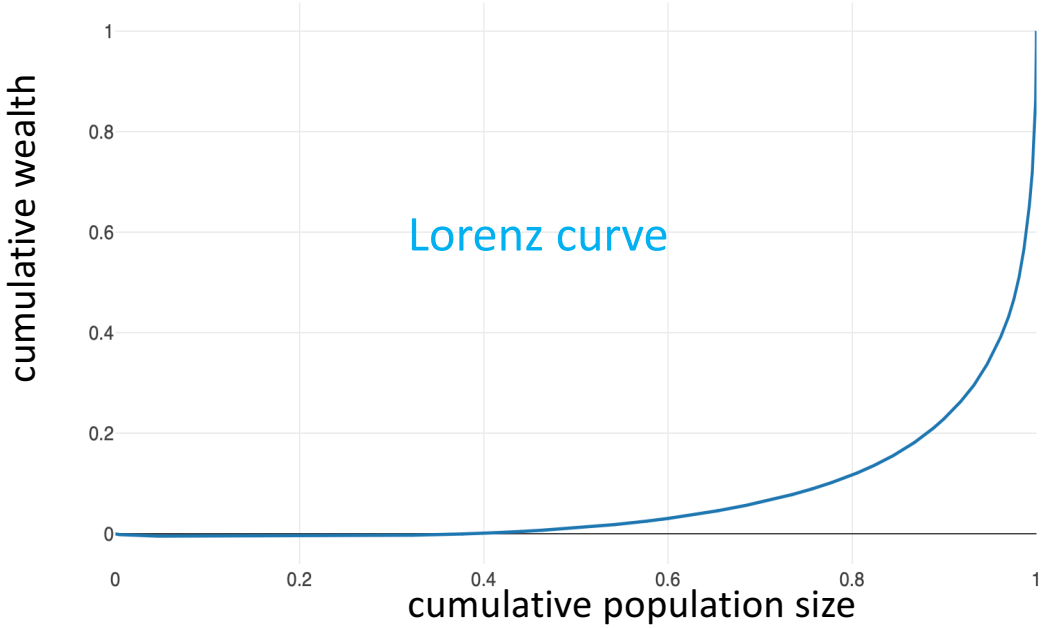
RL w/ hetero agents

https://github.com/shmister/inequality

The problem: top 1% owns 39% of the total wealth
bottom 40% no or negative wealth

what are the causes of the wealth inequality?



The problem: what are the causes of wealth inequality?

Many theories and non-verified opinions:

inheritance,
political/tax system,
access to better information…
individual characteristics

Stanford marshmallow experiment

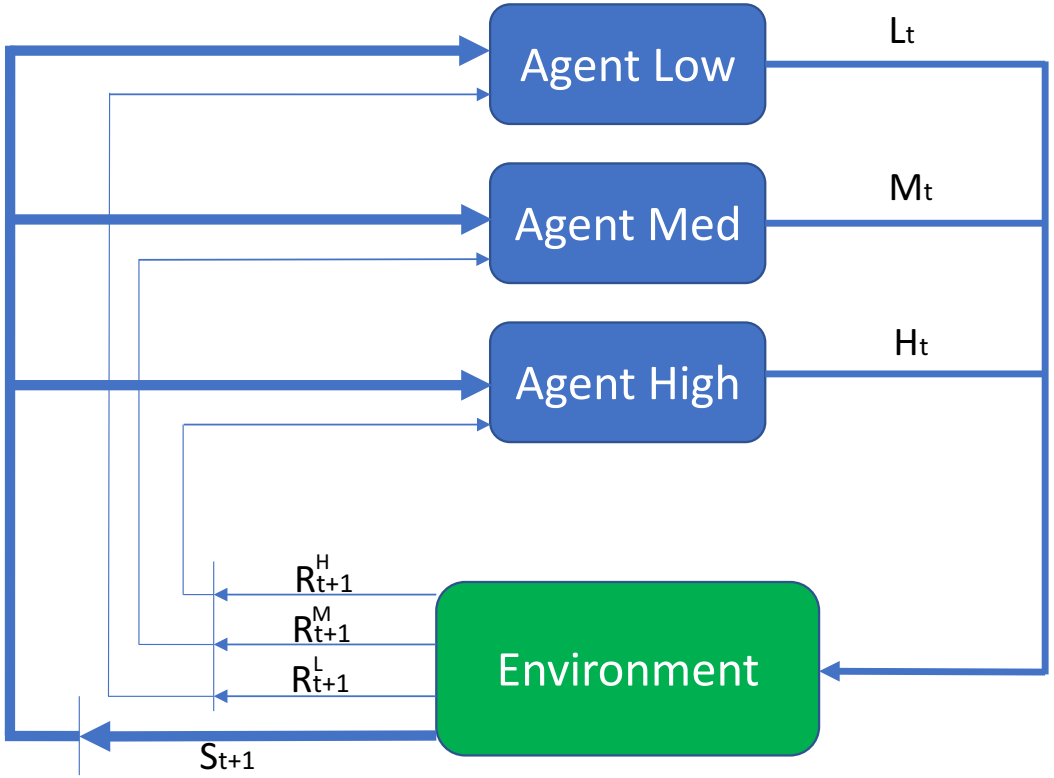
https://en.wikipedia.org/wiki/Stanford_marshmallow_experiment

wait longer

better life outcomes:
SAT scores,
educational attainment,
body mass index (BMI),
and other life measures

although the outcomes of the experiment are challenged now
intuitively, we know we are different

Heterogeneity in discounting and appreciating rewards



$$U_L([s_0, s_1, s_2, \ldots]) = u_L(s_0) + \beta_L u_L(s_1) + \beta_L^2 u_L(s_2) + \ldots$$

$$U_M([s_0, s_1, s_2, \ldots]) = u_M(s_0) + \beta_M u_M(s_1) + \beta_M^2 u_M(s_2) + \ldots$$

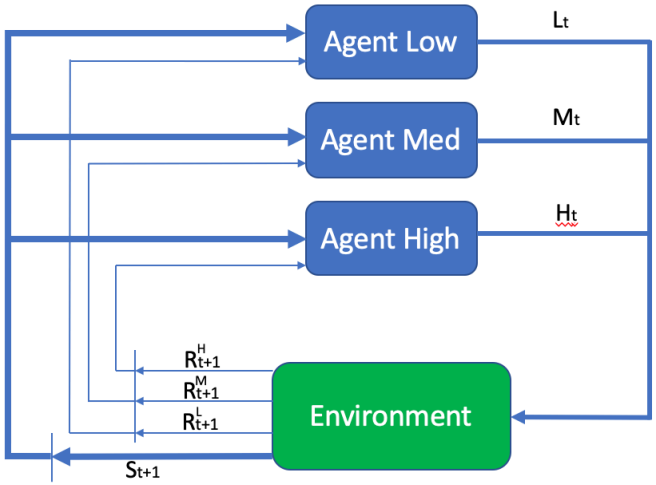
$$U_H([s_0, s_1, s_2, \ldots]) = u_H(s_0) + \beta_H u_H(s_1) + \beta_H^2 u_H(s_2) + \ldots$$

u_H, u_L *and* u_M are different
since the agents appreciate the rewards differently

policy

In this environment, agents receive reward (money) depending on employment (stochastic) and make a decision about the split
a) consume, i.e. immediate reward, b) invest to get future reward

Heterogeneity in discounting and appreciating rewards



$$U_L([s_0, s_1, s_2, \ldots]) = u_L(s_0) + \beta_L u_L(s_1) + \beta_L^2 u_L(s_2) + \ldots$$

$$U_M([s_0, s_1, s_2, \ldots]) = u_M(s_0) + \beta_M u_M(s_1) + \beta_M^2 u_M(s_2) + \ldots$$

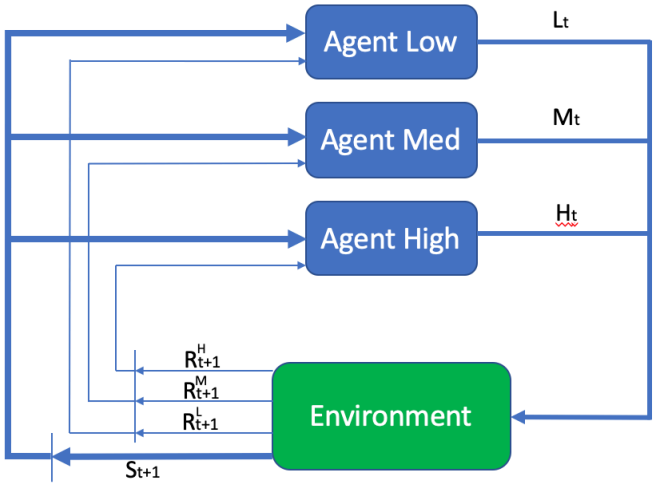
$$U_H([s_0, s_1, s_2, \ldots]) = u_H(s_0) + \beta_H u_H(s_1) + \beta_H^2 u_H(s_2) + \ldots$$

$$\beta_L < \beta_M < \beta_H$$

$$u = u(\gamma), \text{ where } \gamma \text{ is sensitivity to risk}$$

$$\gamma_L < \gamma_M < \gamma_H$$

Heterogeneity in discounting and appreciating rewards



$$U_L([s_0, s_1, s_2, \ldots]) = u_L(s_0) + \beta_L u_L(s_1) + \beta_L^2 u_L(s_2) + \ldots$$

$$U_M([s_0, s_1, s_2, \ldots]) = u_M(s_0) + \beta_M u_M(s_1) + \beta_M^2 u_M(s_2) + \ldots$$

$$U_H([s_0, s_1, s_2, \ldots]) = u_H(s_0) + \beta_H u_H(s_1) + \beta_H^2 u_H(s_2) + \ldots$$

$$\beta_L < \beta_M < \beta_H$$

$$u = u(\gamma), \qquad \text{where } \gamma \text{ is sensitivity to risk}$$

$$\gamma_L < \gamma_M < \gamma_H$$

Lt, Mt and *Ht* are policies that define what part to invest *k'* and what part to consume *c.*

Then the policies are *k' = f(k, F(s), D(k)).*

Agents:

three types (low, medium and high)
change their type stochastically with transition probabilities (MC)
each type has many agents

exploration: due to changing types, explore different paths
exploitation: due to big number of agents, exploit the same policy

Environment:

deterministic + stochastic to mimic recessions and expansions

Solution (Autonomous Learning Laboratory slide):

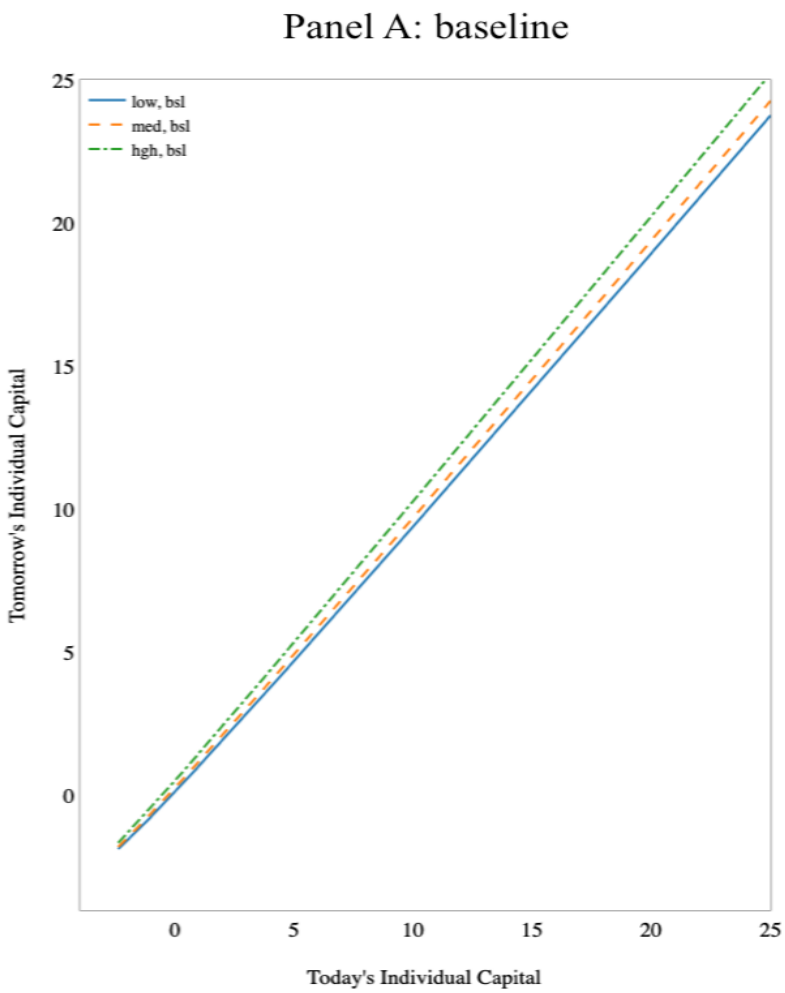
$$U_{i+1}^{\pi} \leftarrow R(s) + \gamma \sum_{s'} P(s' \mid s, \pi(s)) U_i^{\pi}(s')$$

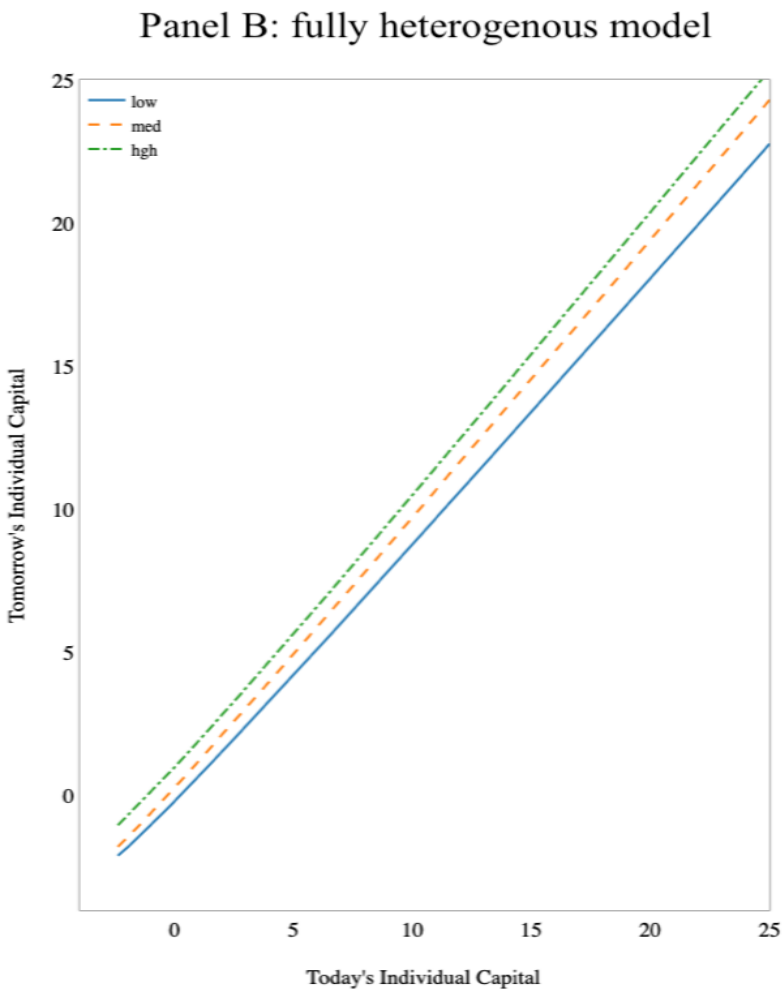
$$\pi_s^* = \arg\max_{\pi} U^{\pi}(s)$$

Policy Iteration

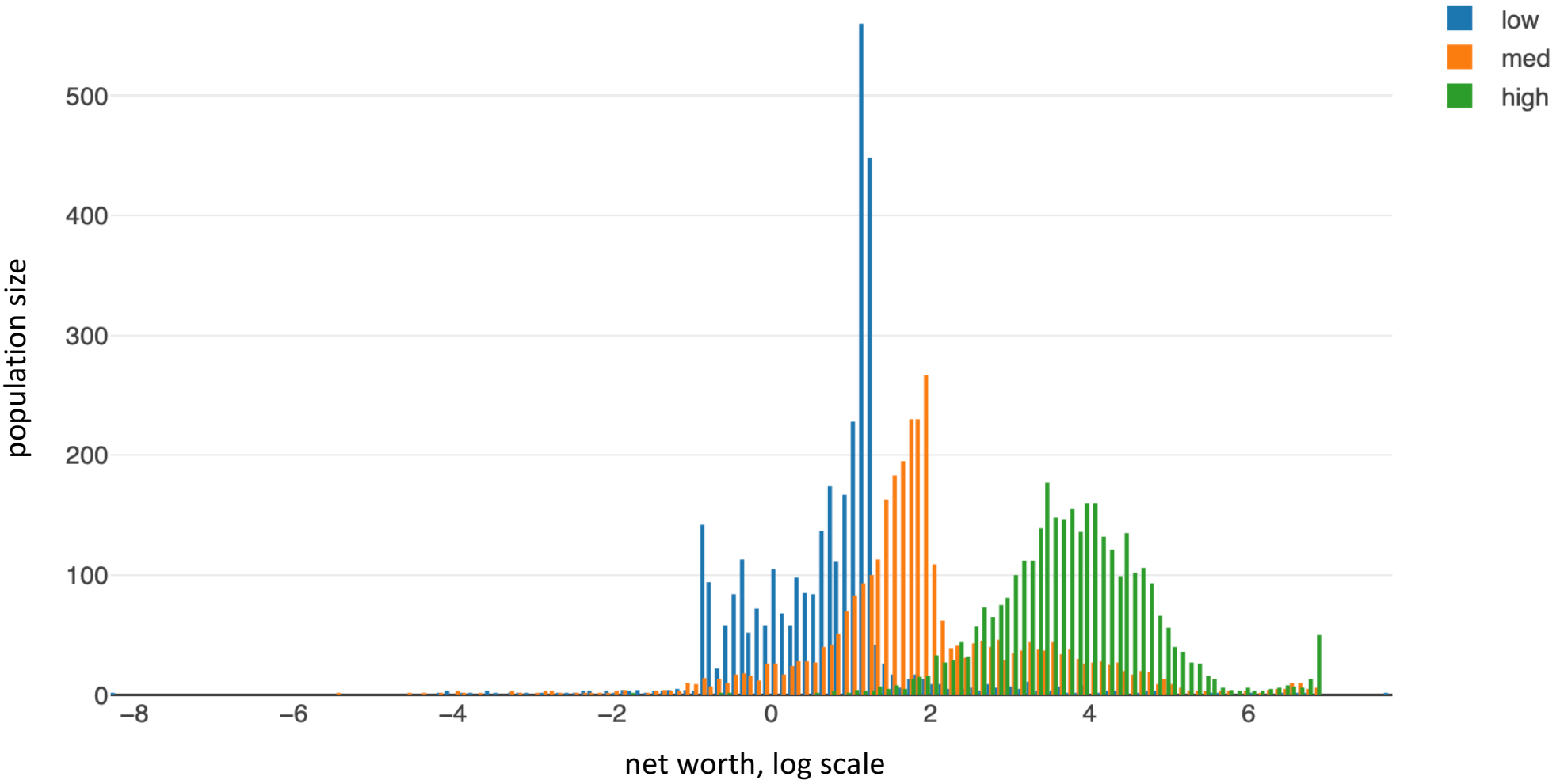
- Policy iteration interleaves two steps:
  - Policy evaluation: Given a policy, compute the utility of each state for that policy
  - Policy improvement: Calculate a new MEU policy
- Terminate when the policy doesn't change the utilities.
- Guaranteed to converge to an optimal policy

Results: learned policies





Results: wealth distribution



Results: cumulative wealth distribution