The Hiring Problem:

Going Beyond Secretaries

Sergei Vassilvitskii (Yahoo!)

Andrei Broder (Yahoo!)
Adam Kirsch (Harvard)
Ravi Kumar (Yahoo!)
Michael Mitzenmacher (Harvard)
Eli Upfal (Brown)

The Secretary Problem

Interview n candidates for a position one at a time. After each interview decide if the candidate is the best.

Goal: maximize the probability of choosing the best candidate.

The Secretary Problem

Interview n candidates for a position one at a time. After each interview decide if the candidate is the best.

Goal: maximize the probability of choosing the best candidate.

The Secretary Problem

Interview n candidates for a position one at a time. After each interview decide if the candidate is the best.

Goal: maximize the probability of choosing the best candidate.

The Secretary Problem

Interview n candidates for a position one at a time. After each interview decide if the candidate is the best.

Goal: maximize the probability of choosing the best candidate.

The Secretary Problem

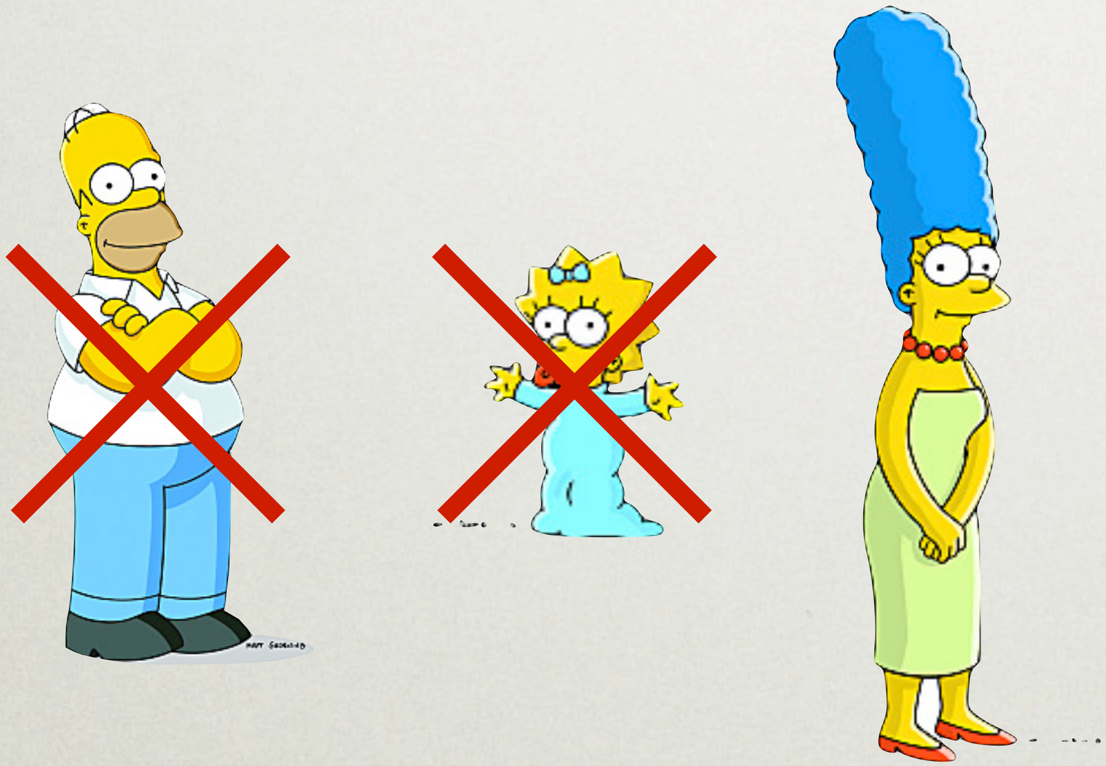
Interview n candidates for a position one at a time. After each interview decide if the candidate is the best.

Goal: maximize the probability of choosing the best candidate.

The Secretary Problem

Interview n candidates for a position one at a time. After each interview decide if the candidate is the best.

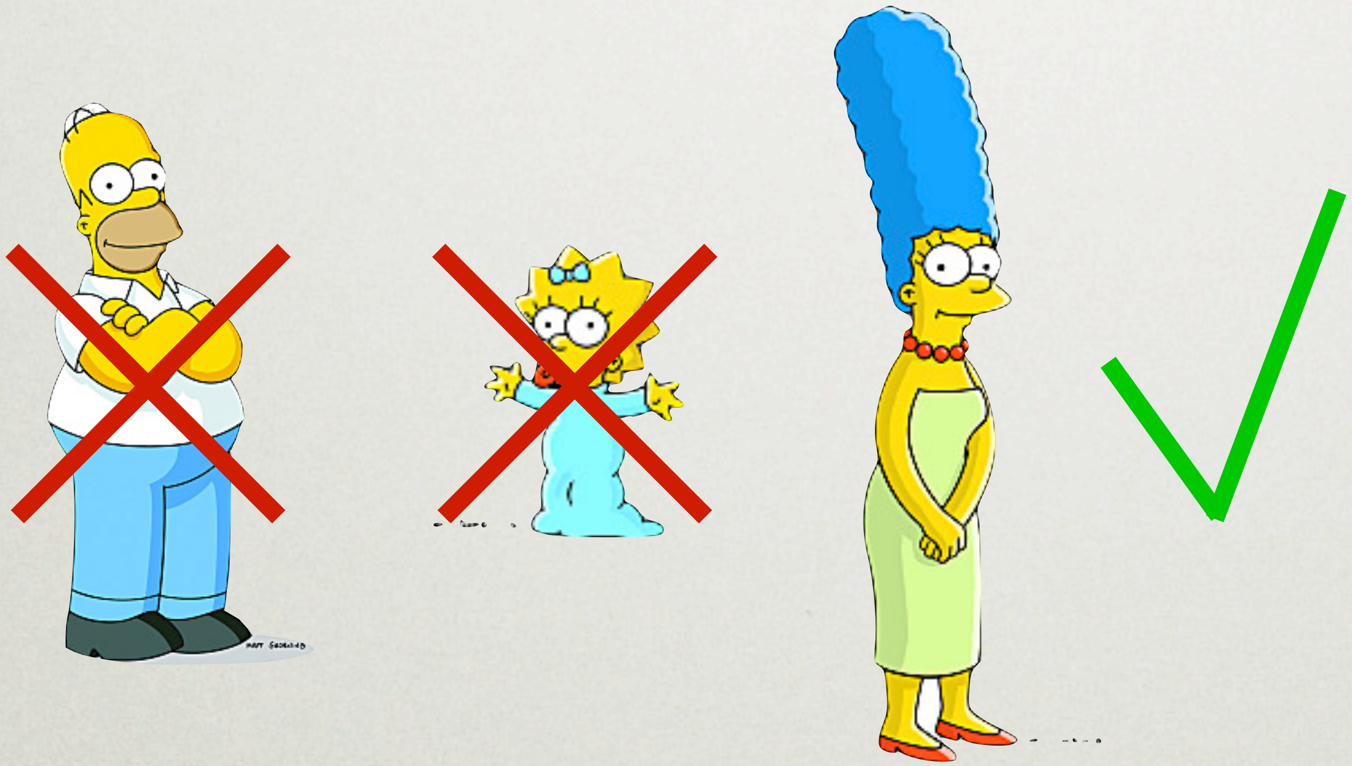
Goal: maximize the probability of choosing the best candidate.



The Secretary Problem

Interview n candidates for a position one at a time. After each interview decide if the candidate is the best.

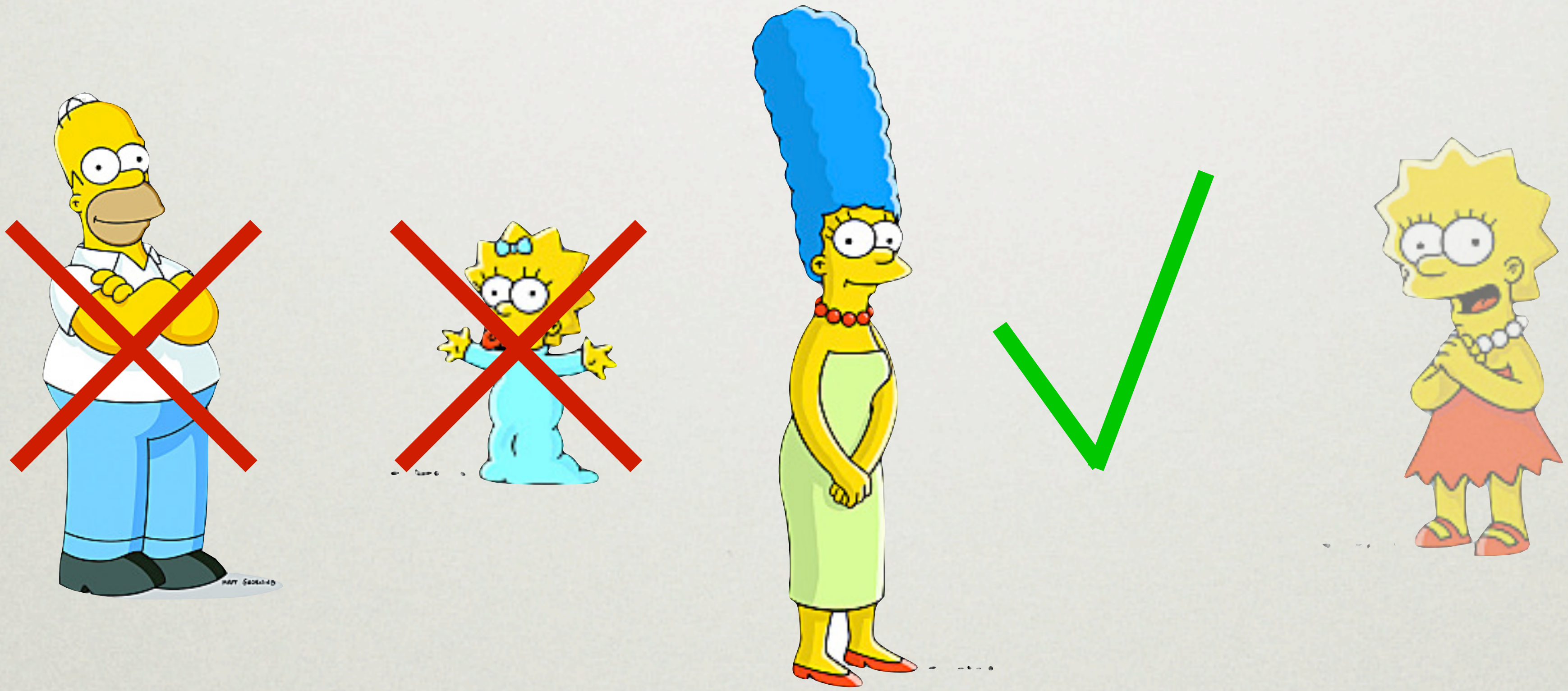
Goal: maximize the probability of choosing the best candidate.



The Secretary Problem

Interview n candidates for a position one at a time. After each interview decide if the candidate is the best.

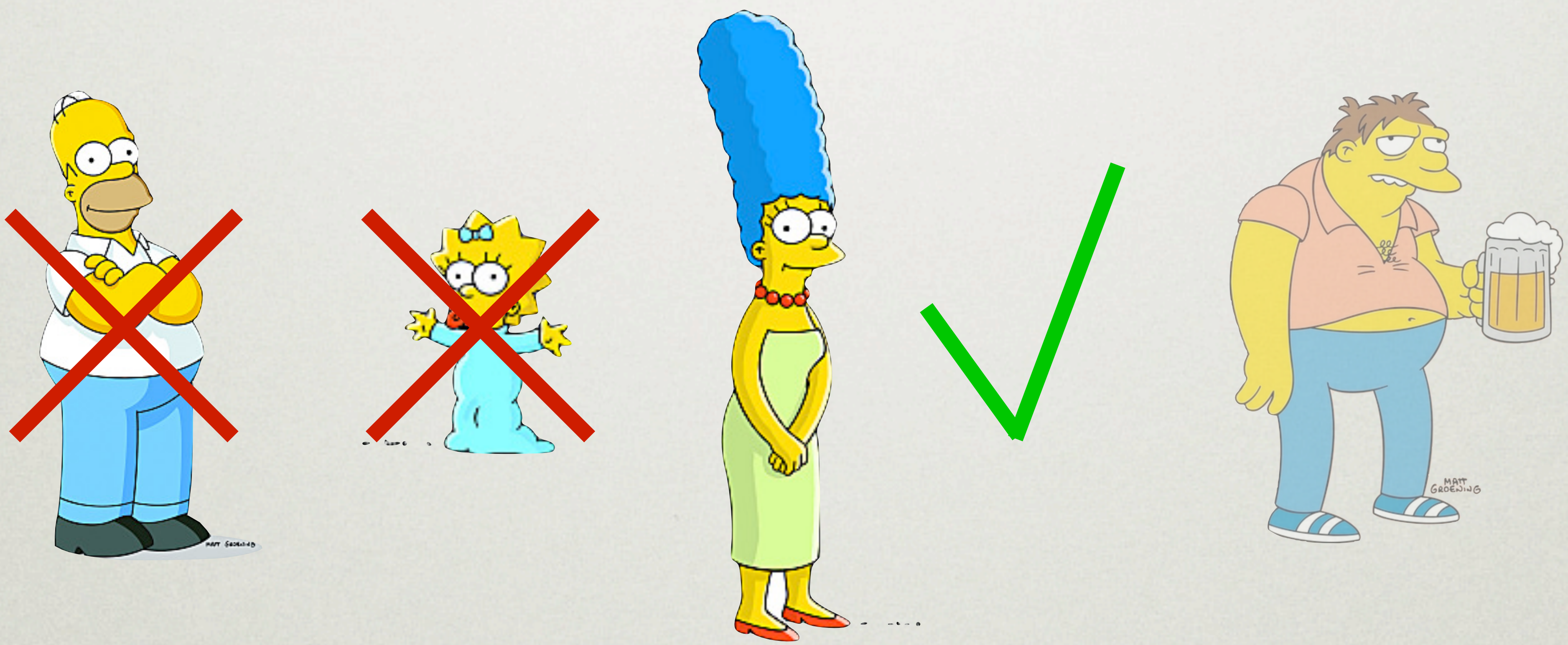
Goal: maximize the probability of choosing the best candidate.



The Secretary Problem

Interview n candidates for a position one at a time. After each interview decide if the candidate is the best.

Goal: maximize the probability of choosing the best candidate.



The Secretary Problem

Interview n candidates for a position one at a time. After each interview decide if the candidate is the best.

Goal: maximize the probability of choosing the best candidate.

This is not about hiring secretaries, but about decision making under uncertainty.

The Hiring Problem

The Hiring Problem

A startup is growing and is hiring many employees:

- Want to hire good employees
- Can’t wait for the perfect candidate

The Hiring Problem

A startup is growing and is hiring many employees:

- Want to hire good employees
- Can't wait for the perfect candidate

Many potential objectives.

- Explore the tradeoff between number of interviews & the average quality.

The Hiring model

Candidates arrive one at a time.

Assume all have iid uniform(0,1) quality scores - For applicant i denote it by i_q.

(Can deal with other distributions, not this talk)

The Hiring model

Candidates arrive one at a time.

Assume all have iid uniform(0,1) quality scores - For applicant i denote it by i_q.

- (Can deal with other distributions, not this talk)

During the interview:

- Observe i_q
- Decide whether to hire or reject

Strategies

Strategies

Hire above a threshold.

Strategies

Hire above a threshold.

Hire above the minimum or maximum.

Strategies

Hire above a threshold.

Hire above the minimum or maximum.

Lake Wobegon Strategies:

> "Lake Wobegon: where all the women are strong, all the men are good looking, and all the children are above average"

Strategies

Hire above a threshold.

Hire above the minimum or maximum.

Lake Wobegon Strategies:

- Hire above the average (mean or median)

Strategies

Hire above a threshold.

Hire above the minimum or maximum.

Lake Wobegon Strategies:

- Hire above the average

Side note: [Google Research Blog - March '06]:

- "... only hire candidates who are above the mean of the current employees..."

Threshold Hiring

Set a threshold t, hire if $i_q \geq t$.

Threshold Hiring

Set a threshold t, hire if $i_q \geq t$.

Threshold Hiring

Set a threshold t, hire if $i_q \geq t$.

Threshold Hiring

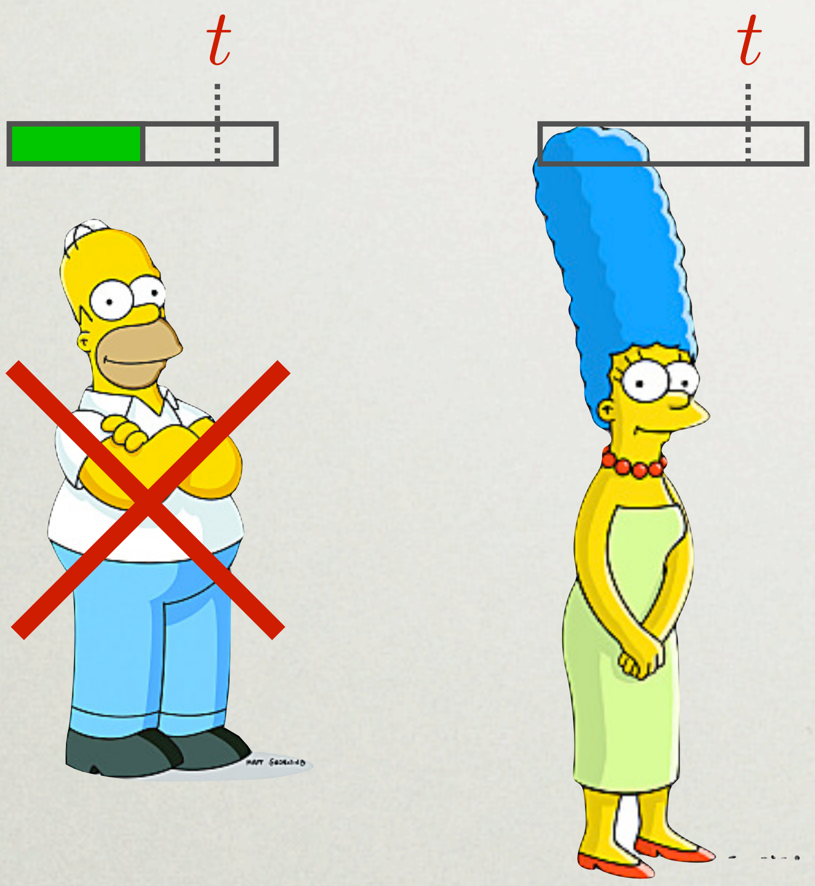
Set a threshold t, hire if $i_q \geq t$.

Threshold Hiring

Set a threshold t, hire if $i_q \geq t$.

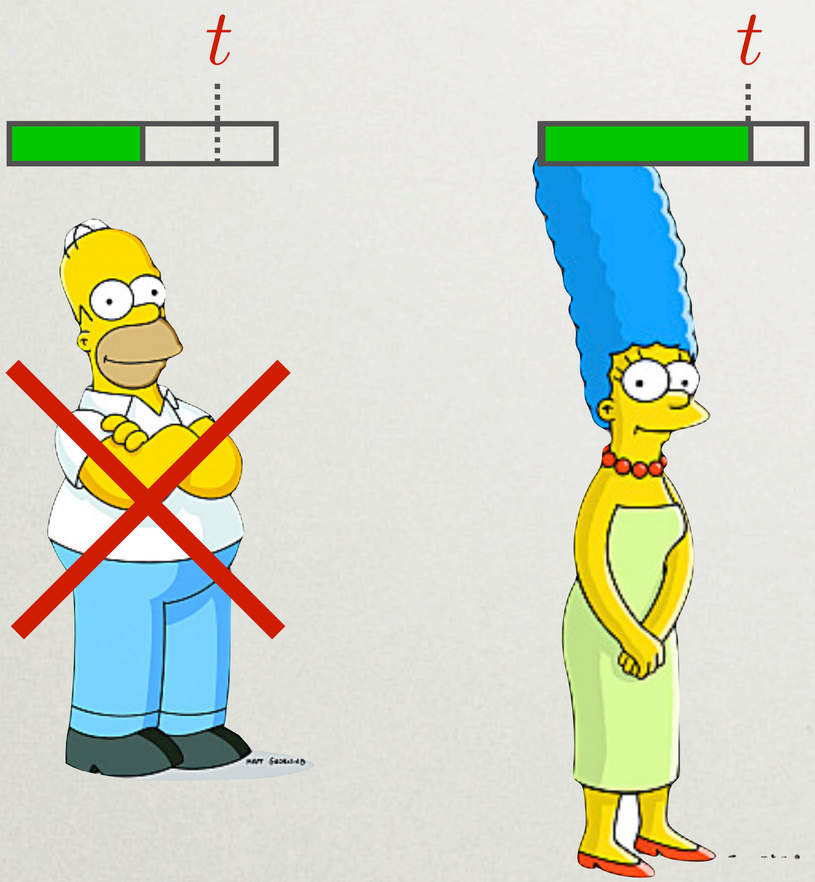
Threshold Hiring

Set a threshold t, hire if $i_q \geq t$.



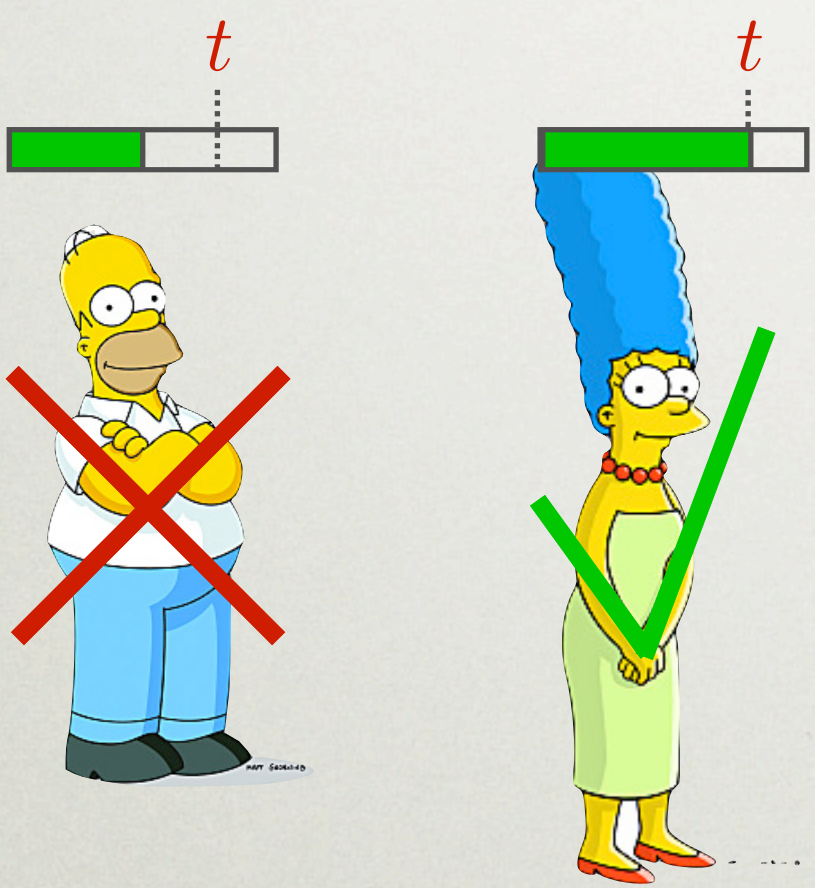
Threshold Hiring

Set a threshold t, hire if $i_q \geq t$.



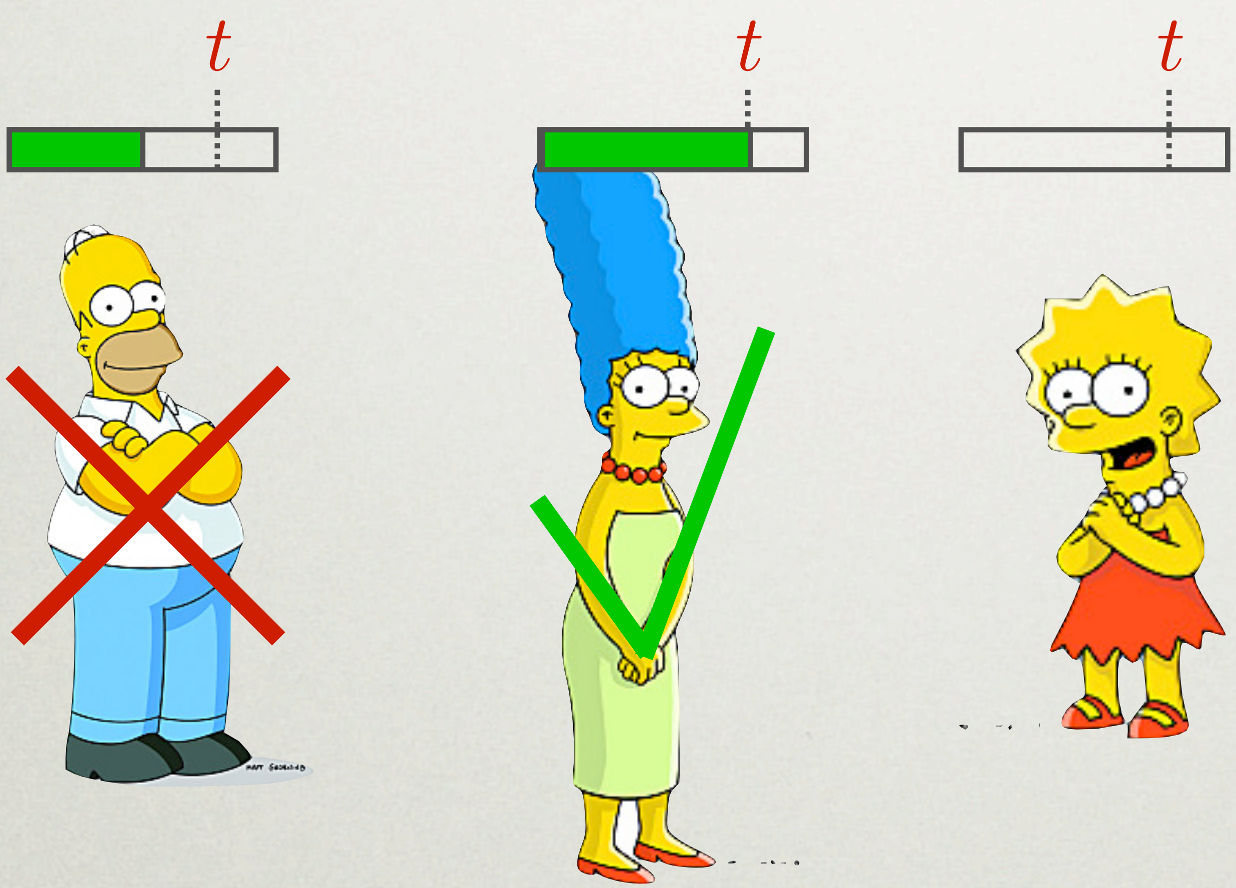
Threshold Hiring

Set a threshold t, hire if $i_q \geq t$.



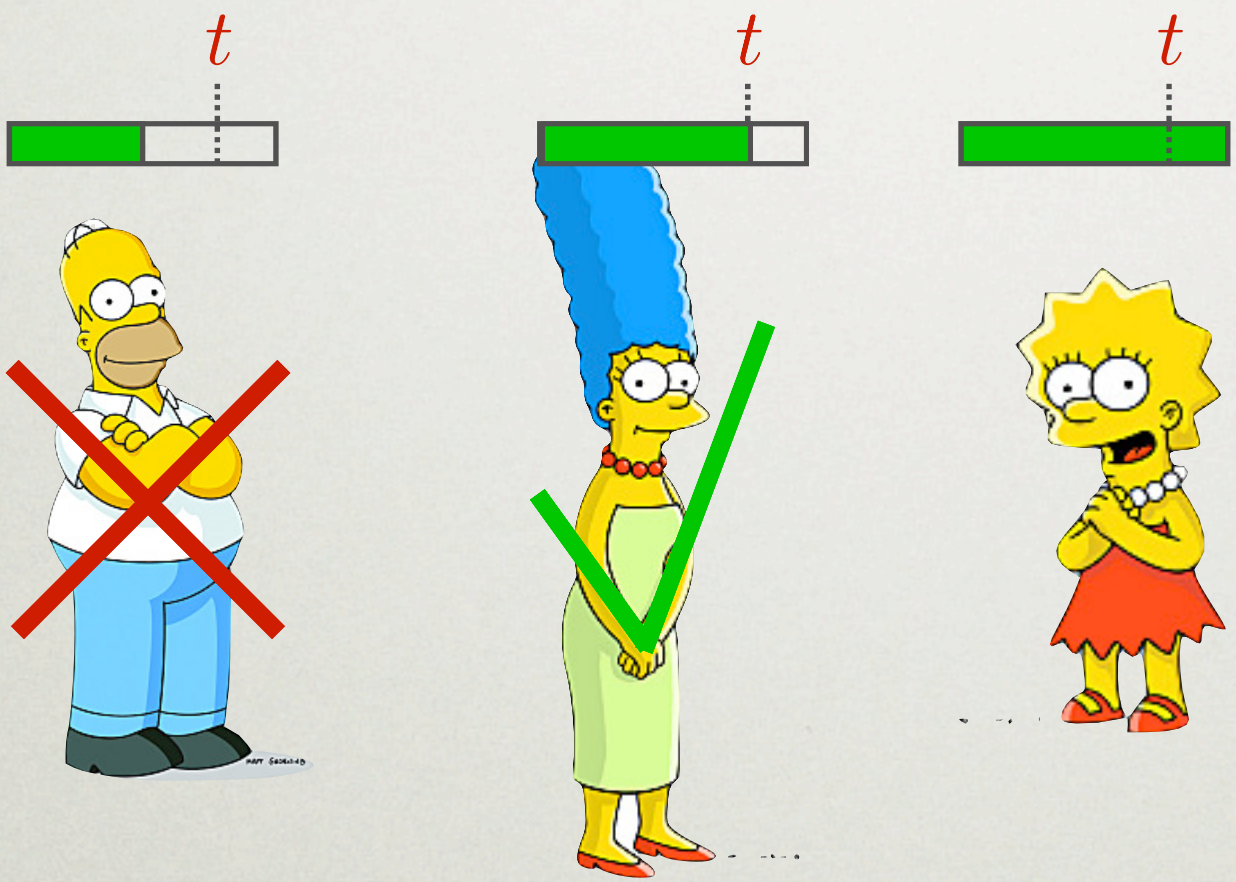
Threshold Hiring

Set a threshold t, hire if $i_q \geq t$.



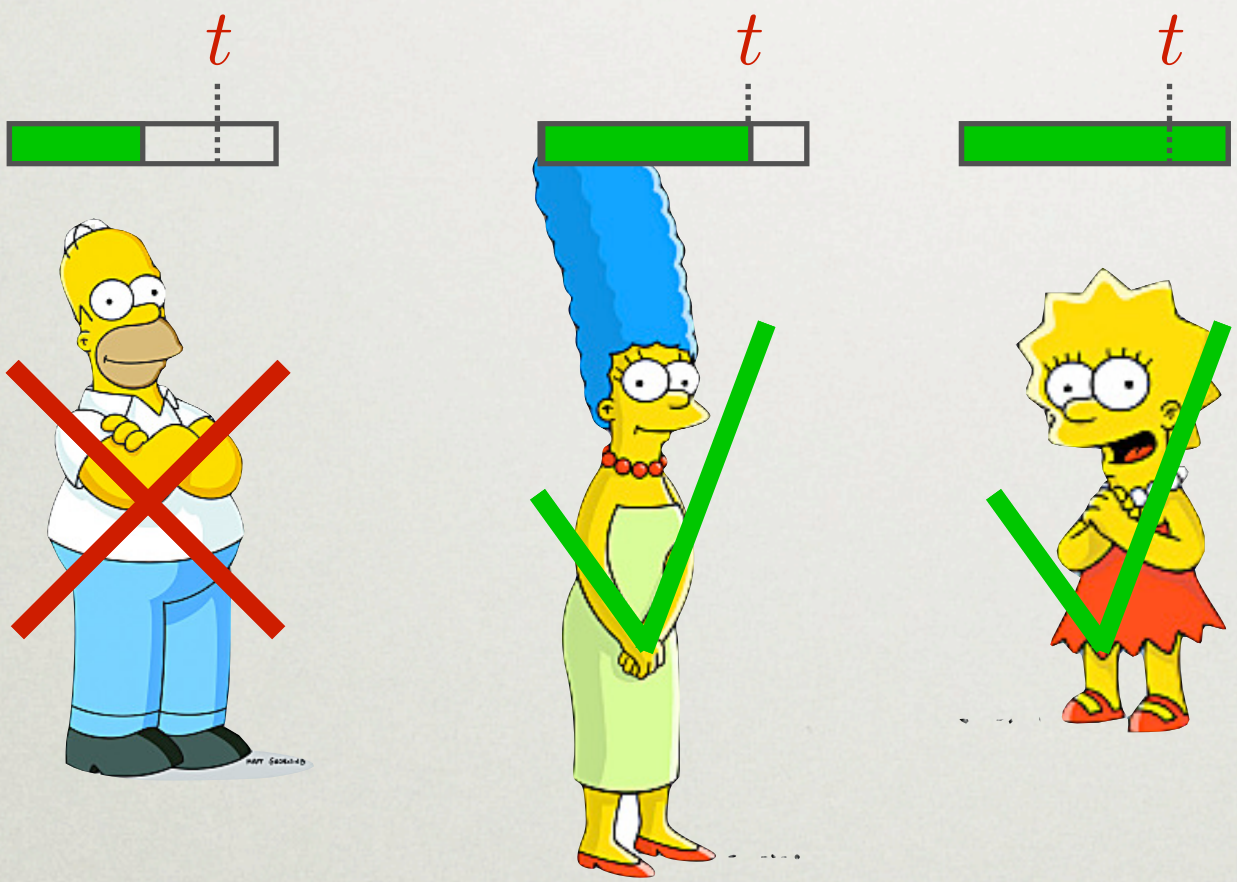
Threshold Hiring

Set a threshold t, hire if $i_q \geq t$.



Threshold Hiring

Set a threshold t, hire if $i_q \geq t$.



Threshold Analysis

Set a threshold t , hire if $i_q \geq t$.

Easy to see that average quality approaches $\frac{1+t}{2}$

Hiring rate $\frac{1}{1-t}$.

Threshold Analysis

Set a threshold t , hire if $i_q \geq t$.

Easy to see that average quality approaches $\frac{1+t}{2}$

Hiring rate $\frac{1}{1-t}$.

Quality stagnates and does not increase with time.

Maximum Hiring

Hire only if better than everyone already hired.

Maximum Hiring

Hire only if better than everyone already hired.

Maximum Hiring

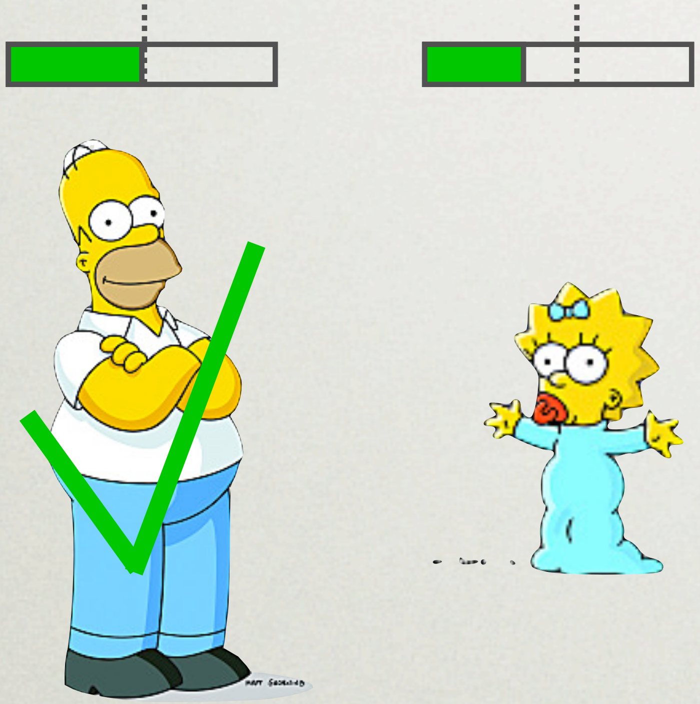
Hire only if better than everyone already hired.

Maximum Hiring

Hire only if better than everyone already hired.

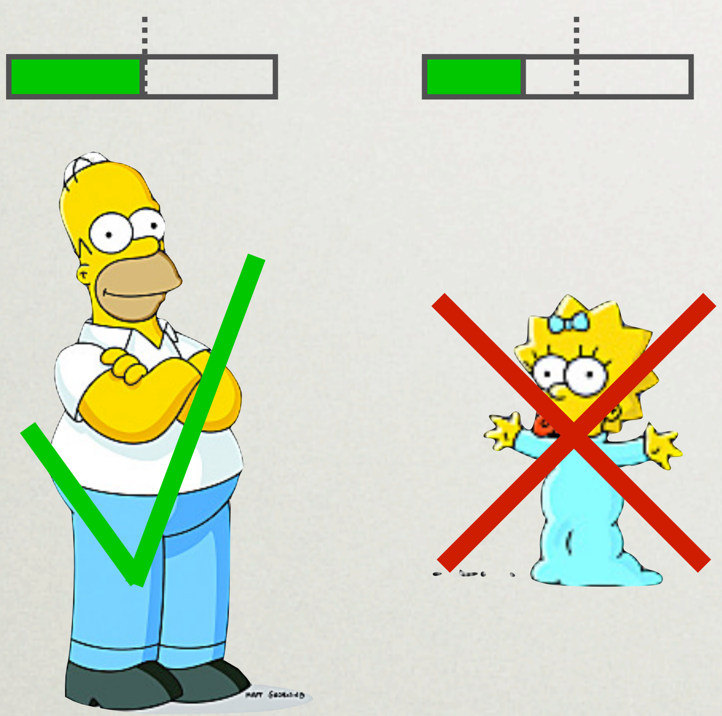
Maximum Hiring

Hire only if better than everyone already hired.



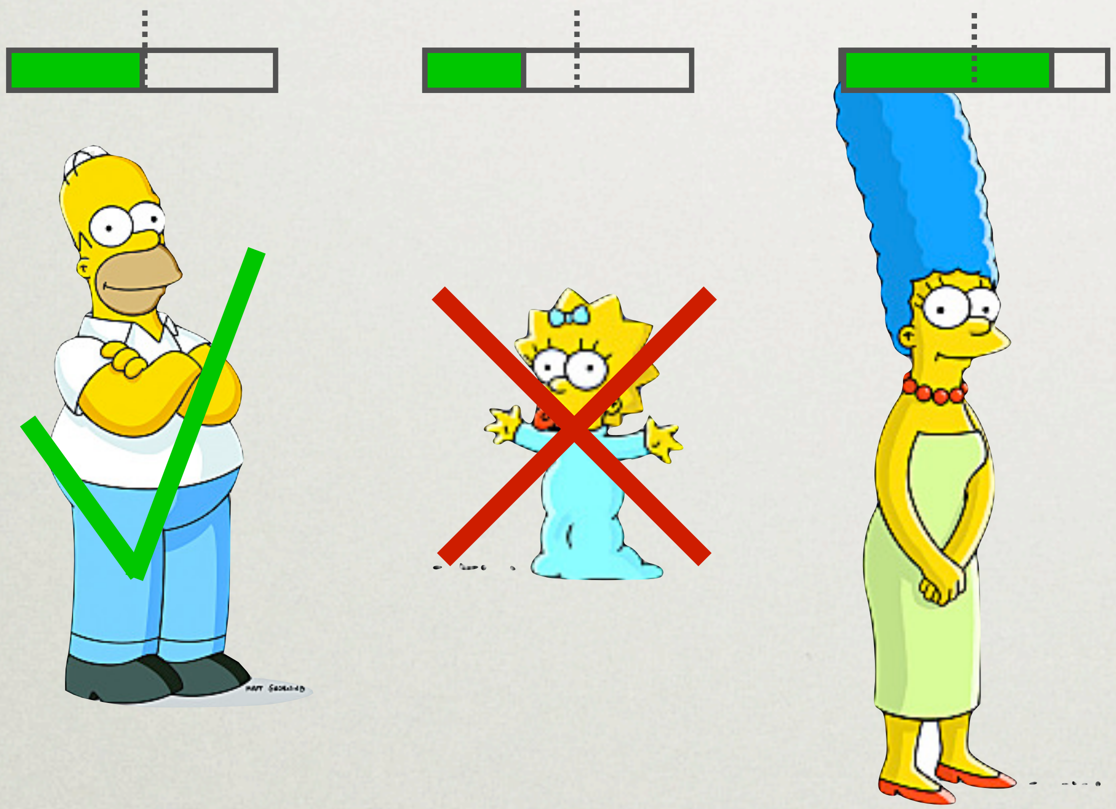
Maximum Hiring

Hire only if better than everyone already hired.



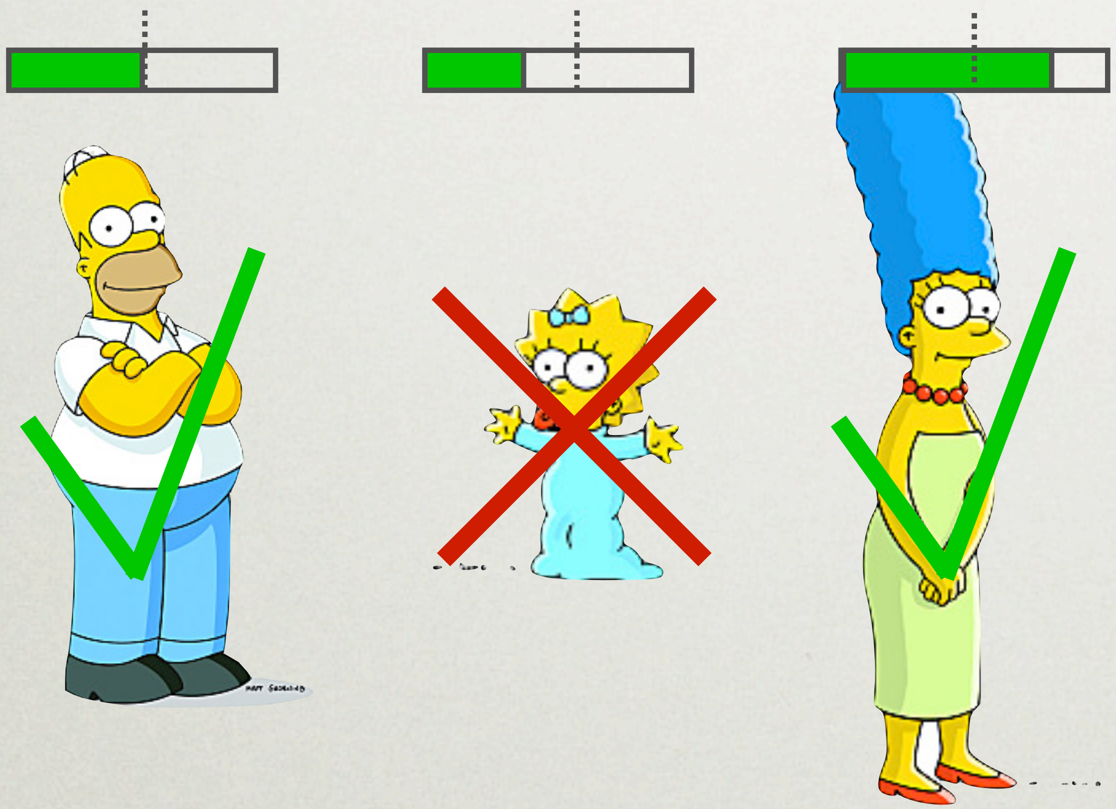
Maximum Hiring

Hire only if better than everyone already hired.



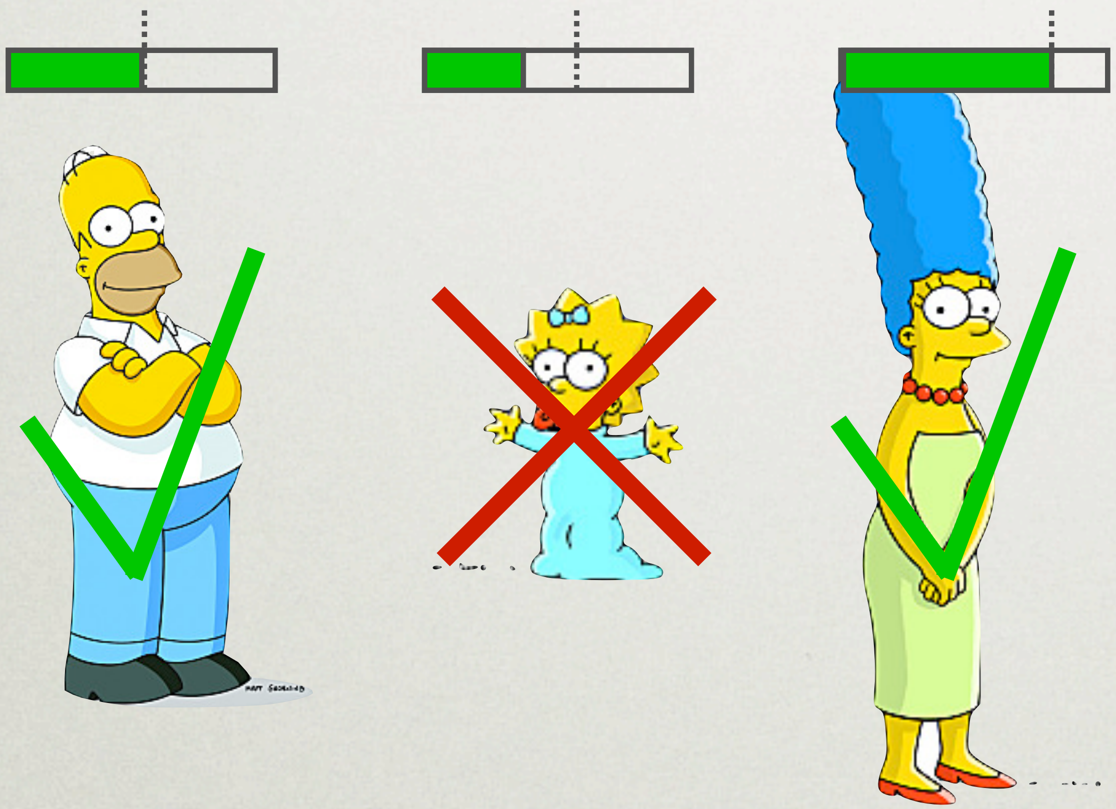
Maximum Hiring

Hire only if better than everyone already hired.



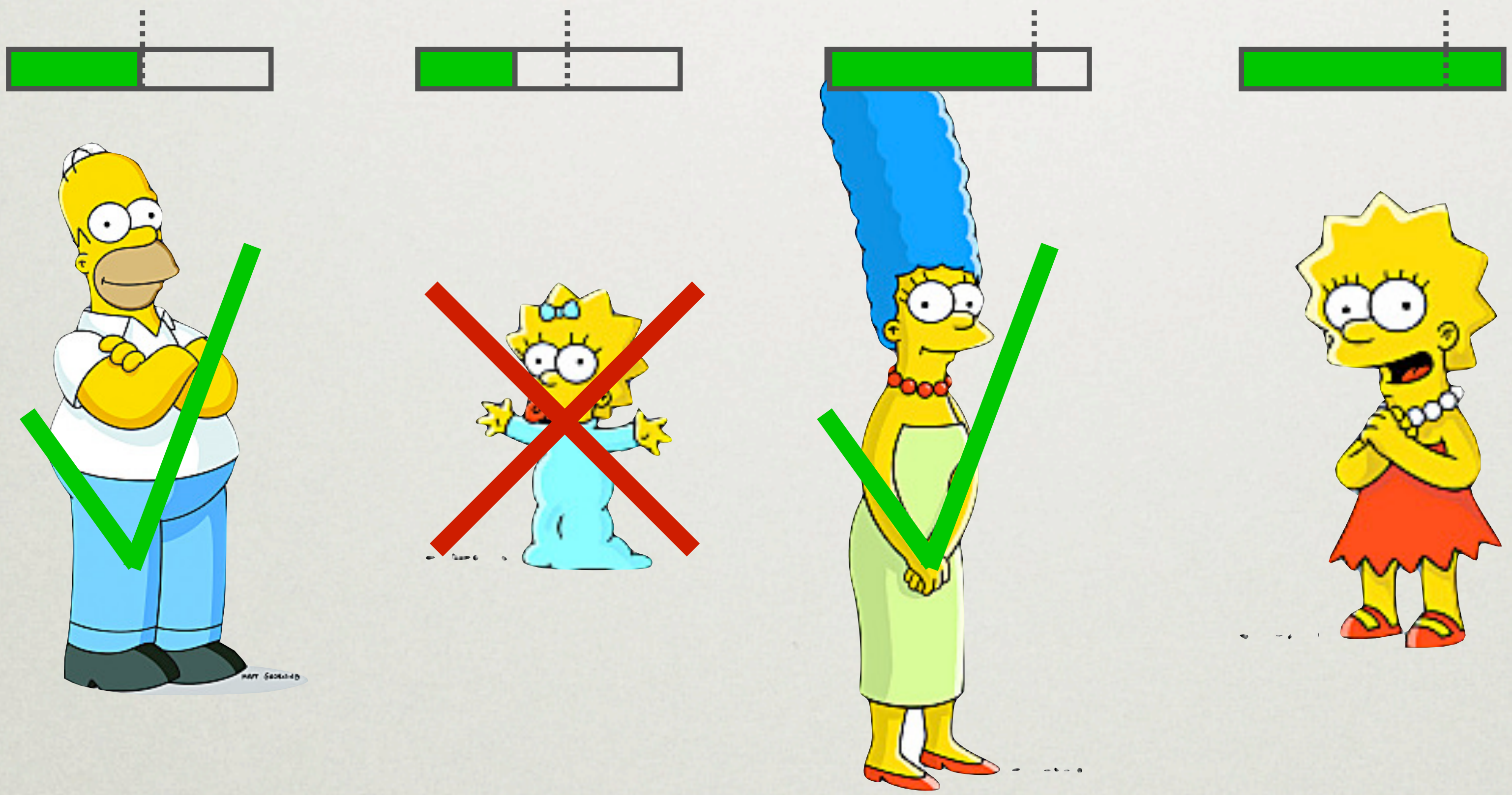
Maximum Hiring

Hire only if better than everyone already hired.



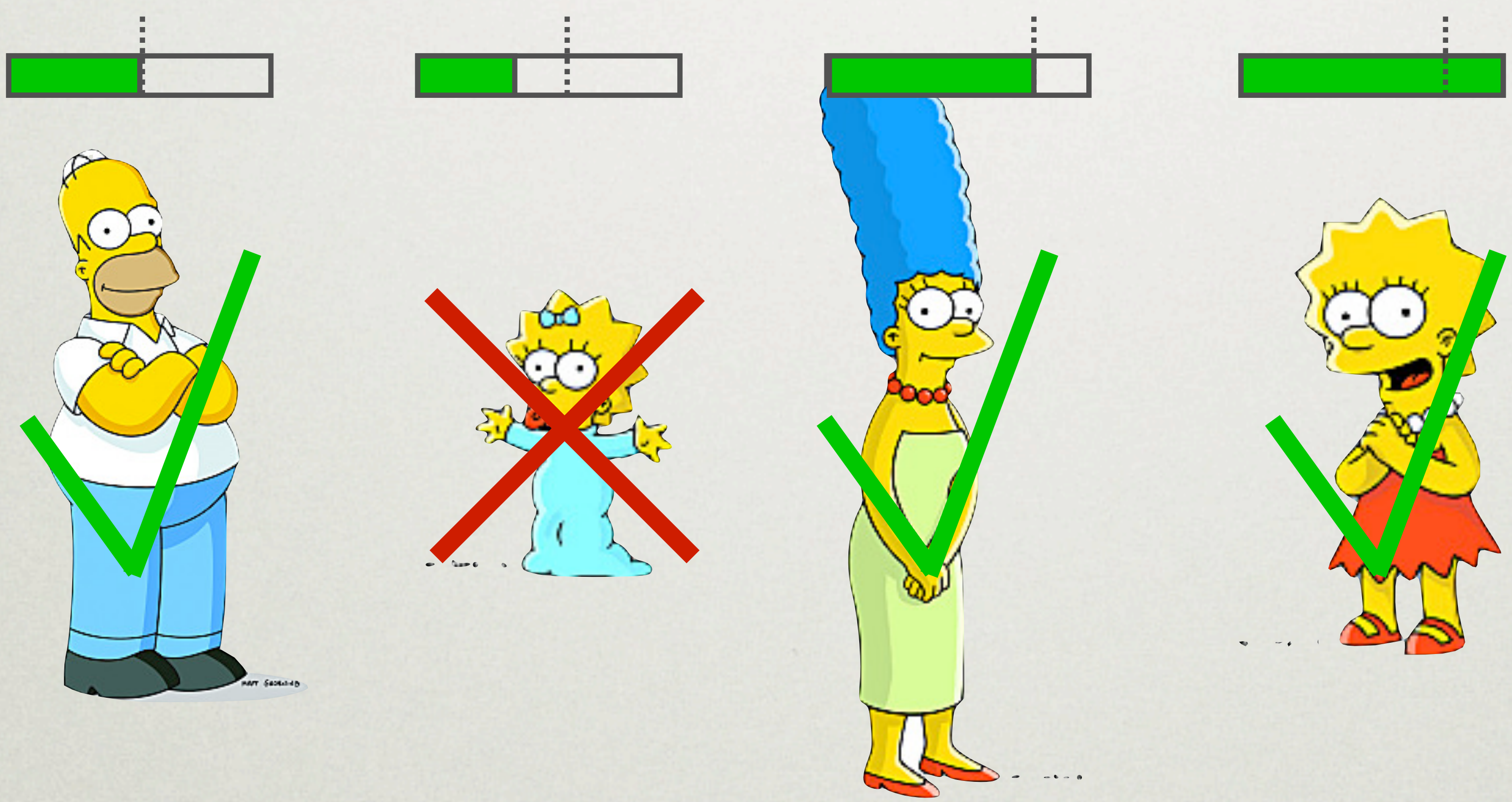
Maximum Hiring

Hire only if better than everyone already hired.



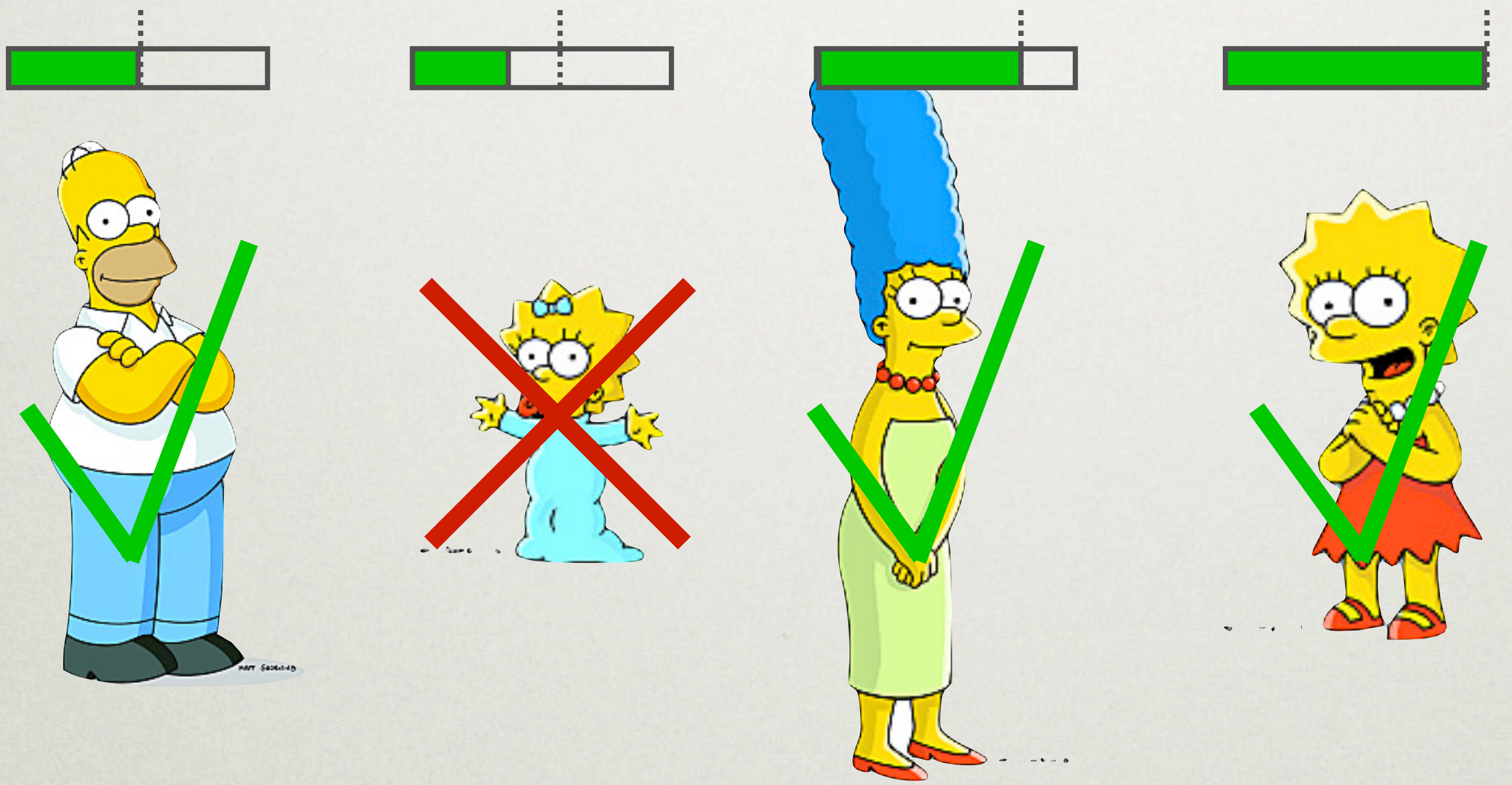
Maximum Hiring

Hire only if better than everyone already hired.



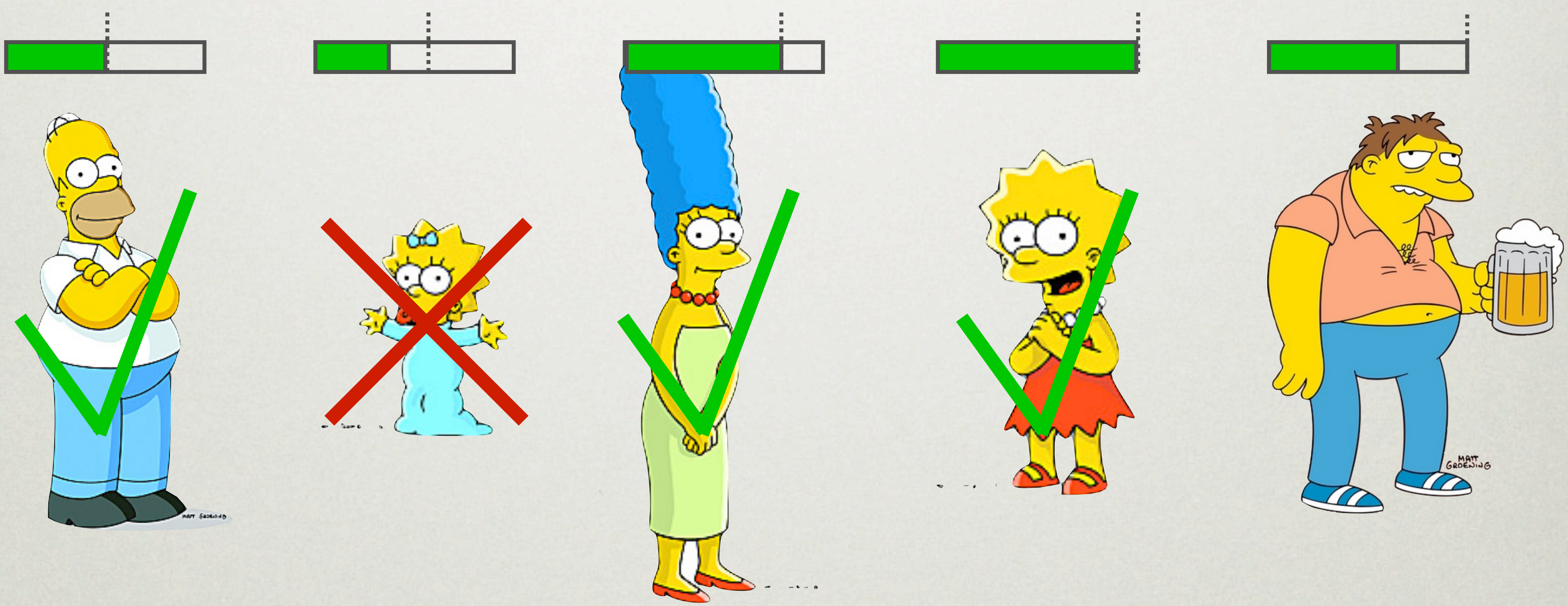
Maximum Hiring

Hire only if better than everyone already hired.



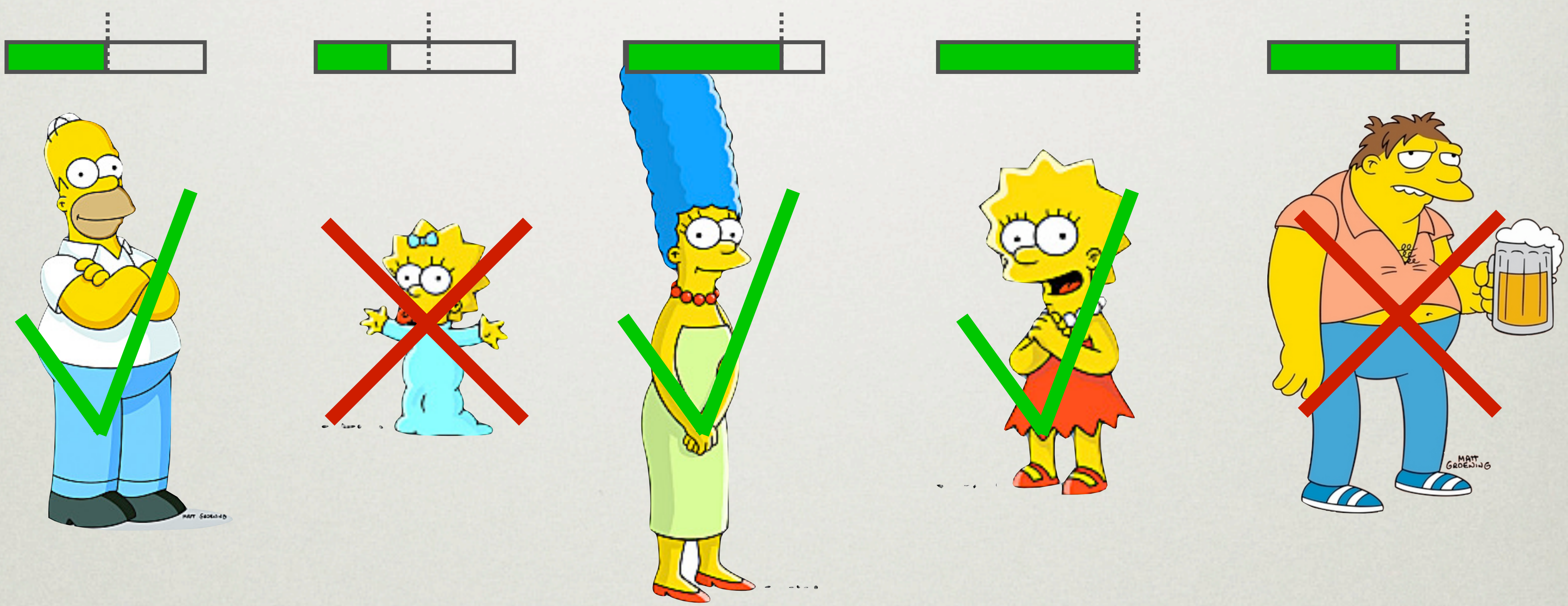
Maximum Hiring

Hire only if better than everyone already hired.



Maximum Hiring

Hire only if better than everyone already hired.



Maximum Analysis

Start with employee of quality q

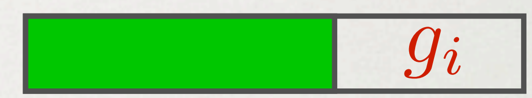
Let h_i be the i-th candidate hired

Maximum Analysis

Start with employee of quality q

Let h_i be the i-th candidate hired

Focus on the gap: $g_i = 1 - (h_i)_q$

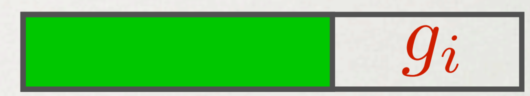


Maximum Analysis

Start with employee of quality q

Let h_i be the i-th candidate hired

Focus on the gap: $g_i = 1 - (h_i)_q$



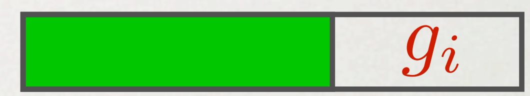
Conditioned on g_{n-1}:

Maximum Analysis

Start with employee of quality q

Let h_i be the i-th candidate hired

Focus on the gap: $g_i = 1 - (h_i)_q$



Conditioned on g_{n-1}:

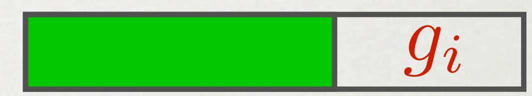
$$g_n \sim Unif(0, g_{n-1})$$

Maximum Analysis

Start with employee of quality q

Let h_i be the i-th candidate hired

Focus on the gap: $g_i = 1 - (h_i)_q$



Conditioned on g_{n-1}:

$$g_n \sim Unif(0, g_{n-1})$$

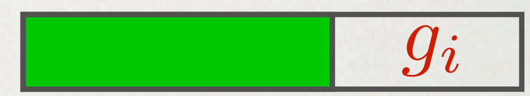
$$E[g_n | g_{n-1}] = \frac{g_{n-1}}{2}$$

Maximum Analysis

Start with employee of quality q

Let h_i be the i-th candidate hired

Focus on the gap: $g_i = 1 - (h_i)_q$



Conditioned on g_{n-1}:

$$g_n \sim Unif(0, g_{n-1})$$

$$E[g_n | g_{n-1}] = \frac{g_{n-1}}{2}$$

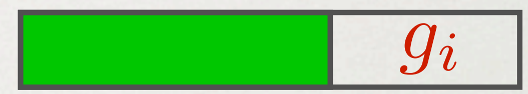
$$E[g_n] = \frac{1-q}{2^n}$$

Maximum Analysis

Start with employee of quality q

Let h_i be the i-th candidate hired

Focus on the gap: $g_i = 1 - (h_i)_q$



Conditioned on g_{n-1}:

$$g_n \sim Unif(0, g_{n-1})$$

$$E[g_n | g_{n-1}] = \frac{g_{n-1}}{2}$$

$$E[g_n] = \frac{1-q}{2^n}$$

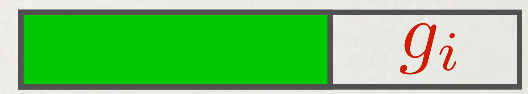
Very high quality!

Maximum Analysis

Start with employee of quality q

Let h_i be the i-th candidate hired

Focus on the gap: $g_i = 1 - (h_i)_q$



Conditioned on g_{n-1}:

$$g_n \sim Unif(0, g_{n-1})$$

$$E[g_n | g_{n-1}] = \frac{g_{n-1}}{2}$$

$$E[g_n] = \frac{1-q}{2^n}$$

Extremely slow hiring!

Lake Wobegon Strategies

Lake Wobegon Strategies

Above the mean:

Average quality after n hires: $1 - \frac{1}{\sqrt{n}}$

Lake Wobegon Strategies

Above the mean:

Average quality after n hires: $1 - \dfrac{1}{\sqrt{n}}$

Above the median:

Median quality after n hires: $1 - \dfrac{1}{n}$

Lake Wobegon Strategies

Above the mean:

Average quality after n hires: $1 - \frac{1}{\sqrt{n}}$

Above the median:

Median quality after n hires: $1 - \frac{1}{n}$

Surprising:

Tight concentration is not possible

Hiring above mean converges to a log-normal distribution

Hiring Above Mean

Start with employee of quality q

Let h_i be the i-th candidate hired

Focus on the gap: $g_i = 1 - (h_i)_q$

Hiring Above Mean

Start with employee of quality q

Let h_i be the i-th candidate hired

Focus on the gap: $g_i = 1 - (h_i)_q$

Conditioned on g_n : $(i_{n+1})_q \sim Unif(1 - g_n, 1)$

Hiring Above Mean

Start with employee of quality q

Let h_i be the i-th candidate hired

Focus on the gap: $g_i = 1 - (h_i)_q$

Conditioned on g_n : $(i_{n+1})_q \sim Unif(1 - g_n, 1)$

Therefore: $$g_{n+1} \sim \frac{n+1}{n+2} g_n + \frac{1}{n+2} Unif(0, g_n)$$

Hiring Above Mean

Start with employee of quality q

Let h_i be the i-th candidate hired

Focus on the gap: $g_i = 1 - (h_i)_q$

Conditioned on g_n : $(i_{n+1})_q \sim Unif(1 - g_n, 1)$

Therefore: $$g_{n+1} \sim \frac{n+1}{n+2} g_n + \frac{1}{n+2} Unif(0, g_n)$$

$$= g_n \left(1 - \frac{Unif(0,1)}{n+2}\right)$$

Hiring Above Mean

$$g_{n+1} = g_n \left(1 - \frac{Unif(0,1)}{n+2}\right)$$

Hiring Above Mean

$$g_{n+1} = g_n \left(1 - \frac{Unif(0,1)}{n+2}\right)$$

Expand: $$g_{n+t} = g_n \prod_{i=1}^{t} \left(1 - \frac{Unif(0,1)}{n+i+1}\right)$$

Hiring Above Mean

$$g_{n+1} = g_n \left(1 - \frac{Unif(0,1)}{n+2}\right)$$

Expand: $$g_{n+t} = g_n \prod_{i=1}^{t} \left(1 - \frac{Unif(0,1)}{n+i+1}\right)$$

Therefore:

$$g_n \sim (1-q) \prod_{i=1}^{n} \left(1 - \frac{Unif(0,1)}{i+1}\right)$$

Hiring Above Mean

$$g_{n+1} = g_n \left(1 - \frac{Unif(0,1)}{n+2}\right)$$

Expand: $$g_{n+t} = g_n \prod_{i=1}^{t} \left(1 - \frac{Unif(0,1)}{n+i+1}\right)$$

Therefore:

$$g_n \sim (1-q) \prod_{i=1}^{n} \left(1 - \frac{Unif(0,1)}{i+1}\right)$$

$$E[g_n] = (1-q) \prod_{i=1}^{n} \left(1 - \frac{1}{2(i+1)}\right) = \Theta(\frac{1}{\sqrt{n}})$$

Hiring Above Mean

Conclusion:

Average quality after n hires: $1 - \Theta(\frac{1}{\sqrt{n}})$

Time to hire n employees: $\Theta(n^{3/2})$

Very weak concentration results.

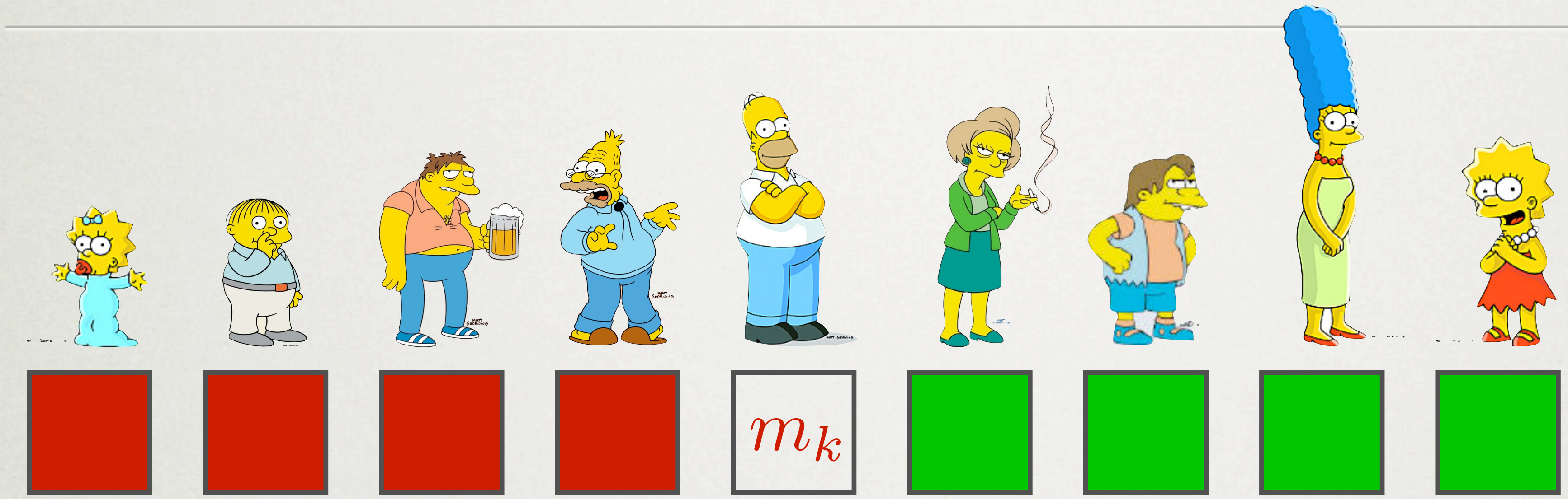
Hiring Above Median

Start with employee of quality q

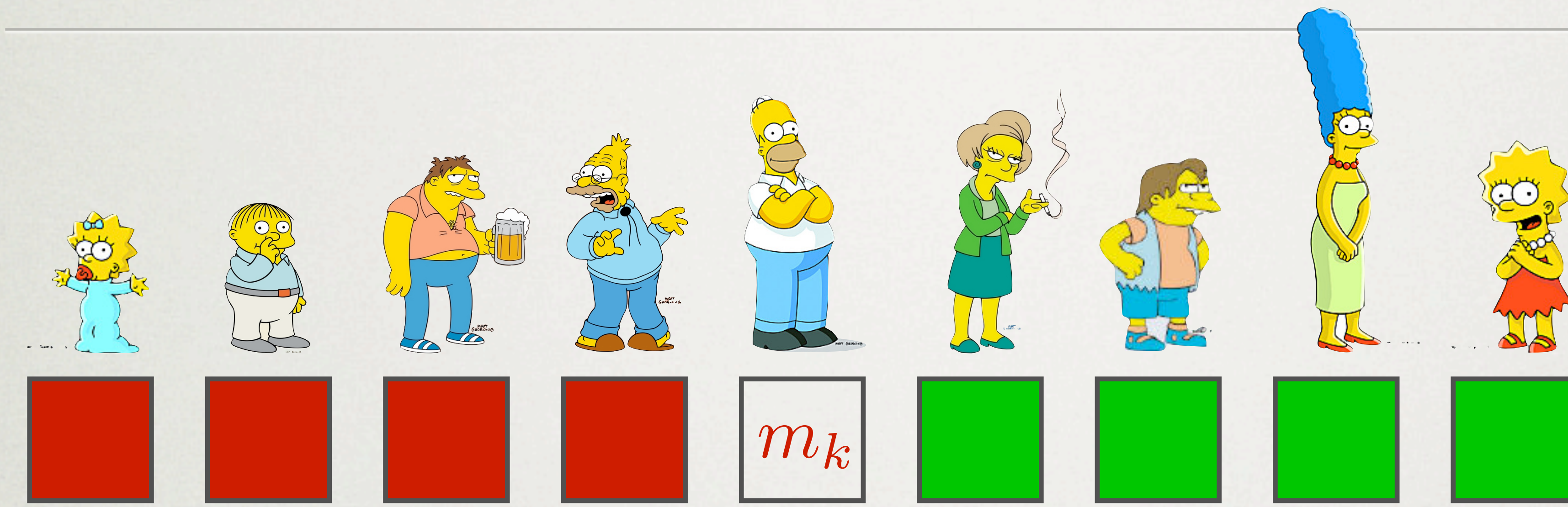
When we have $2k+1$ employees.

- Compute median m_k of the scores
- Hire next 2 applicants with scores above m_k

Hiring Above Median

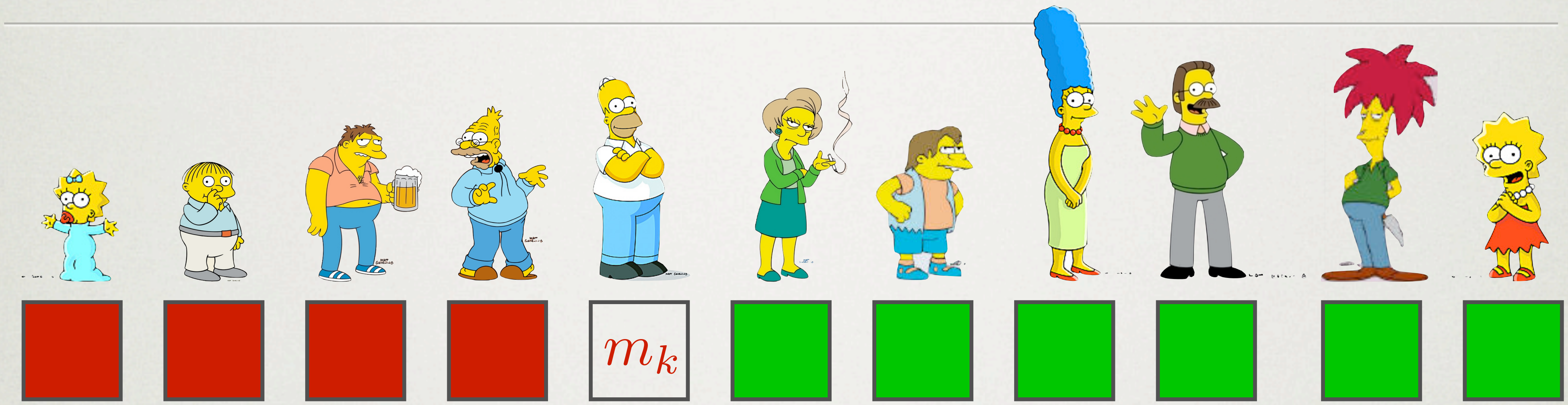


Hiring Above Median

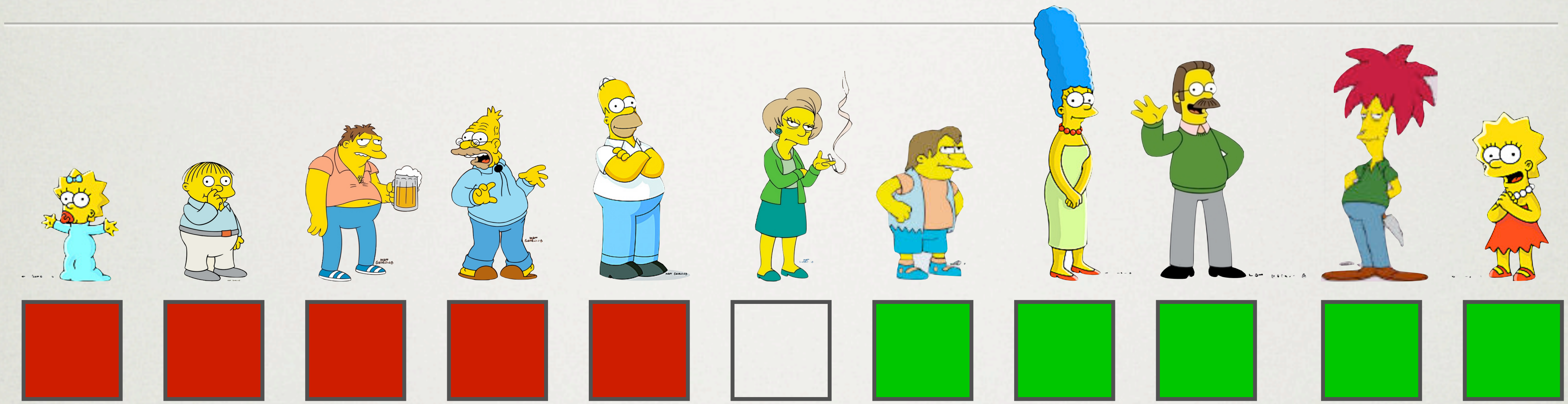


New hires:  with quality above m_k

Hiring Above Median

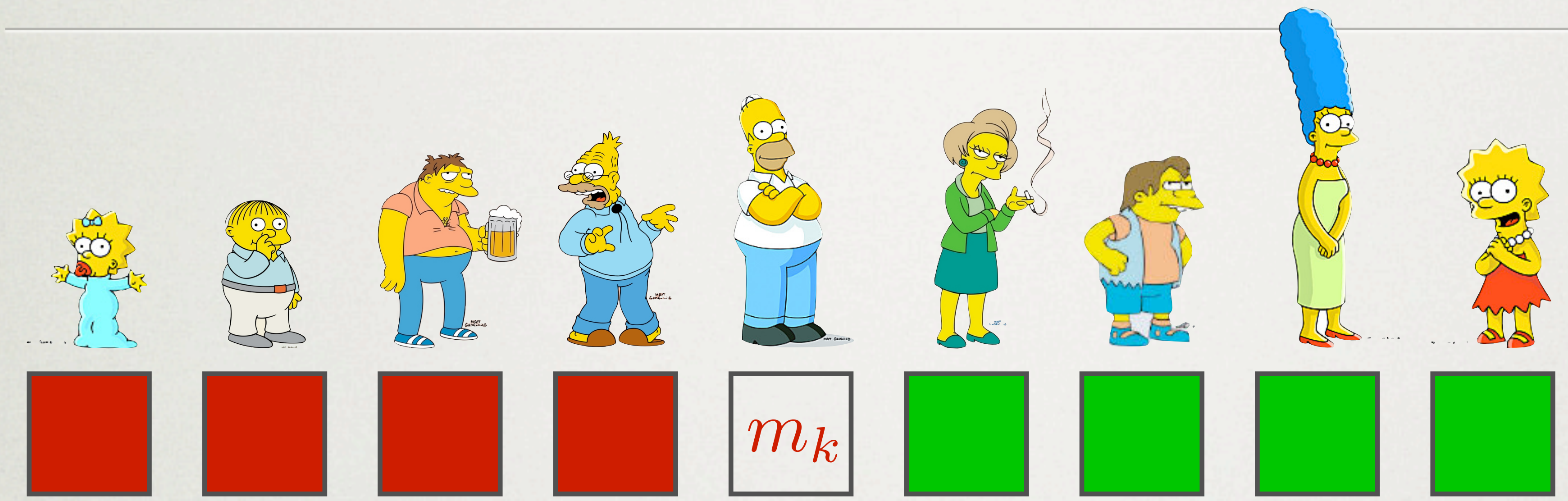


Hiring Above Median

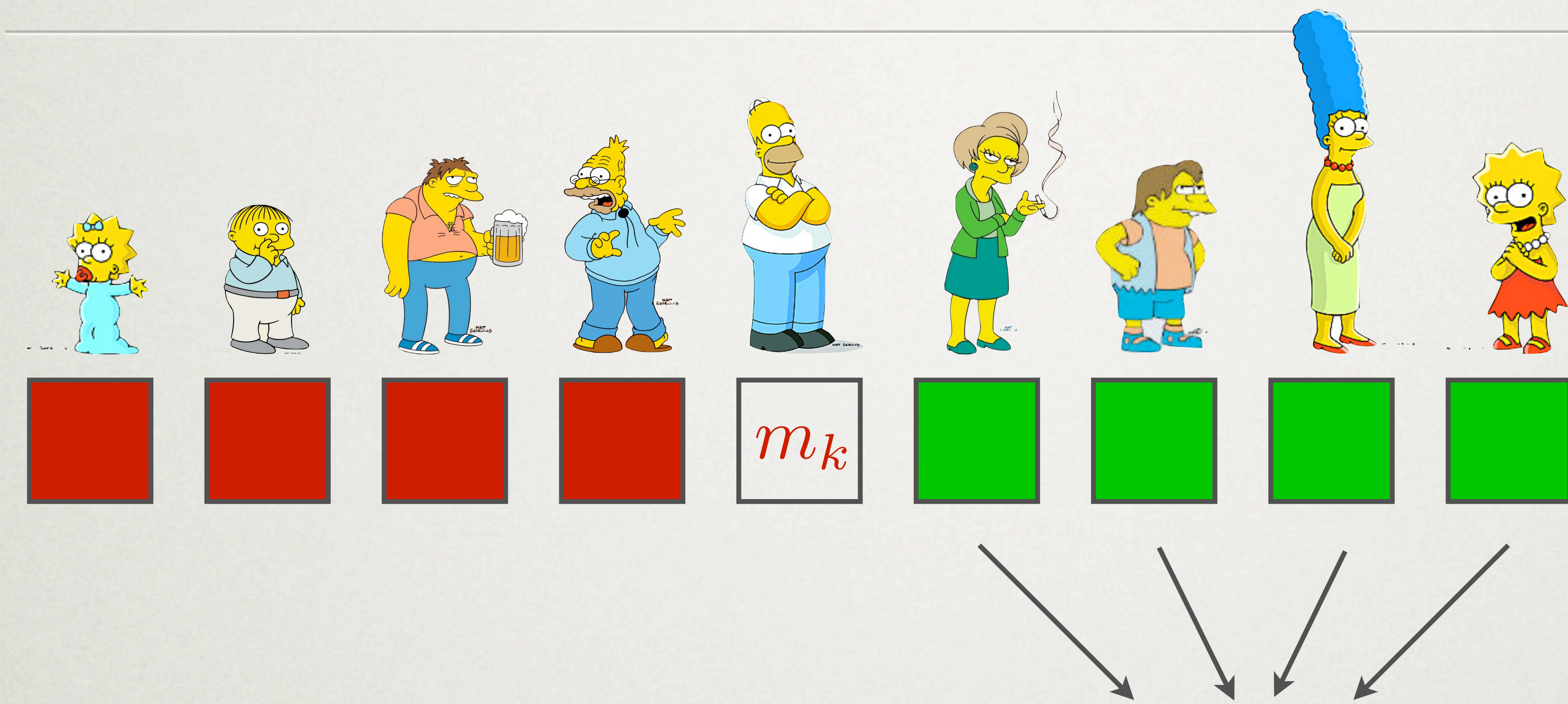


m_{k+1}

Hiring Above Median

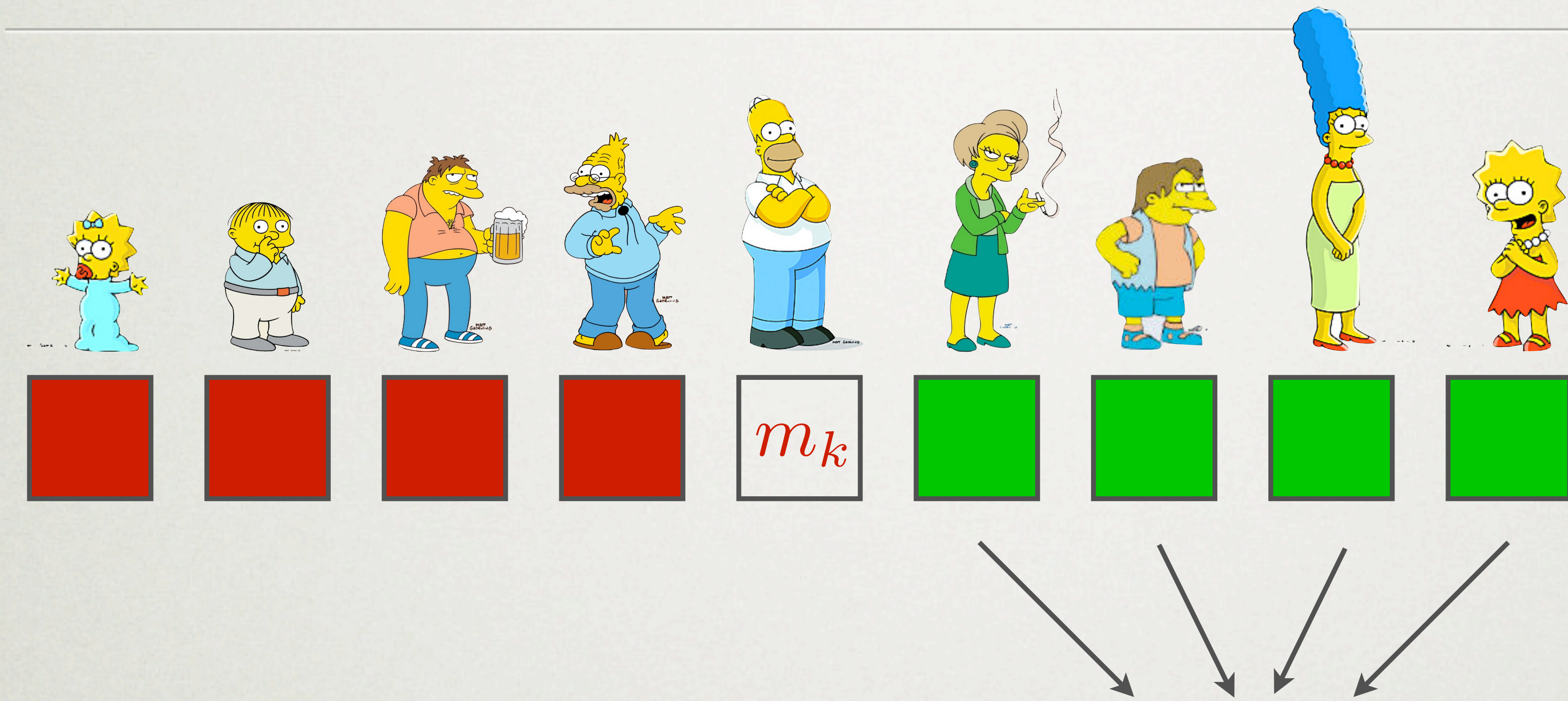


Hiring Above Median



Inductive hypothesis: $Unif(m_k, 1)$

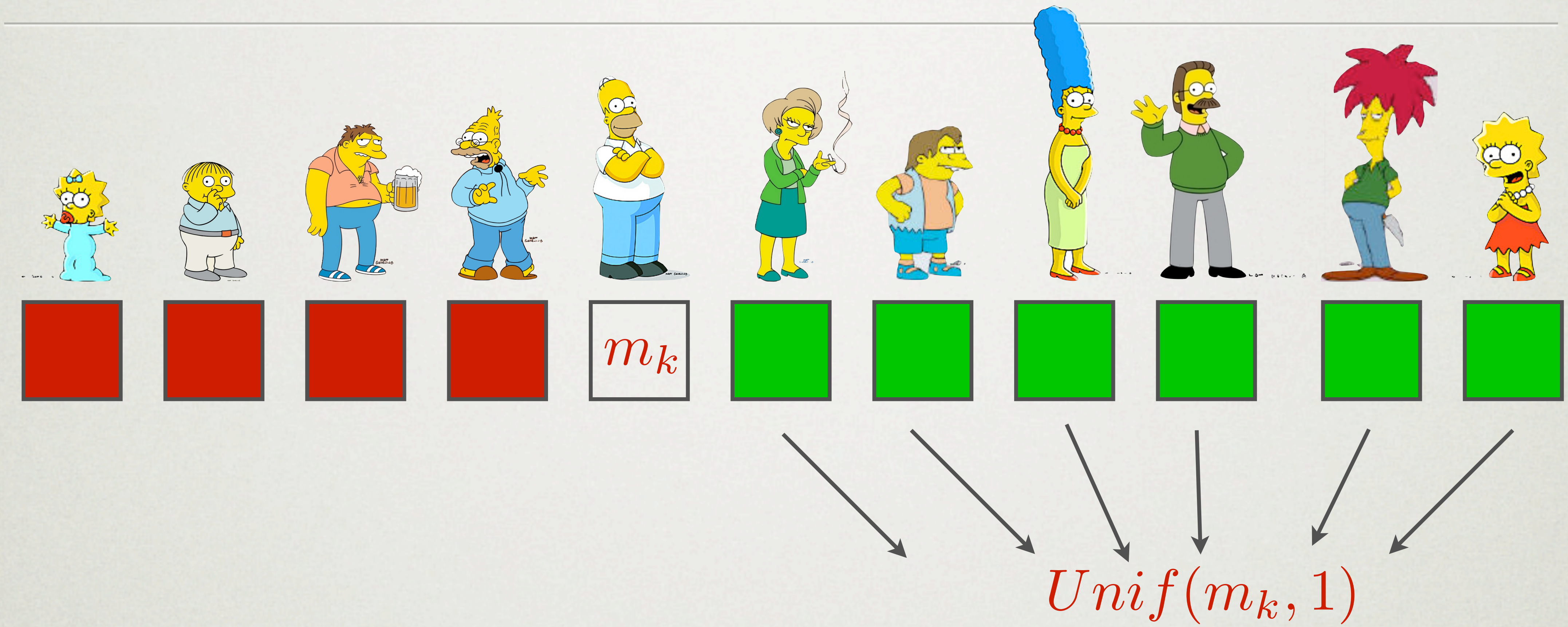
Hiring Above Median



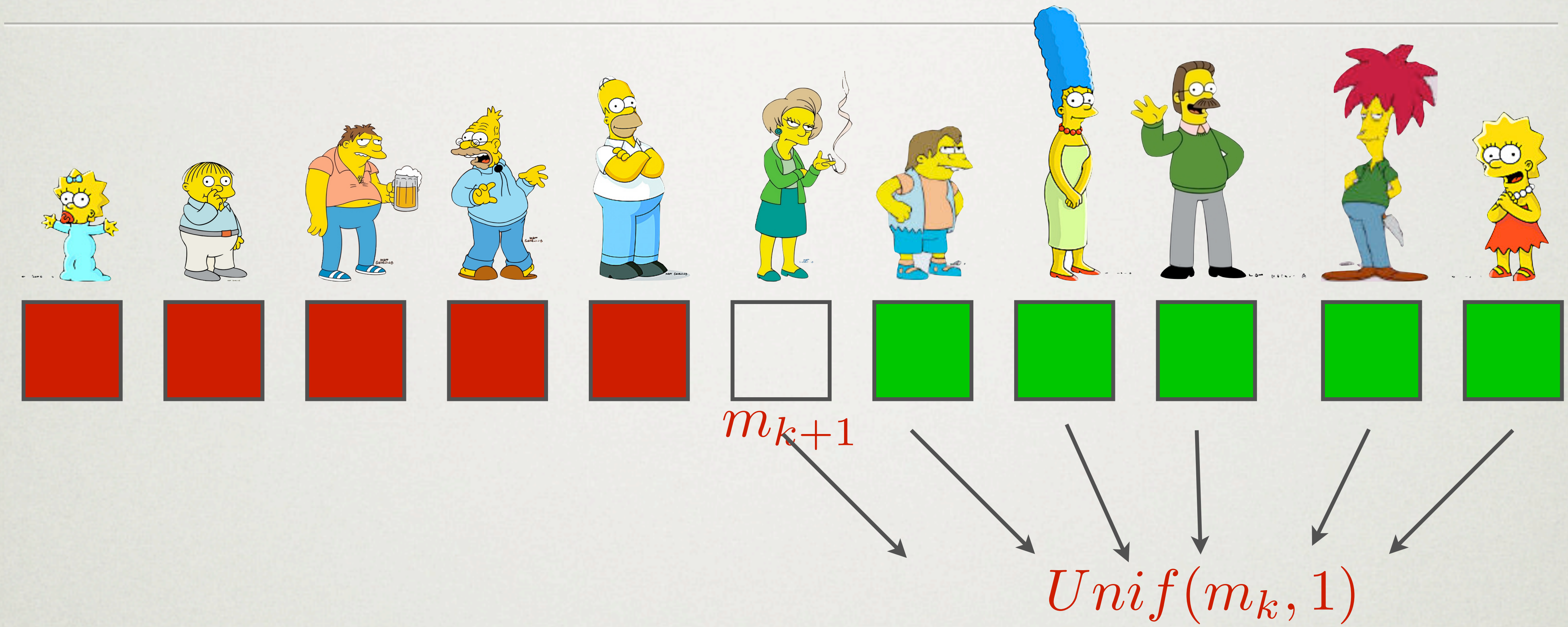
Inductive hypothesis: $Unif(m_k, 1)$

New hires: are $Unif(m_k, 1)$

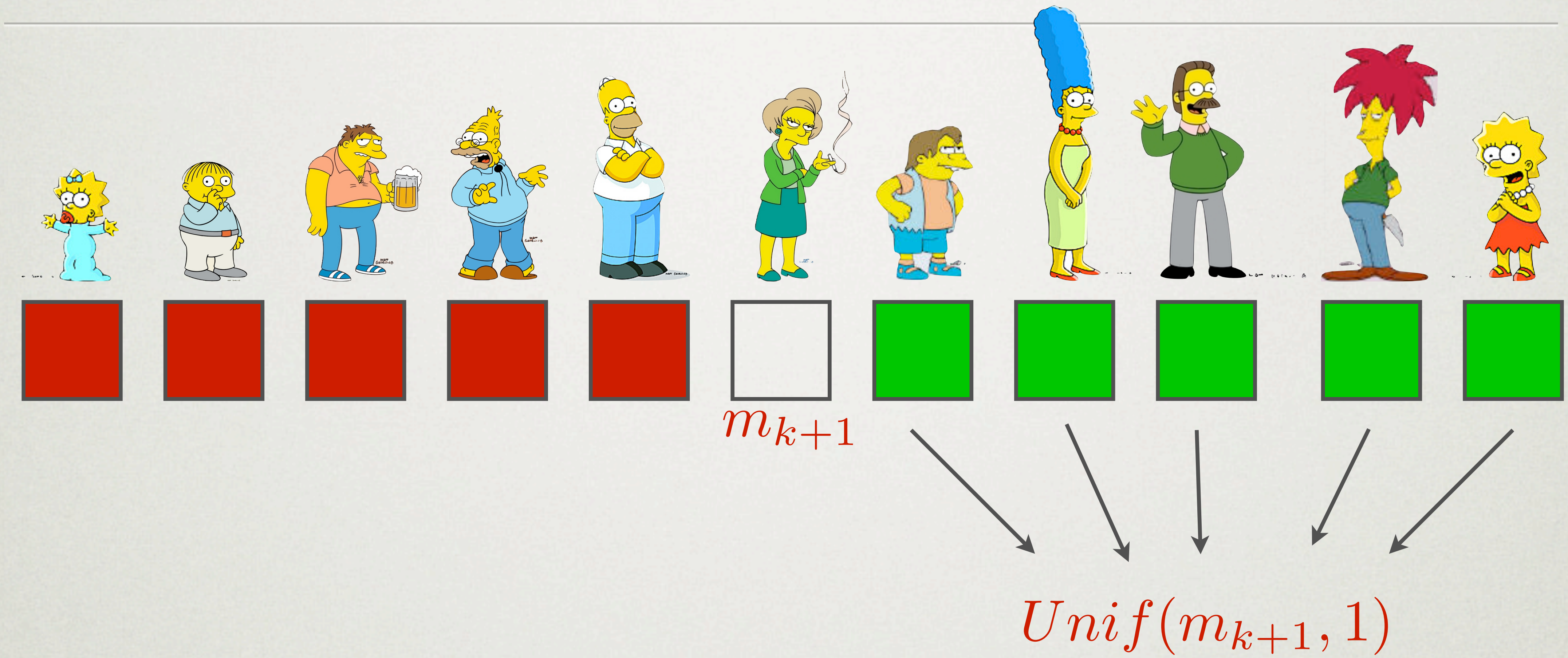
Hiring Above Median



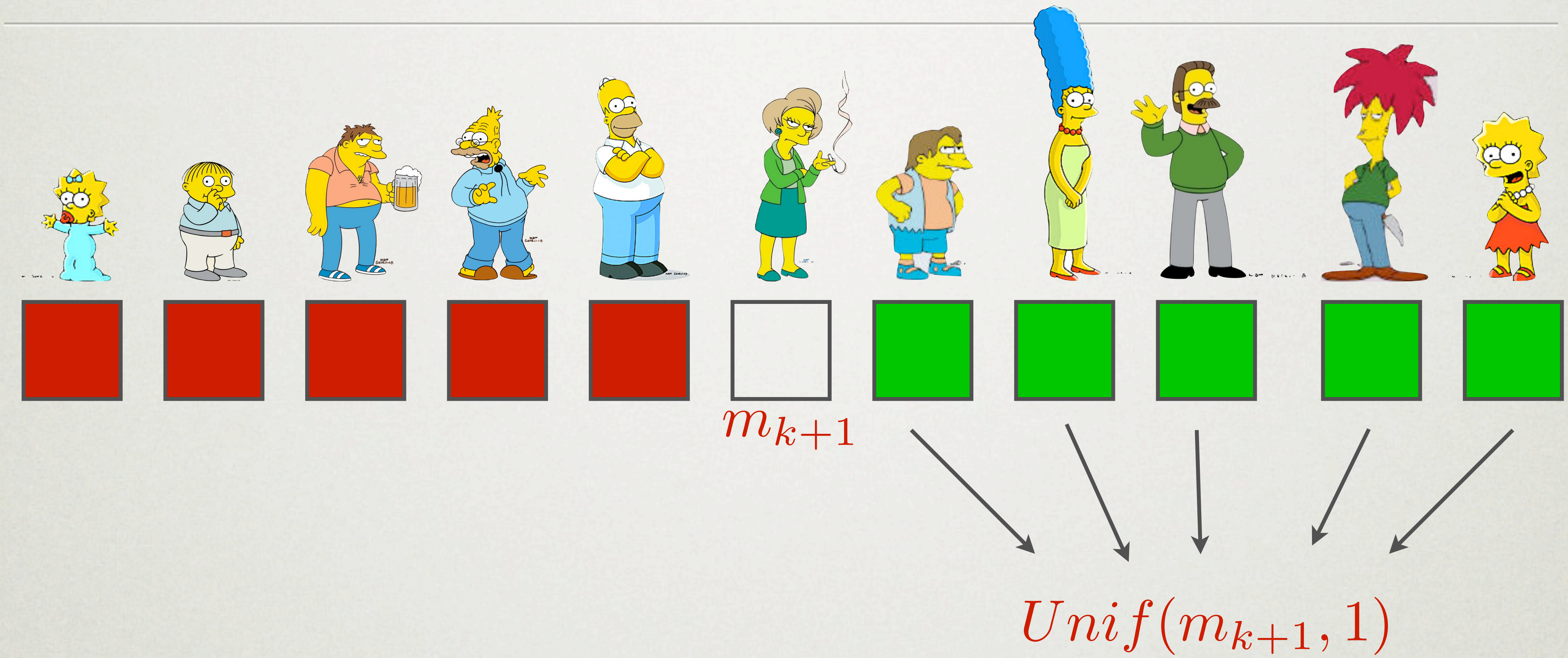
Hiring Above Median



Hiring Above Median

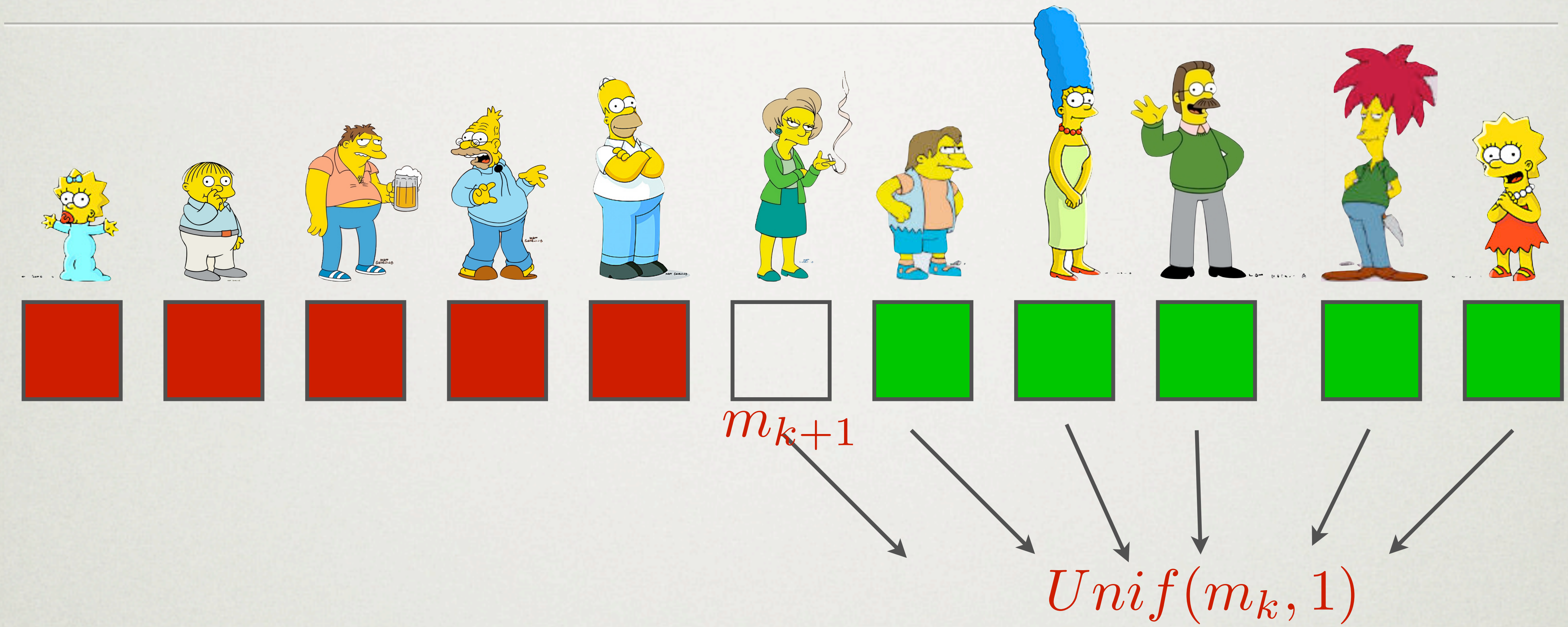


Hiring Above Median

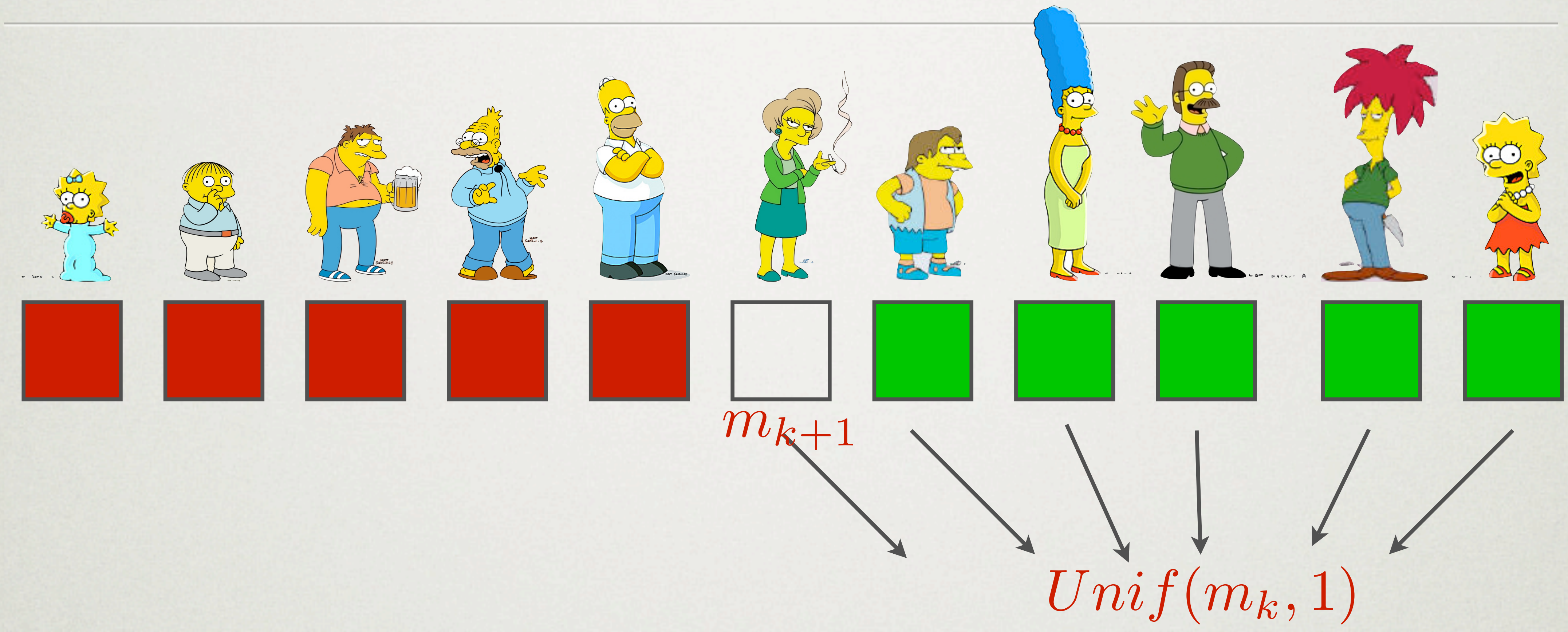


Induction holds.

Hiring Above Median

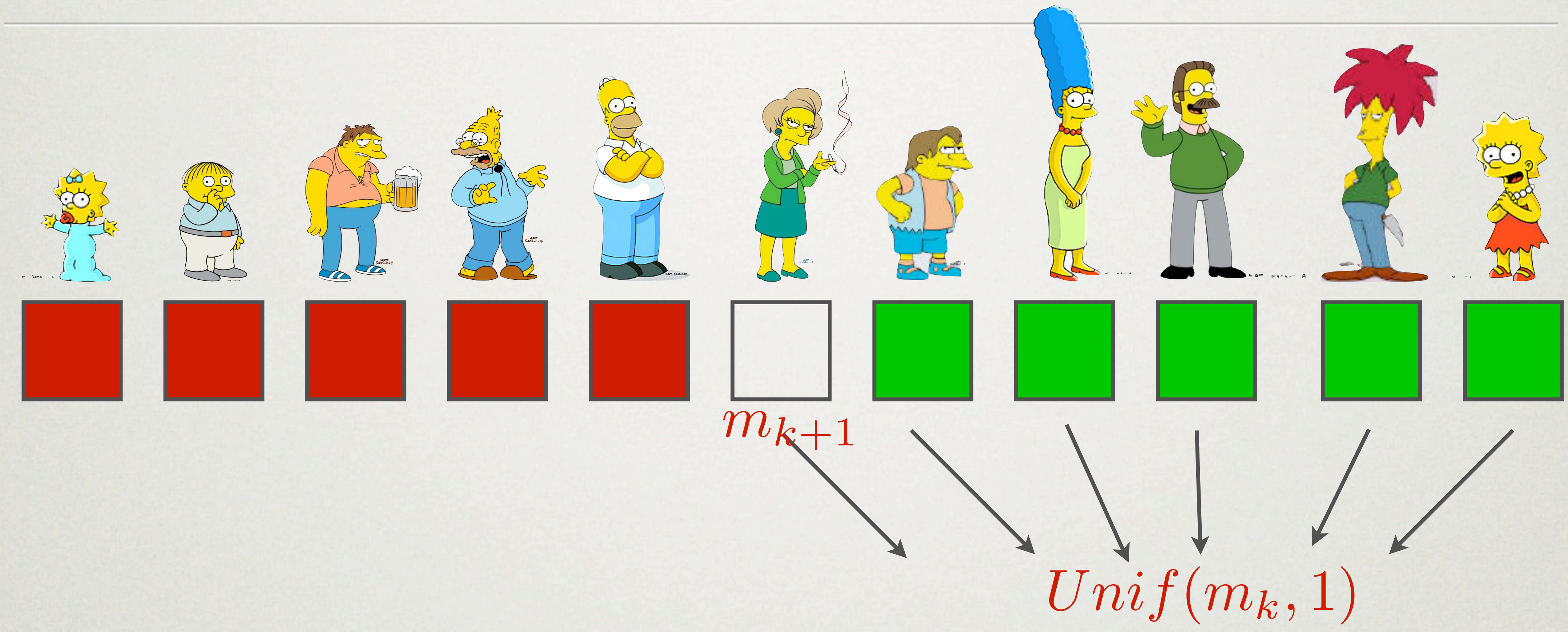


Hiring Above Median



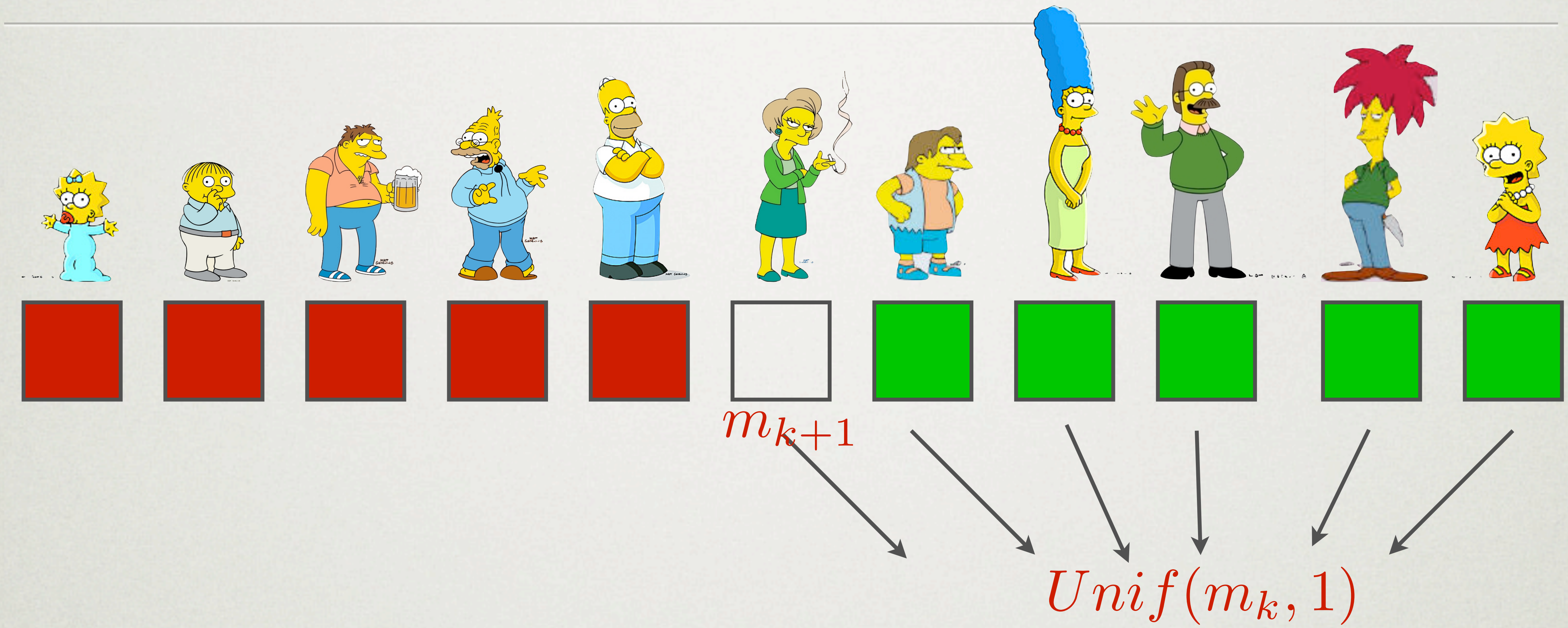
The median:

Hiring Above Median



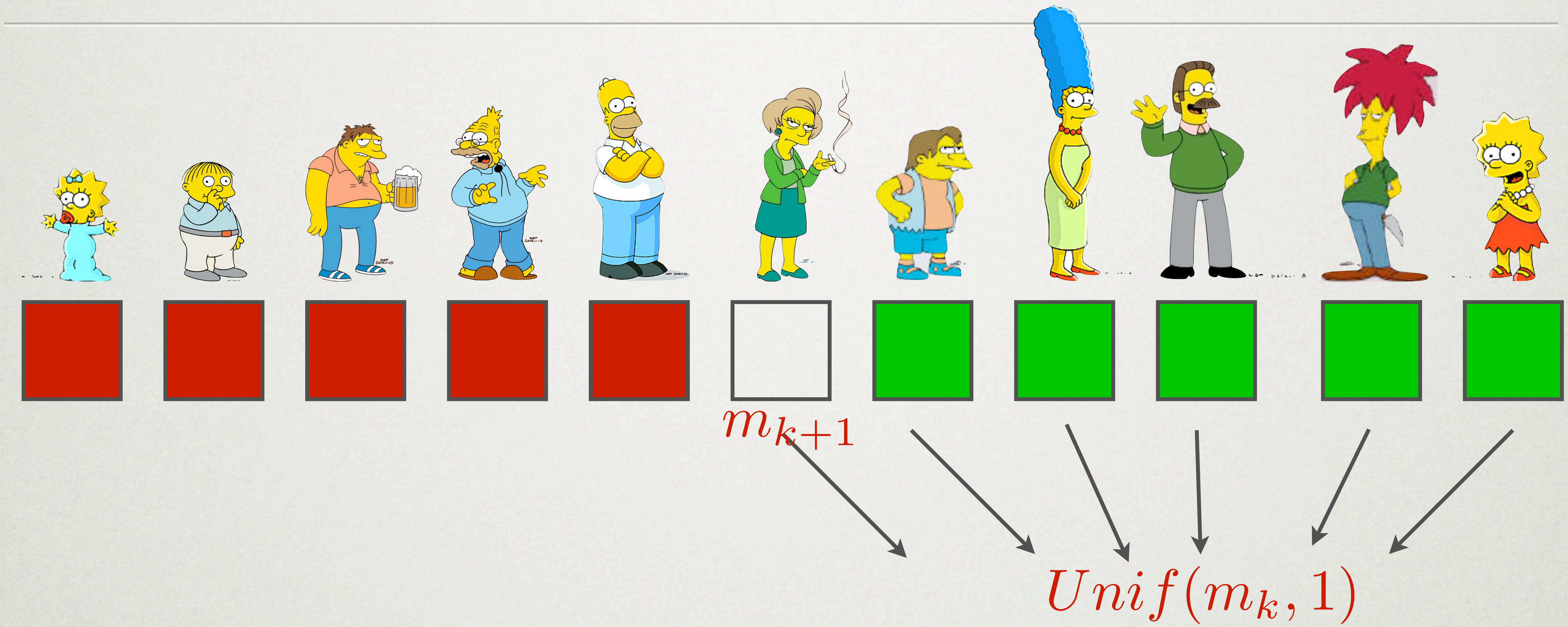
The median: smallest of $k+1$ uniform r.v., each $Unif(m_k, 1)$

Hiring Above Median



The median: smallest of $k+1$ uniform r.v., each $Unif(0, g'_k)$

Hiring Above Median



The median: smallest of $k+1$ uniform r.v., each $Unif(0, g'_k)$

$$g'_{k+1} | g'_k \sim g'_k Beta(k+2, 1)$$

Hiring Above Median

Expand (like the means): $g'_k \sim g \prod_{i=1}^{k} Beta(i+1, 1)$

Hiring Above Median

Expand (like the means): $g'_k \sim g \prod_{i=1}^{k} Beta(i+1, 1)$

$$E[g'_n] = \Theta(\frac{1}{n})$$

Hiring Above Median

Expand (like the means): $g'_k \sim g \prod_{i=1}^{k} Beta(i+1,1)$

$$E[g'_n] = \Theta(\frac{1}{n})$$

Caveat: this is the median gap! The mean gap is: $\Theta(\frac{\log n}{n})$

Hiring Above Median

Conclusion:

Average quality after n hires: $1 - \Theta(\frac{\log n}{n})$

Time to hire n employees: $\Theta(n^2)$

Again, very weak concentration results.

Extensions

Interview preprocessing:

- Self selection: quality may increase over time

Errors:

- Noisy estimates on i_q scores.

Firing:

- Periodically fire bottom 10% (Jack Welch Strategy)
- Similar analysis to the median.

Many more...

Thank You