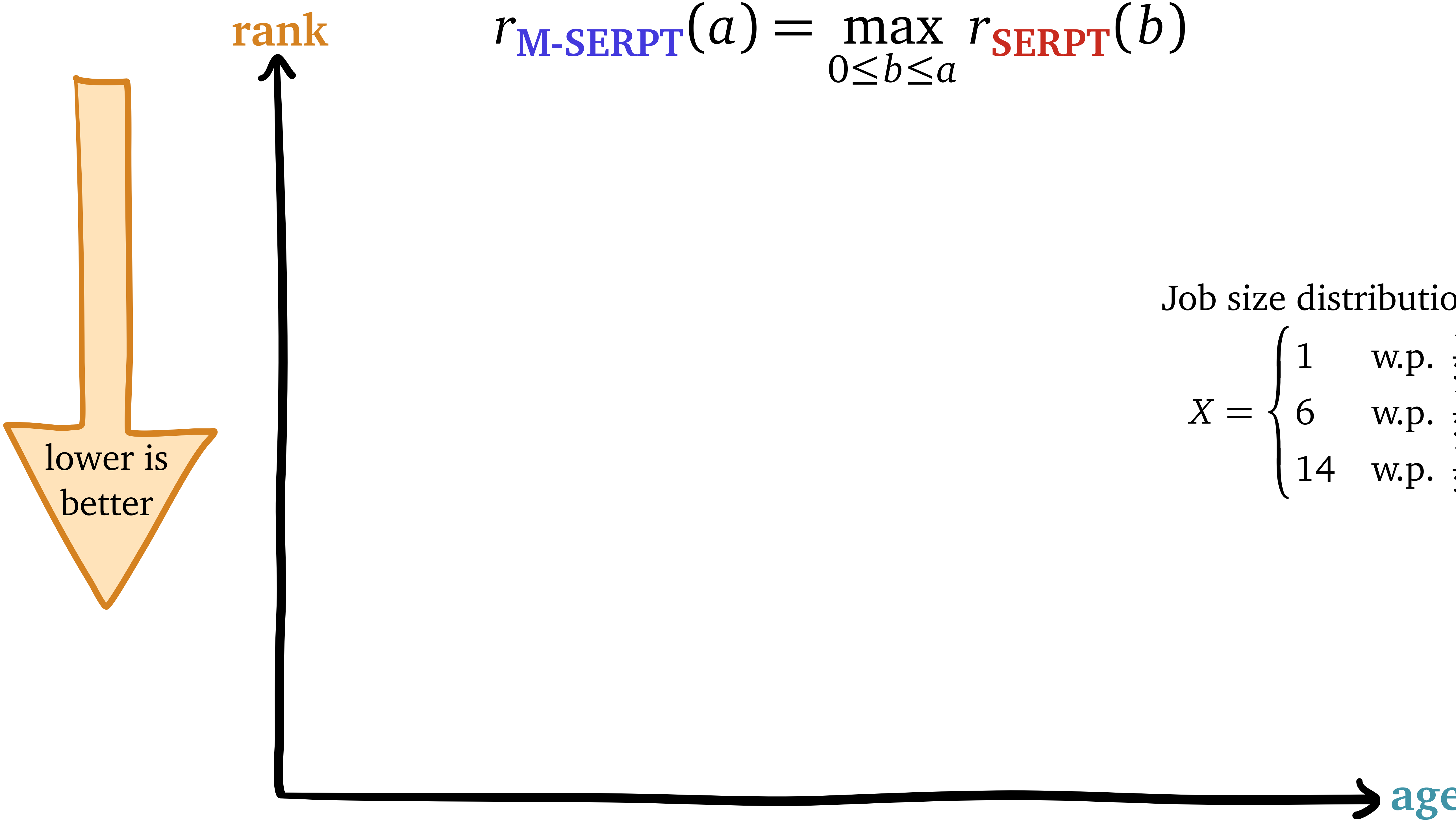
M-SERPT Example

Job size distribution:

$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

M-SERPT Example

$$r_{\text{M-SERPT}}(a) = \max_{0 \le b \le a} r_{\text{SERPT}}(b)$$

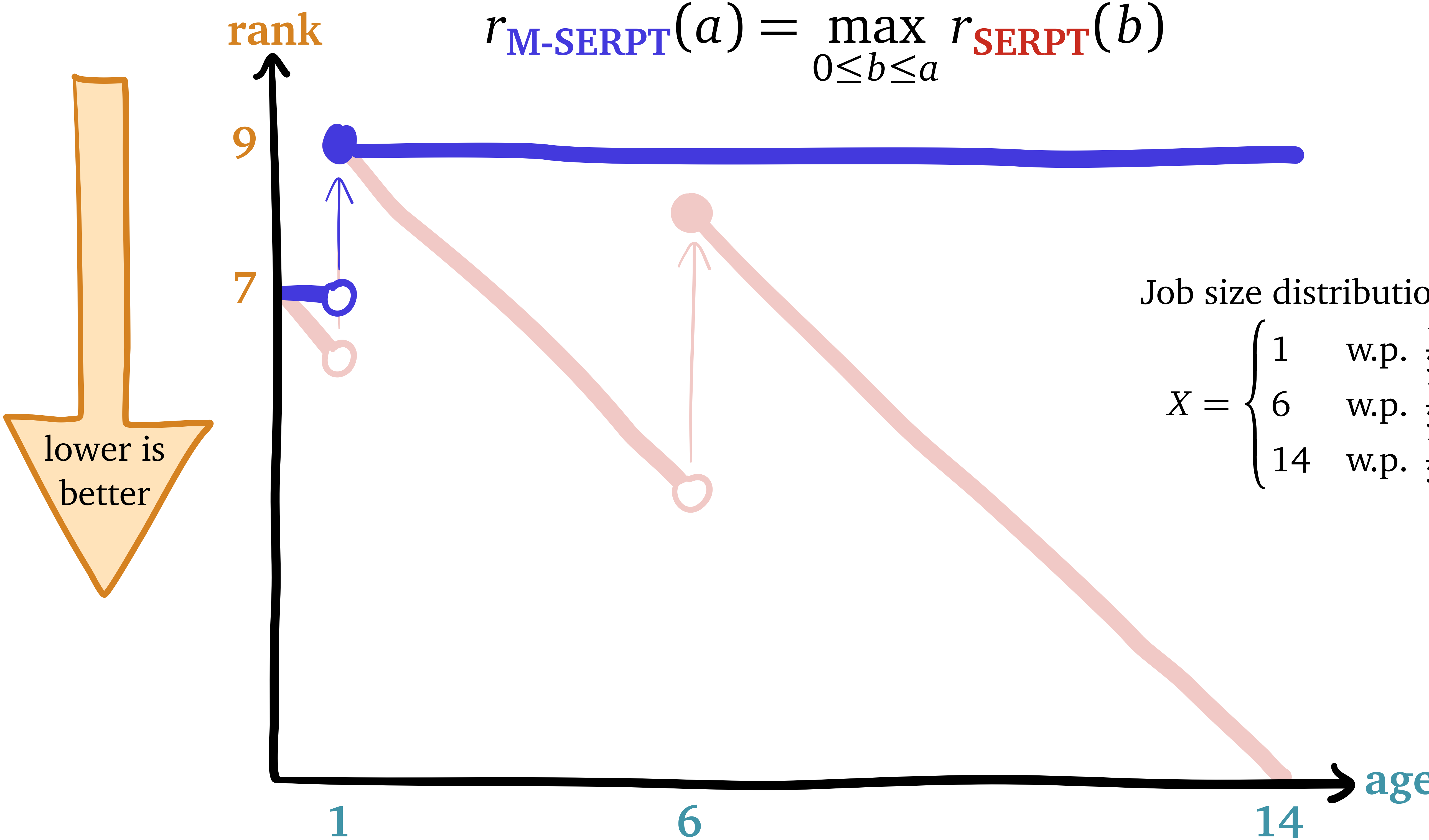


Job size distribution:

$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

M-SERPT Example

$$r_{\text{M-SERPT}}(a) = \max_{0 \le b \le a} r_{\text{SERPT}}(b)$$



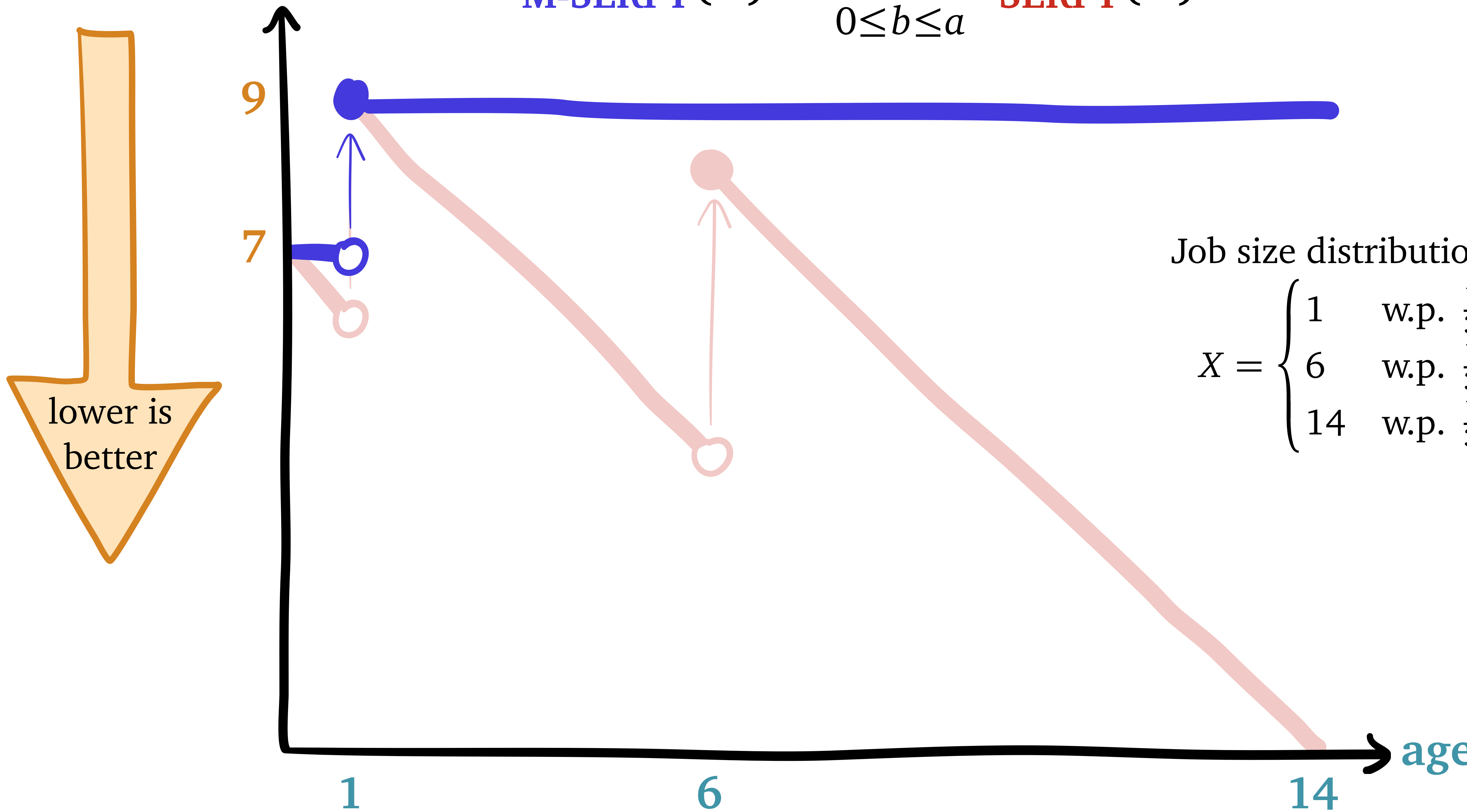
Job size distribution:

$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

M-SERPT Example

monotonic

$$r_{\text{M-SERPT}}(a) = \max_{0 \le b \le a} r_{\text{SERPT}}(b)$$



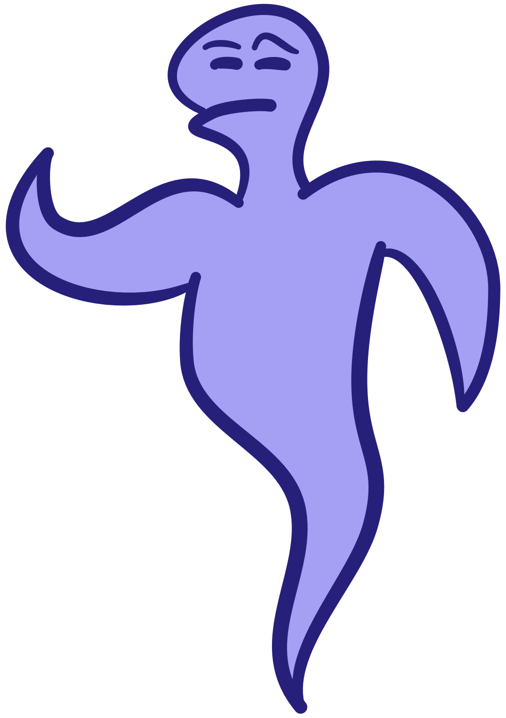
Job size distribution:

$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

M-SERPT

A new policy with both!

- *simple* definition like **SERPT**
- *provable* guarantee on $\mathbf{E}[T]$ like **Gittins**



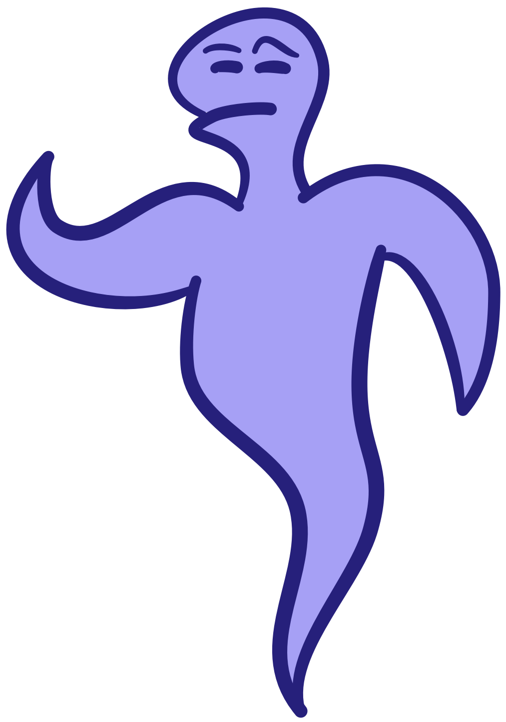
M-SERPT

A new policy with both!

- *simple* definition like **SERPT**
- *provable* guarantee on $\mathbf{E}[T]$ like **Gittins**

Definition:

$$r_{\text{M-SERPT}}(a) = \max_{0 \le b \le a} r_{\text{SERPT}}(b)$$



M-SERPT

A new policy with both!

- *simple* definition like **SERPT**
- *provable* guarantee on $\mathbf{E}[T]$ like **Gittins**

Definition:

$$r_{\text{M-SERPT}}(a) = \max_{0 \le b \le a} r_{\text{SERPT}}(b)$$

Theorem:

$$\frac{\mathbf{E}[T_{\text{M-SERPT}}]}{\mathbf{E}[T_{\text{Gittins}}]} \le 5$$

M-SERPT

A new policy with both!

- *simple* definition like **SERPT**
- *provable* guarantee on $\mathbf{E}[T]$ like **Gittins**

Definition:

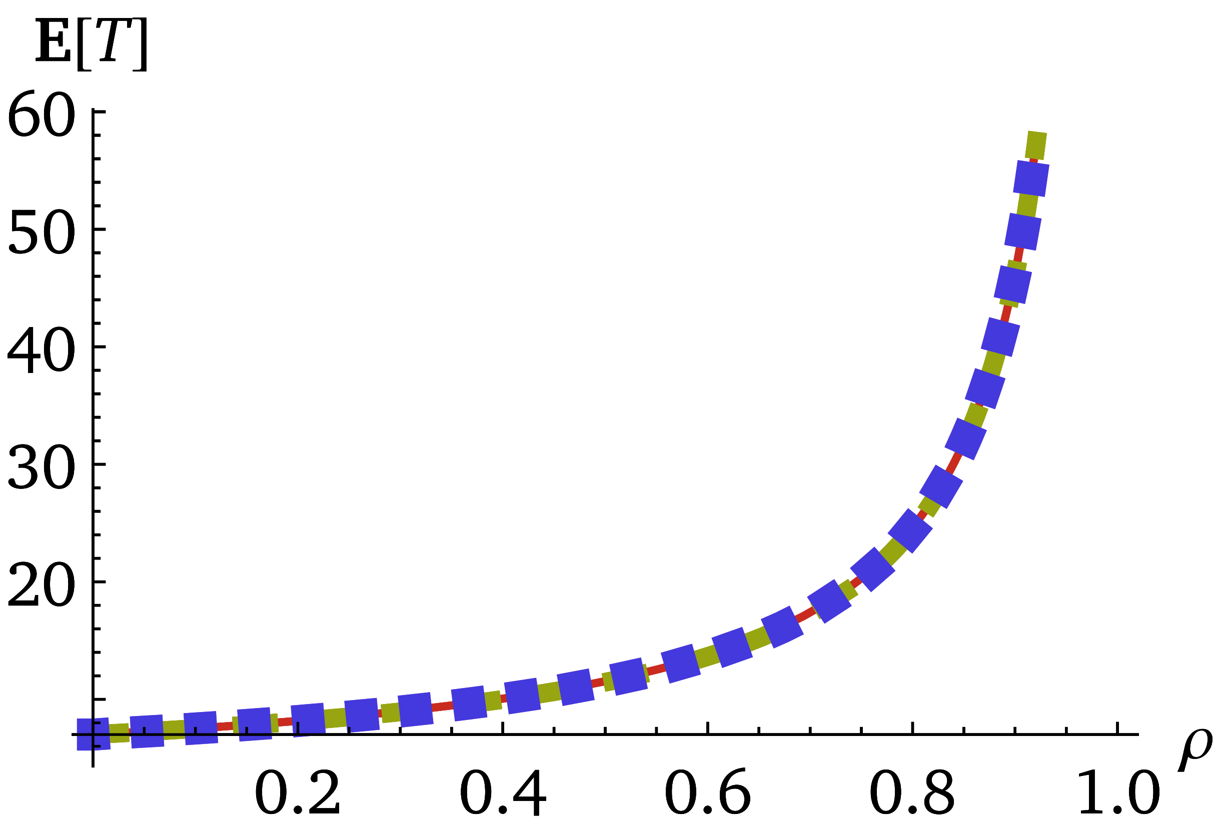
$$r_{\text{M-SERPT}}(a) = \max_{0 \le b \le a} r_{\text{SERPT}}(b)$$

Th

smaller at low load
first constant ratio

$$\frac{\mathbf{E}[T_{\text{M-SERPT}}]}{\mathbf{E}[T_{\text{Gittins}}]} \le 5$$

M-SERPT Performance



Job size distribution:

$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{3} \\ 6 & \text{w.p. } \frac{1}{3} \\ 14 & \text{w.p. } \frac{1}{3} \end{cases}$$

SERPT

Gittins

M-SERPT

Questions:

1. Why **M-SERPT**? Why not **SERPT**?
2. Why the factor of 5?

Response Time Decomposition

Response Time Decomposition

$$T = Q + R$$

Response Time Decomposition

$$T = Q + R$$

T = response time

Response Time Decomposition

"*queue* delay"

$$T = Q + R$$

T = response time

Response Time Decomposition

"*queue* delay"

$$T = Q + R$$

T = response time

Q = delays due to jobs in *queue*

Response Time Decomposition

"*queue* delay"

$$T = Q + R$$

"*run* delay"

T = response time

Q = delays due to jobs in *queue*

Response Time Decomposition

“*queue* delay”

$$T = Q + R$$

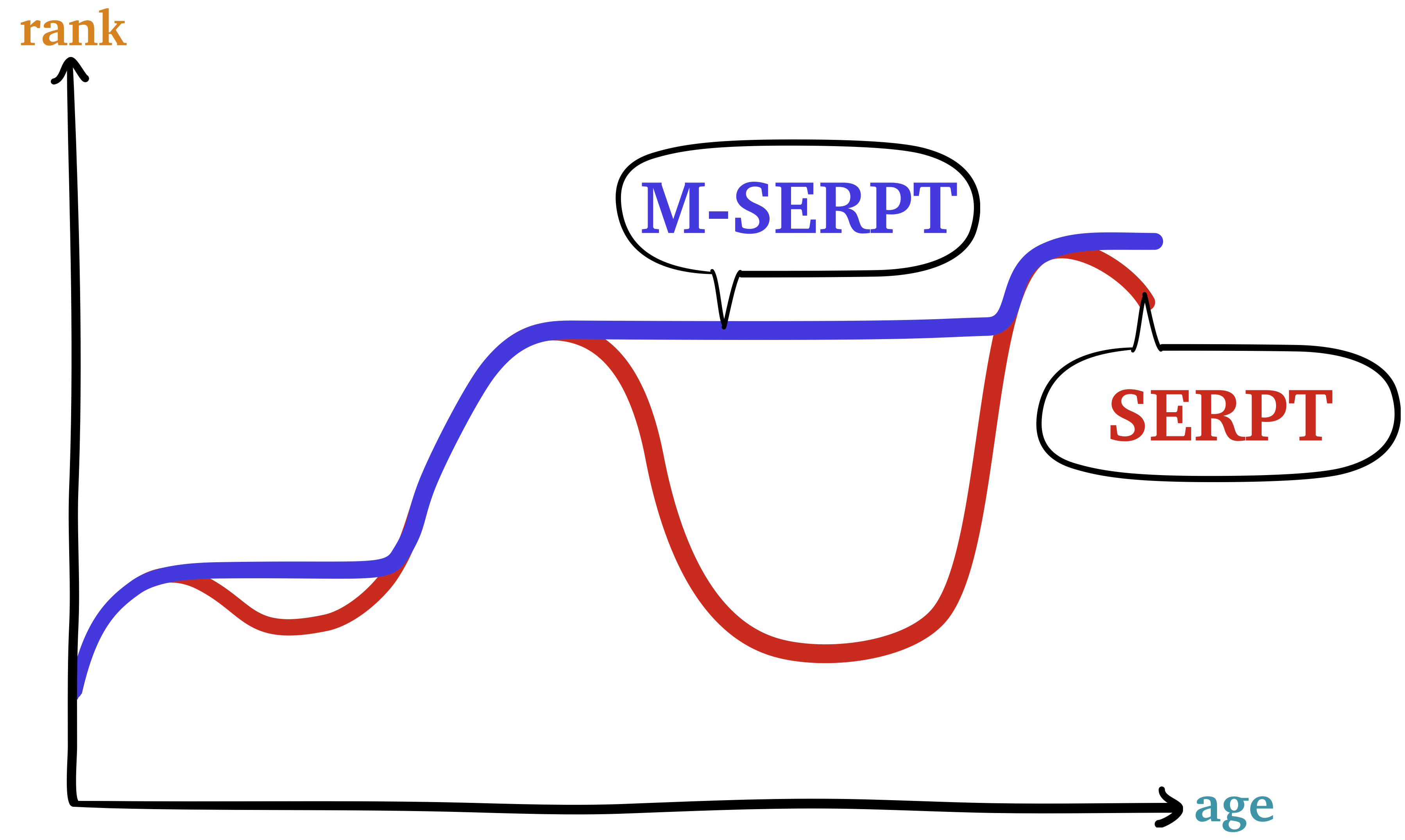
“*run* delay”

T = response time

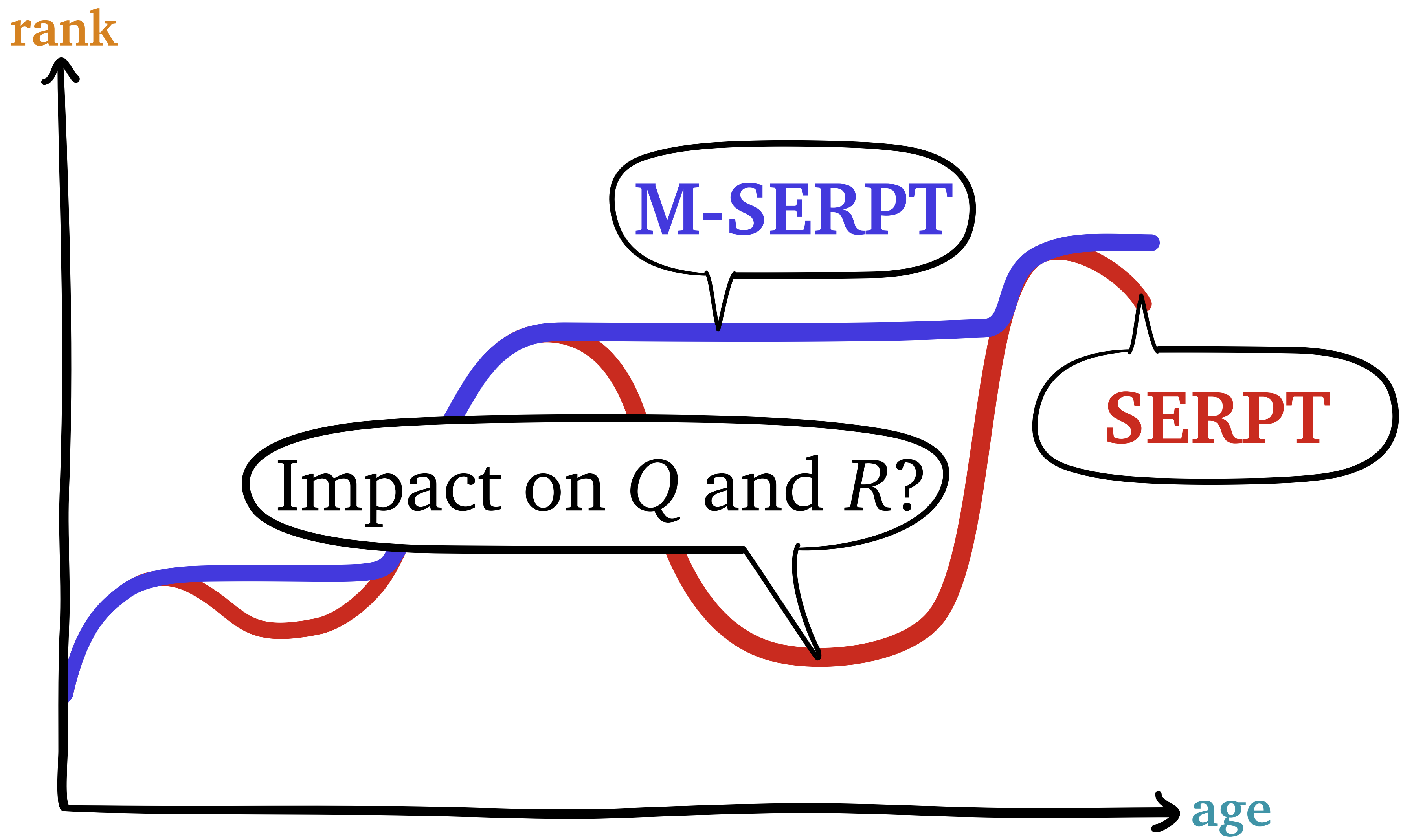
Q = delays due to jobs in *queue*

R = my size + delays while I’m *running*

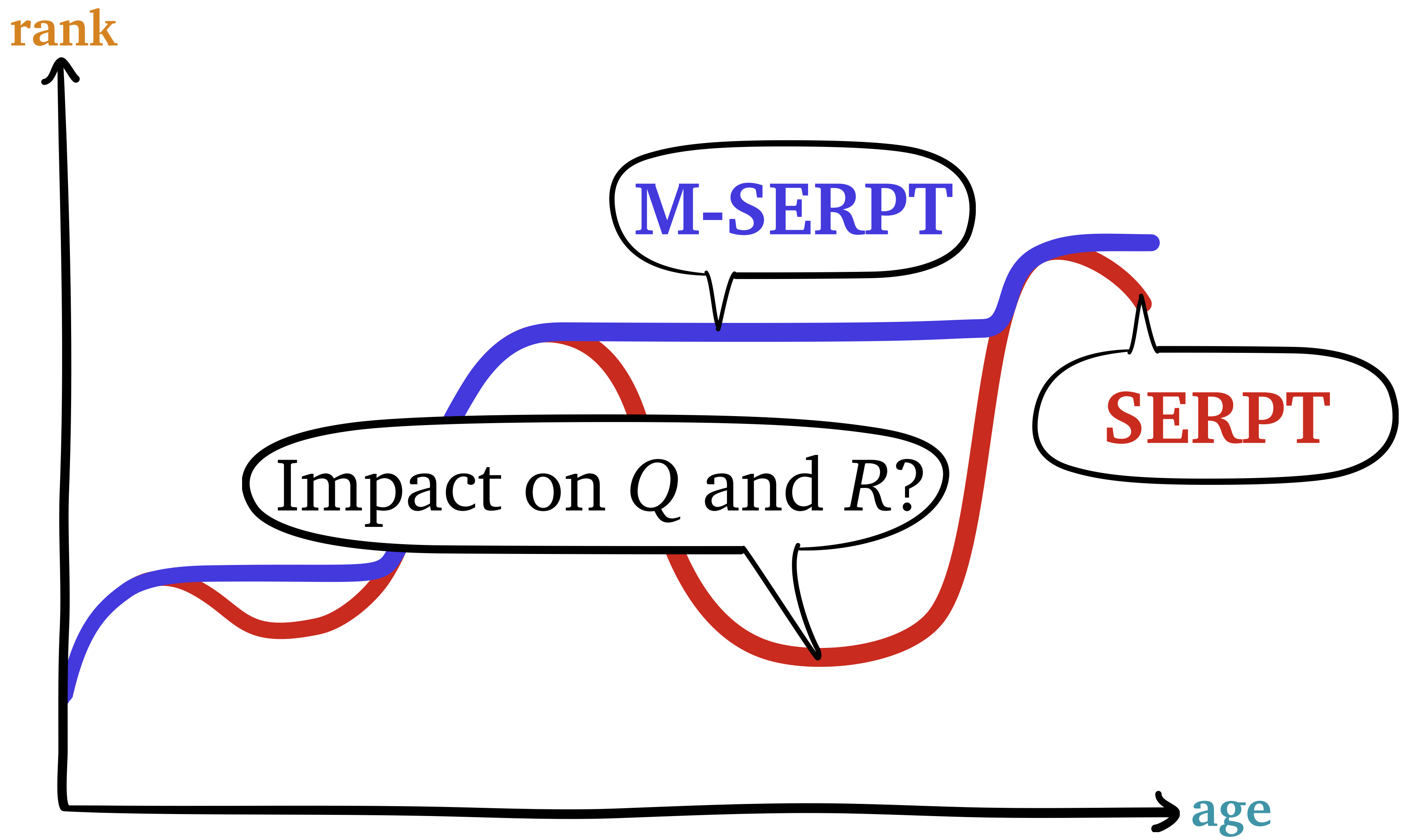
SERPT vs. M-SERPT



SERPT vs. M-SERPT



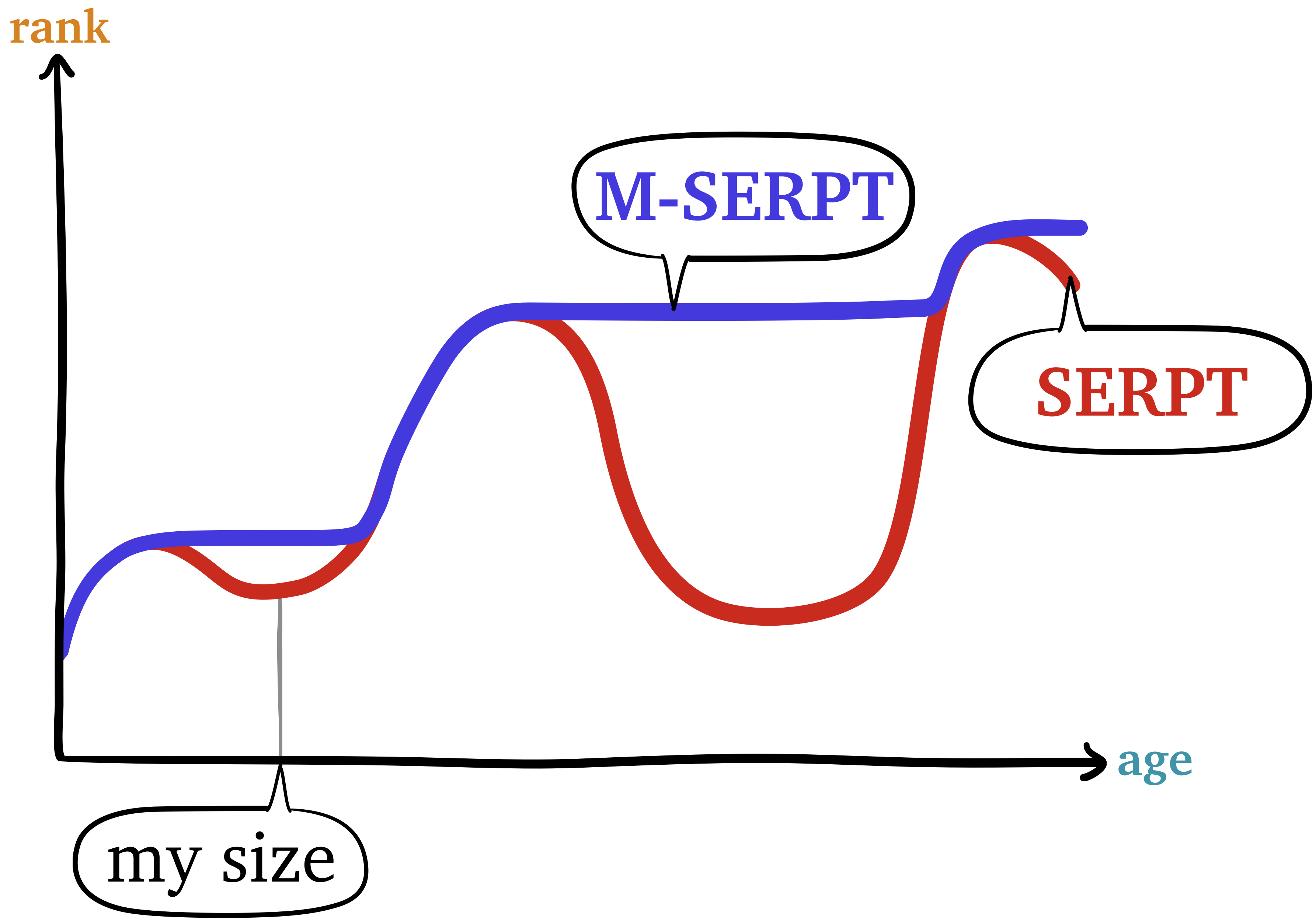
SERPT vs. M-SERPT



$$\mathbf{E}[Q_{\text{SERPT}}] \; ? \; \mathbf{E}[Q_{\text{M-SERPT}}]$$

$$\mathbf{E}[R_{\text{SERPT}}] \; ? \; \mathbf{E}[R_{\text{M-SERPT}}]$$

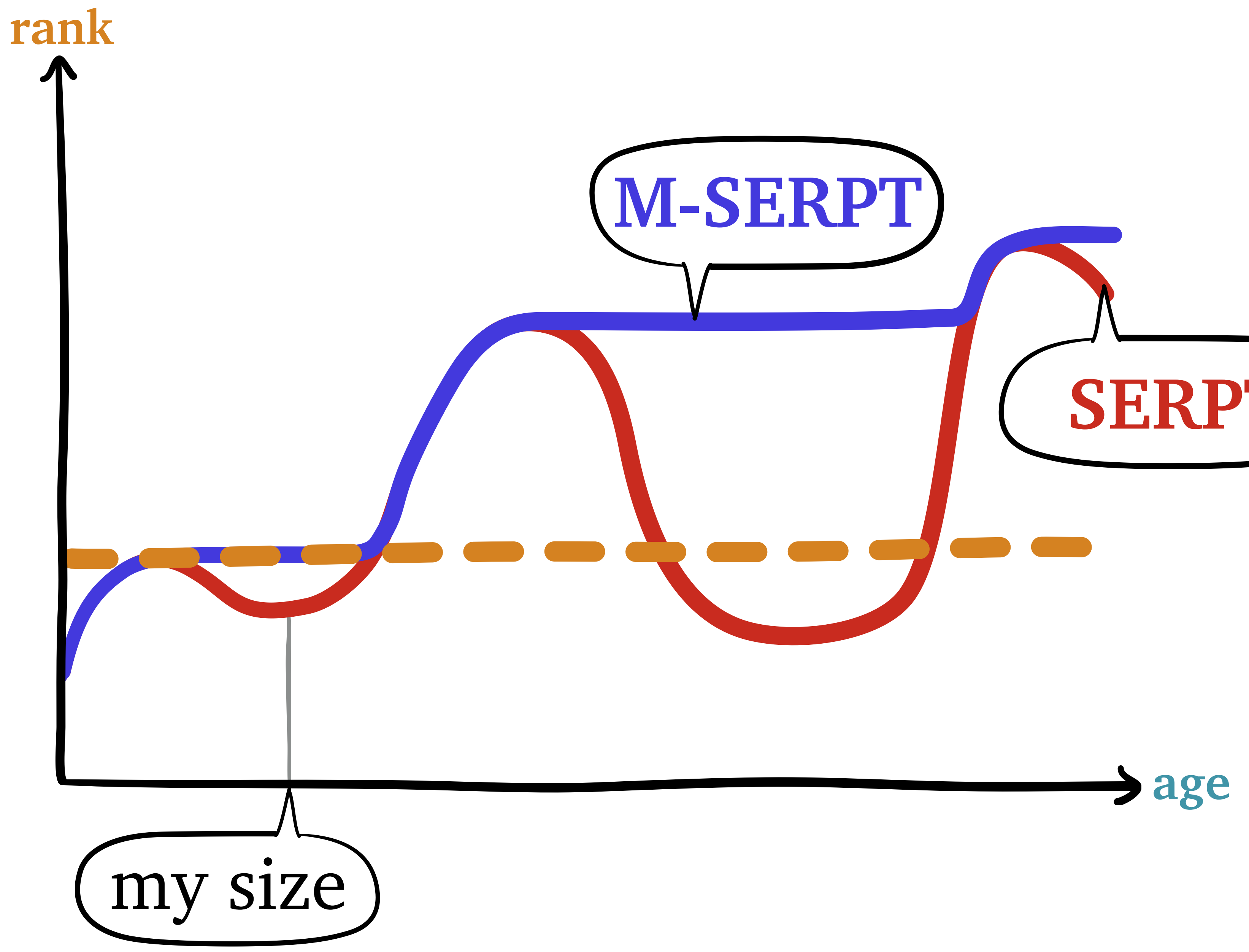
SERPT vs. M-SERPT



$$\mathbf{E}[Q_{\text{SERPT}}] \ ? \ \mathbf{E}[Q_{\text{M-SERPT}}]$$

$$\mathbf{E}[R_{\text{SERPT}}] \ ? \ \mathbf{E}[R_{\text{M-SERPT}}]$$

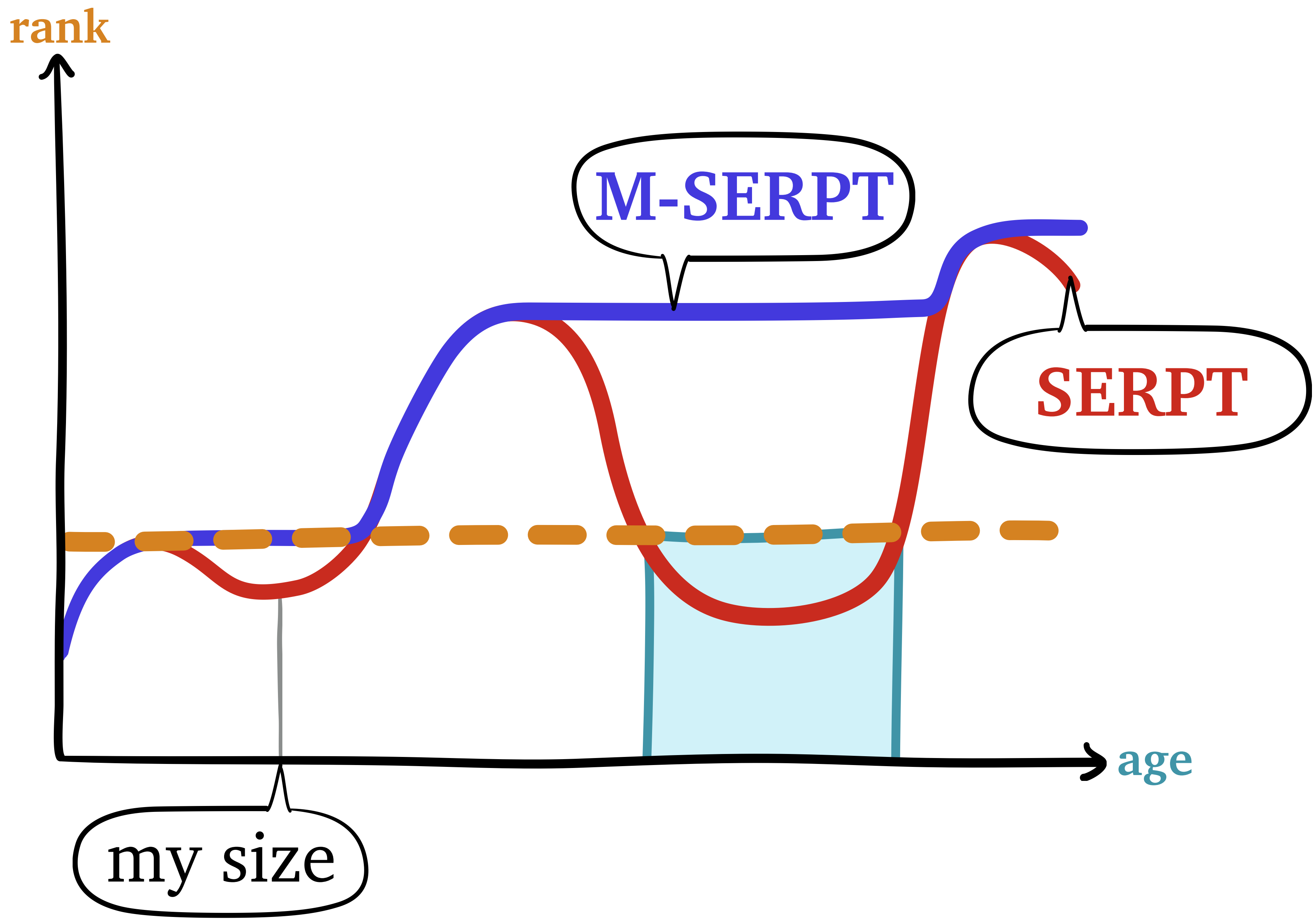
SERPT vs. M-SERPT



$$\mathbf{E}[Q_{\text{SERPT}}] \ ? \ \mathbf{E}[Q_{\text{M-SERPT}}]$$

$$\mathbf{E}[R_{\text{SERPT}}] \ ? \ \mathbf{E}[R_{\text{M-SERPT}}]$$

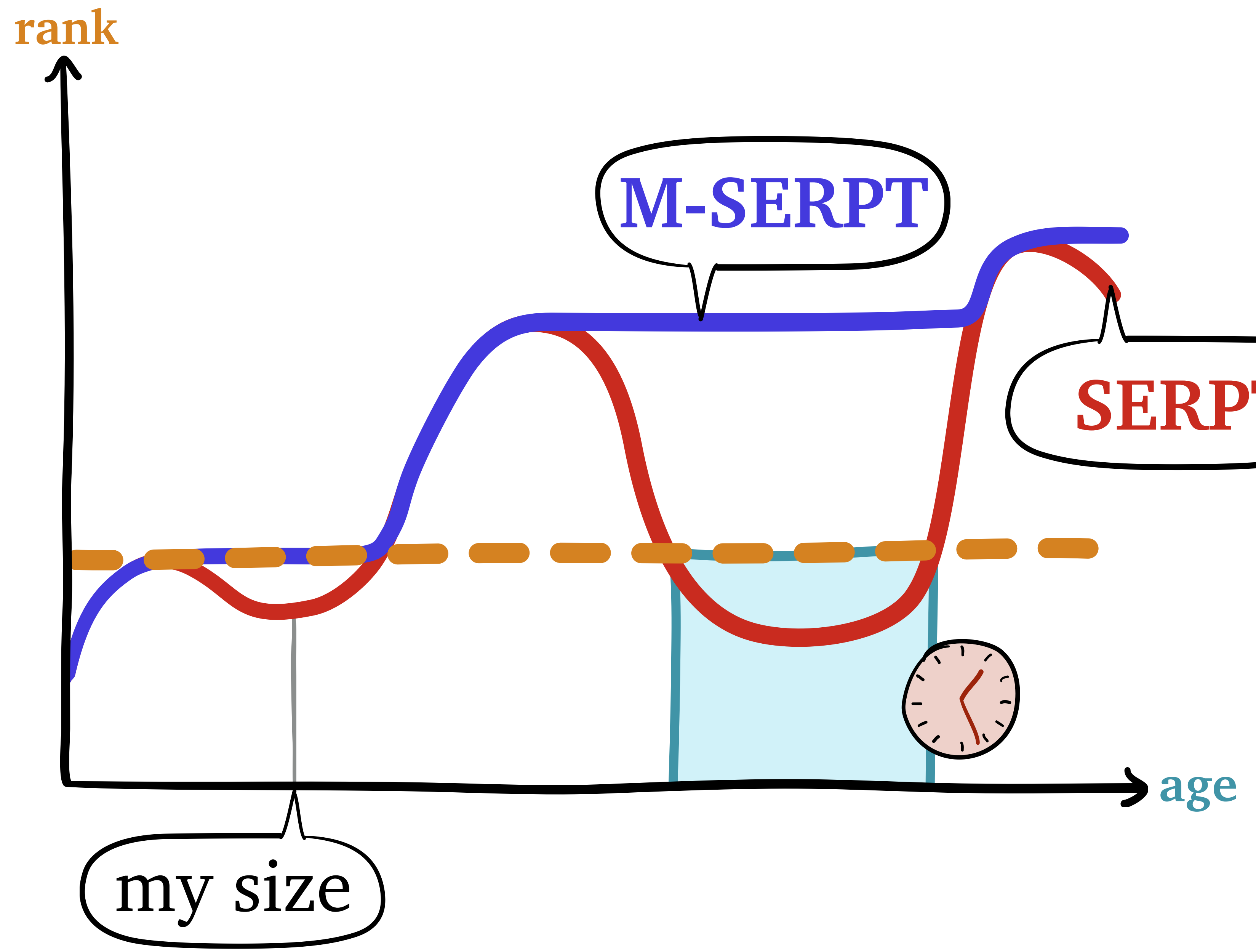
SERPT vs. M-SERPT



$$\mathbf{E}[Q_{\text{SERPT}}] \ ? \ \mathbf{E}[Q_{\text{M-SERPT}}]$$

$$\mathbf{E}[R_{\text{SERPT}}] \ ? \ \mathbf{E}[R_{\text{M-SERPT}}]$$

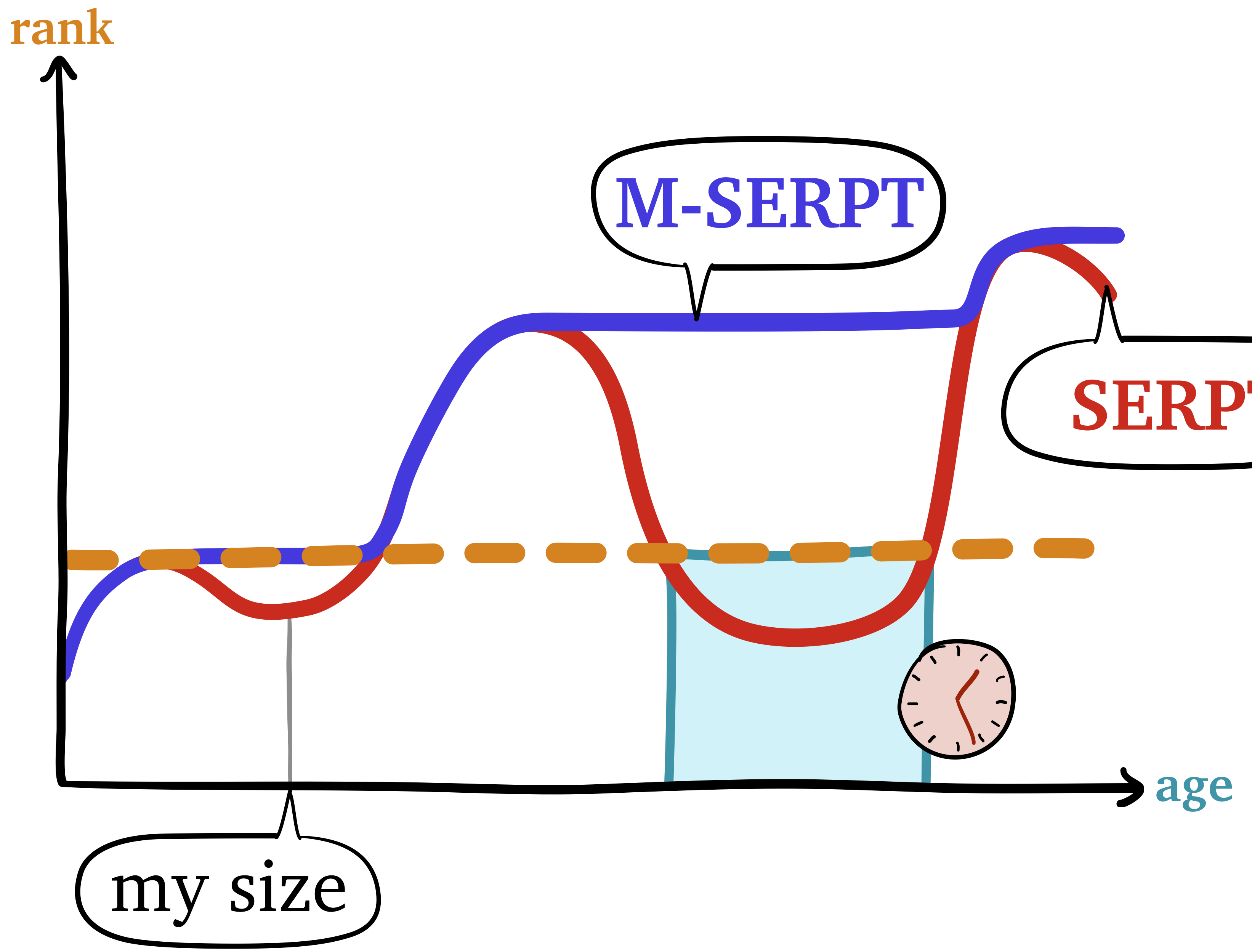
SERPT vs. M-SERPT



$\mathbf{E}[Q_{\text{SERPT}}]$? $\mathbf{E}[Q_{\text{M-SERPT}}]$

$\mathbf{E}[R_{\text{SERPT}}]$? $\mathbf{E}[R_{\text{M-SERPT}}]$

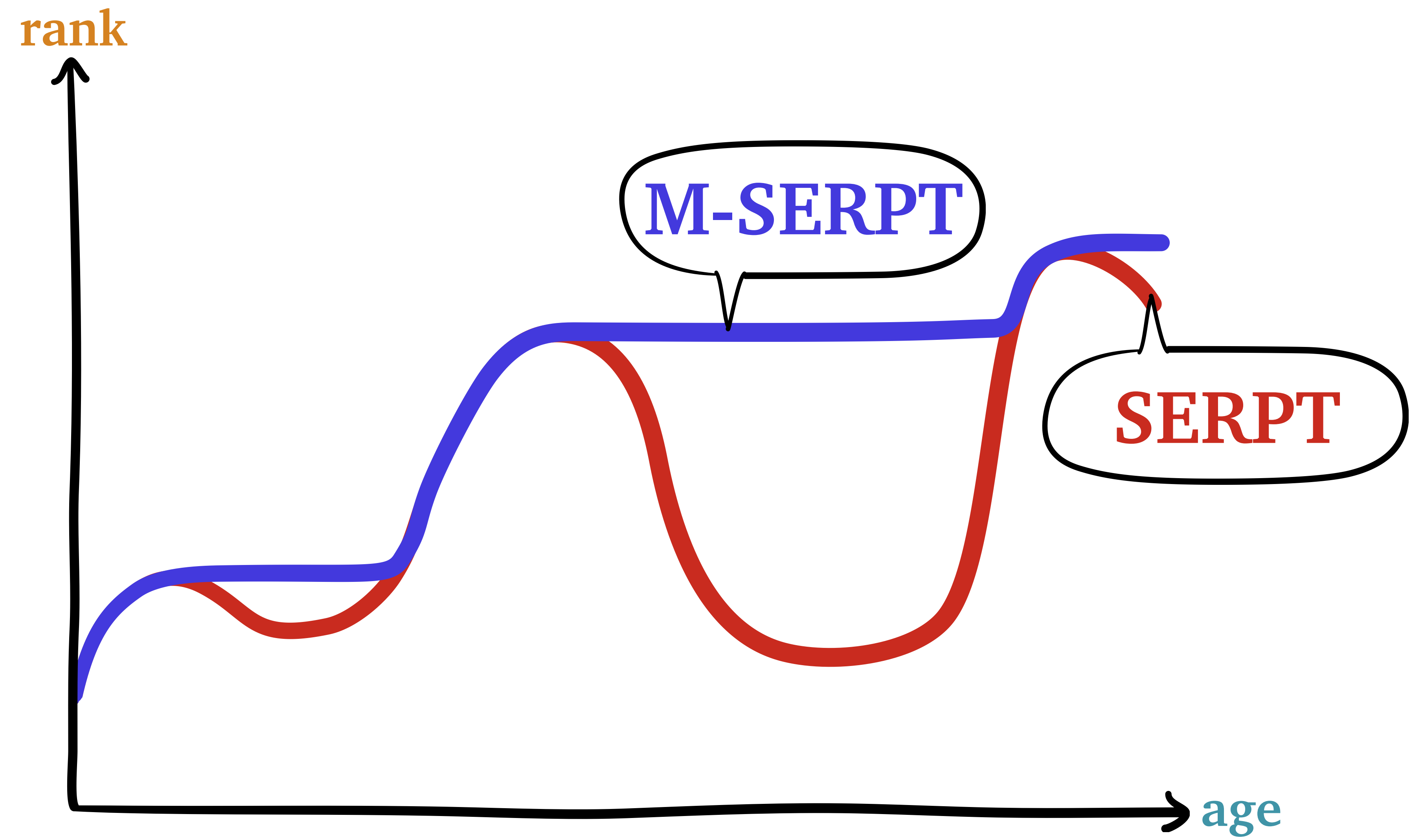
SERPT vs. M-SERPT



$$\mathbf{E}[Q_{\text{SERPT}}] \geq \mathbf{E}[Q_{\text{M-SERPT}}]$$

$$\mathbf{E}[R_{\text{SERPT}}] \ ? \ \mathbf{E}[R_{\text{M-SERPT}}]$$

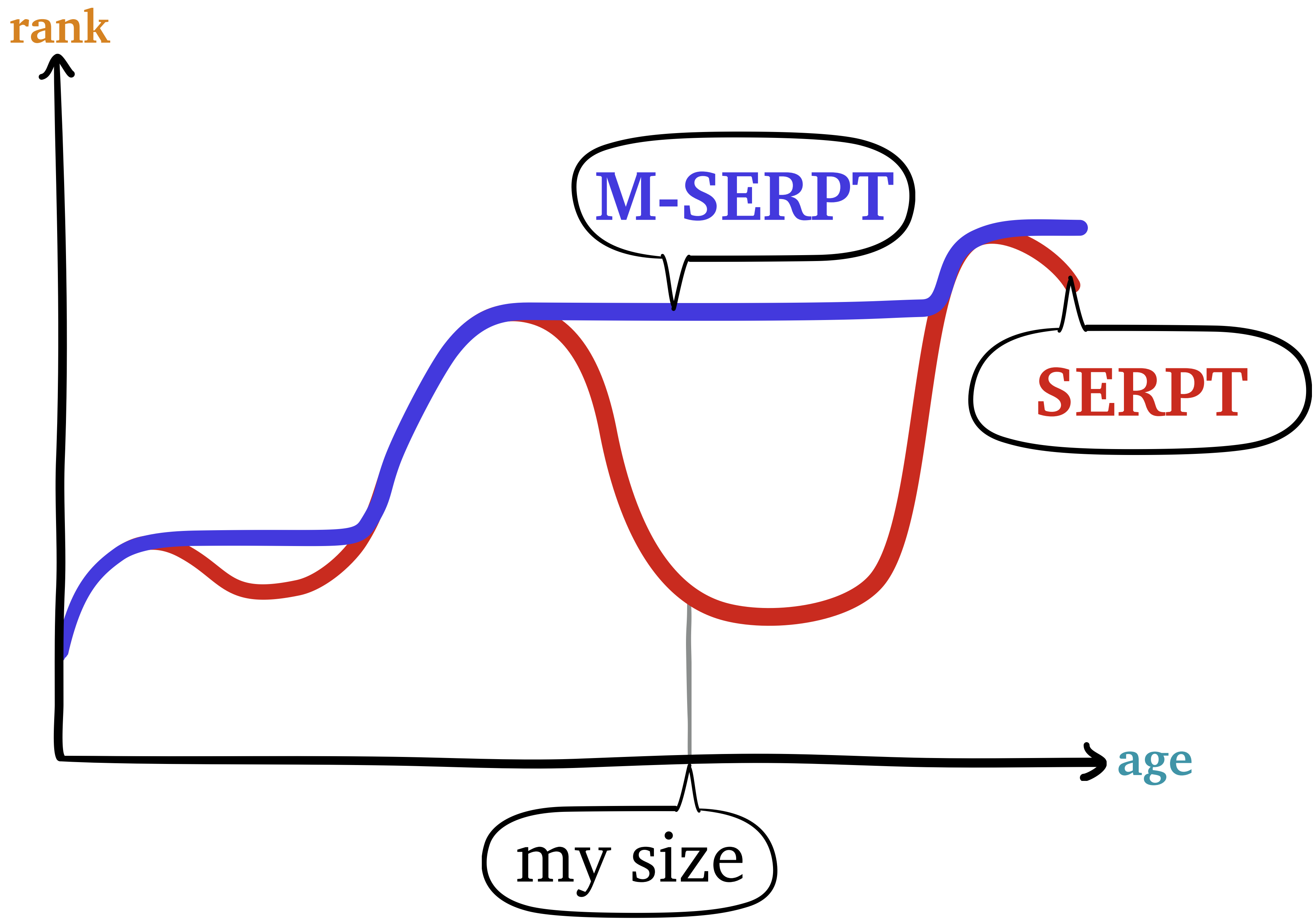
SERPT vs. M-SERPT



$$\mathbf{E}[Q_{\text{SERPT}}] \geq \mathbf{E}[Q_{\text{M-SERPT}}]$$

$$\mathbf{E}[R_{\text{SERPT}}] \;?\; \mathbf{E}[R_{\text{M-SERPT}}]$$

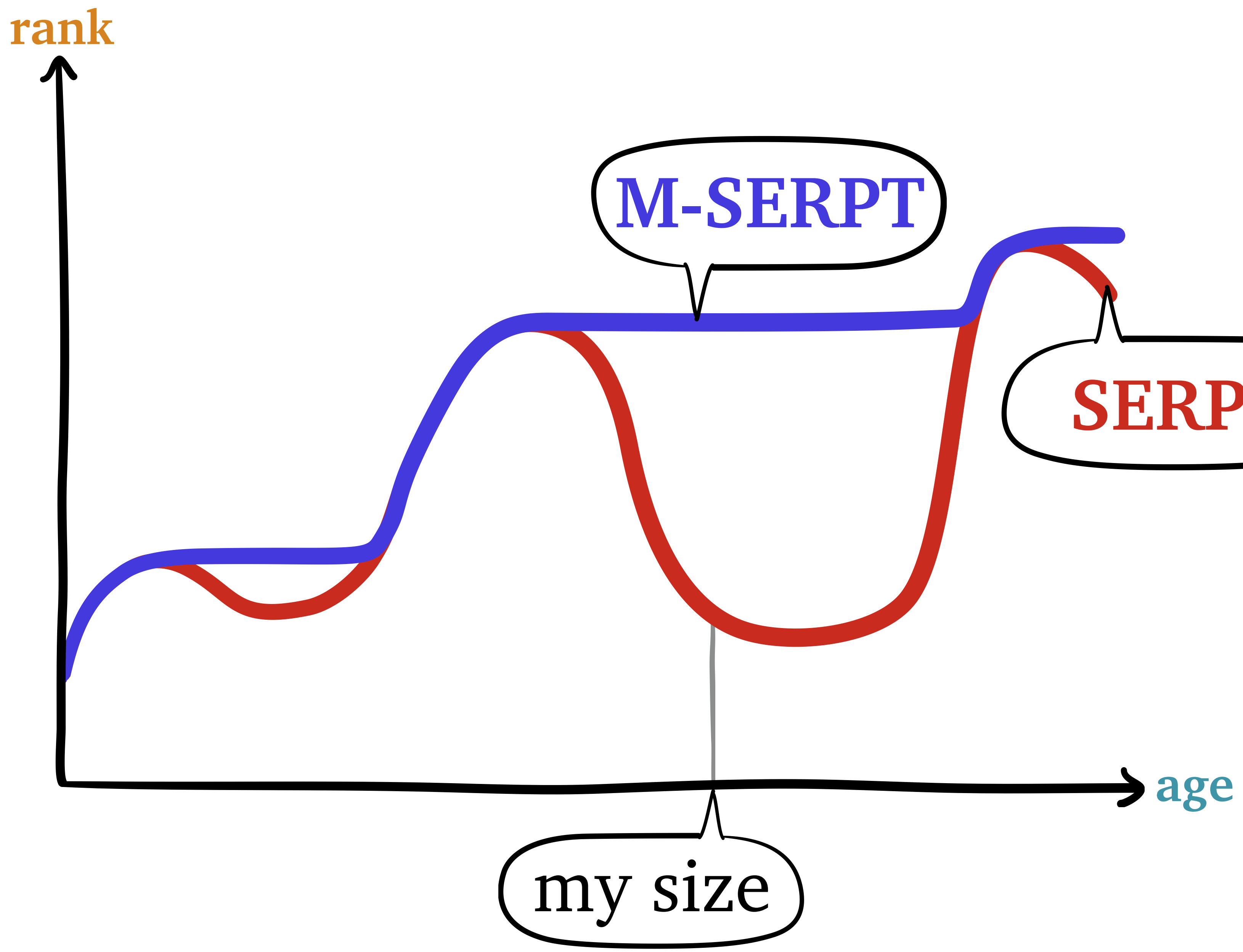
SERPT vs. M-SERPT



$$\mathbf{E}[Q_{\text{SERPT}}] \geq \mathbf{E}[Q_{\text{M-SERPT}}]$$

$$\mathbf{E}[R_{\text{SERPT}}] \ ? \ \mathbf{E}[R_{\text{M-SERPT}}]$$

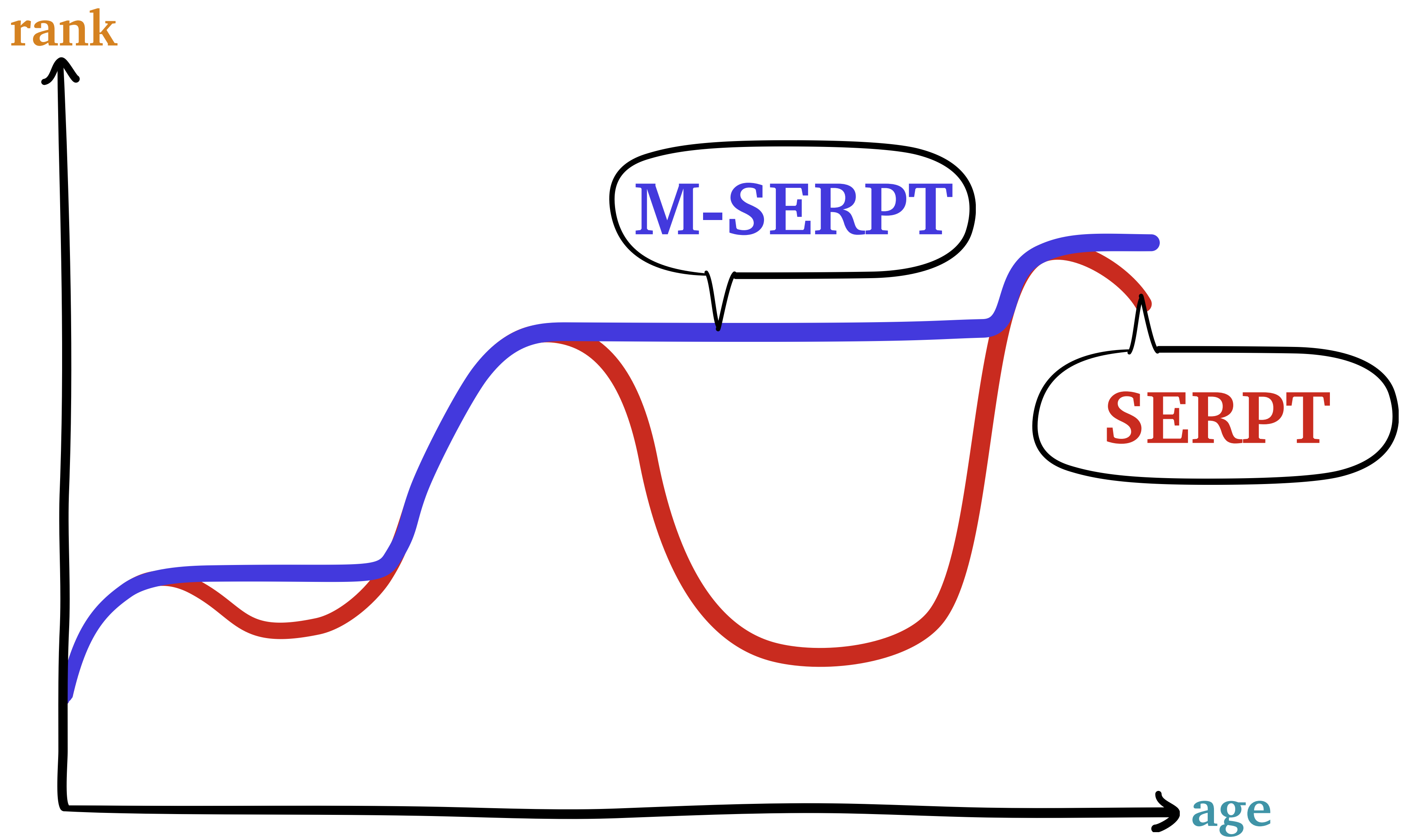
SERPT vs. M-SERPT



$$\mathbf{E}[Q_{\text{SERPT}}] \geq \mathbf{E}[Q_{\text{M-SERPT}}]$$

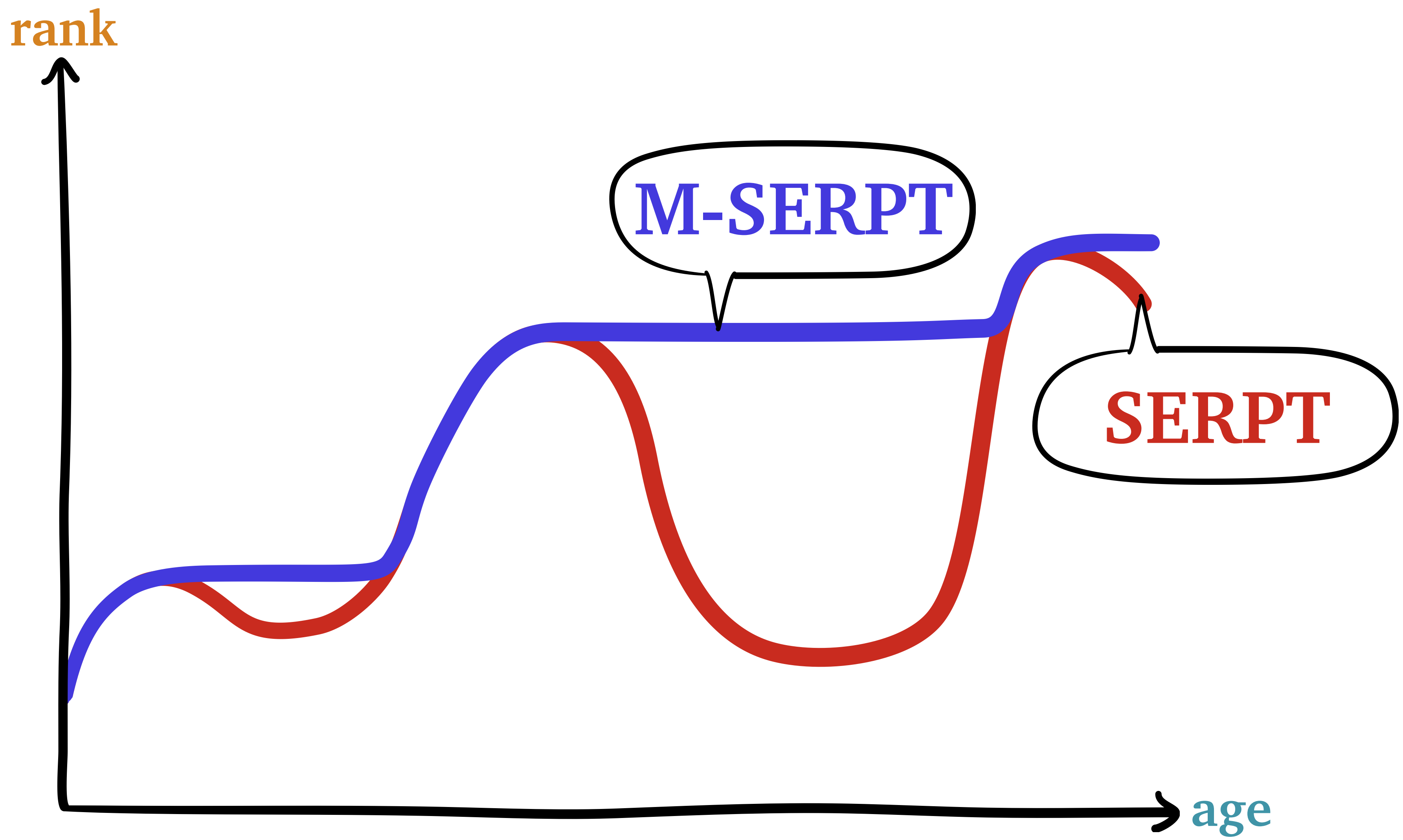
$$\mathbf{E}[R_{\text{SERPT}}] \leq \mathbf{E}[R_{\text{M-SERPT}}]$$

SERPT vs. M-SERPT



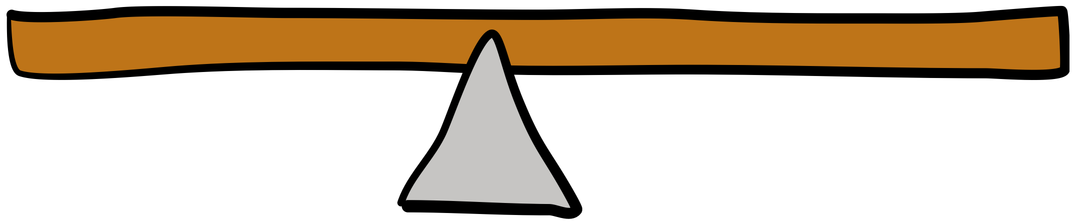
$$\mathbf{E}[Q_{\text{SERPT}}] \geq \mathbf{E}[Q_{\text{M-SERPT}}]$$
$$\mathbf{E}[R_{\text{SERPT}}] \leq \mathbf{E}[R_{\text{M-SERPT}}]$$

SERPT vs. M-SERPT

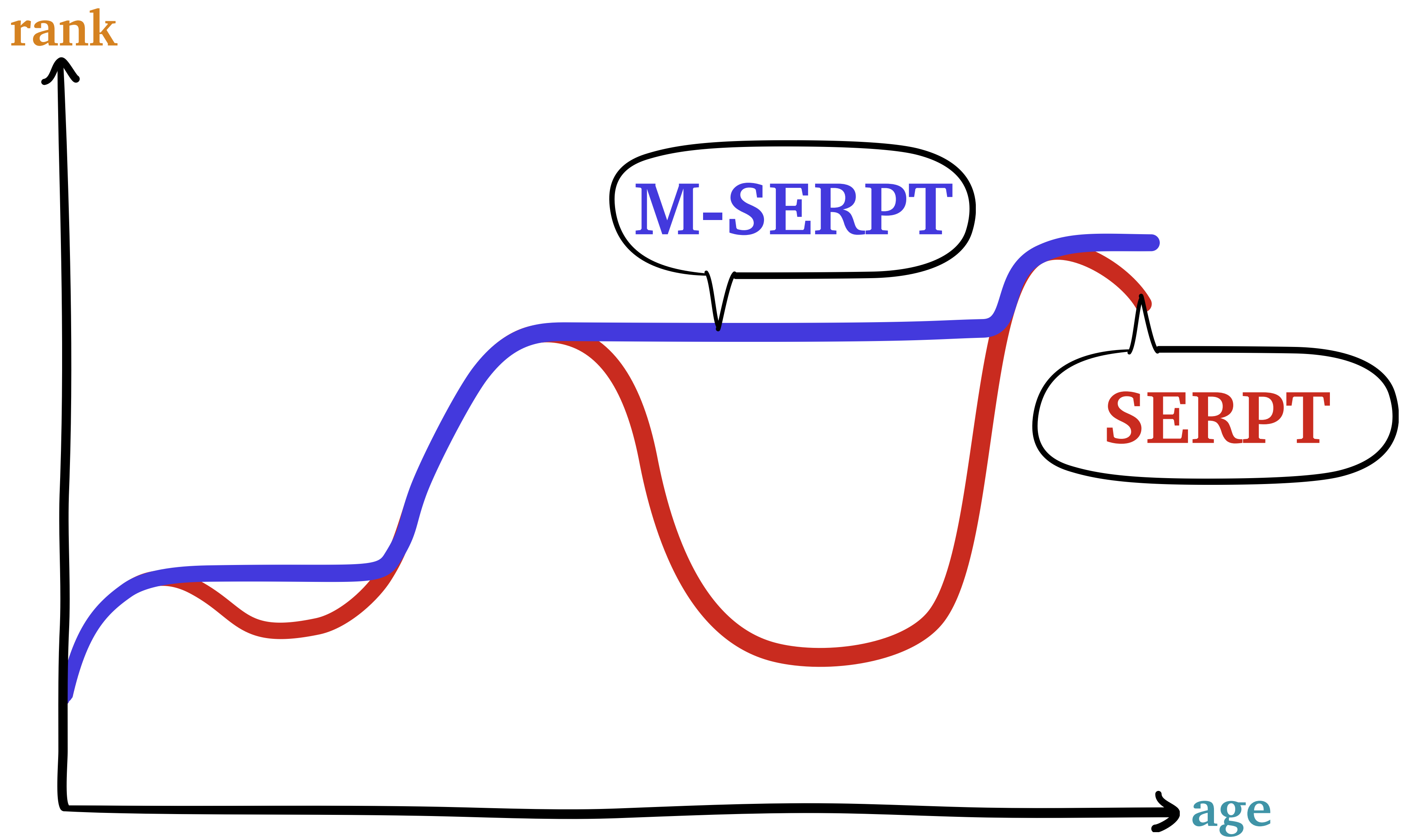


$$\mathbf{E}[Q_{\text{SERPT}}] \geq \mathbf{E}[Q_{\text{M-SERPT}}]$$

$$\mathbf{E}[R_{\text{SERPT}}] \leq \mathbf{E}[R_{\text{M-SERPT}}]$$

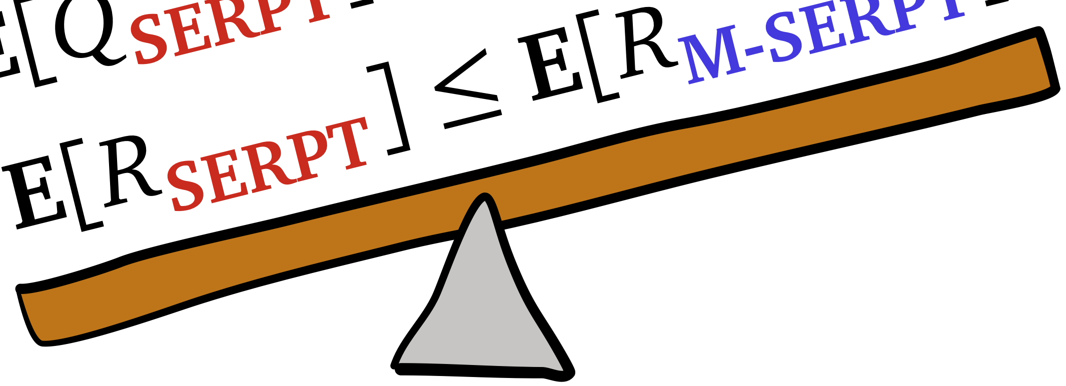


SERPT vs. M-SERPT

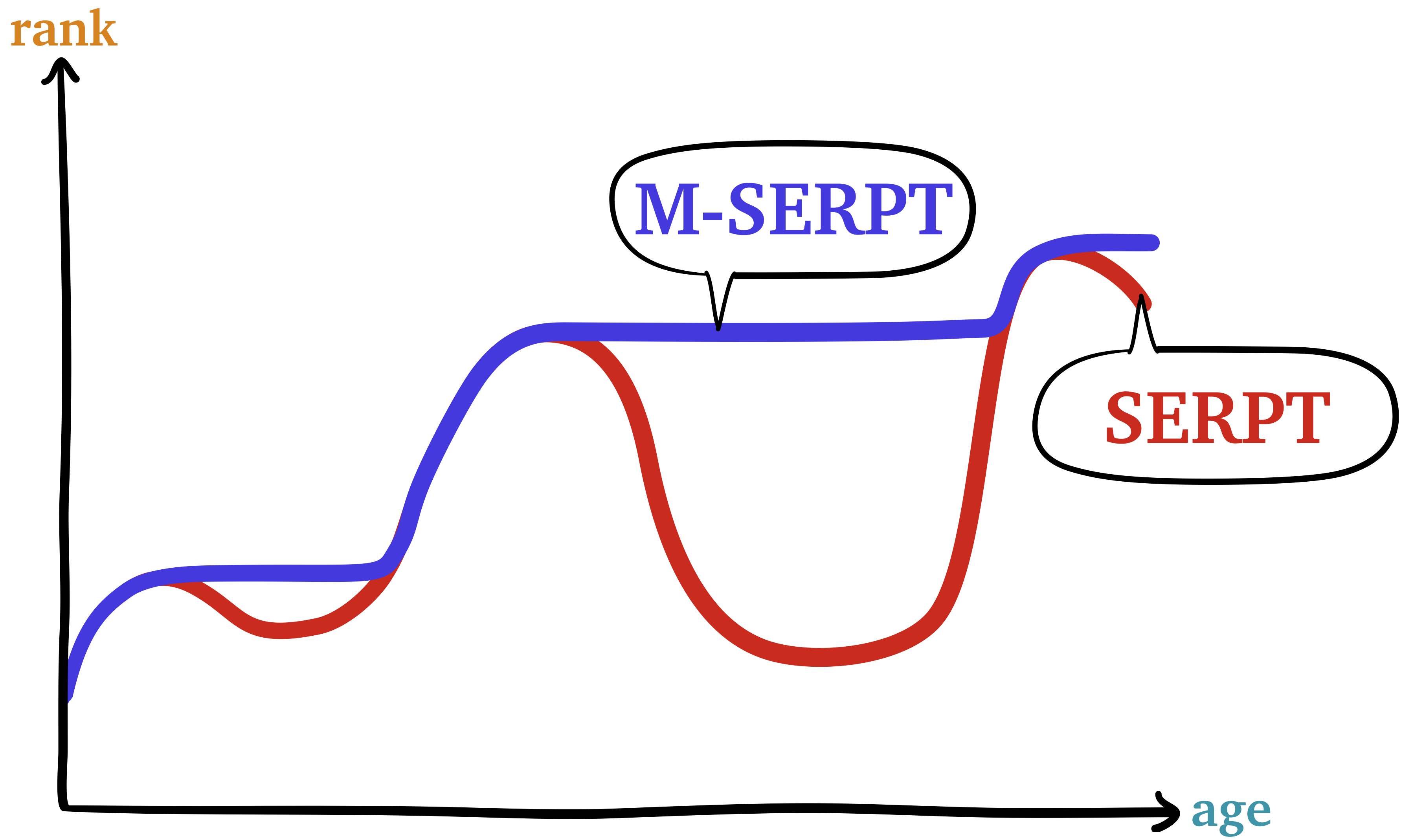


$$\mathbf{E}[Q_{\text{SERPT}}] \geq \mathbf{E}[Q_{\text{M-SERPT}}]$$

$$\mathbf{E}[R_{\text{SERPT}}] \leq \mathbf{E}[R_{\text{M-SERPT}}]$$

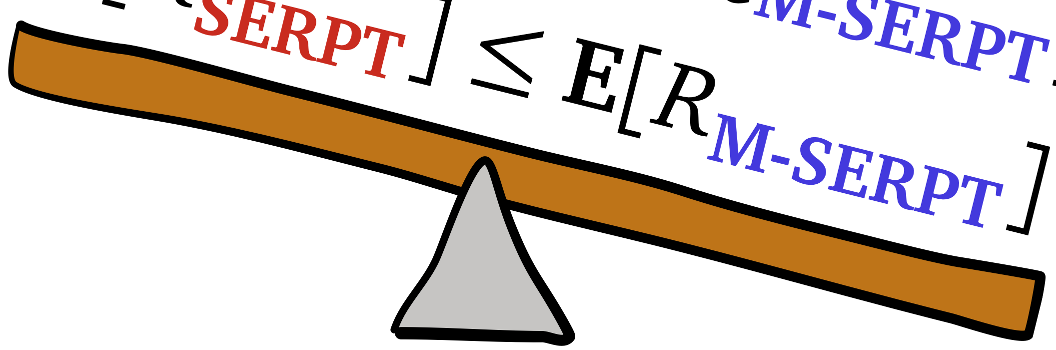


SERPT vs. M-SERPT

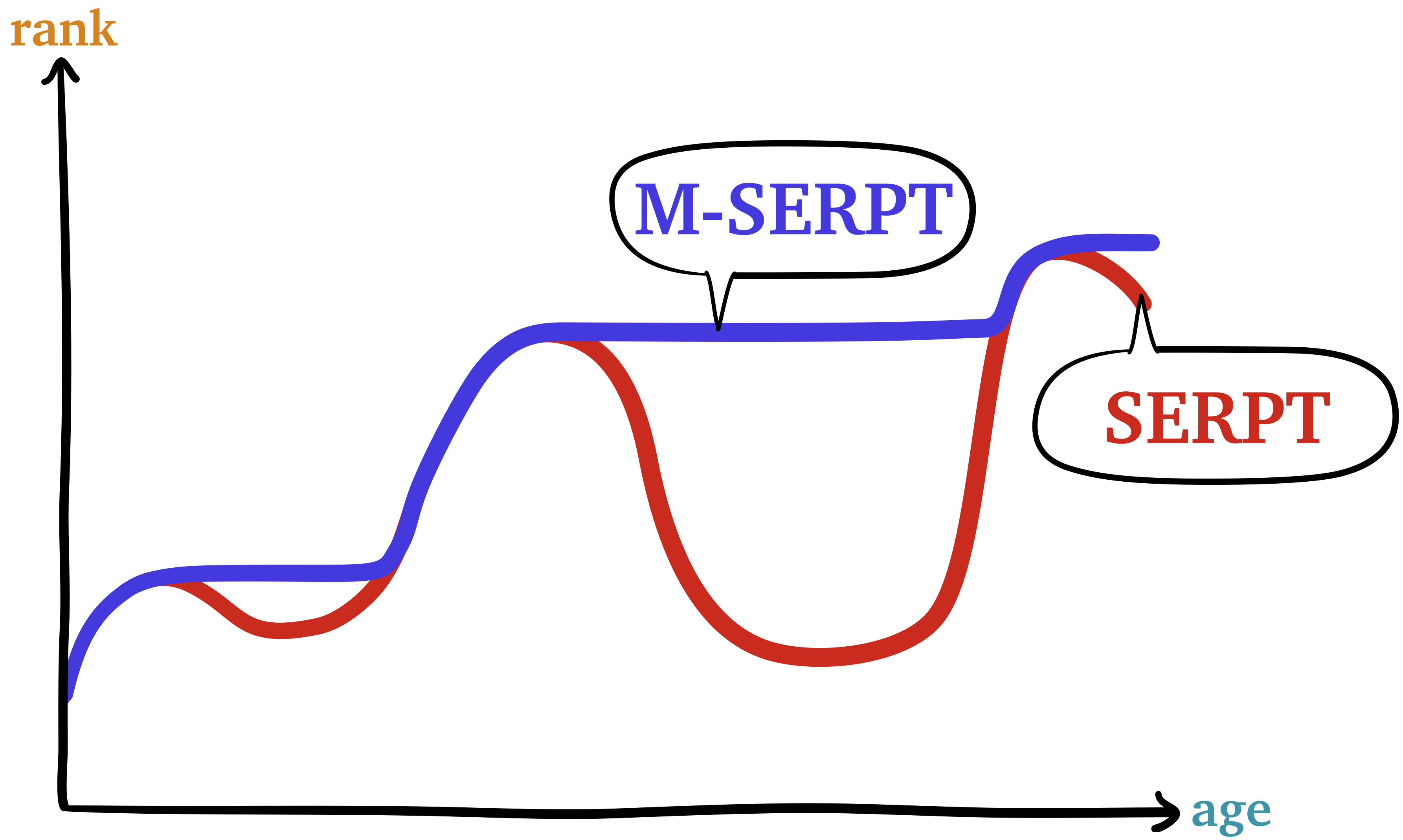


$$\mathbf{E}[Q_{\text{SERPT}}] \geq \mathbf{E}[Q_{\text{M-SERPT}}]$$

$$\mathbf{E}[R_{\text{SERPT}}] \leq \mathbf{E}[R_{\text{M-SERPT}}]$$

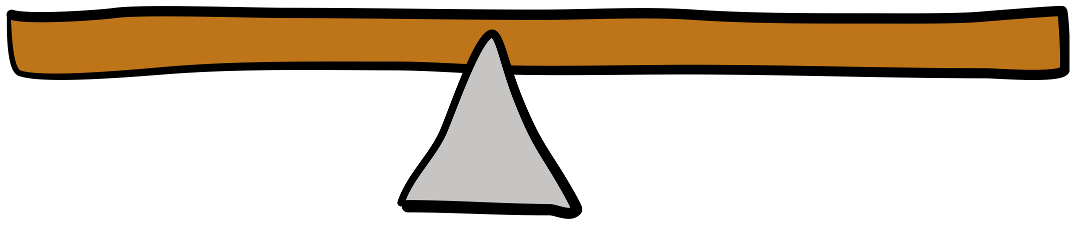


SERPT vs. M-SERPT

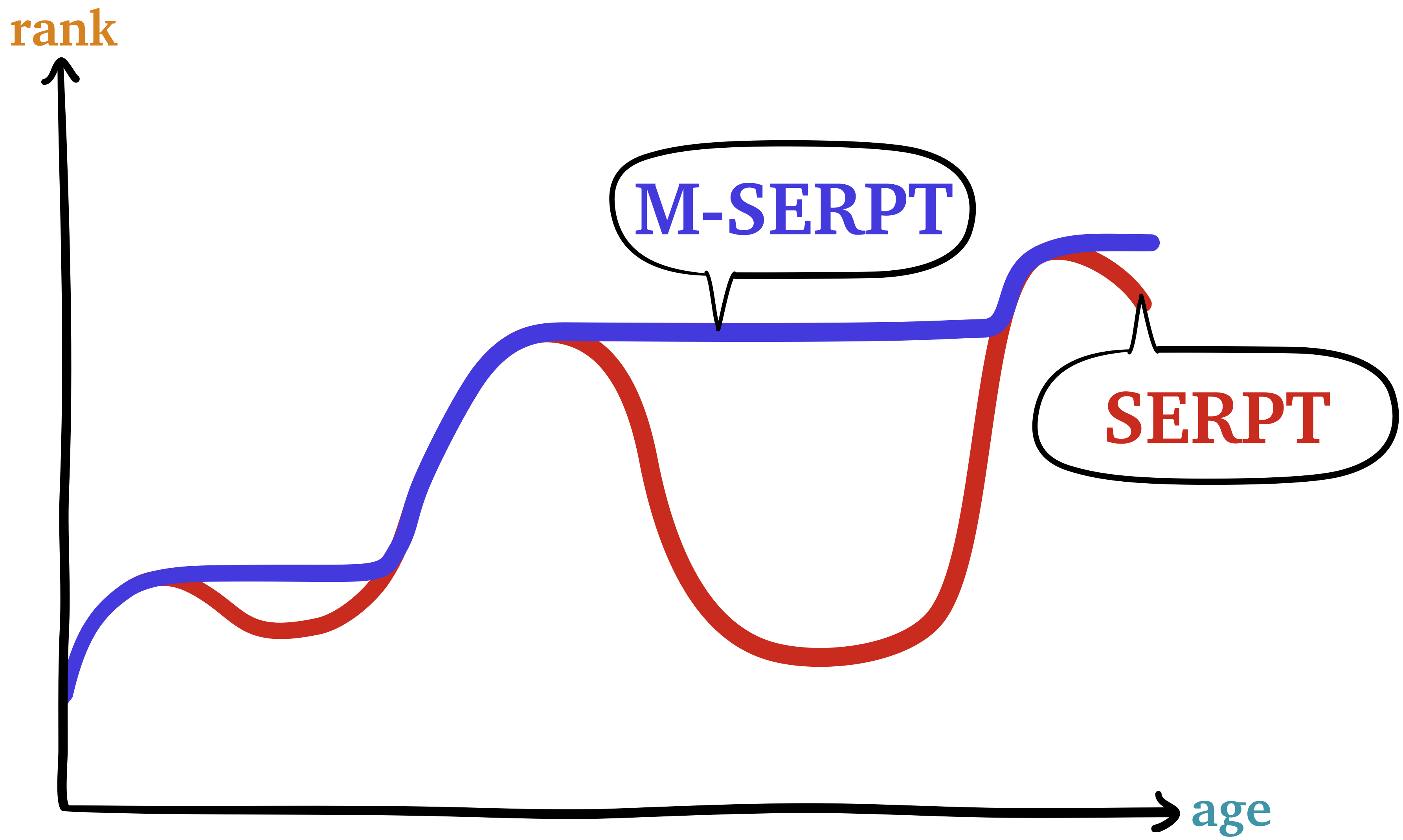


$$\mathbf{E}[Q_{\text{SERPT}}] \geq \mathbf{E}[Q_{\text{M-SERPT}}]$$

$$\mathbf{E}[R_{\text{SERPT}}] \leq \mathbf{E}[R_{\text{M-SERPT}}]$$

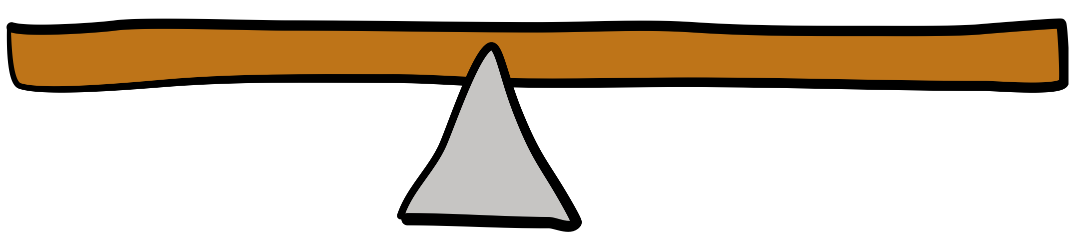


SERPT vs. M-SERPT



$$\mathbf{E}[Q_{\text{SERPT}}] \geq \mathbf{E}[Q_{\text{M-SERPT}}]$$

$$\mathbf{E}[R_{\text{SERPT}}] \leq \mathbf{E}[R_{\text{M-SERPT}}]$$



M-SERPT's Factor of 5

M-SERPT's Factor of 5

$$\mathbf{E}[Q_{\textbf{M-SERPT}}] \leq 2 \cdot \mathbf{E}[Q_{\textbf{Gittins}}]$$

M-SERPT's Factor of 5

$$\mathbf{E}[Q_{\text{M-SERPT}}] \leq 2 \cdot \mathbf{E}[Q_{\text{Gittins}}]$$

$$\mathbf{E}[R_{\text{M-SERPT}}] \leq \mathbf{E}[Q_{\text{M-SERPT}}] + \frac{\mathbf{E}[X]}{\rho} \log \frac{1}{1-\rho}$$

M-SERPT's Factor of 5

$$\mathbf{E}[Q_{\textbf{M-SERPT}}] \leq 2 \cdot \mathbf{E}[Q_{\textbf{Gittins}}]$$

$$\begin{aligned} \mathbf{E}[R_{\textbf{M-SERPT}}] &\leq \mathbf{E}[Q_{\textbf{M-SERPT}}] + \frac{\mathbf{E}[X]}{\rho} \log \frac{1}{1-\rho} \\ &\leq 2 \cdot \mathbf{E}[Q_{\textbf{Gittins}}] + \mathbf{E}[T_{\textbf{SRPT}}] \end{aligned}$$

M-SERPT's Factor of 5

$\leq 2 \cdot \mathbf{E}[T_{\text{Gittins}}]$

$$\mathbf{E}[Q_{\text{M-SERPT}}] \leq 2 \cdot \mathbf{E}[Q_{\text{Gittins}}]$$

$$\mathbf{E}[R_{\text{M-SERPT}}] \leq \mathbf{E}[Q_{\text{M-SERPT}}] + \frac{\mathbf{E}[X]}{\rho} \log \frac{1}{1-\rho}$$

$$\leq 2 \cdot \mathbf{E}[Q_{\text{Gittins}}] + \mathbf{E}[T_{\text{SRPT}}]$$

M-SERPT's Factor of 5

$\leq 2 \cdot \mathbf{E}[T_{\text{Gittins}}]$

$$\mathbf{E}[Q_{\text{M-SERPT}}] \leq 2 \cdot \mathbf{E}[Q_{\text{Gittins}}]$$

$$\mathbf{E}[R_{\text{M-SERPT}}] \leq \mathbf{E}[Q_{\text{M-SERPT}}] + \frac{\mathbf{E}[X]}{\rho} \log \frac{1}{1-\rho}$$

$$\leq 2 \cdot \mathbf{E}[Q_{\text{Gittins}}] + \mathbf{E}[T_{\text{SRPT}}]$$

$\leq 2 \cdot \mathbf{E}[T_{\text{Gittins}}]$

M-SERPT's Factor of 5

$\leq 2 \cdot \mathbf{E}[T_{\text{Gittins}}]$

$$\mathbf{E}[Q_{\text{M-SERPT}}] \leq 2 \cdot \mathbf{E}[Q_{\text{Gittins}}]$$

$$\mathbf{E}[R_{\text{M-SERPT}}] \leq \mathbf{E}[Q_{\text{M-SERPT}}] + \frac{\mathbf{E}[X]}{\rho} \log \frac{1}{1-\rho}$$

$$\leq 2 \cdot \mathbf{E}[Q_{\text{Gittins}}] + \mathbf{E}[T_{\text{SRPT}}]$$

$\leq 2 \cdot \mathbf{E}[T_{\text{Gittins}}]$ $\leq \mathbf{E}[T_{\text{Gittins}}]$

M-SERPT's Factor of 5

$\leq 2 \cdot \mathbf{E}[T_{\text{Gittins}}]$

$$\mathbf{E}[Q_{\text{M-SERPT}}] \leq 2 \cdot \mathbf{E}[Q_{\text{Gittins}}]$$

$$\mathbf{E}[R_{\text{M-SERPT}}] \leq \mathbf{E}[Q_{\text{M-SERPT}}] + \frac{\mathbf{E}[X]}{\rho} \log \frac{1}{1-\rho}$$

$$\leq 2 \cdot \mathbf{E}[Q_{\text{Gittins}}] + \mathbf{E}[T_{\text{SRPT}}]$$

$\leq 2 \cdot \mathbf{E}[T_{\text{Gittins}}]$

$\leq \mathbf{E}[T_{\text{Gittins}}]$

$$\frac{\mathbf{E}[T_{\text{M-SERPT}}]}{\mathbf{E}[T_{\text{Gittins}}]} \leq 2 + 2 + 1 = 5$$

M-SERPT's Factor of 5

$$\mathbf{E}[Q_{\text{M-SERPT}}] \le 2 \cdot \mathbf{E}[Q_{\text{Gittins}}]$$

$$\begin{aligned}\mathbf{E}[R_{\text{M-SERPT}}] &\le \mathbf{E}[Q_{\text{M-SERPT}}] + \frac{\mathbf{E}[X]}{\rho} \log \frac{1}{1-\rho} \\ &\le 2 \cdot \mathbf{E}[Q_{\text{Gittins}}] + \mathbf{E}[T_{\text{SRPT}}]\end{aligned}$$

$$\frac{\mathbf{E}[T_{\text{M-SERPT}}]}{\mathbf{E}[T_{\text{Gittins}}]} \le 2 + 2 + 1 = 5$$

M-SERPT's Factor of 5

tight bound

$$\mathbf{E}[Q_{\text{M-SERPT}}] \leq 2 \cdot \mathbf{E}[Q_{\text{Gittins}}]$$

$$\mathbf{E}[R_{\text{M-SERPT}}] \leq \mathbf{E}[Q_{\text{M-SERPT}}] + \frac{\mathbf{E}[X]}{\rho} \log \frac{1}{1-\rho}$$

$$\leq 2 \cdot \mathbf{E}[Q_{\text{Gittins}}] + \mathbf{E}[T_{\text{SRPT}}]$$

$$\frac{\mathbf{E}[T_{\text{M-SERPT}}]}{\mathbf{E}[T_{\text{Gittins}}]} \leq 2 + 2 + 1 = 5$$

M-SERPT's Factor of 5

tight bound

$$\mathbf{E}[Q_{\text{M-SERPT}}] \le 2 \cdot \mathbf{E}[Q_{\text{Gittins}}]$$

$$\mathbf{E}[R_{\text{M-SERPT}}] \le \mathbf{E}[Q_{\text{M-SERPT}}] + \frac{\mathbf{E}[X]}{\rho} \log \frac{1}{1-\rho}$$

$$\le 2 \cdot \mathbf{E}[Q_{\text{Gittins}}] + \mathbf{E}[T_{\text{SRPT}}]$$

loose bound

$$\frac{\mathbf{E}[T_{\text{M-SERPT}}]}{\mathbf{E}[T_{\text{Gittins}}]} \le 2 + 2 + 1 = 5$$

M-SERPT's Factor of 5

tight bound

$$\mathbf{E}[Q_{\text{M-SERPT}}] \leq 2 \cdot \mathbf{E}[Q_{\text{Gittins}}]$$

$$\mathbf{E}[R_{\text{M-SERPT}}] \leq \mathbf{E}[Q_{\text{M-SERPT}}] + \frac{\mathbf{E}[X]}{\rho} \log \frac{1}{1-\rho}$$

$$\leq 2 \cdot \mathbf{E}[Q_{\text{Git}}$$

loose bound

Open problem:

$$\frac{\mathbf{E}[T_{\text{M-SERPT}}]}{\mathbf{E}[T_{\text{Gittins}}]} \leq 2?$$

$$\frac{\mathbf{E}[T_{\text{M-SERPT}}]}{\mathbf{E}[T_{\text{Gittins}}]} \leq 2 + 2 + 1 = 5$$

M-SERPT

first scheduling policy with

- simple rank function
- constant approximation ratio for $\mathbf{E}[T]$

M-SERPT

first scheduling policy with

- simple rank function
- constant approximation ratio for $\mathbf{E}[T]$

$$\frac{\mathbf{E}[T_{\text{M-SERPT}}]}{\mathbf{E}[T_{\text{Gittins}}]} \le 5$$

M-SERPT

first scheduling policy with

- simple rank function
- constant approximation ratio for $\mathbf{E}[T]$

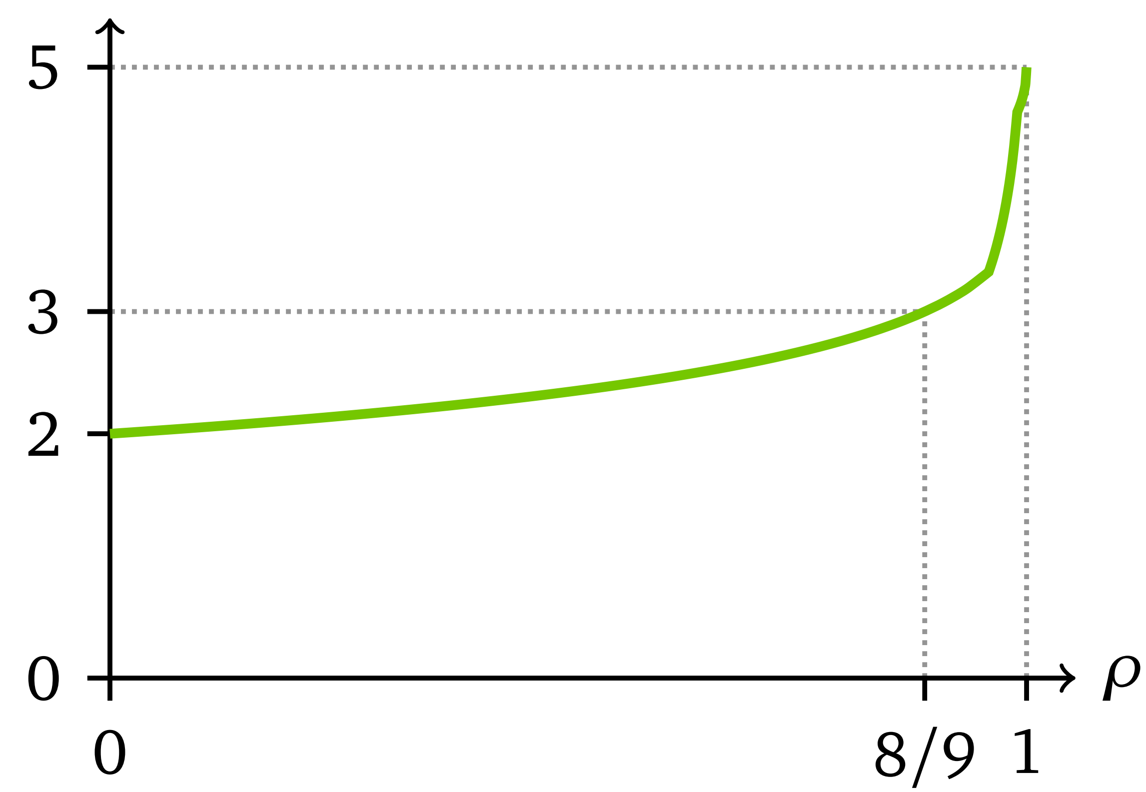
$$\frac{\mathbf{E}[T_{\text{M-SERPT}}]}{\mathbf{E}[T_{\text{Gittins}}]} \leq \begin{cases} \dfrac{4}{1+\sqrt{1+\rho}} & 0 \leq \rho < 0.96 \\ \dfrac{1}{\rho} \log \dfrac{1}{1-\rho} & 0.96 \leq \rho < 0.99 \\ 1 + \dfrac{4}{1+\sqrt{1+\rho}} & 0.99 \leq \rho < 1 \end{cases}$$

M-SERPT

first scheduling policy with

- simple rank function
- constant approximation ratio for $\mathbf{E}[T]$

$$\frac{\mathbf{E}[T_{\text{M-SERPT}}]}{\mathbf{E}[T_{\text{Gittins}}]} \le \begin{cases} \dfrac{4}{1+\sqrt{1+\rho}} & 0 \le \rho < 0.96 \\ \dfrac{1}{\rho}\log\dfrac{1}{1-\rho} & 0.96 \le \rho < 0.99 \\ 1+\dfrac{4}{1+\sqrt{1+\rho}} & 0.99 \le \rho < 1 \end{cases}$$

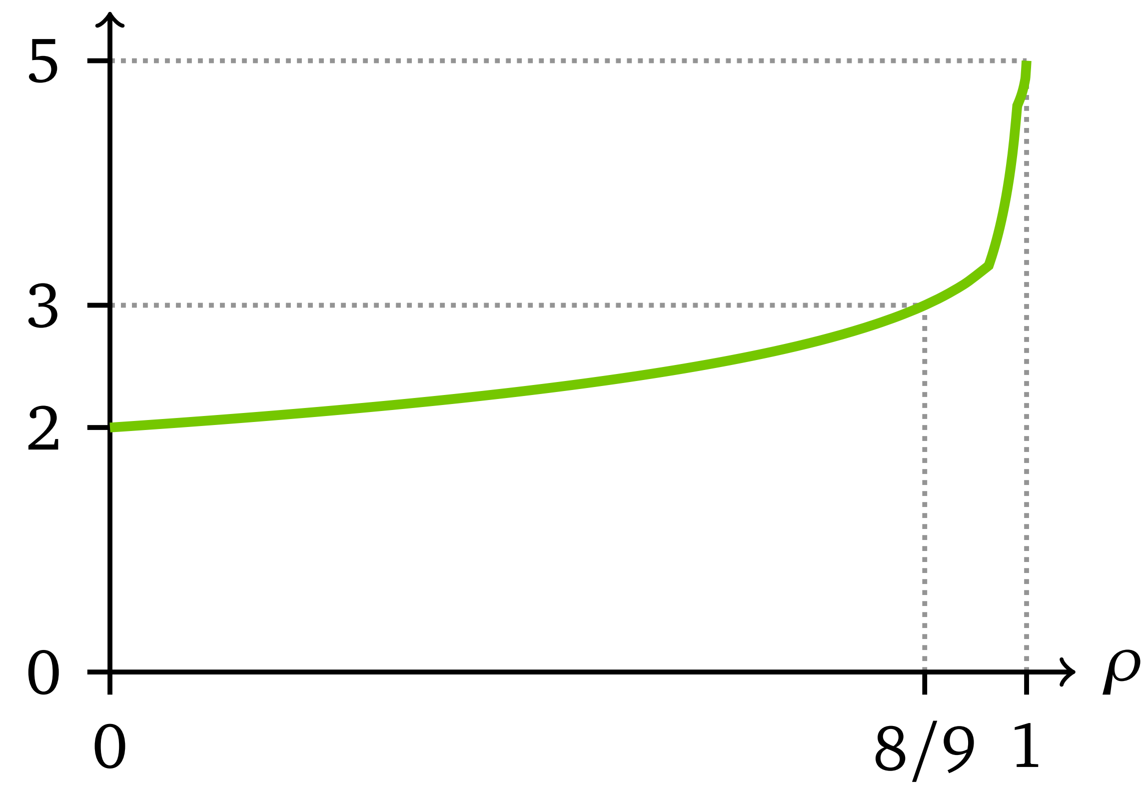


M-SERPT

first scheduling policy with

- simple rank function
- constant approximation ratio for $\mathbf{E}[T]$

$$\frac{\mathbf{E}[T_{\text{M-SERPT}}]}{\mathbf{E}[T_{\text{Gittins}}]} \le \begin{cases} \dfrac{4}{1+\sqrt{1+\rho}} & 0 \le \rho < 0.96 \\ \dfrac{1}{\rho}\log\dfrac{1}{1-\rho} & 0.96 \le \rho < 0.99 \\ 1+\dfrac{4}{1+\sqrt{1+\rho}} & 0.99 \le \rho < 1 \end{cases}$$



Get in touch: `zscully@cs.cmu.edu`

Bonus Slides

What Makes Gittins Hard?

$$r_{\text{Gittins}}(a) = \inf_{\Delta > 0} \frac{\mathbf{E}[\min\{X - a, \Delta\} \mid X > a]}{\mathbf{P}[X - a \leq \Delta \mid X > a]}$$

What Makes Gittins Hard?

$$r_{\text{Gittins}}(a) = \inf_{\Delta > 0} \frac{\mathbf{E}[\min\{X - a, \Delta\} \mid X > a]}{\mathbf{P}[X - a \leq \Delta \mid X > a]}$$

Need to optimize Δ at every age a

What Makes Gittins Hard?

$$r_{\text{Gittins}}(a) = \inf_{\Delta>0} \frac{\mathbf{E}[\min\{X-a,\Delta\} \mid X>a]}{\mathbf{P}[X-a\le\Delta \mid X>a]}$$

Need to optimize Δ at every age a

Takes $\Omega(n^2)$ time for n support points

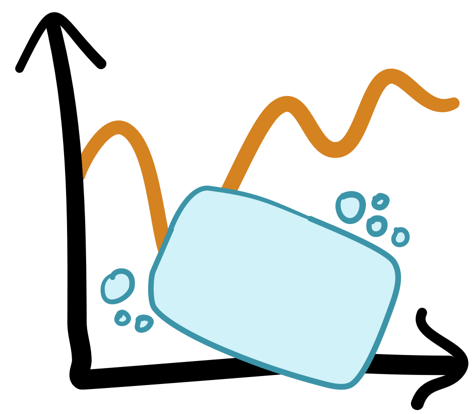
Proof Approach

Proof Approach

SOAP

Scully, Harchol-Balter,
Scheller-Wolf (2018)

Proof Approach



SOAP

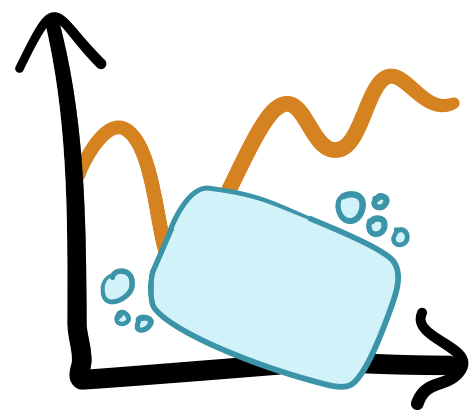
Scully, Harchol-Balter,
Scheller-Wolf (2018)

job size distribution X
rank function r

→

expression for $\mathbf{E}[T]$

Proof Approach



SOAP

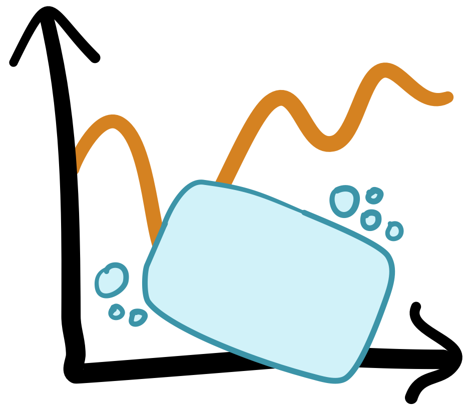
Scully, Harchol-Balter, Scheller-Wolf (2018)

job size distribution X
rank function r

→ expression for $\mathbf{E}[T]$

For **M-SERPT** and **Gittins**, can't write $\mathbf{E}[T]$ nicely

Proof Approach



SOAP

Scully, Harchol-Balter, Scheller-Wolf (2018)

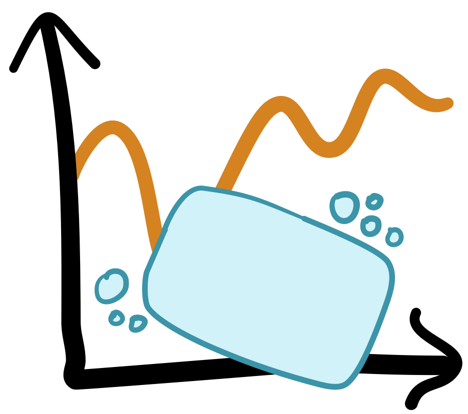
job size distribution X
rank function r

→ expression for $\mathbf{E}[T]$

r depends on X

For **M-SERPT** and **Gittins**, can't write $\mathbf{E}[T]$ nicely

Proof Approach



SOAP

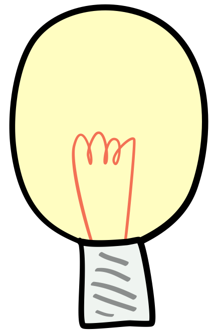
Scully, Harchol-Balter, Scheller-Wolf (2018)

job size distribution X
rank function r

→ expression for $\mathbf{E}[T]$

r depends on X

For **M-SERPT** and **Gittins**, can't write $\mathbf{E}[T]$ nicely



Compare $\mathbf{E}[T]$ expressions to each other