Improving Human Decision-Making with Machine Learning

Park Sinchaisri
Berkeley Haas

with Hamsa Bastani (Wharton)
& Osbert Bastani (Penn)

Learning is Costly

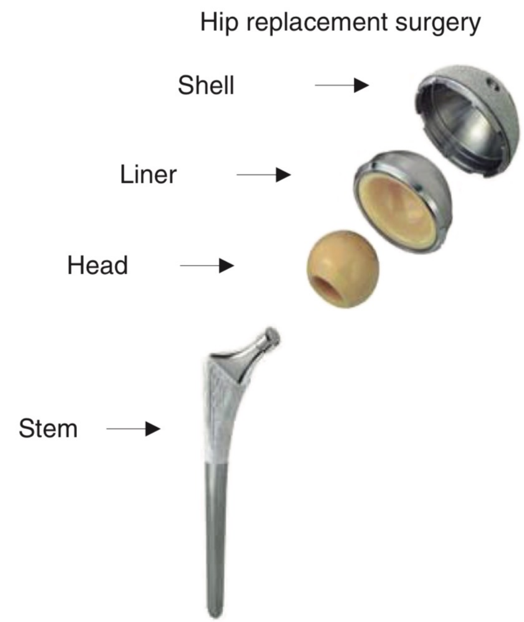
2+ years

to be fully productive

\$1,286/worker

training expenses

- Training Magazine 2019



New device = **+32.4%**

surgery duration

- Ramdas et al. 2018

Also – Tucker et al 2002, Ibanez et al 2017, Gurvich et al 2019, Bavafa & Jonasson 2020, Bloom et al 2020, ...

Learning from Experts

+10.9%

productivity

- Song et al. 2018

Also – Chan et al 2014, Herkenhoff et al 2018, Tan & Netessine 2019, Jarosch et al 2019, ...

Trace Data is Everywhere

Physicians

• ROACH,TRISTIN	Fibrinogen, INR, PT, PTT AMD_996304_76	MILLER,ALEX,MD *status: Unreviewed*	05•19•17
• ROACH,TRISTIN	Lipitor 80 mg	MILLER,ALEX,MD *status: Unreviewed*	05•18•17
• LEON,ERIN	Geriatric Wellness Visit	JONES,CAMERON,MD *status: Unreviewed*	05•16•17
• BECK,ALIVIA	Zocor 20 mg	JACK,JACK,MD *status: Unreviewed, held*	05•18•17
NORTON,BETHANY	Norvasc 10 mg	MILLER,ALEX,MD *status: Unreviewed*	05•18•17
MONTGOMERY,BLAINE	Glucophage 850 mg	OSHEA,JAMIE,MD *reviewed by: PPMD_AKN... status: Reviewed*	05•18•17
KLECK,MICHAEL	Office Visit - Abbreviated	JONES,CAMERON,MD *reviewed by: SUSAN status: Reviewed*	05•12•17
MCARDLE,HELEN	Office Visit - Mobile	JONES,CAMERON,MD *status: Unreviewed*	05•12•17

Uber Drivers

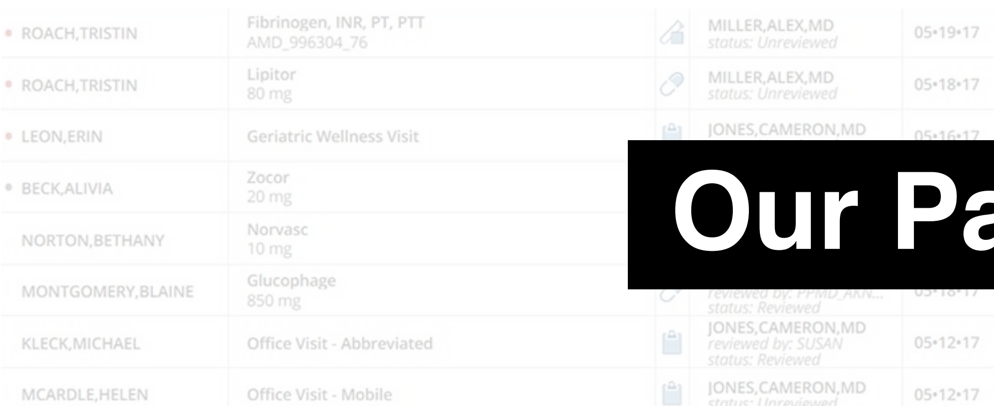
Trace data → Tips

Noisy, high-volume data
hard to extract insights

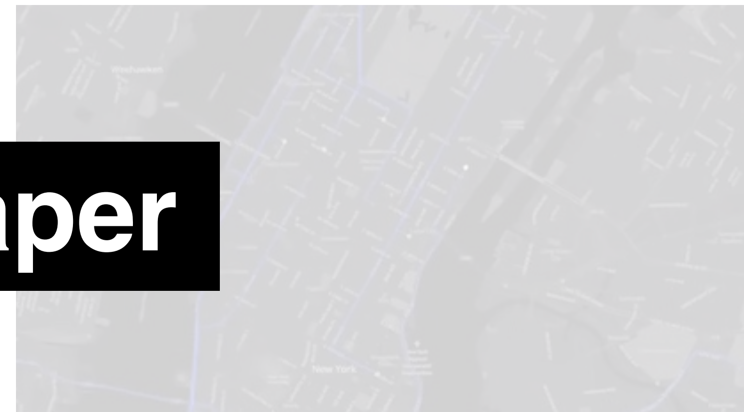
Key Q: can insights from ML
improve human decision-making?

Trace Data is Everywhere

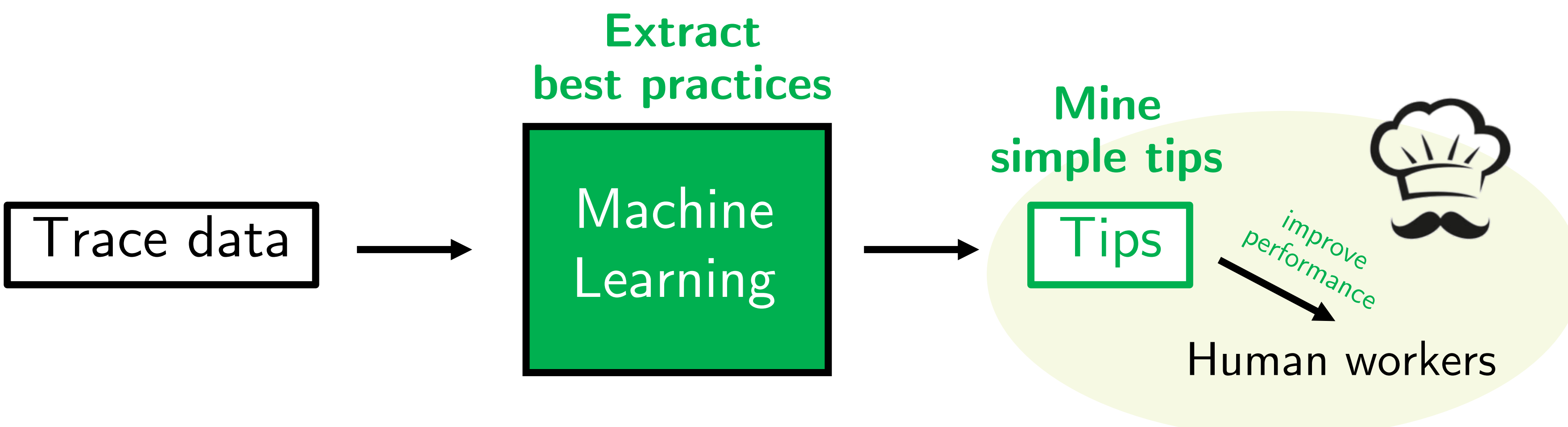
Physicians



Uber Drivers



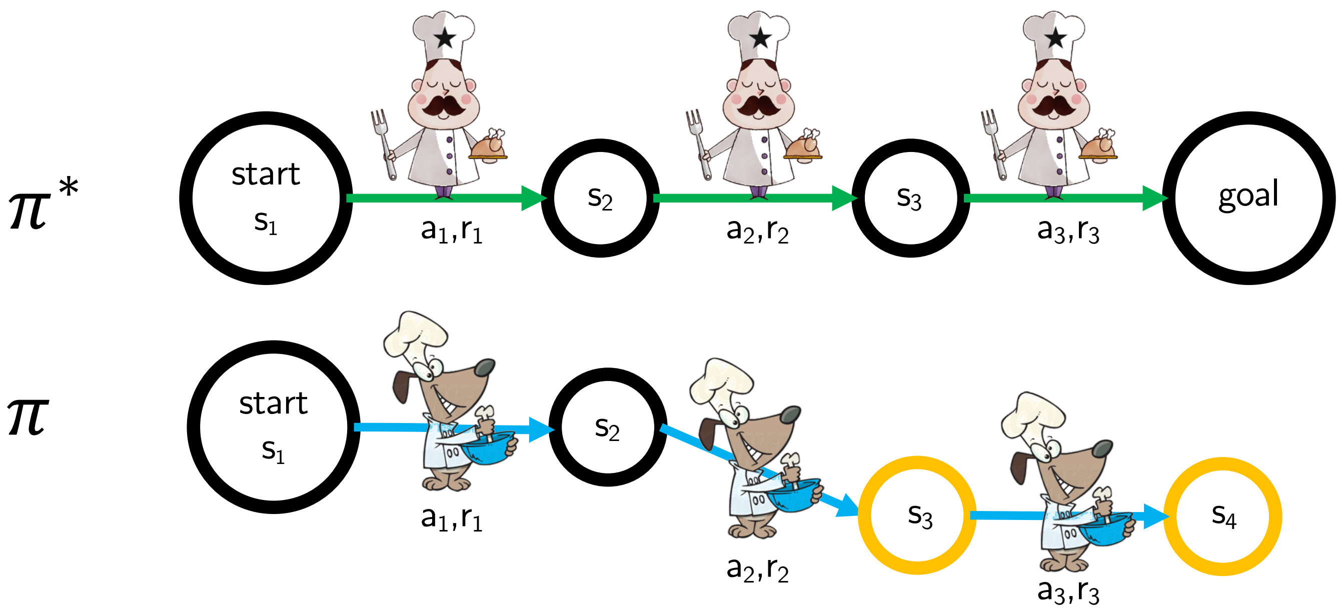
Our Paper



Problem Formulation

MDP Formulation:
Optimal policy and *human* make sequences of decisions

$$\mathcal{M} = (S, A, R, P, \gamma)$$



Problem Formulation

Input:
Trace data $\hat{d}_h$ from trainee

Human

sequences of state-action-reward tuples
$\{(s_1, a_1, r_1), (s_2, a_2, r_2), \ldots, (s_T, a_T, r_T)\}$

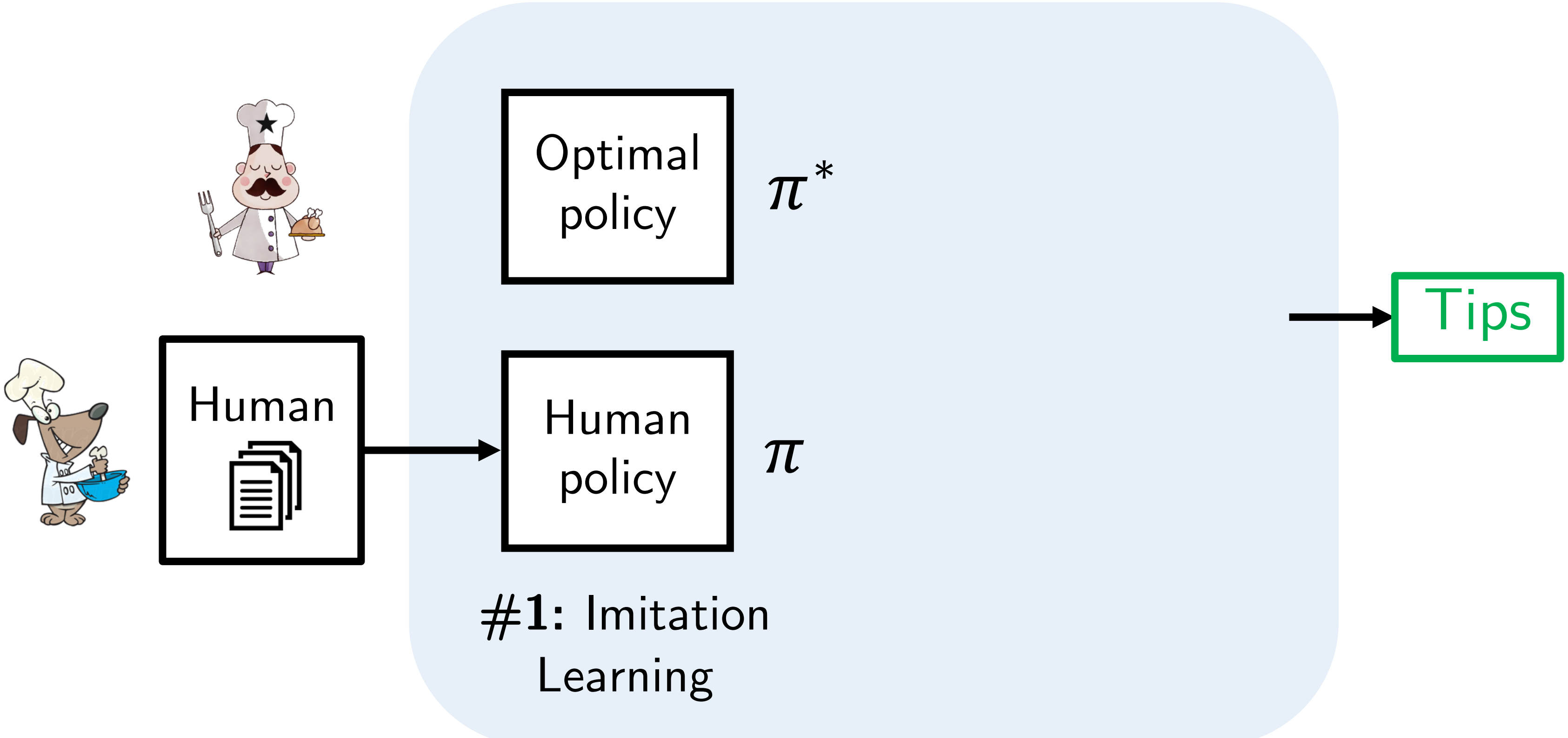
Tips

if **[state constraint]** then **[action]**

Caruana et al. 2015, Letham et al. 2015,
Bertsimas and Dunn 2017

Our Approach

$$\mathcal{M} = (S, A, R, P, \gamma)$$



Ng & Russell 2000, Abbeel & Ng 2004,
Ziebart et al. 2008, Ross & Bagnell 2011

Value function $V^{\pi}(s)$ is the cumulative reward obtained by using policy π from state s

$$V^{\pi}(s) = \mathbb{E}[\,\sum_{t=0}^{T} R(s_t, a_t) \mid s_0 = s, a_t = \pi(s_t)\,]$$

Step 1: Imitation Learning

Q function $Q^{\pi}(s,a)$ is the reward obtained by taking action a in state s and using policy π thereafter

$$Q^{\pi}(s,a) = \mathbb{E}_{s' \sim p(s'|s,a)}[V^{\pi}(s')]$$

\- Watkins & Dayan 1992

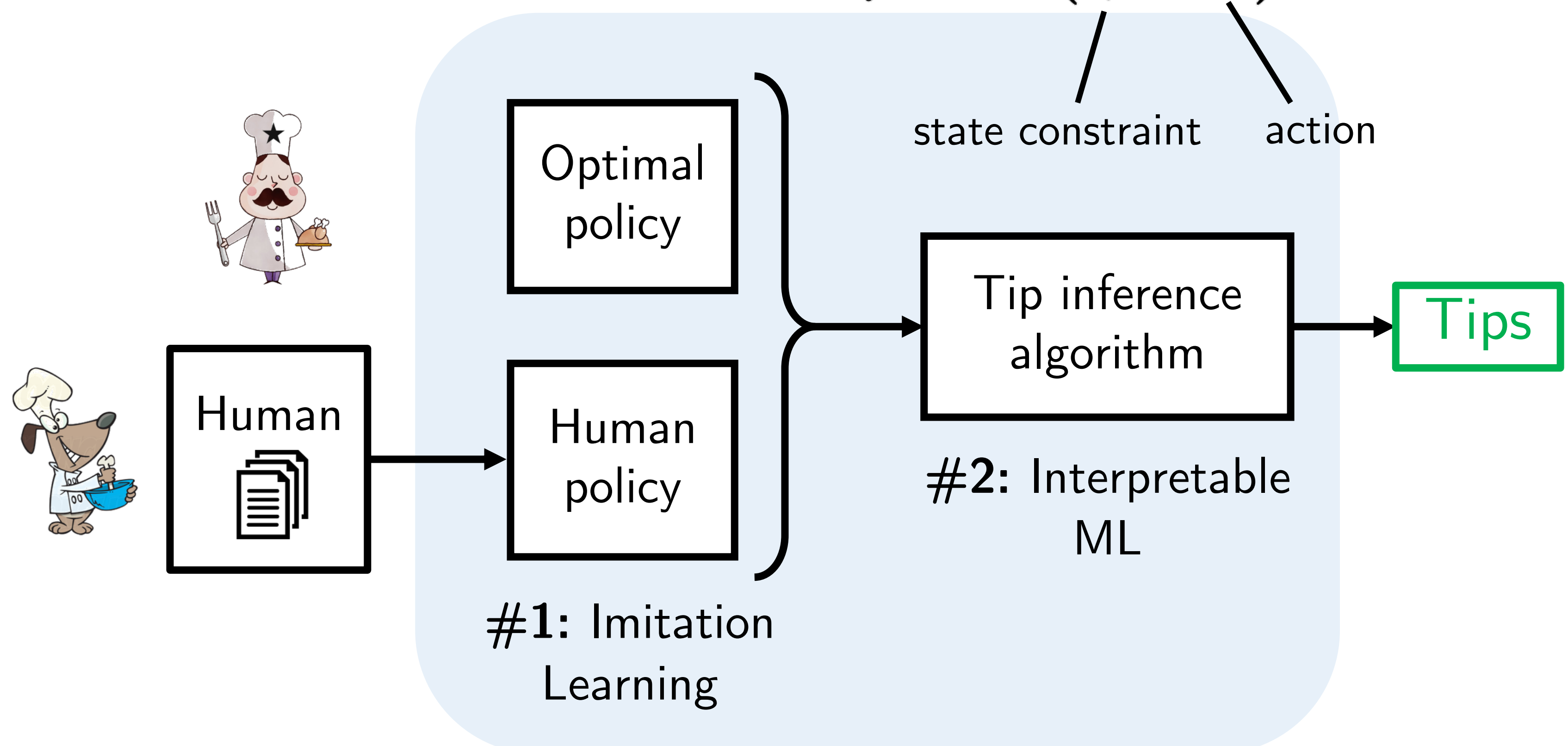
Parametrize policy using DNN π_θ
Optimize θ using the policy gradient algorithm

\- Williams 1992, Sutton 2000

Learn using supervised learning $\hat{Q}^{\pi}_{\theta}(s,a) \approx Q^{\pi}(s,a)$

Our Approach

$$\rho = (\psi, a)$$ "If ψ, then a"



Caruana et al. 2015, Letham et al. 2015

Step 2: Interpretable RL

- **Algorithm:** Choose tip ρ that maximizes the objective

$$J(\rho) = \underbrace{V^{\pi_{\mathrm{h}} \oplus \rho}(s_0)}_{\text{Human policy + tip}} - \underbrace{V^{\pi_{\mathrm{h}}}(s_0)}_{\text{Only human policy}}$$

  - $\pi_{\mathrm{h}} \oplus \rho$ denotes overriding the human policy with tip ρ.
  - $V^{\pi}(s)$ is the value function and s_0 is the initial state.
  - J measures the improvement in human reward

- **Intuition:** Want tips that maximize performance
  - Assumes the human follows the tip exactly

- **Challenge:** Hard to estimate $V^{\pi_{\mathrm{h}} \oplus \rho}$

Step 2: Interpretable RL

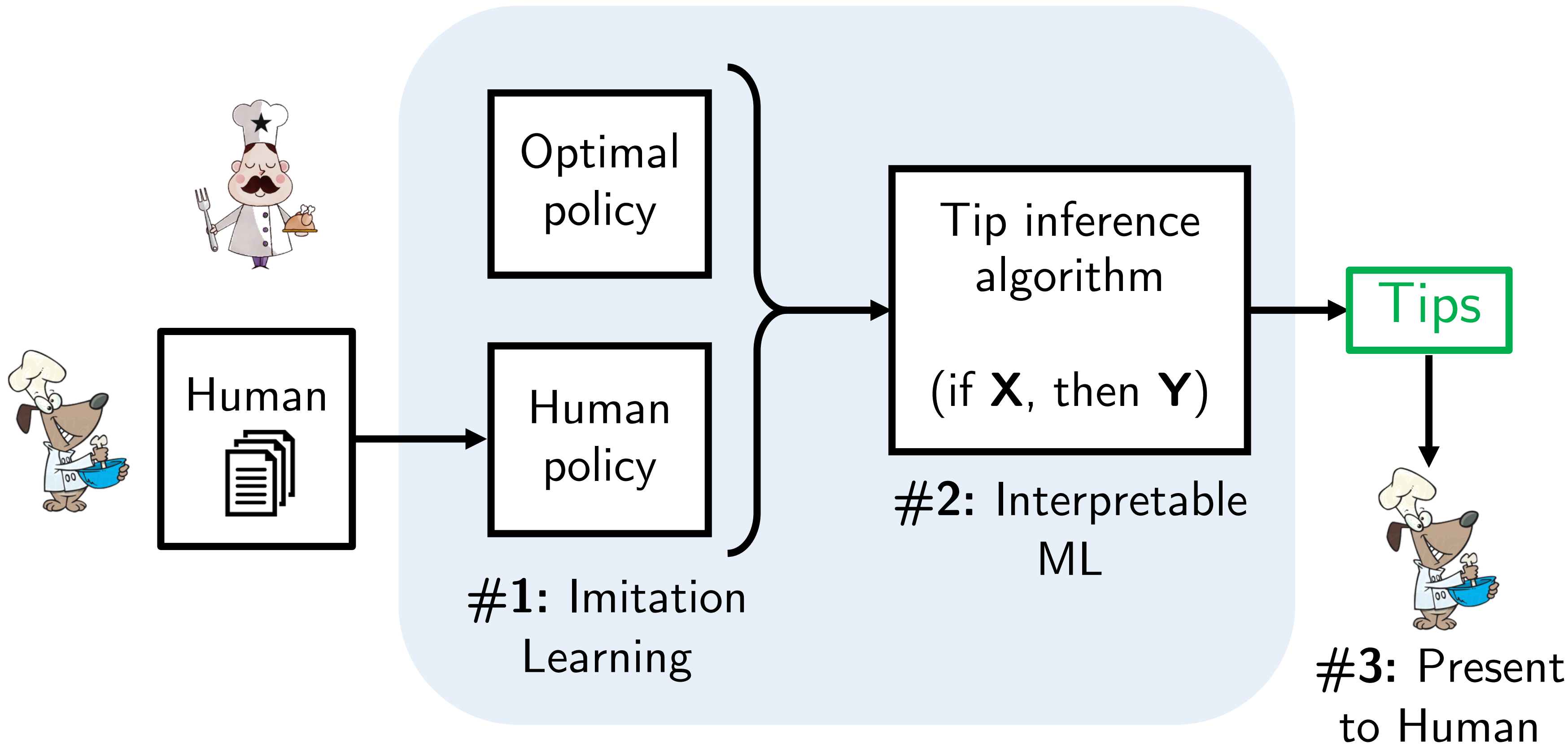
- **Key Lemma:** We have

$$J(\rho) \approx \mathbb{E}_{(s,a)\sim D_{\pi_{\mathrm{h}}\oplus\rho}}[Q^*(s, a \oplus \rho) - Q^*(s,a)]$$

  - Q^* is the optimal policy's Q function
  - D_π is the state-action distribution of policy π
  - $a \oplus \rho$ overrides the human action a if the tip is applicable in state s

- Rank tips based on expected improvement

Our Approach



Queueing Game

Burger Queen

x 4 within 50 ticks

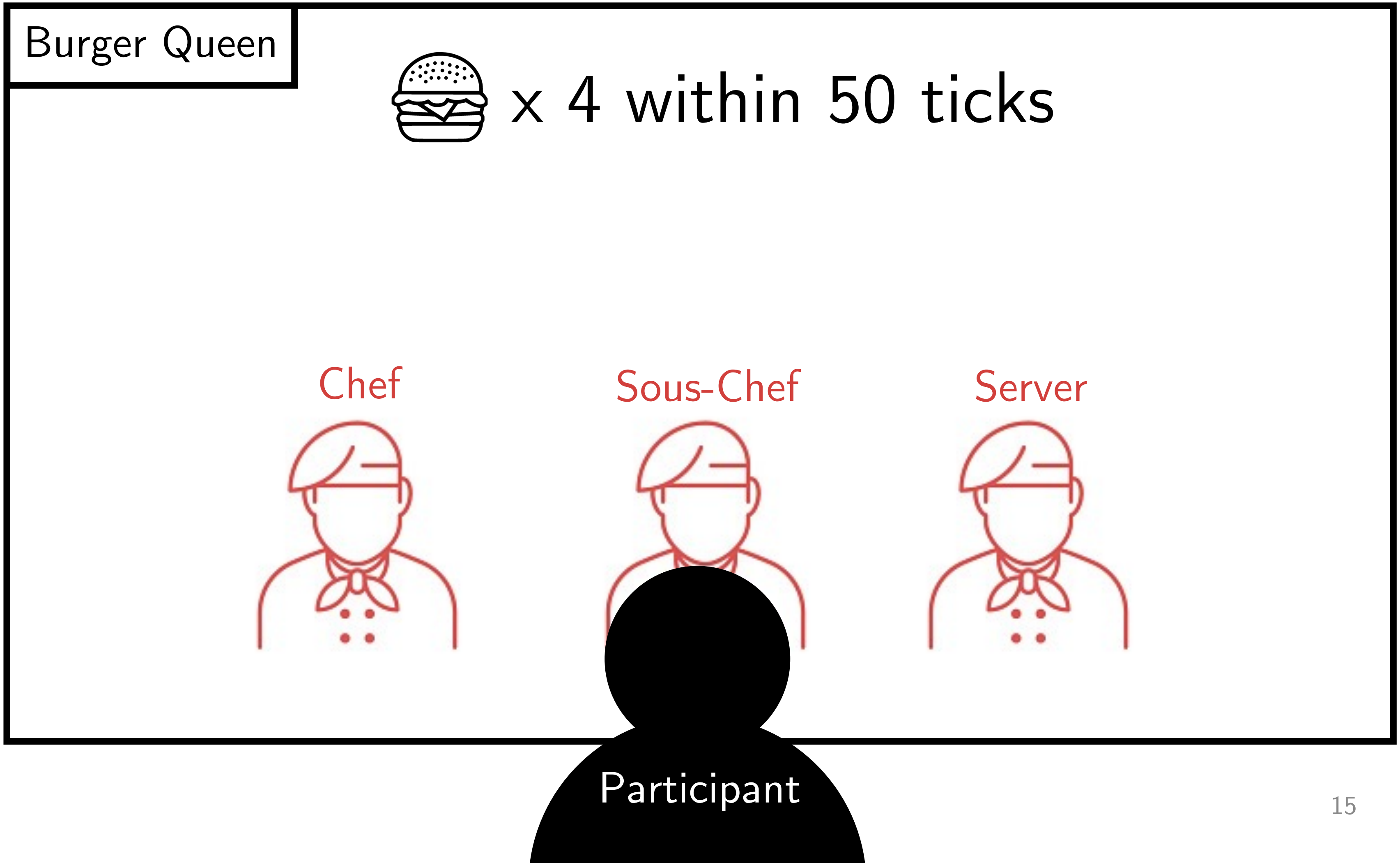
Making Burgers

Chop meat (2 ticks) → Cook burger (10 ticks) → Plate (2 ticks)

Participant

Pre-registered at
https://aspredicted.org/blind.php?x=8ye5cb

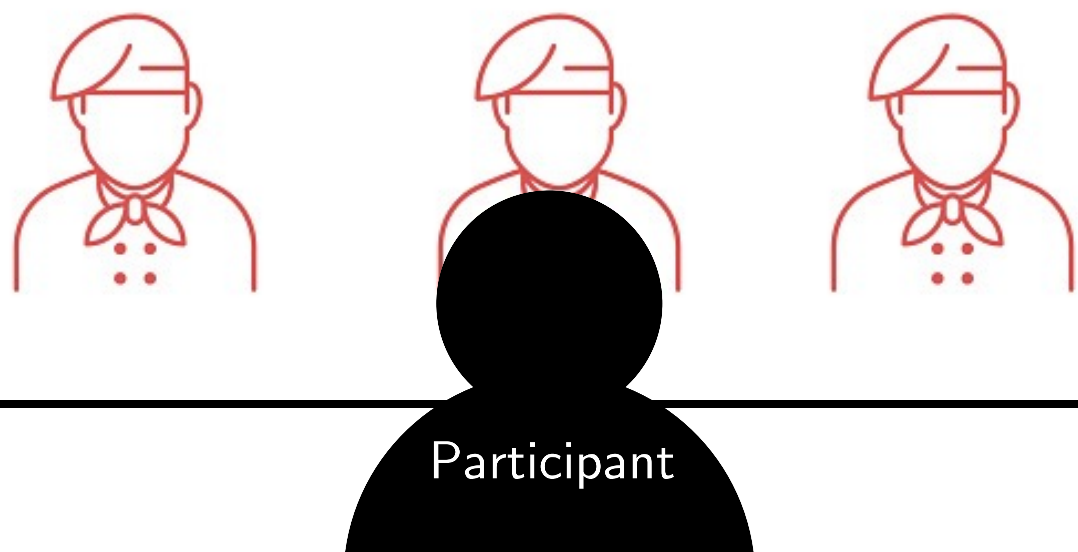
Queueing Game



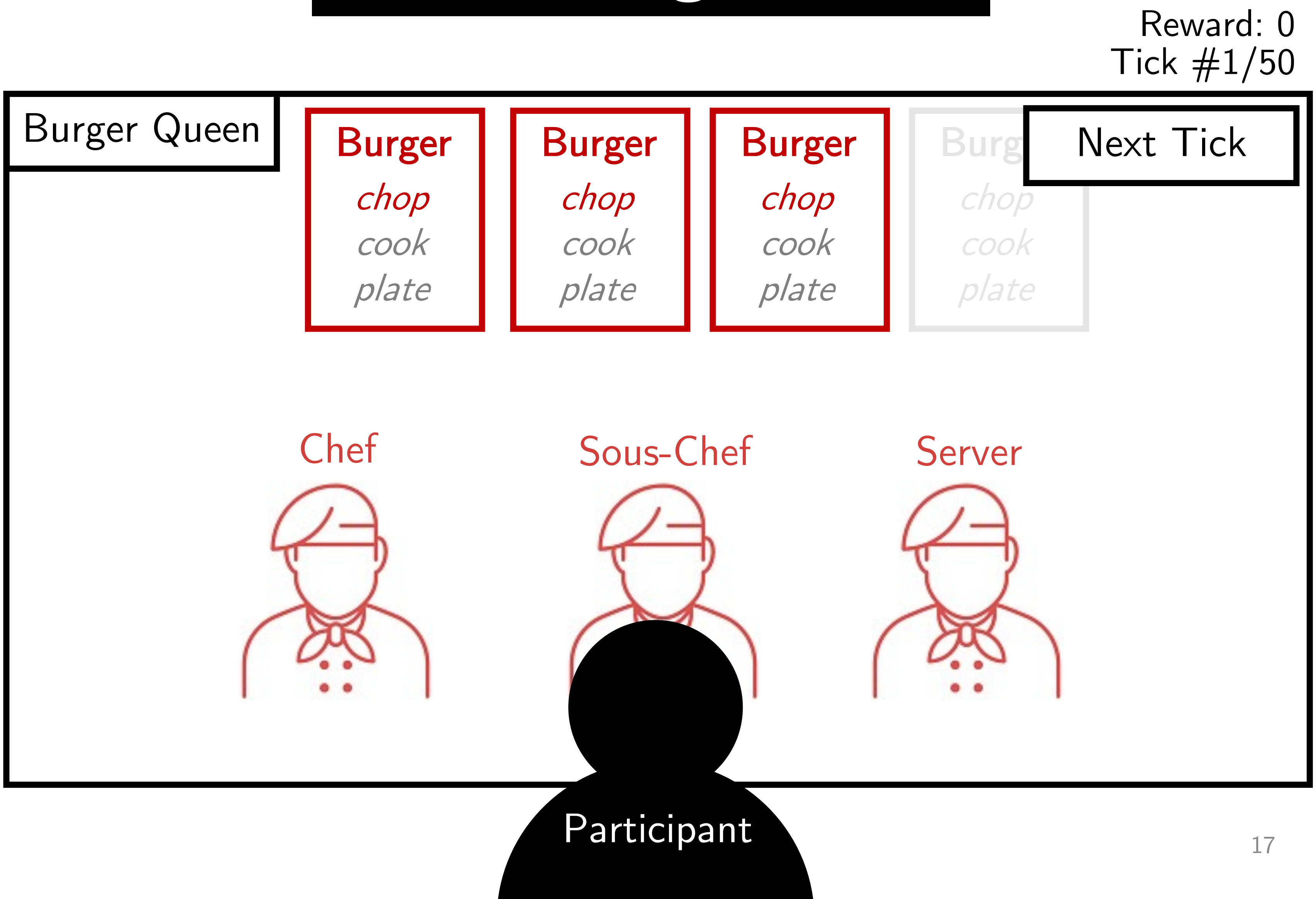
Queueing Game

Burger Queen

Chopping:	Fast	Average	Slow
Cooking:	Fast	Average	Slow
Plating:	Slow	Average	Fast
	Chef	Sous-Chef	Server

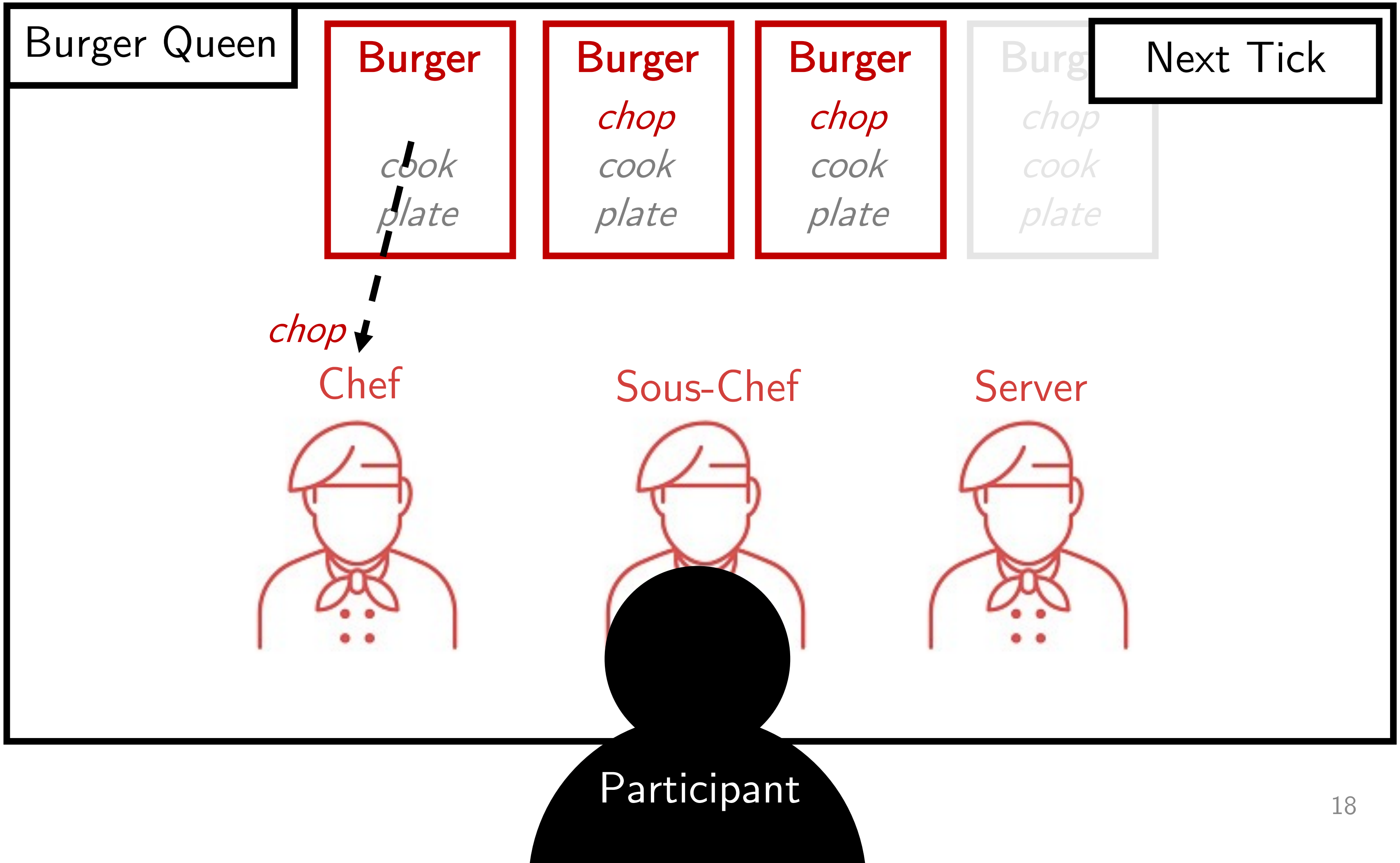


Queueing Game



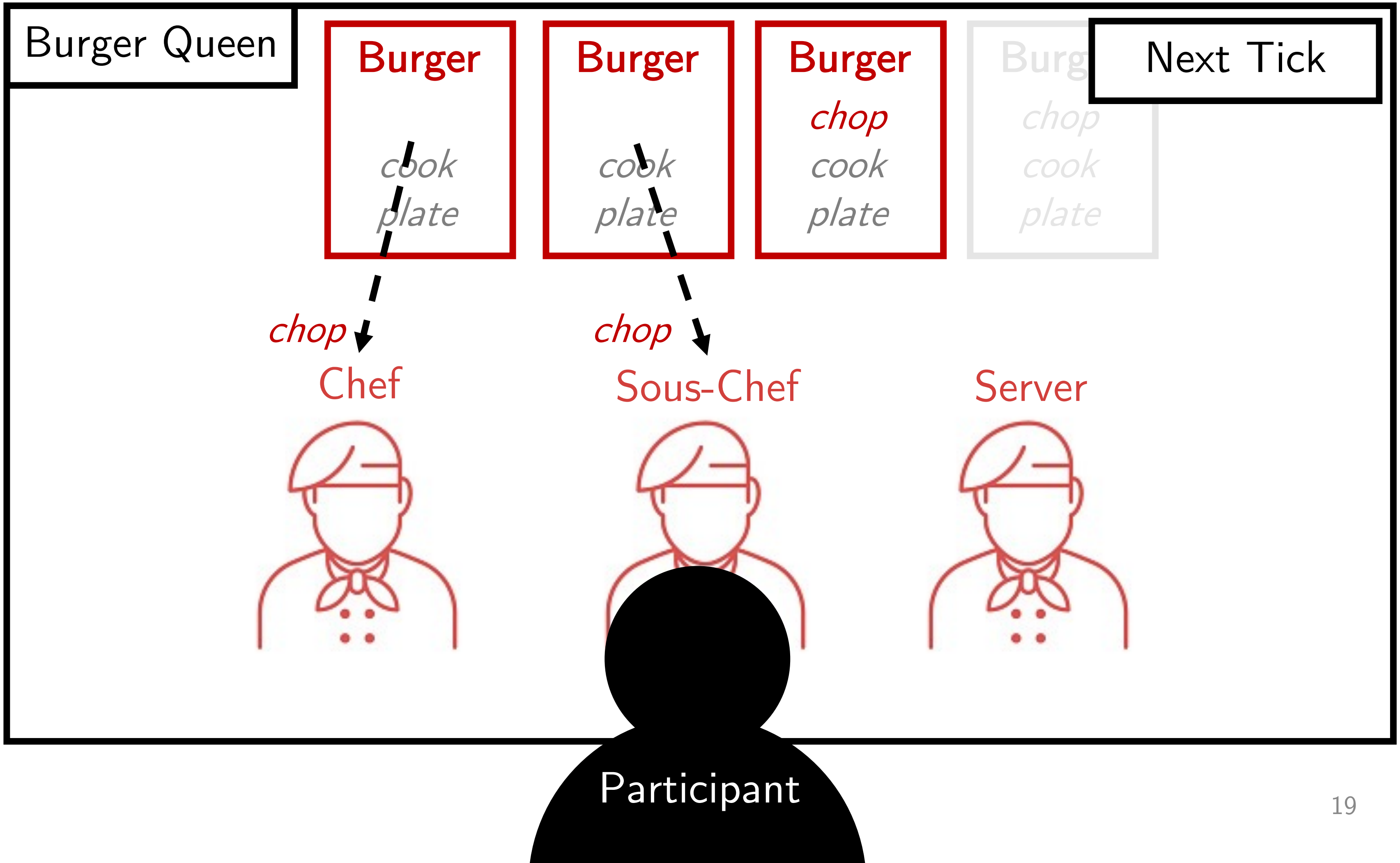
Queueing Game

Reward: 0
Tick #1/50



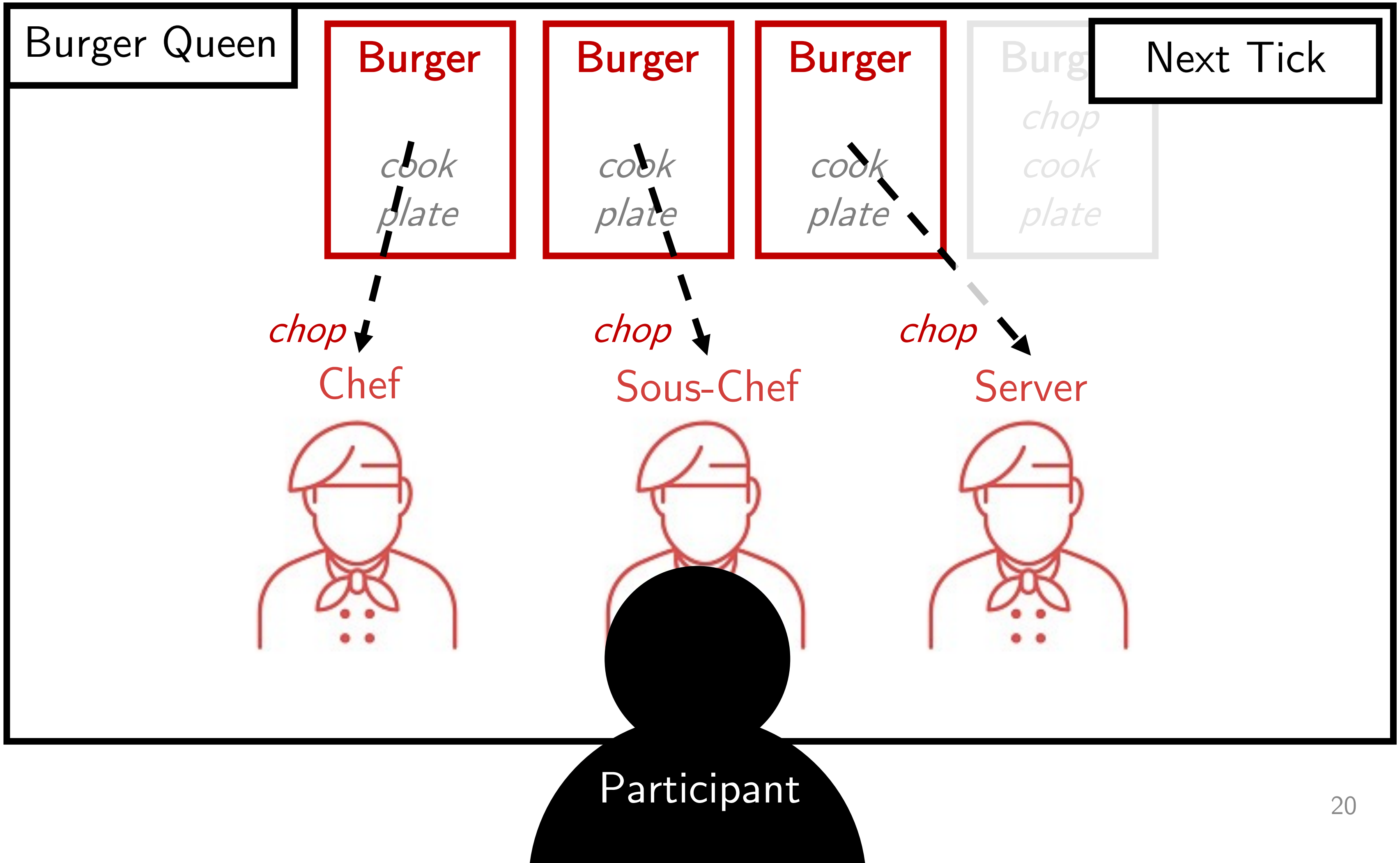
Queueing Game

Reward: 0
Tick #1/50



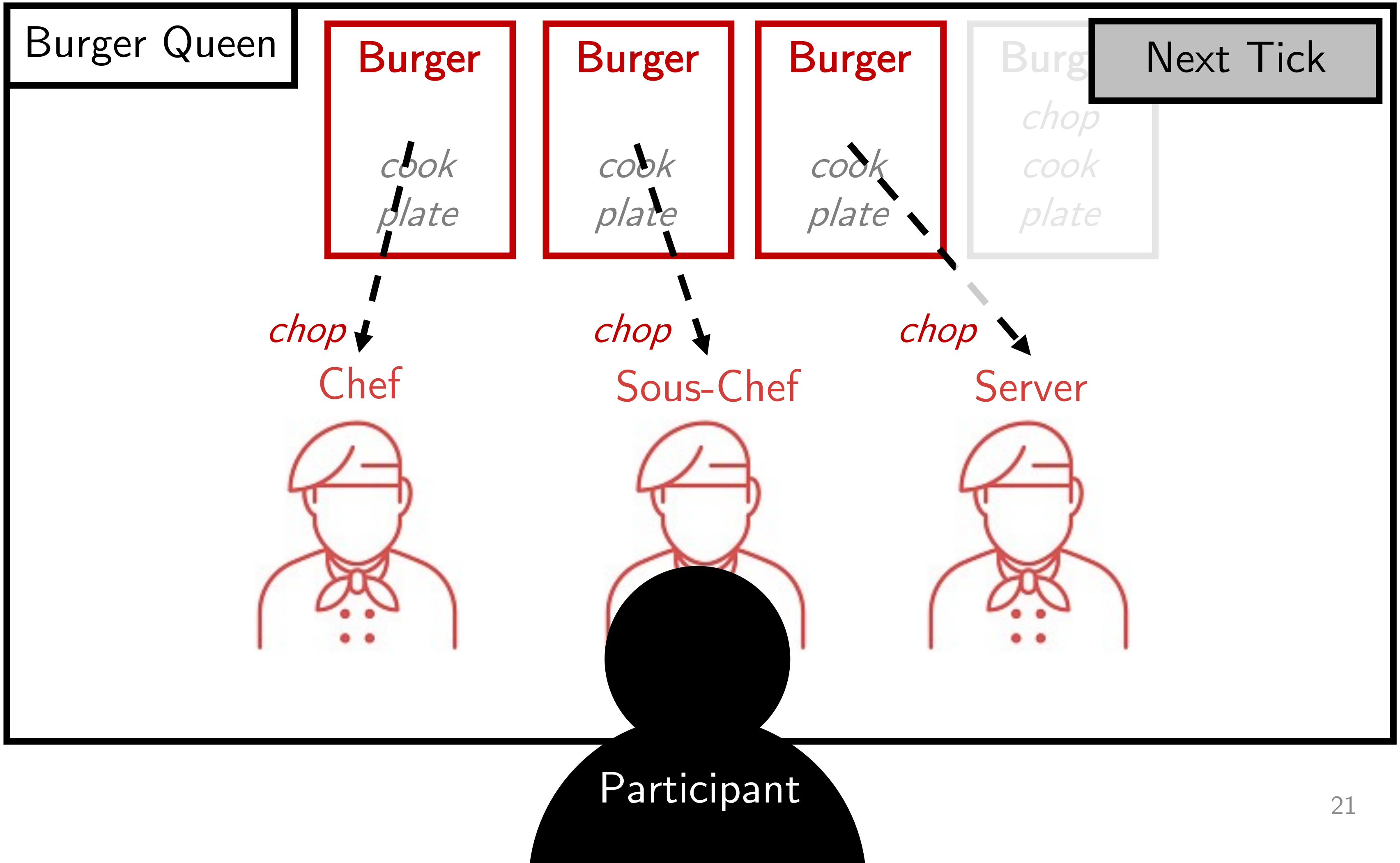
Queueing Game

Reward: 0
Tick #1/50

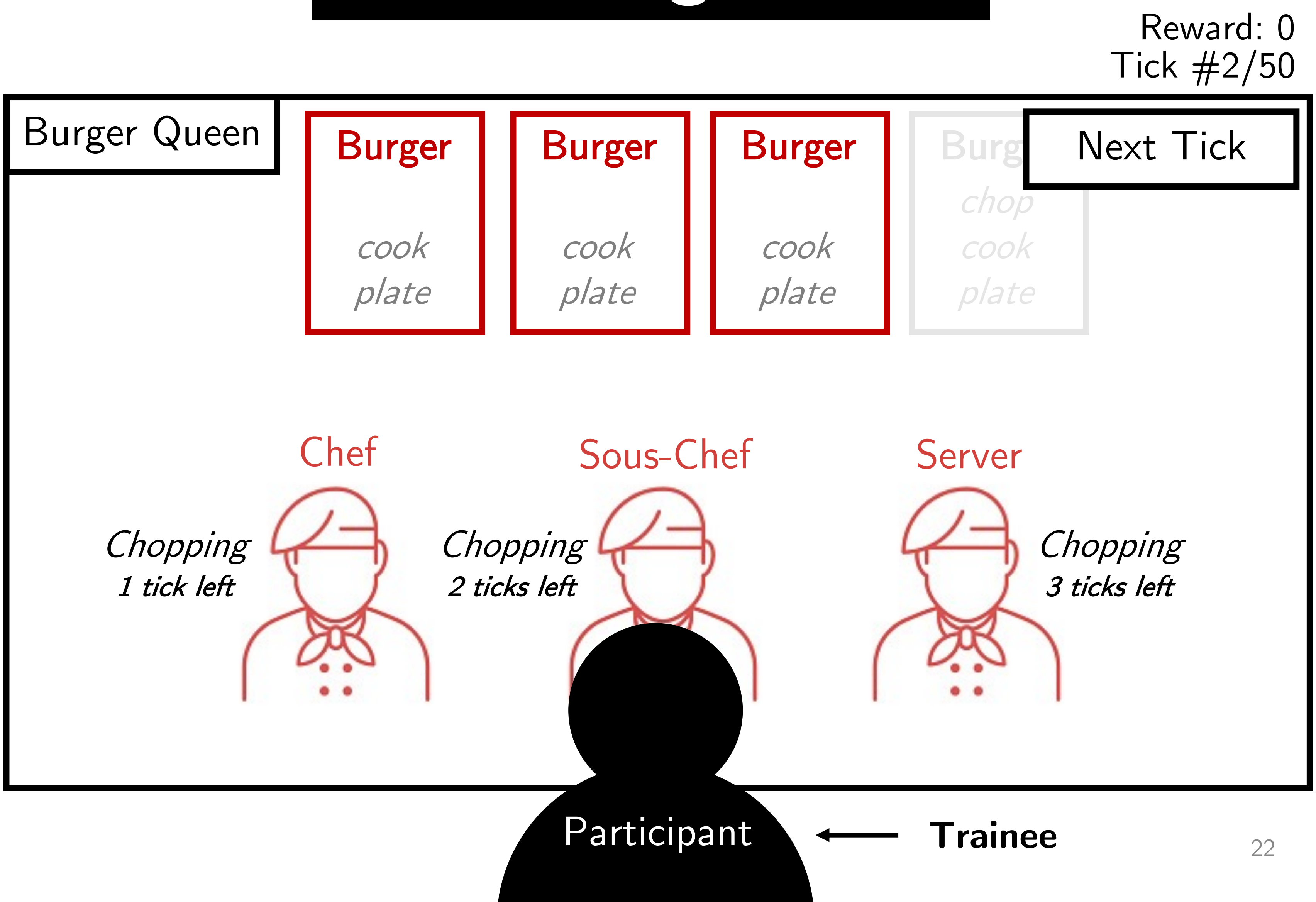


Queueing Game

Reward: 0
Tick #1/50



Queueing Game

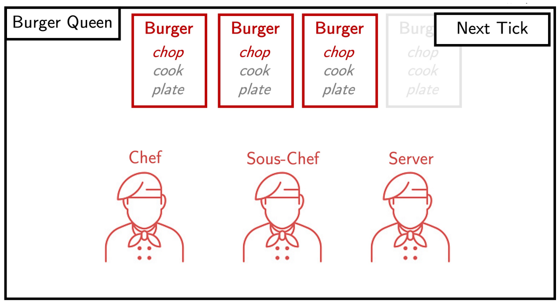


User Study Design

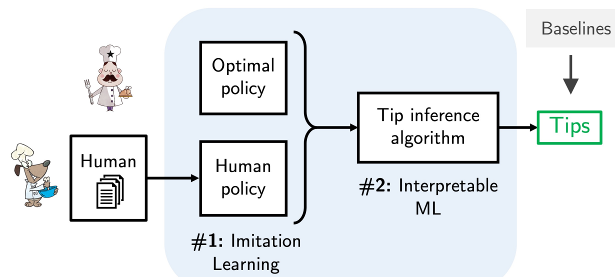
N = 1400

Phase I

N = 200

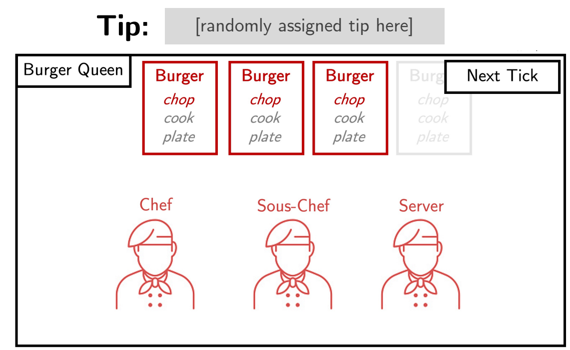


Gather trace data



Tip inference

Phase II



Tip evaluation

Environment

Normal

Disrupted

in the middle

Pre-registered at https://aspredicted.org/blind.php?x=8ye5cb

Design Disrupted Configuration

x 4 within 50 ticks

+ Our Tip

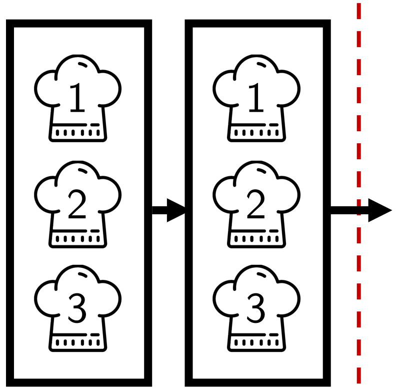
"Chef should never plate"

Round 1 | Round 2

Chef

Sous-Chef

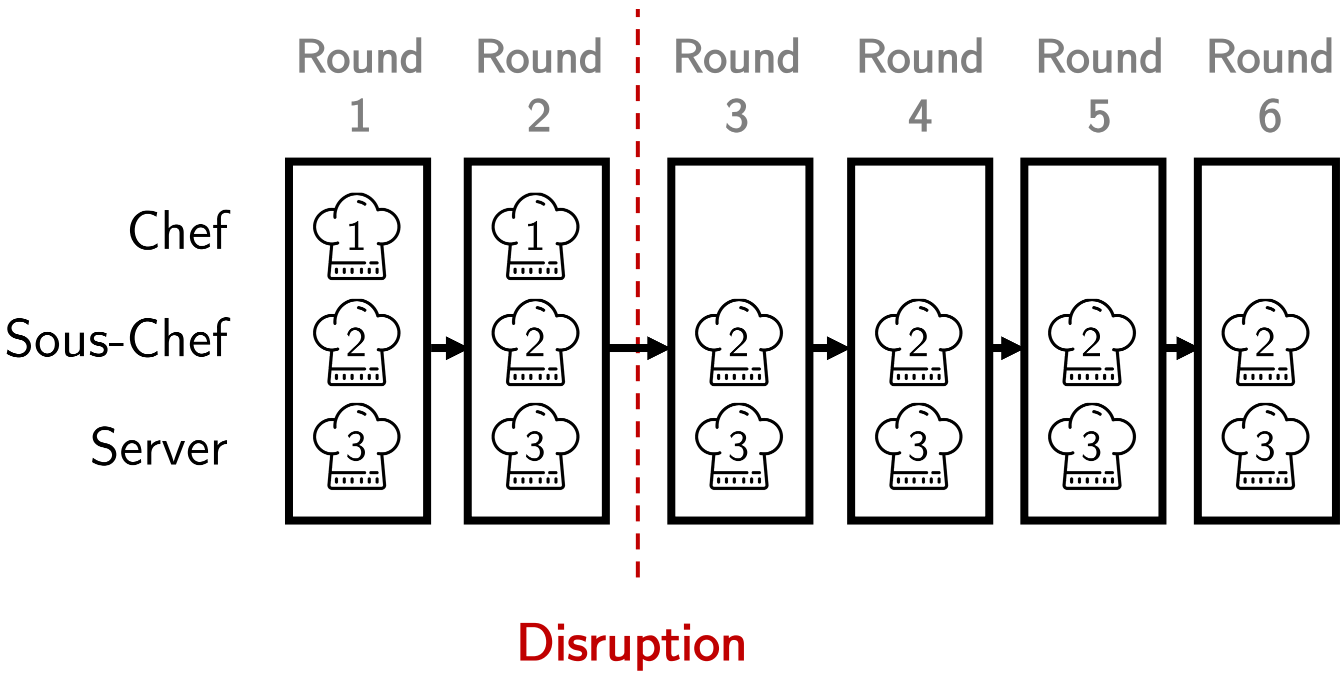
Server



Unfortunately, the Chef is on vacation during this round.
(Who travels these days...)

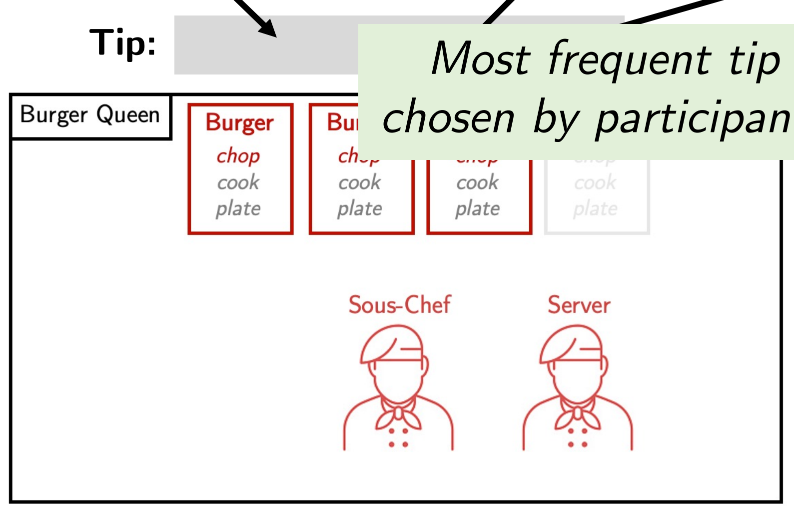
Disruption

Chopping:	Fast	Average	Slow
Cooking:	Fast	Average	Slow
Plating:	Slow	Average	Fast
	Chef	Sous-Chef	Server



Phase I Inferred Tips

Control	Algorithm	Human	Baseline
- No tip -	Server should cook twice	Server should cook once	Sous-Chef should plate twice



Most frequent tip chosen by participants

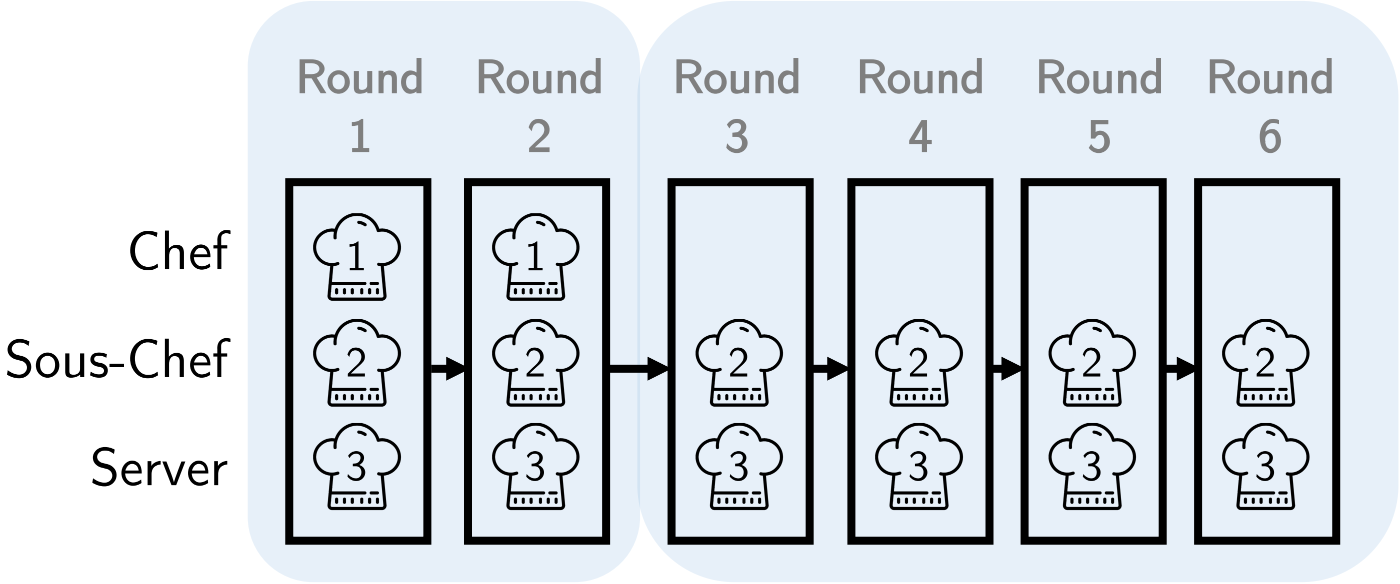
Most frequent s-a deviation b/w optimal and trainee policies

Amazon Mechanical Turk, N = 172

mean age 36.4, 62% female

Algorithm vs Human

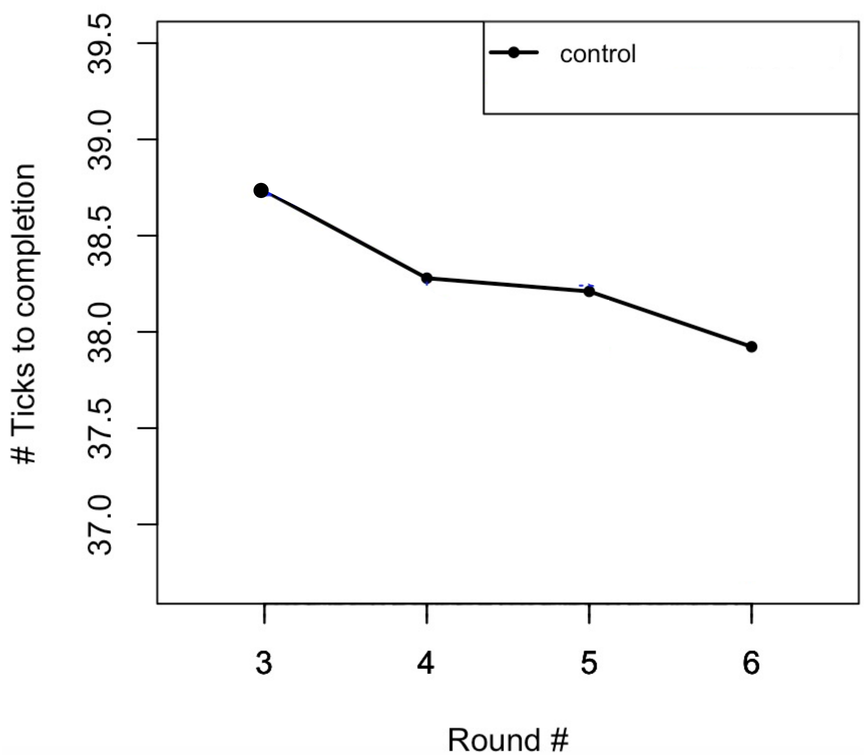
Algorithm	Human
Server should cook twice	Server should cook once



"Server shouldn't cook"

Results People Improve Over Time

Ticks to completion

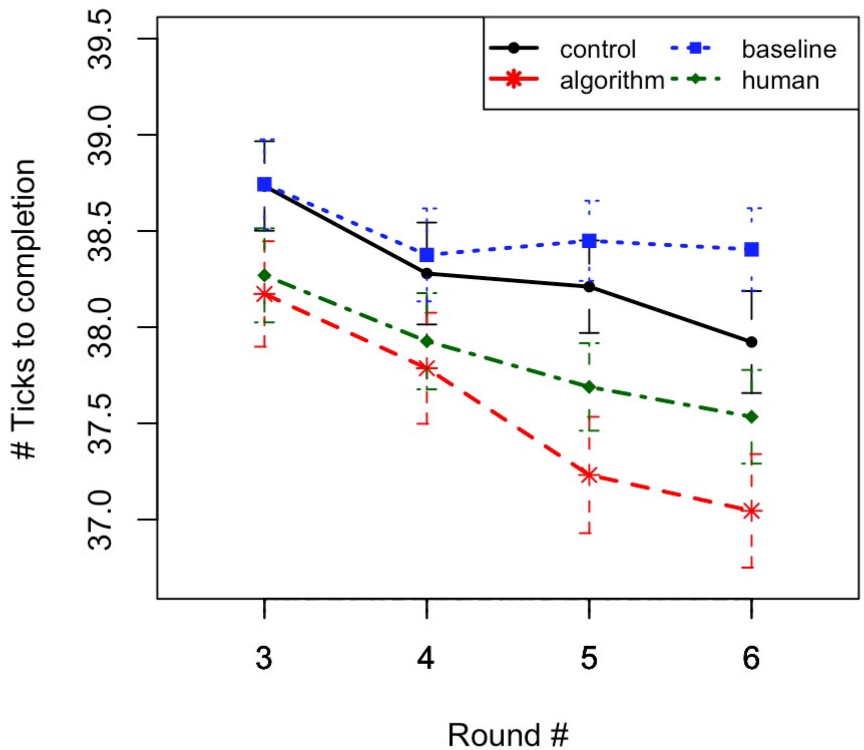


Amazon Mechanical Turk, N = 1,011

mean age 34.9, 60% female

Results Our Tip Improves Performance

Ticks to completion



One-sided T-Tests

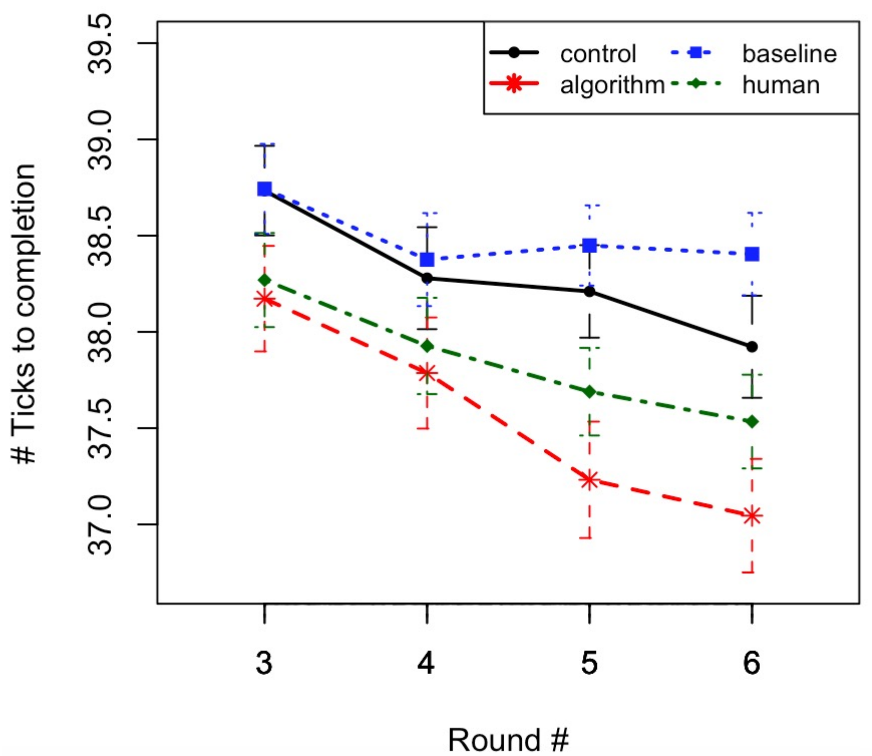
Algorithm *beats* Control ($p = 0.000008$)
Algorithm *beats* Human ($p = 0.006$)
Algorithm *beats* Baseline ($p < 1e\text{-}12$)

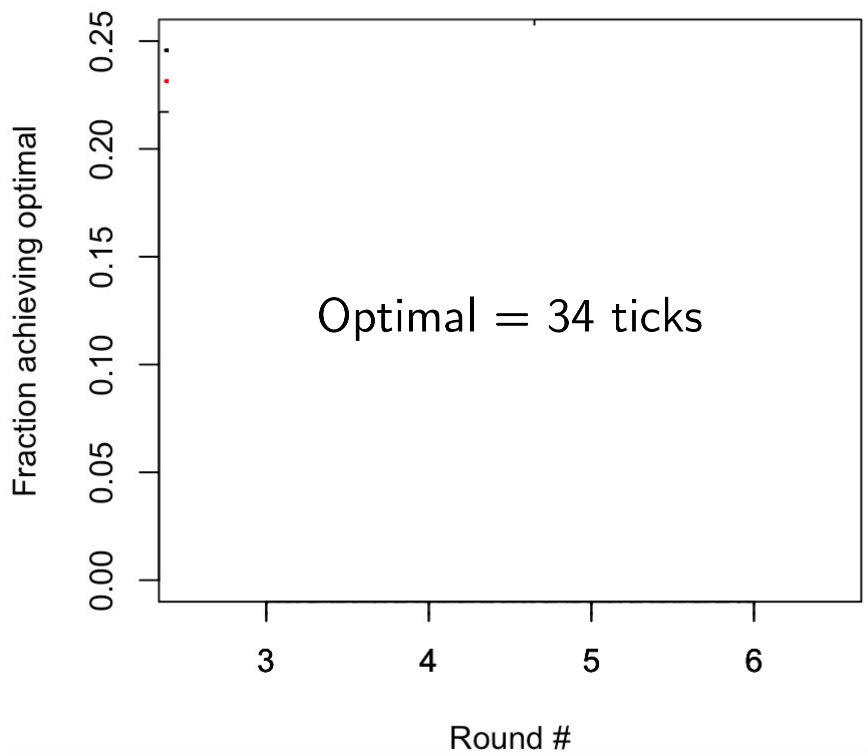
Amazon Mechanical Turk, N = 1,011
mean age 34.9, 60% female

Results

Ticks to completion

Fraction achieving optimal



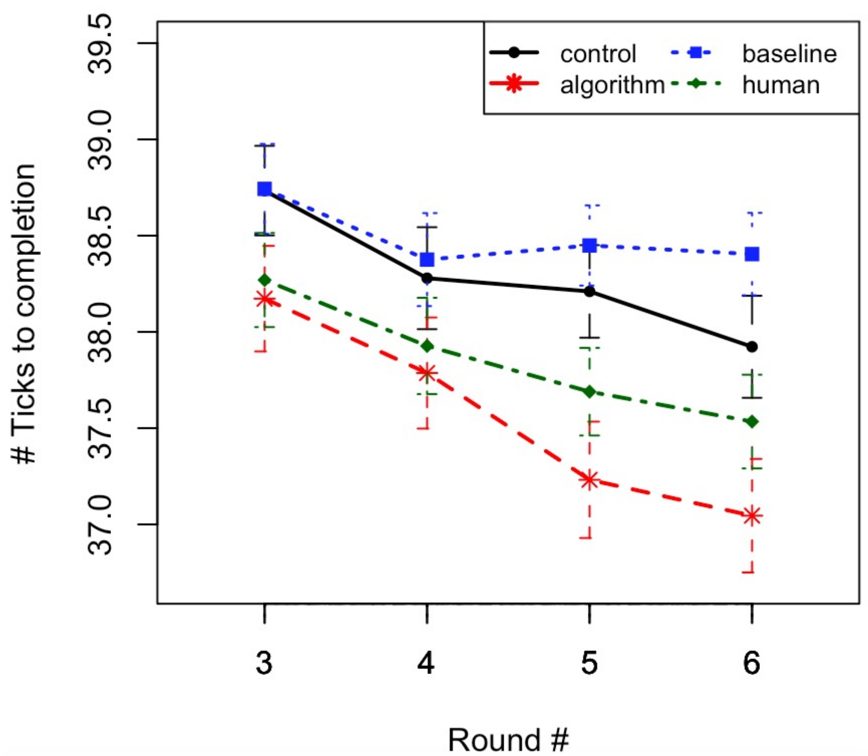


Amazon Mechanical Turk, N = 1,011

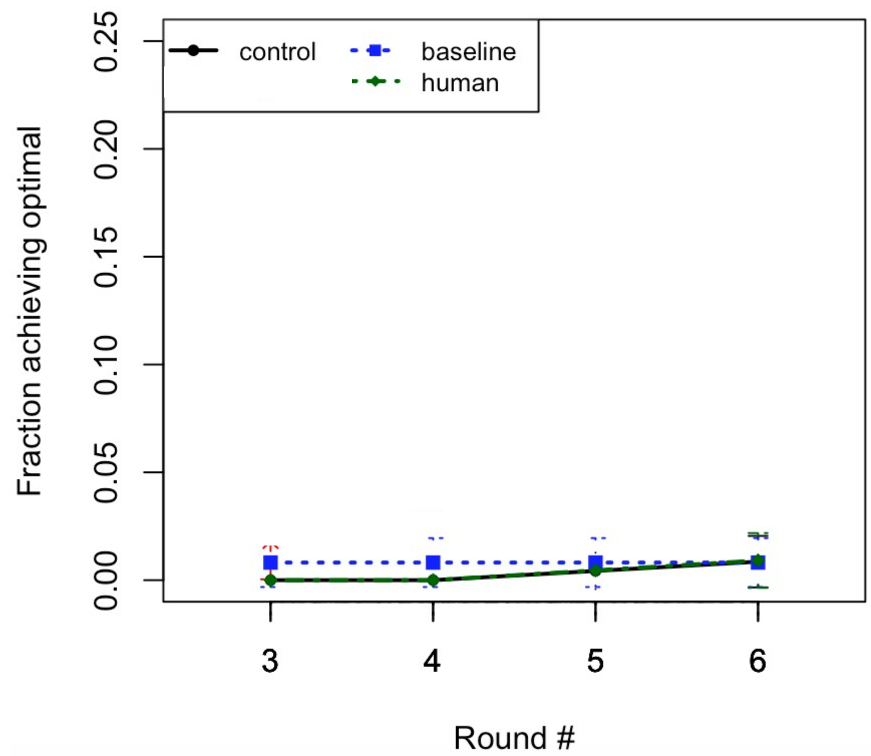
mean age 34.9, 60% female

Results Difficult to Reach Optimal

Ticks to completion



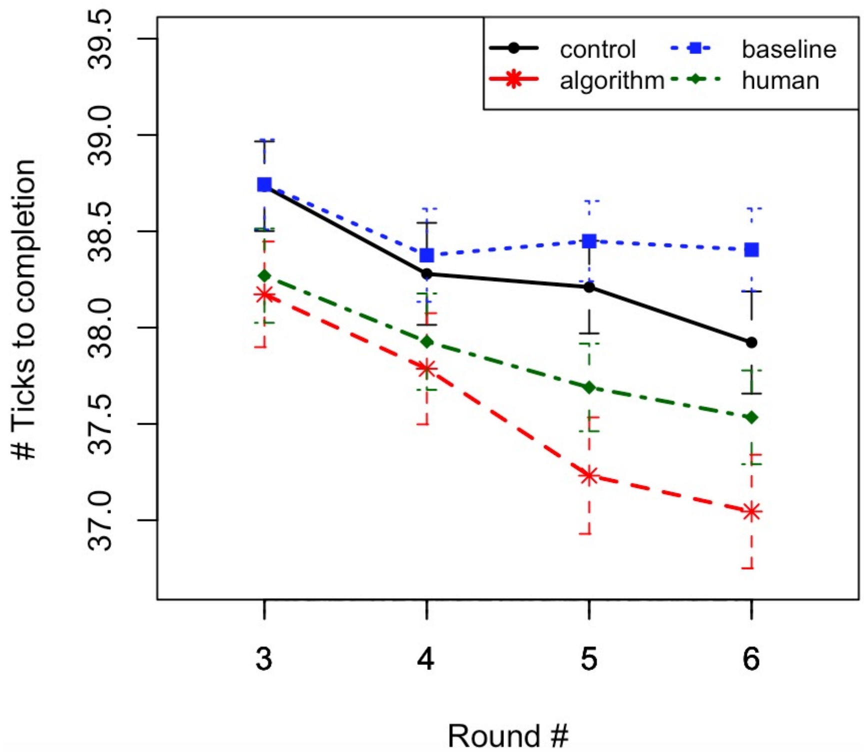
Fraction achieving optimal



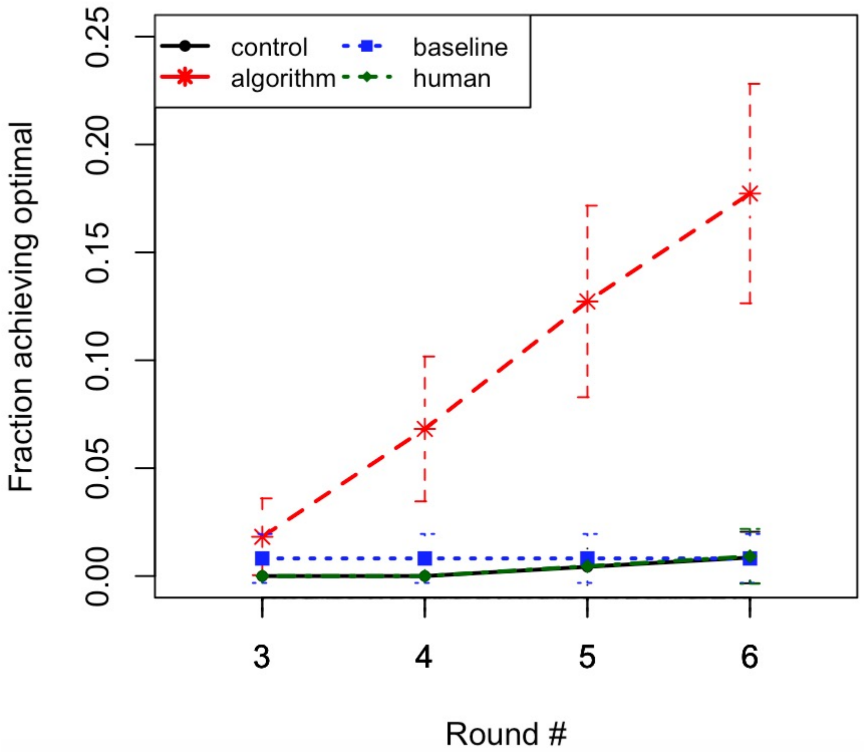
Amazon Mechanical Turk, N = 1,011
mean age 34.9, 60% female

Results Our Tip Helps Reach Optimal

Ticks to completion



Fraction achieving optimal

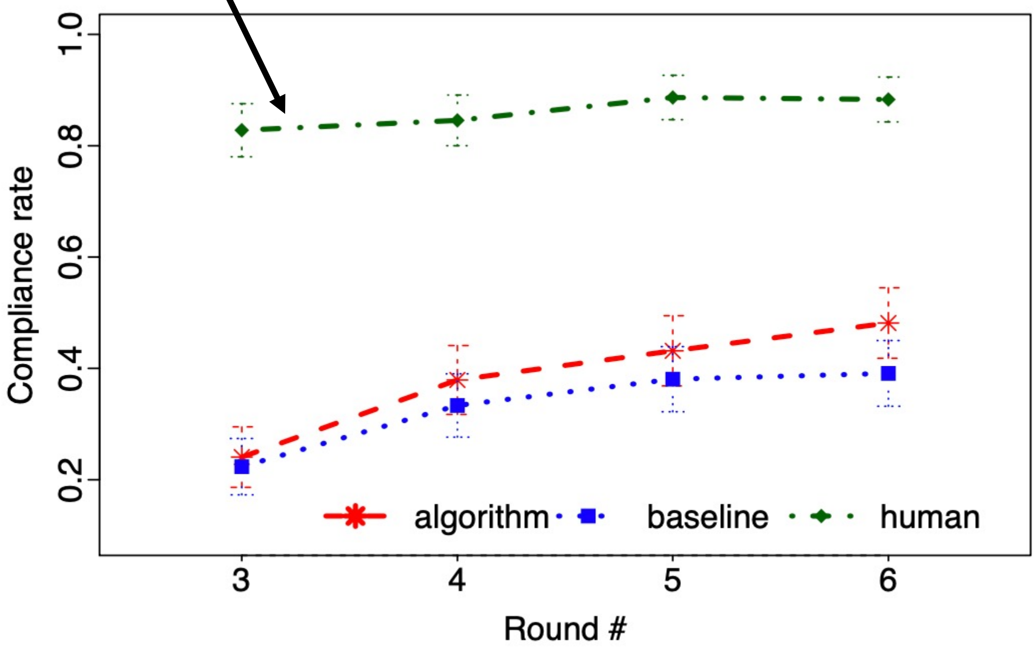


Amazon Mechanical Turk, N = 1,011

mean age 34.9, 60% female

Results Compliance

Participants comply more with intuitive `Human` tip



26% Positive, 17% Negative

"I felt that tip was **valid**."

R_1rvkYTwgAjD0z4z

"It helped because she could cook one burger but **any more than that and your ticks would be too high**."

R_d6YSuigdikyaNdT

"I thought it was **smart** and I used it exclusively."

R_beijQ8guDyExa5r

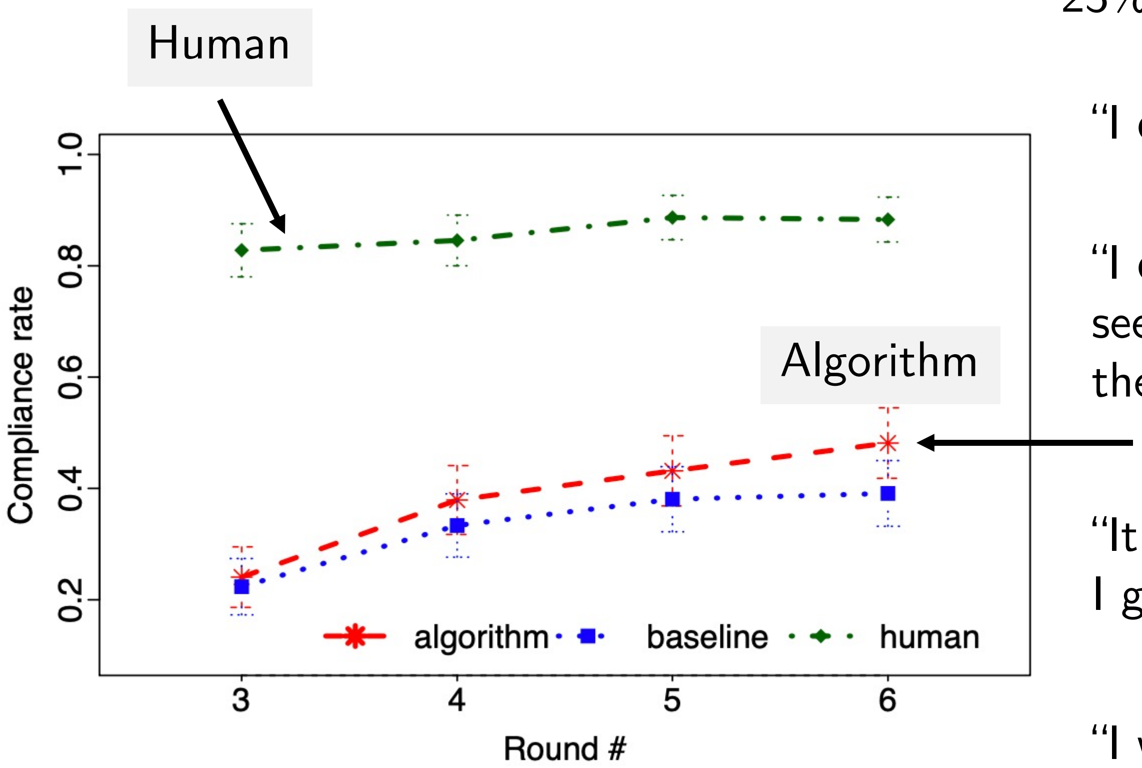
"It was **accurate**, and I implemented it."

R_1pA8wDYgWc9hblt

Amazon Mechanical Turk, N = 1,011

mean age 34.9, 60% female

Results Against Counterintuitive Tips



23% Positive, **33% Negative**

"I didn't think it was right."
R_3EgrcrQouPcb1fS

"I didn't follow it because it seemed counter intuitive since they're slow."
R_10HkPUkR6o0qDFT

"It didn't make sense and in fact I got worse trying to use it,"
R_2YD5x6BL7mhCYEP

"I wasn't sure how to use it."
R_2s0UA1omAifrFgx

Amazon Mechanical Turk, N = 1,011
mean age 34.9, 60% female

Results Learning Beyond Tips

Structure of Optimal Policy

	Chop	Cook	Plate	
Sous-Chef 2	3	2	2	times
Server 3	1	2	2	times

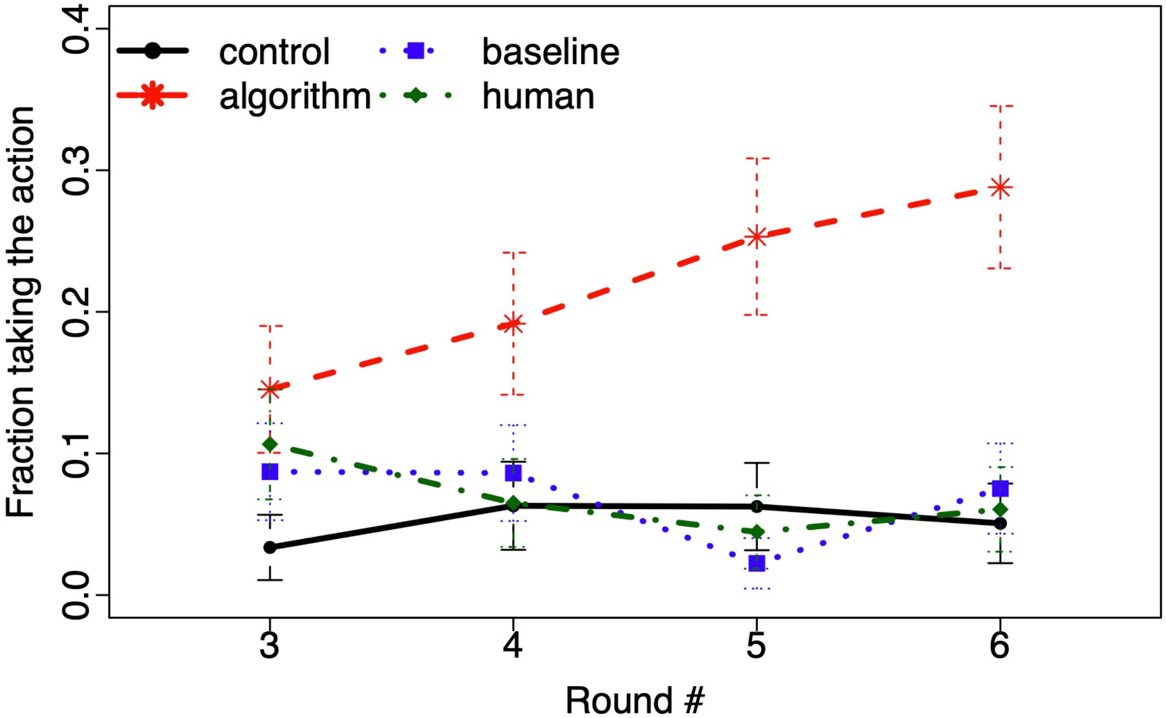
Algorithm

Baseline

Results Learning Beyond Tips

Our tip effectively led people to the states they can learn other optimal strategies

Sous-Chef chops 3 times

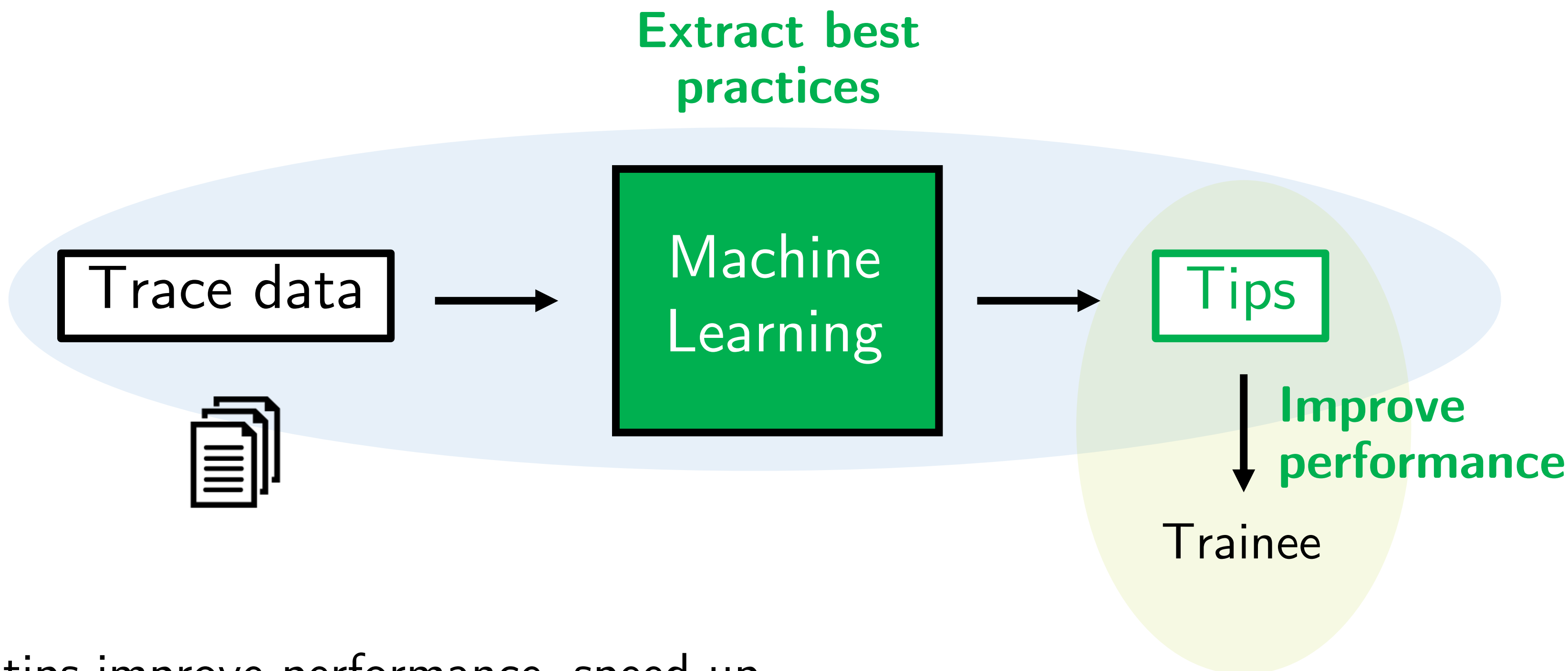


Amazon Mechanical Turk, N = 1,011

mean age 34.9, 60% female

Summary

ML to automatically extract simple tips that help people improve in a dynamic way



Our tips improve performance, speed up learning, help adapt to disruption, and uncover other optimal strategies

Hamsa Bastani, Osbert Bastani, **Park Sinchaisri** (parksinchaisri@berkeley.edu/ wichinpong.com)

Potential Issues

- Compliance to tips, "algorithm aversion"
  (e.g., Dawes et al 1989, Dietvorst et al 2015)
- Interpretability
- Learning curve