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

Learning is Costly

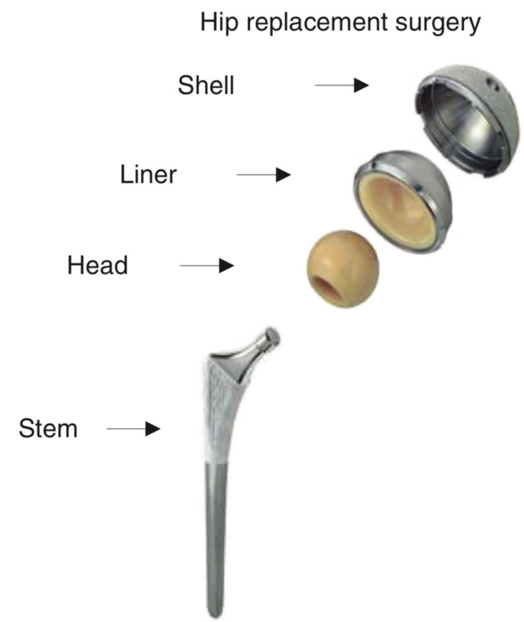
2+ years

to be fully productive

$1,286/worker

training expenses

- Training Magazine 2019



New device = **+32.4%**

surgery duration

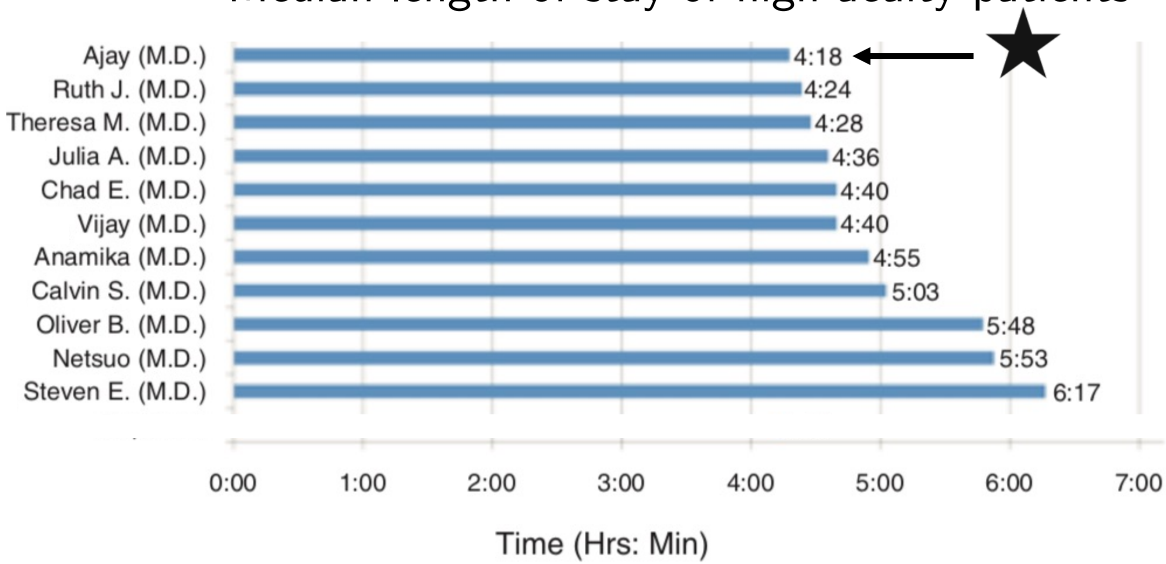
- Ramdas et al. 2018

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

Learning from Experts

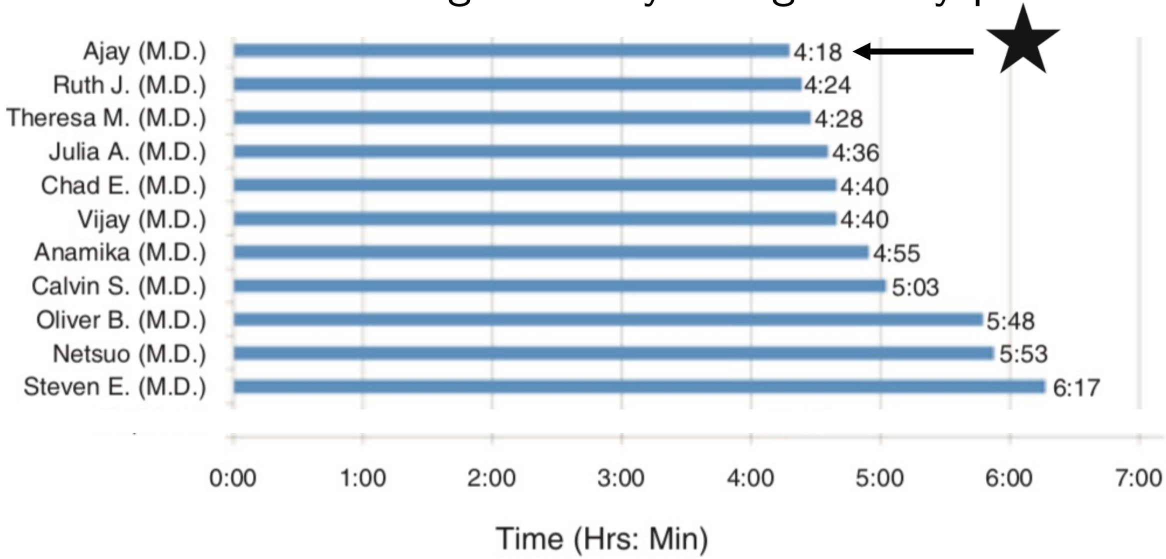
Learning from Experts

Median length of stay of high acuity patients



Learning from Experts

Median length of stay of high acuity patients



+10.9%

productivity

- Song et al. 2018

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

Learning from Experts

+10.9%

productivity

- Song et al. 2018

Learning from Experts

+10.9%

productivity

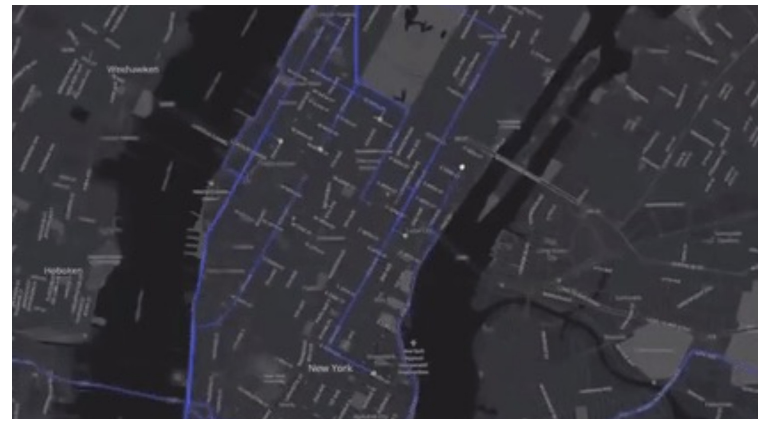
- Song et al. 2018

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

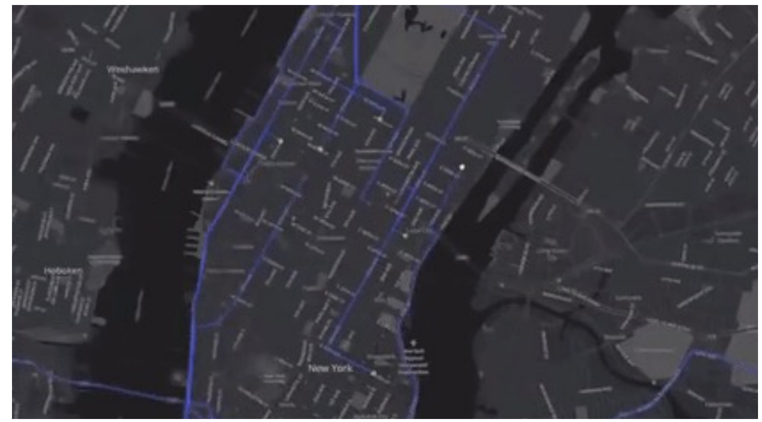


Trace Data is Everywhere

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

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

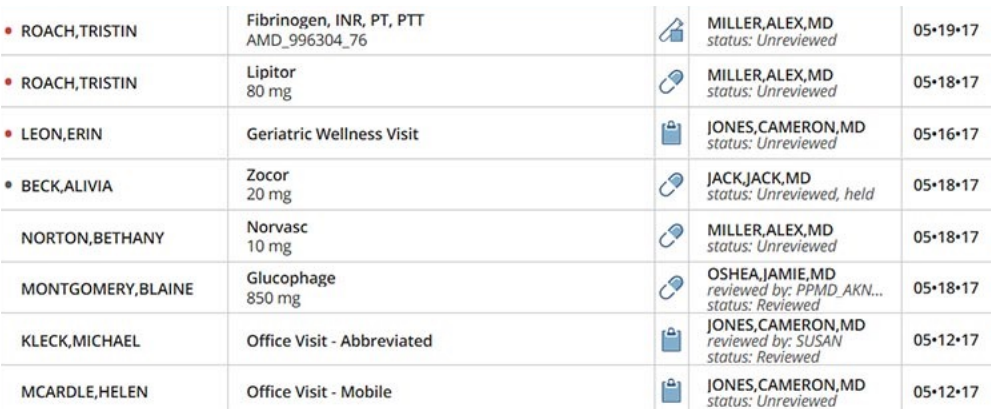
Uber Drivers

Trace data → Tips

Noisy, high-volume data
hard to extract insights

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

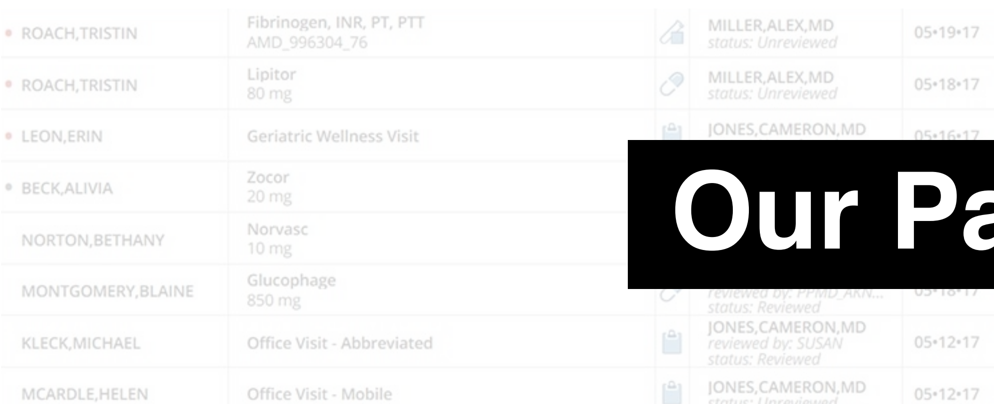
Noisy, high-volume data
hard to extract insights

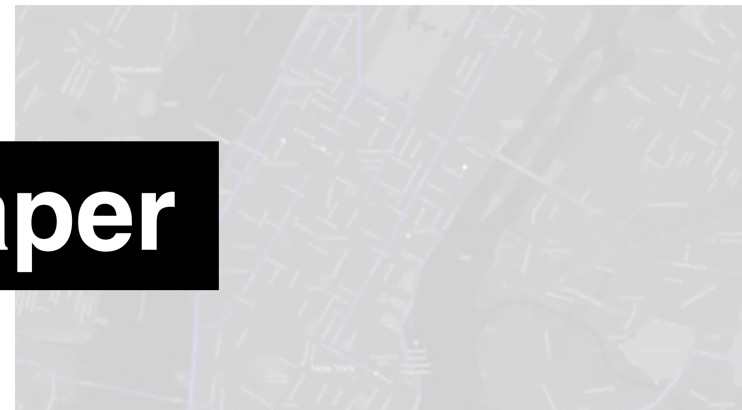
Key Q: can insights from ML
improve human decision-making?

Trace Data is Everywhere

Physicians

Uber Drivers





Our Paper

Trace data → Machine Learning → Tips

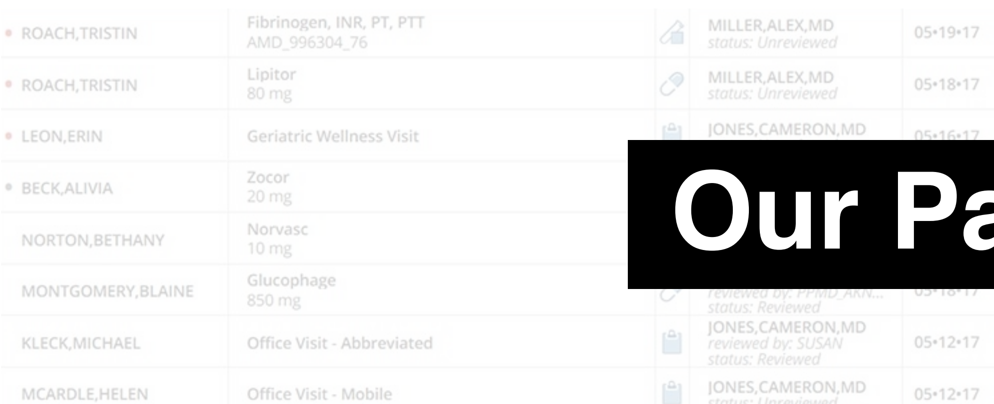
Extract best practices

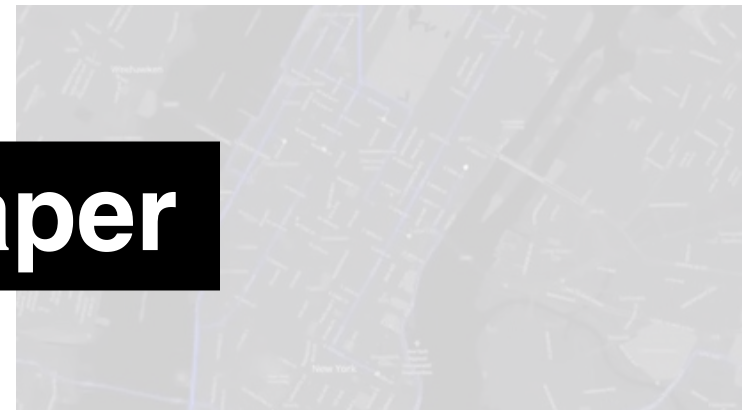
Mine simple tips

Trace Data is Everywhere

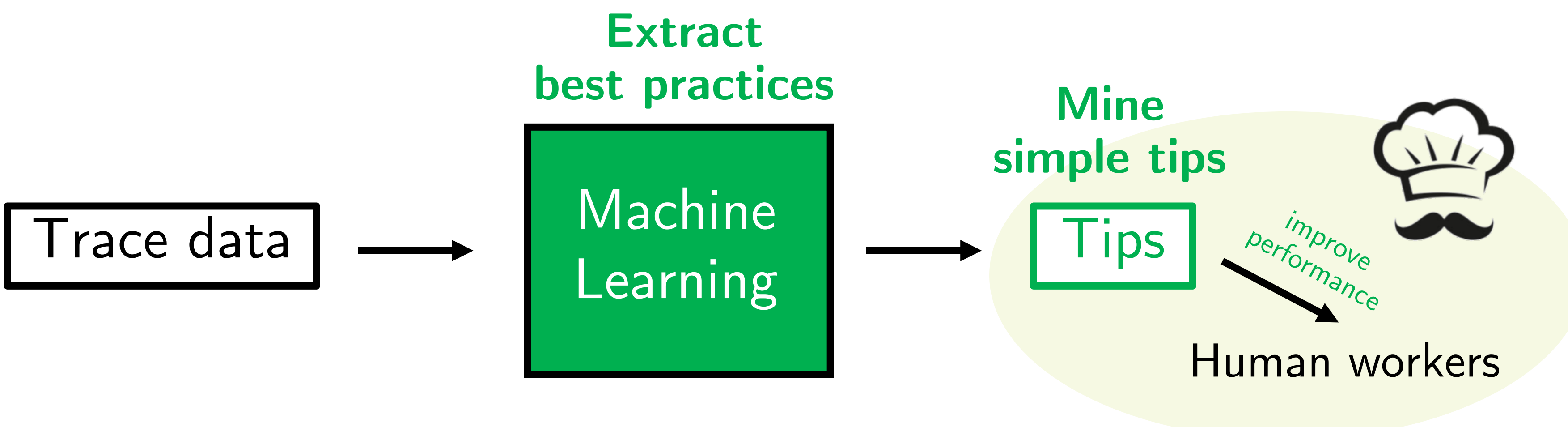
Physicians

Uber Drivers

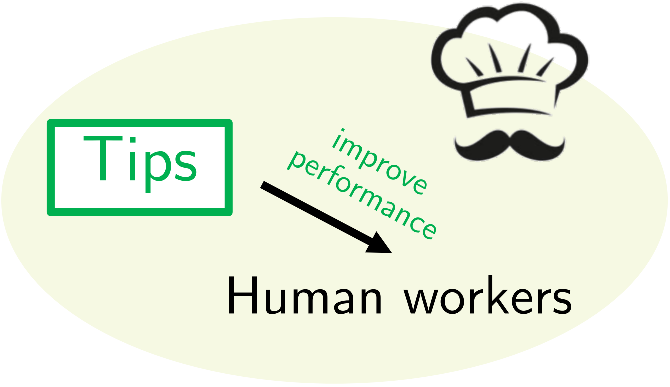




Our Paper



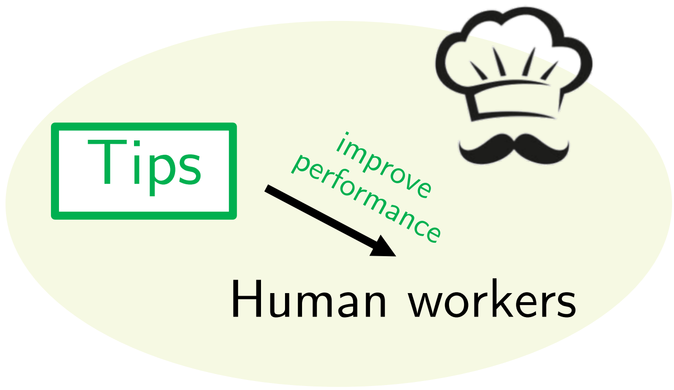
Potential Issues



Potential Issues

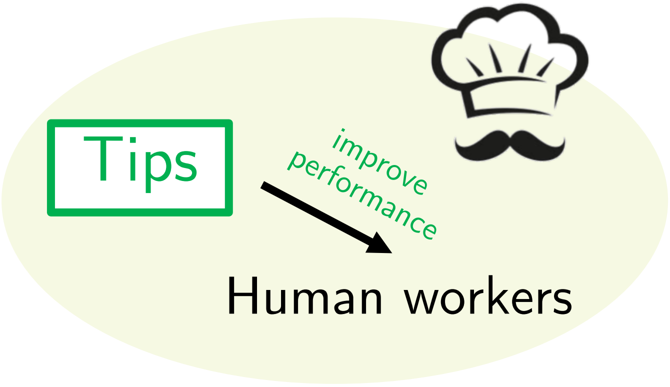
- Compliance to tips, "algorithm aversion"

  (e.g., Dawes et al 1989, Dietvorst et al 2015)



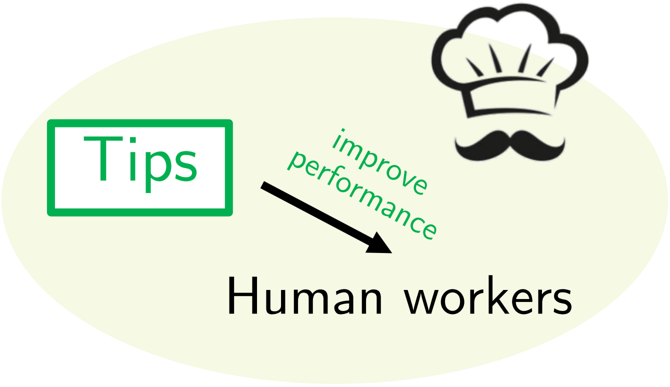
Potential Issues

- Compliance to tips, “algorithm aversion” (e.g., Dawes et al 1989, Dietvorst et al 2015)
- Interpretability

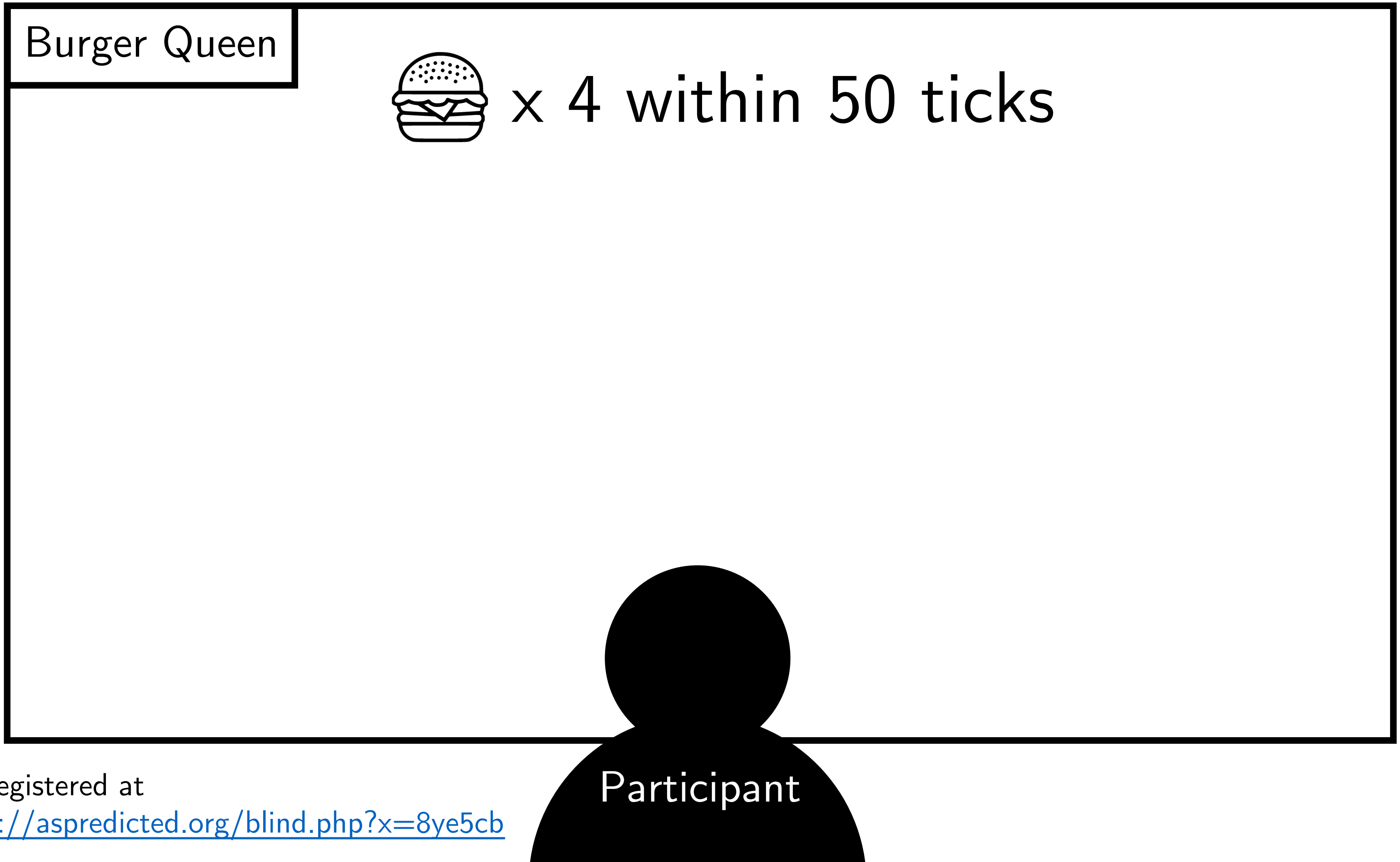


Potential Issues

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

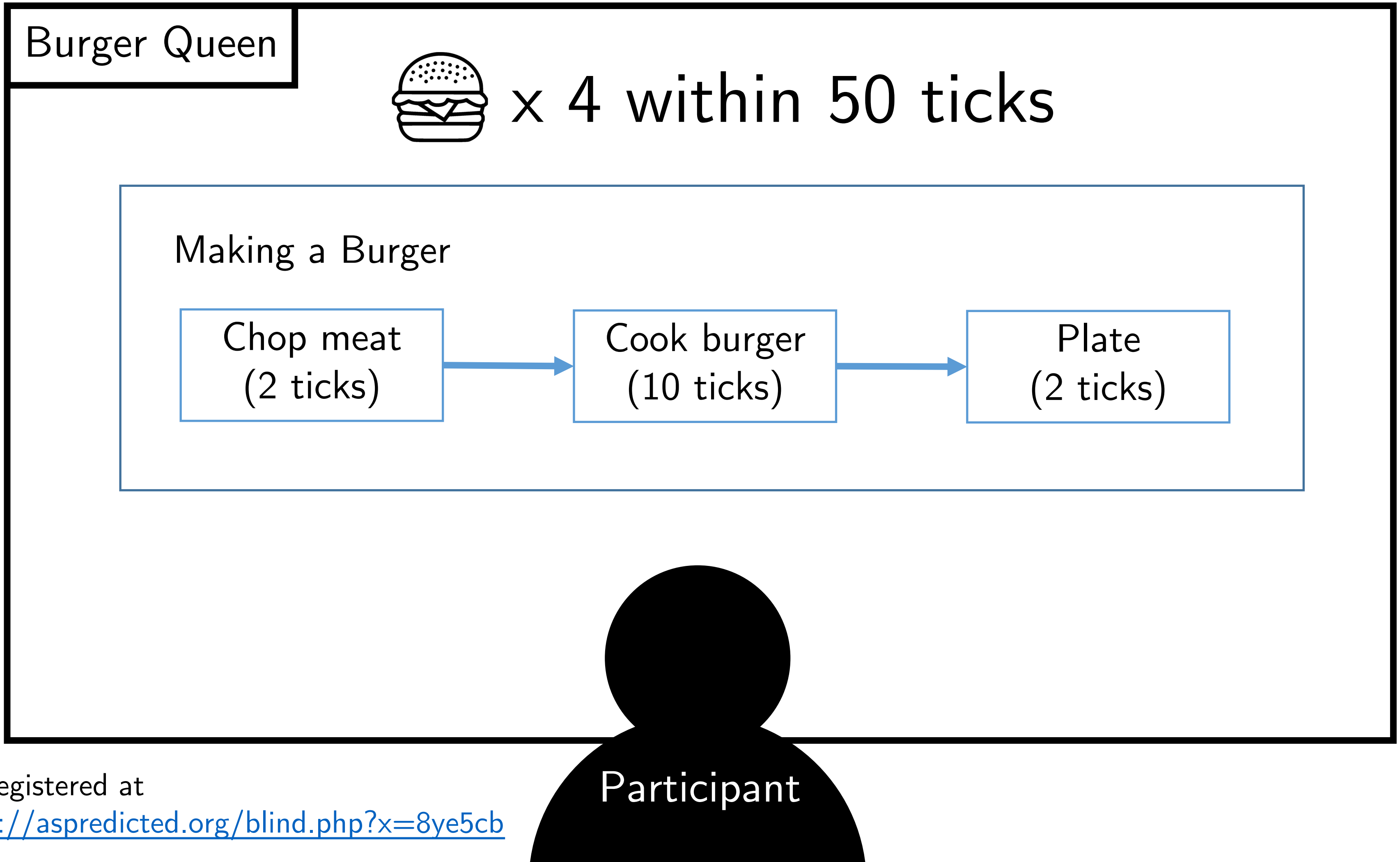


Queueing Game



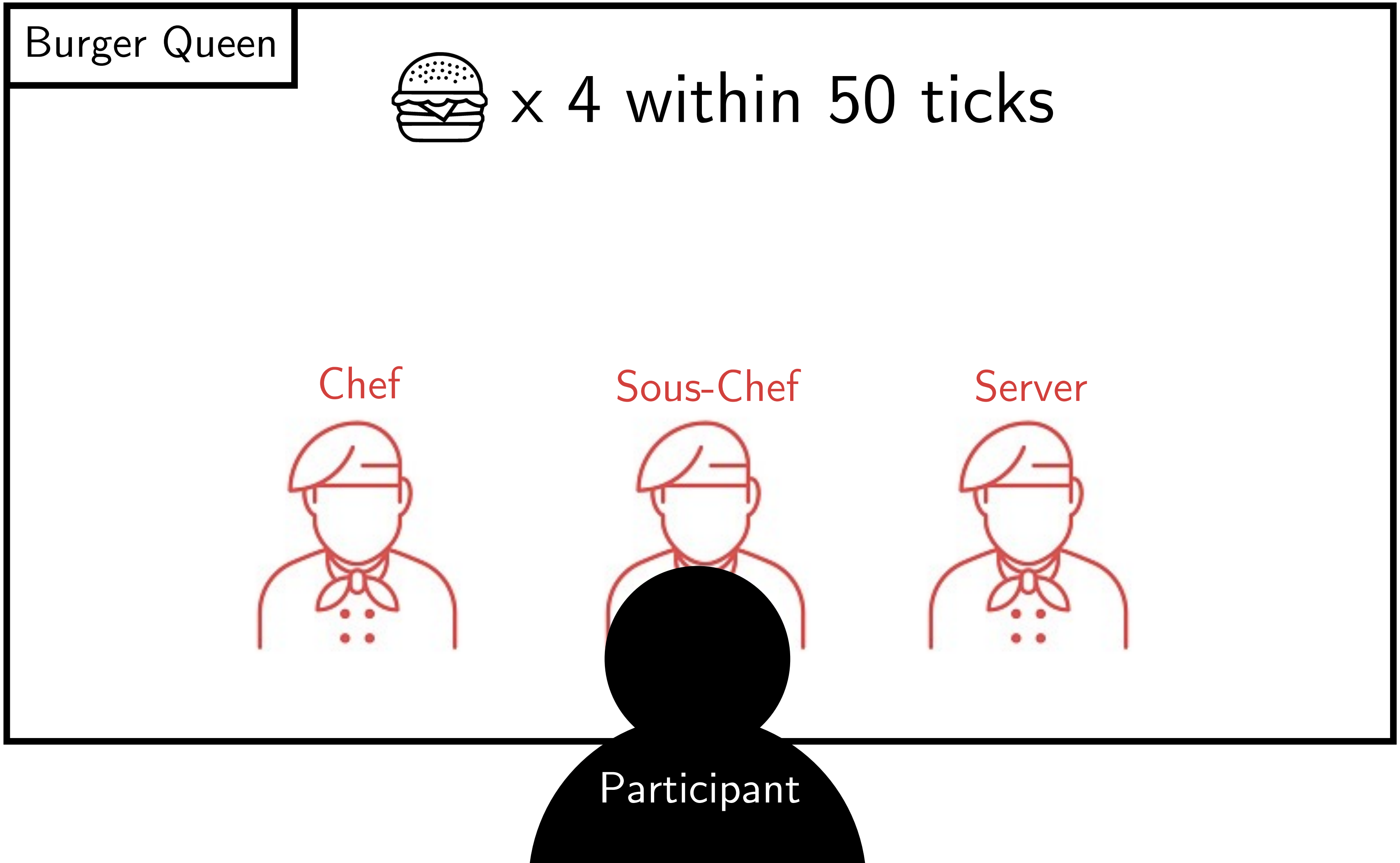
Pre-registered at
https://aspredicted.org/blind.php?x=8ye5cb

Queueing Game



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

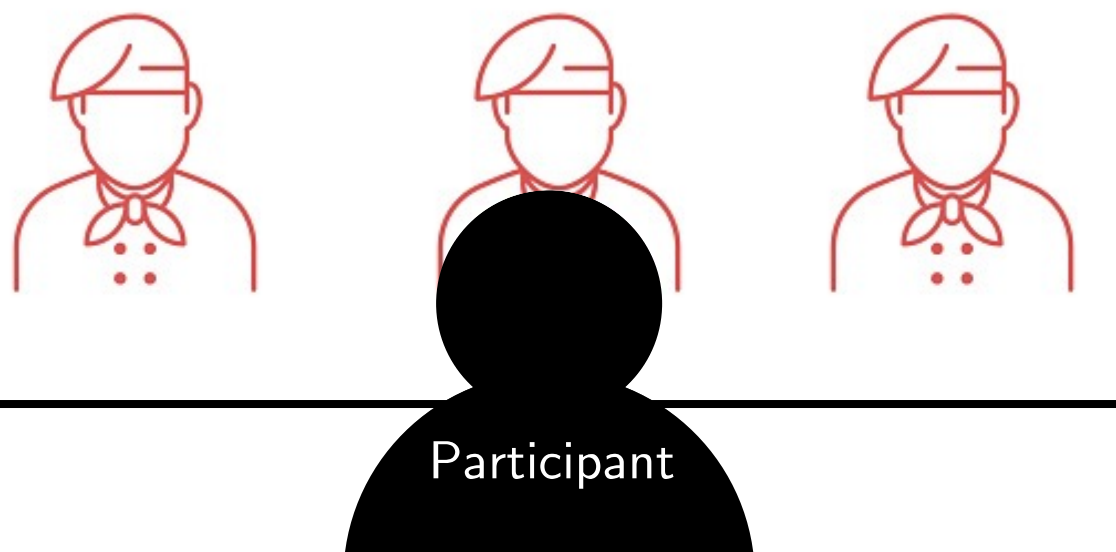
Queueing Game



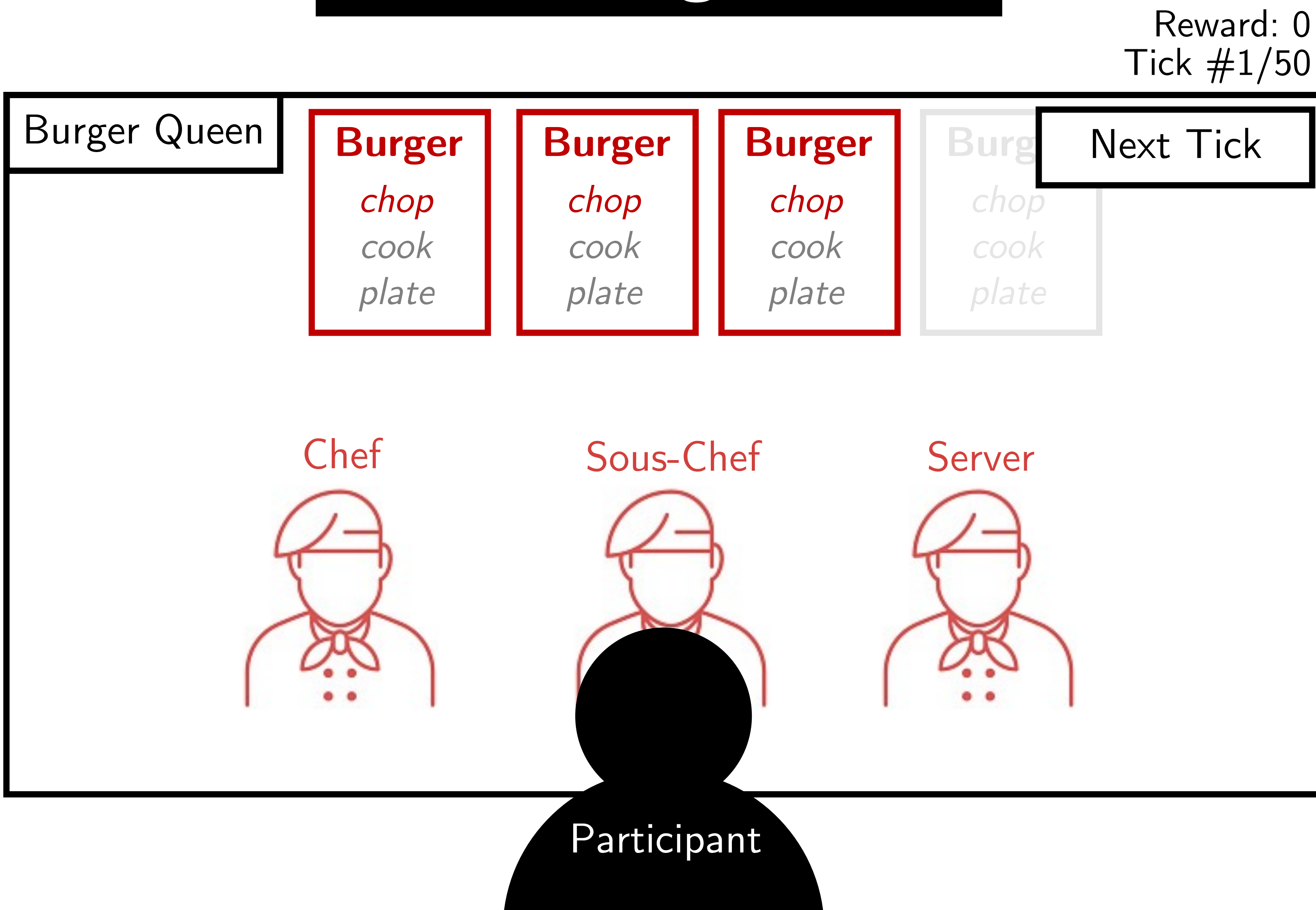
Queueing Game

Burger Queen

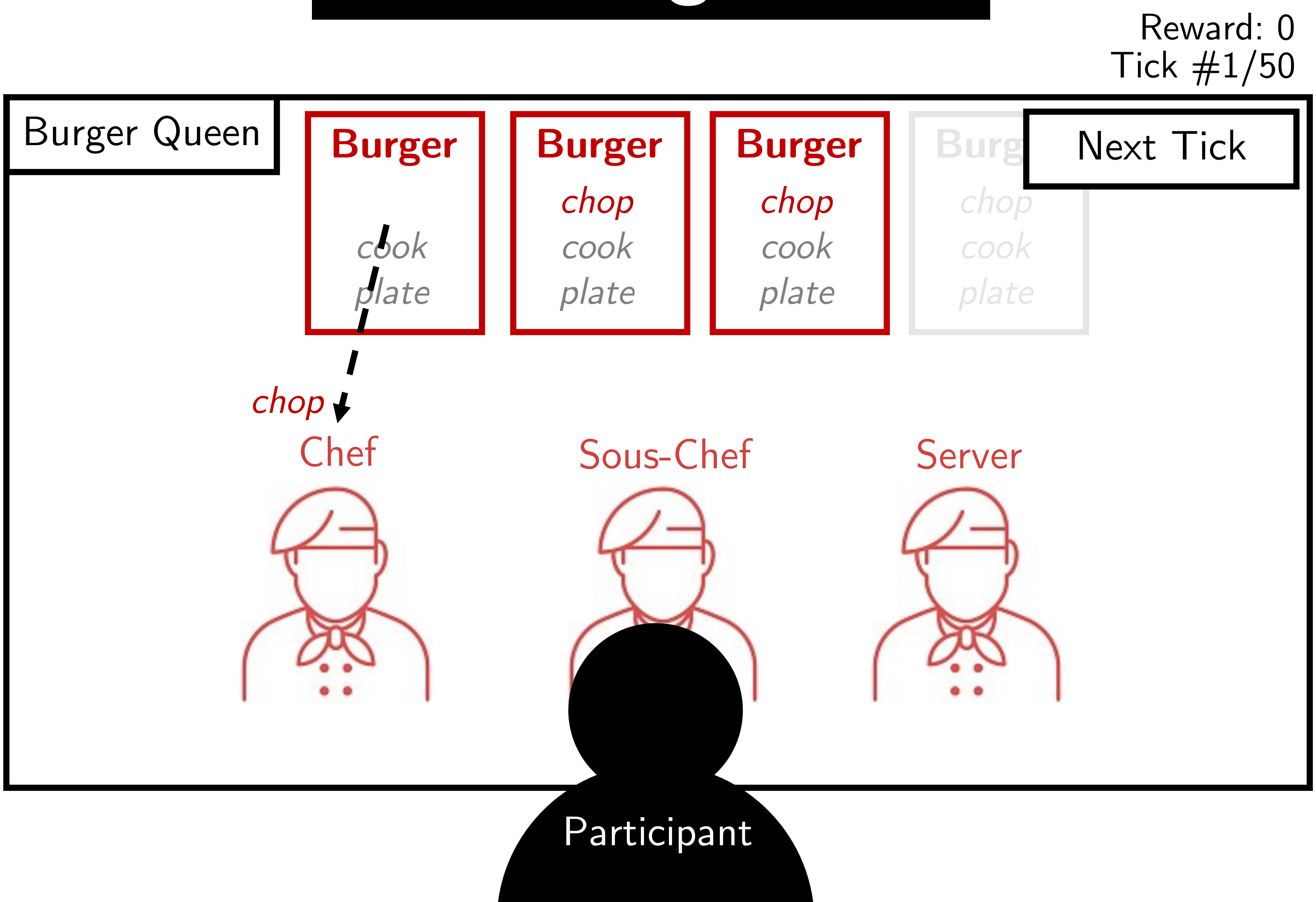
	Chef	Sous-Chef	Server
Chopping:	Fast	Average	Slow
Cooking:	Fast	Average	Slow
Plating:	Slow	Average	Fast



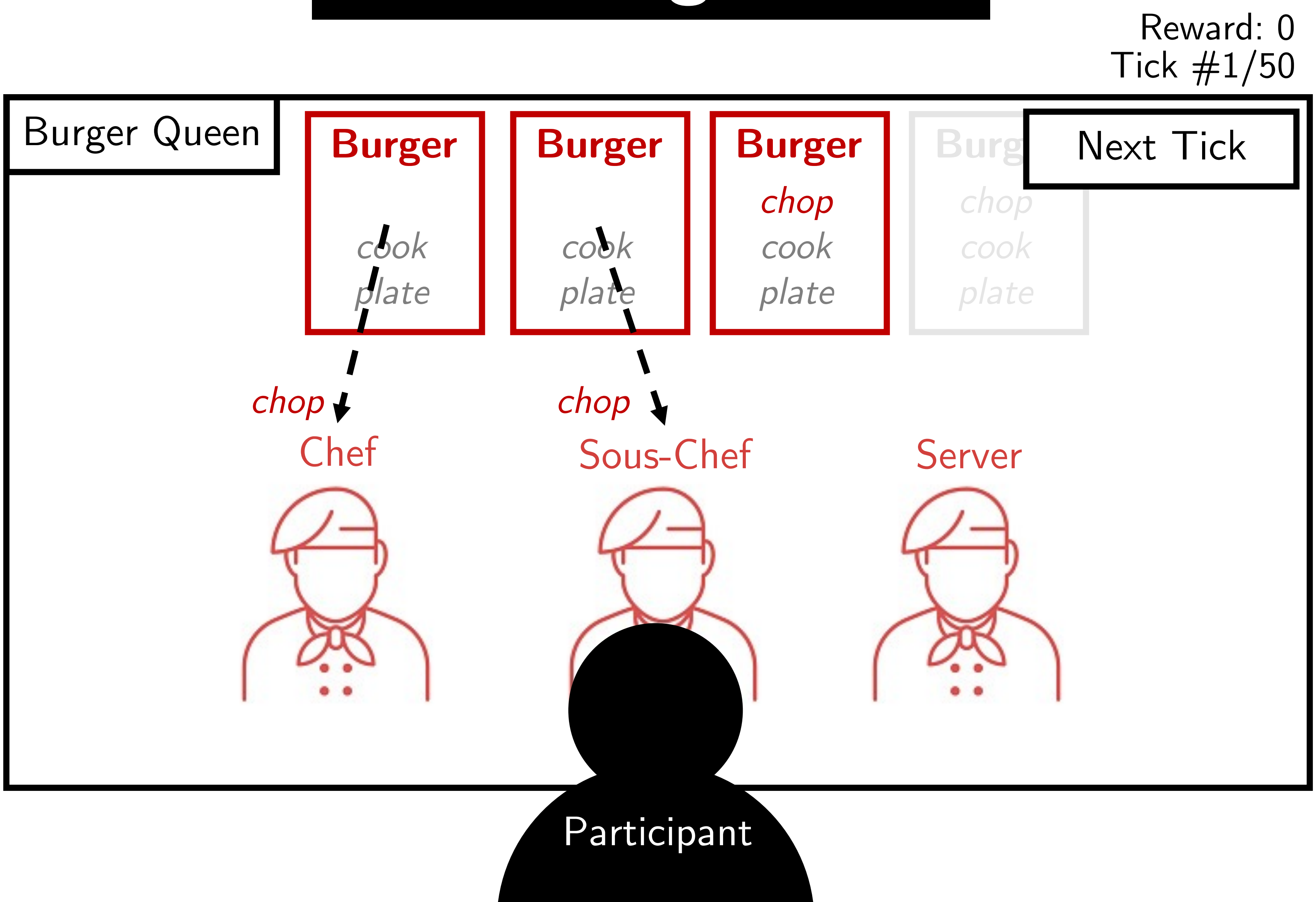
Queueing Game



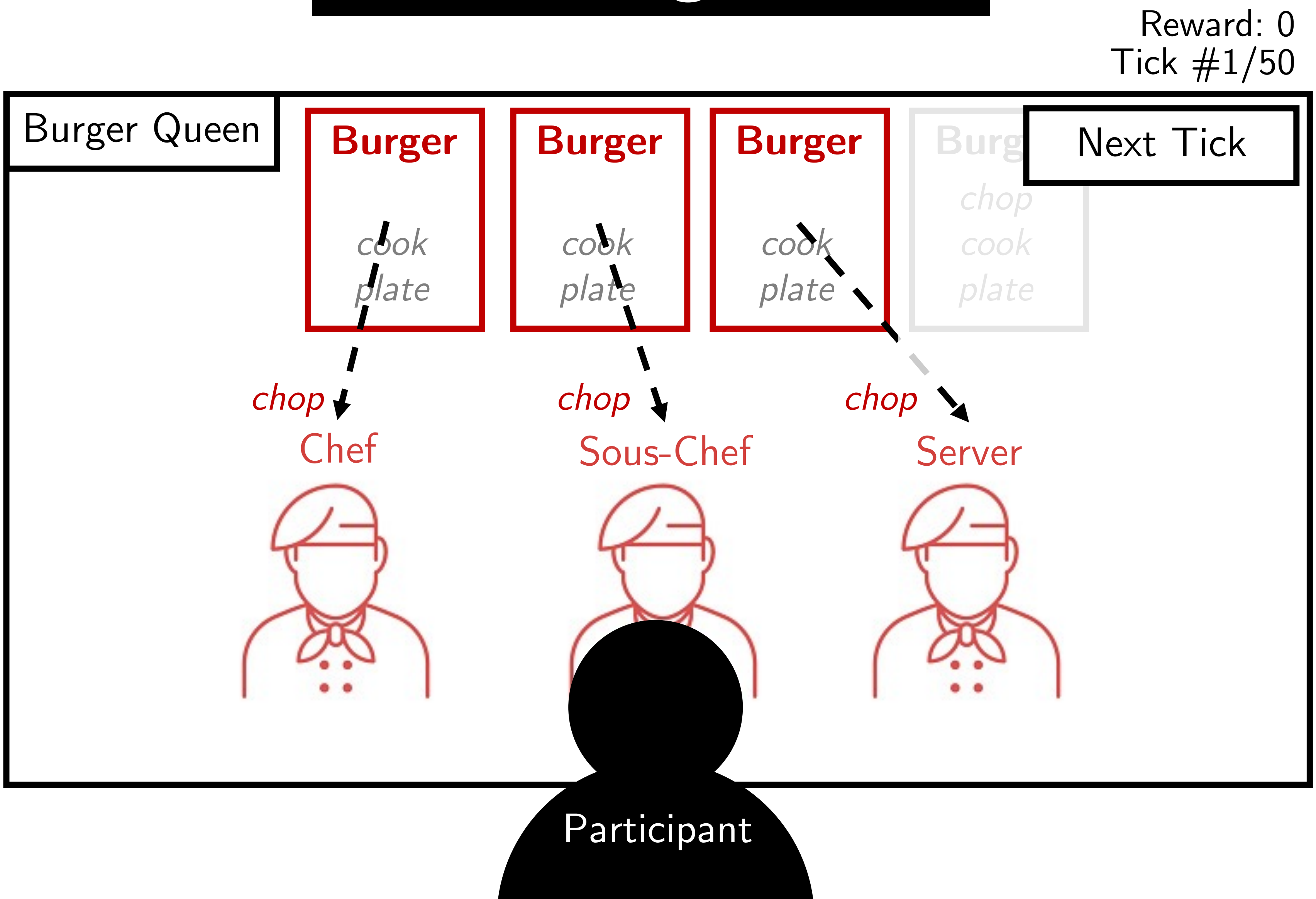
Queueing Game



Queueing Game

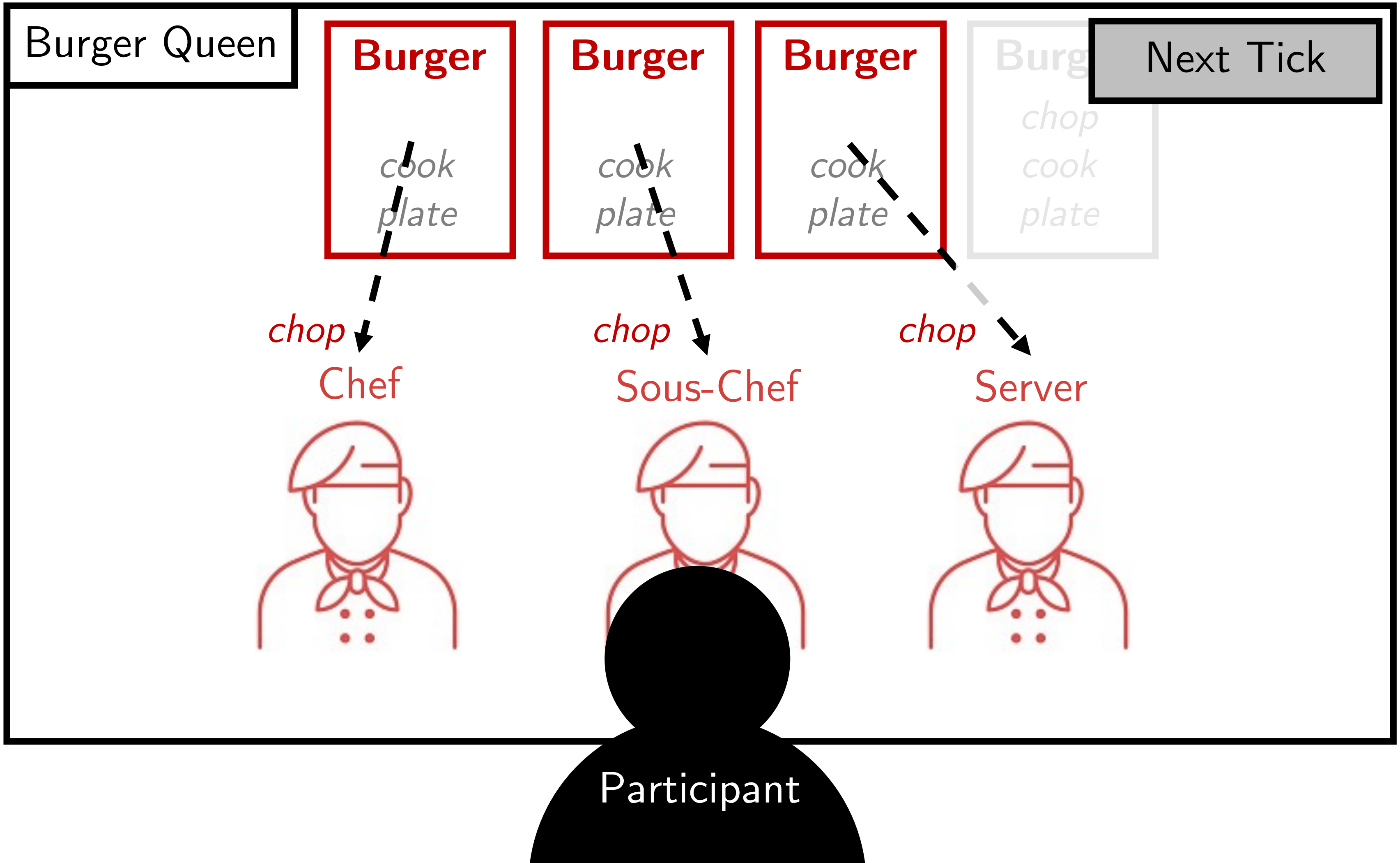


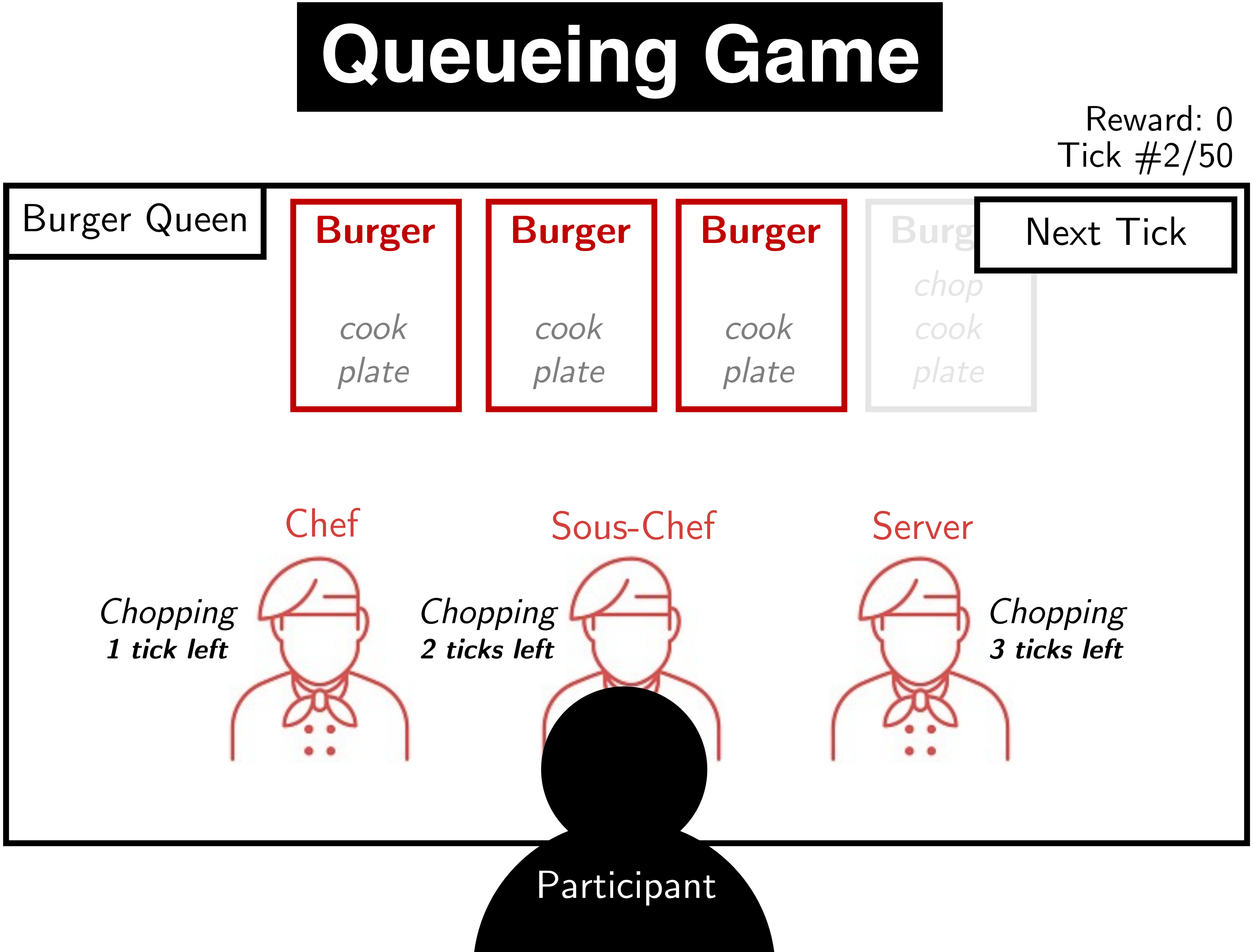
Queueing Game



Queueing Game

Reward: 0
Tick #1/50




Queueing Game
Reward: 0
Tick #2/50
Burger Queen
Burger
cook
plate
Burger
cook
plate
Burger
cook
plate
Burg
chop
cook
plate
Next Tick
Chef
Sous-Chef
Server
Chopping
1 tick left
Chopping
2 ticks left
Chopping
3 ticks left
Participant

Problem Formulation

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

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

Problem Formulation

MDP Formulation:

Optimal policy and *human* make sequences of decisions

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

Problem Formulation

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

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

π^* start s_1

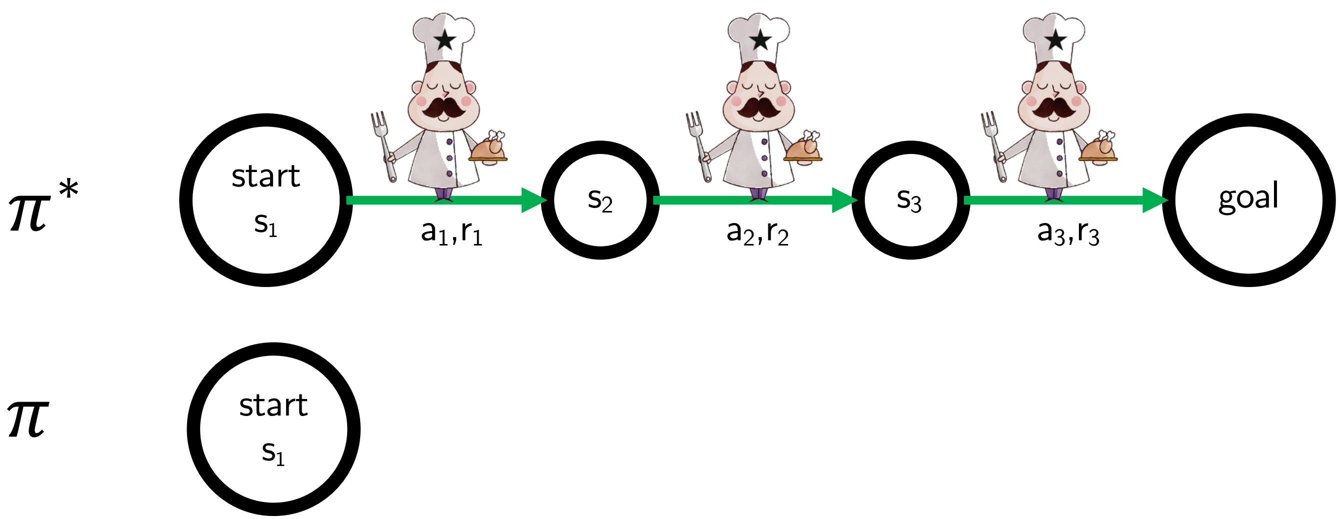
π start s_1

Problem Formulation

MDP Formulation:

Optimal policy and *human* make sequences of decisions

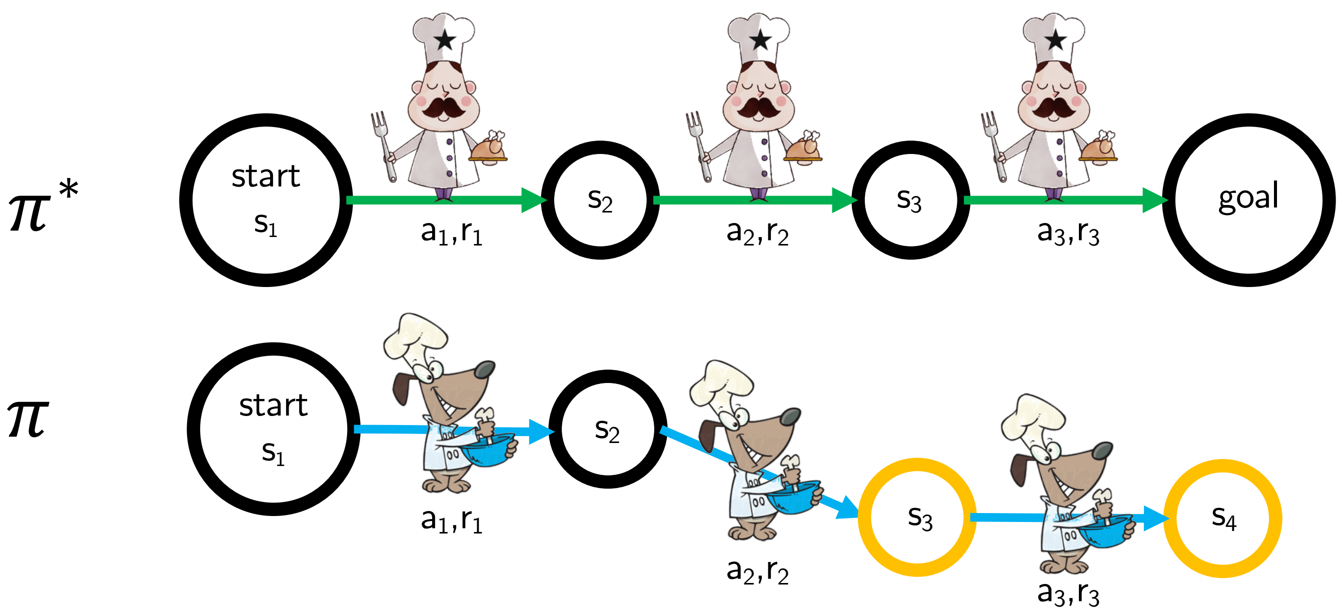
$$\mathcal{M} = (S, A, R, P, \gamma)$$



Problem Formulation

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

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

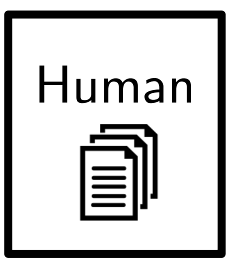


Our Approach

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

sequences of state-action-reward tuples

$\{(s_1, a_1, r_1), (s_2, a_2, r_2), \dots, (s_T, a_T, r_T)\}$

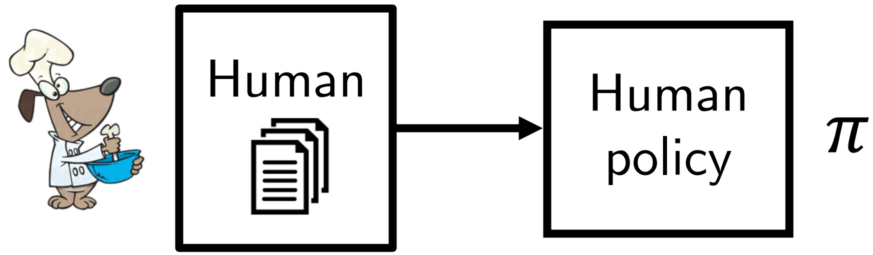


Our Approach

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

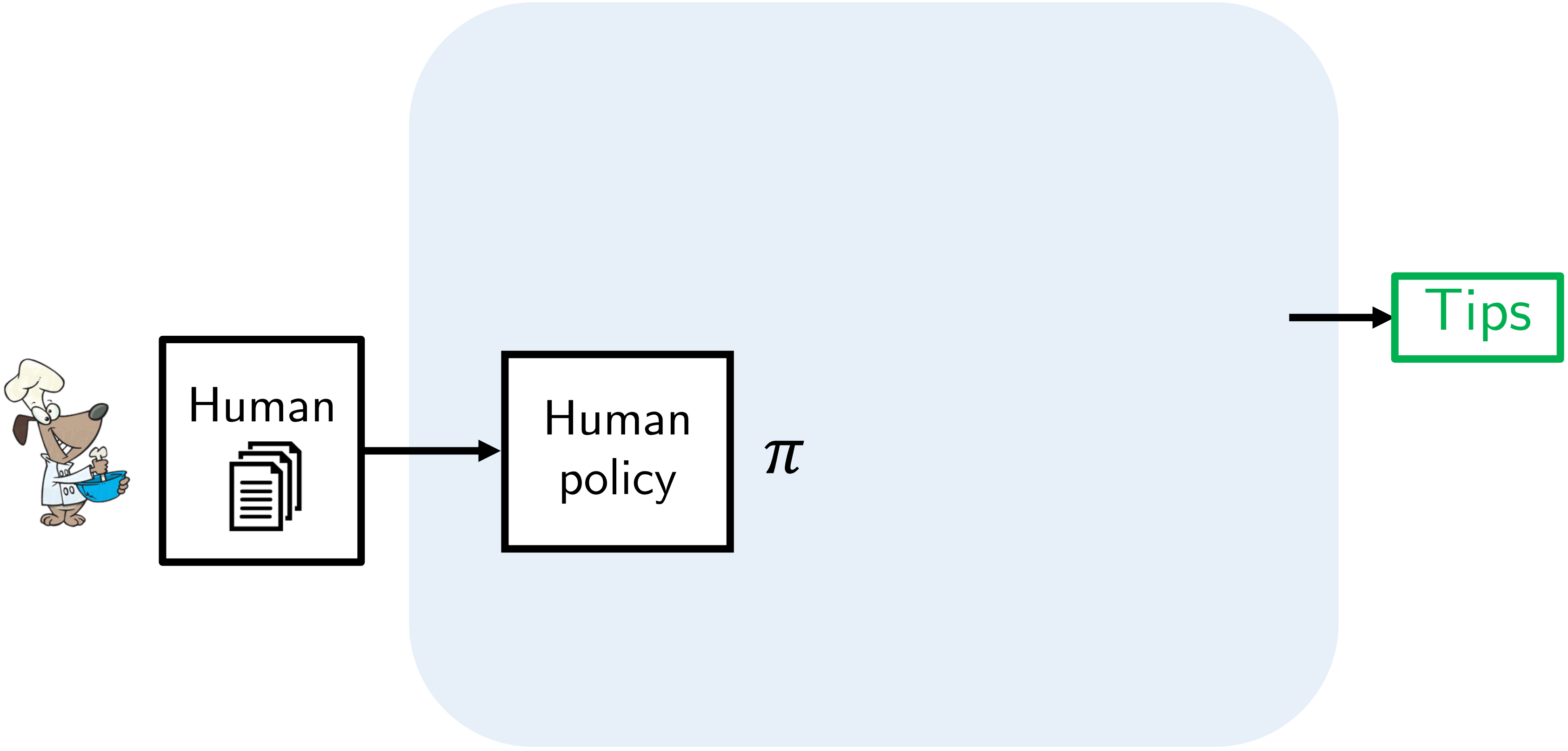
sequences of state-action-reward tuples

$$\{(s_1, a_1, r_1), (s_2, a_2, r_2), \ldots, (s_T, a_T, r_T)\}$$



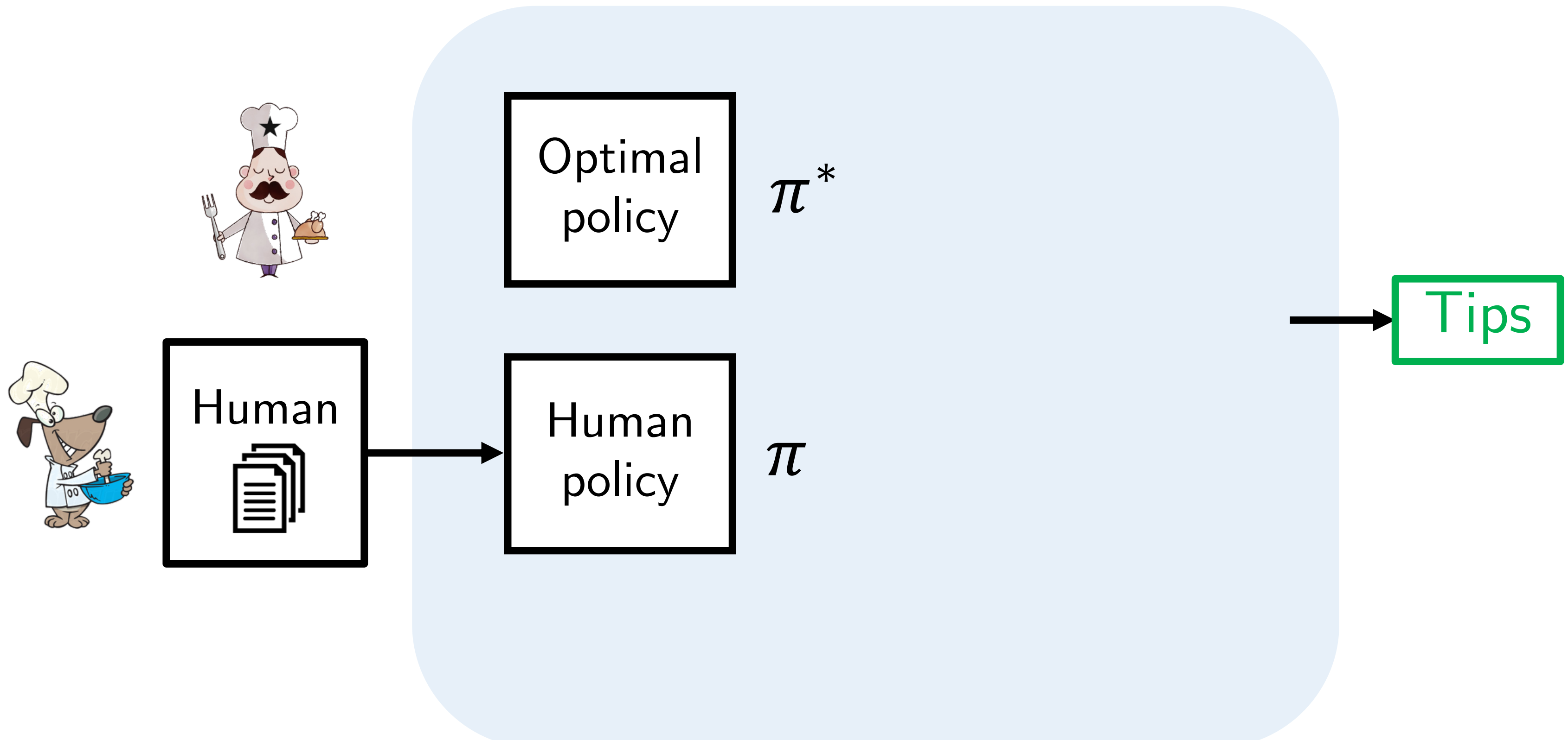
Our Approach

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



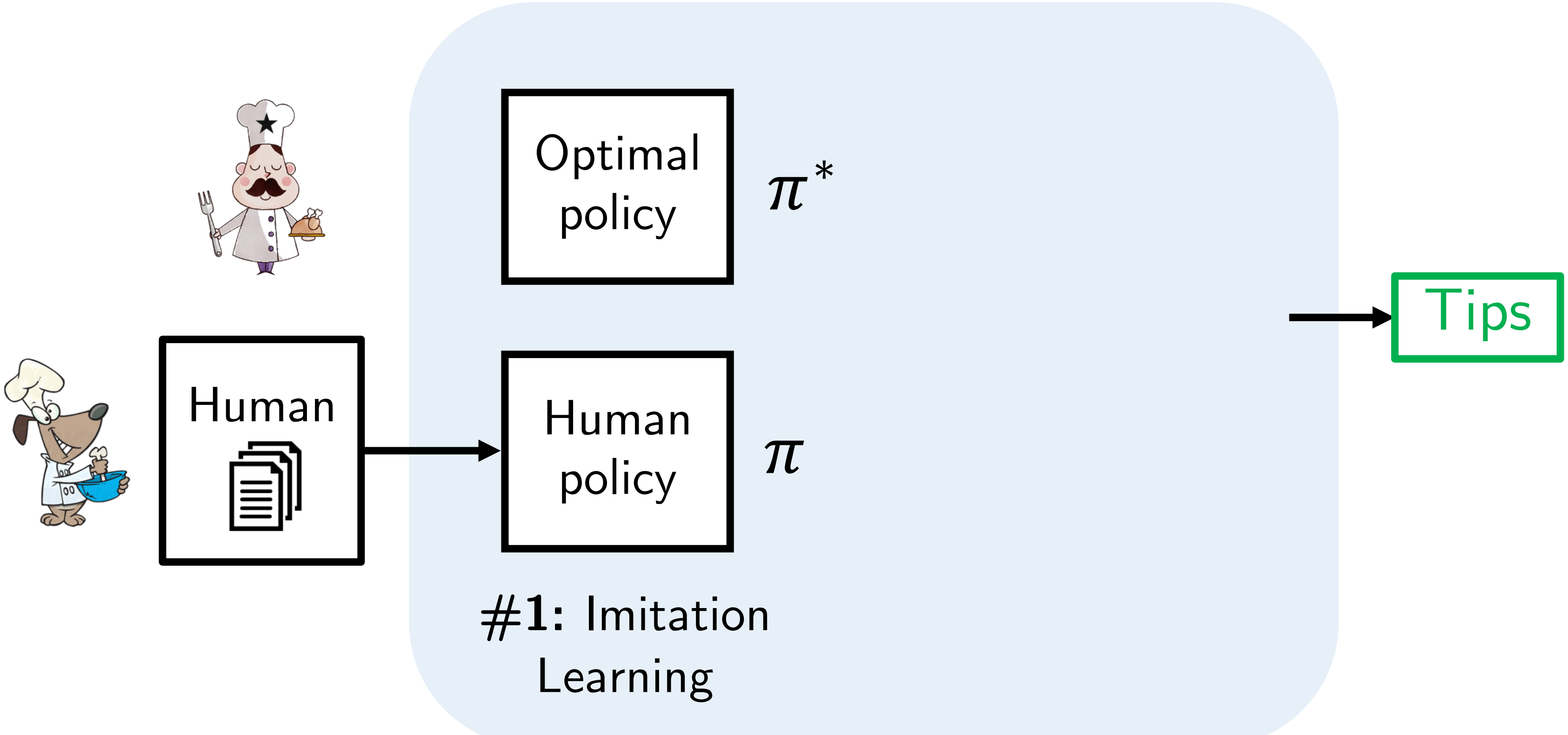
Our Approach

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



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

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

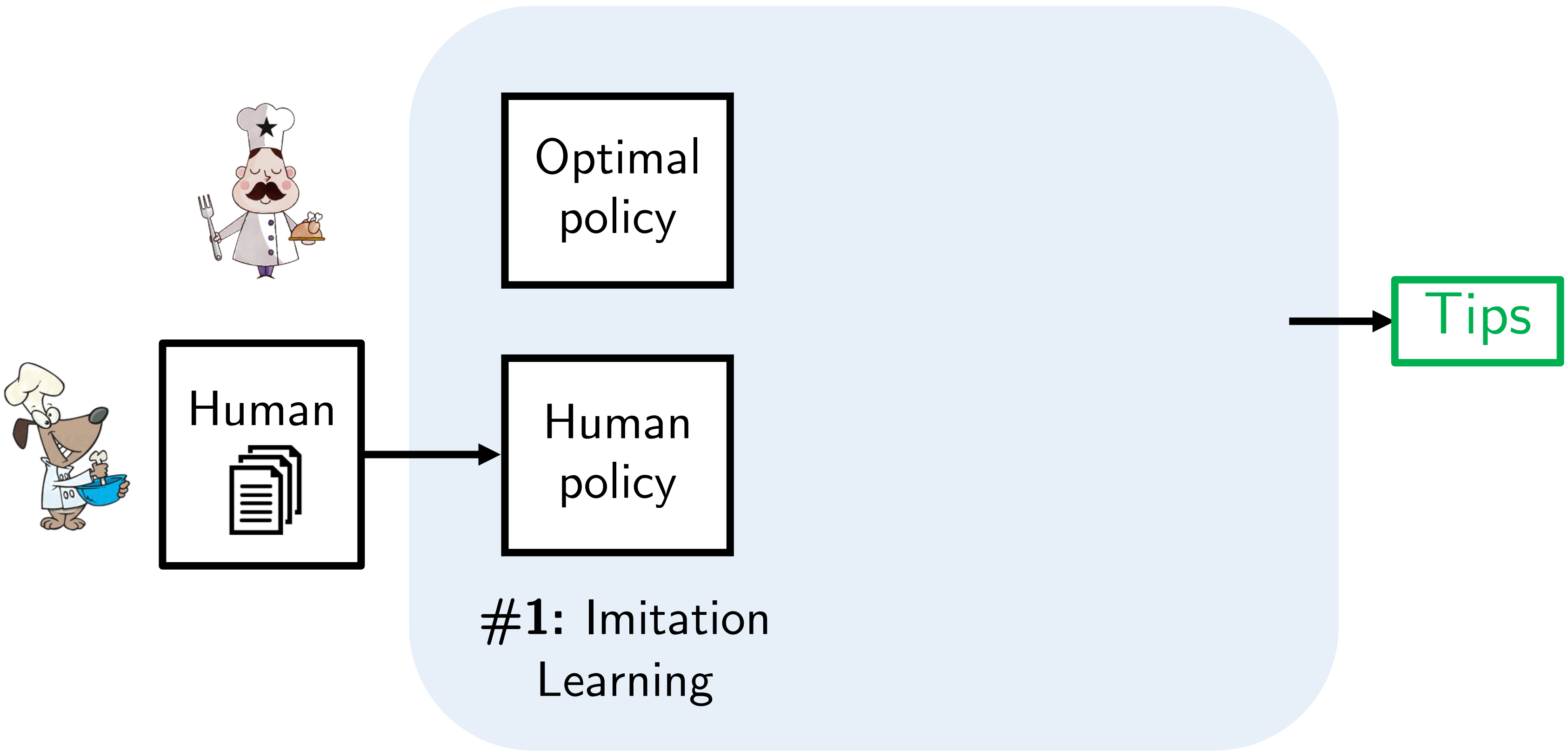
$$Q^\pi(s,a) = \mathbb{E}_{s' \sim p(s'|s,a)}[V^\pi(s')]$$

- Watkins & Dayan 1992

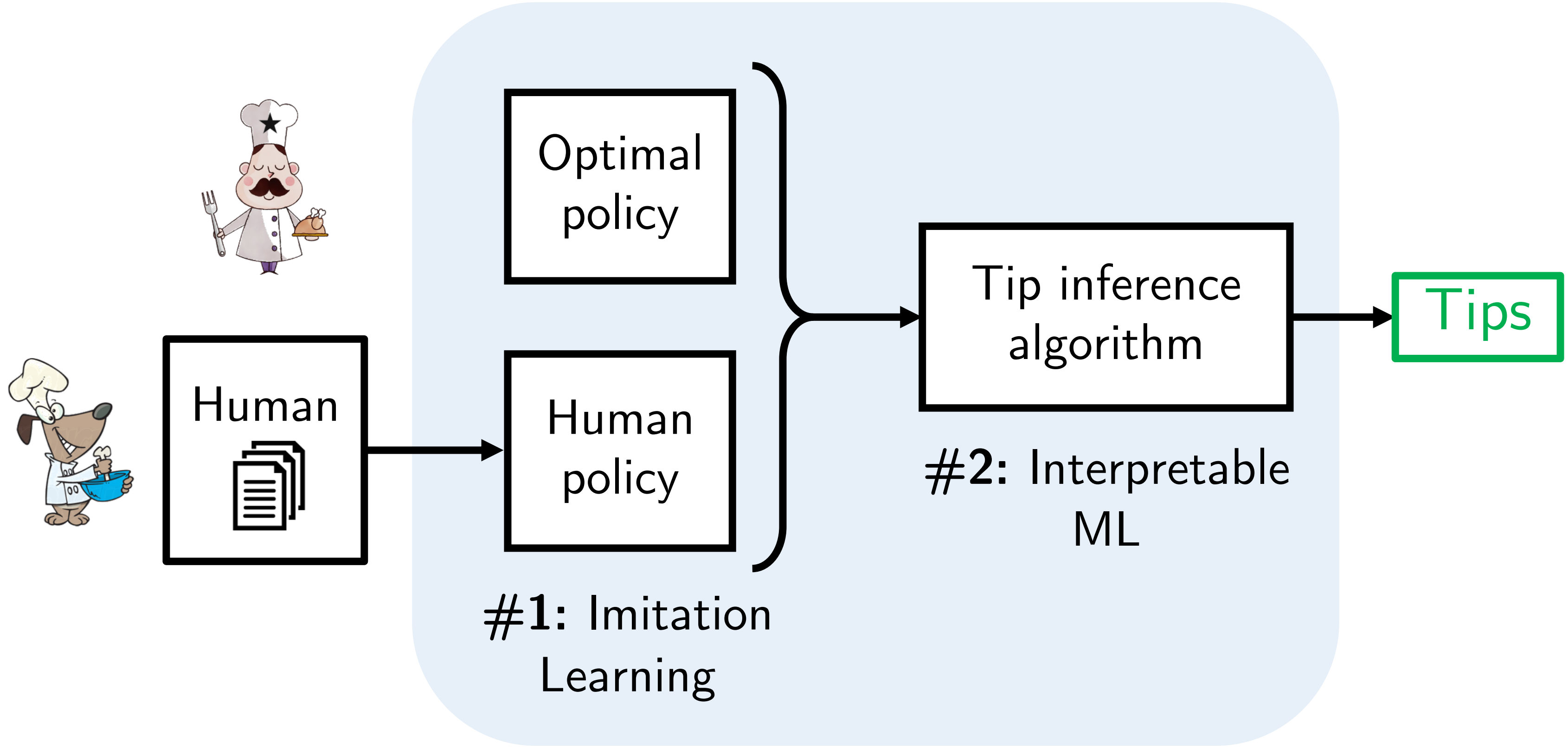
- Learn using supervised learning on trace data obtained using π

$$\hat{Q}^\pi_\theta(s,a) \approx Q^\pi(s,a)$$

Our Approach

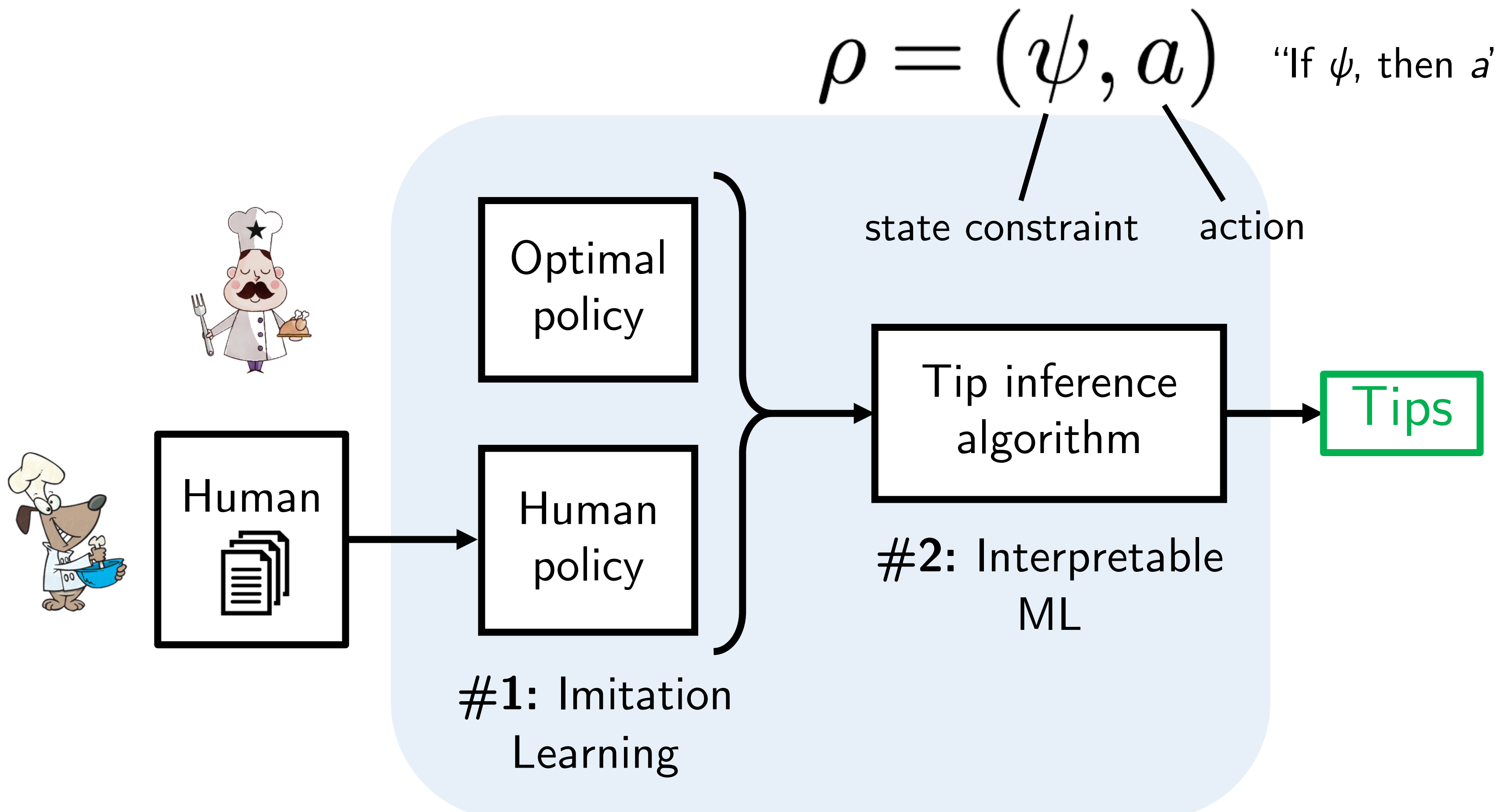


Our Approach



Caruana et al. 2015, Letham et al. 2015

Our Approach



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.
  - J measures the improvement in human reward

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.
  - J measures the improvement in human reward

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

Step 2: Interpretable RL

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

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

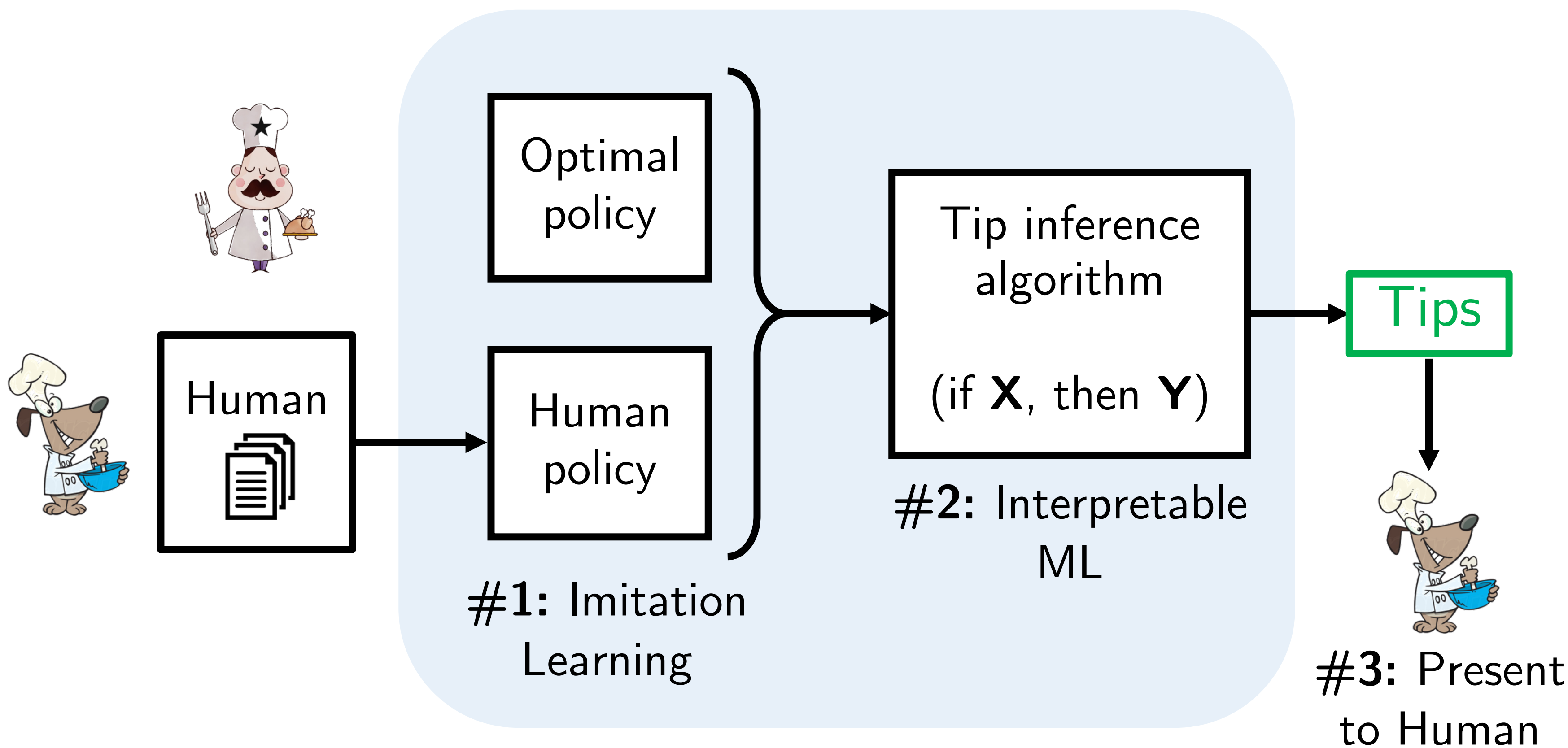
  - $\pi_{\mathrm{h}} \oplus \rho$ denotes overriding the human policy with tip ρ.
  - J measures the improvement in human reward

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

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

  - 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

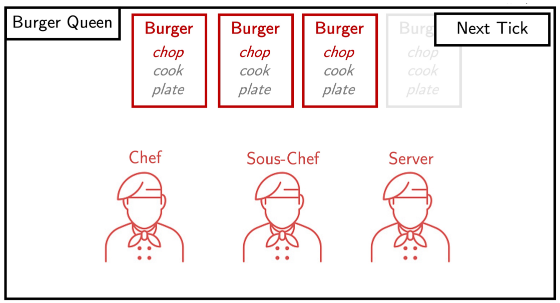
Our Approach



User Study Design

Phase I

N = 200



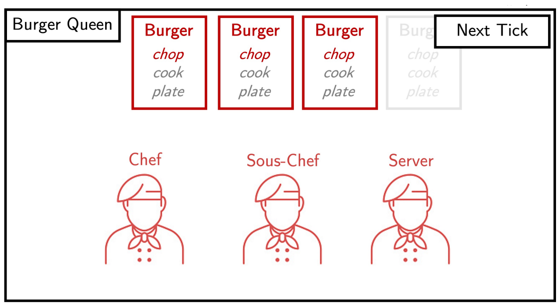
Gather trace data

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

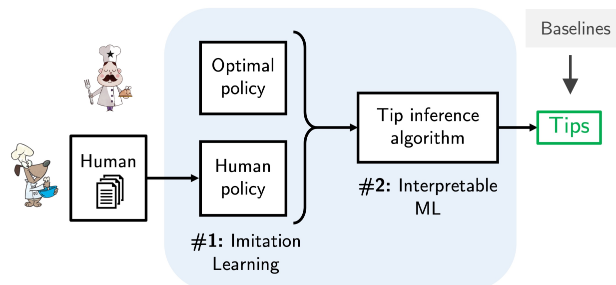
User Study Design

Phase I

N = 200



Gather trace data



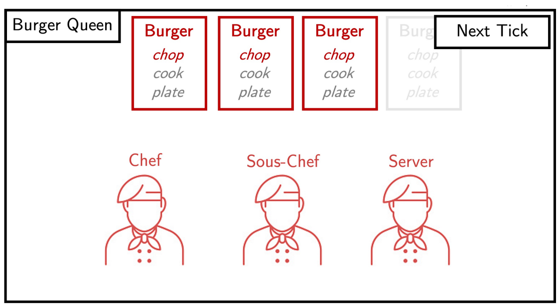
Tip inference

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

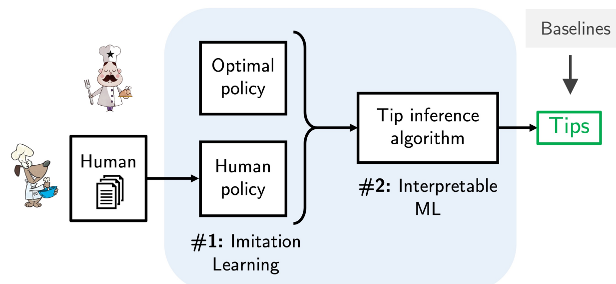
User Study Design

Phase I

N = 200



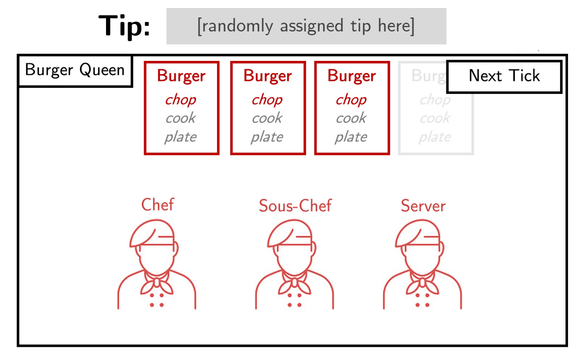
Gather trace data



Tip inference

Phase II

N = 1400



Tip evaluation

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

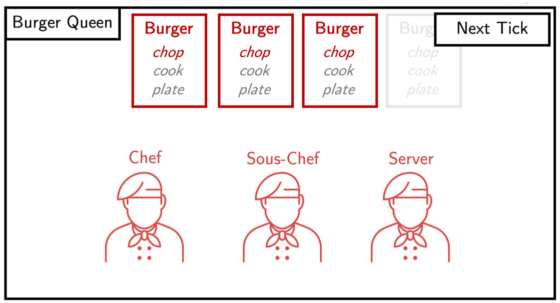
User Study Design

N = 1400

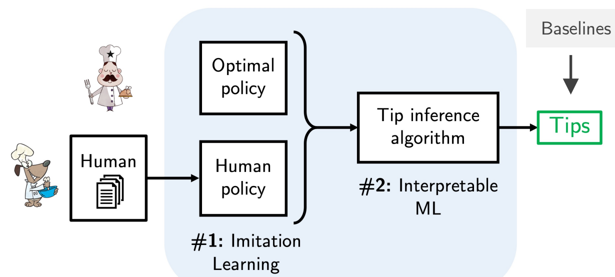
Phase I

N = 200

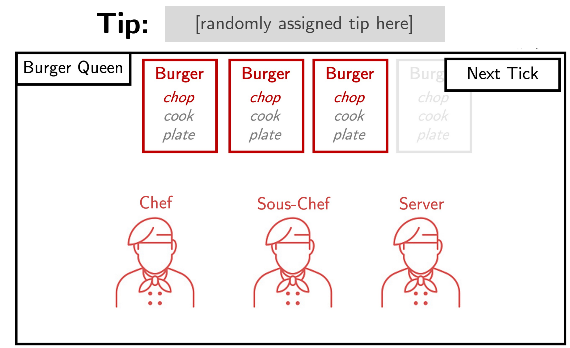
Phase II



Gather trace data



Tip inference



Tip evaluation

Environment

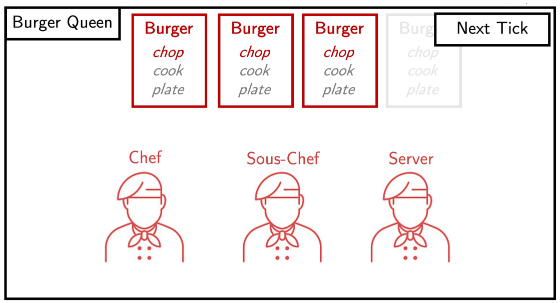
Pre-registered at https://aspredicted.org/blind.php?x=8ye5cb

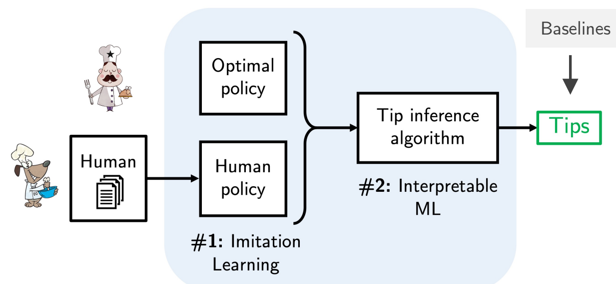
User Study Design

N = 1400

Phase I

N = 200

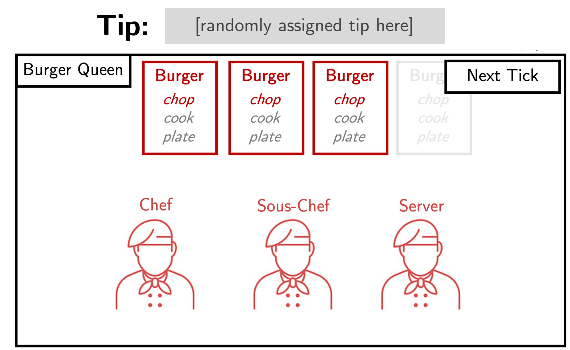




Gather trace data

Tip inference

Phase II



Tip evaluation

Environment

Normal

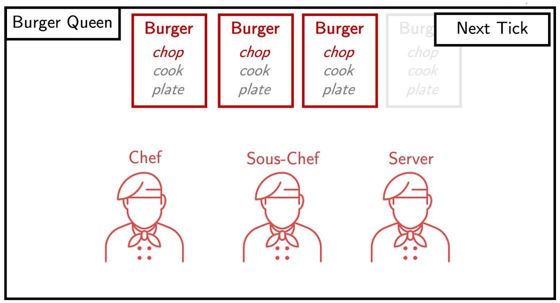
Pre-registered at https://aspredicted.org/blind.php?x=8ye5cb

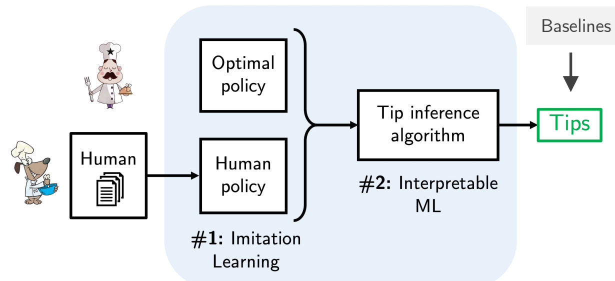
User Study Design

N = 1400

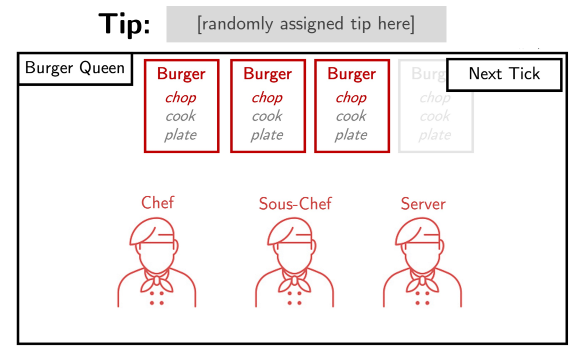
Phase I

N = 200





Phase II



Gather trace data

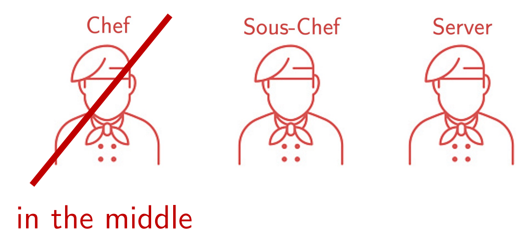
Tip inference

Tip evaluation

Environment

Normal

Disrupted



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

Design Tip Inference

$$\text{if } (\text{order} = o \;\wedge\; \text{subtask} = s \;\wedge\; \text{virtual worker} = w) \text{ then } (\text{assign } (o, s) \text{ to } w),$$

If chopping for Burger #1 and chef are available,
then assign chopping Burger #1 to chef

Design Tip Inference

$$\text{if } (\text{order} = o \ \wedge \ \text{subtask} = s \ \wedge \ \text{virtual worker} = w) \text{ then } (\text{assign } (o, s) \text{ to } w),$$

If chopping for Burger #1 and chef are available,
then assign chopping Burger #1 to chef

$$\text{if } (\text{order} = \text{burger}_1 \ \wedge \ \text{subtask} = \text{cooking} \ \wedge \ \text{virtual worker} = \text{chef}) \text{ then } (\text{assign } (\text{burger}_1, \text{cooking}) \text{ to chef}),$$

$$\text{if } (\text{order} = \text{burger}_2 \ \wedge \ \text{subtask} = \text{cooking} \ \wedge \ \text{virtual worker} = \text{chef}) \text{ then } (\text{assign } (\text{burger}_2, \text{cooking}) \text{ to chef}),$$

Design Tip Inference

$$\text{if } (\text{order} = o \ \wedge \ \text{subtask} = s \ \wedge \ \text{virtual worker} = w) \text{ then } (\text{assign } (o, s) \text{ to } w),$$

If chopping for Burger #1 and chef are available,
then assign chopping Burger #1 to chef

$$\text{if } (\text{order} = \text{burger}_1 \ \wedge \ \text{subtask} = \text{cooking} \ \wedge \ \text{virtual worker} = \text{chef}) \text{ then } (\text{assign } (\text{burger}_1, \text{cooking}) \text{ to chef}),$$

$$\text{if } (\text{order} = \text{burger}_2 \ \wedge \ \text{subtask} = \text{cooking} \ \wedge \ \text{virtual worker} = \text{chef}) \text{ then } (\text{assign } (\text{burger}_2, \text{cooking}) \text{ to chef}),$$

$$\downarrow$$

Assign cooking to chef 2 times

Design Tip Inference

$$\text{if (order} = o \ \wedge \ \text{subtask} = s \ \wedge \ \text{virtual worker} = w) \text{ then (assign } (o, s) \text{ to } w),$$

If chopping for Burger #1 and chef are available,
then assign chopping Burger #1 to chef

$$\text{if (order} = \text{burger}_1 \ \wedge \ \text{subtask} = \text{cooking} \ \wedge \ \text{virtual worker} = \text{chef}) \text{ then (assign } (\text{burger}_1, \text{cooking}) \text{ to chef}),$$

$$\text{if (order} = \text{burger}_2 \ \wedge \ \text{subtask} = \text{cooking} \ \wedge \ \text{virtual worker} = \text{chef}) \text{ then (assign } (\text{burger}_2, \text{cooking}) \text{ to chef}),$$

↓

Assign cooking to chef 2 times

↓

Chef should cook twice

Design Disrupted Configuration

🍔 x 4 within 50 ticks

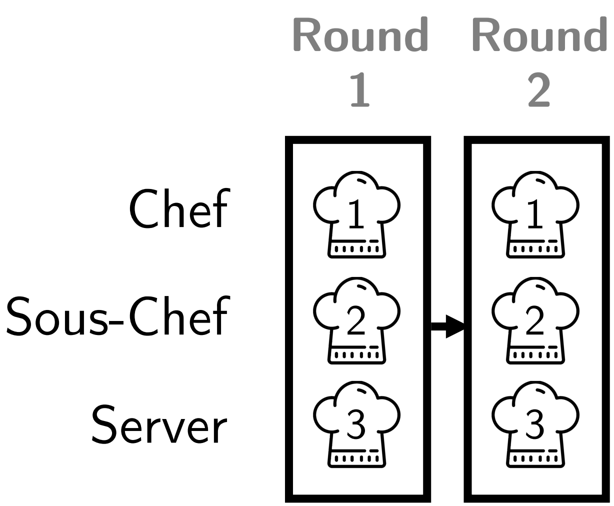
Pre-registered at https://aspredicted.org/blind.php?x=8ye5cb

Design Disrupted Configuration

x 4 within 50 ticks

+ Our Tip

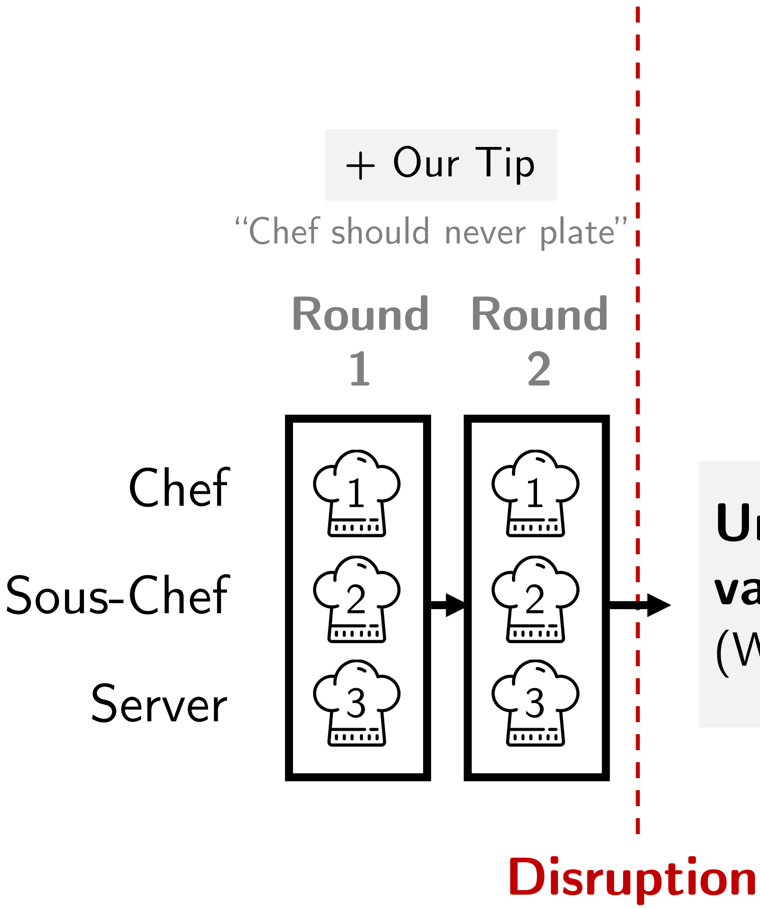
"Chef should never plate"



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

Design Disrupted Configuration

x 4 within 50 ticks

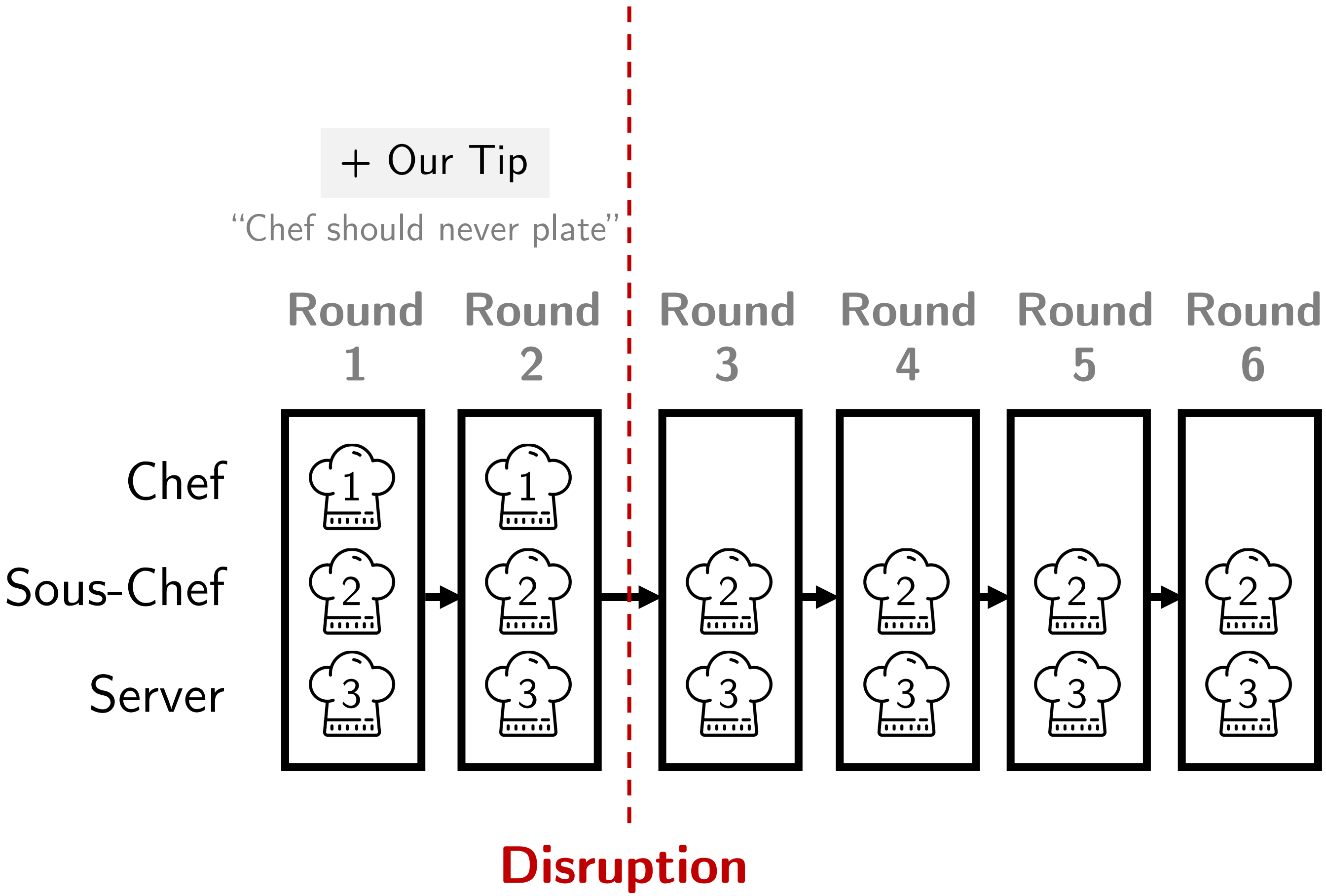


Unfortunately, the Chef is on vacation during this round.
(Who travels these days...)

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

Design Disrupted Configuration

🍔 x 4 within 50 ticks



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

Phase I Inferred Tips

Algorithm

Server
should cook twice

Amazon Mechanical Turk, N = 172
mean age 36.4, 62% female

Phase I Inferred Tips

Algorithm	Human
Server should cook twice	
	Most frequent tip chosen by participants

Amazon Mechanical Turk, N = 172
mean age 36.4, 62% female

Phase I Inferred Tips

Algorithm	Human
Server should cook twice	Server should cook once

Most frequent tip chosen by participants

Amazon Mechanical Turk, N = 172
mean age 36.4, 62% female

Phase I Inferred Tips

Algorithm	Human	Baseline
Server should cook twice	Server should cook once	
	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

Phase I Inferred Tips

Algorithm	Human	Baseline
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

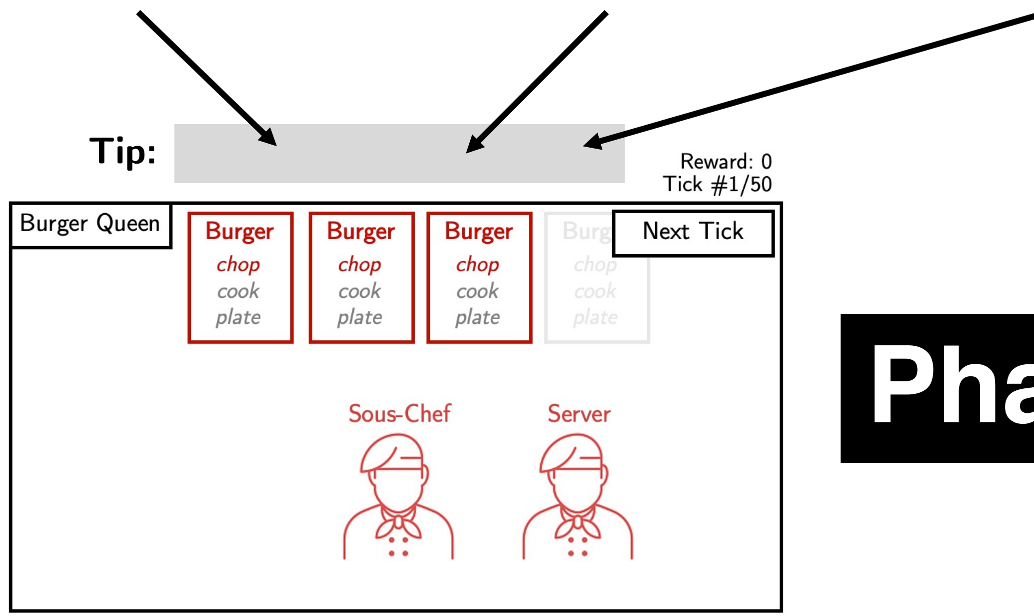
Amazon Mechanical Turk, N = 172

mean age 36.4, 62% female

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

Phase II

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



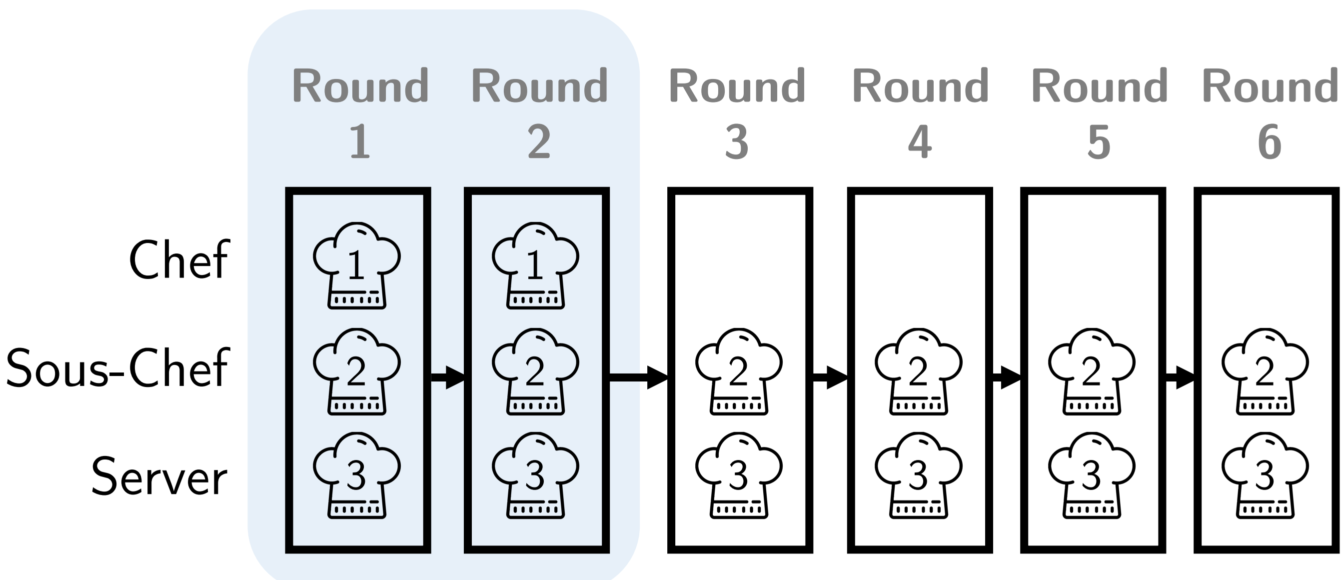
Phase II

Algorithm vs Human

Algorithm	Human
Server should cook twice	Server should cook once

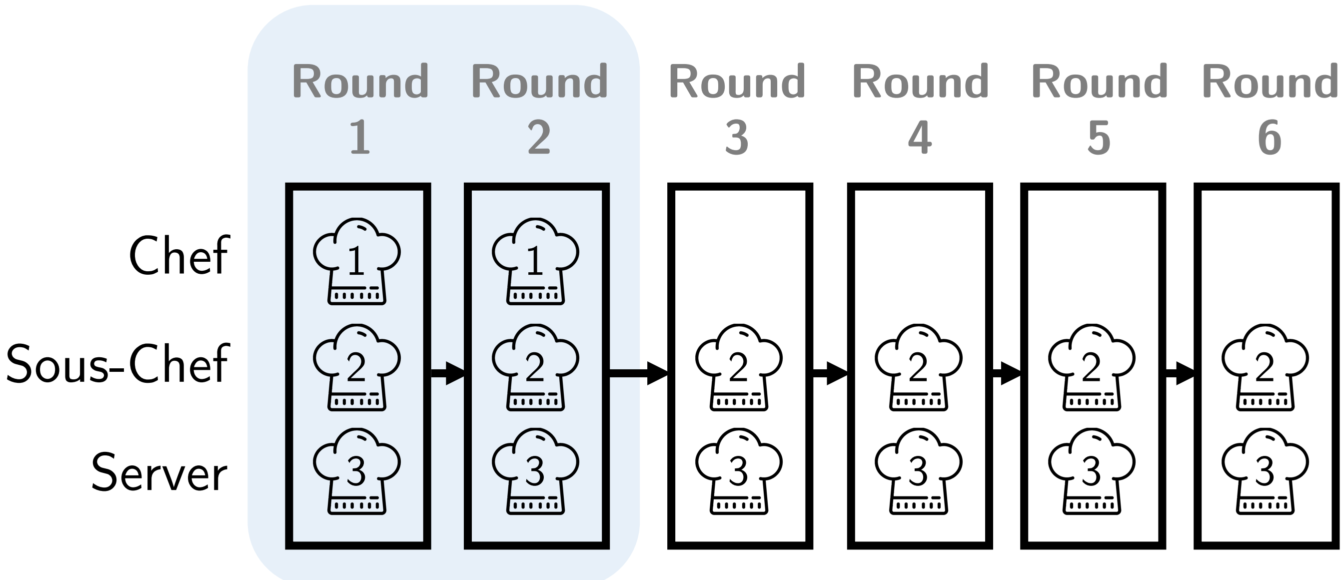
Algorithm vs Human

Algorithm	Human
Server should cook twice	Server should cook once



Algorithm vs Human

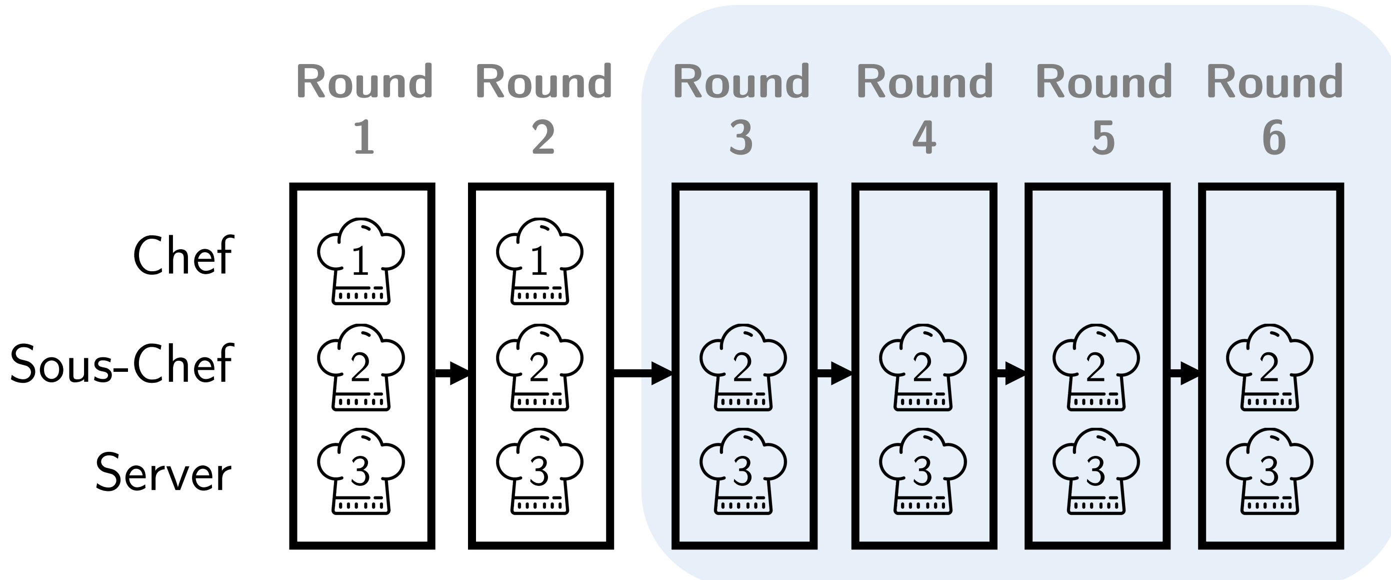
Algorithm	Human
Server should cook twice	Server should cook once



“Server shouldn’t cook”

Algorithm vs Human

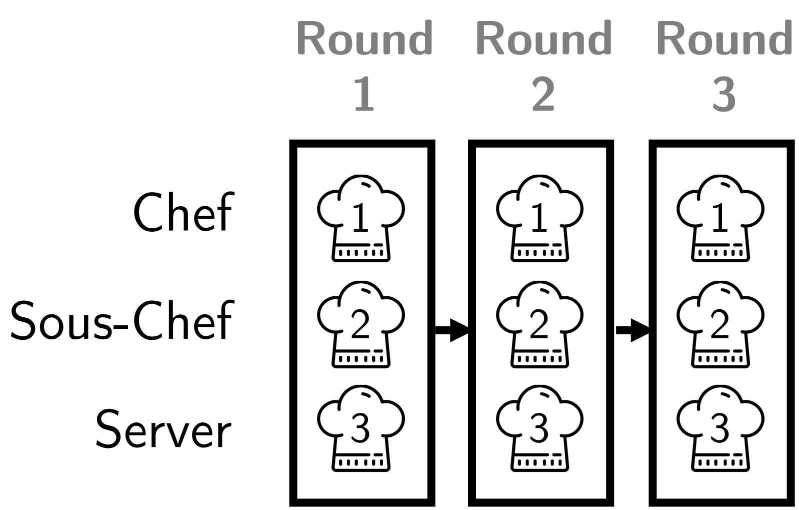
Algorithm	Human
Server should cook twice	Server should cook once



"Server shouldn't cook"

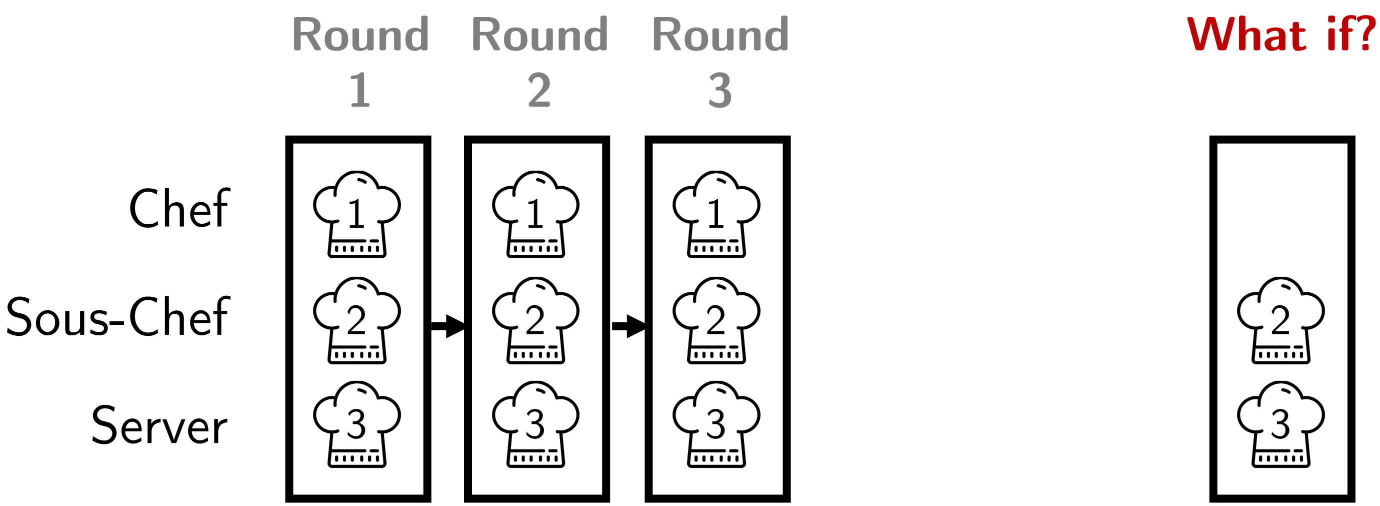
Algorithm vs Human

Algorithm	Human
Server should cook twice	Server should cook once



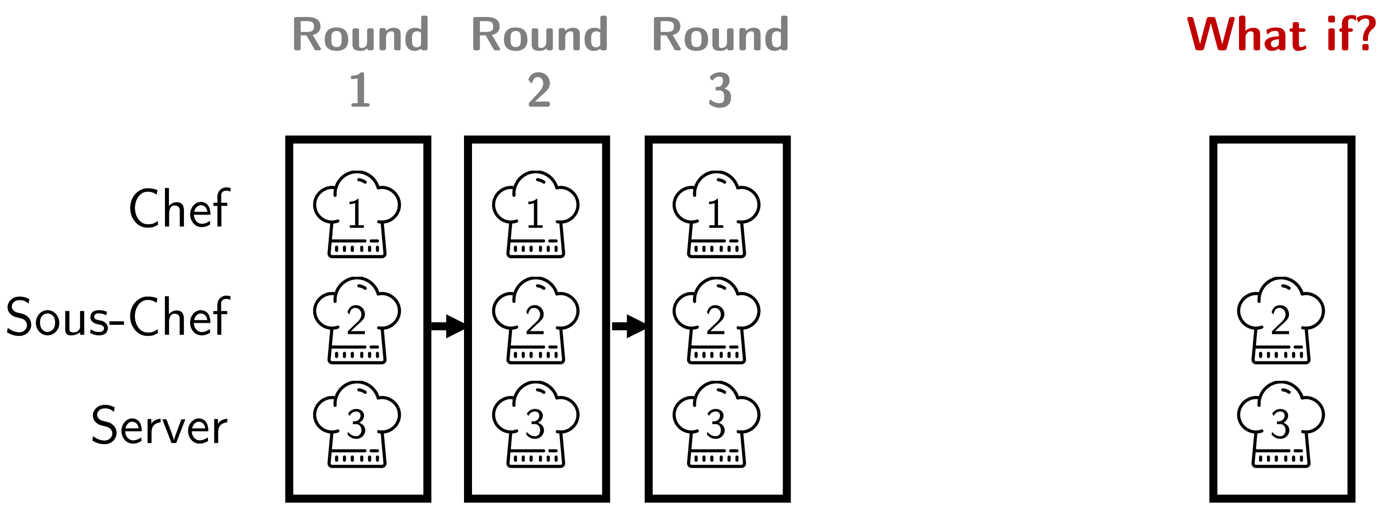
Algorithm vs Human

Algorithm	Human
Server should cook twice	Server should cook once



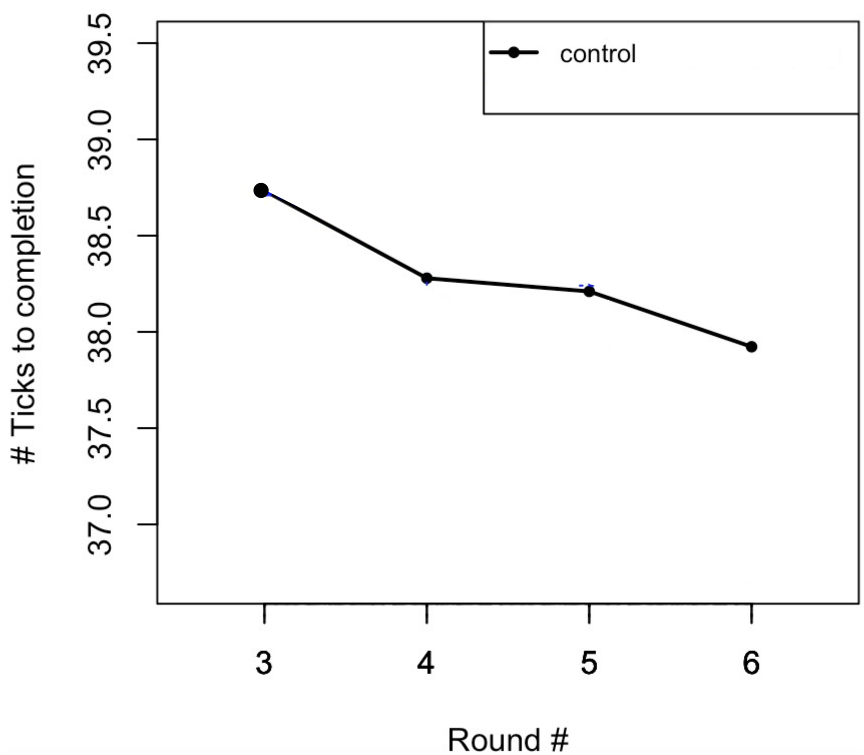
Algorithm vs Human

Algorithm	Human	Hypothetical
Server should cook twice	Server should cook once	Server shouldn't cook



Results People Improve Over Time

Ticks to completion

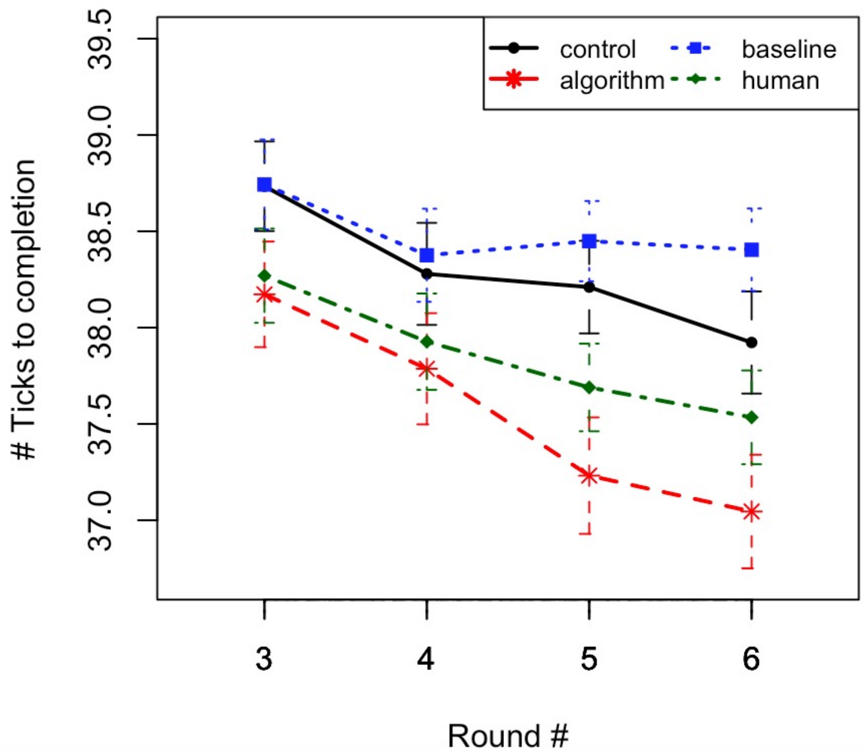


Amazon Mechanical Turk, N = 1,011

mean age 34.9, 60% female

Results Our Tip Improves Performance

Ticks to completion



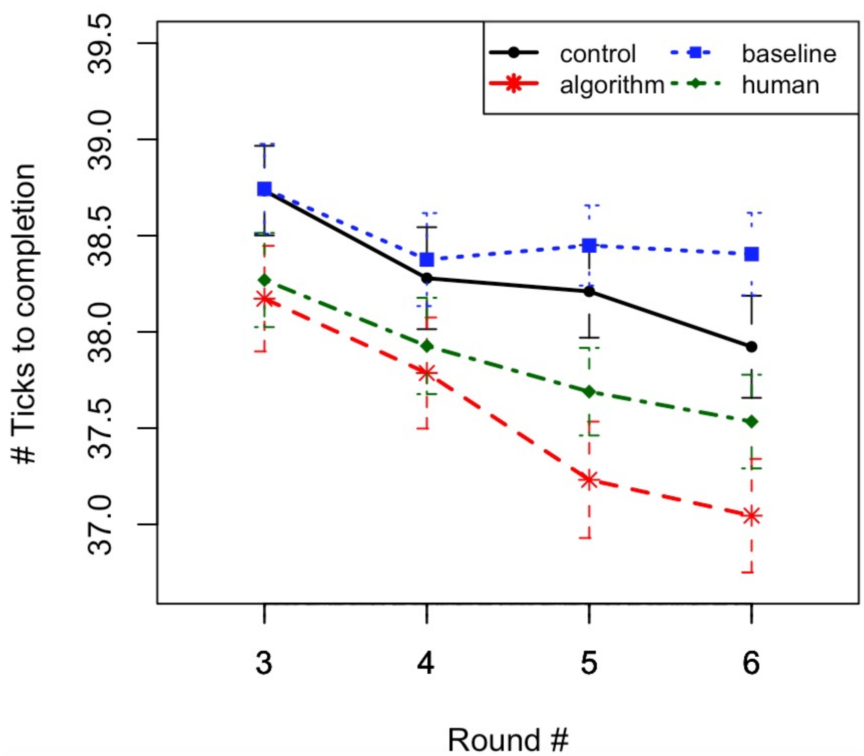
One-sided T-Tests

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

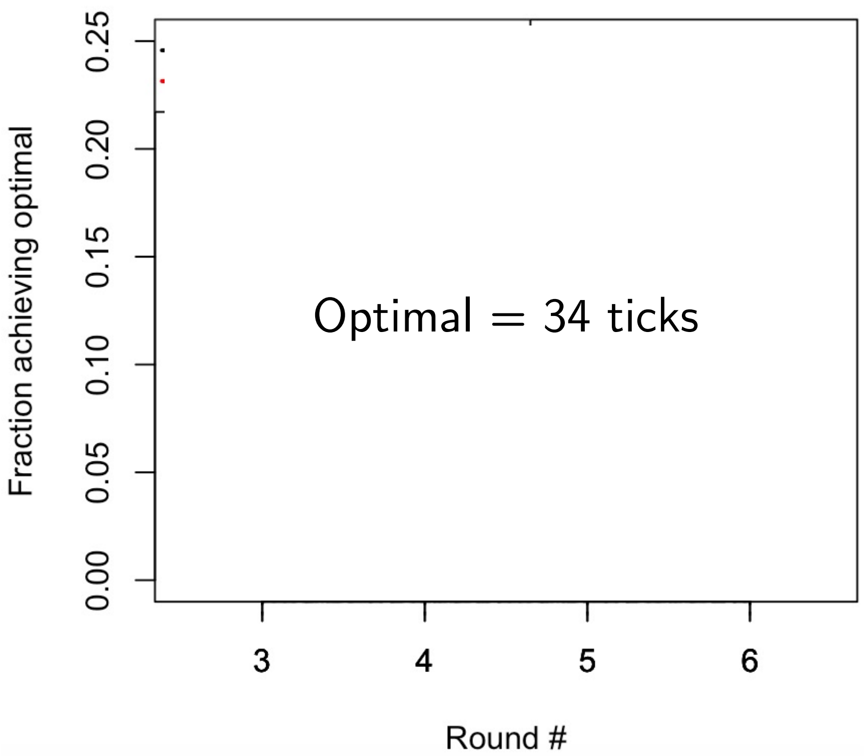
Amazon Mechanical Turk, N = 1,011
mean age 34.9, 60% female

Results

Ticks to completion



Fraction achieving optimal

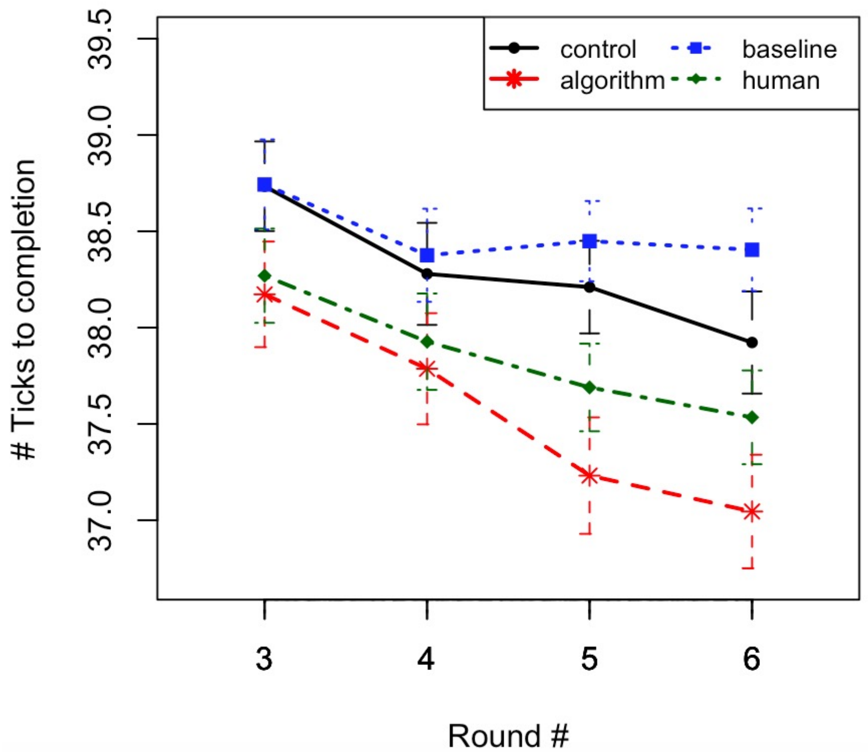


Amazon Mechanical Turk, N = 1,011

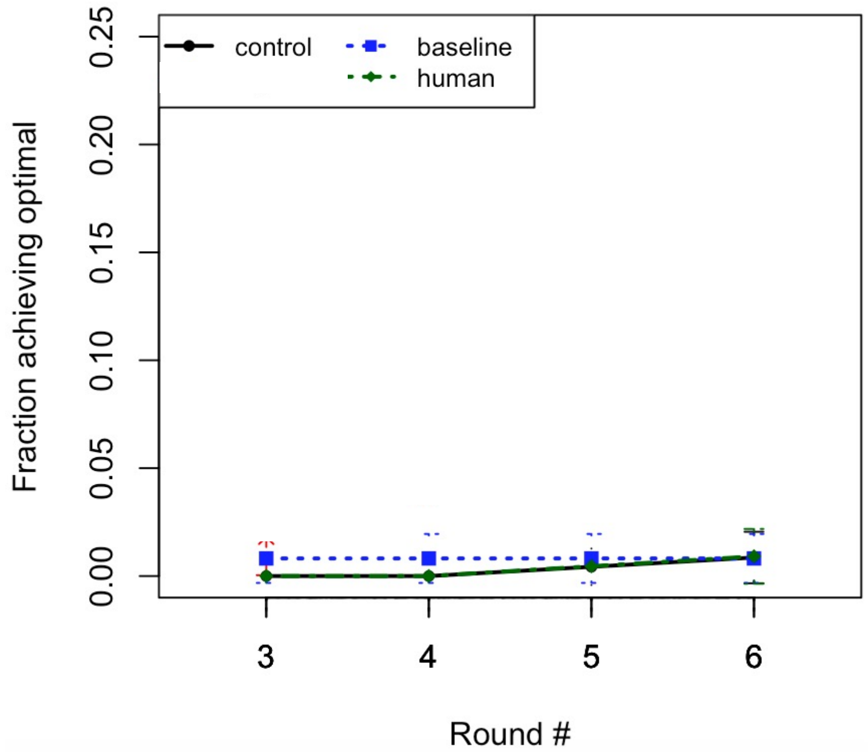
mean age 34.9, 60% female

Results Difficult to Reach Optimal

Ticks to completion



Fraction achieving optimal

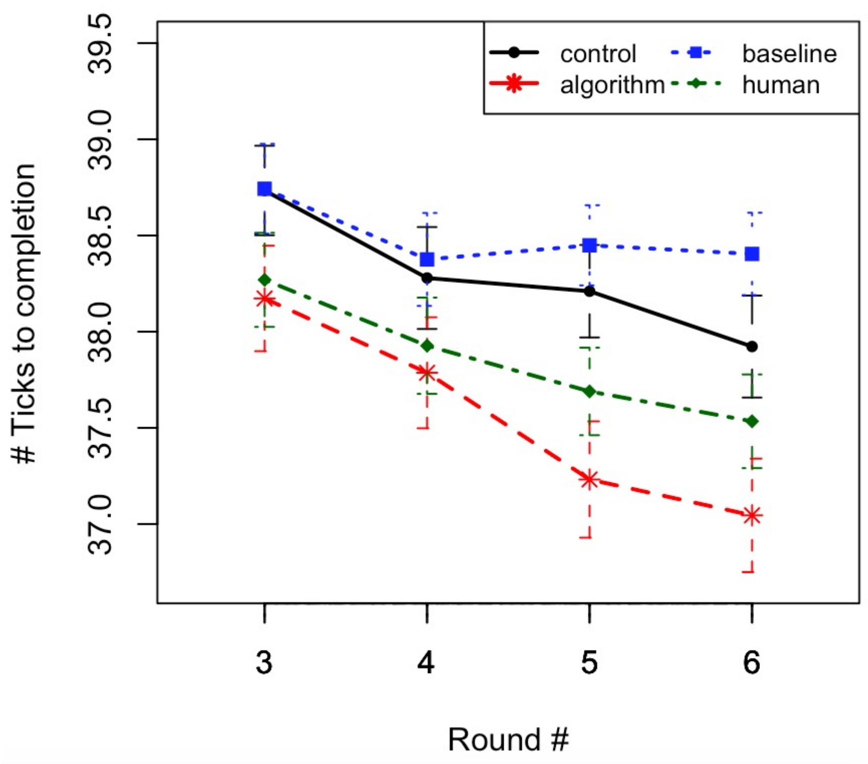


Amazon Mechanical Turk, N = 1,011

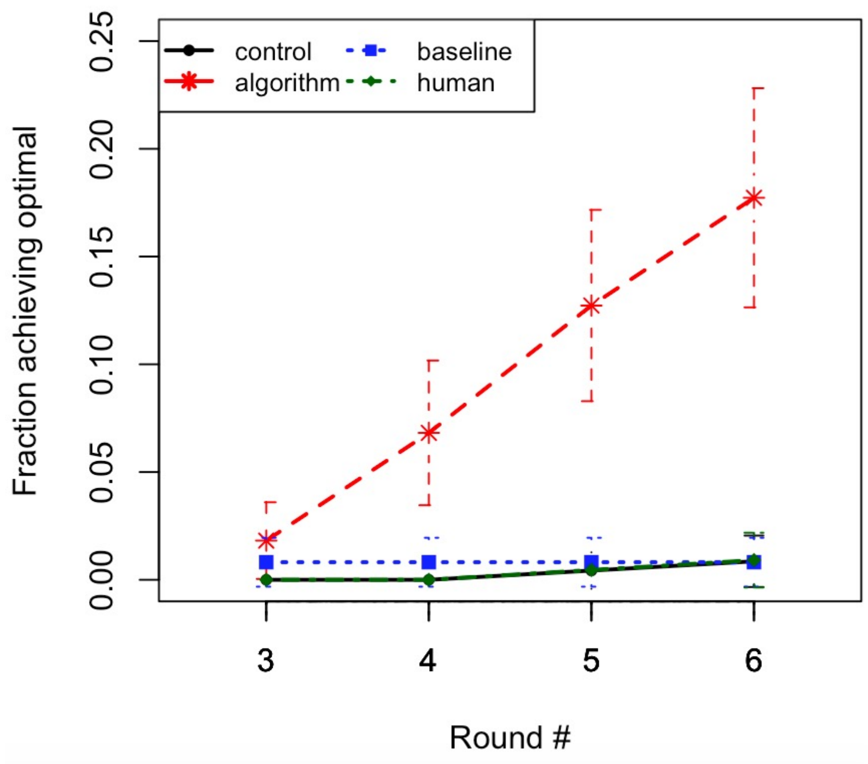
mean age 34.9, 60% female

Results Our Tip Helps Reach Optimal

Ticks to completion



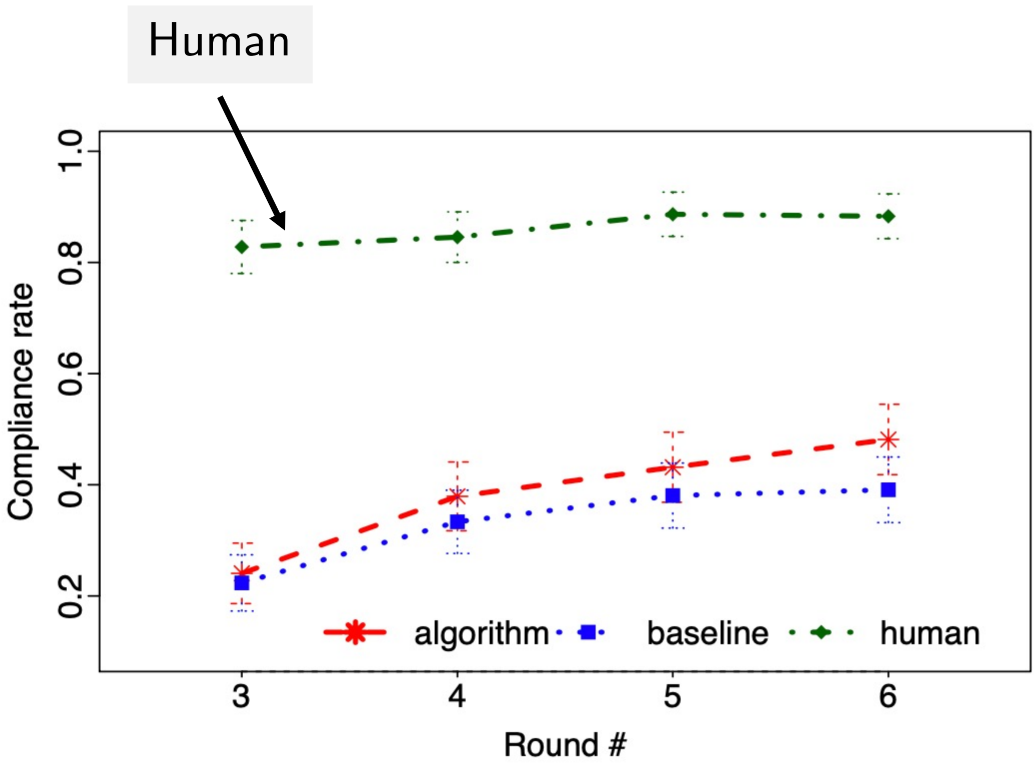
Fraction achieving optimal



Amazon Mechanical Turk, N = 1,011

mean age 34.9, 60% female

Results Complying with Intuitive Tip

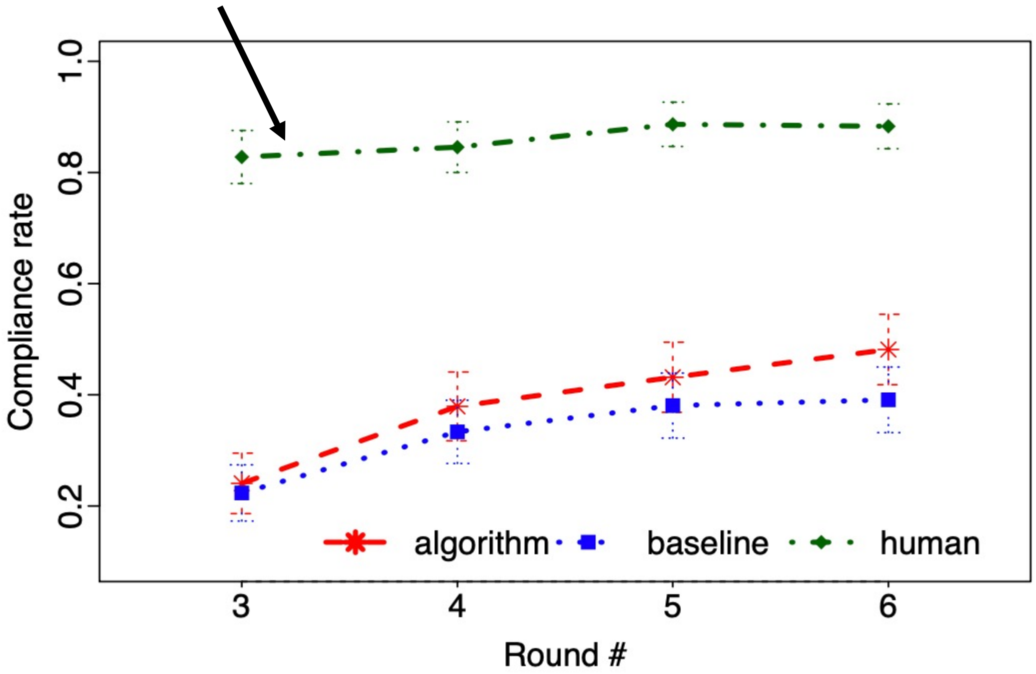


Amazon Mechanical Turk, N = 1,011

mean age 34.9, 60% female

Results Complying with Intuitive Tip

Human 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

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

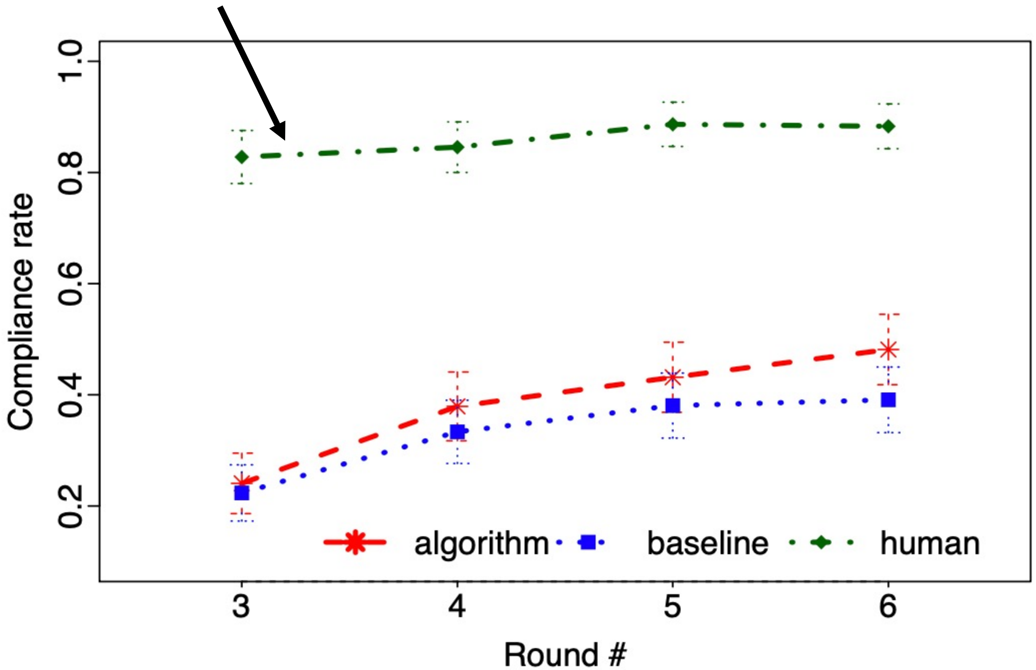
R_1pA8wDYgWc9hblt

Amazon Mechanical Turk, N = 1,011

mean age 34.9, 60% female

Results Complying with Intuitive Tip

Human 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

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

R_1pA8wDYgWc9hblt

"It stunk honestly. **The server takes forever to cook."**

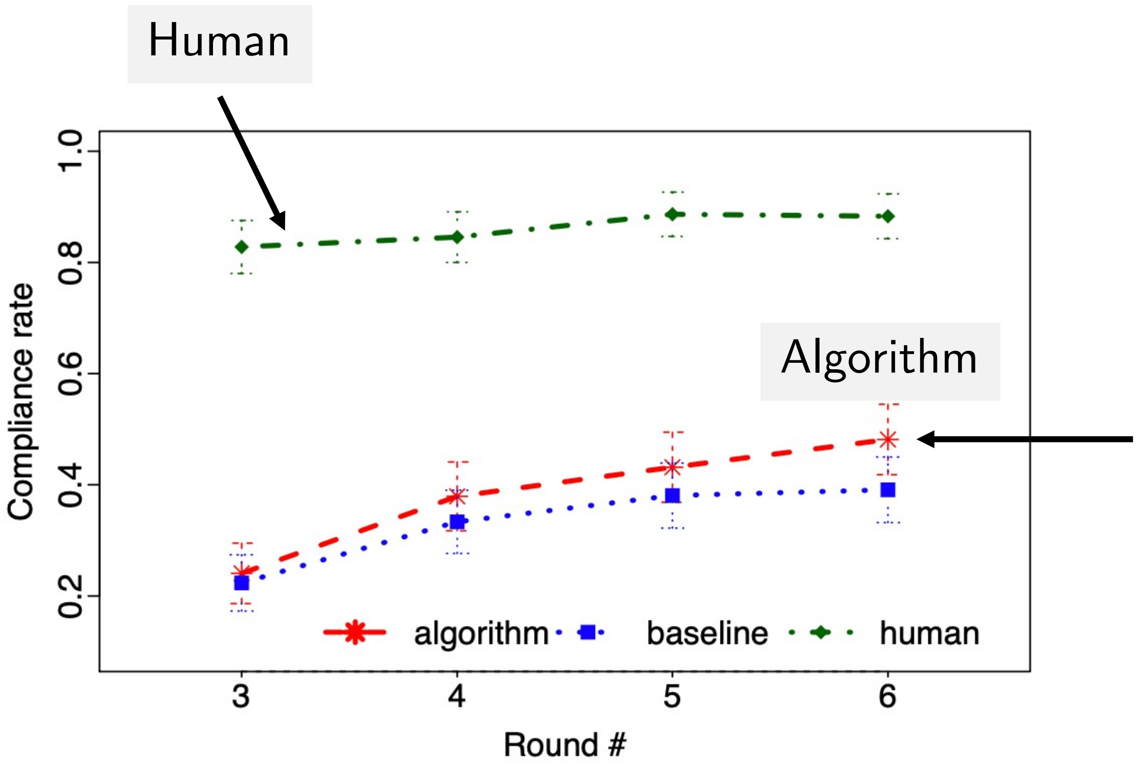
R_beijQ8guDyExa5r

"I used the tip but **I don't think it was helpful.** The server took long to cook."

Amazon Mechanical Turk, N = 1,011

mean age 34.9, 60% female

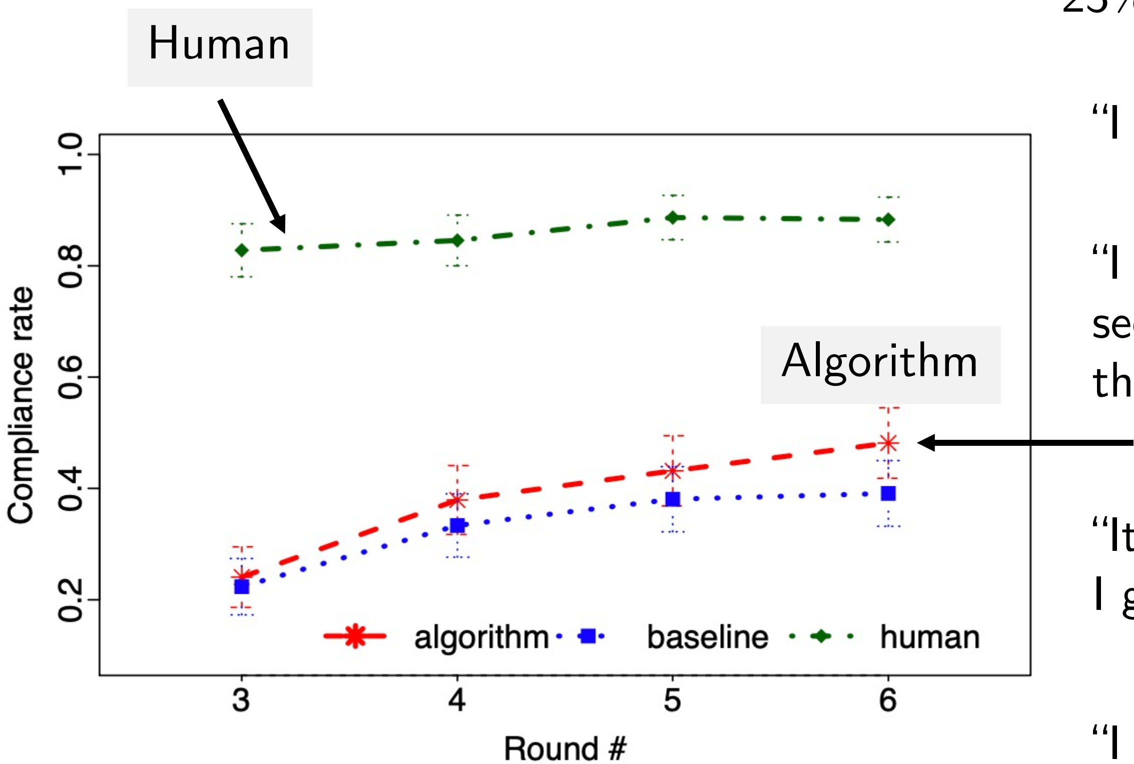
Results Against Counterintuitive Tips



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

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

Structure of Optimal Policy

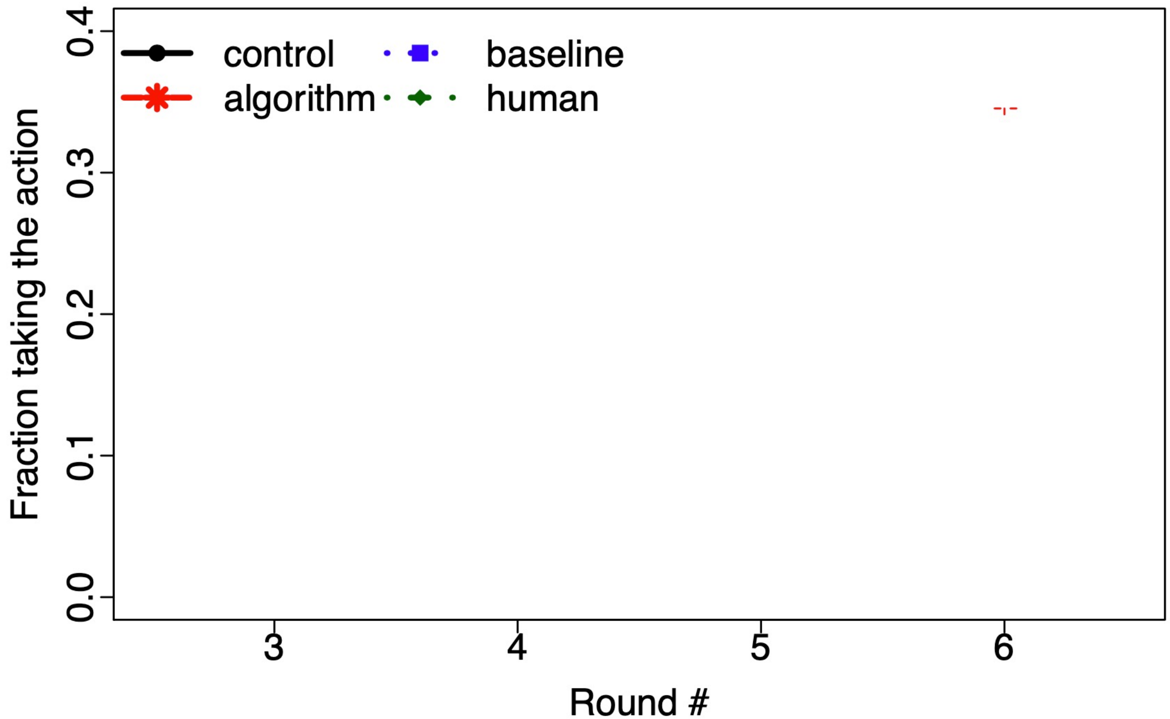
		Chop	Cook	Plate	
Sous-Chef	2	3	2	2	times
Server	3	1	2	2	times

Algorithm

Baseline

Results Learning Beyond Tips

Sous-Chef
chops 3 times



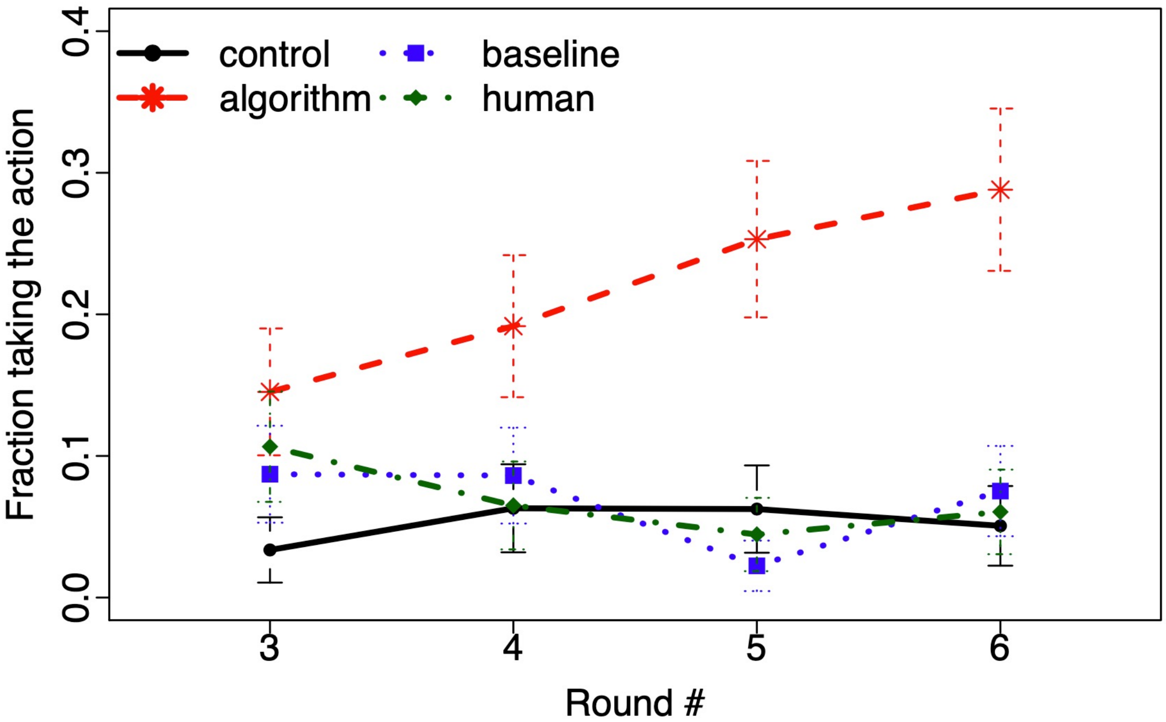
Amazon Mechanical Turk, N = 1,011

mean age 34.9, 60% female

Results Learning Beyond Tips

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

Sous-Chef chops 3 times

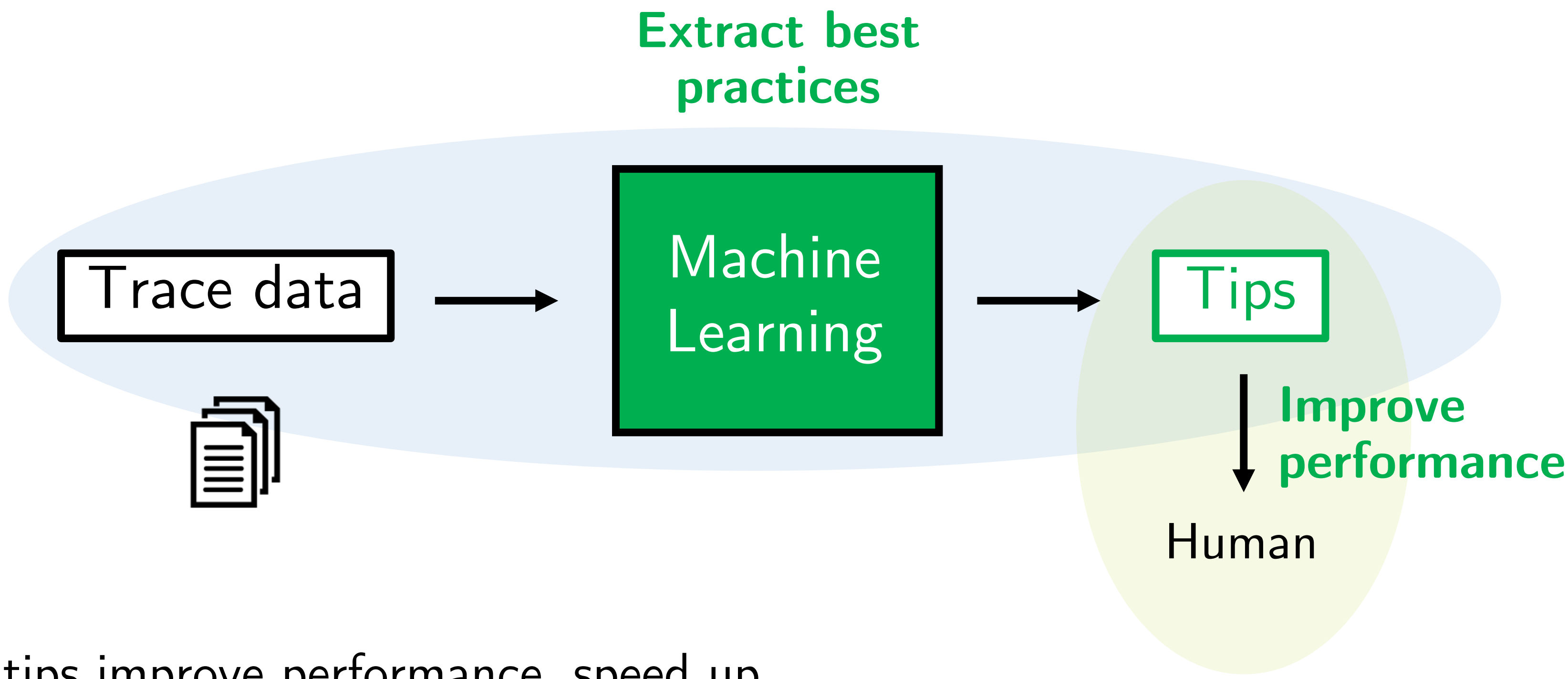


Amazon Mechanical Turk, N = 1,011

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

Feedback (tips) very welcome!

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

Our Approach

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

Optimal policy π^*

Human

Human policy π

Tips

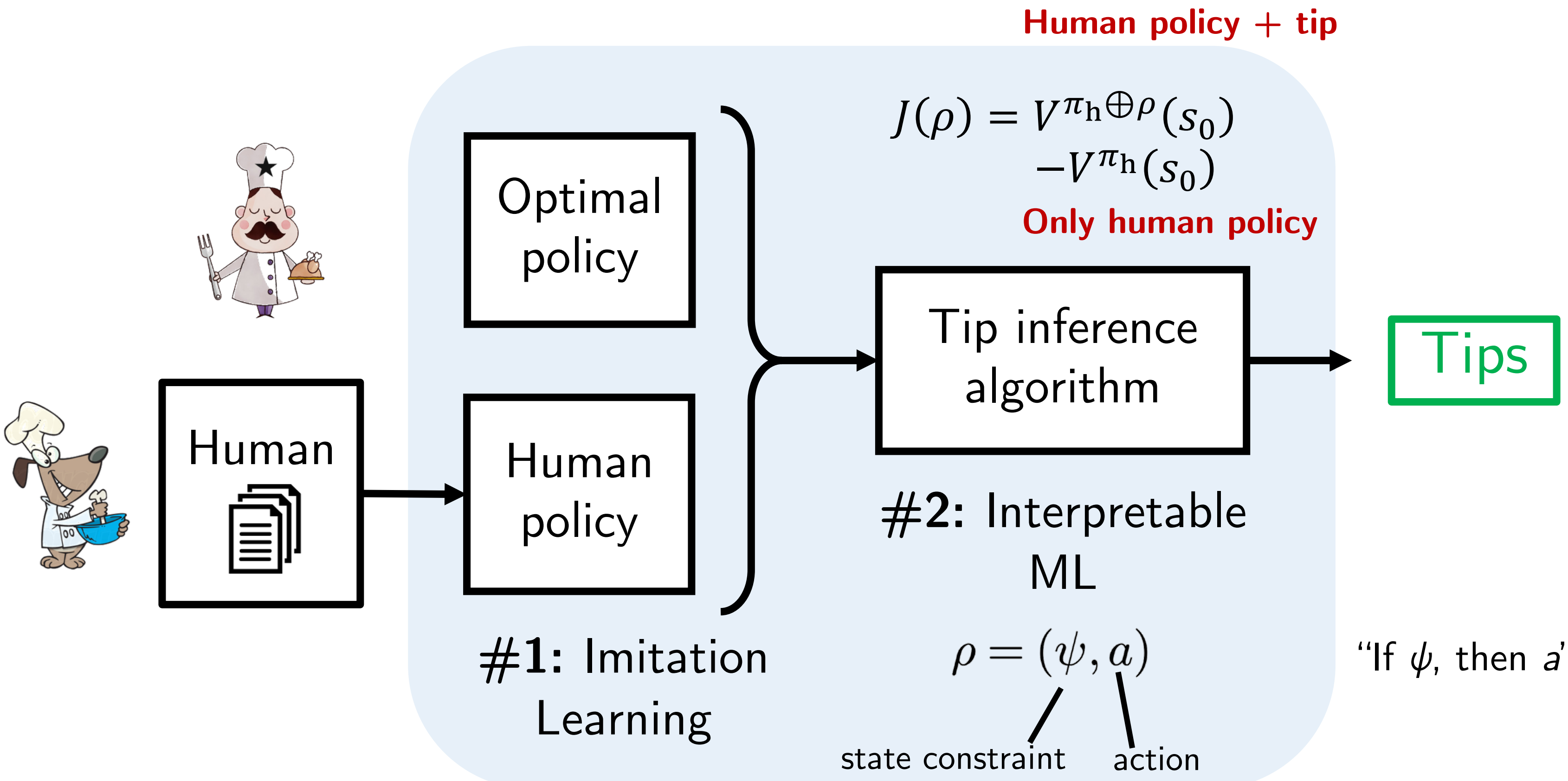
#1: Imitation Learning

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

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

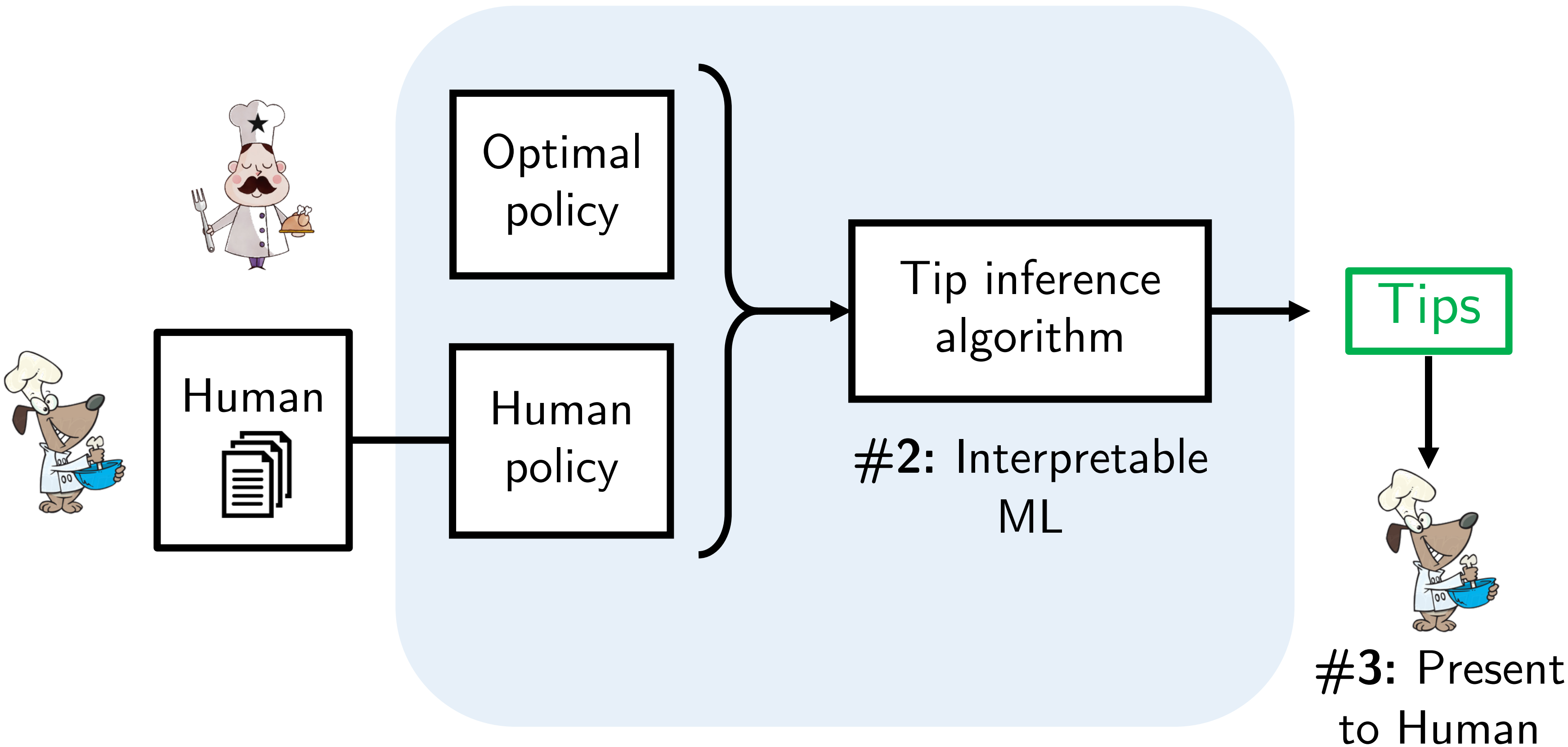
Ng & Russell 2000, Abbeel & Ng 2004, Ziebart et al. 2008, Ross & Bagnell 2011

Our Approach



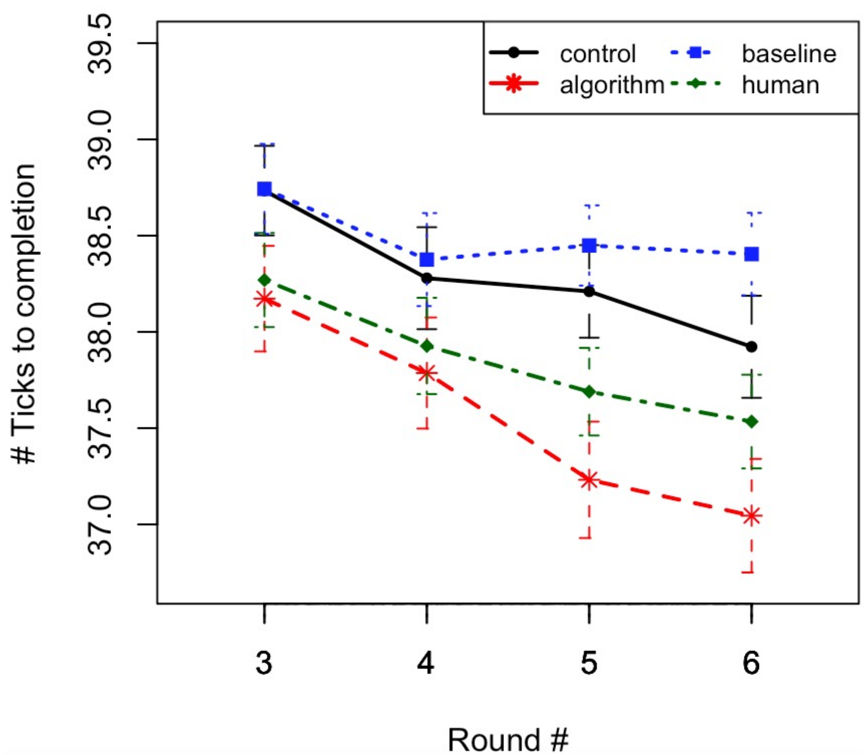
Caruana et al. 2015, Letham et al. 2015

Our Approach

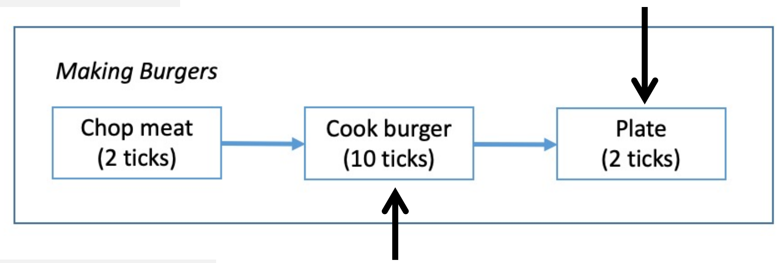


Results Good Tip = Consequential

Ticks to completion



Baseline Sous-Chef should plate twice



Algorithm Server should cook twice

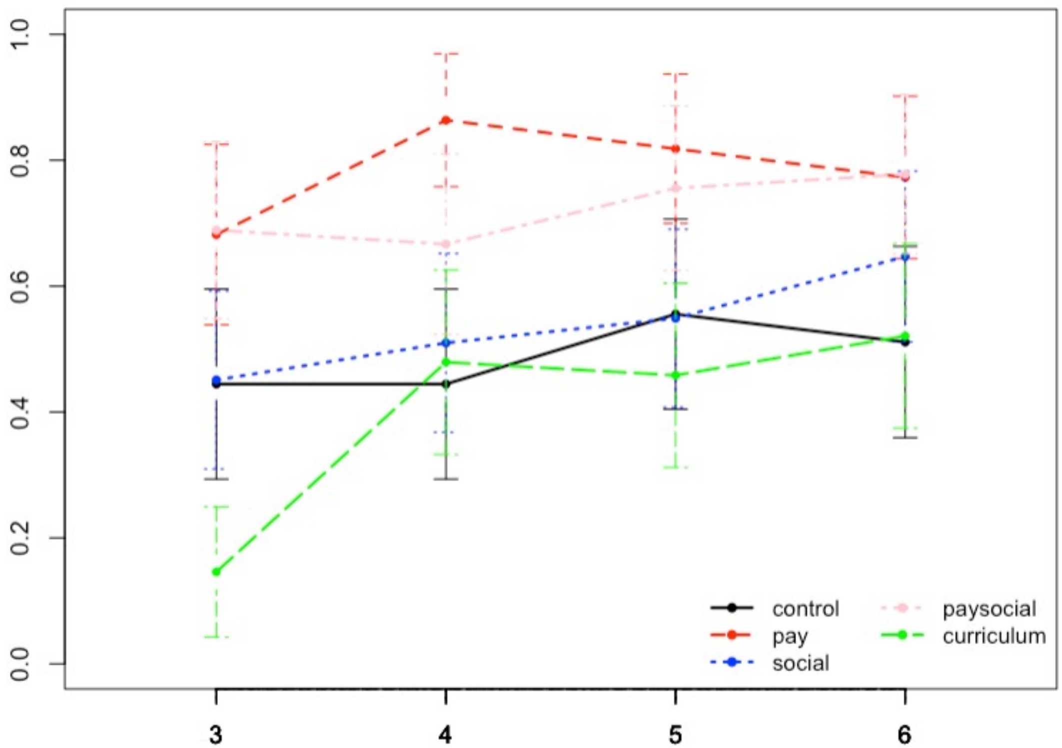
Amazon Mechanical Turk, N = 1,011

mean age 34.9, 60% female

In Progress Improving Compliance

Interventions based on incentives, social information, and pace

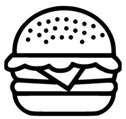
Fraction of participants complying with the optimal tip

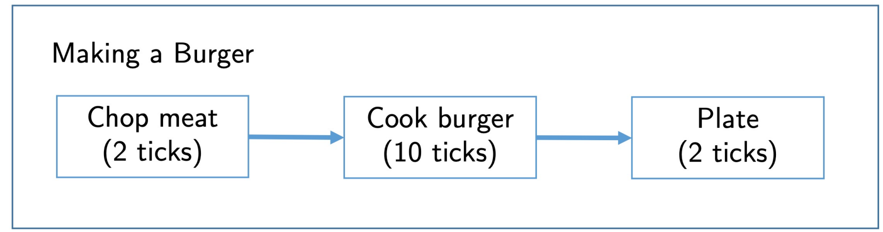


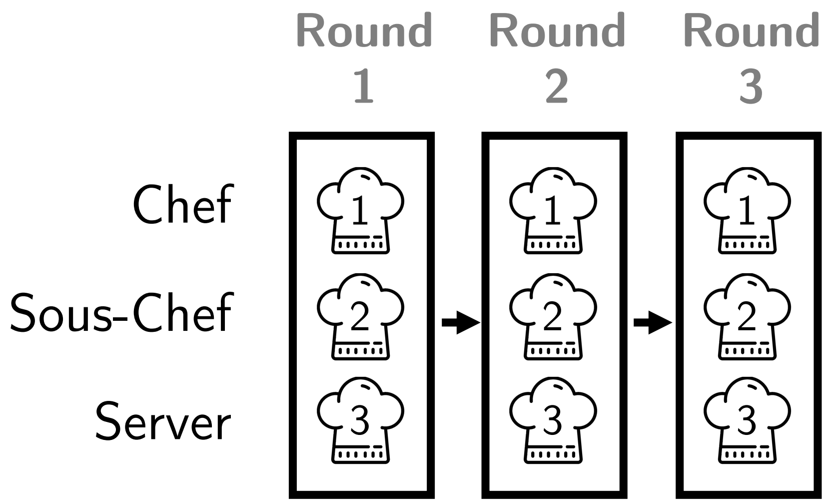
Round

Amazon Mechanical Turk, N = 250

Design Normal Configuration

 x 4 within 50 ticks





Amazon Mechanical Turk, N = 183 ⟵ **Phase I**

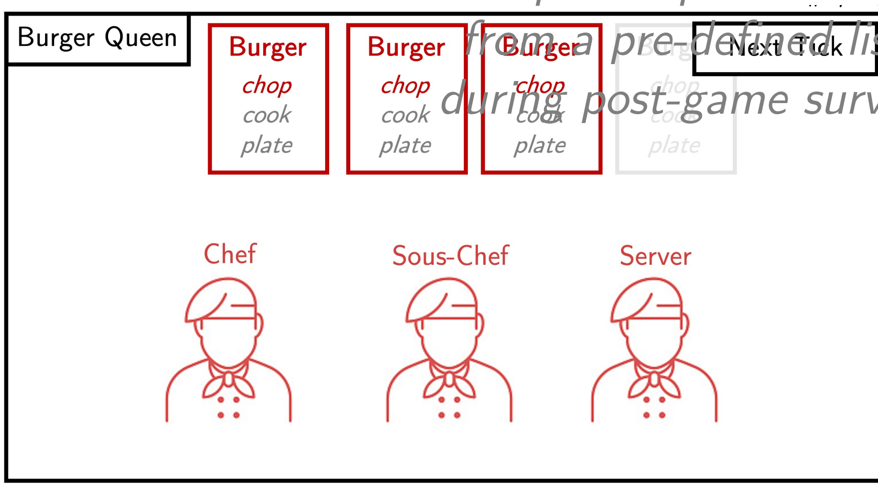
mean age 34.6, 57% female

Phase I Inferred Tips

Control	Algorithm	Human	Baseline
- No tip -	Chef should never plate	Strategically leave some worker(s) idle	Chef should chop once

Tip:

Most frequent tipchosen by users from a pre-defined list during post-game survey



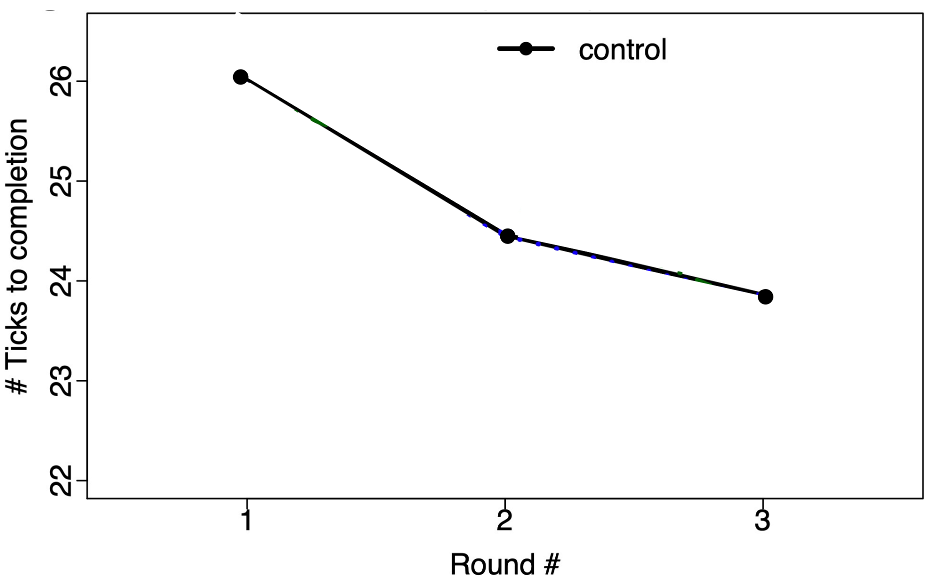
Phase II

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

Amazon Mechanical Turk, N = 183
mean age 34.6, 57% female

Results People Improve Over Time

\# Ticks to completion

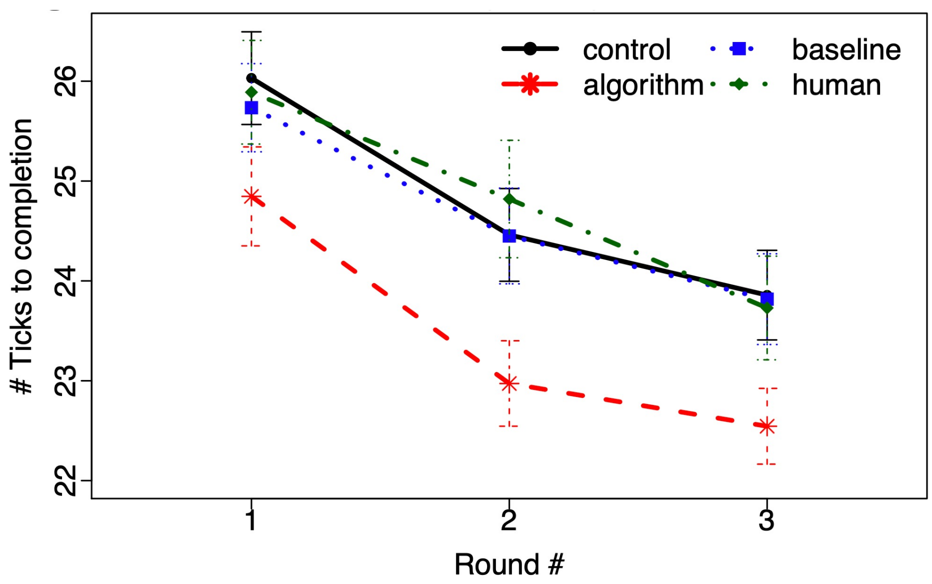


Amazon Mechanical Turk, N = 1,317

mean age 33.3, 51% female

Results Our Tip Improves Performance

Ticks to completion



One-sided T-Tests

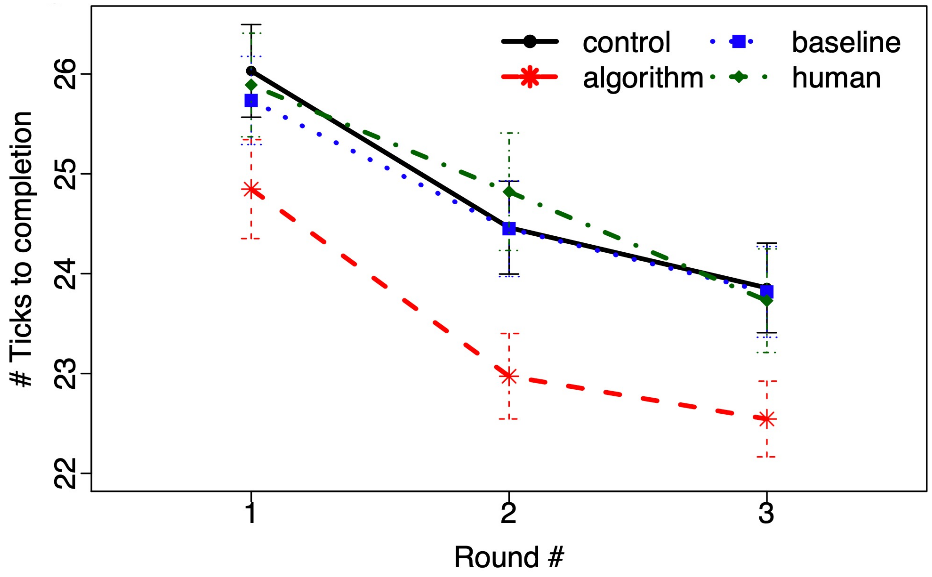
Algorithm *beats* Control ($p < 0.0001$)
Algorithm *beats* Human ($p = 0.0002$)
Algorithm *beats* Baseline ($p < 0.0001$)

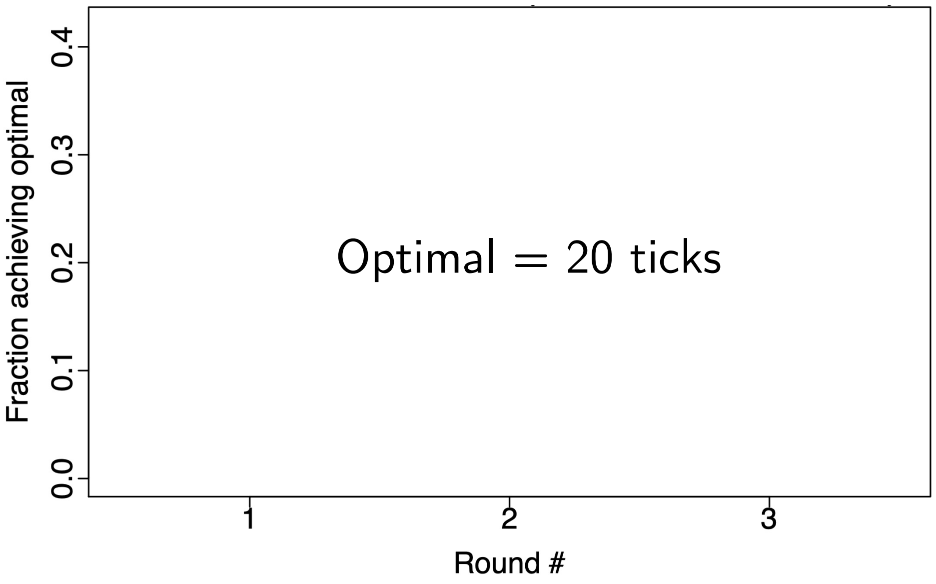
Amazon Mechanical Turk, N = 1,317
mean age 33.3, 51% female

Results

Ticks to completion

Fraction achieving optimal



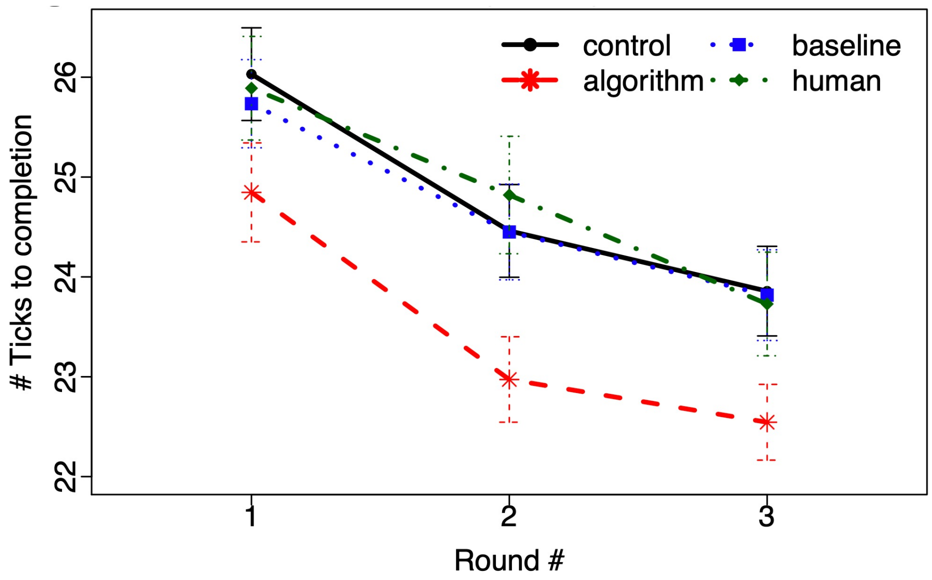


Amazon Mechanical Turk, N = 1,317

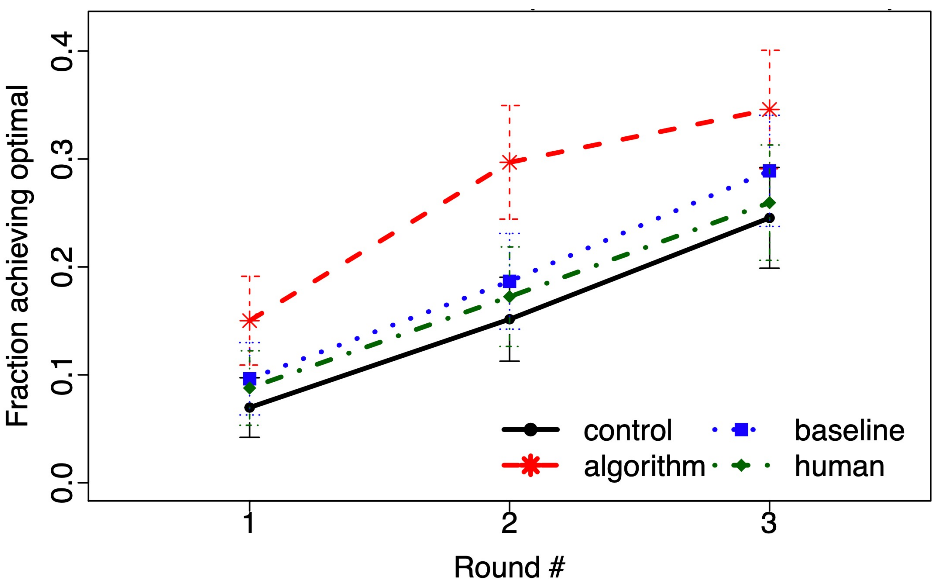
mean age 33.3, 51% female

Results Our Tip Helps Reach Optimal

Ticks to completion



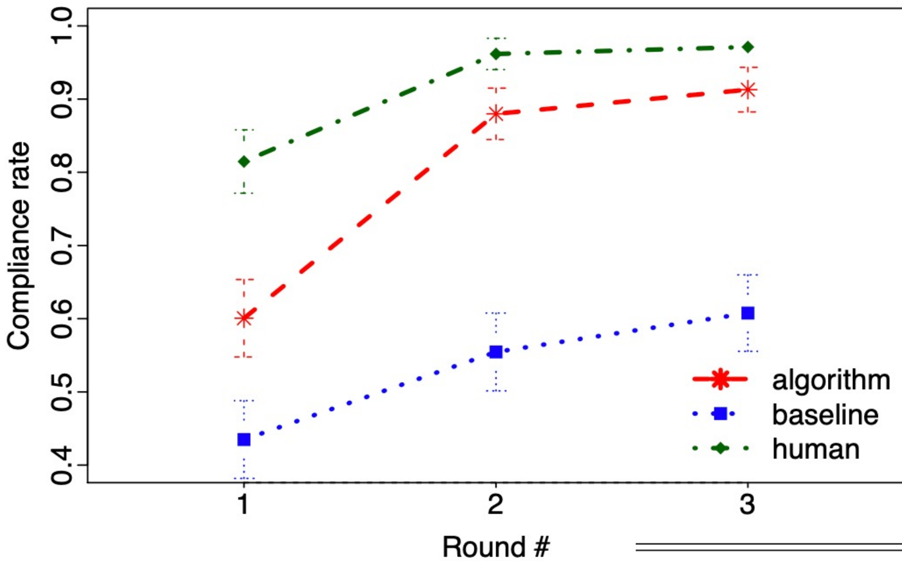
Fraction achieving optimal



Amazon Mechanical Turk, N = 1,317

mean age 33.3, 51% female

Results More Compliance over Time



Normal	Algorithm *"Chef shouldn't plate"*	Baseline *"Chef chops once"*	Human *"Leave some idle"*
(N1) Positive	25.87%	16.33%	29.23%
(N2) Negative	4.20%	5.44%	1.92%
(N3) Neutral	53.85%	51.70%	48.08%

Amazon Mechanical Turk, N = 1,317
mean age 33.3, 51% female

Discussion

- Compliance depends on whether human understands how to operationalize the tip
- Adoption of tips takes time
- Even optimal tips that focus on non-consequential actions can hurt performance

Algorithm

+ “It’s very helpful. It made me focus on making sure the server cooked more even if that was not his obvious strength.”

+ “I did not listen to it at first because I didn’t believe that it would actually help but it did.”

- “I think it was a bad tip. I couldn’t figure out how to incorporate it successfully.”

“I didn’t use it for round 1, used it for round 2 and it made me do worse, so round 3 I tried it again and was still unable to do well, so round 4 I ignored it.”

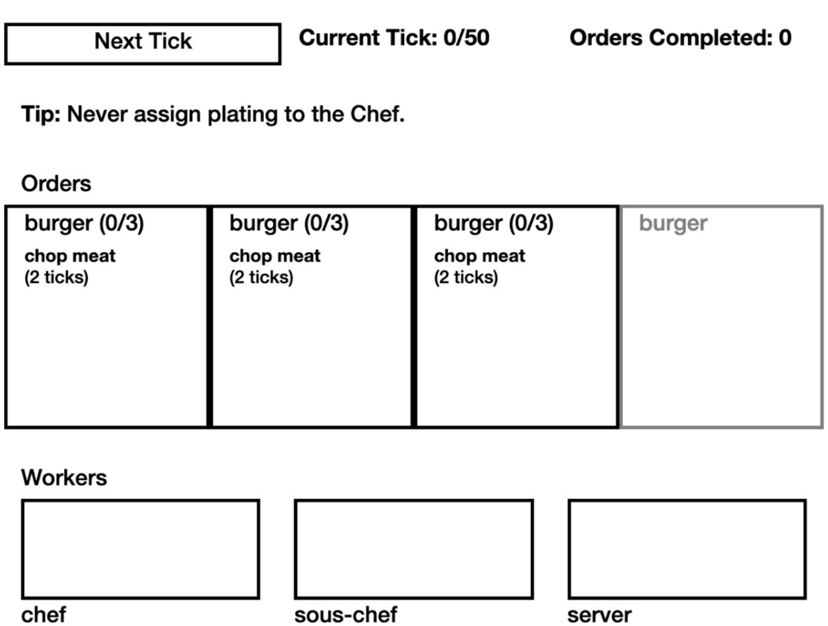
Human

+ “I thought it was smart and I used it exclusively.”

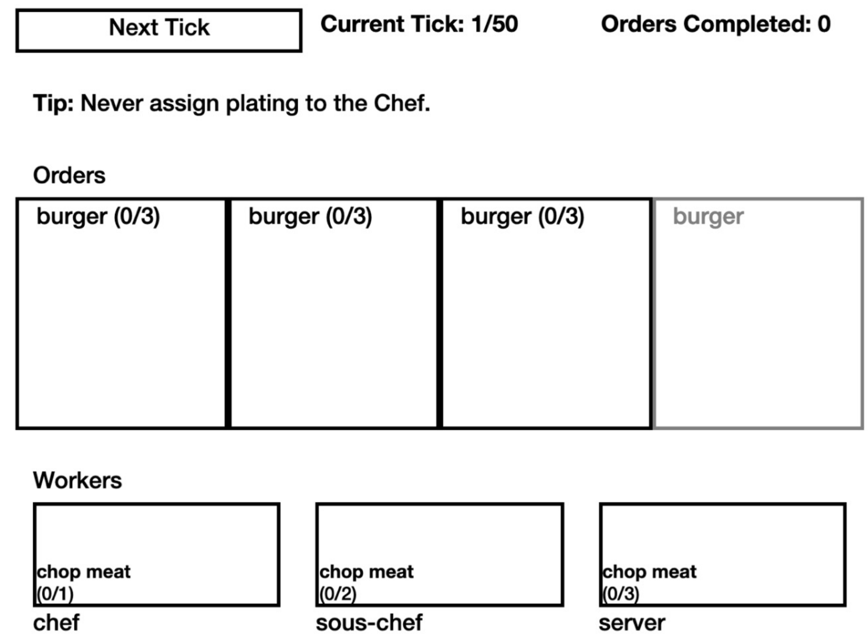
+ “It helped because she could cook one burger but any more than that and your ticks would be too high.”

- “It stunk honestly. The server takes forever to cook.”

- “I used the tip but I don’t think it was helpful. The server took long to cook.”

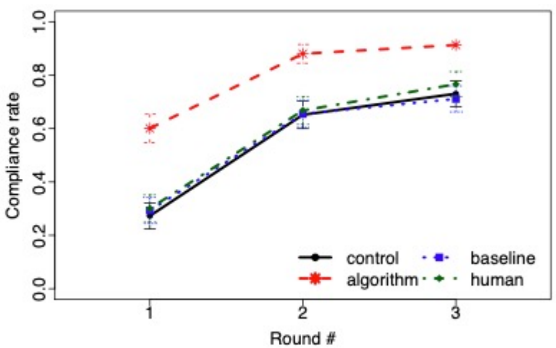


(a) The initial state where users observe available subtasks, median times to completion, and three idle virtual workers. The interface also shows the current tick, time limit, current progress, and potential tip.

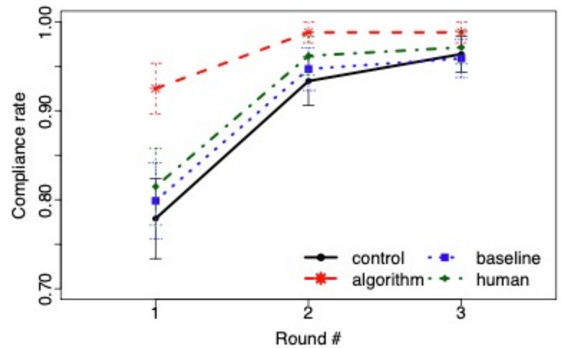


(b) The next state after all three previously available subtasks were assigned to the virtual workers and the true completion times were realized, revealing different levels of virtual workers' skills.

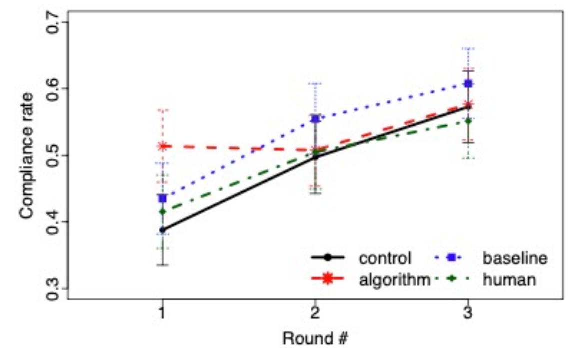
Learning Beyond Tips



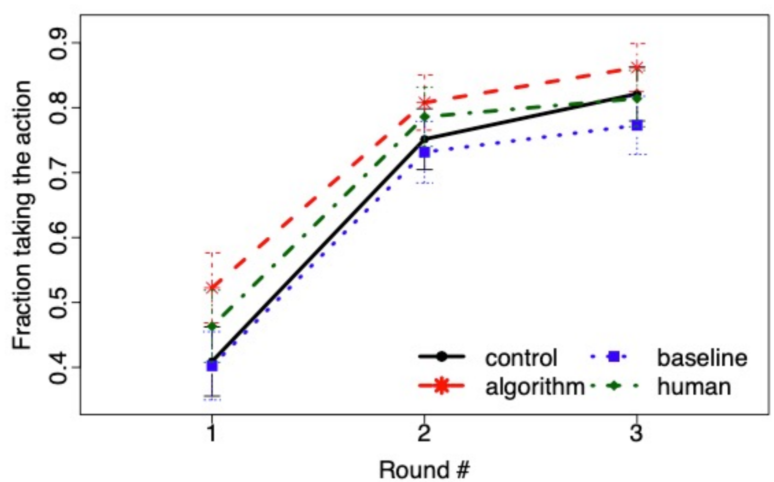
(a) Algorithm: "Chef shouldn't plate"



(b) Human: "Leave some idle"



(c) Baseline: "Chef chops once"

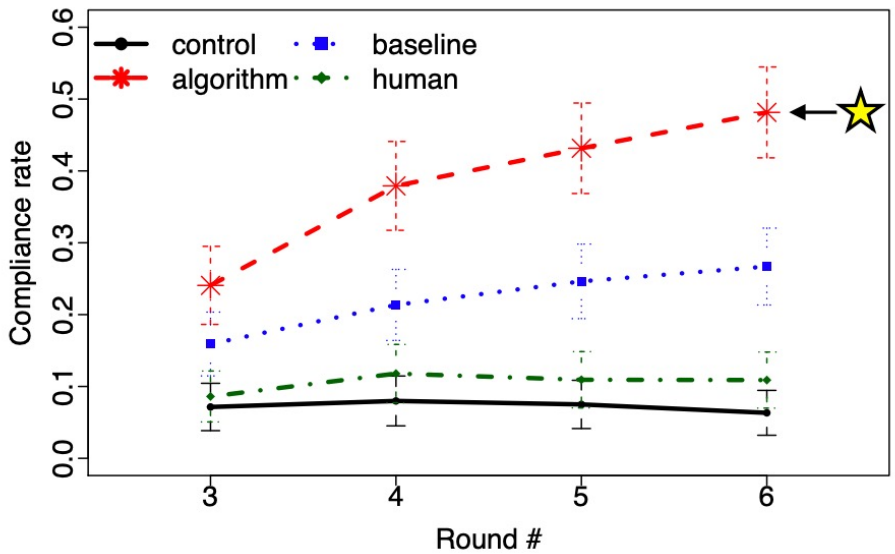


(a) Fully-staffed: "Server shouldn't cook"

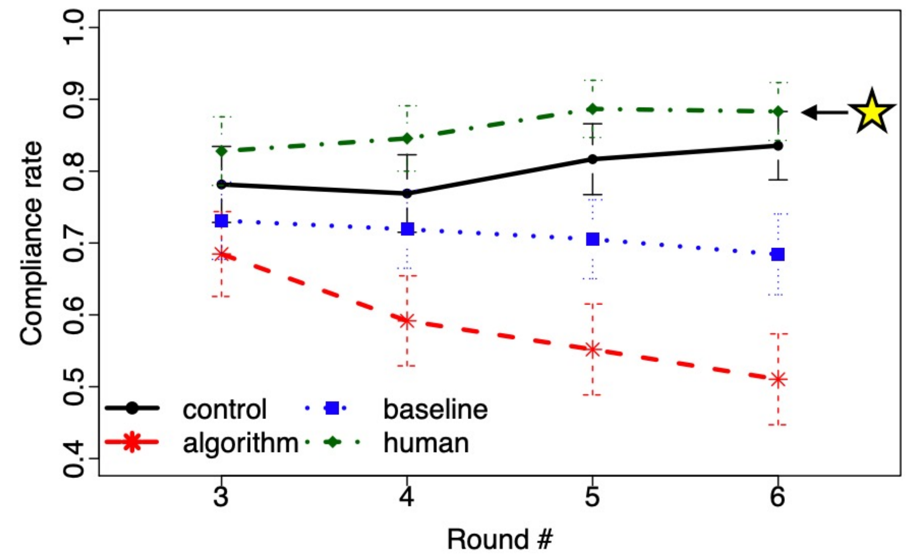
Optimal Policies

Fully-staffed scenario: In this scenario, the participant has access to all three virtual workers. The optimal number of steps needed to complete this scenario is 20 ticks. The key insights to achieving optimal performance are: (i) all three workers should be assigned to chopping in the first time step, (ii) the chef must cook three of the burgers and the sous-chef must cook one (i.e., the second burger), (iii) the server should never cook and must be kept idle when the third burger becomes available for cooking; they should instead wait to be assigned to plating the first cooked burger, (iv) the chef should never plate, (v) the sous-chef must plate exactly one of the burgers, and (vi) none of the three workers should be left idle except in the previous cases.

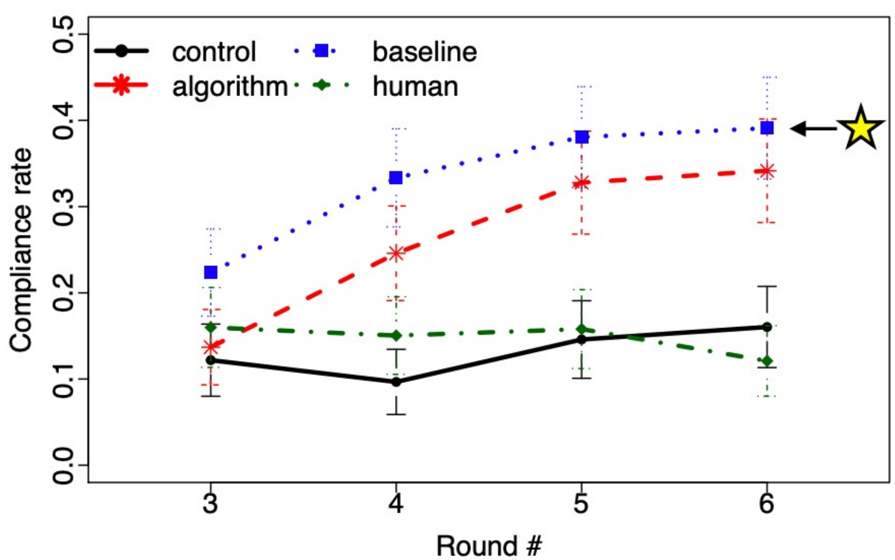
Understaffed scenario: In this scenario, the participant has access to only two virtual workers—namely, the sous-chef and the server. The optimal number of steps needed to complete this scenario is 34 ticks. The keys insights to achieving the optimal performance are: (i) both workers should be assigned to chopping in the first time step, (ii) the sous-chef and the server must cook two burgers each, even though the server is very slow at cooking, (iii) the sous-chef must choose chopping over cooking after finishing her first chopping task, (iv) the server's first three tasks must be chopping, cooking, and cooking, in that order, (v) the sous-chef must chop three of the four burgers and the server must chop one, (vi) both workers must plate two burgers each, even though the sous-chef is slower at plating than the server, (vii) the second cooked burger must not be served until the third and fourth burgers are cooked, and (viii) both workers must be kept busy at all times.



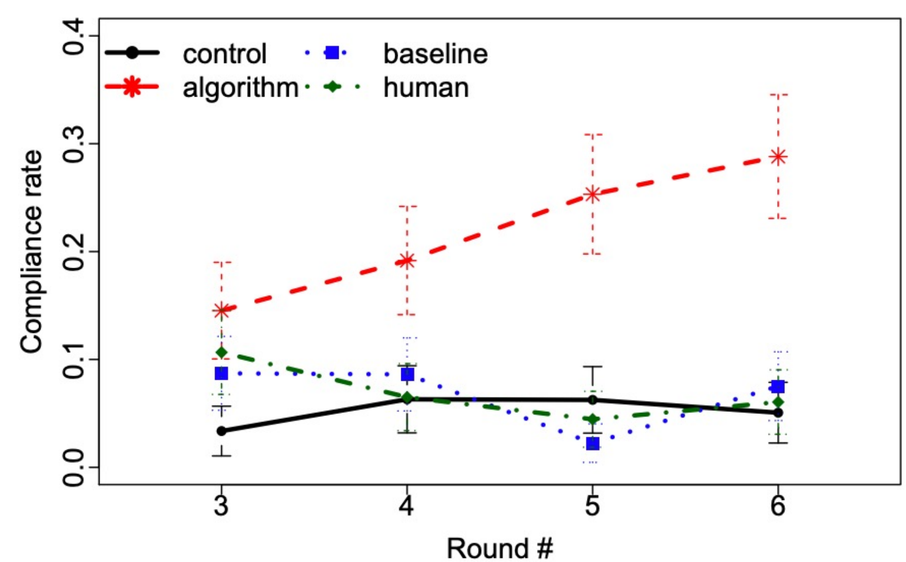
(a) Algorithm Tip: "Server cooks twice"



(b) Human Tip: "Server cooks once"



(c) Baseline Tip: "Sous-chef plates twice"



(d) Unshown Tip: "Server chops once"

Additional Results

	Phase I: Normal	Phase II: Normal	Phase I: Disrupted	Phase II: Disrupted
Total	183	1,317	172	1,011
Mean age [range]	34.6 [18, 76]	33.3 [18, 74]	34 [19, 76]	34.9 [16, 84]
Female	57.38%	51.03%	61.63%	60.14%
$\geq$ 2-year degree	73.22%	67.73%	77.91%	70.43%
Median duration	18.82 minutes	20.50 min	27.80 min	26.80 min
Found the game difficult	60.66%	50.04%	70.93%	64.99%
Never played similar games	45.36%	43.82%	46.51%	43.52%

Normal	Algorithm *"Chef shouldn't plate"*	Baseline *"Chef chops once"*	Human *"Leave some idle"*
(N1) Positive	25.87%	16.33%	29.23%
(N2) Negative	4.20%	5.44%	1.92%
(N3) Neutral	53.85%	51.70%	48.08%

Table 7: Participants' coded feedback on the provided tips (normal configuration).

Disrupted	Algorithm *"Server cooks twice"*	Baseline *"Sous-chef plates twice"*	Human *"Server cooks once"*
(D1) Positive	23.10%	10.19%	25.87%
(D2) Negative	33.10%	37.58%	16.78%
(D3) Neutral	32.76%	42.99%	47.90%

Table 8: Participants' coded feedback on the provided tips (disrupted configuration).

Next Steps

- Testing the algorithm with real data
- Extend to teams: collaboration/competition
- Combat algorithm aversion/improve compliance