

# Improving Human Sequential Decision-Making with Reinforcement Learning

Park Sinchaisri  
UC Berkeley Haas

Stanford GSB OIT Seminar  
March 6, 2024



# Research Overview

1

## Operations for the Future of Work

# Research Overview

1

## Operations for the Future of Work

### Gig Workers' Decision-Making

- Behavioral & economic drivers on workers' labor decisions/scheduling  
(with Gad Allon, Maxime Cohen, M&SOM 2023)
- Multihoming, incentive schemes  
(with Gad Allon, Maxime Cohen, Ken Moon, under review at Management Science)

### Gig Workers' Learning

- Optimizing task selection/assignment to improve learning  
(with Shunan Jiang, work in progress)

# Research Overview

1

## Operations for the Future of Work

### Gig Workers' Decision-Making

- Behavioral & economic drivers on workers' labor decisions/scheduling  
(with Gad Allon, Maxime Cohen, M&SOM 2023)
- Multihoming, incentive schemes  
(with Gad Allon, Maxime Cohen, Ken Moon, under review at Management Science)

2

## Human-AI Interfaces in Operations

### Gig Workers' Learning

- Optimizing task selection/assignment to improve learning  
(with Shunan Jiang, work in progress)

# Research Overview

1

## Operations for the Future of Work

### Gig Workers' Decision-Making

- Behavioral & economic drivers on workers' labor decisions/scheduling  
(with Gad Allon, Maxime Cohen, M&SOM 2023)
- Multihoming, incentive schemes  
(with Gad Allon, Maxime Cohen, Ken Moon, under review at Management Science)

### Gig Workers' Learning

- Optimizing task selection/assignment to improve learning  
(with Shunan Jiang, work in progress)

2

## Human-AI Interfaces in Operations

### Tips for Sequential Decision-Making

# Research Overview

1

## Operations for the Future of Work

### Gig Workers' Decision-Making

- Behavioral & economic drivers on workers' labor decisions/scheduling  
(with Gad Allon, Maxime Cohen, M&SOM 2023)
- Multihoming, incentive schemes  
(with Gad Allon, Maxime Cohen, Ken Moon, under review at Management Science)

### Gig Workers' Learning

- Optimizing task selection/assignment to improve learning  
(with Shunan Jiang, work in progress)

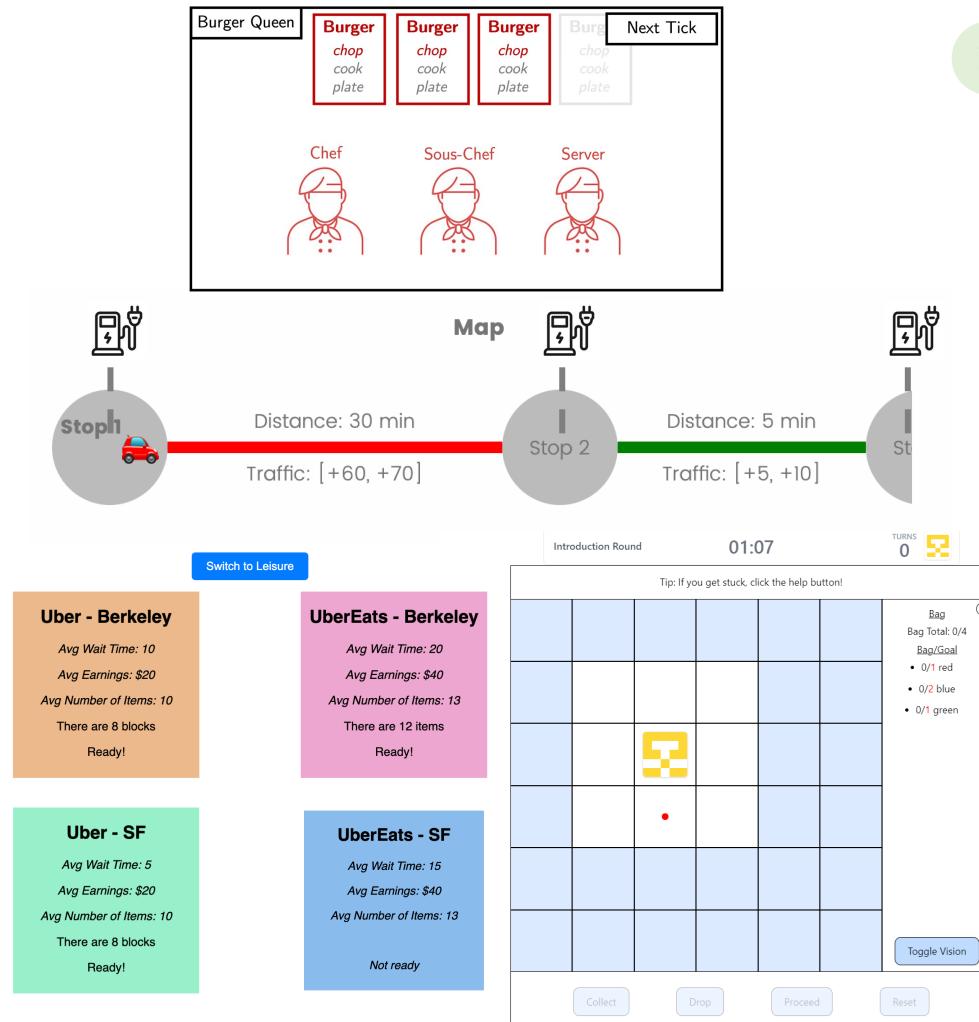
2

## Human-AI Interfaces in Operations

### Tips for Sequential Decision-Making

- Learning best practices from data and mining simple advice  
(with Hamsa Bastani, Osbert Bastani, Minor Revision at Management Science)
- Precision of advice in uncertain environment  
(with Philippe Blaettchen, preparing for submission)
- Characterizing non-adoption  
(with David Lee, preparing for submission)
- Generative AI and productivity  
(with Sam Keppler, Clare Snyder, under review at CSCW)
- Pricing competing products  
(with Olivia Natan, work in progress)

# Research Overview



## 2 Human-AI Interfaces in Operations

### Tips for Sequential Decision-Making

- Learning best practices from data and mining simple advice  
(with Hamsa Bastani, Osbert Bastani, Minor Revision at Management Science)
- Precision of advice in uncertain environment  
(with Philippe Blaettchen, preparing for submission)
- Characterizing non-adoption  
(with David Lee, preparing for submission)
- Generative AI and productivity  
(with Sam Keppler, Clare Snyder, under review at CSCW)
- Pricing competing products  
(with Olivia Natan, work in progress)

# Learning is Costly

**2+ years**

to be fully productive

**\$1,286/worker**

training expenses

- Training Magazine 2019

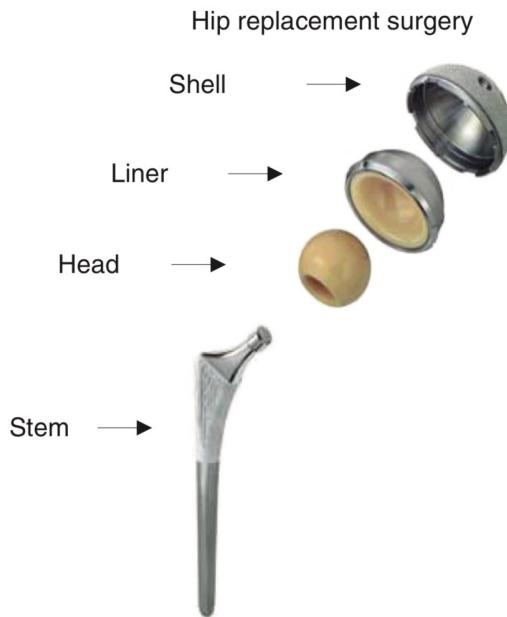
# Learning is Costly

**2+ years**  
to be fully productive

**\$1,286/worker**

training expenses

- Training Magazine 2019



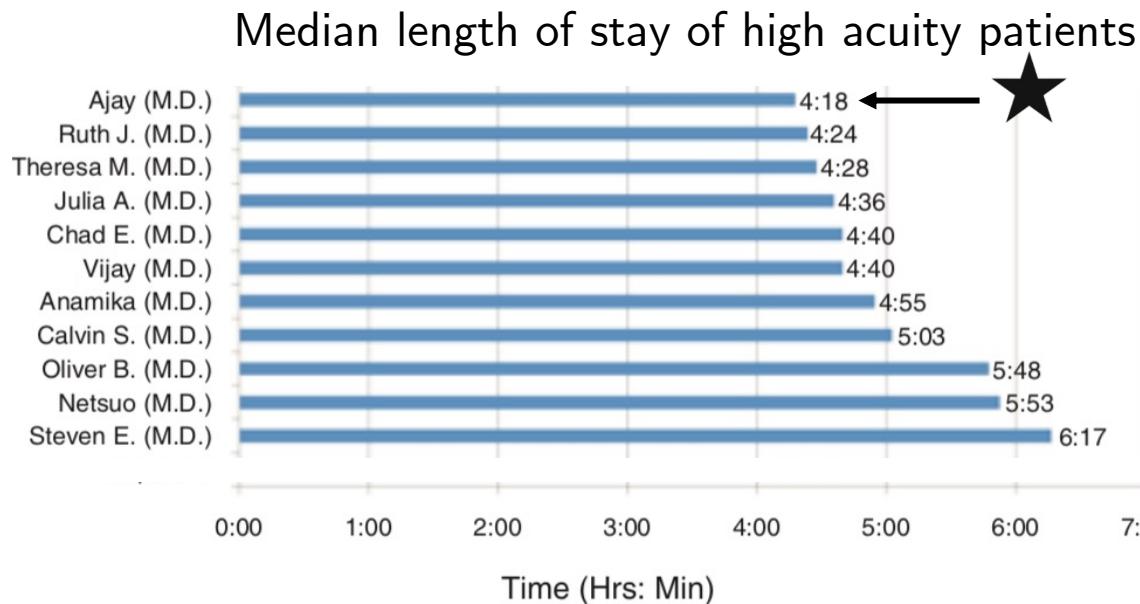
New device = **+32.4%**  
surgery duration

- Ramdas et al. 2018

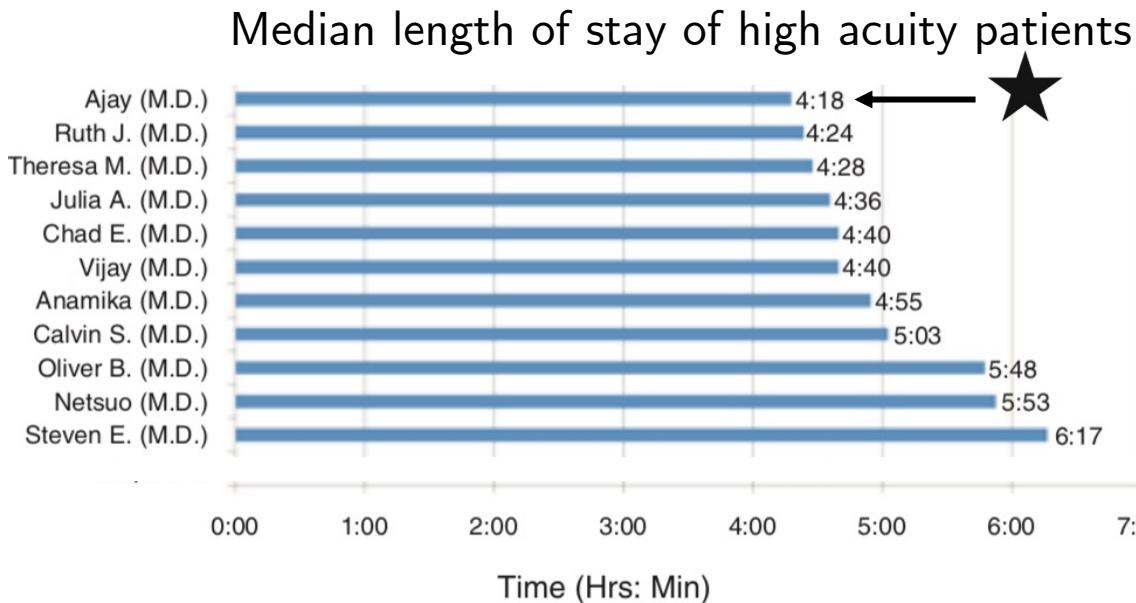
Also – Tucker et al 2002, Ibanez et al 2017, Gurvich et al 2019,  
Bavafa & Jonasson 2020, Bloom et al 2020, ...

# Learning from Experts

# Learning from Experts



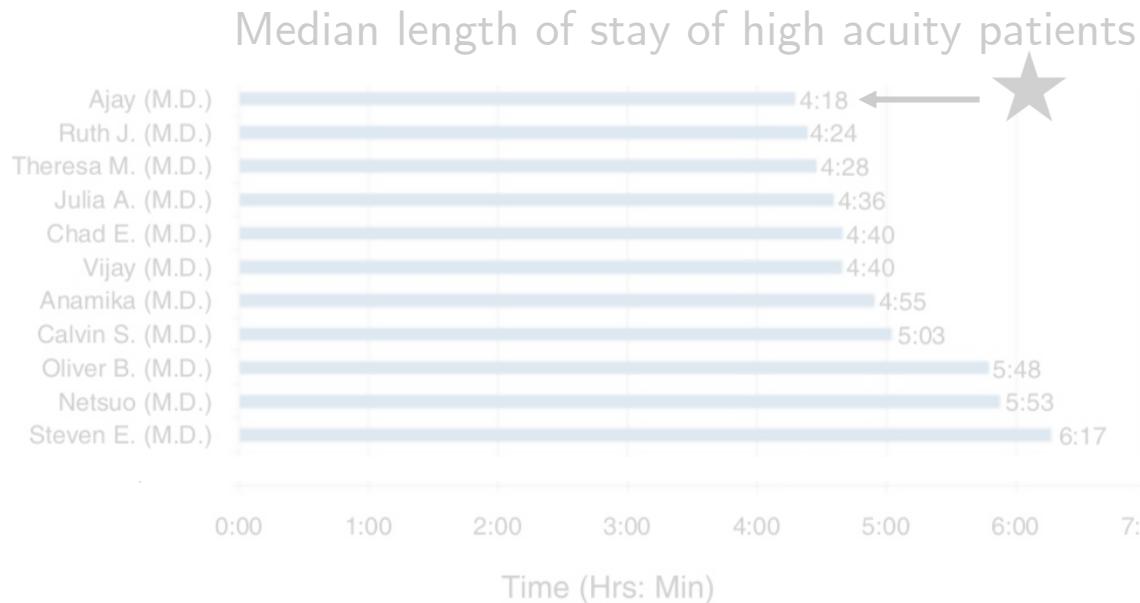
# Learning from Experts



+10.9%  
productivity  
- Song et al. 2018

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

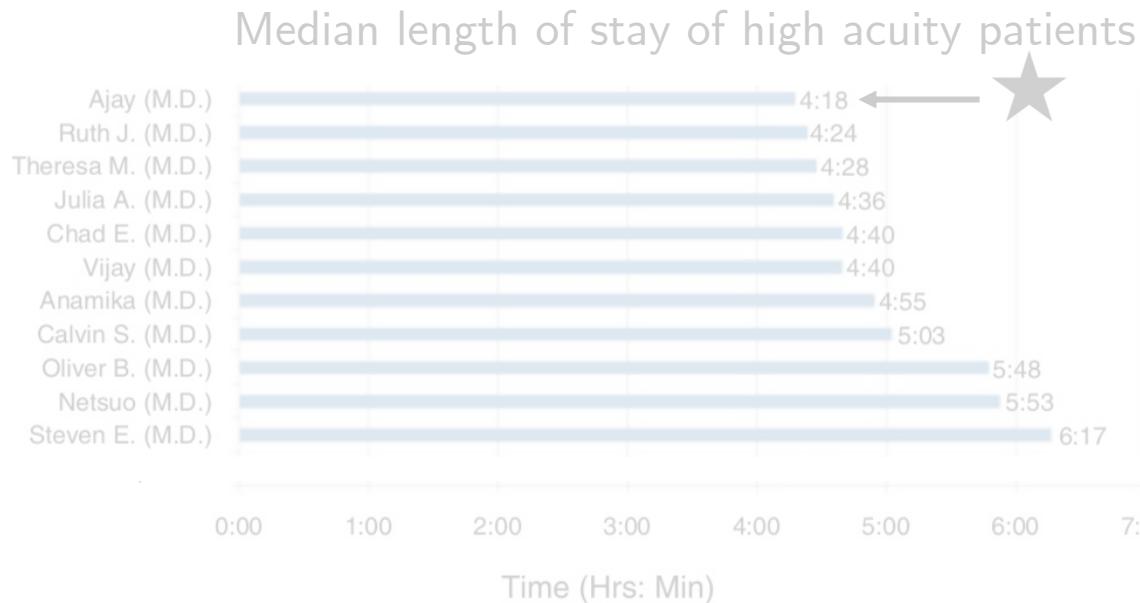
# Learning from Experts



+10.9%  
productivity  
- Song et al. 2018



# Learning from Experts



+10.9%  
productivity  
- Song et al. 2018

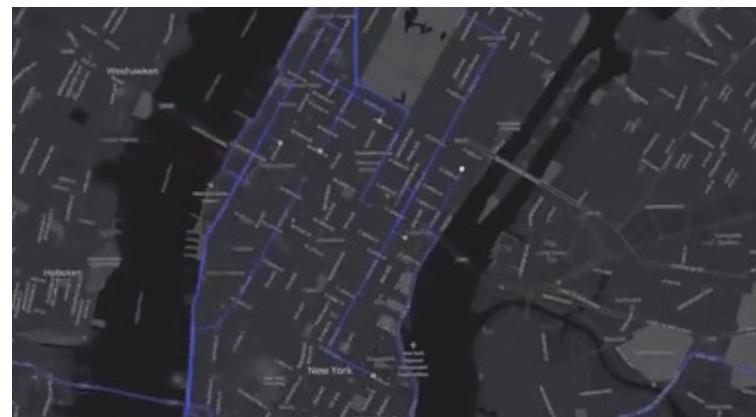


# Trace Data is Everywhere

## Physicians

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



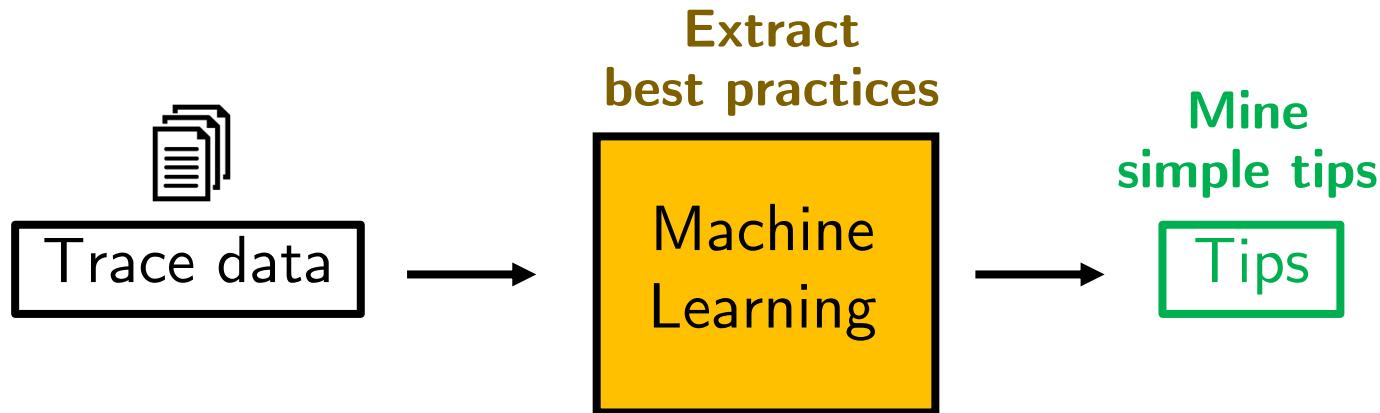
Trace data



Tips

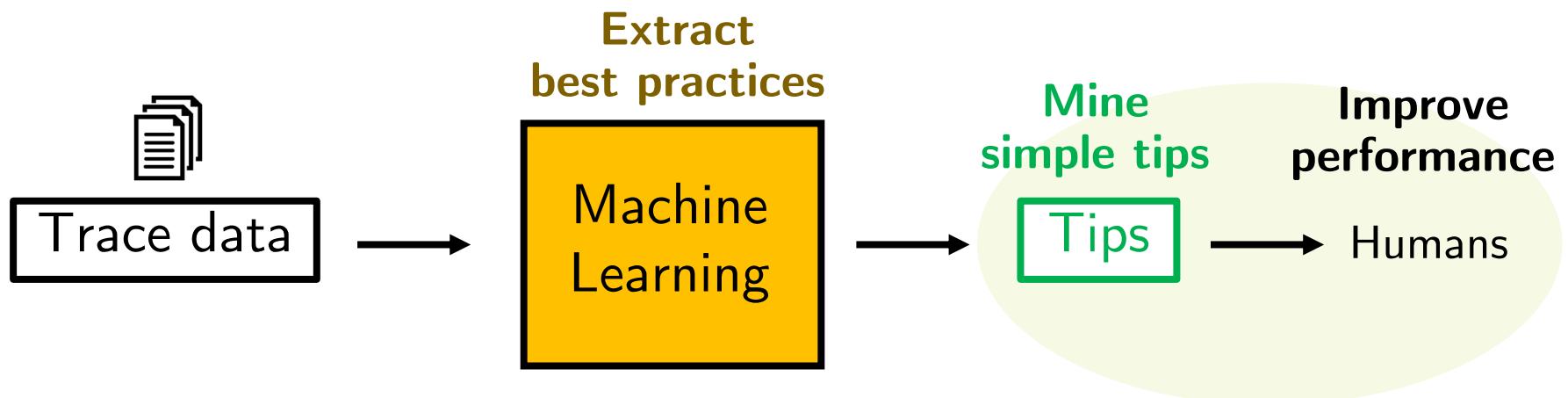
# How to Help Humans

## Improve Their Decision-Making?



# How to Help Humans

## Improve Their Decision-Making?



# How to Help Humans

## Improve Their Decision-Making?

### Outline:



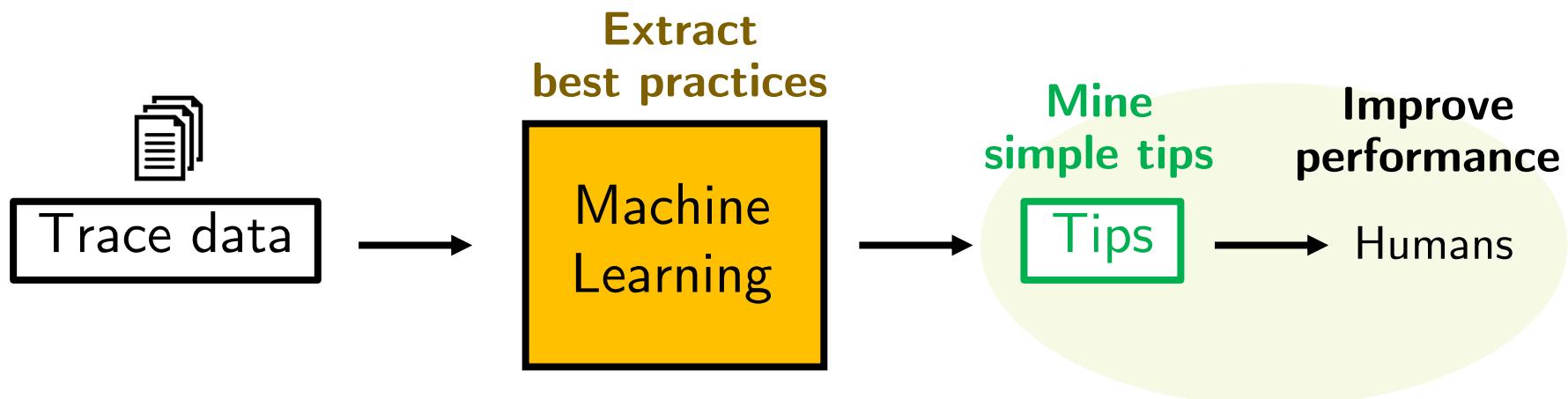
Experimental Design 🍔

Tip Inference

Results: Performance/Compliance

Improving Compliance

with Hamsa Bastani & Osbert Bastani  
*Minor Revision @ Management Science*



# How to Help Humans

## Improve Their Decision-Making?

### Outline:



Experimental Design 🍔

Tip Inference

Results: Performance/Compliance

Improving Compliance

with Hamsa Bastani & Osbert Bastani  
*Minor Revision @ Management Science*

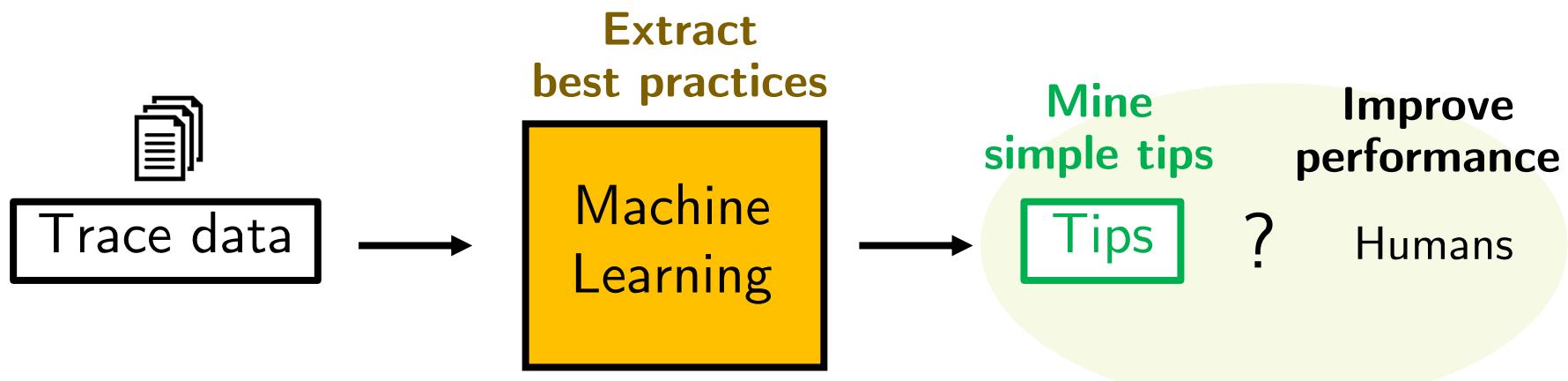
Experimental Design ⚡

+ Precision of Tips

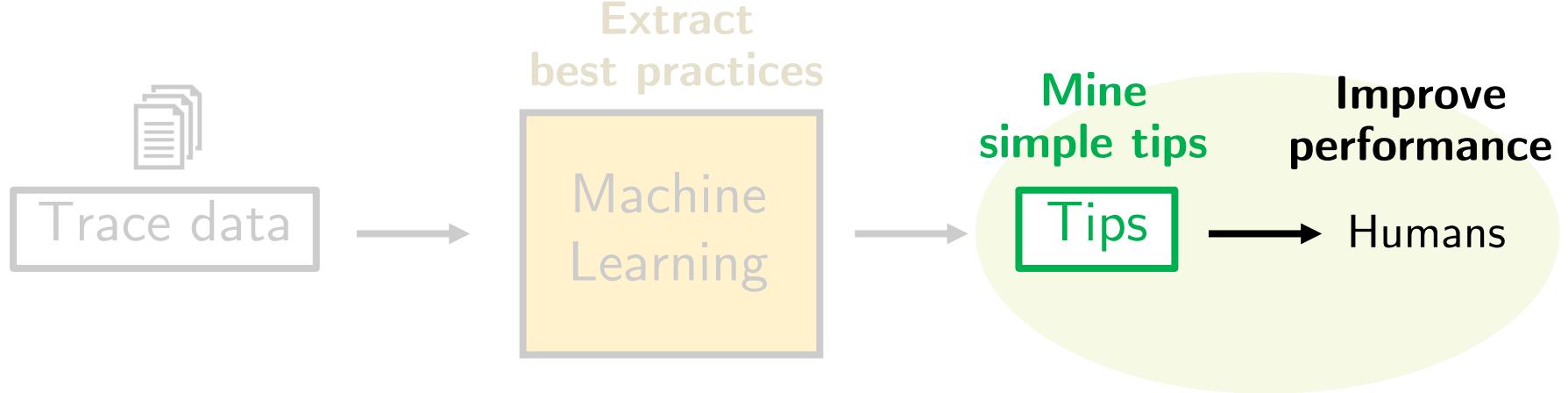
+ Learning Post-Tip



with Philippe Blaettchen  
*Preparation for submission*

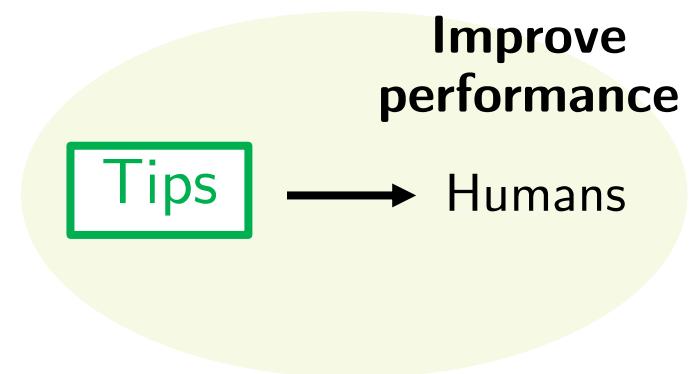


# Potential Issues



# Potential Issues

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



# Potential Issues

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

BUSINESS

## Will the public accept the fatal mistakes of self-driving cars?



By Steven Overly

February 20, 2017 at 6:48 p.m. EST

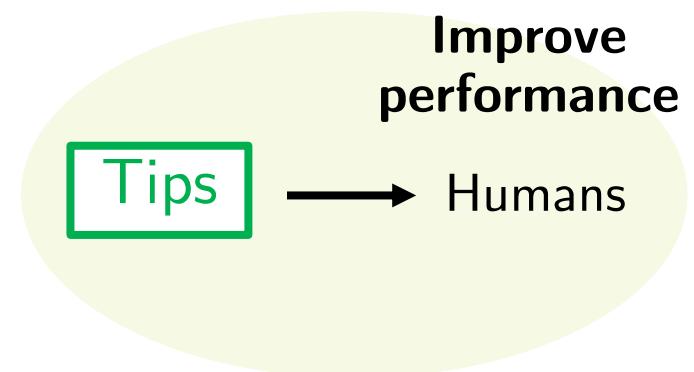
Tips

Improve performance

→ Humans

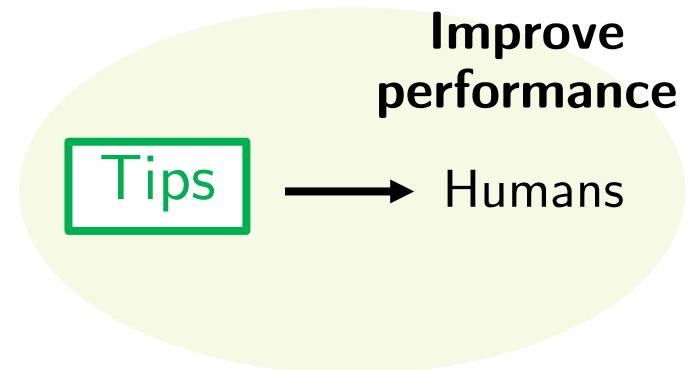
# Potential Issues

- Compliance to tips, “algorithm aversion” (e.g., Dietvorst et al 2015)
- Private information (e.g., Ibanez et al 2018, Balakrishnan et al 2022, Sun et al 2022)



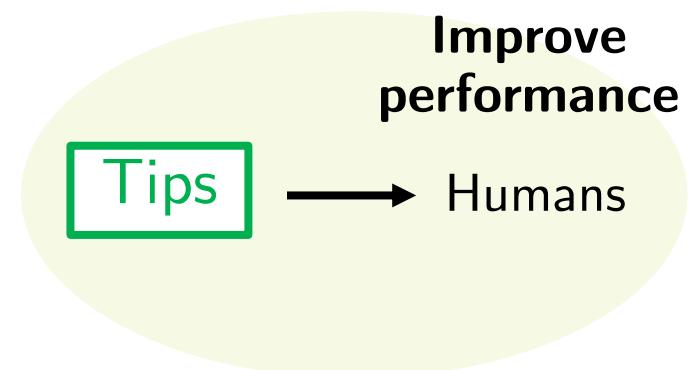
# Potential Issues

- Compliance to tips, “algorithm aversion” (e.g., Dietvorst et al 2015)
- Private information (e.g., Ibanez et al 2018, Balakrishnan et al 2022, Sun et al 2022)



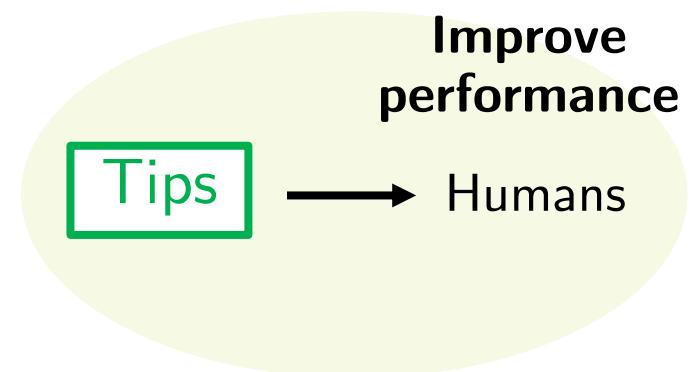
# Potential Issues

- Compliance to tips, “algorithm aversion” (e.g., Dietvorst et al 2015)
- Private information (e.g., Ibanez et al 2018, Balakrishnan et al 2022, Sun et al 2022)
- Interpretability, inability to precisely implement



# Potential Issues

- Compliance to tips, “algorithm aversion” (e.g., Dietvorst et al 2015)
- Private information (e.g., Ibanez et al 2018, Balakrishnan et al 2022, Sun et al 2022)
- Interpretability, inability to precisely implement
- Learning curve, spillovers

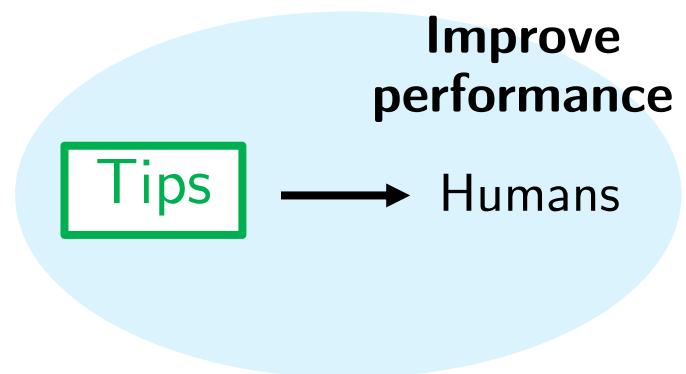


# Potential Issues

- Compliance to tips, “algorithm aversion” (e.g., Dietvorst et al 2015)
- Private information (e.g., Ibanez et al 2018, Balakrishnan et al 2022, Sun et al 2022)
- Interpretability, inability to precisely implement
- Learning curve, spillovers

## What We Did:

Controlled environment  
to observe human learning  
& decision-making



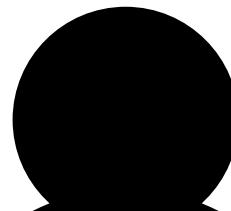
# Study 1:

## Cooking Game

Burger Queen



× 4 within 50 ticks



Participant

Pre-registered at

<https://aspredicted.org/blind.php?x=8ye5cb>

# Study 1:

## Cooking Game

Burger Queen



x 4 within 50 ticks

Making a Burger

Chop meat  
(2 ticks)

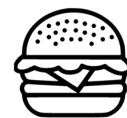
Cook burger  
(10 ticks)

Plate  
(2 ticks)

# Study 1:

## Cooking Game

Burger Queen



x 4 within 50 ticks

Chef



Sous-Chef



Server



Participant

Pre-registered at

<https://aspredicted.org/blind.php?x=8ye5cb>

# Study 1:

## Cooking Game

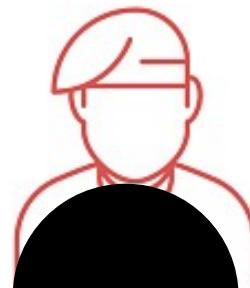
Burger Queen

Chopping:	Fast	Average	Slow
Cooking:	Fast	Average	Slow
Plating:	Slow	Average	Fast

Chef



Sous-Chef



Server



Participant

Pre-registered at

<https://aspredicted.org/blind.php?x=8ye5cb>

# **Study 1:**

# Cooking Game

Reward: 0  
Tick #1/50

# Burger Queen

# Burger

*chop*  
*cook*  
*plate*

# Burger

*chop*  
*cook*  
*plate*

# Burger

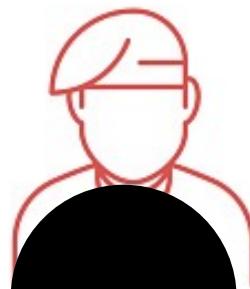
*chop*  
*cook*  
*plate*

## Next Tick

# Chef



## Sous-Chef



Server



Pre-registered at

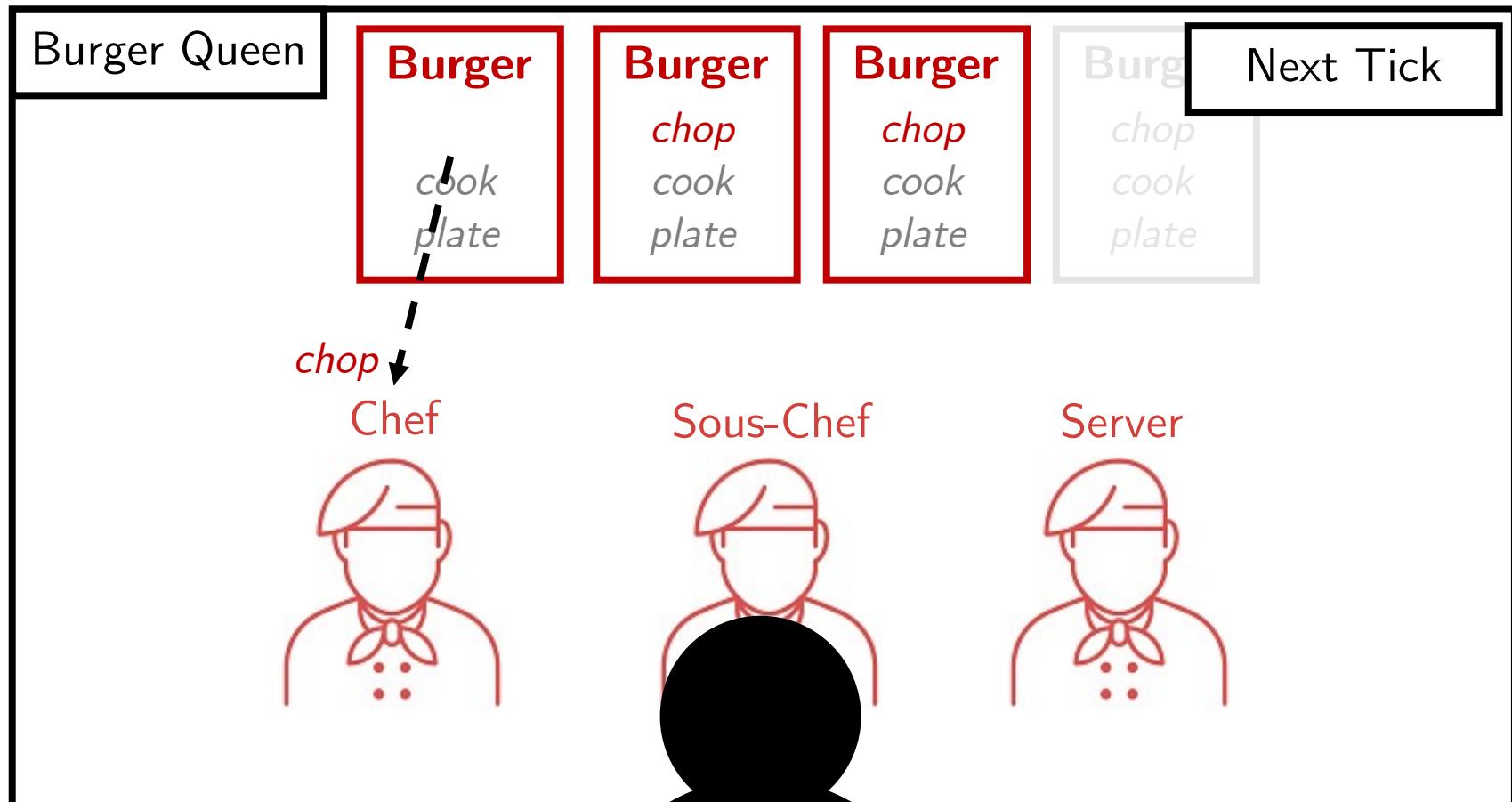
<https://aspredicted.org/blind.php?x=8ye5cb>

## Participant

# **Study 1:**

# Cooking Game

Reward: 0  
Tick #1/50



Pre-registered at

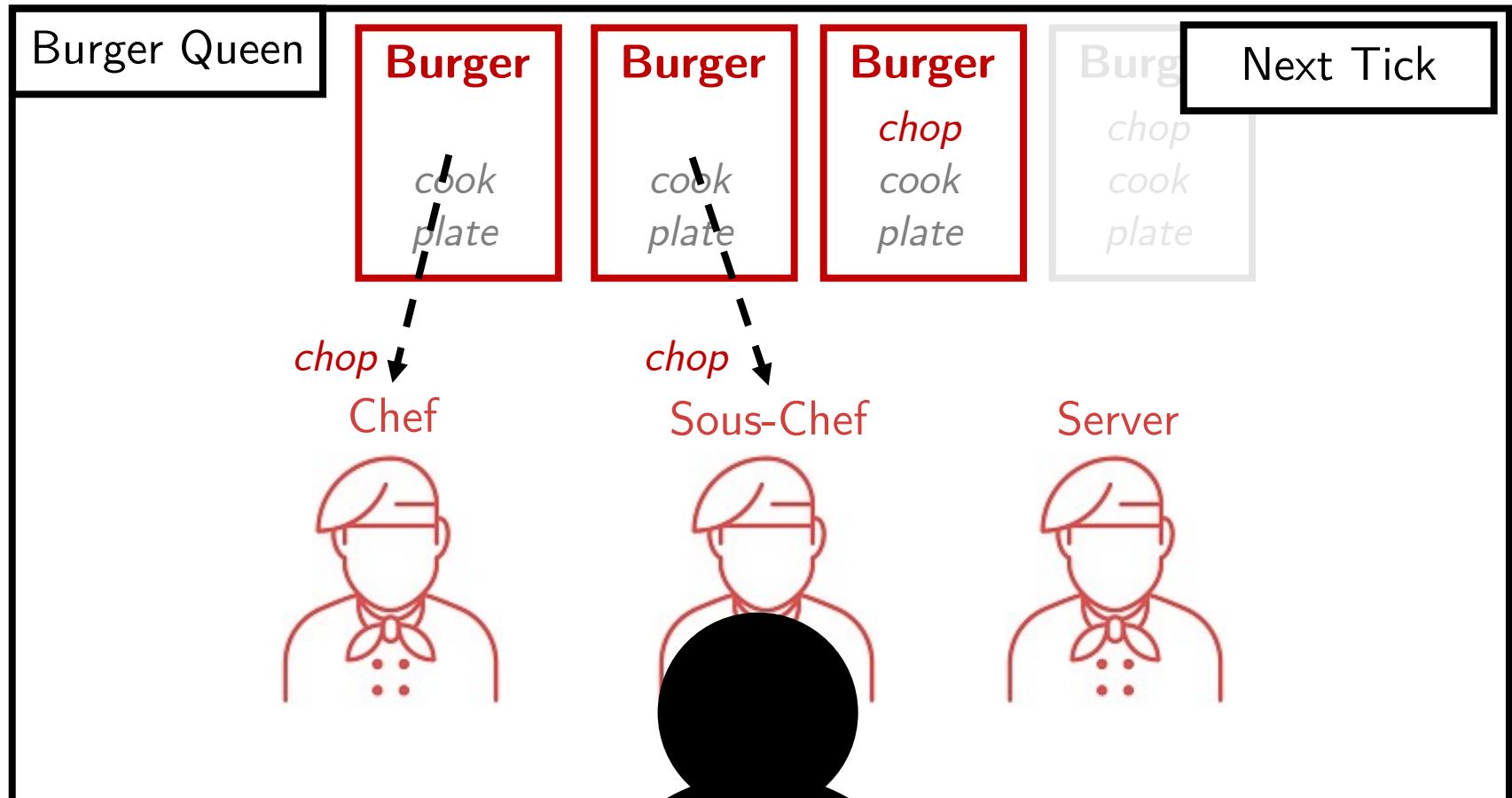
<https://aspredicted.org/blind.php?x=8ye5cb>

## Participant

# Study 1:

## Cooking Game

Reward: 0  
Tick #1/50



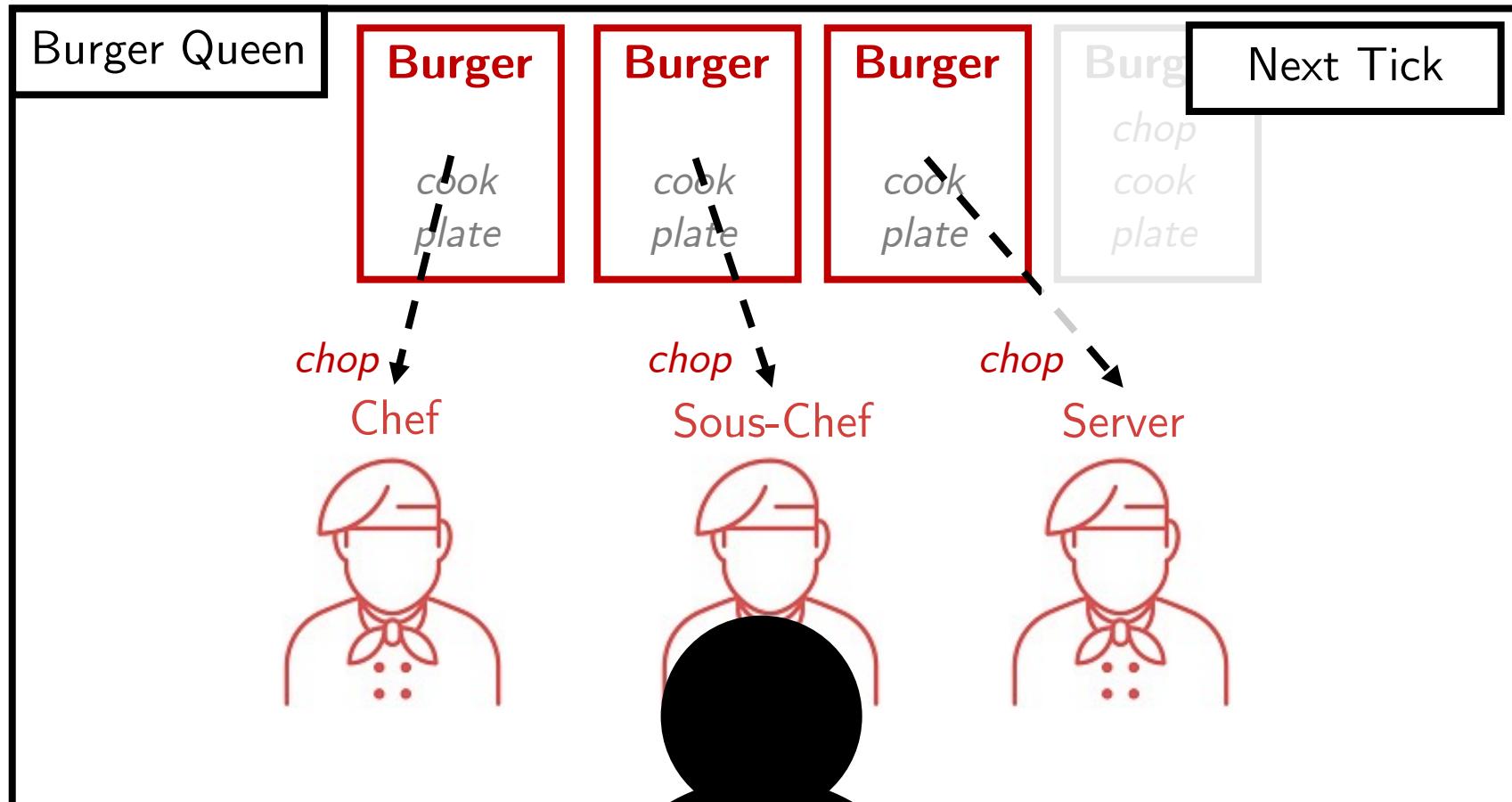
Pre-registered at

<https://aspredicted.org/blind.php?x=8ye5cb>

# Study 1:

## Cooking Game

Reward: 0  
Tick #1/50



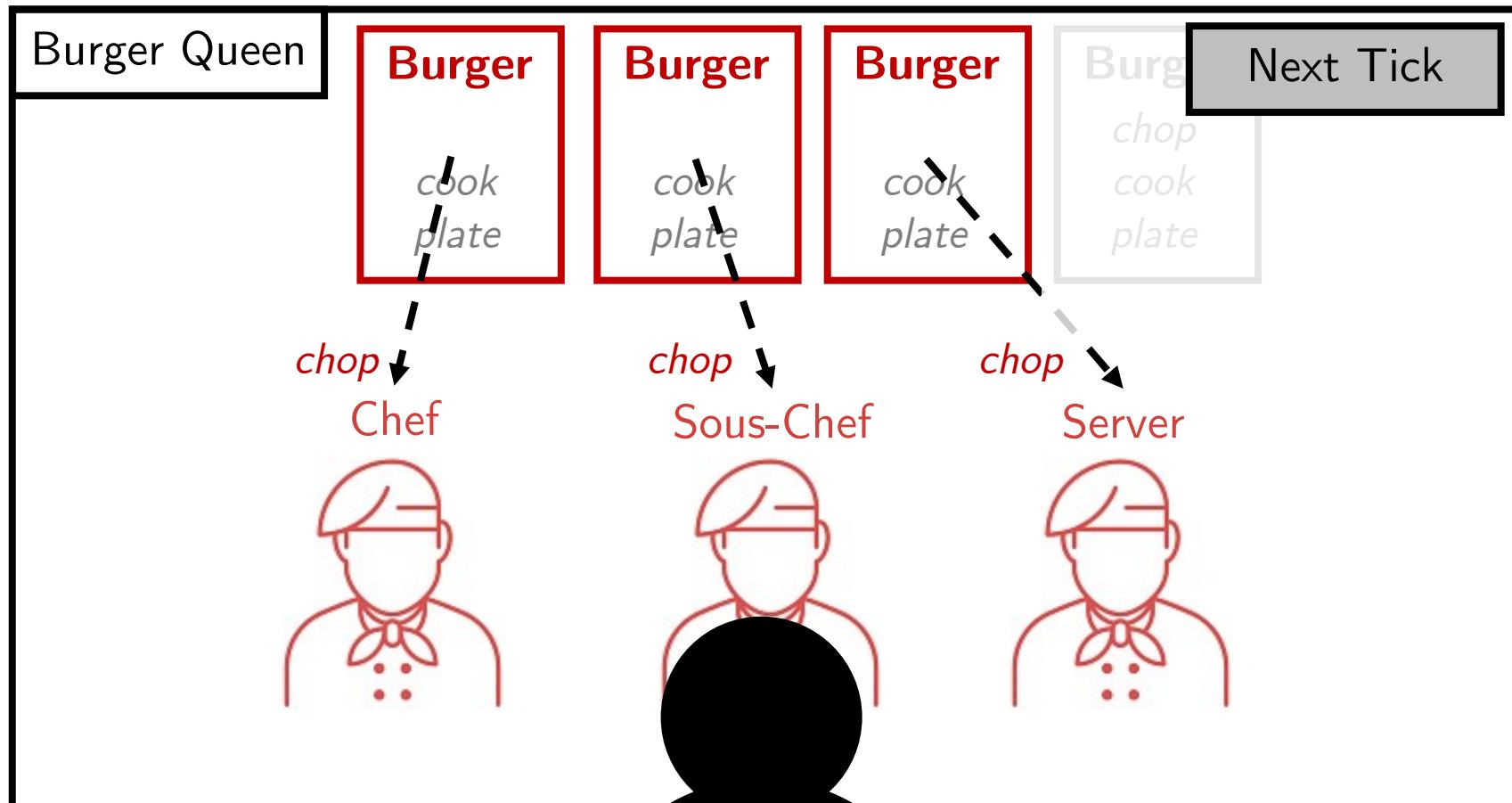
Pre-registered at

<https://aspredicted.org/blind.php?x=8ye5cb>

# Study 1:

## Cooking Game

Reward: 0  
Tick #1/50



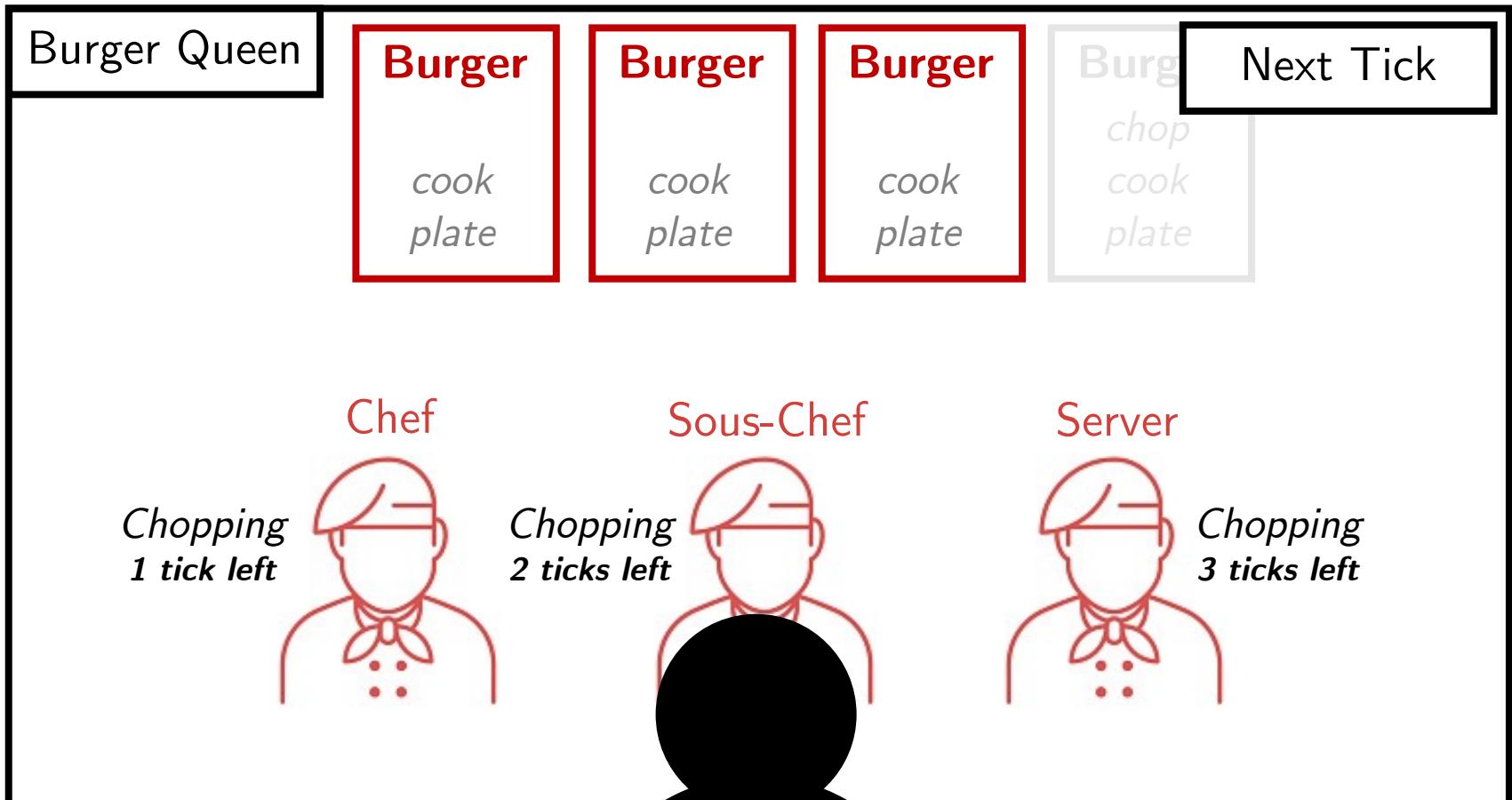
Pre-registered at

<https://aspredicted.org/blind.php?x=8ye5cb>

# Study 1:

## Cooking Game

Reward: 0  
Tick #2/50



# Study 1:

## Design

## Disruption Scenario



x 4 within 50 ticks

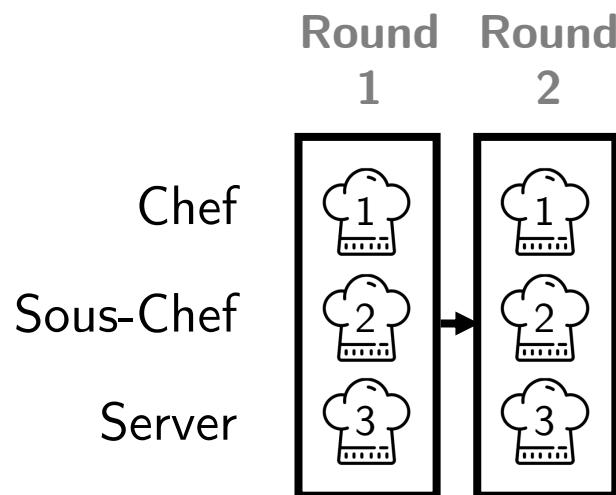
# Study 1:

## Design

## Disruption Scenario



x 4 within 50 ticks



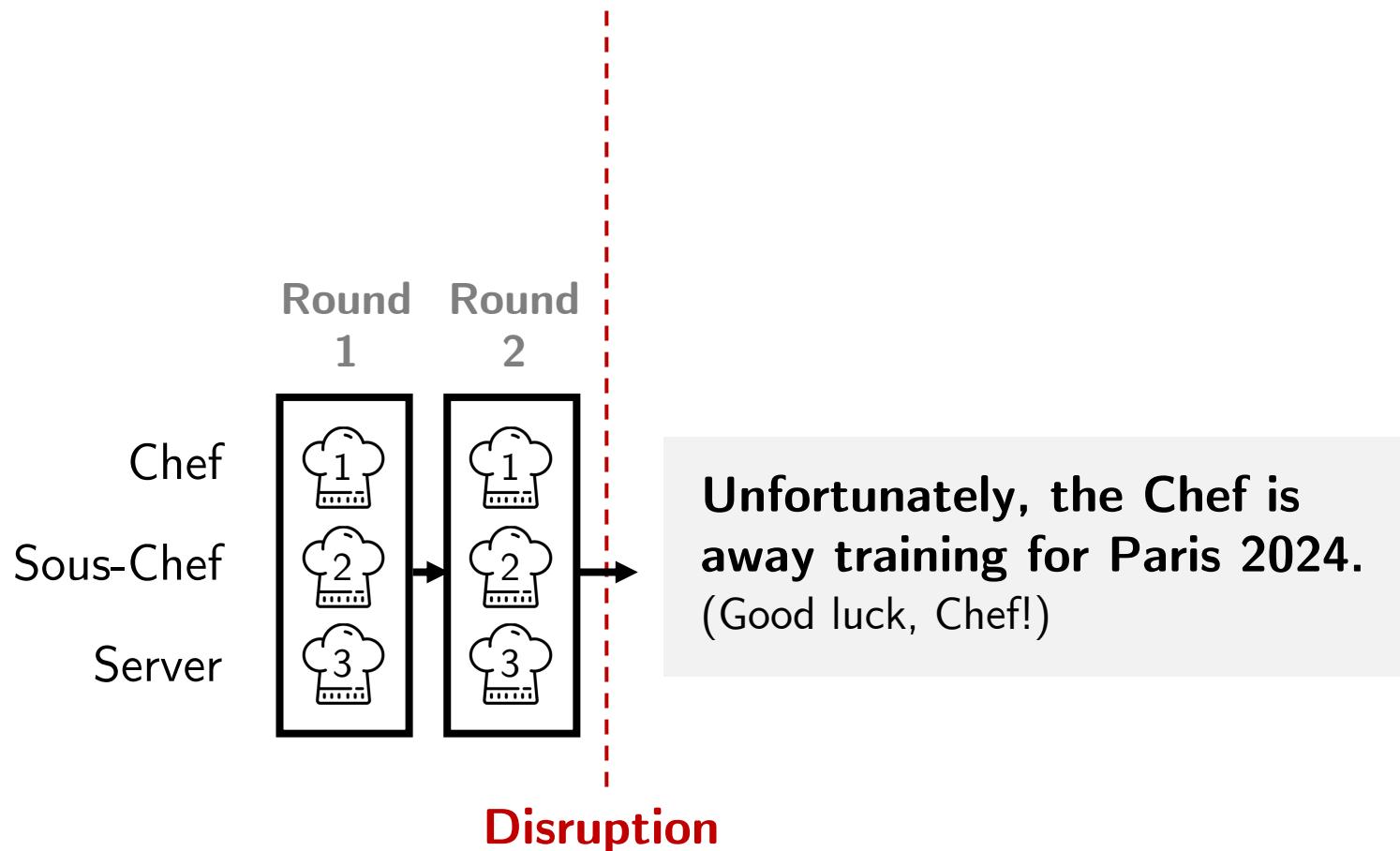
# Study 1:

## Design

## Disruption Scenario



x 4 within 50 ticks



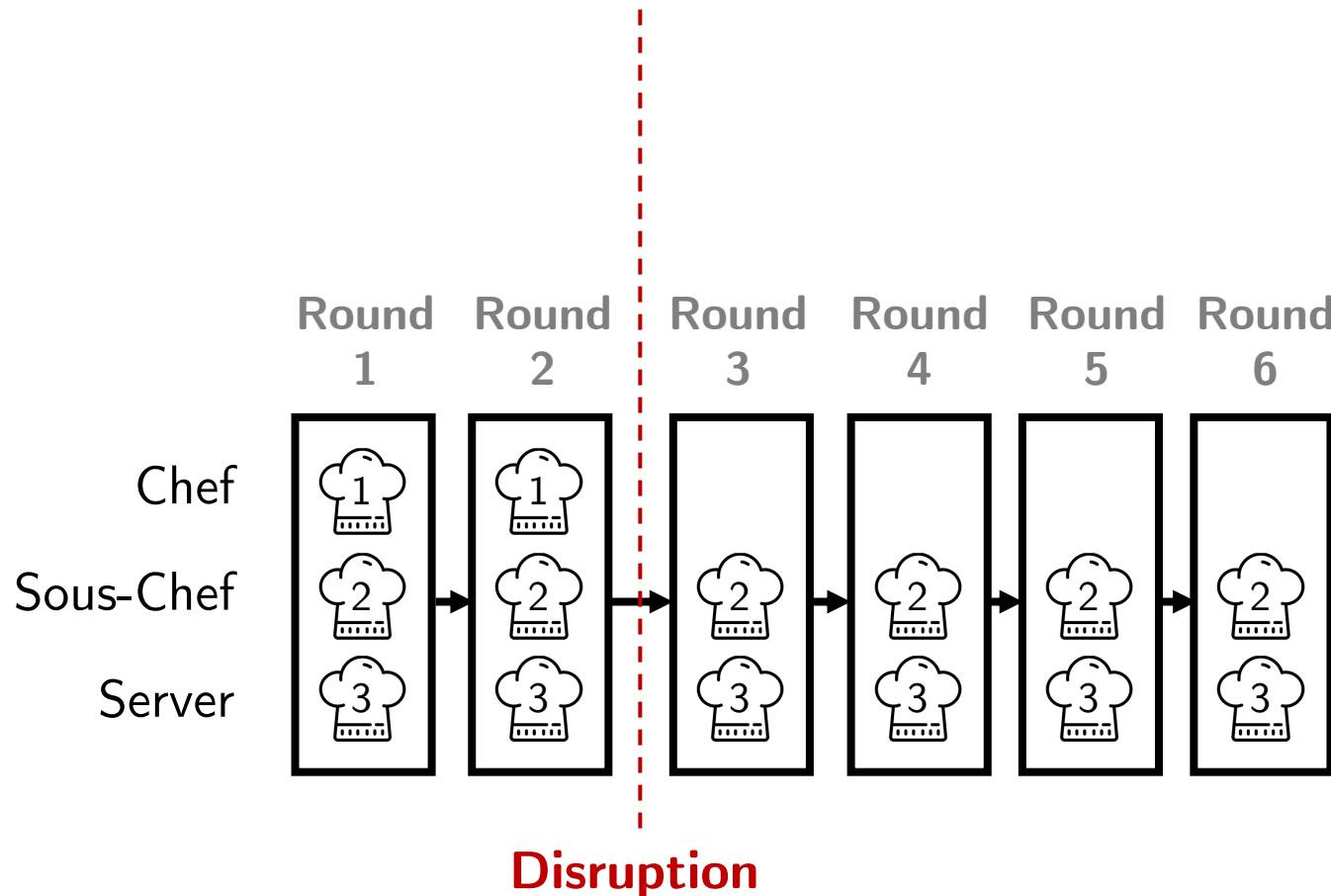
# Study 1:

## Design

## Disruption Scenario



x 4 within 50 ticks



# Problem Formulation

## MDP Formulation:

*Optimal policy and human* make sequences of decisions



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

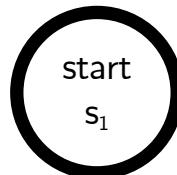
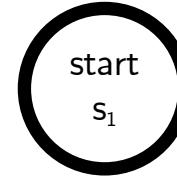
# Problem Formulation

## MDP Formulation:

*Optimal policy and human* make sequences of decisions



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

 $\pi^*$  $\pi$ 

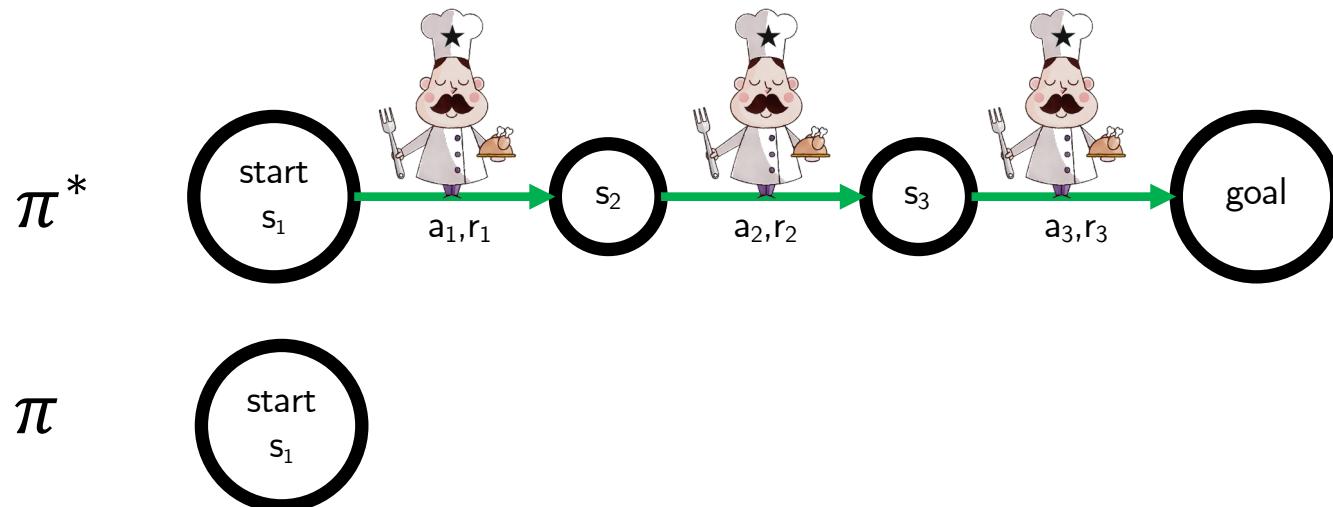
# Problem Formulation

## MDP Formulation:

*Optimal policy and human make sequences of decisions*



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



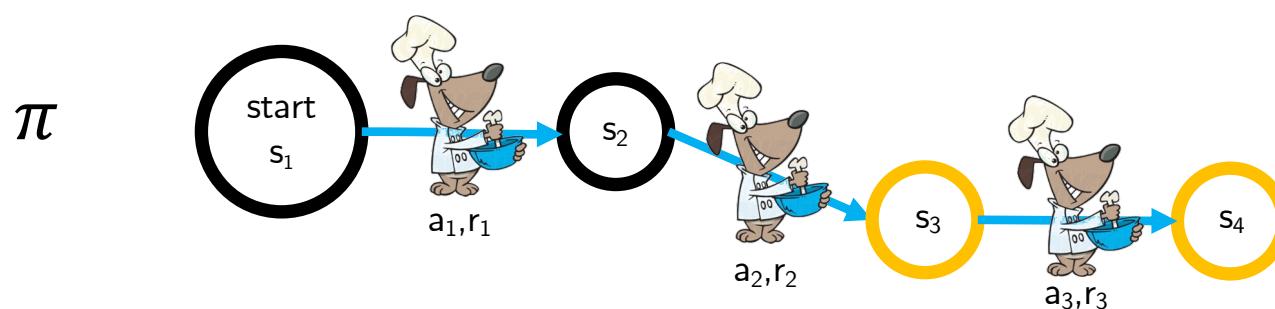
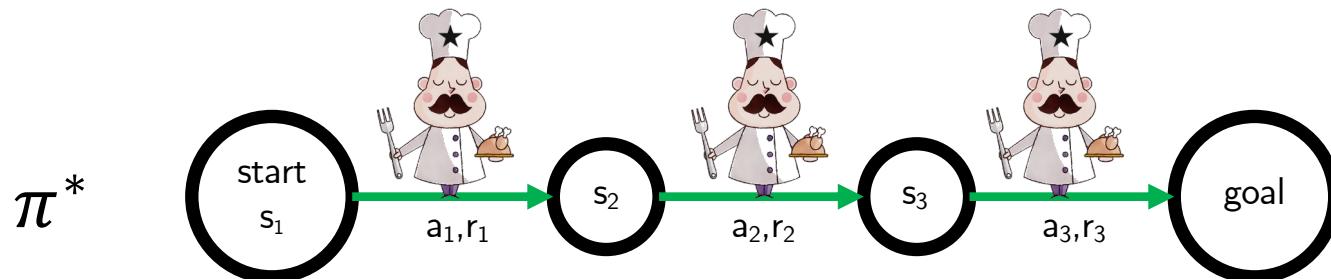
# Problem Formulation

## MDP Formulation:

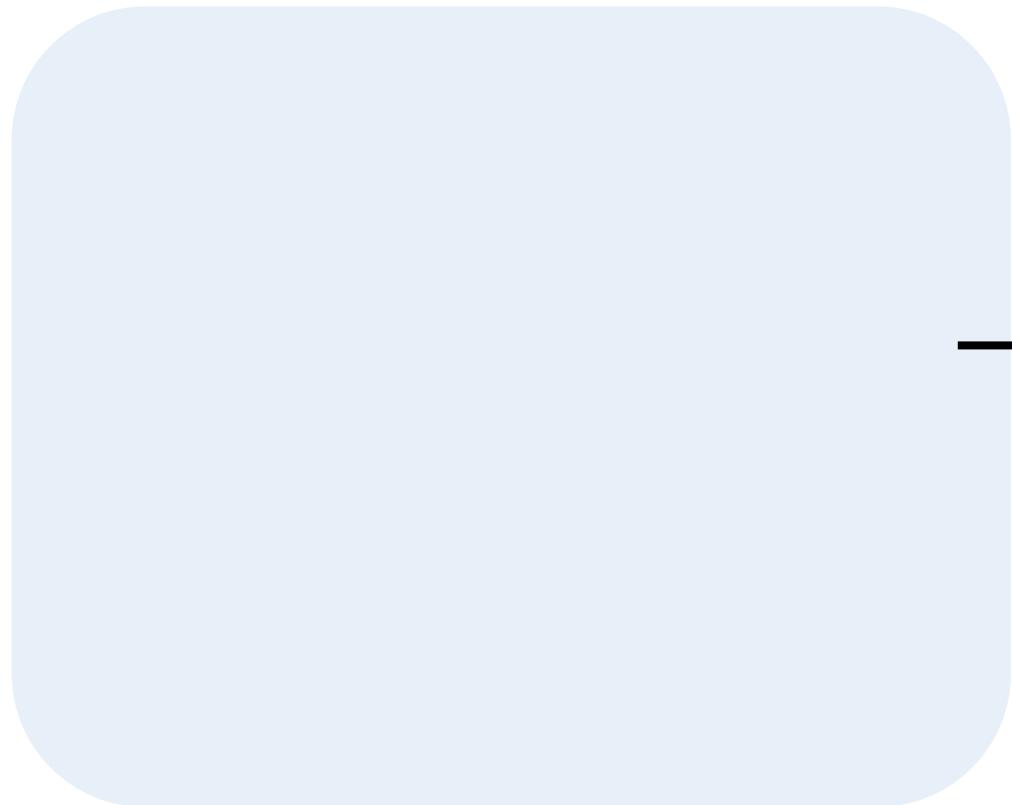
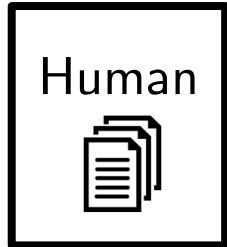
*Optimal policy and human* make sequences of decisions



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

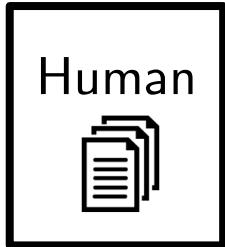


# Our Approach



# Our Approach

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



**Input:**

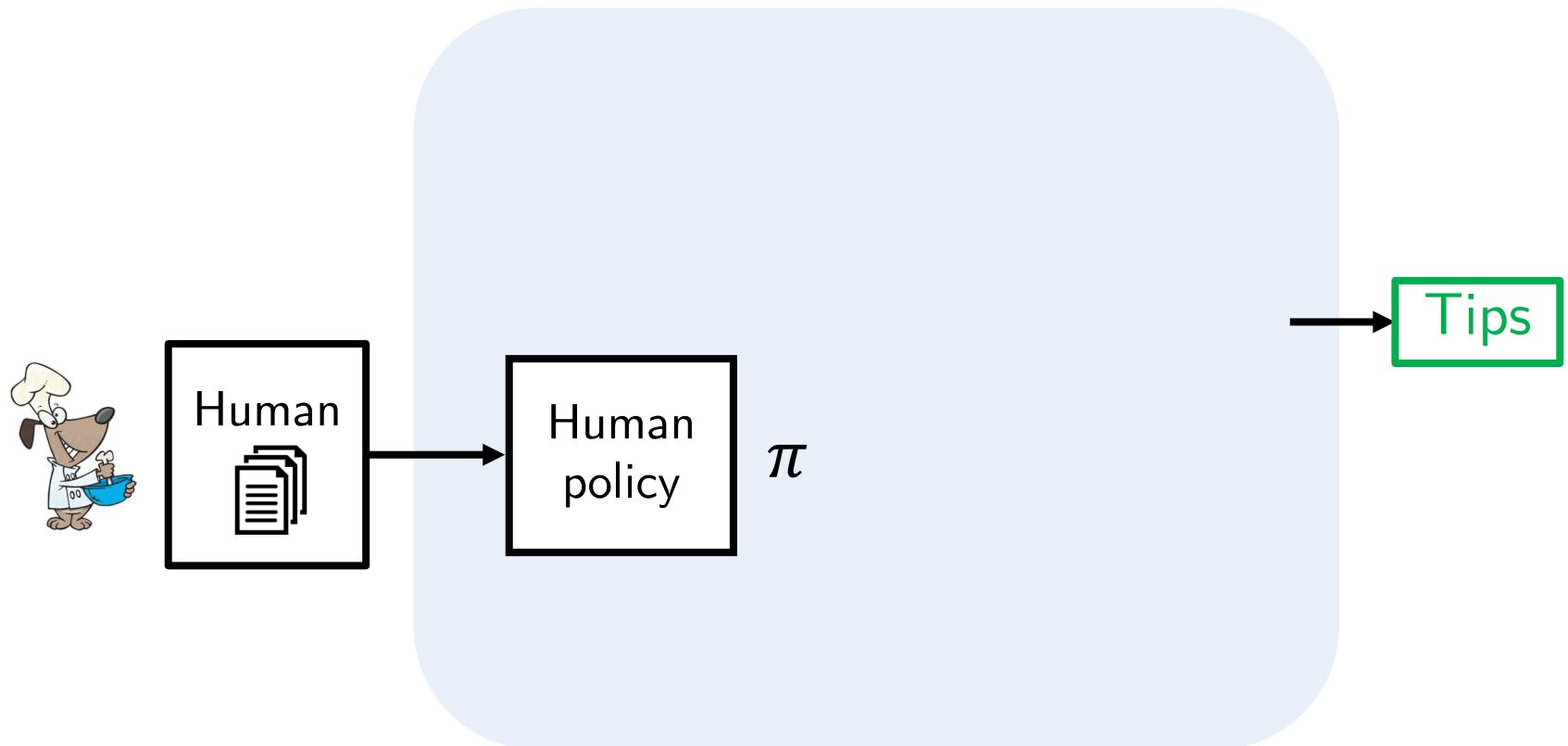
*Trace data  $\hat{d}_h$   
from human*

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

Tips

# Our Approach

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



# Our Approach

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

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

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



$\pi$

# Step 1: Q-Learning

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

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

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

- Watkins & Dayan 1992

# Step 1: Q-Learning

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

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

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

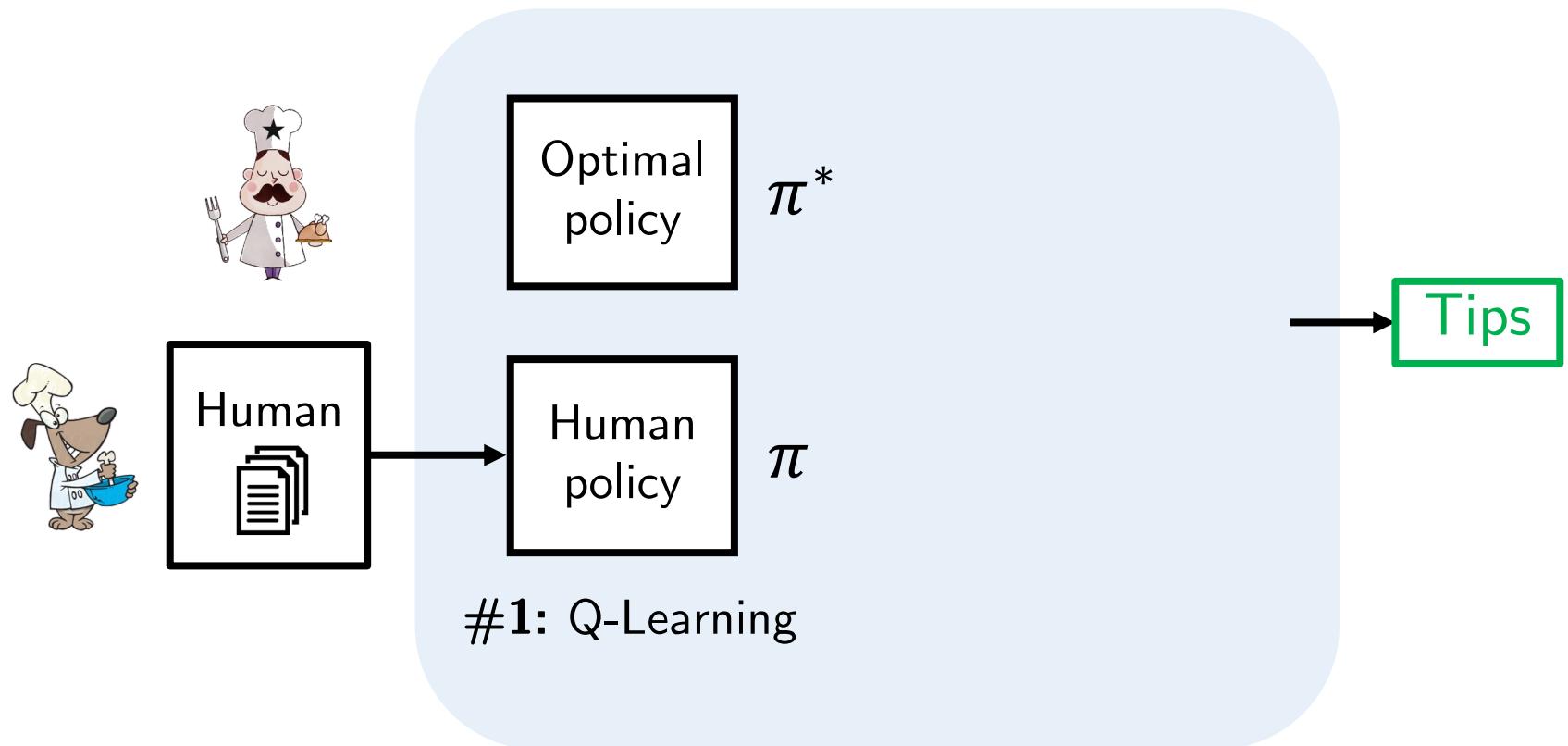
- Watkins & Dayan 1992

- Learn using supervised learning on trace data obtained using  $\pi$

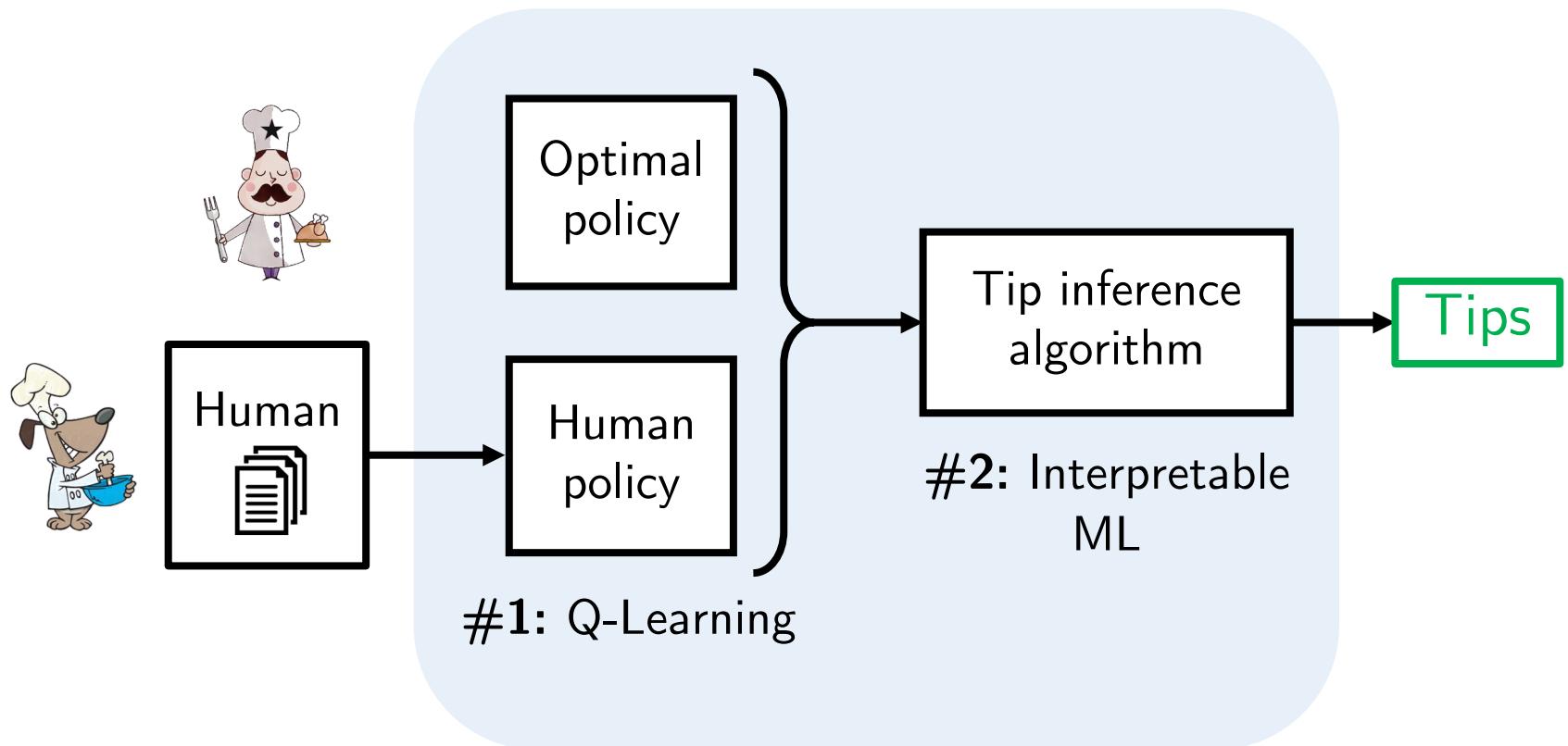
$$\hat{Q}_\theta^\pi(s, a) \approx Q^\pi(s, a)$$

# Our Approach

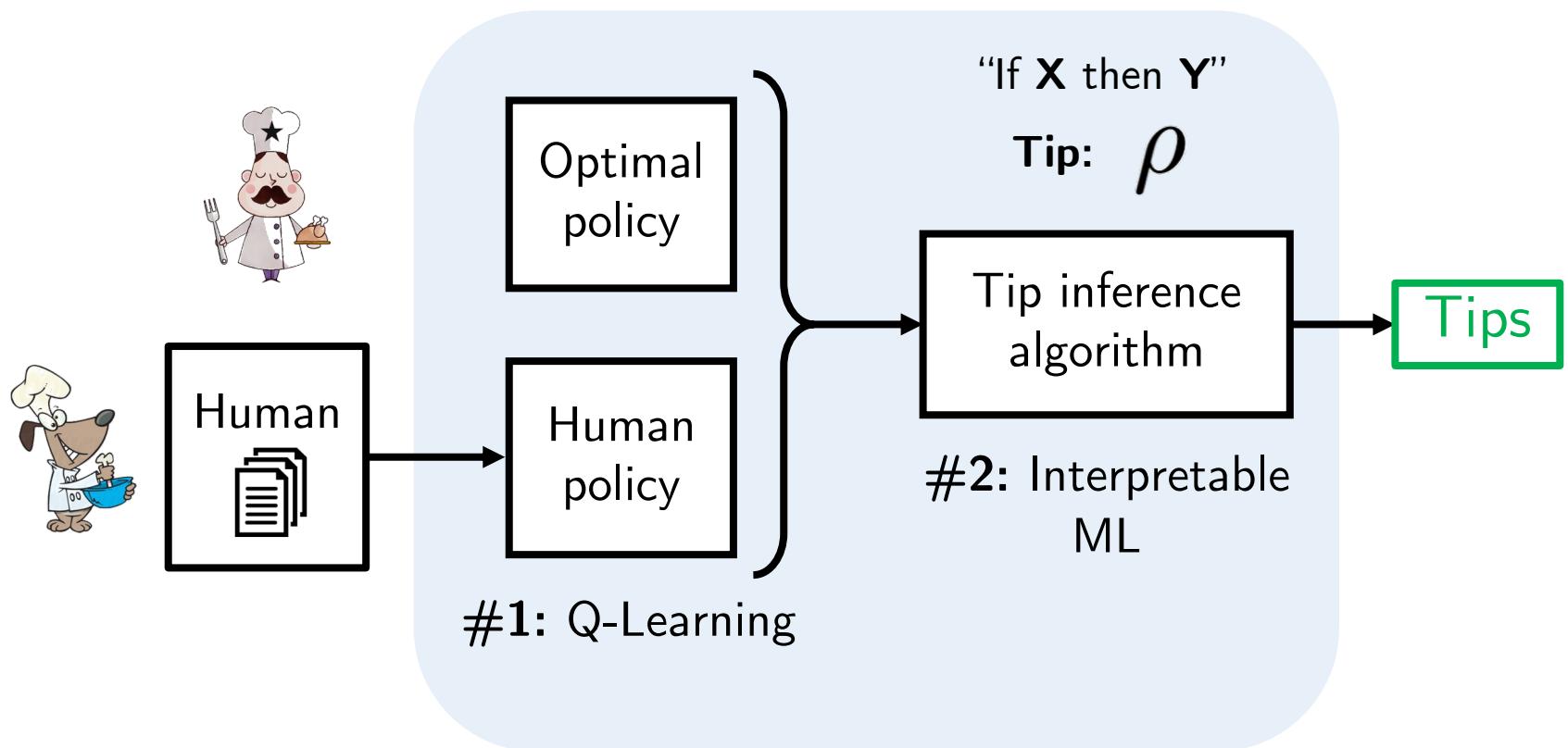
MDP:  $\mathcal{M} = (S, A, R, P, \gamma)$



# Our Approach



# Our Approach



# Step 2: Tip Inference

---

Cumulative reward  
for a given policy

$$J(\pi) = \mathbb{E}_{\zeta \sim D^{(\pi)}} \left[ \sum_{t=1}^T r_t \right]$$

# Step 2: Tip Inference

Cumulative reward  
for a given policy

$$J(\pi) = \mathbb{E}_{\zeta \sim D^{(\pi)}} \left[ \sum_{t=1}^T r_t \right]$$

- **Algorithm:** Choose tip  $\rho$  that maximizes the objective

$$J(\pi_H \oplus \rho) - J(\pi_H)$$

**Human policy + tip**      **Only human policy**

- $\pi_h \oplus \rho$  denotes overriding the human policy with tip  $\rho$ .

# Step 2: Tip Inference

Cumulative reward  
for a given policy

$$J(\pi) = \mathbb{E}_{\zeta \sim D^{(\pi)}} \left[ \sum_{t=1}^T r_t \right]$$

- **Algorithm:** Choose tip  $\rho$  that maximizes the objective

$$J(\pi_H \oplus \rho) - J(\pi_H)$$

**Human policy + tip**      **Only human policy**

- $\pi_h \oplus \rho$  denotes overriding the human policy with tip  $\rho$ .

- **Lemma:**  $J(\pi_H \oplus \rho) - J(\pi_H) \approx$

$$\mathbb{E}_{\zeta \sim D^{(\pi_H)}} \left[ \sum_{t=1}^T Q_t^*(s_t, \pi_H \oplus \rho(s_t)) - Q_t^*(s_t, \pi_H(s_t)) \right]$$

Indirect effect of distribution shift is small; use observed data

Q-network we learned previously!

## Step 2: Tip Inference

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

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

# Step 2: Tip Inference

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

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

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

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

# Step 2: Tip Inference

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

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

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

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



Assign chef to cook the first 2 burgers

# Step 2: Tip Inference

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

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

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

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



Assign chef to cook the first 2 burgers



Chef should cook twice

# Step 2: Tip Inference

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

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

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

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



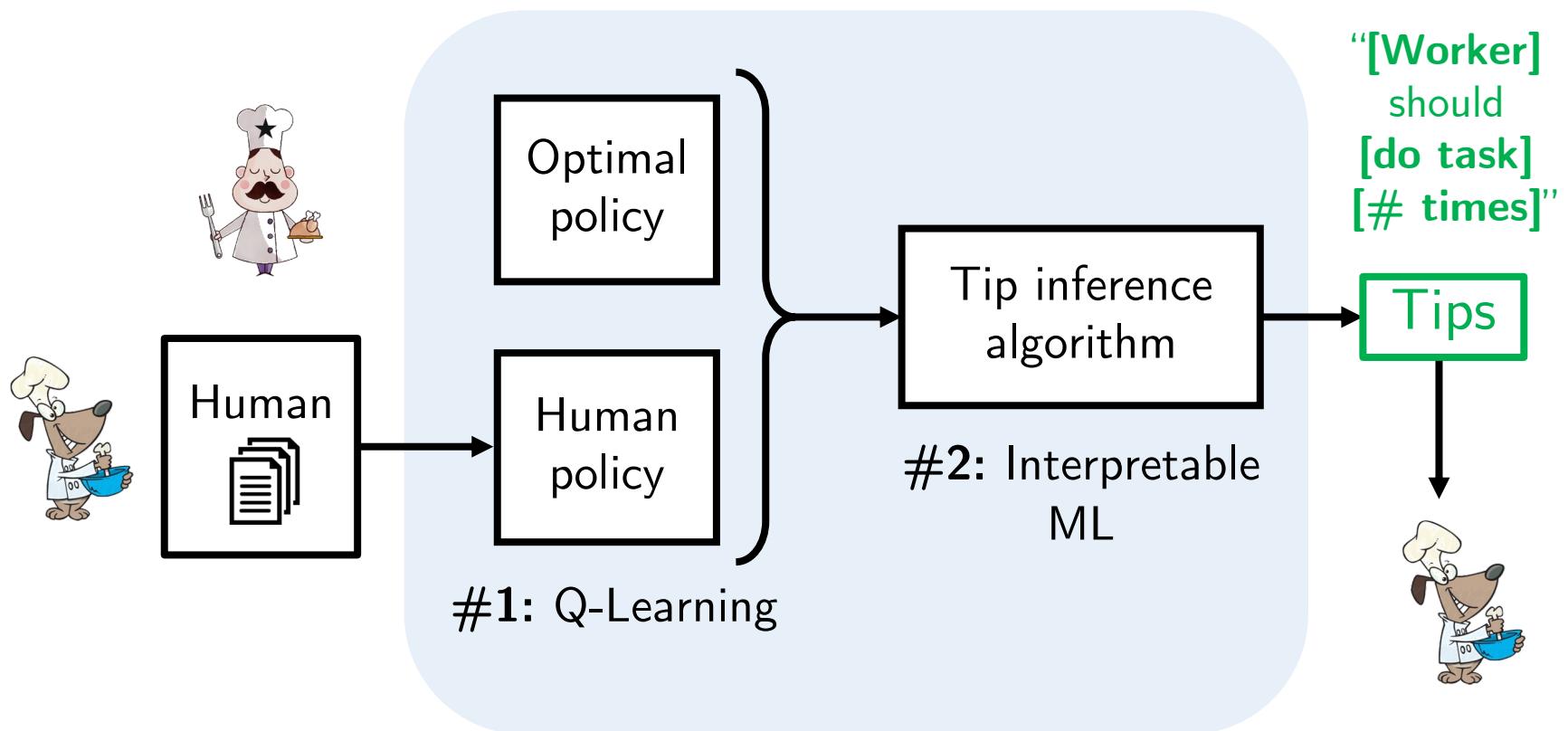
Assign chef to cook the first 2 burgers



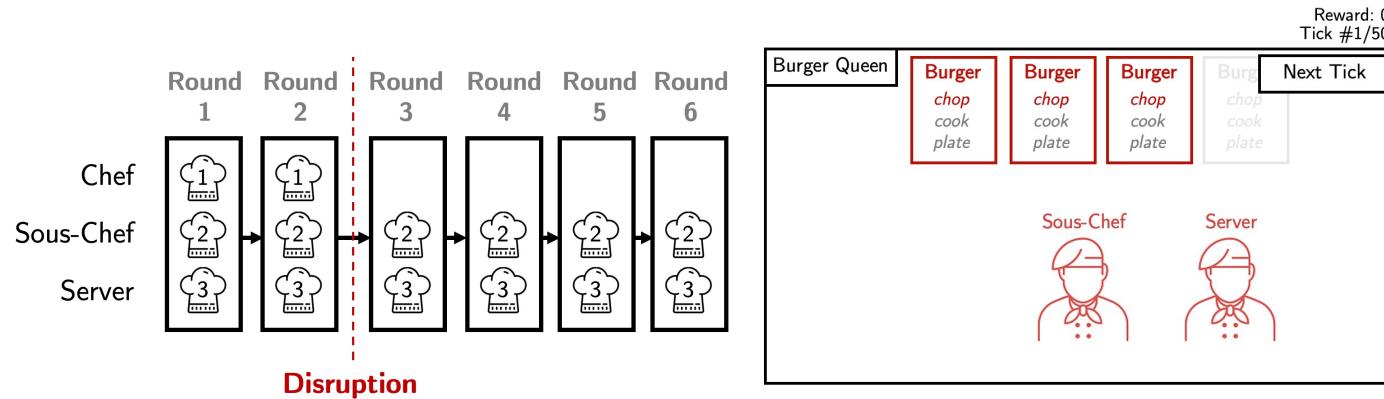
Chef should cook twice

“[Worker] should [do task] [# times]”

# Our Approach

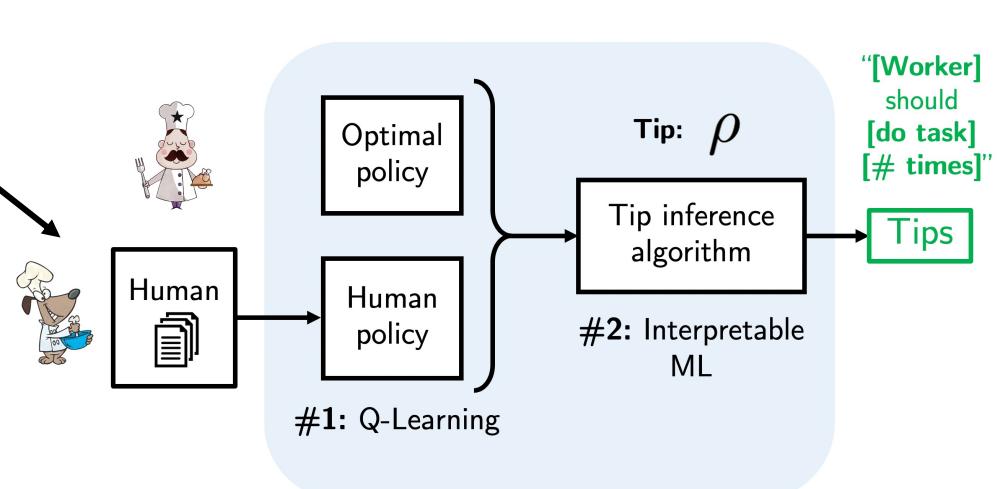
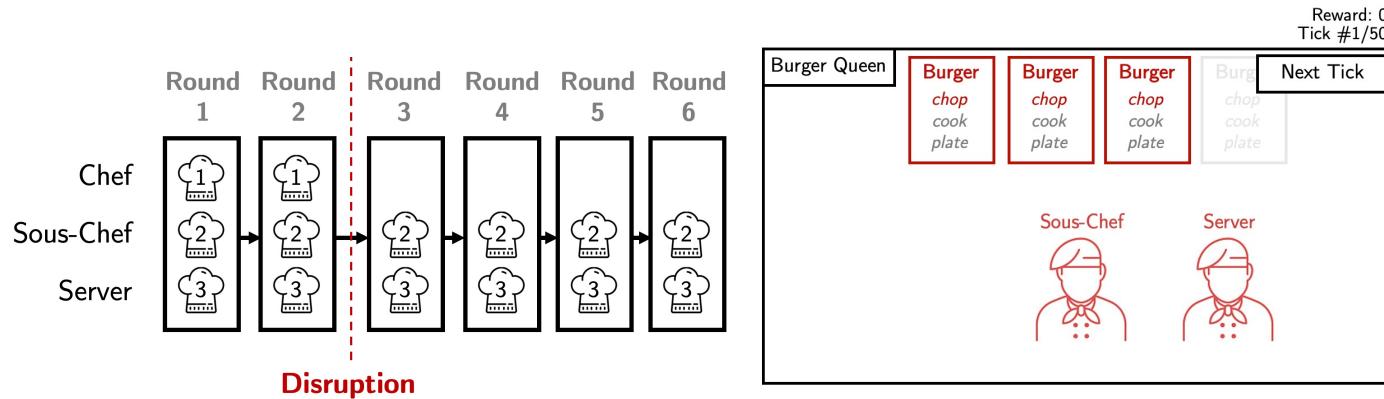


# Phase I Collect Trace Data



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

# Phase I Collect Trace Data



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

Sous-Chef



Server



x 4

# Phase I Inferred Tips

Algorithm

Server  
should cook twice

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

Sous-Chef



Server



x 4

# Phase I Inferred Tips

Algorithm

Human

Server  
should cook twice

*Most frequent tip  
chosen by participants*

Sous-Chef



Server



x 4

# Phase I Inferred Tips

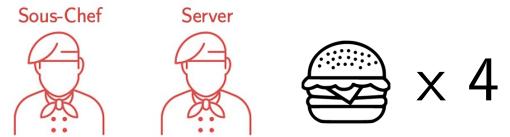
Algorithm

Server  
should cook twice

Human

Server  
should cook once

*Most frequent tip  
chosen by participants*



# Phase I Inferred Tips

Algorithm

Server  
should cook twice

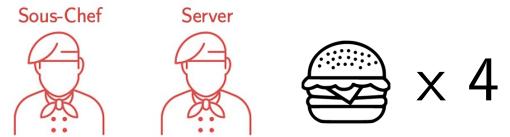
Human

Server  
should cook once

Baseline

*Most frequent tip  
chosen by participants*

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



# 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*

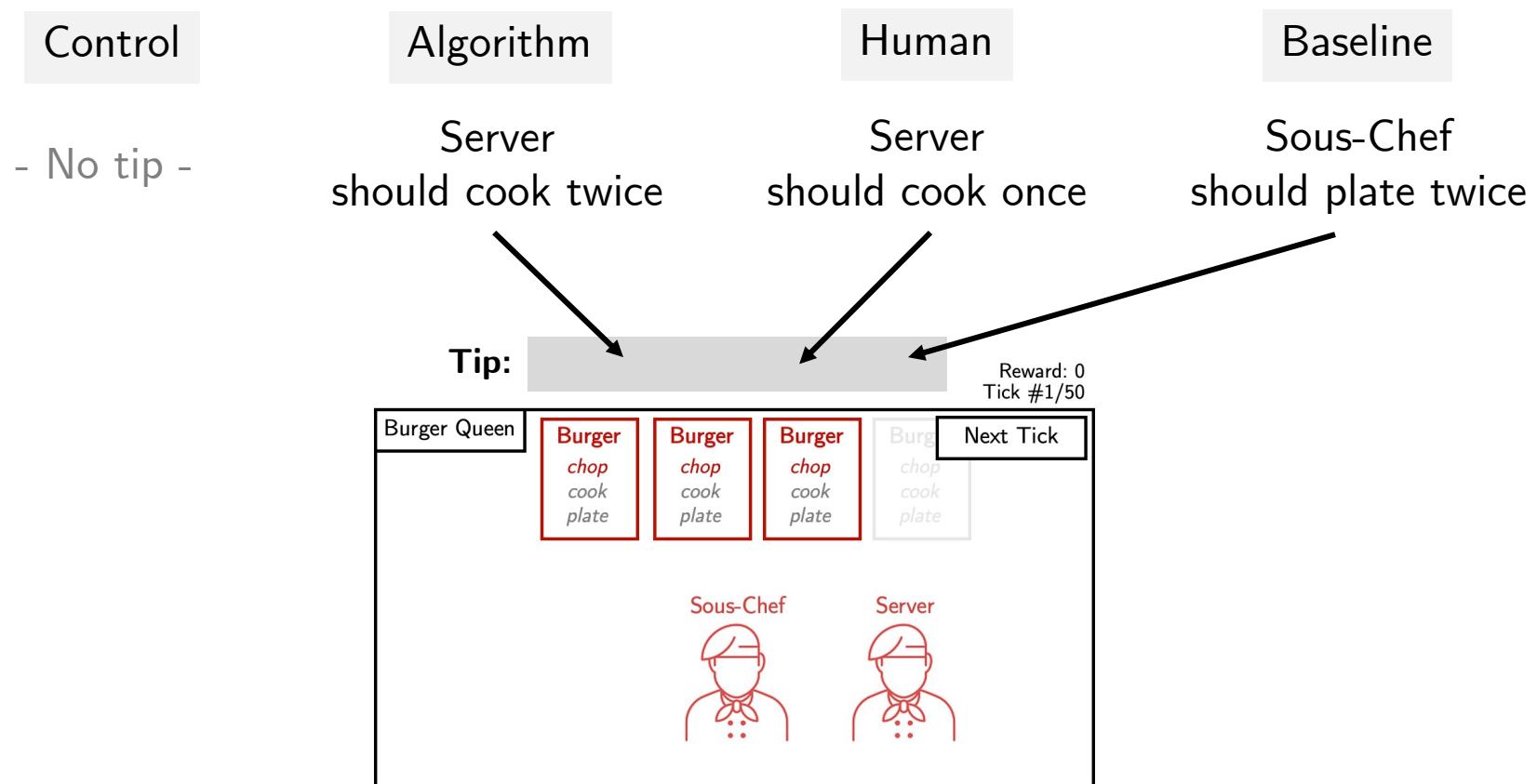
*Most frequent s-a  
deviation b/w optimal  
and trainee policies*

# Phase II Comparing Tips

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

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

# Phase II Comparing Tips



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

# Algorithm vs Human

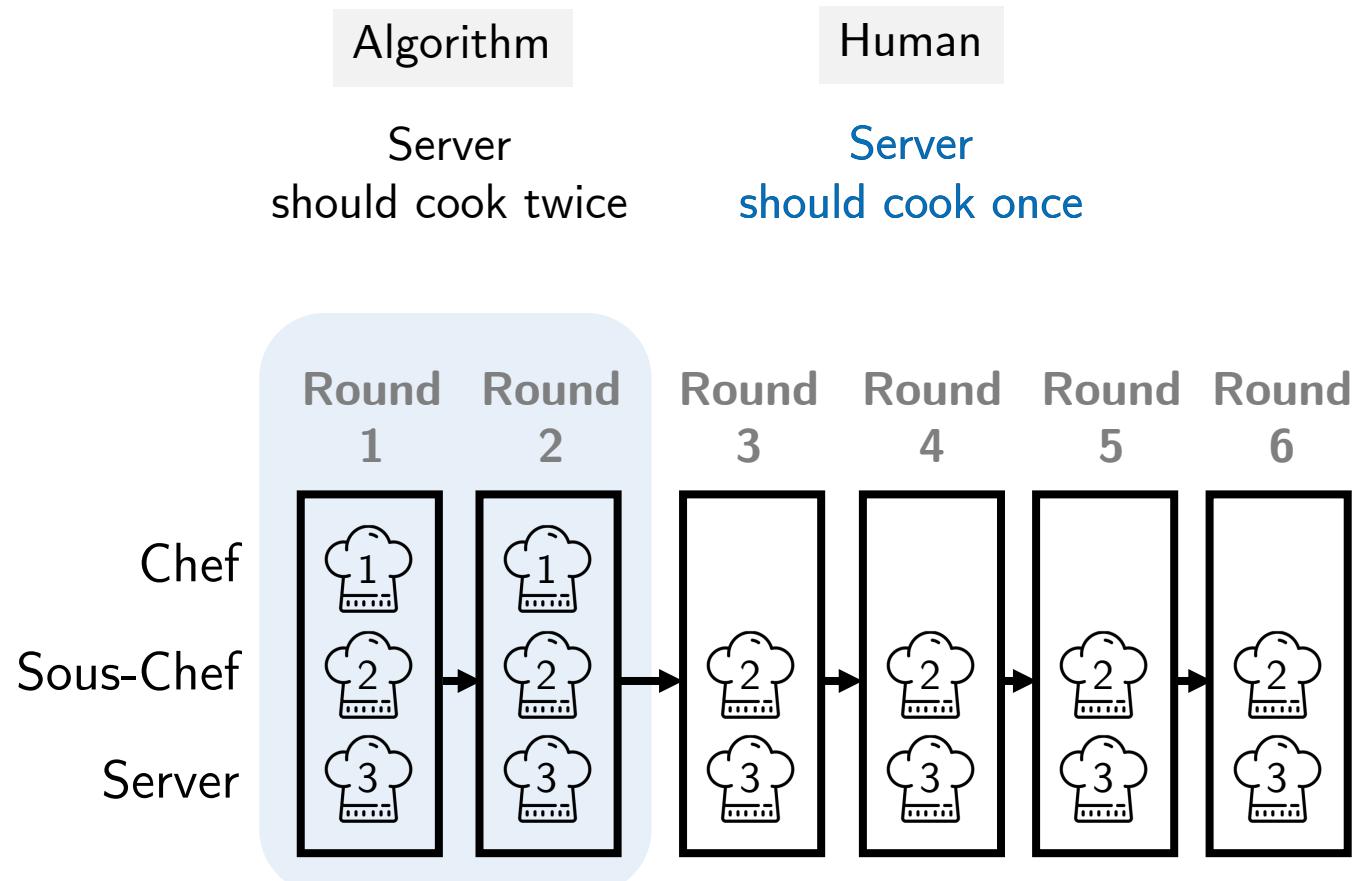
Algorithm

Server  
should cook twice

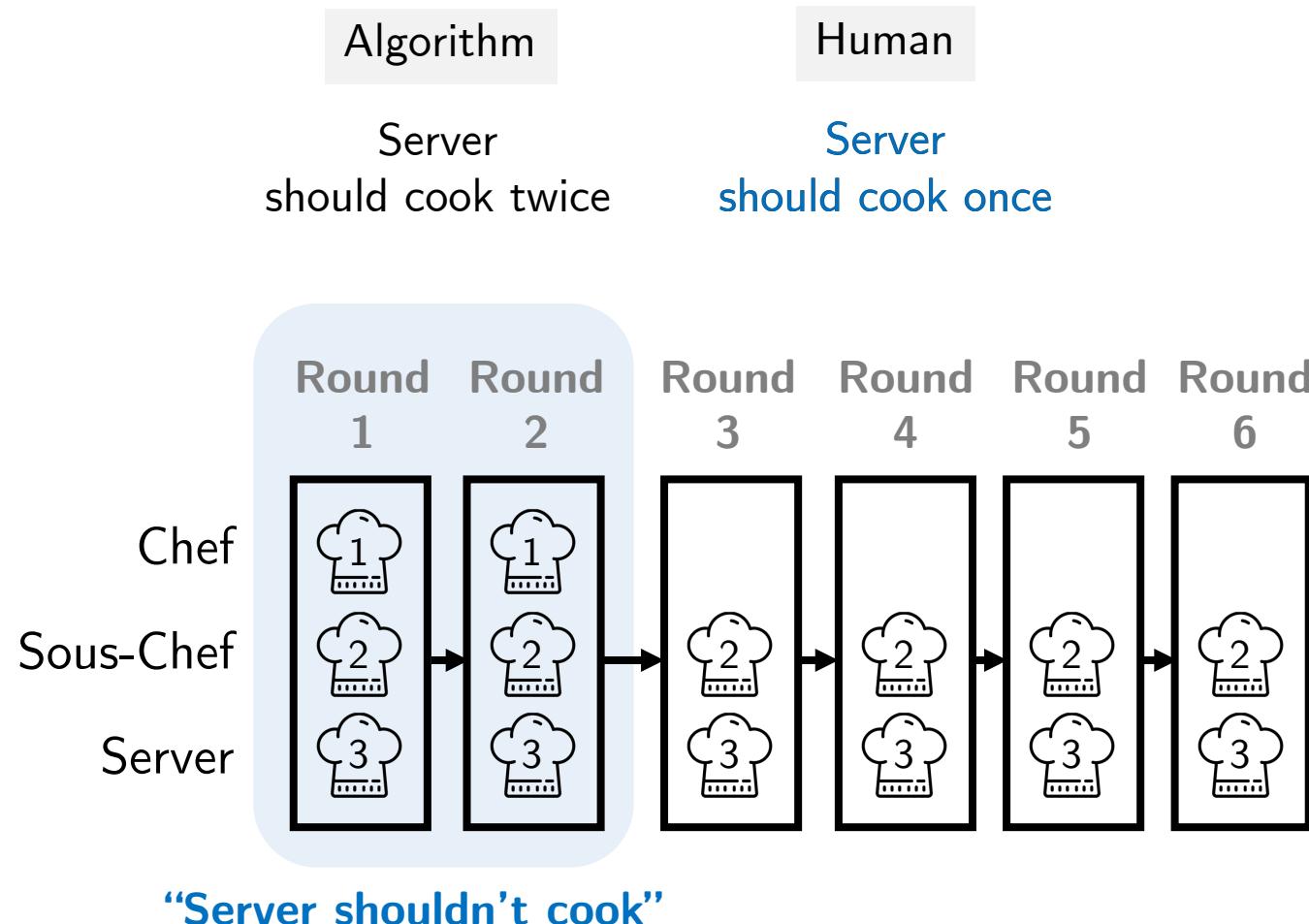
Human

Server  
should cook once

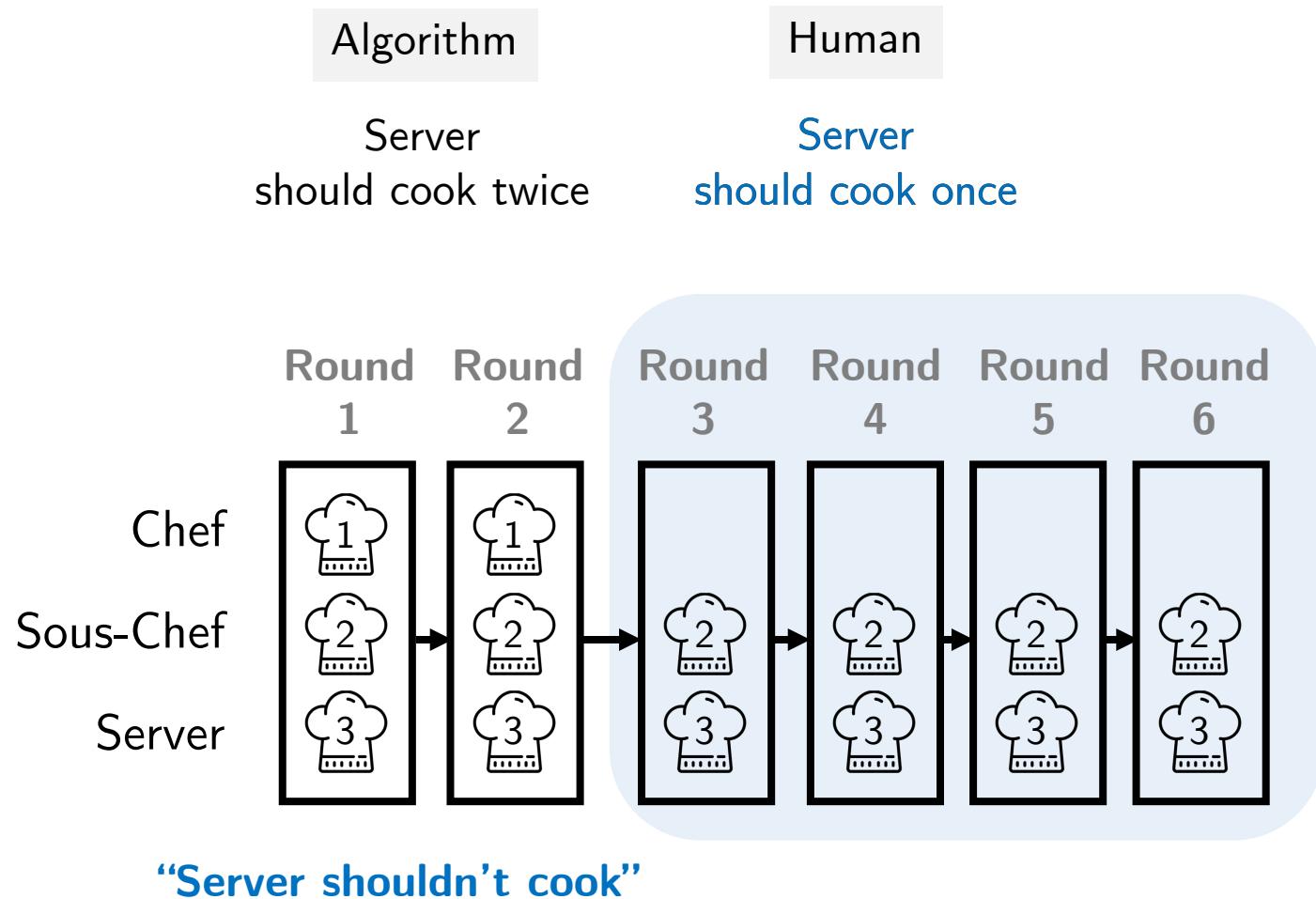
# Algorithm vs Human



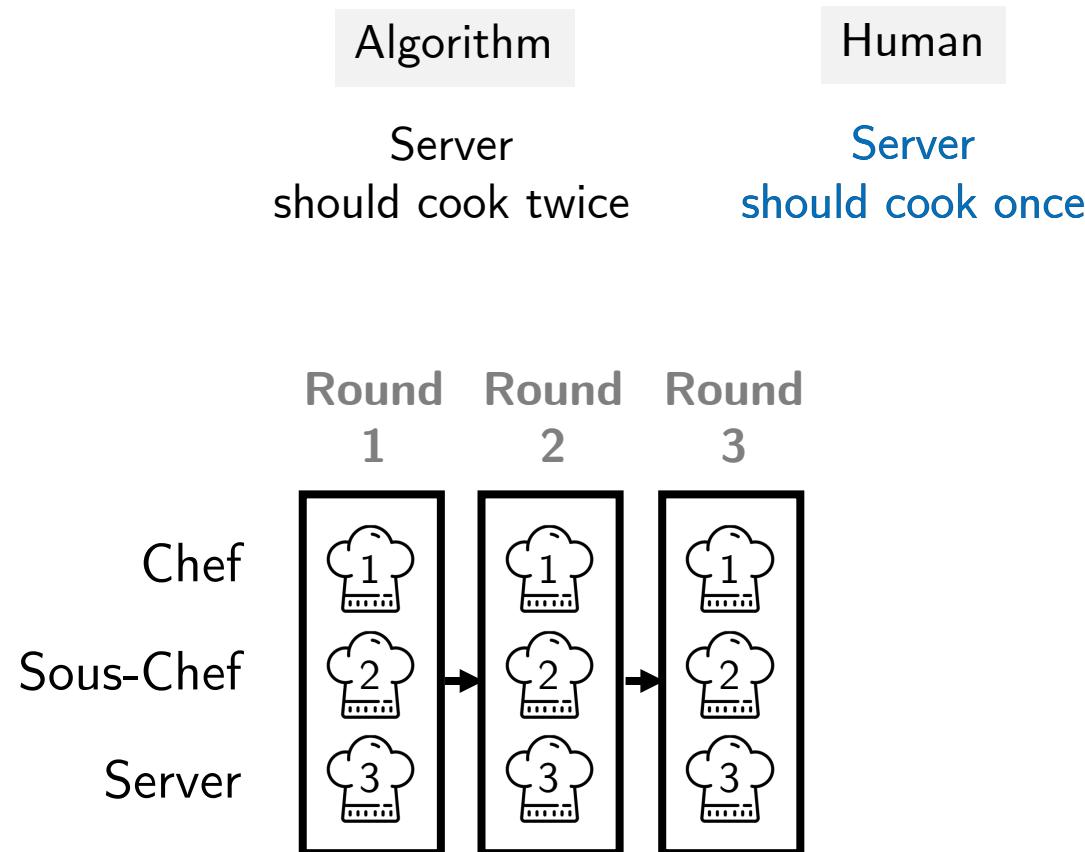
# Algorithm vs Human



# Algorithm vs Human

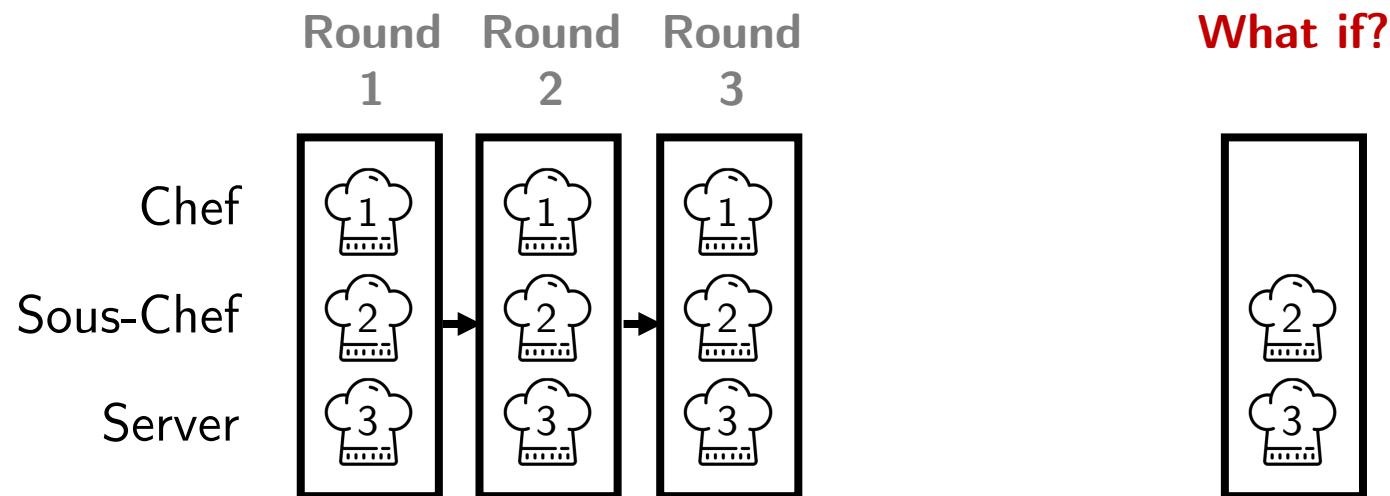


# Algorithm vs Human



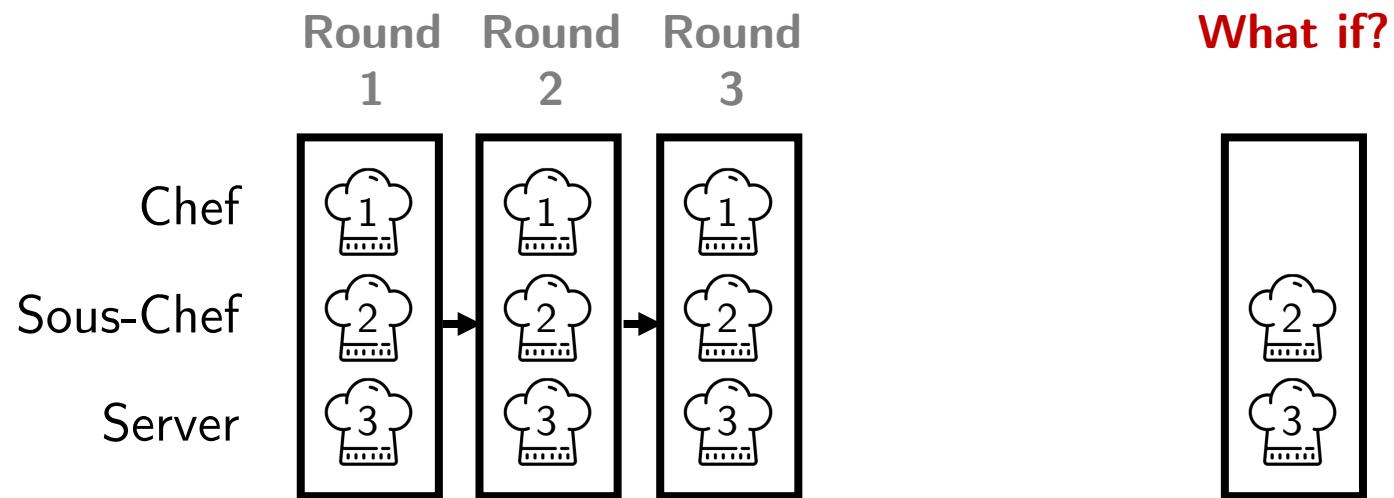
# Algorithm vs Human

Algorithm	Human
Server should cook twice	Server should cook once

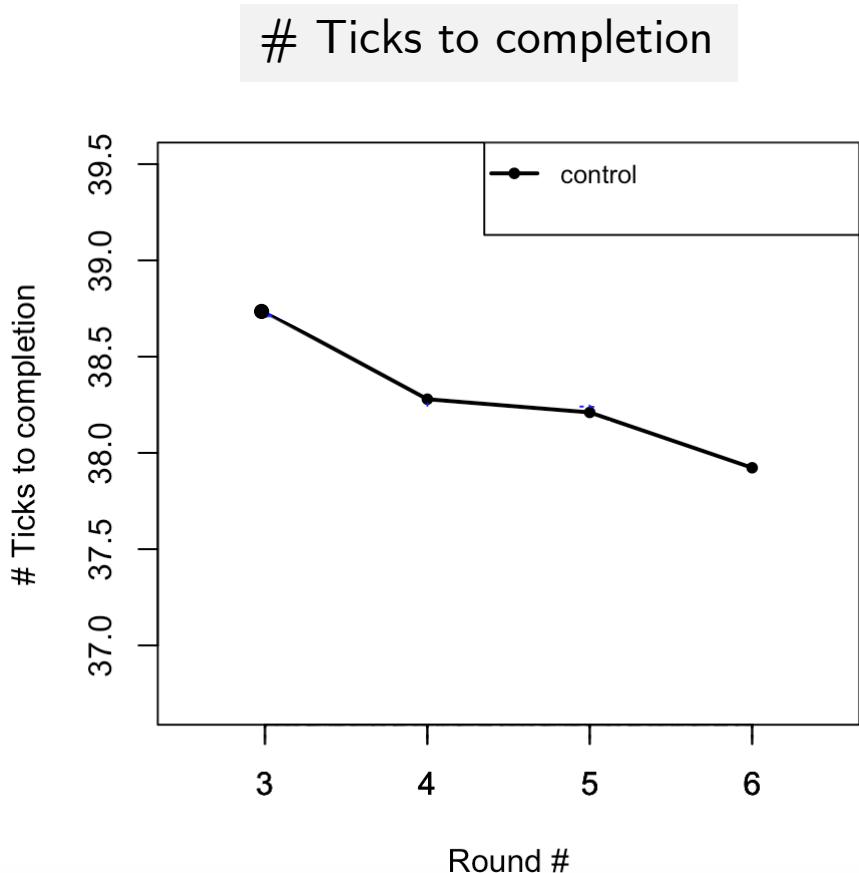


# Algorithm vs Human

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

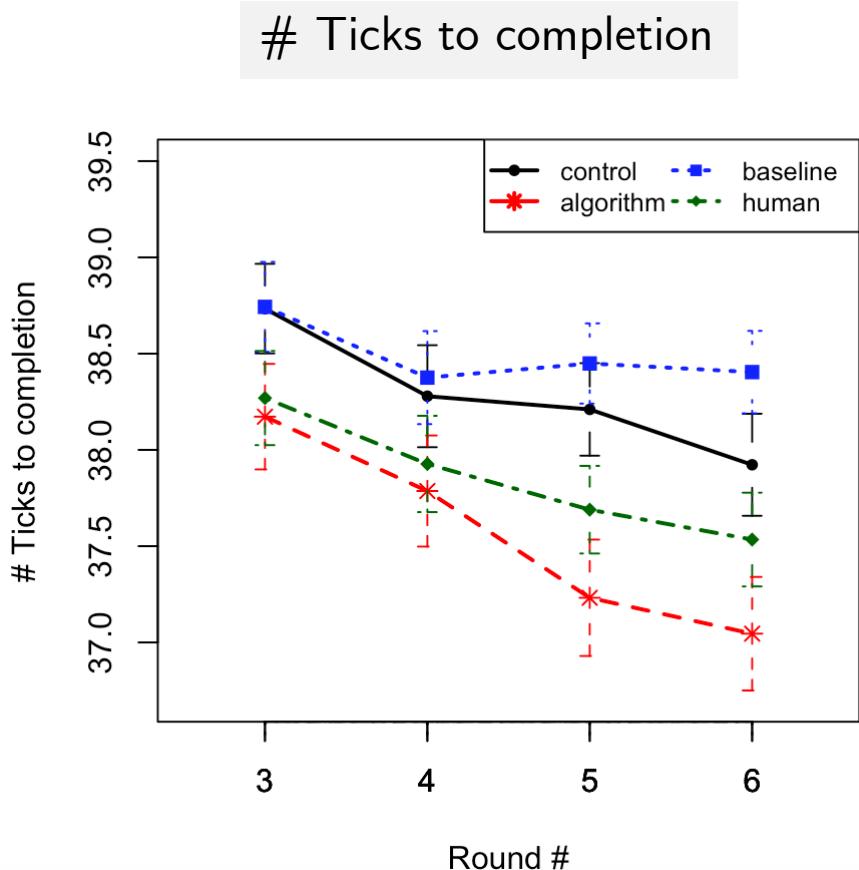


# Study 1: Results People Improve Over Time



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

# Study 1: Results Our Tip Improves Performance

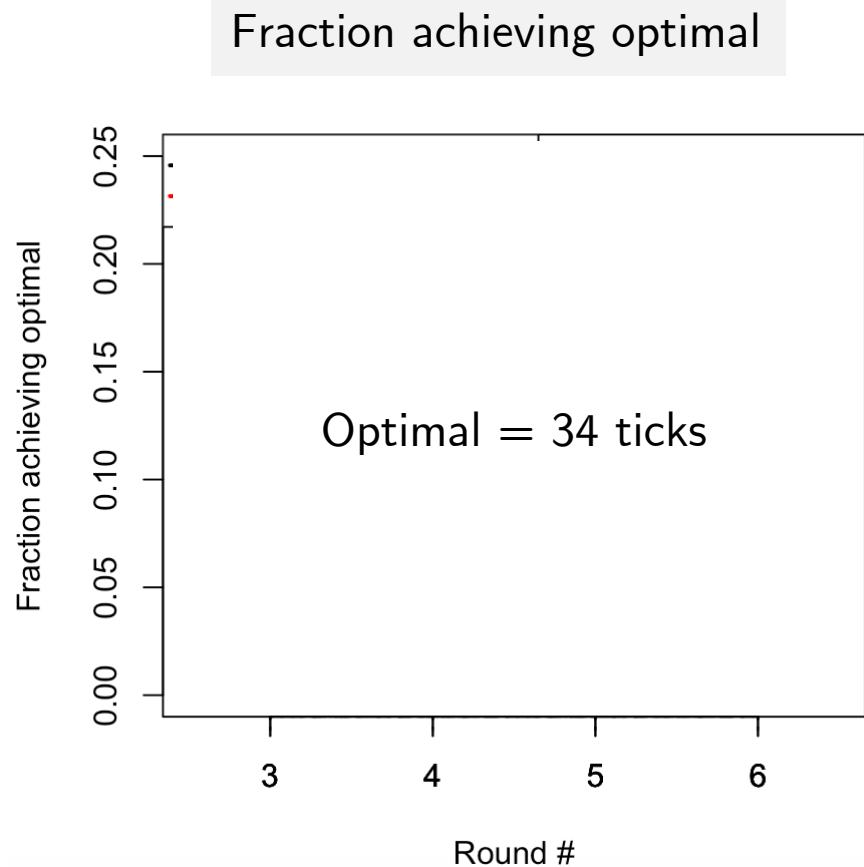
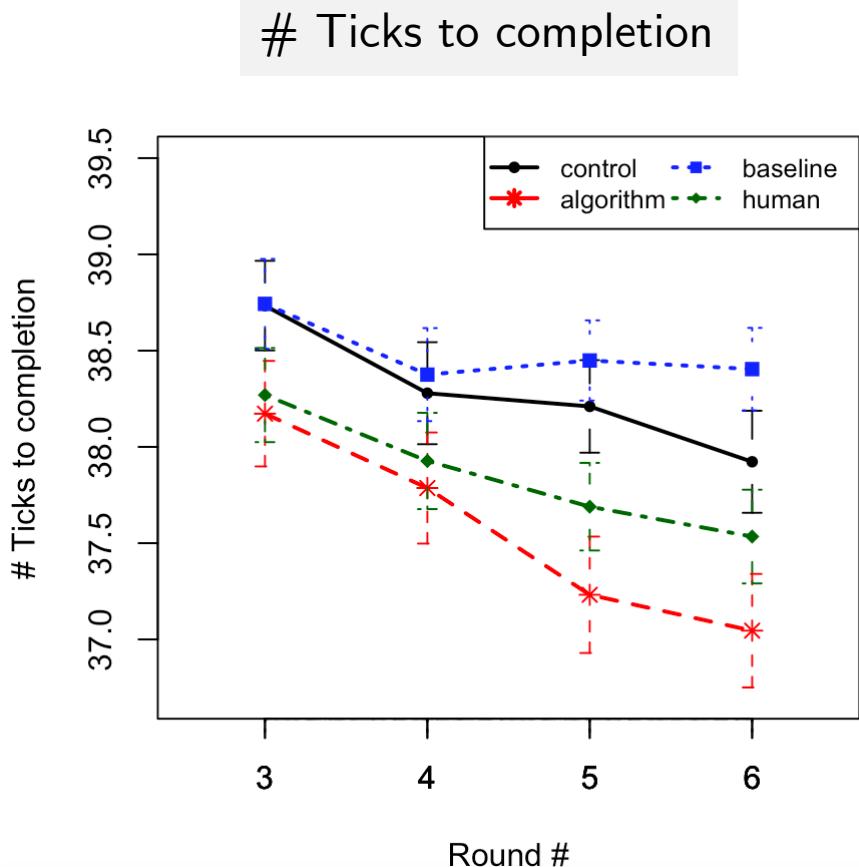


One-sided T-Tests

Algorithm beats Control ( $p = 0.000008$ )  
Algorithm beats Human ( $p = 0.006$ )  
Algorithm beats Baseline ( $p < 1e-12$ )

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

# Study 1: Results



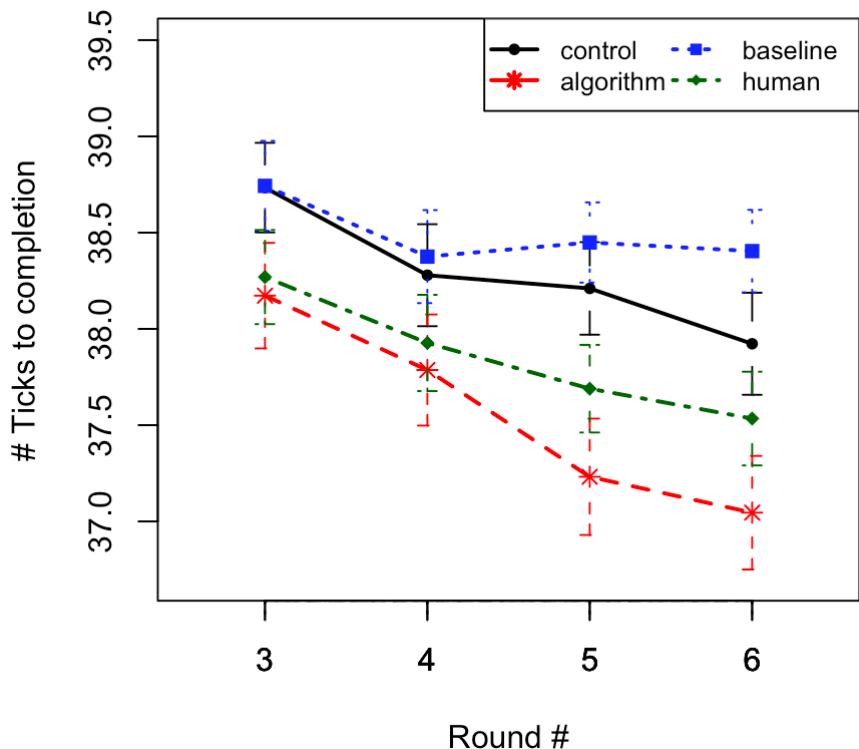
Amazon Mechanical Turk, N = 1,011  
mean age 34.9, 60% female

# Study 1:

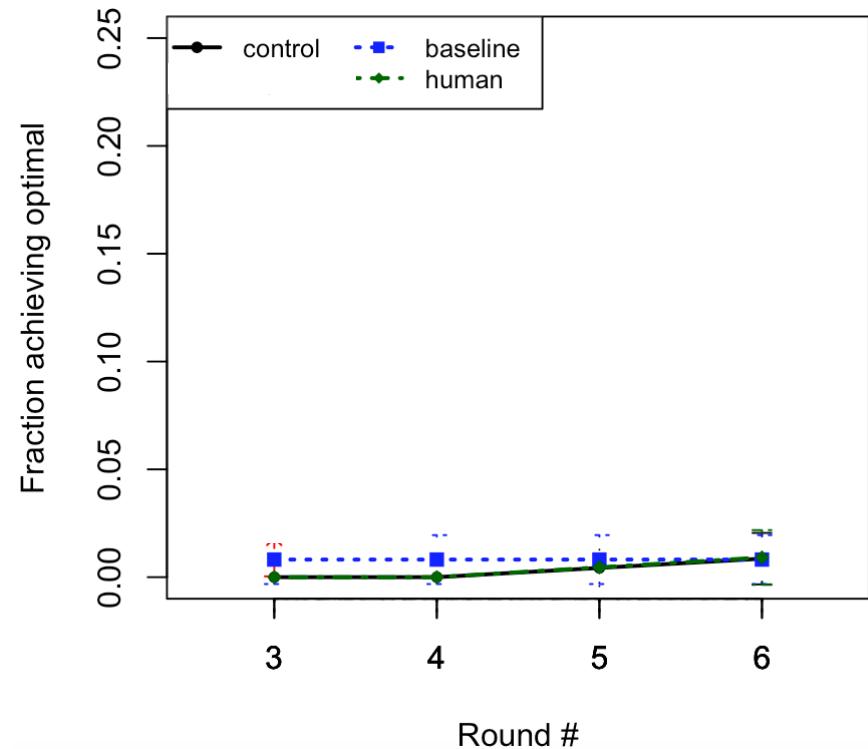
# Results

## Difficult to Reach Optimal

# Ticks to completion



Fraction achieving optimal



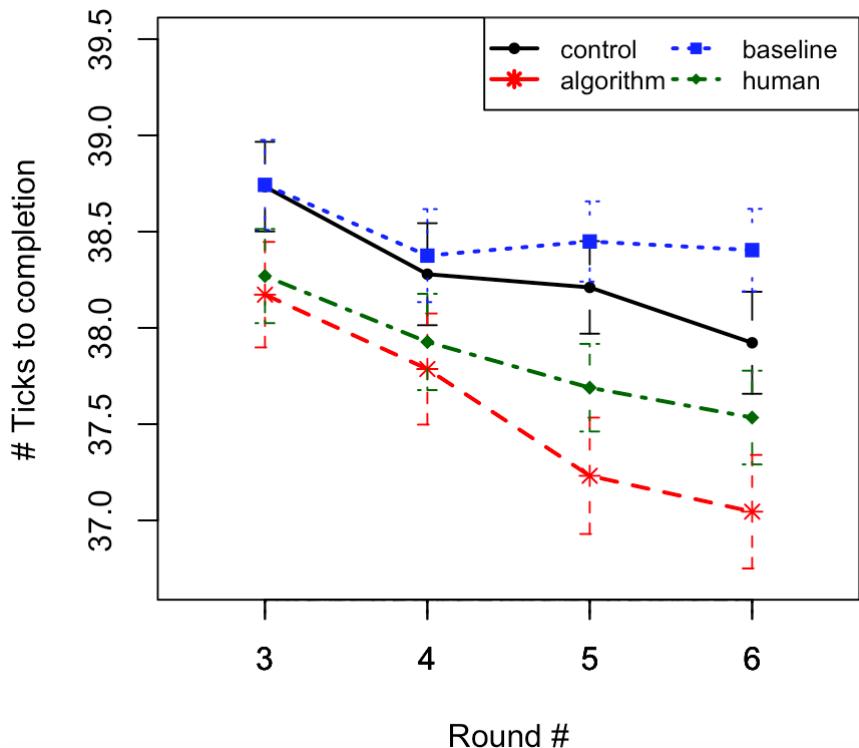
Amazon Mechanical Turk, N = 1,011  
mean age 34.9, 60% female

# Study 1:

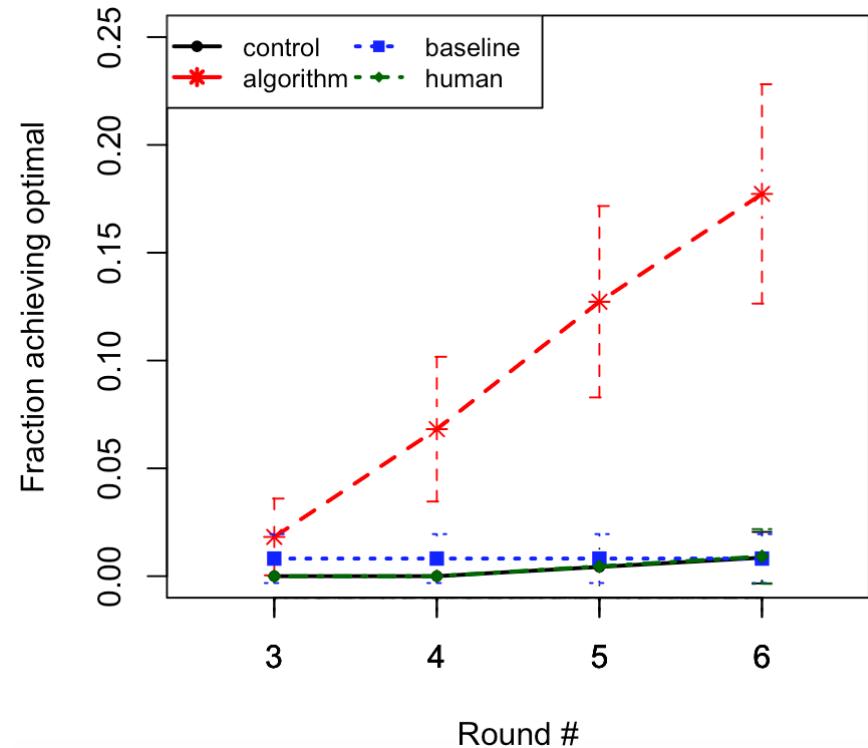
# Results

# Our Tip Helps Reach Optimal

# Ticks to completion

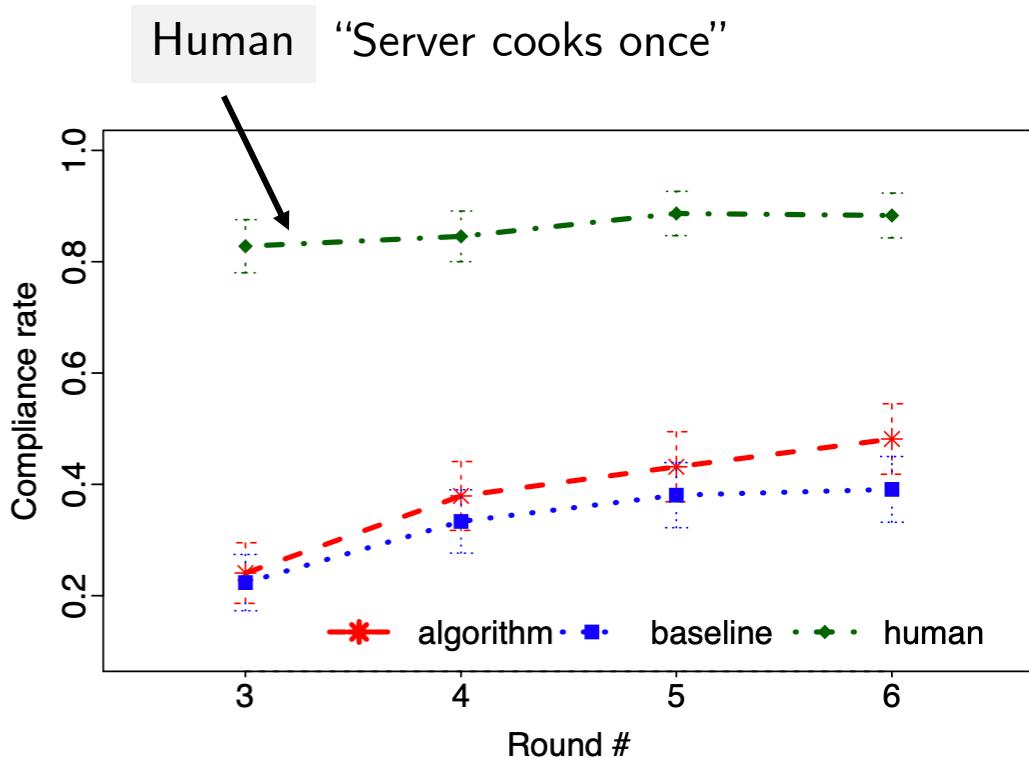


Fraction achieving optimal



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

# Study 1: Results Complying with Intuitive Tip

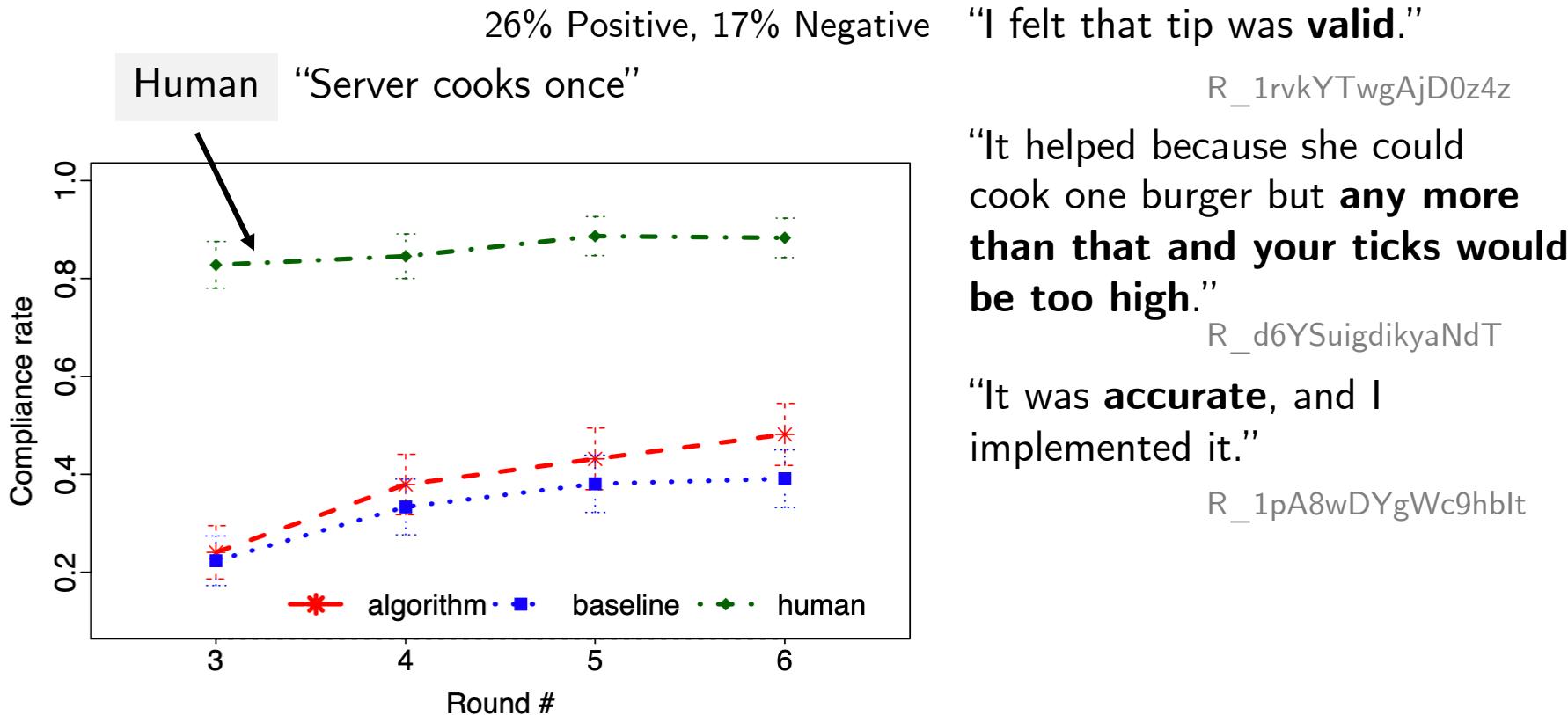


Amazon Mechanical Turk, N = 1,011  
mean age 34.9, 60% female

# Study 1:

# Results

## Complying with Intuitive Tip

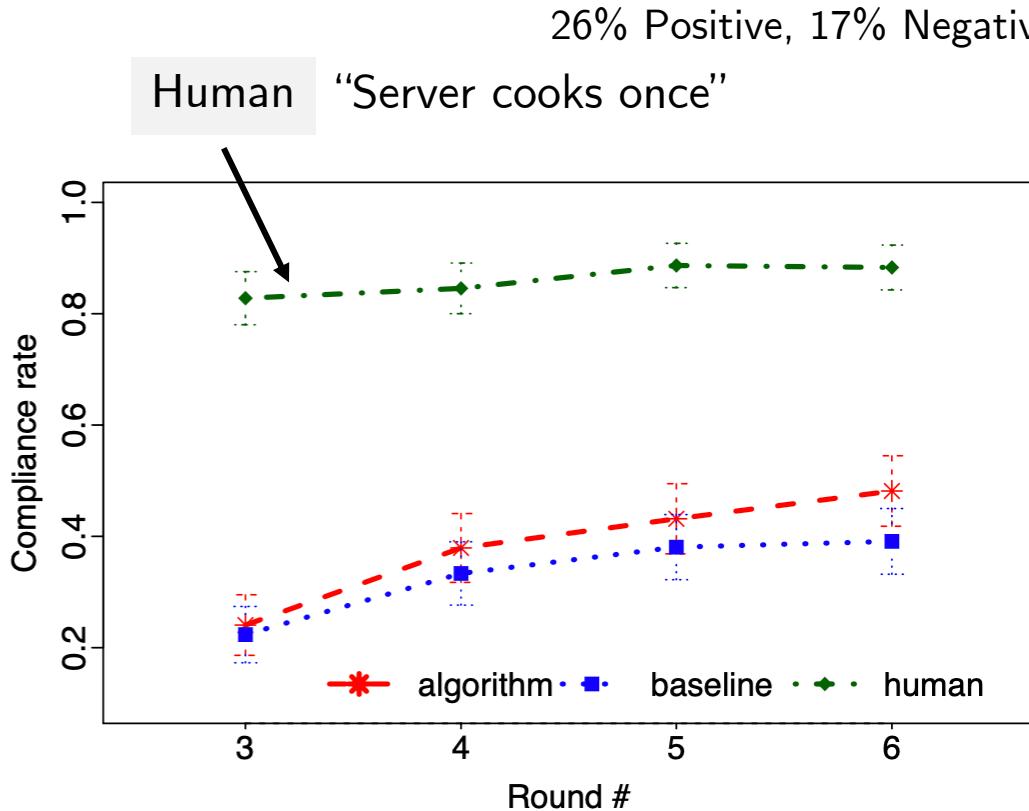


Amazon Mechanical Turk, N = 1,011  
mean age 34.9, 60% female

# Study 1:

# Results

# Complying with Intuitive Tip



“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.”

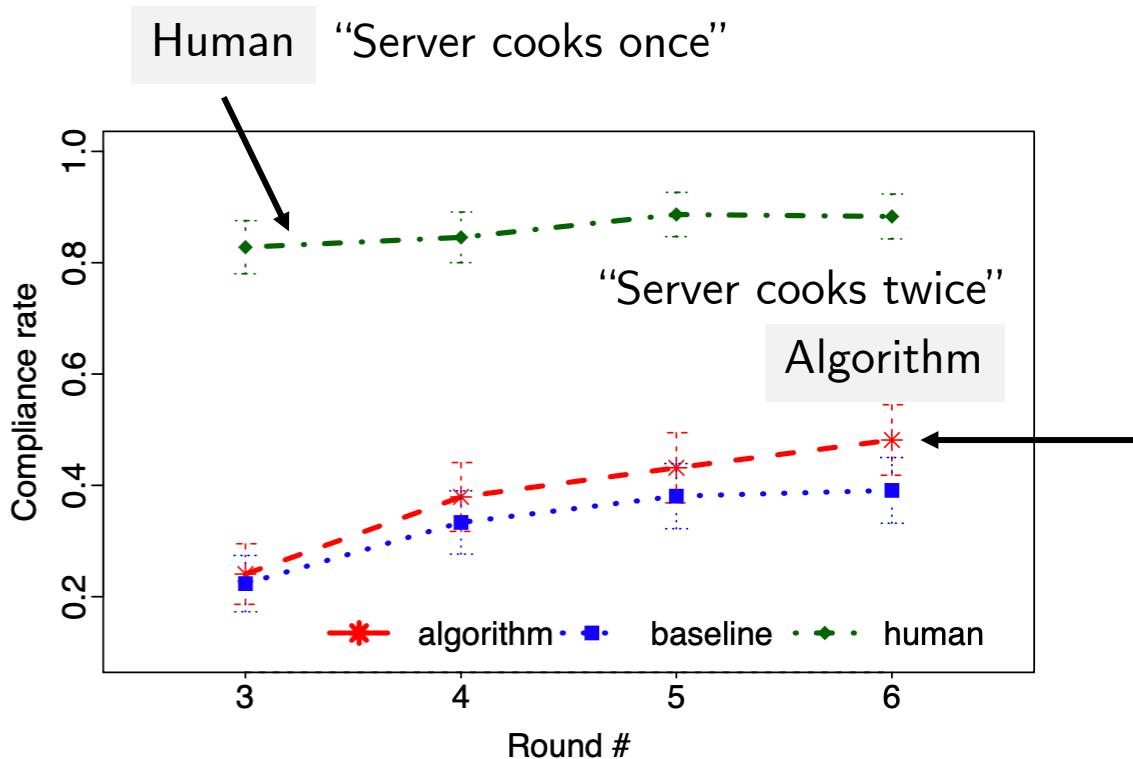
R\_1pA8wDYgWc9hbIt

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

R\_beijQ8guDyExa5r

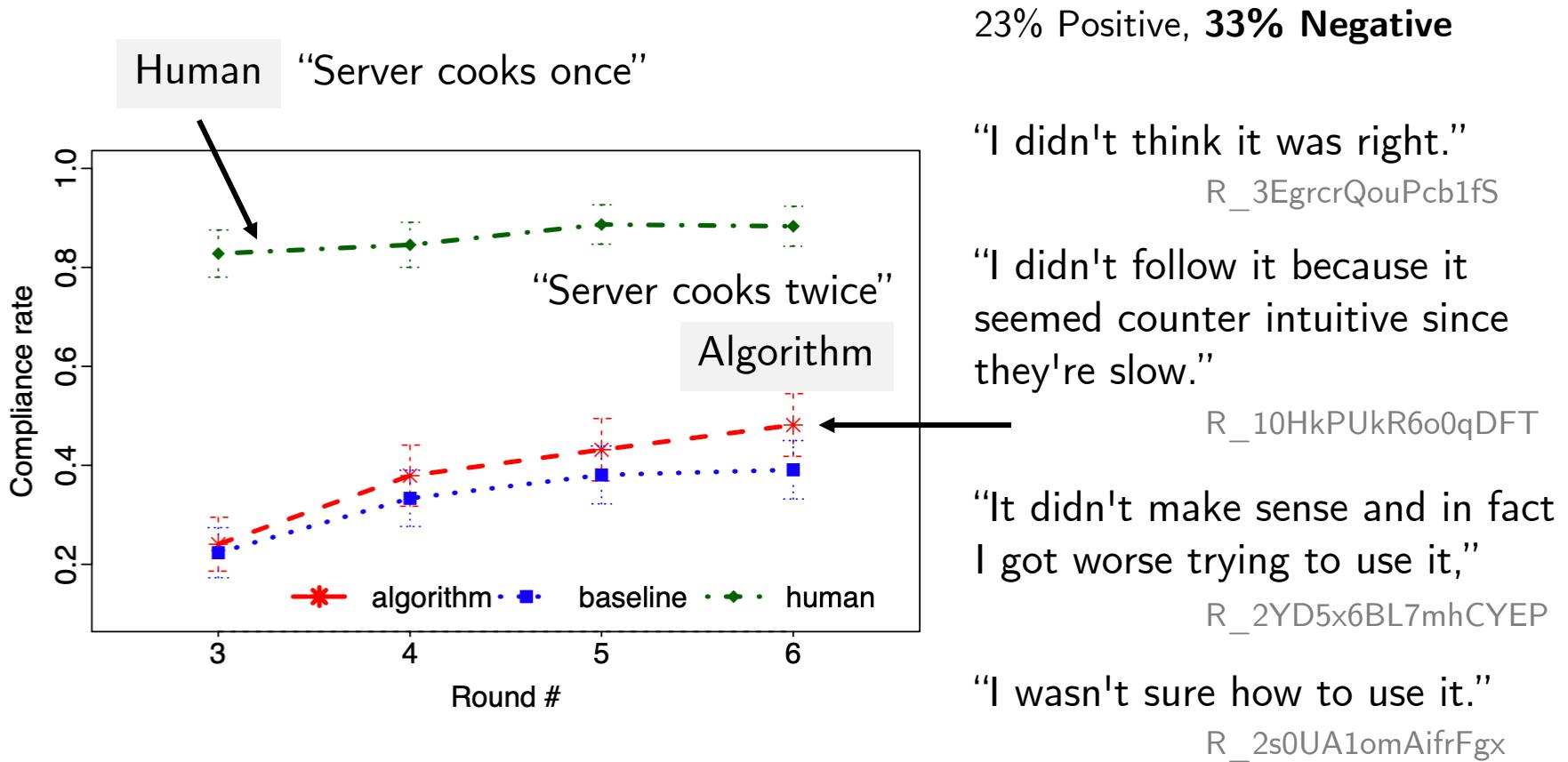
“I used the tip but **I don’t think it was helpful**. The server took long to cook.”

# Study 1: Results Against Counterintuitive Tips



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

# Study 1: Results Against Counterintuitive Tips



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

# Improving Compliance?

# Improving Compliance

## Social information

### Here's how you compare to neighbors



Aug 21, 2015 - Sep 20, 2015

This is based on 87 similar homes within approx. 4 miles. Efficient neighbors are the 20% who use the least amount of electricity.  
See back for details.



You're using more than  
your neighbors.



8% more electricity  
than average neighbors

Allcott 2011, *Journal of Public Economics*

# Improving Compliance

## Social information

“The majority of best players adopted this rule [Server Cook Twice], enabling them to achieve the optimal performance of 34 ticks.”

in all 4 disrupted rounds (3-6)

# Improving Compliance

Social information

“The majority of best players adopted this rule [Server Cook Twice], enabling them to achieve the optimal performance of 34 ticks.”

in all 4 disrupted rounds (3-6)

“Pay” – incentive to try

# Improving Compliance

Social information

“The majority of best players adopted this rule [Server Cook Twice], enabling them to achieve the optimal performance of 34 ticks.”

in all 4 disrupted rounds (3-6)

“Pay” – incentive to try

“You’ll earn the maximum bonus if server cooks twice in this round.”

in rounds 3-4, back to original scheme in rounds 5-6

# Improving Compliance

Social information

“The majority of best players adopted this rule [Server Cook Twice], enabling them to achieve the optimal performance of 34 ticks.”

in all 4 disrupted rounds (3-6)

“Pay” – incentive to try

“You’ll earn the maximum bonus if server cooks twice in this round.”

in rounds 3-4, back to original scheme in rounds 5-6

“Curriculum” – pacing learning

# Improving Compliance

Social information

“The majority of best players adopted this rule [Server Cook Twice], enabling them to achieve the optimal performance of 34 ticks.”

in all 4 disrupted rounds (3-6)

“Pay” – incentive to try

“You’ll earn the maximum bonus if server cooks twice in this round.”

in rounds 3-4, back to original scheme in rounds 5-6

“Curriculum” – pacing learning

Algorithm

Server  
should cook twice

Human

Server  
should cook once

Hypothetical

Server  
shouldn’t cook

# Improving Compliance

Social information

“The majority of best players adopted this rule [Server Cook Twice], enabling them to achieve the optimal performance of 34 ticks.”

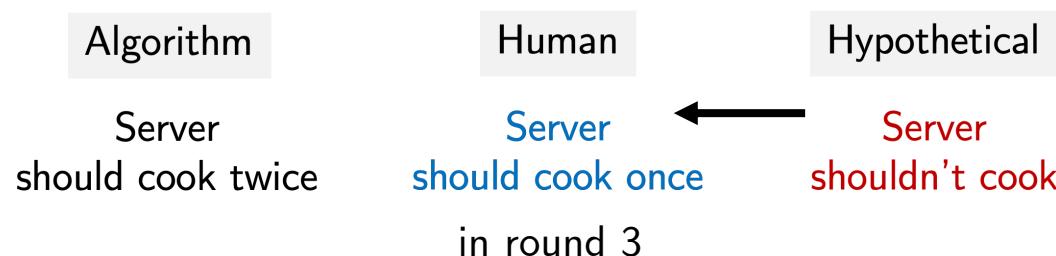
in all 4 disrupted rounds (3-6)

“Pay” – incentive to try

“You’ll earn the maximum bonus if server cooks twice in this round.”

in rounds 3-4, back to original scheme in rounds 5-6

“Curriculum” – pacing learning



# Improving Compliance

Social information

“The majority of best players adopted this rule [Server Cook Twice], enabling them to achieve the optimal performance of 34 ticks.”

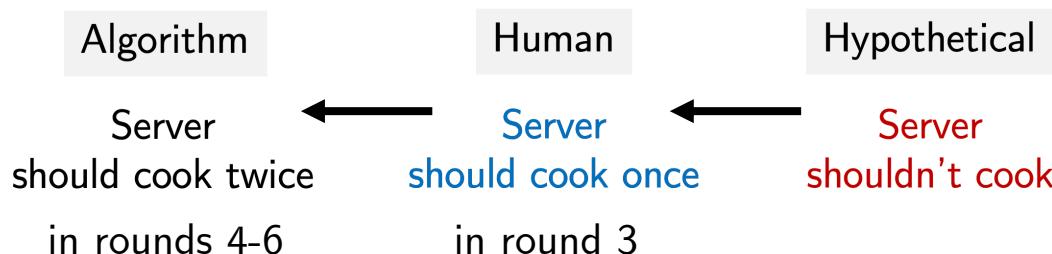
in all 4 disrupted rounds (3-6)

“Pay” – incentive to try

“You’ll earn the maximum bonus if server cooks twice in this round.”

in rounds 3-4, back to original scheme in rounds 5-6

“Curriculum” – pacing learning



# Improving Compliance

Social information

“The majority of best players adopted this rule [Server Cook Twice], enabling them to achieve the optimal performance of 34 ticks.”

“Pay + Social”

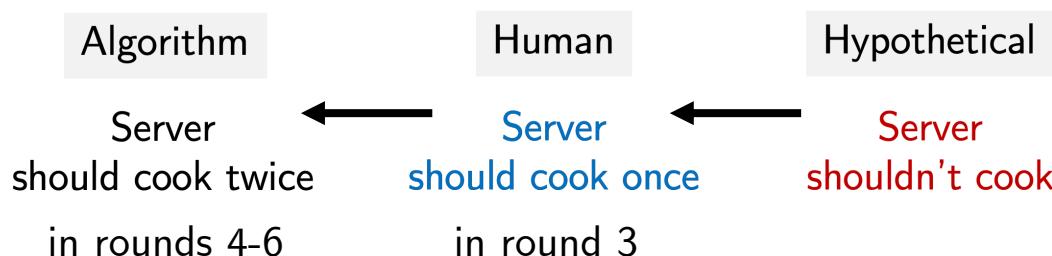
in all 4 disrupted rounds (3-6)

“Pay” – incentive to try

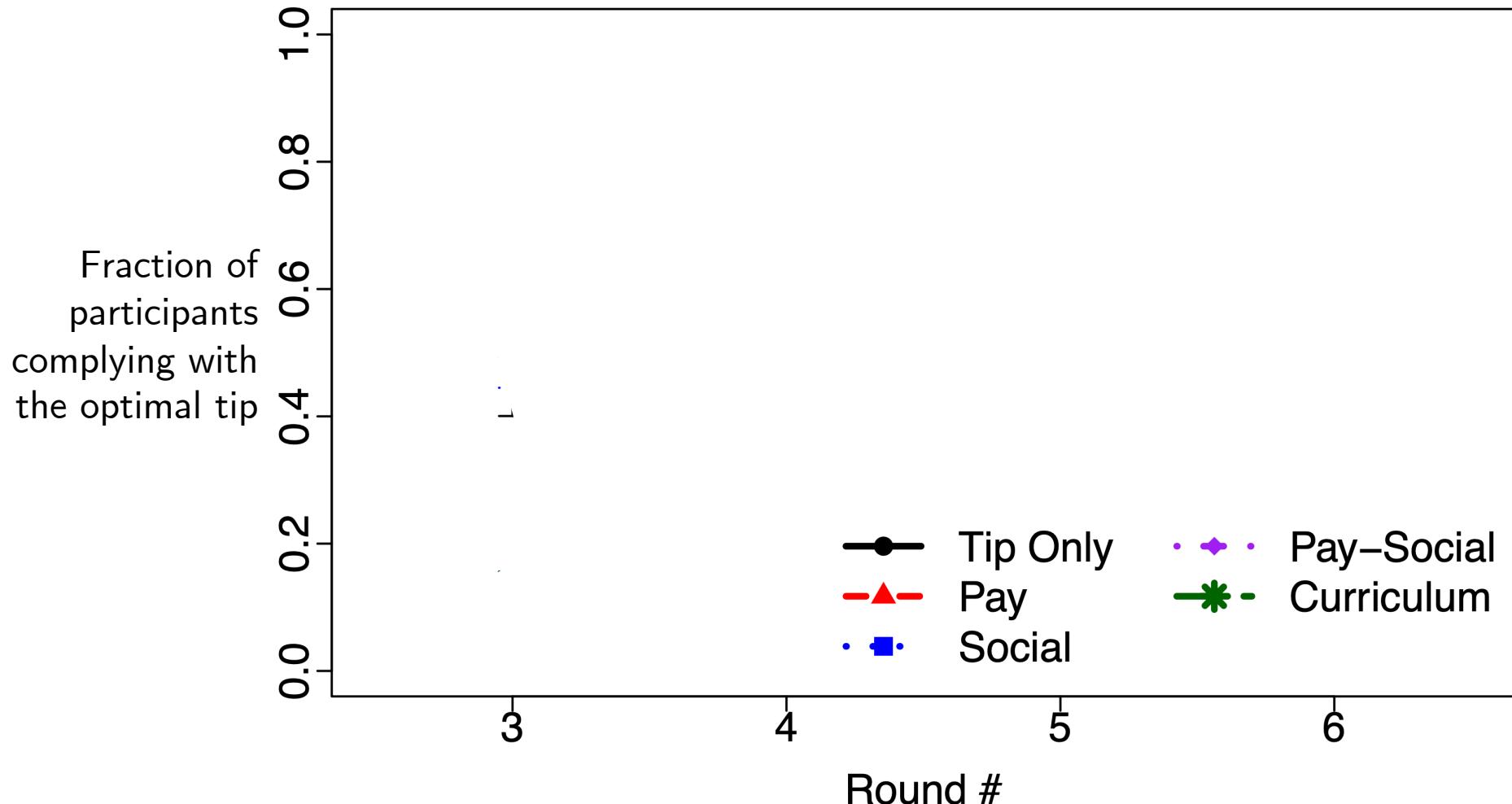
“You’ll earn the maximum bonus if server cooks twice in this round.”

in rounds 3-4, back to original scheme in rounds 5-6

“Curriculum” – pacing learning

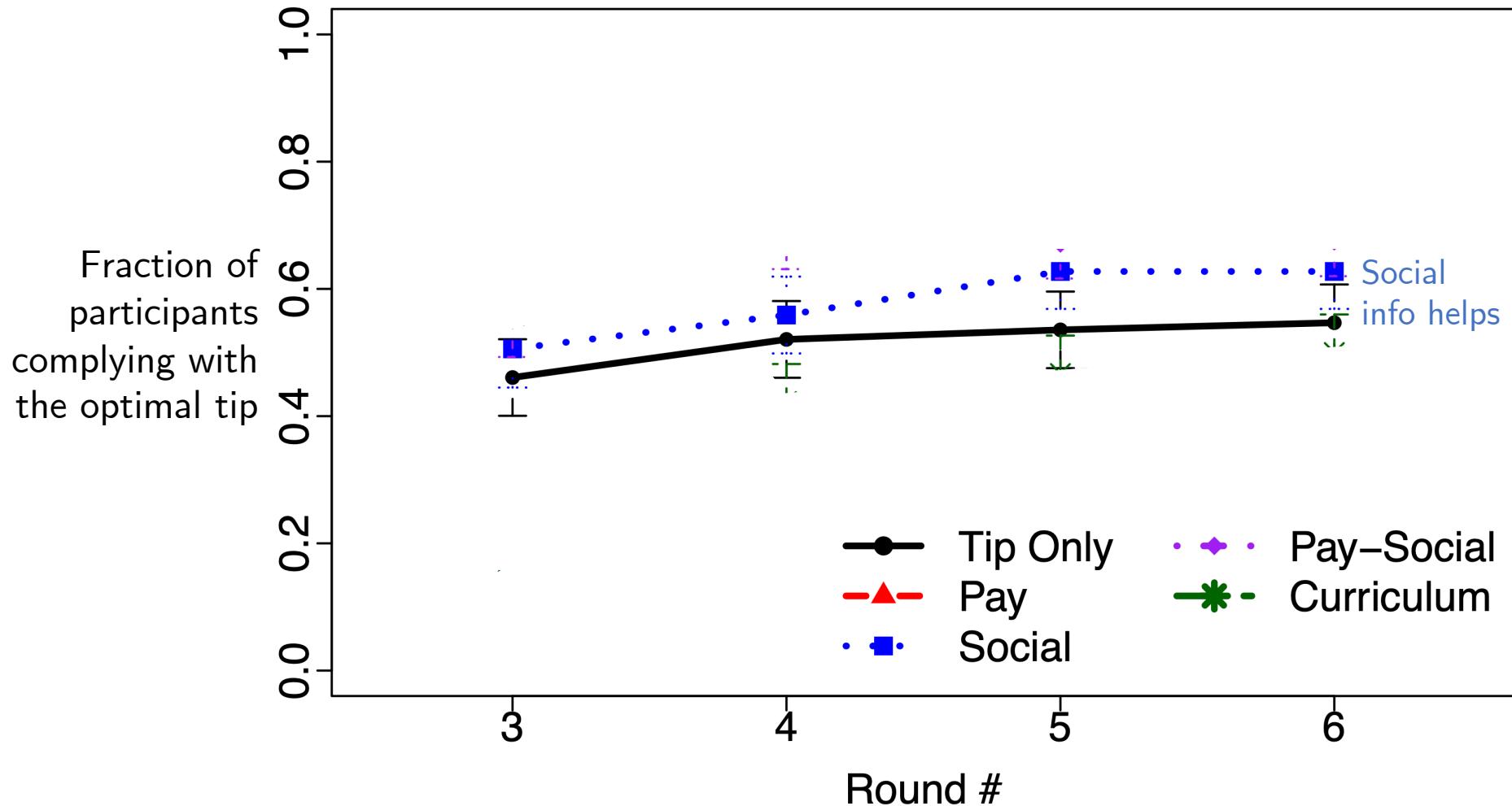


# Improving Compliance

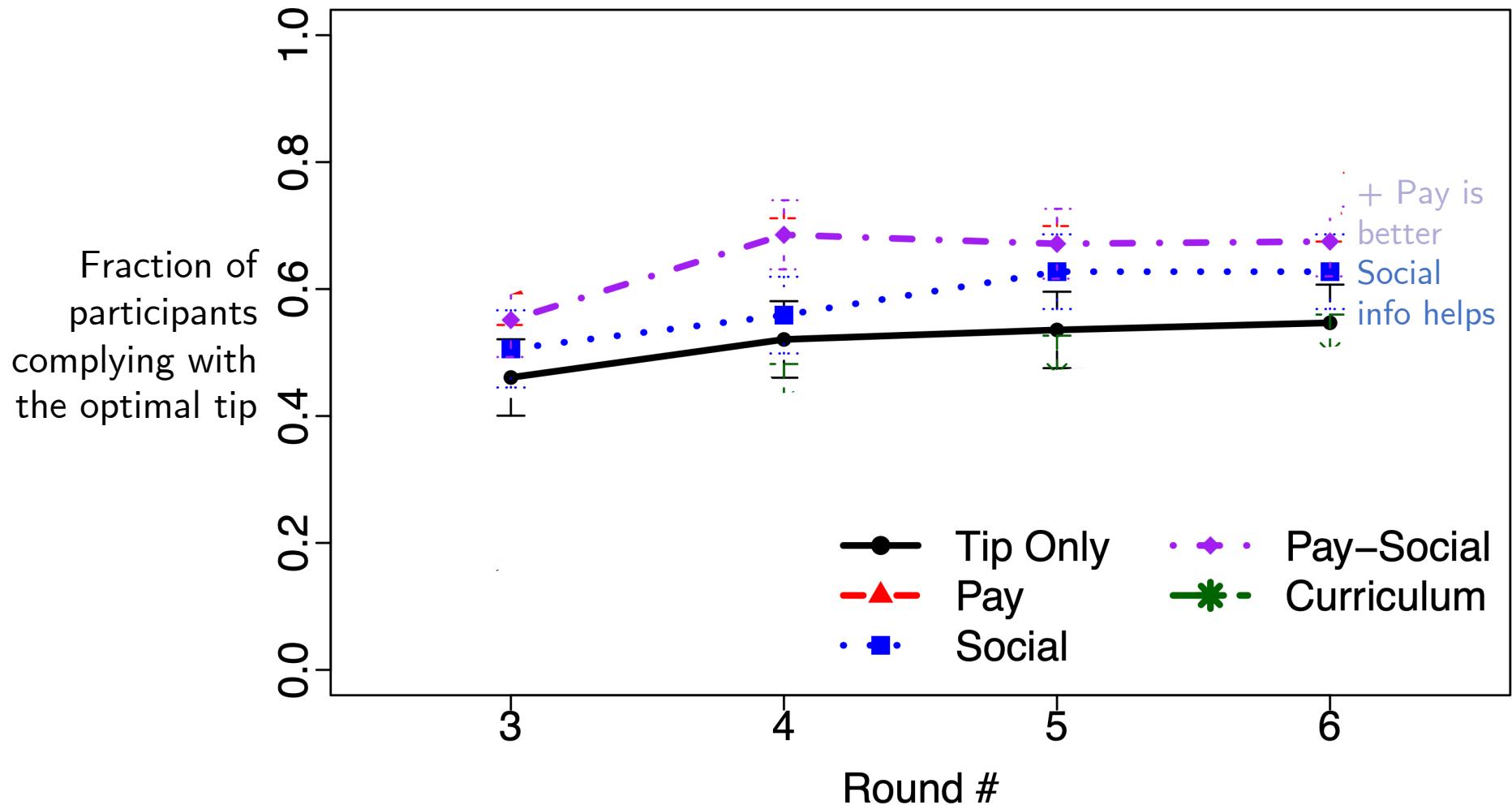


Amazon Mechanical Turk, N = 1,416

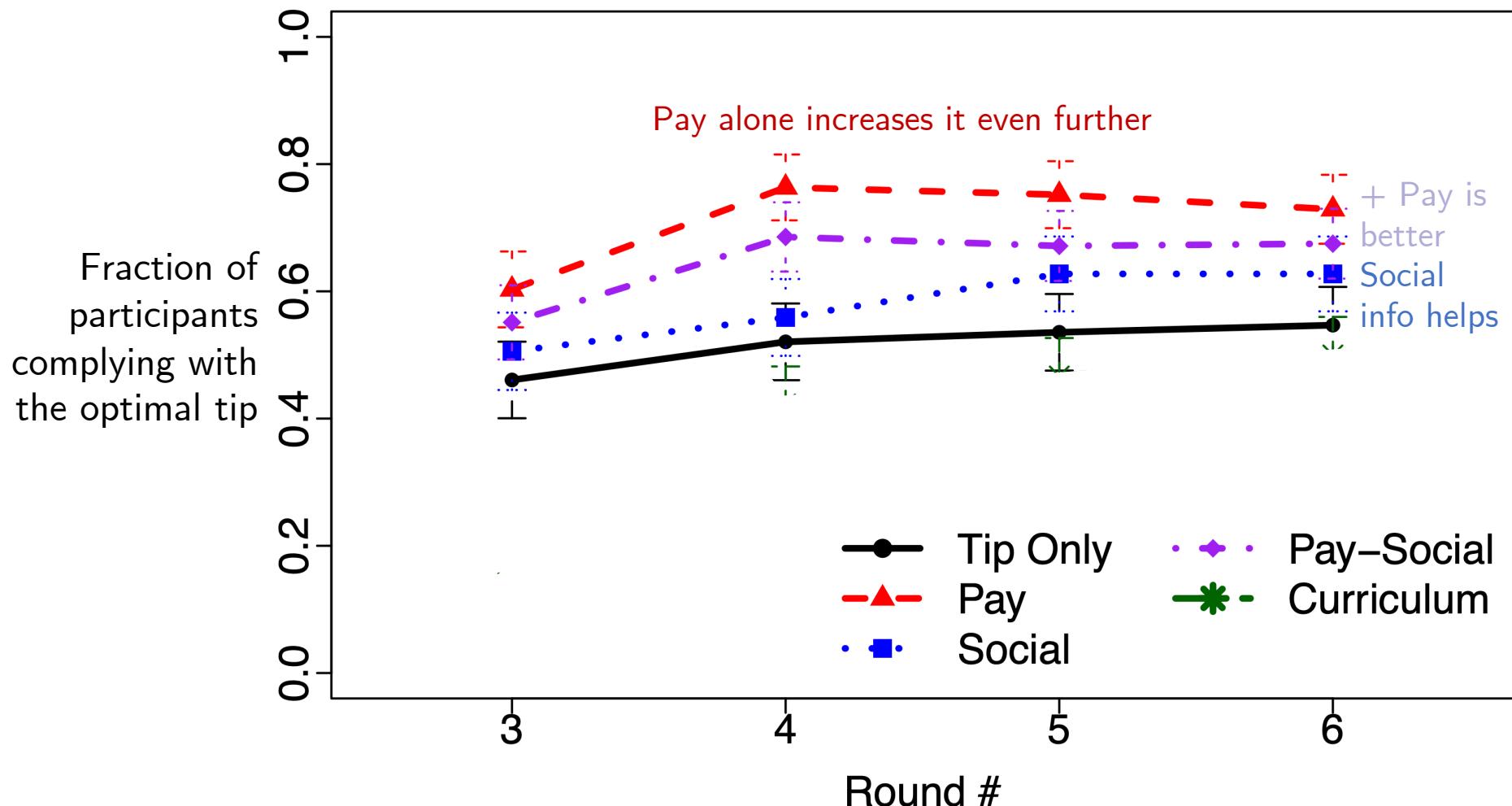
# Improving Compliance



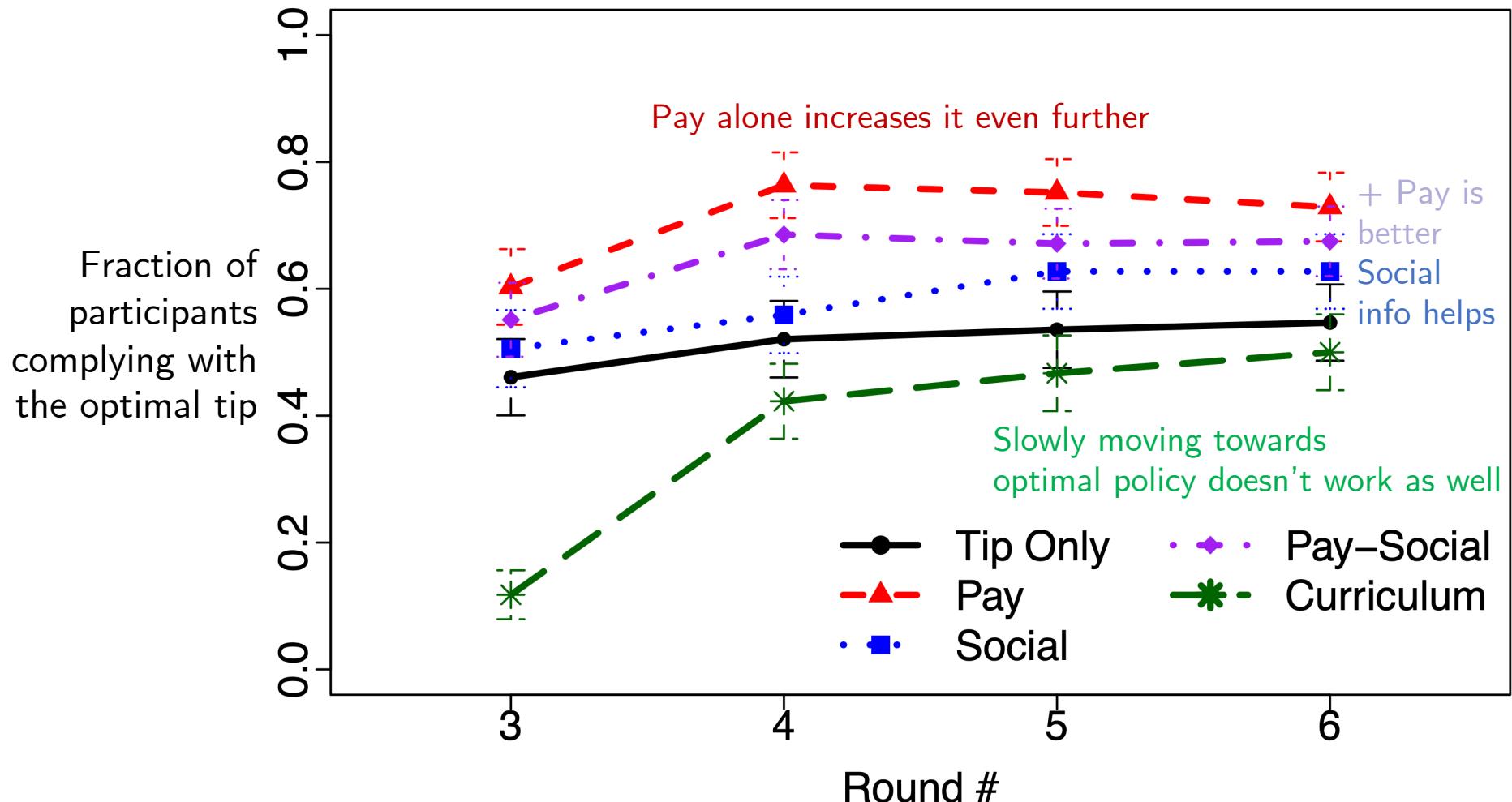
# Improving Compliance



# Improving Compliance



# Improving Compliance



# Summary

ML framework to leverage behavioral trace data to infer simple tips that help humans



# Summary

ML framework to leverage behavioral trace data to infer simple tips that help humans



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

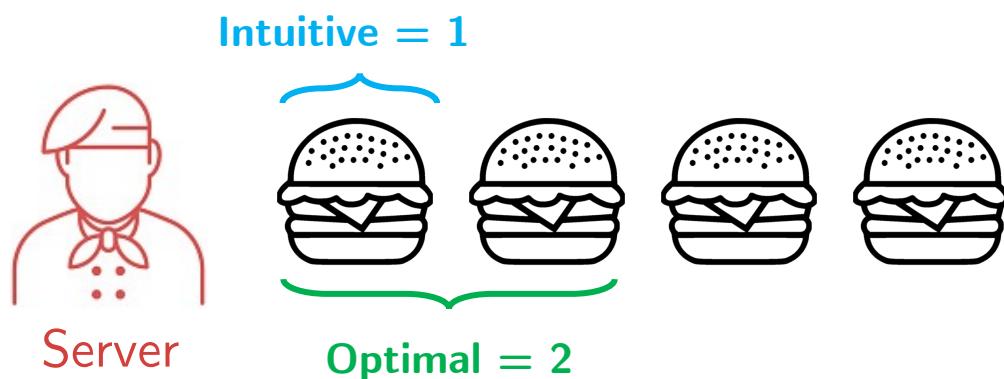
Performance/compliance tradeoff

with Hamsa Bastani & Osbert Bastani  
*Minor Revision @ Management Science*

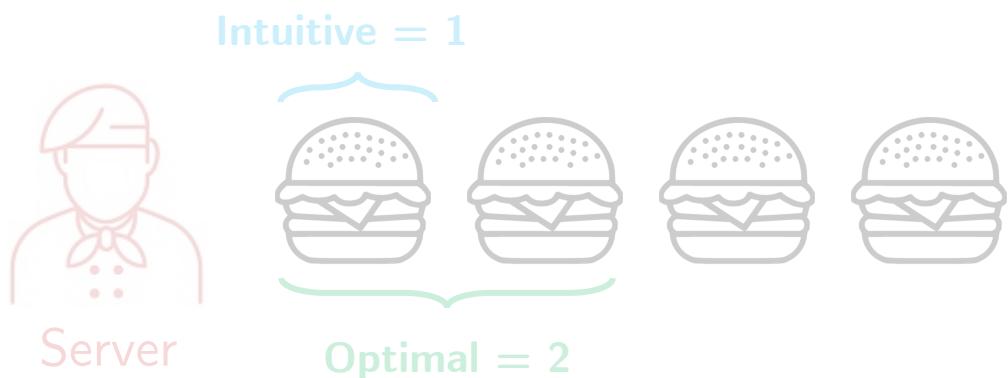


(Available at: [bit.ly/tipspaper](https://bit.ly/tipspaper))

# Study 1:

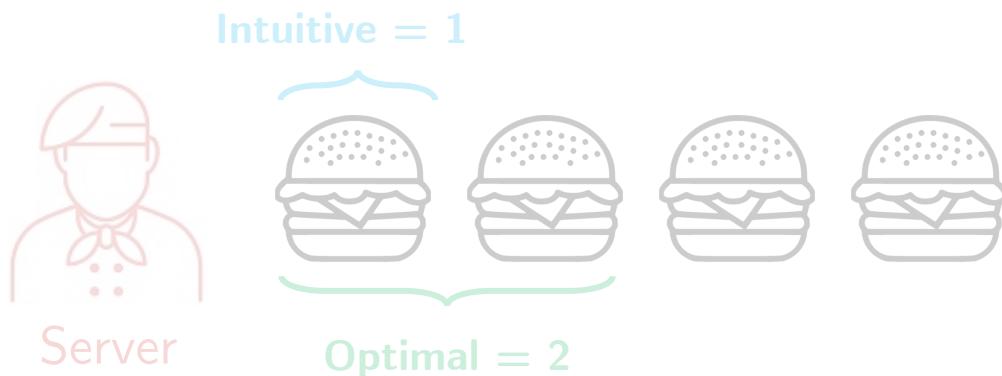


# What If Decisions Are More Granular?



# What If Decisions Are More Granular?

## Also, What If Things Are Uncertain?



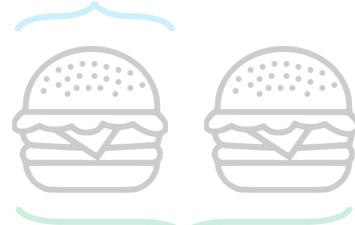
## Study 2:

**What If Decisions Are More Granular?**

**Also, What If Things Are Uncertain?**

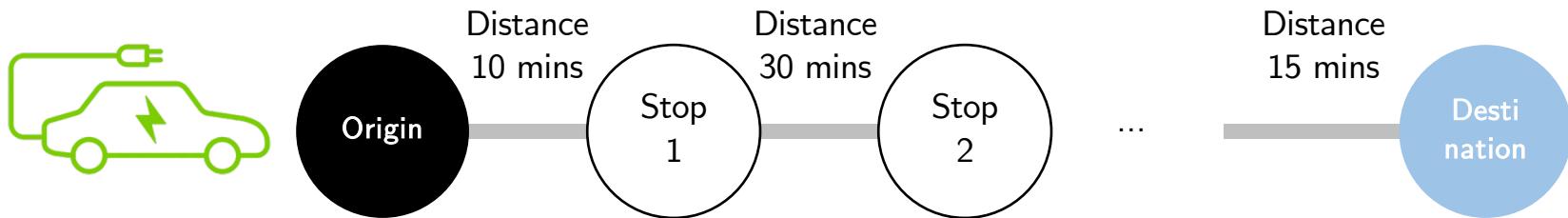


Intuitive = 1



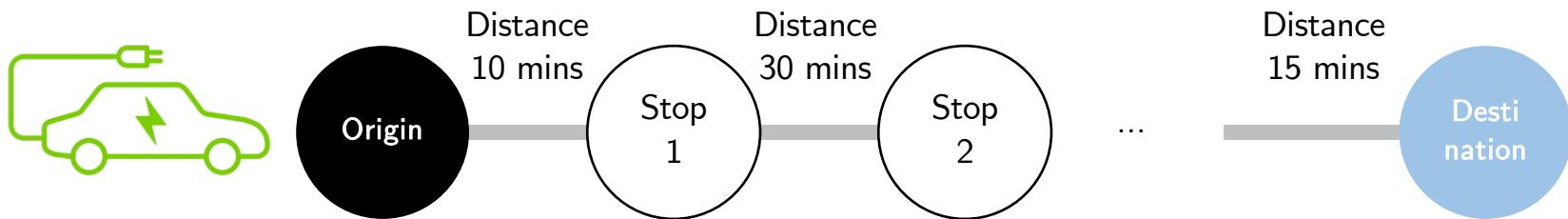
## Study 2:

# EV Charging Game



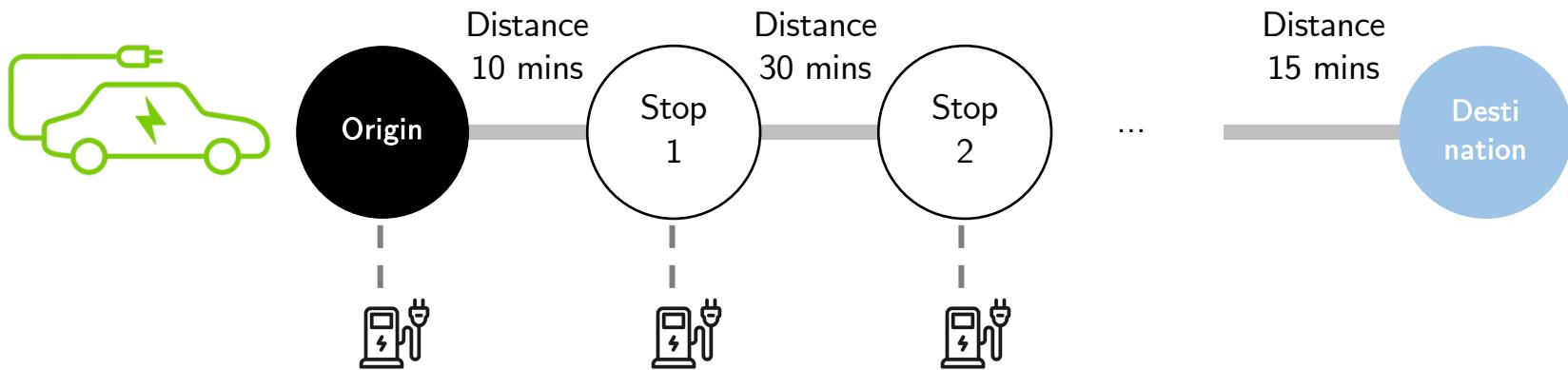
## Study 2:

# EV Charging Game



**Goal:** Get to destination as fast as possible (**1% charge can travel 1 minute**)

# EV Charging Game

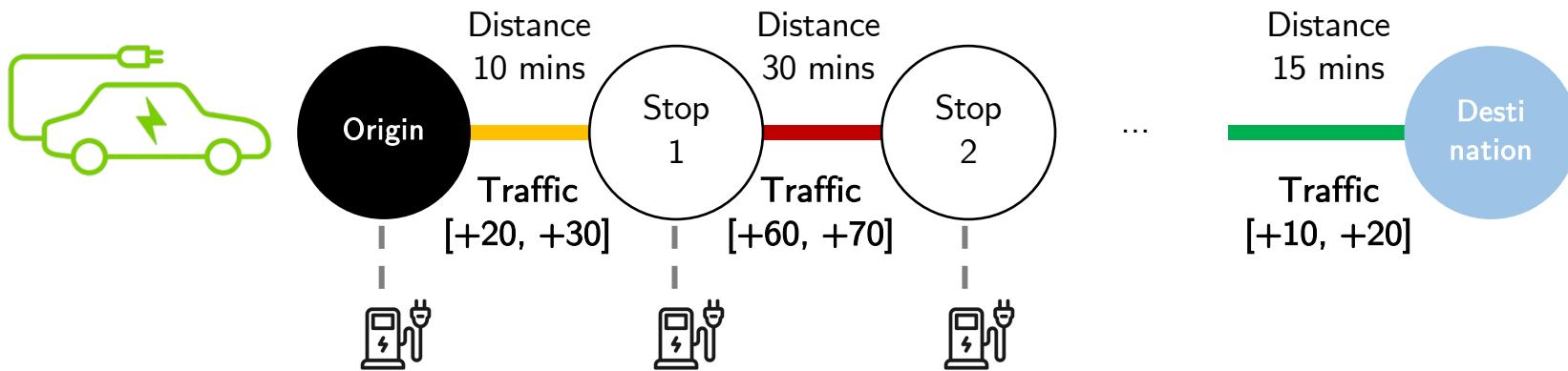


**Goal:** Get to destination as fast as possible (**1% charge can travel 1 minute**)

**Challenges:** Even without traffic, the amount of initial charge won't be enough

## Study 2:

# EV Charging Game

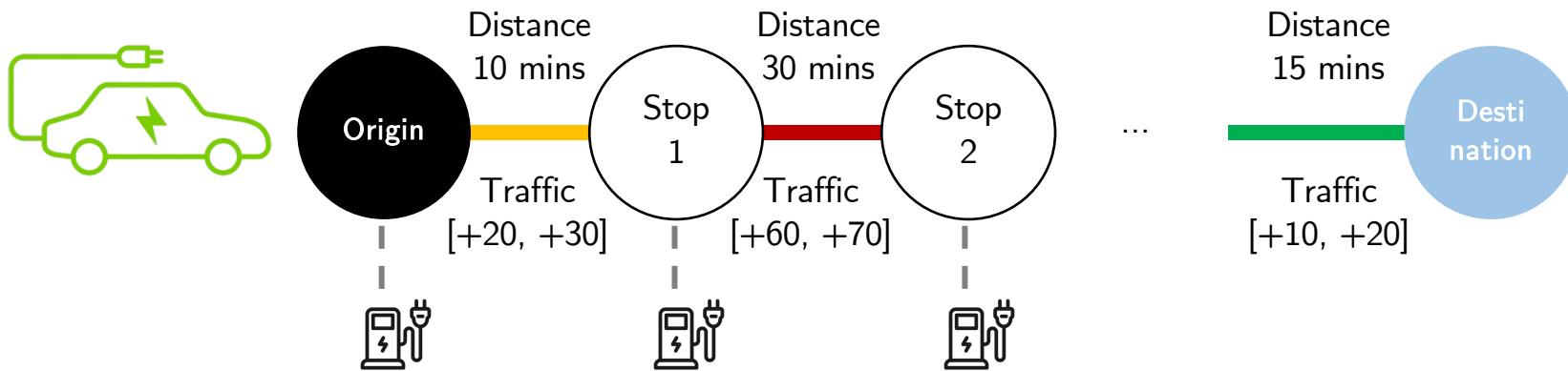


**Goal:** Get to destination as fast as possible (**1% charge can travel 1 minute**)

**Challenges:** Even without traffic, the amount of initial charge won't be enough  
+ Uncertain traffic, uniformly distributed within the specified range

## Study 2:

# EV Charging Game



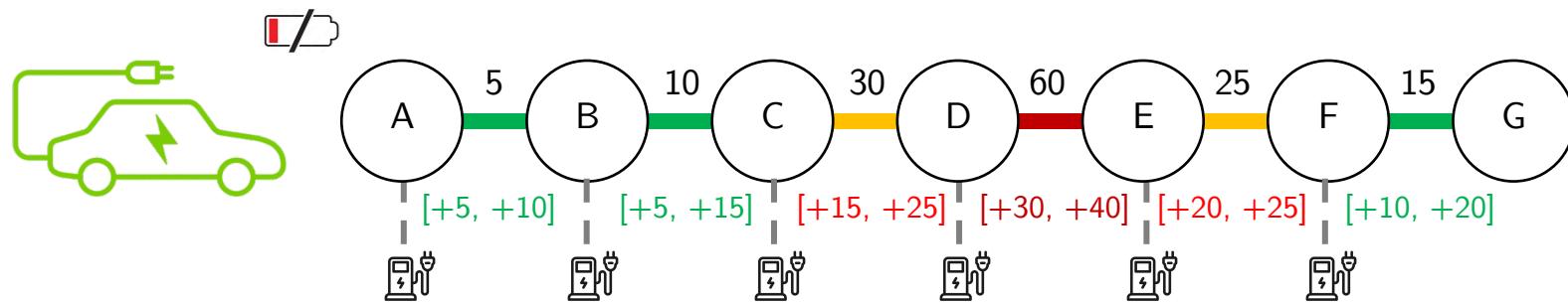
**Goal:** Get to destination as fast as possible (**1% charge can travel 1 minute**)

**Challenges:** Even without traffic, the amount of initial charge won't be enough

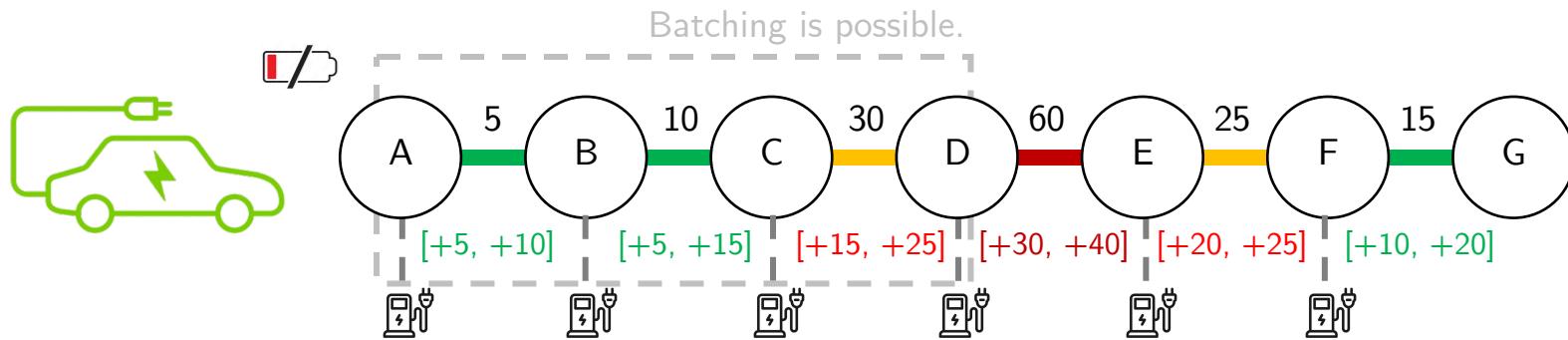
- + Uncertain traffic, uniformly distributed within the specified range
- + Cost to exit to charge (**+30 mins**) and nonlinear charging time
- + Penalty if running out of charge in the middle (**+300 mins**)

# Design

## To Batch, or Not to Batch

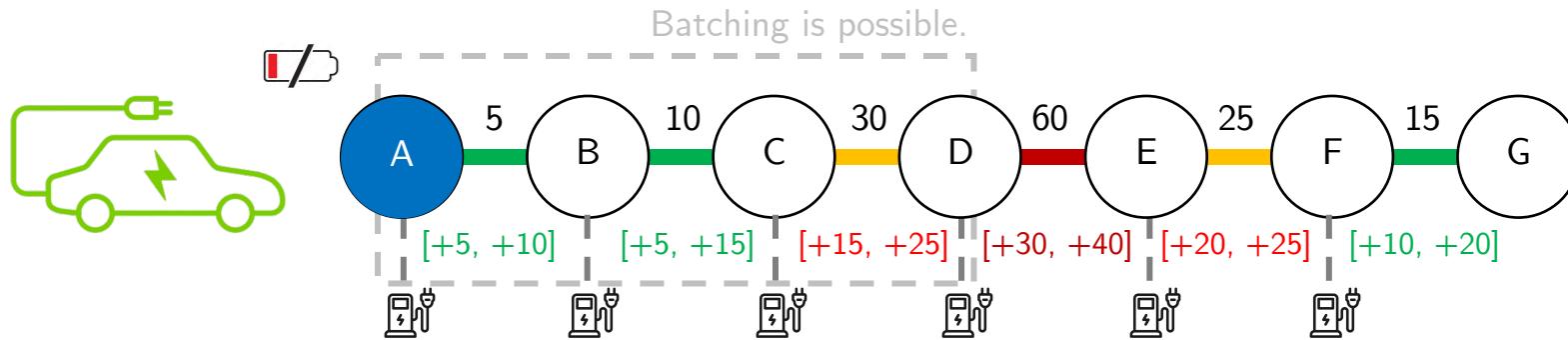


# Design To Batch, or Not to Batch



# Design

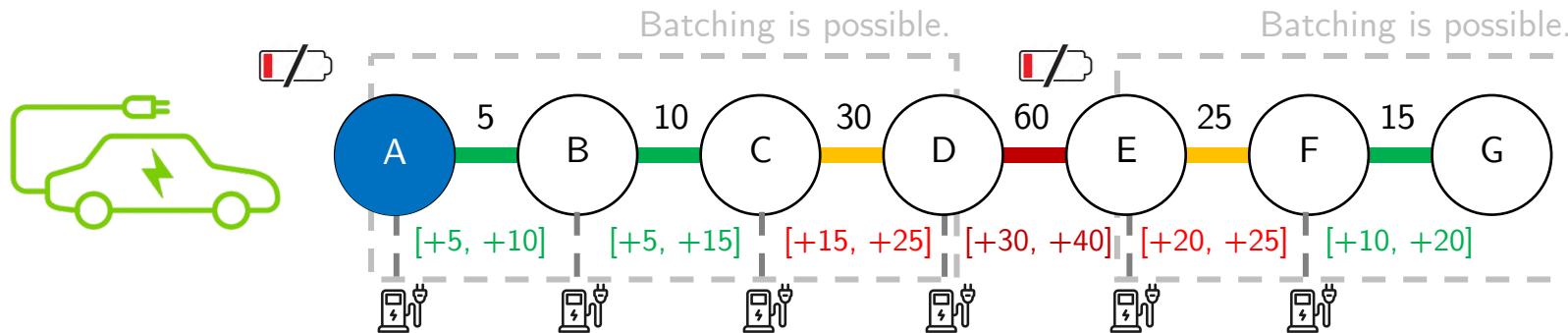
## To Batch, or Not to Batch



Optimal is to “batch” the required charges  
for the next two stops ( $A \rightarrow C$ )  
rather than just  $A \rightarrow B$  or further batch  $A \rightarrow D$ .

# Design

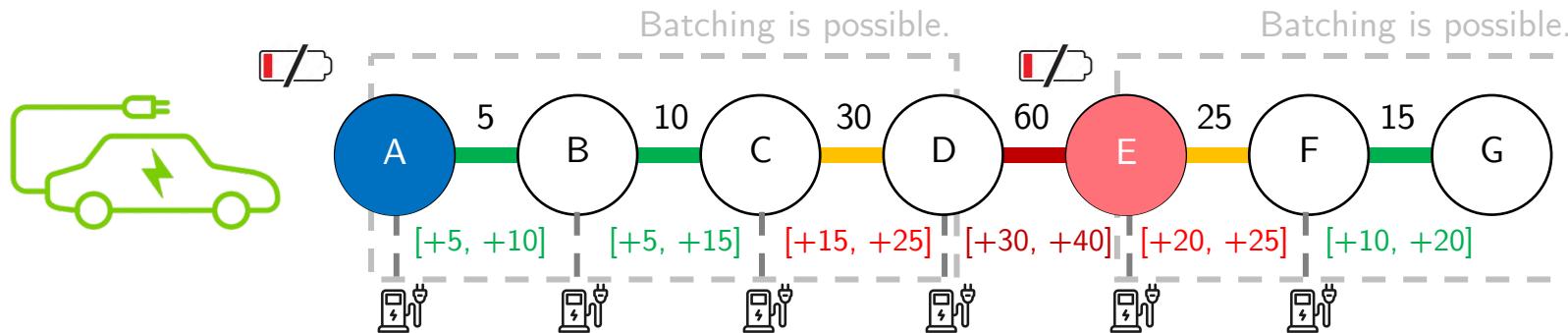
## To Batch, or Not to Batch



Optimal is to “batch” the required charges  
for the next two stops ( $A \rightarrow C$ )  
rather than just  $A \rightarrow B$  or further batch  $A \rightarrow D$ .

# Design

## To Batch, or Not to Batch



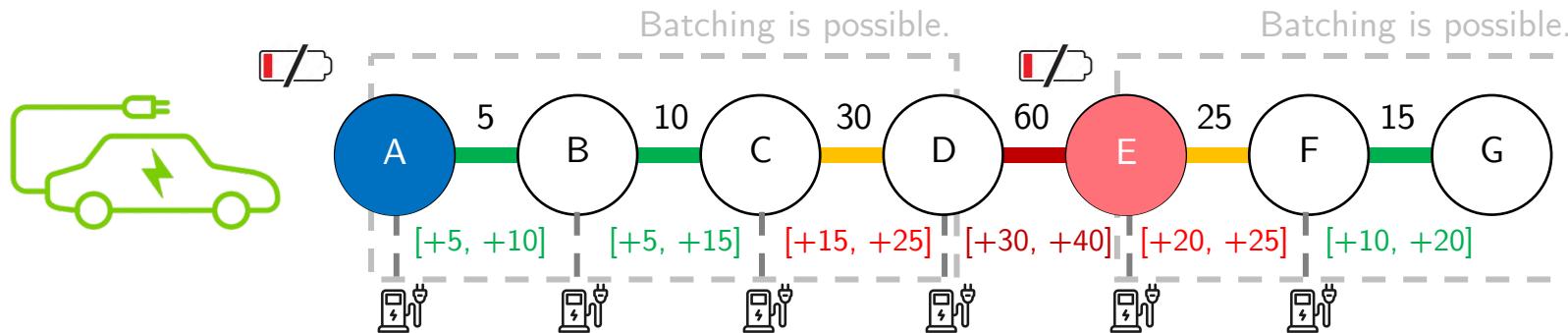
Optimal is to “batch” the required charges for the next two stops ( $A \rightarrow C$ ) rather than just  $A \rightarrow B$  or further batch  $A \rightarrow D$ .



Optimal is to “split” = only charge for the next stop ( $E \rightarrow F$ ) rather than batch  $E \rightarrow G$ .

# Study 2A:

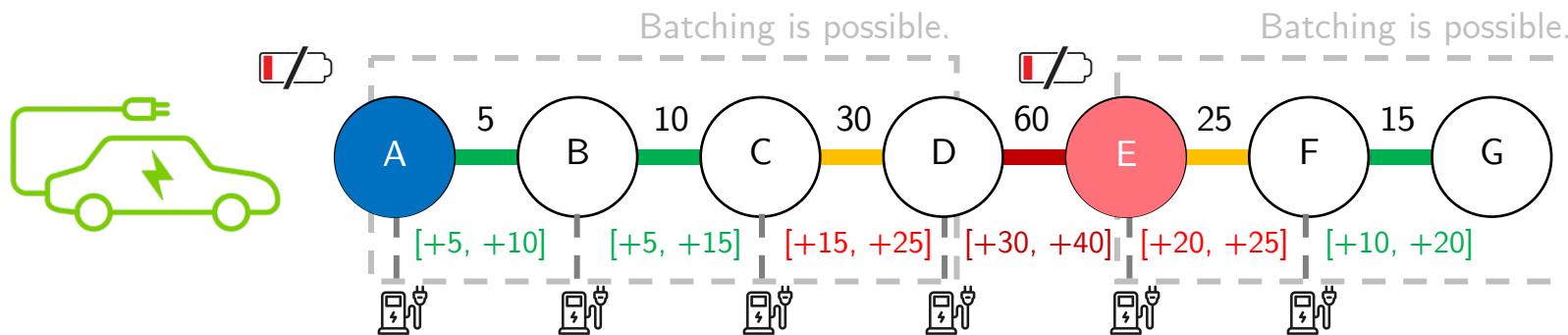
## Design Flow



Performance: Time to destination

# Study 2A:

## Design Flow



Performance: Time to destination

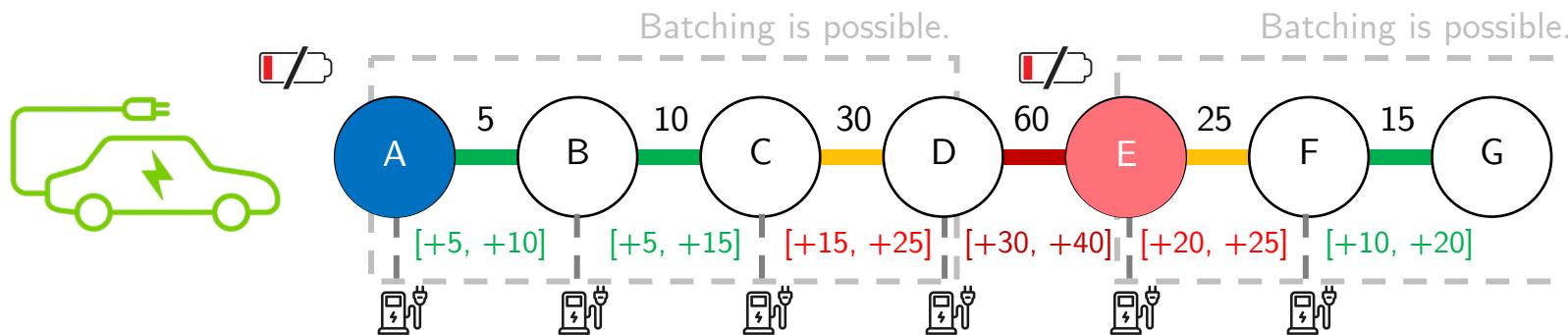
Round 1

Round 2

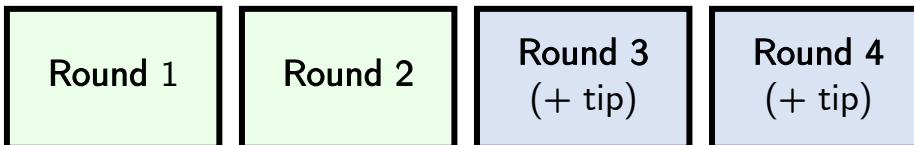
Realized traffic will be different across rounds but drawn from the same distributions

# Study 2A:

## Design Flow



Performance: Time to destination

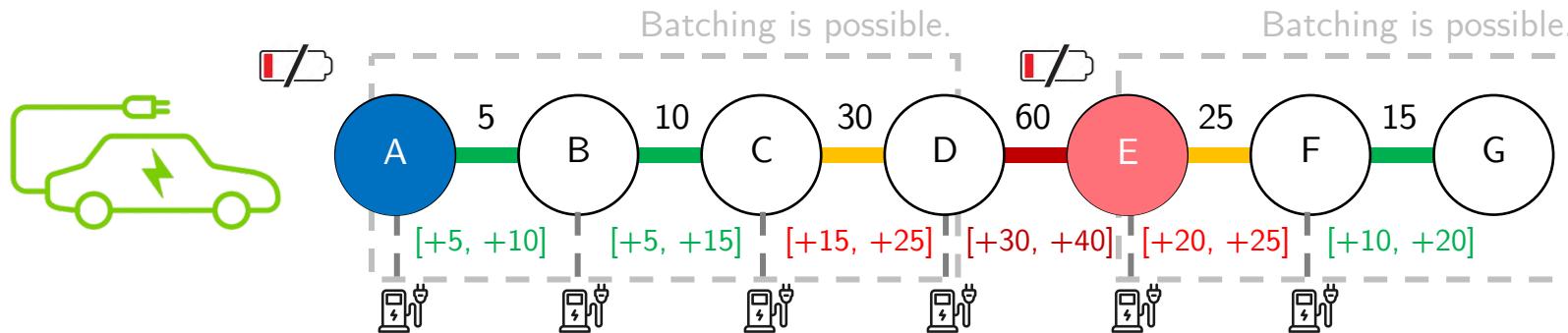


Realized traffic will be different across rounds but drawn from the same distributions

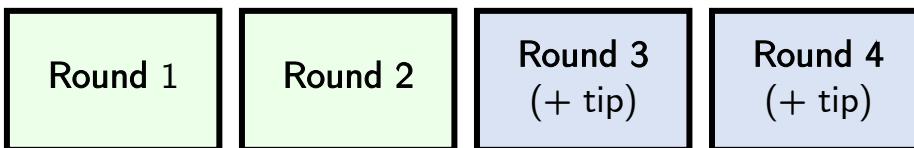
# Study 2A:

# Design

## Tip Conditions: *Precise vs Broad*



Performance: Time to destination



Realized traffic will be different across rounds but drawn from the same distributions

At every stop, we present either...

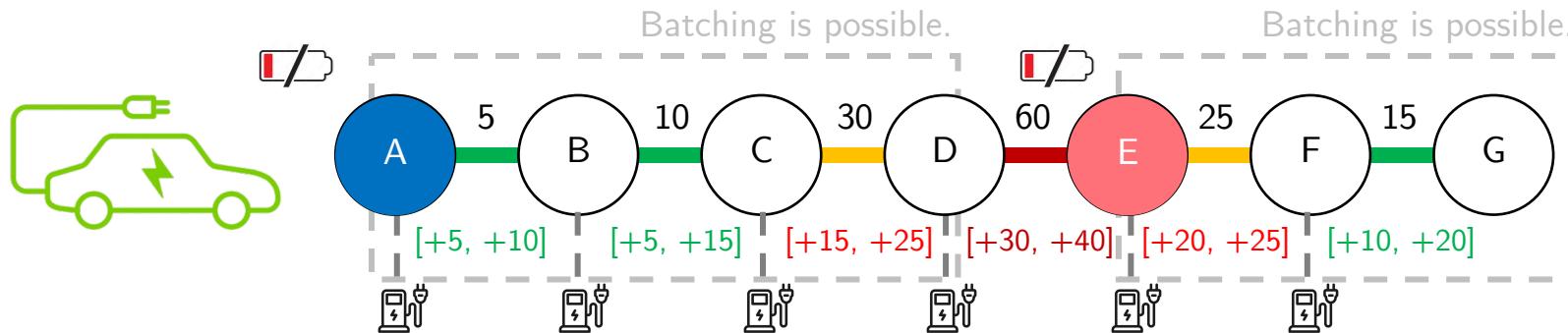
**Precise Tip:**

You should exit and charge X%

# Study 2A:

# Design

## Tip Conditions: *Precise vs Broad*



Performance: Time to destination



Realized traffic will be different across rounds but drawn from the same distributions

At every stop, we present either...

**Precise Tip:**

You should exit and charge X%

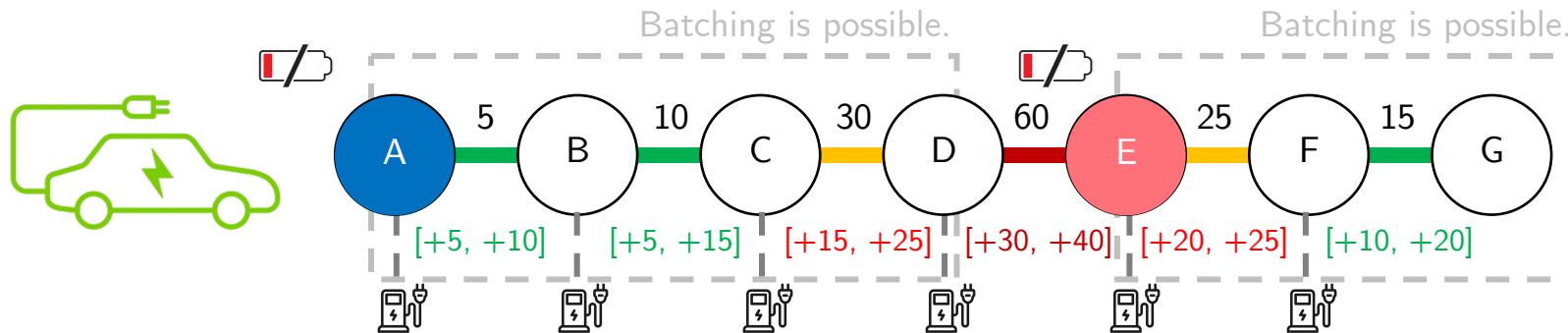
**Broad Tip:**

"Split" You should charge enough for this segment

# Study 2A:

# Design

## Tip Conditions: *Precise vs Broad*



Performance: Time to destination



Realized traffic will be different across rounds but drawn from the same distributions

At every stop, we present either...

**Precise Tip:**

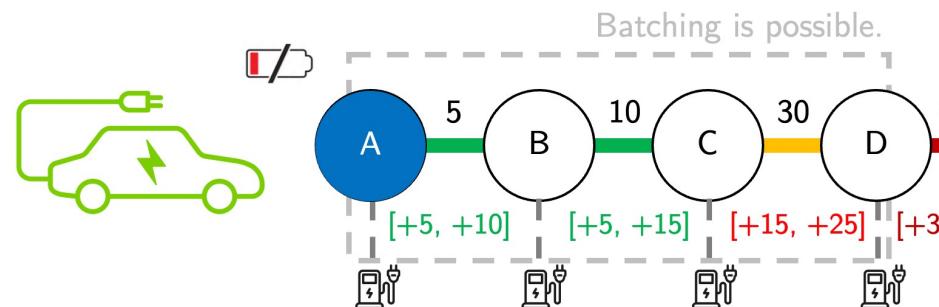
You should exit and charge X%

**Broad Tip:**

“Split” You should charge enough for this segment

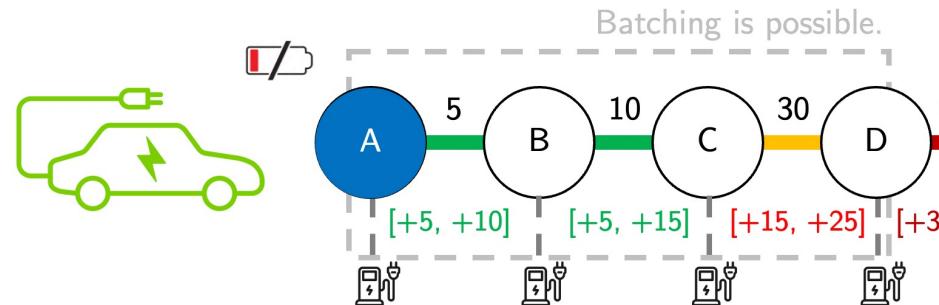
“Batch” You should charge enough for this segment + next one

# Study 2A: Results



Exit A: Forced to charge  
Optimal: Batch A → C  
rather than just A→B  
or further batch A→D.  
("aftercharge" = 25-40)

# Study 2A: Results



Exit A: Forced to charge  
Optimal: Batch A → C  
rather than just A → B  
or further batch A → D.  
("aftercharge" = 25-40)

## Aftercharge

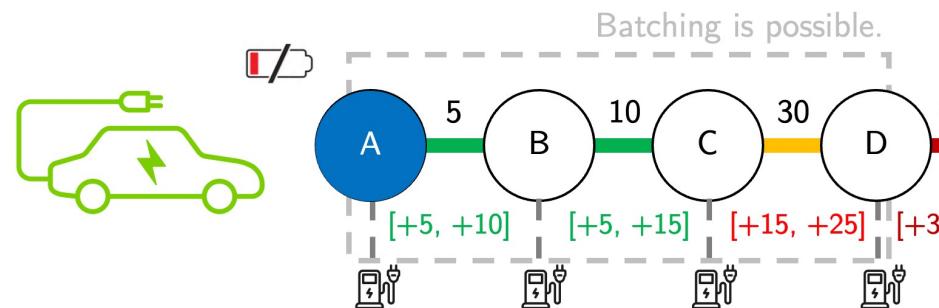
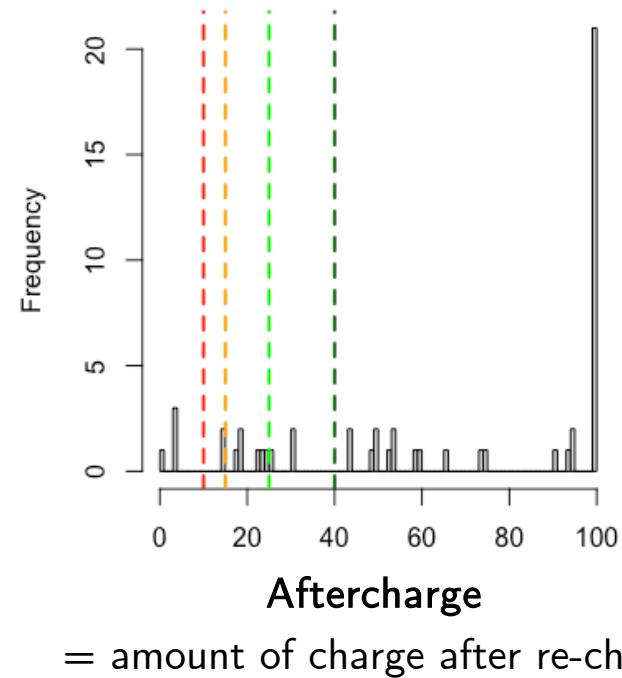
= amount of charge after re-charging

# Study 2A:

# Results

## Wide Range of Decisions

Round 1

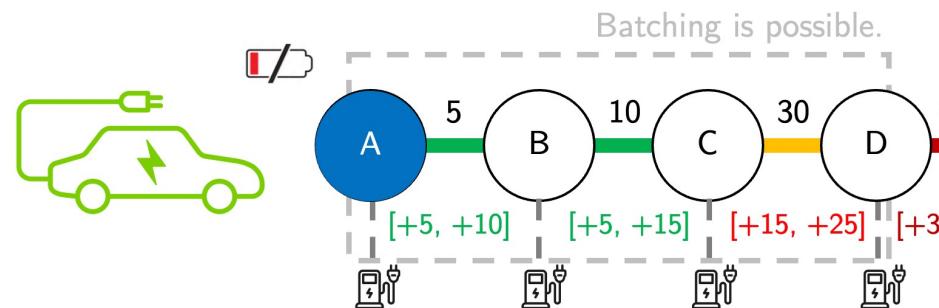
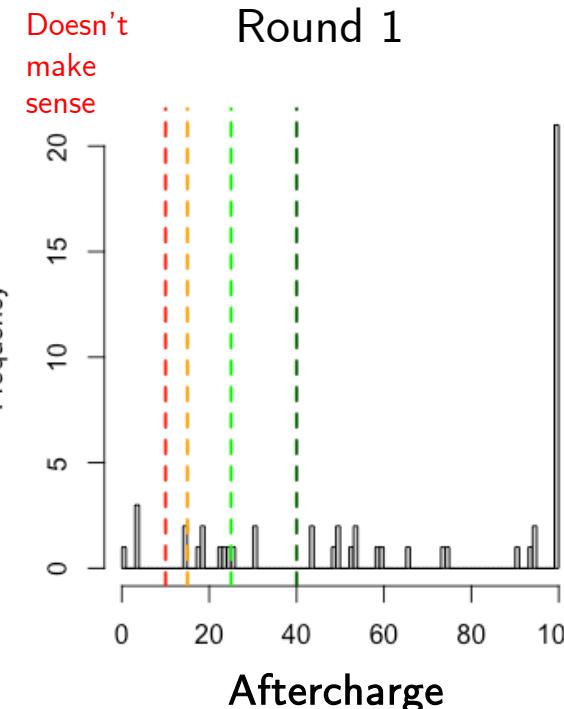


Exit A: Forced to charge  
Optimal: Batch A → C  
rather than just A→B  
or further batch A→D.  
("aftercharge" = 25-40)

# Study 2A:

# Results

## Wide Range of Decisions

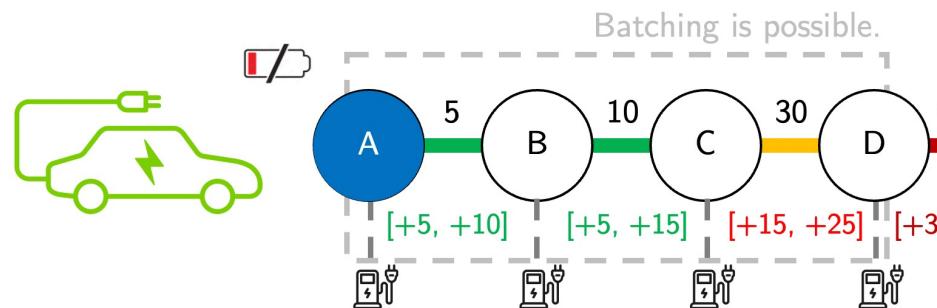
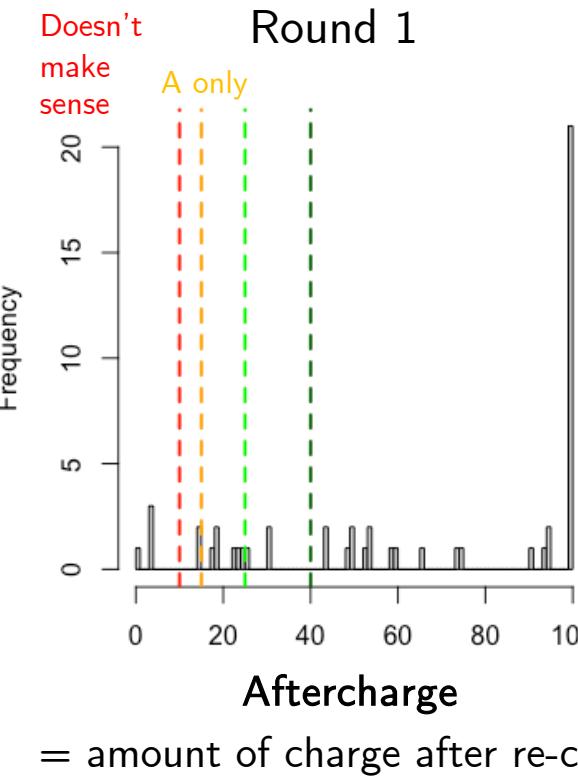


Exit A: Forced to charge  
Optimal: Batch A → C  
rather than just A → B  
or further batch A → D.  
("aftercharge" = 25-40)

# Study 2A:

# Results

## Wide Range of Decisions

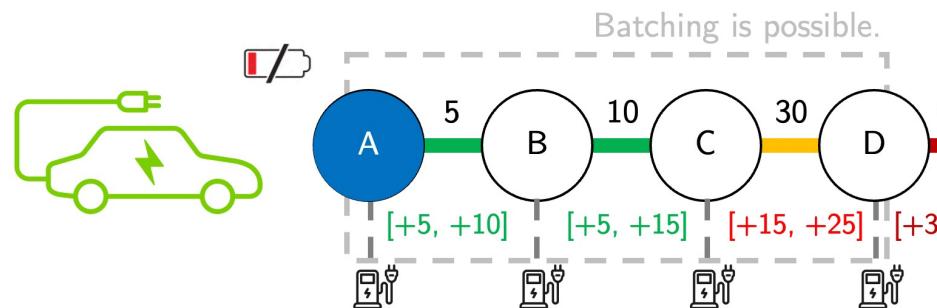
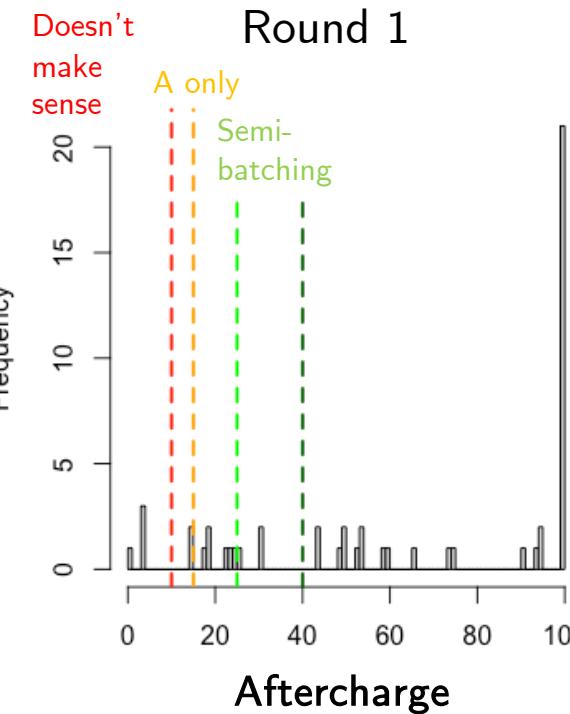


Exit A: Forced to charge  
Optimal: Batch A → C  
rather than just A→B  
or further batch A→D.  
("aftercharge" = 25-40)

# Study 2A:

# Results

## Wide Range of Decisions

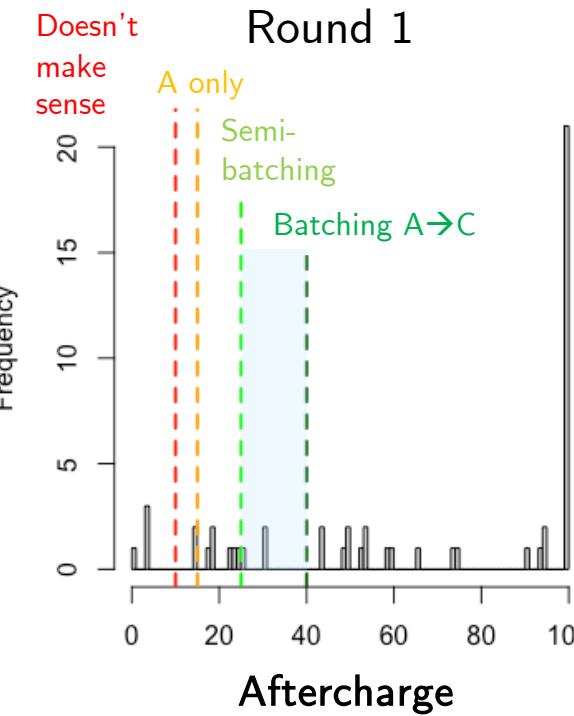


Exit A: Forced to charge  
Optimal: Batch A → C  
rather than just A → B  
or further batch A → D.  
("aftercharge" = 25-40)

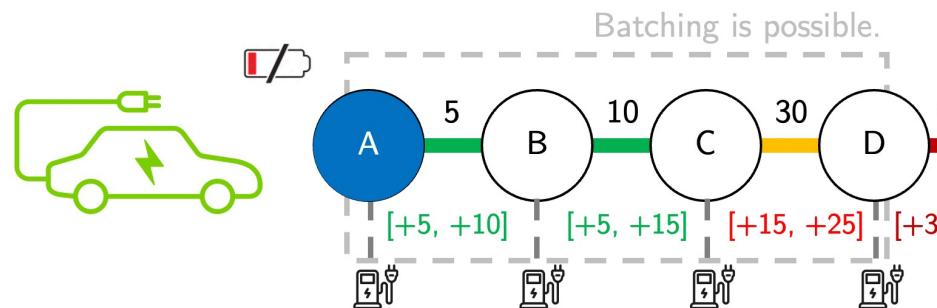
# Study 2A:

# Results

## Wide Range of Decisions



= amount of charge after re-charging

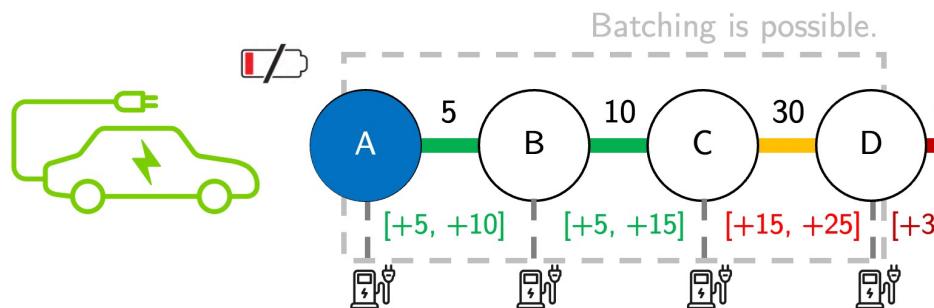


Exit A: Forced to charge  
Optimal: Batch A → C  
rather than just A→B  
or further batch A→D.  
("aftercharge" = 25-40)

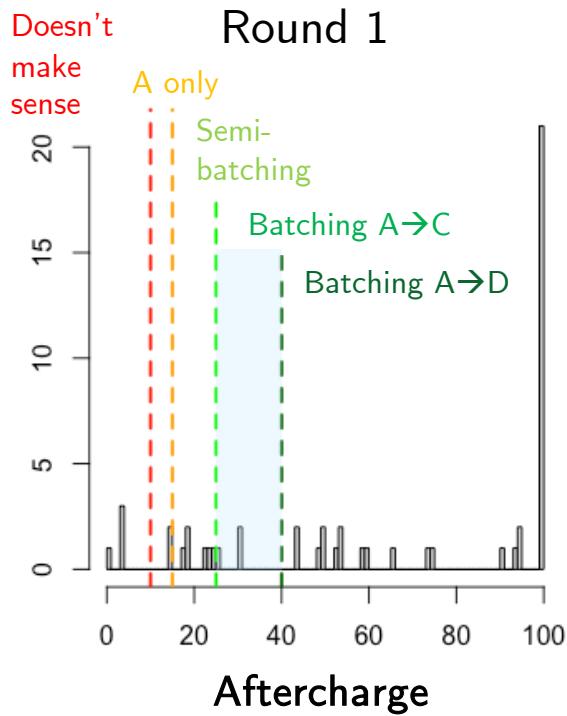
# Study 2A:

# Results

## Wide Range of Decisions



Round 1

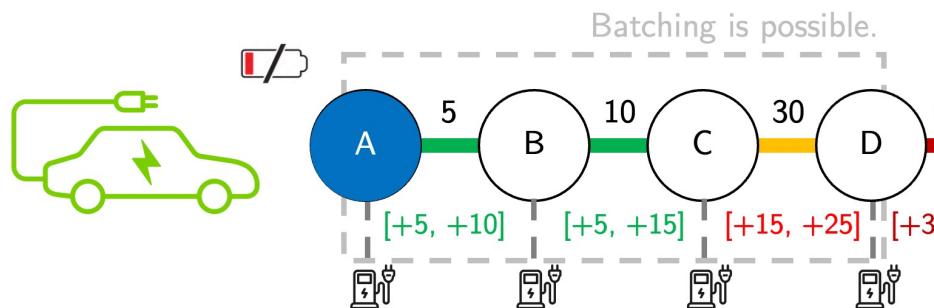


= amount of charge after re-charging

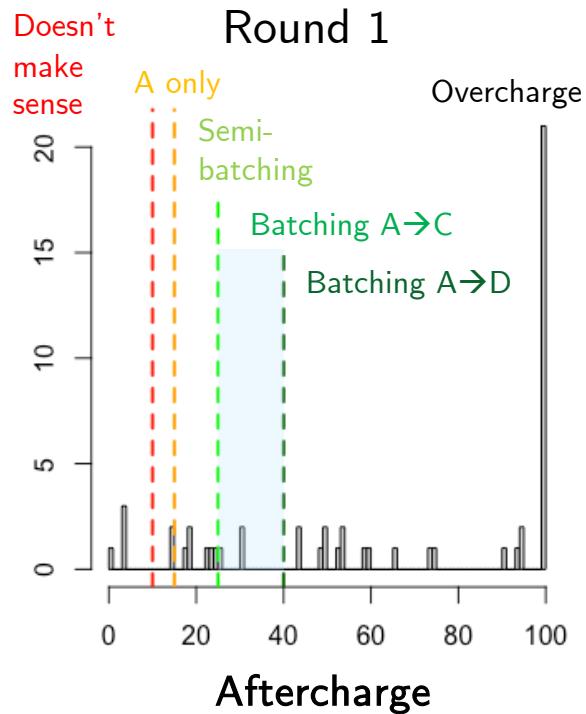
# Study 2A:

# Results

## Wide Range of Decisions



Round 1



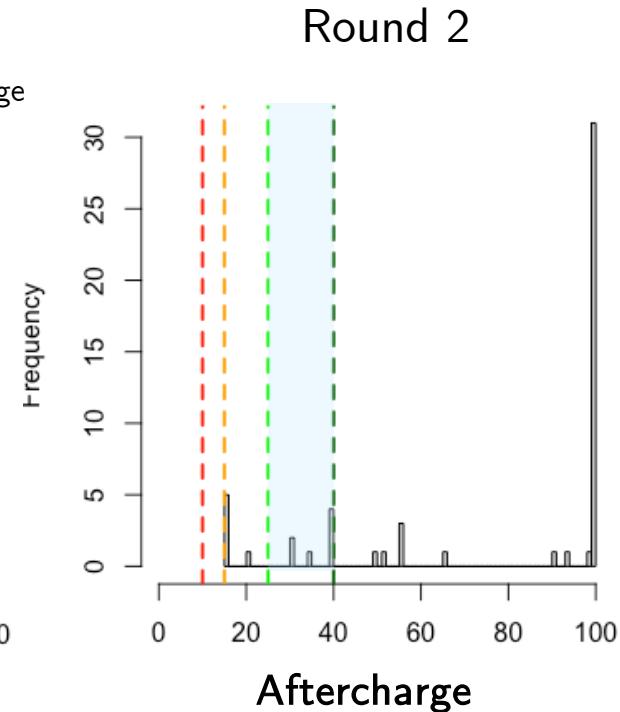
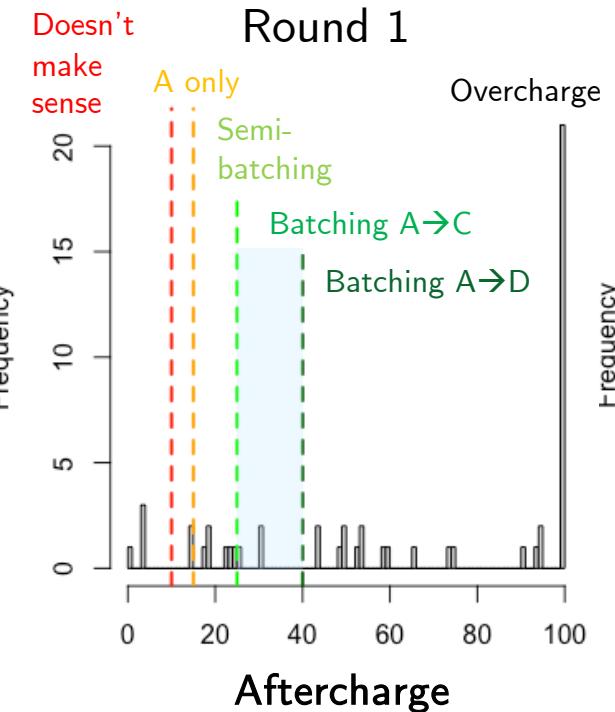
= amount of charge after re-charging

Exit A: Forced to charge  
Optimal: Batch A → C  
rather than just A→B  
or further batch A→D.  
("aftercharge" = 25-40)

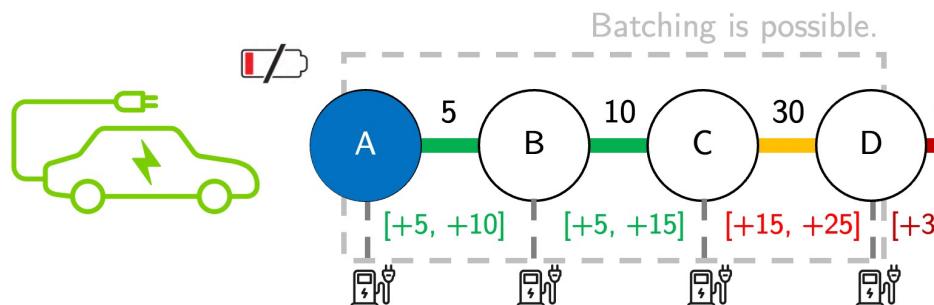
# Study 2A:

# Results

## People Learned w/o Tip



= amount of charge after re-charging

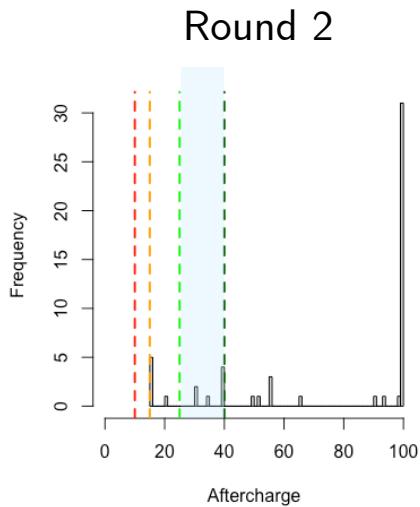


Exit A: Forced to charge  
Optimal: Batch A → C  
rather than just A→B  
or further batch A→D.  
("aftercharge" = 25-40)

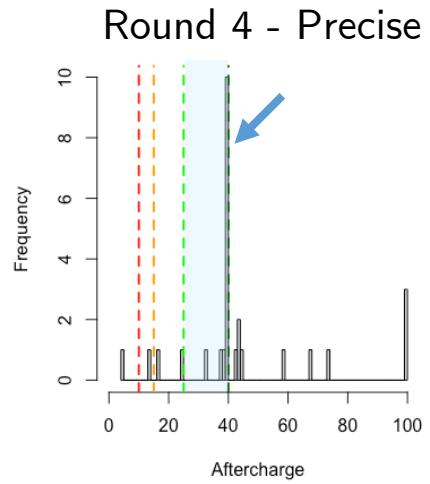
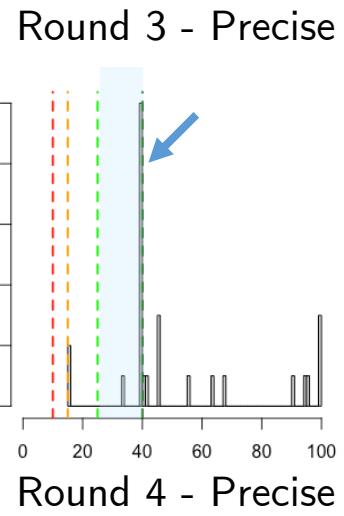
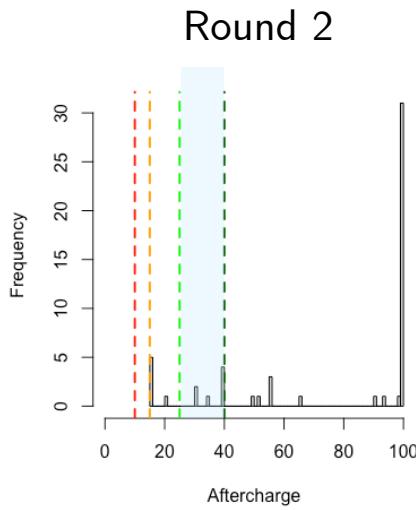
In one round, everyone learned to do at least semi-batching for this exit.

# Results

## What Happened with Tip?

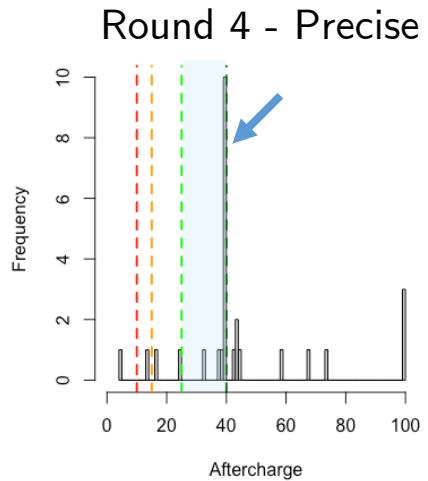
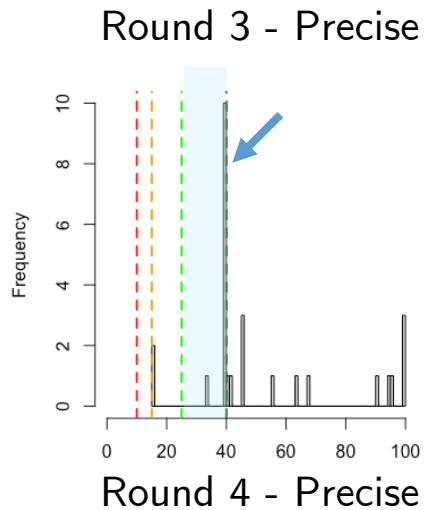
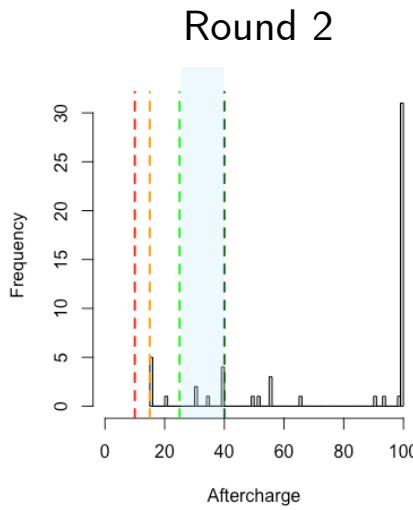


# Results



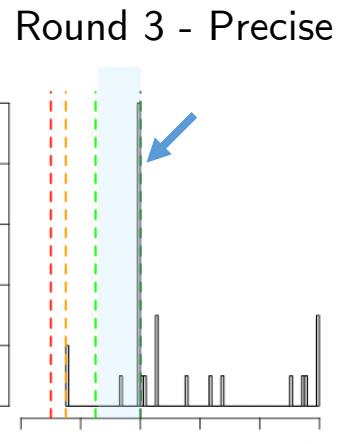
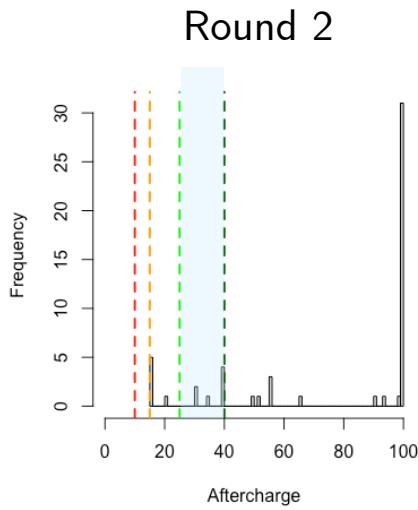
# Results

## Precise Tip Worked (as Expected)

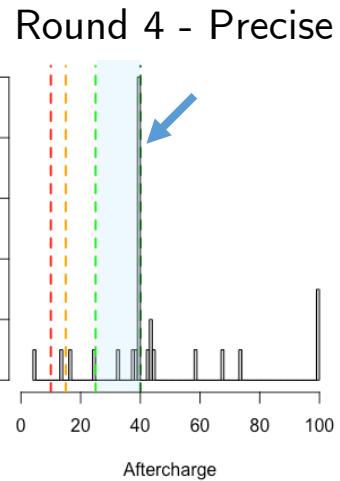


The Precise tip  
successfully nudged  
participants to  
batch A → C

# Results



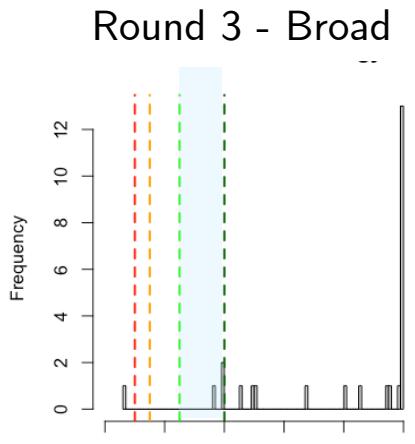
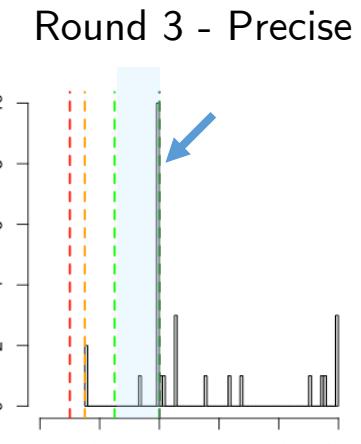
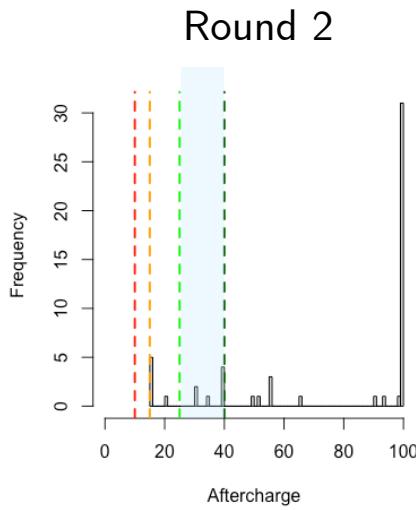
**Broad tip**  
“You should charge  
enough for this segment  
and the next one”



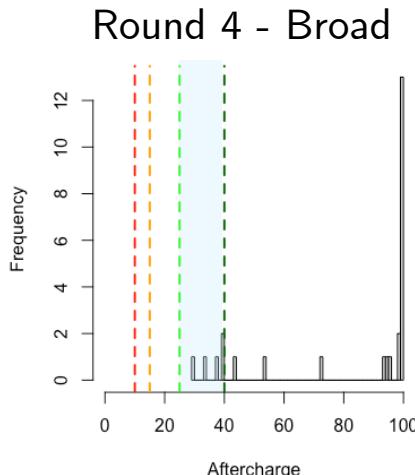
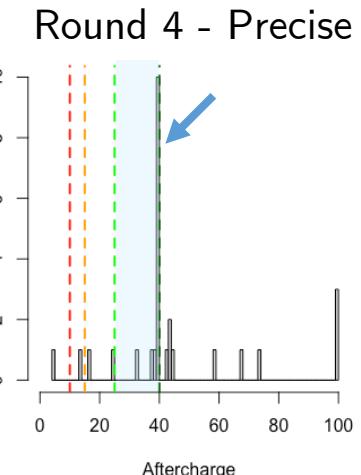
The Precise tip  
successfully nudged  
participants to  
batch A → C

# Results

## Broad Tip Kind of Failed



**Broad tip**  
“You should charge  
enough for this segment  
and the next one”

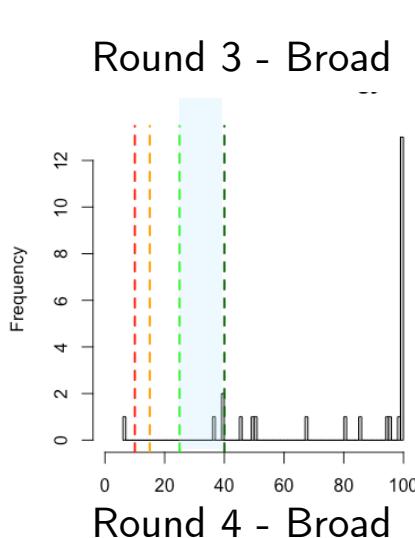
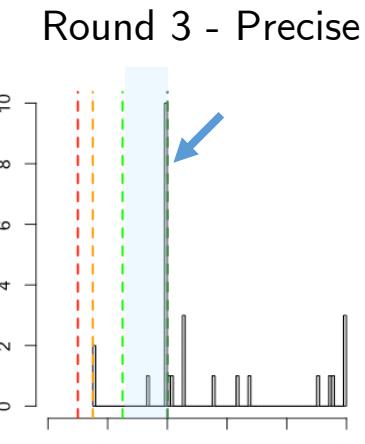
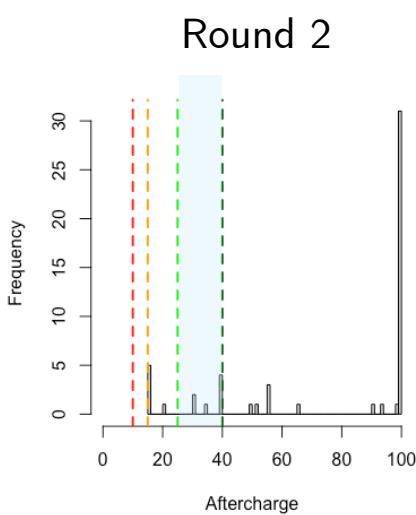


The **Precise** tip  
successfully nudged  
participants to  
batch A → C

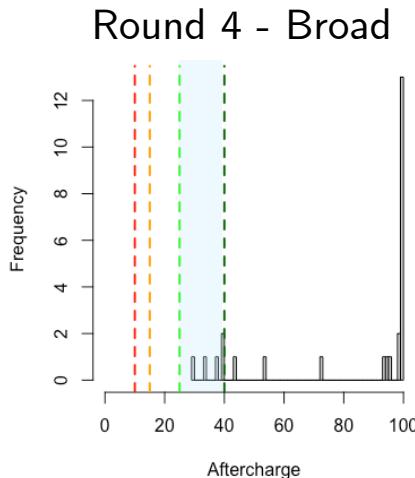
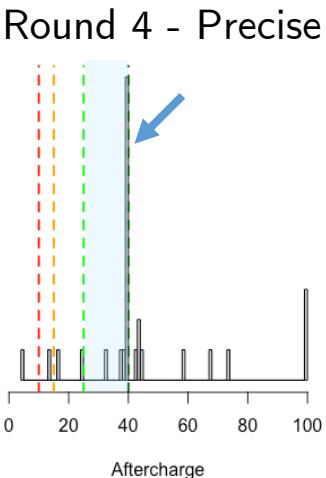
The **Broad** tip had  
low compliance even  
though it was clear  
what the participant  
should do

# Results

## Broad Tip Kind of Failed



**Broad tip**  
“You should charge  
enough for this segment  
and the next one”



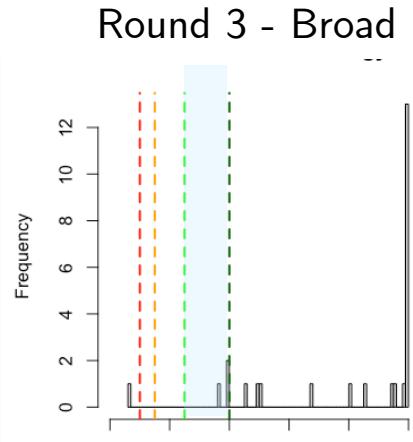
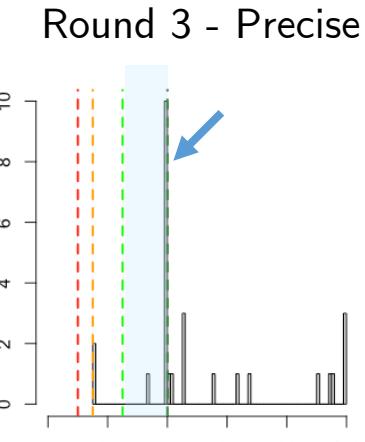
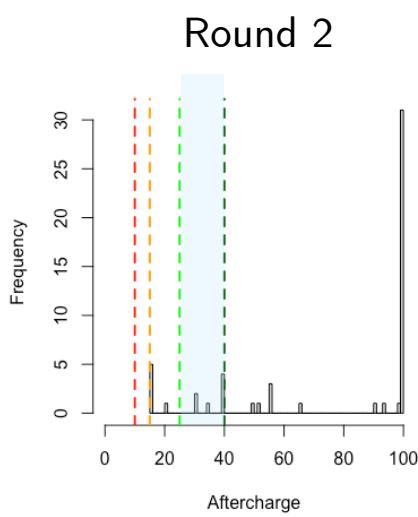
The **Precise** tip  
successfully nudged  
participants to  
batch A → C

The **Broad** tip had  
low compliance even  
though it was clear  
what the participant  
should do

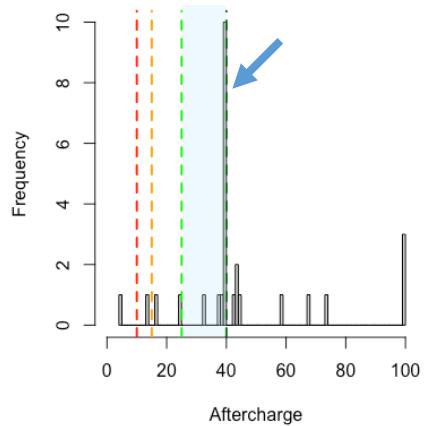
Q: Could it be due to  
the tip being perceived  
as counterintuitive?

# Results

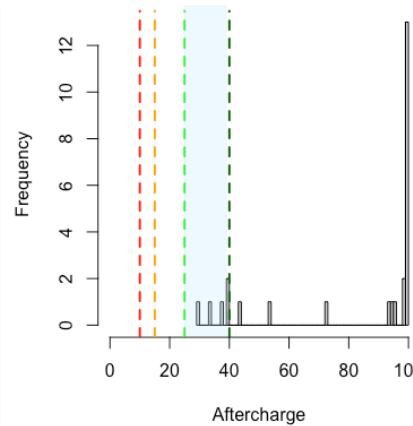
## Broad Tip Kind of Failed



Round 4 - Precise



Round 4 - Broad



The Precise tip successfully nudged participants to batch A → C

**Broad tip**  
“You should charge enough for this segment and the next one”

The Broad tip had low compliance even though it was clear what the participant should do

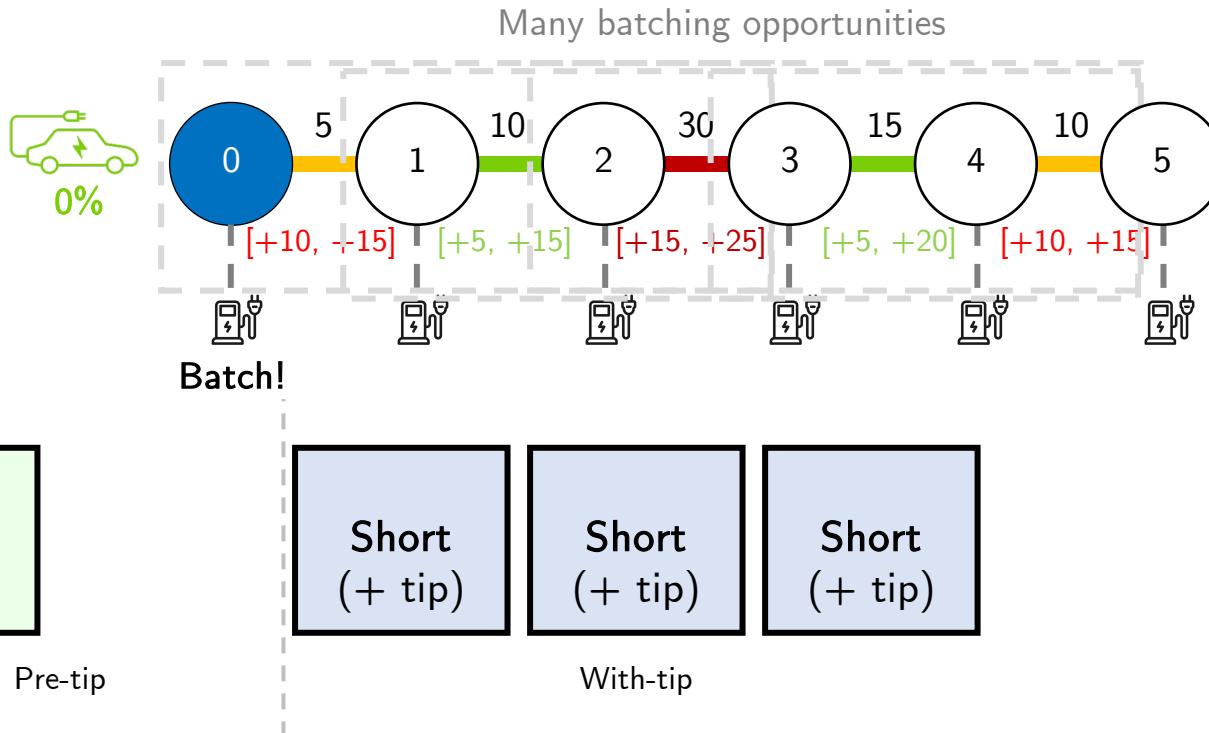
Q: Could it be due to the tip being perceived as counterintuitive?

Q: Did Precise tip help people learn?

**But Did We Help Them  
Improve Their Decision-Making?**

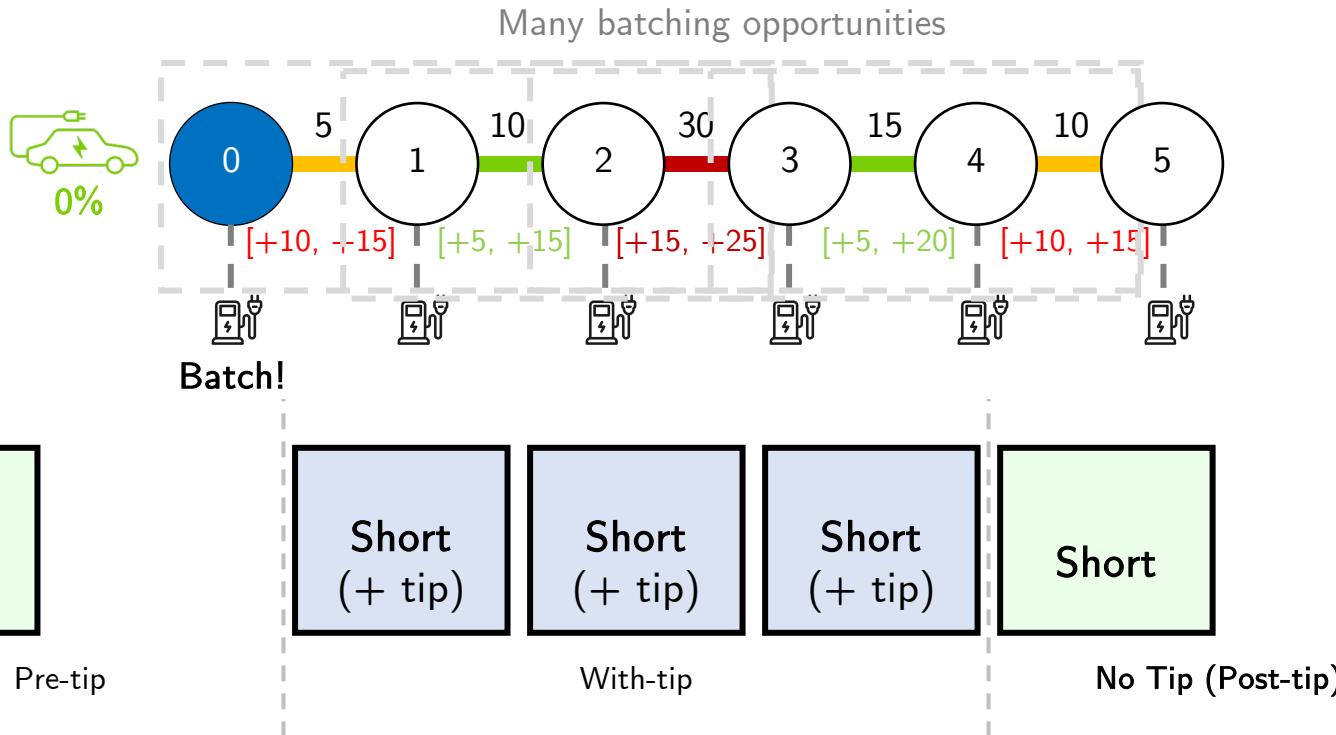
# Study 2B:

# Design Short Map

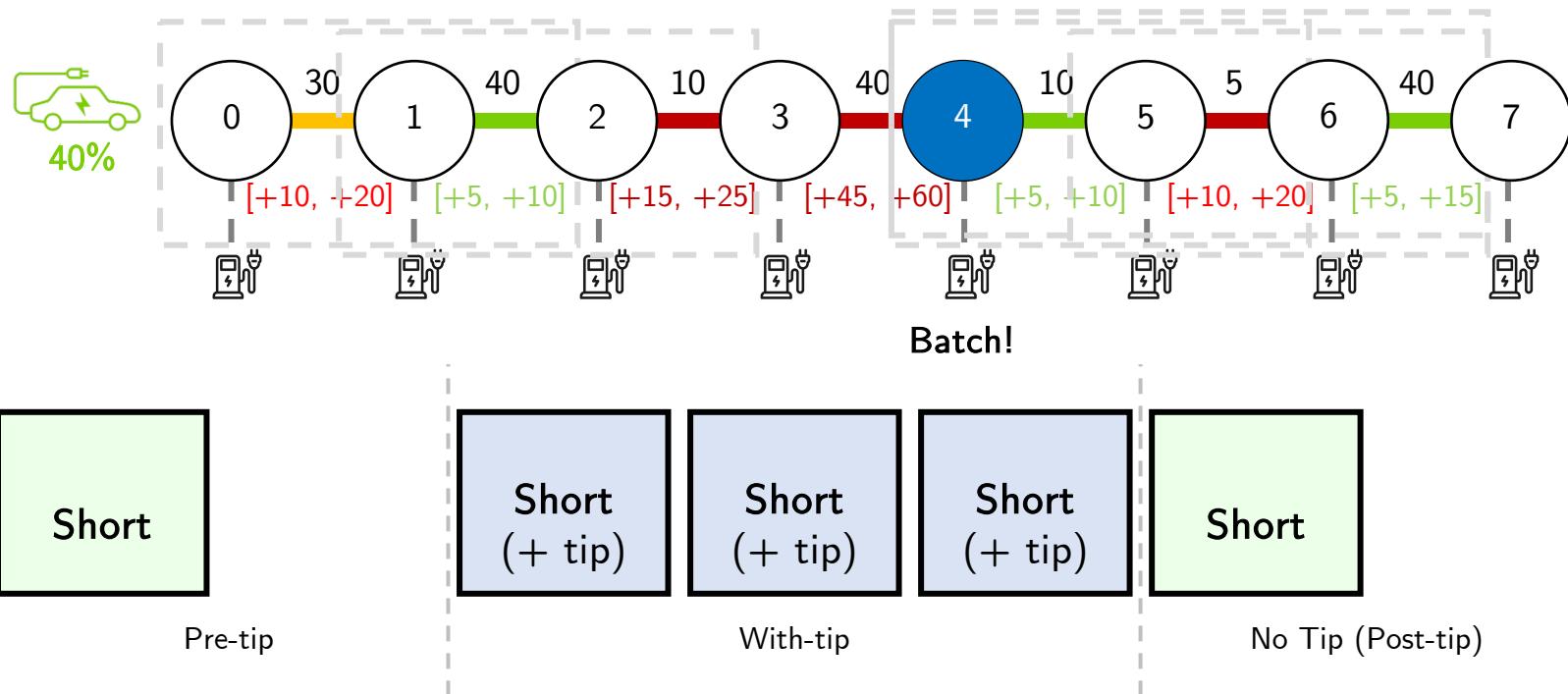


# Study 2B:

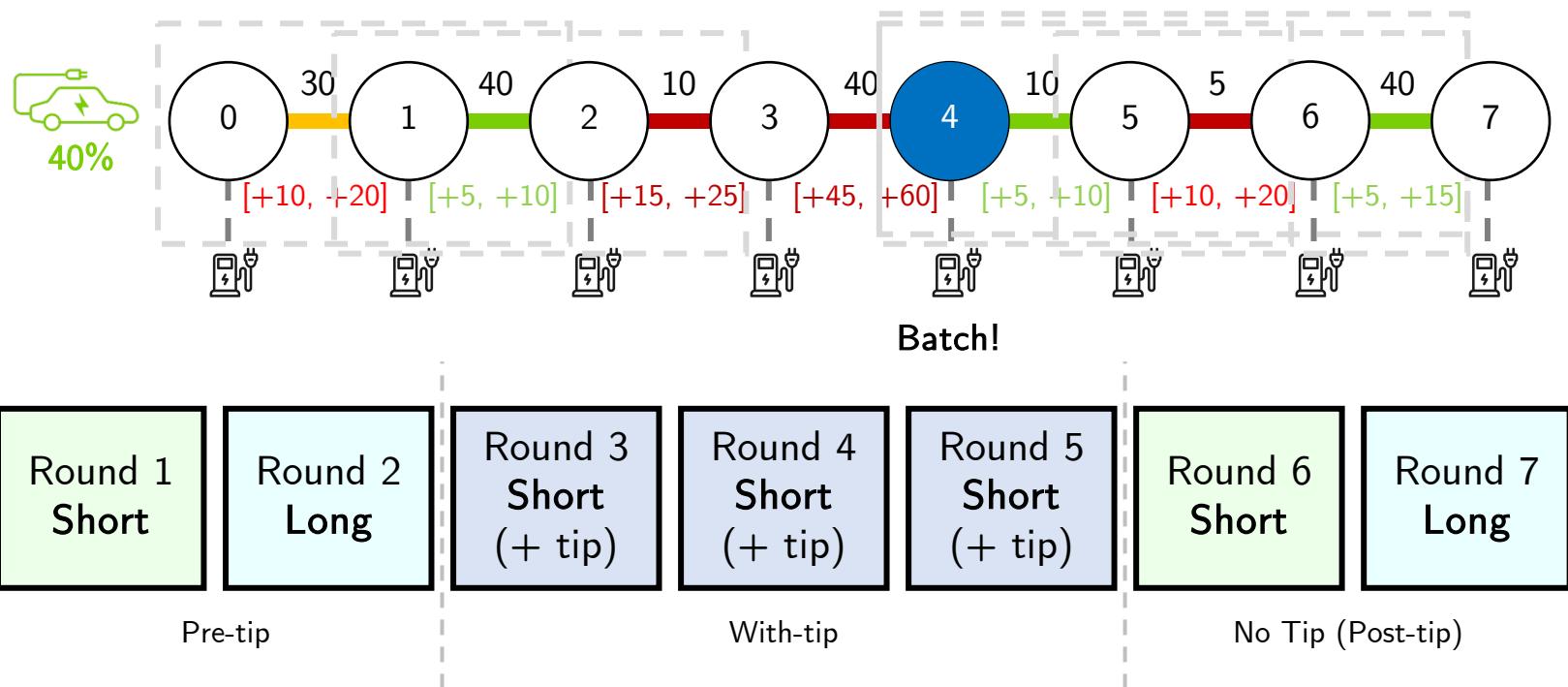
# Design Short Map



# Study 2B: Design + Long Map (New Environment)



# Study 2B: Design + Long Map (New Environment)



# Study 2B:

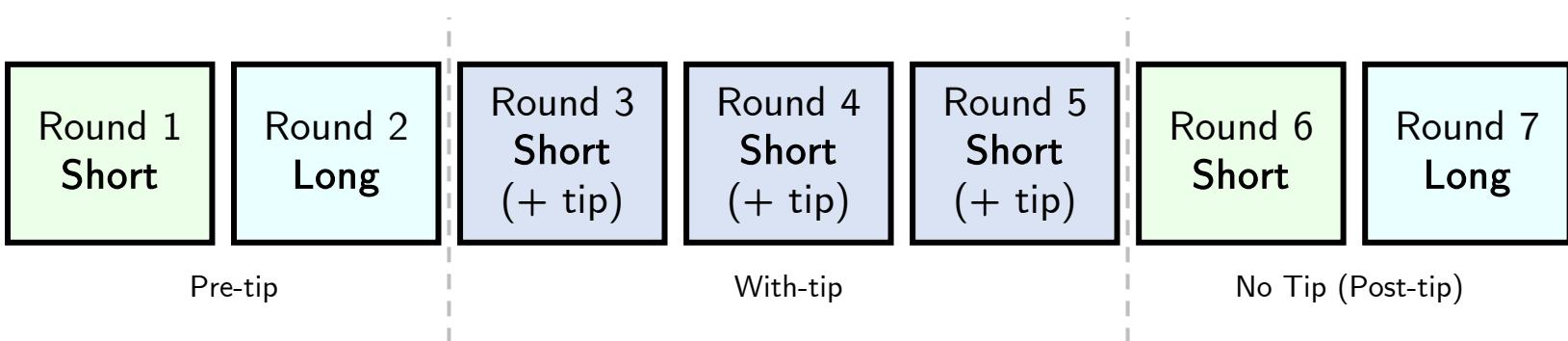
## Design

2

tip precision

x

2

centered / skewed  
realized traffic

# Study 2B:

## Design

2

x

2

tip precision

centered / skewed  
realized traffic

Precise Tip:

You should exit  
and charge X%

“Split”

Broad Tip:

You should charge enough  
for this segment

“Batch”

You should charge enough  
for this segment + next oneRound 1  
ShortRound 2  
LongRound 3  
Short  
(+ tip)Round 4  
Short  
(+ tip)Round 5  
Short  
(+ tip)Round 6  
ShortRound 7  
Long

Pre-tip

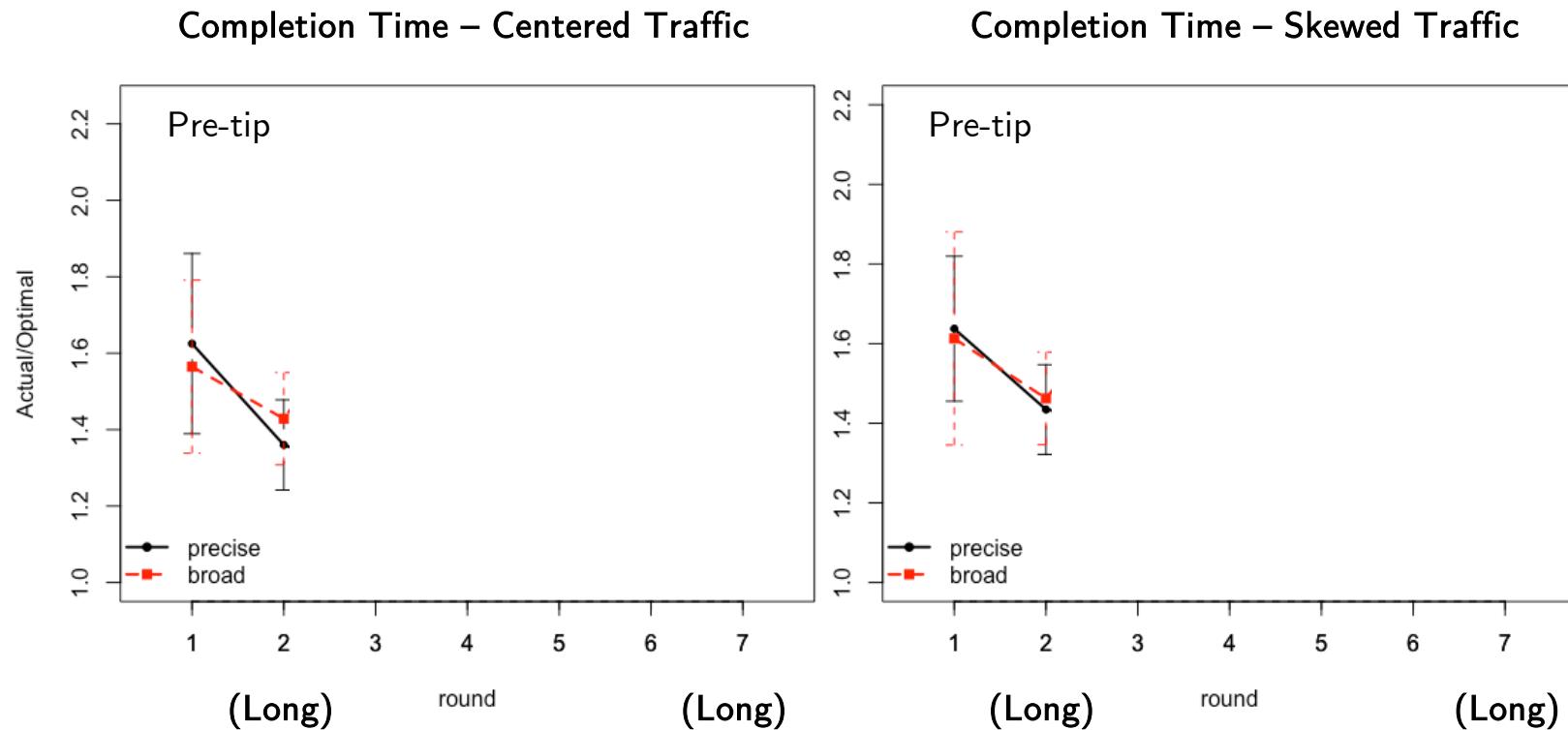
With-tip

No Tip (Post-tip)

# Study 2B:

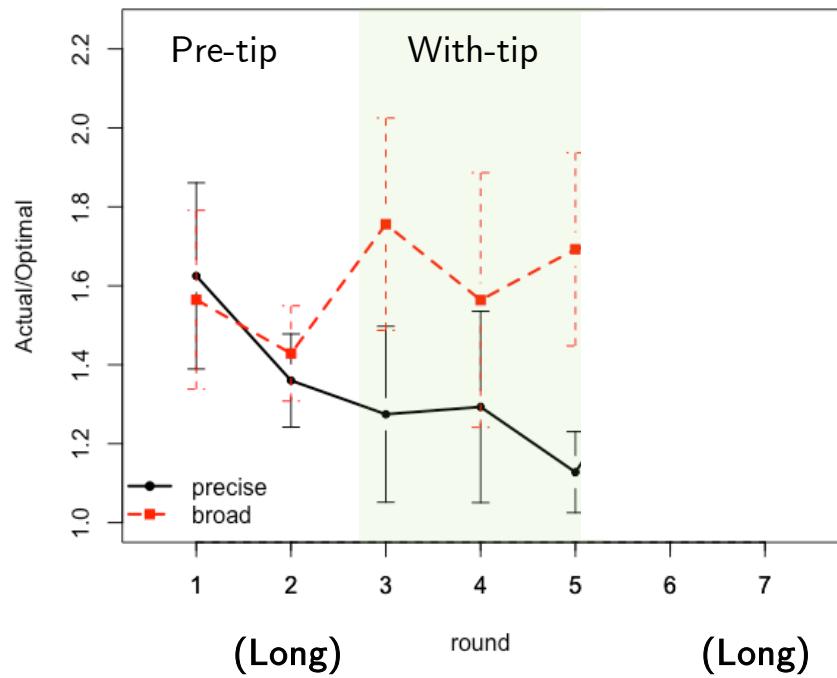
# Results

## Performance Across Rounds

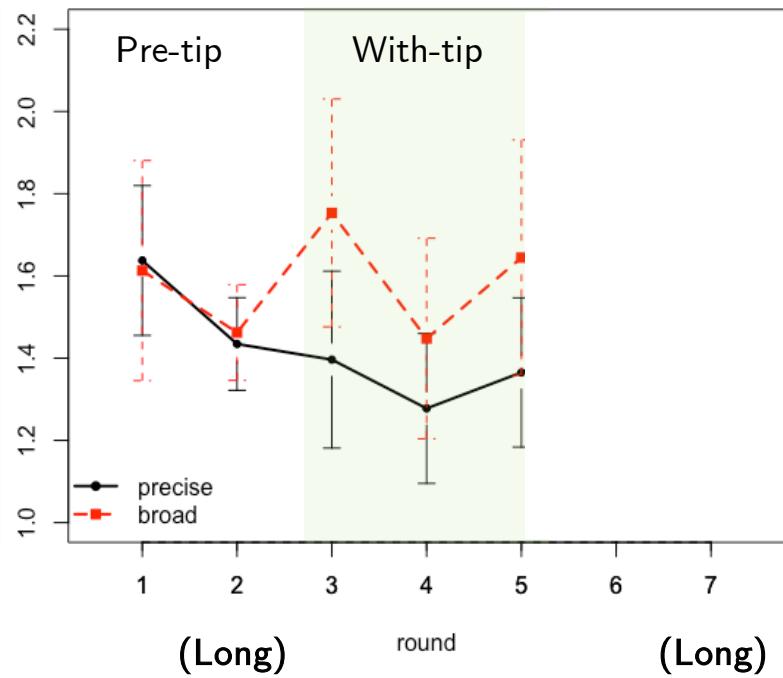


# Study 2B: Results

Completion Time – Centered Traffic



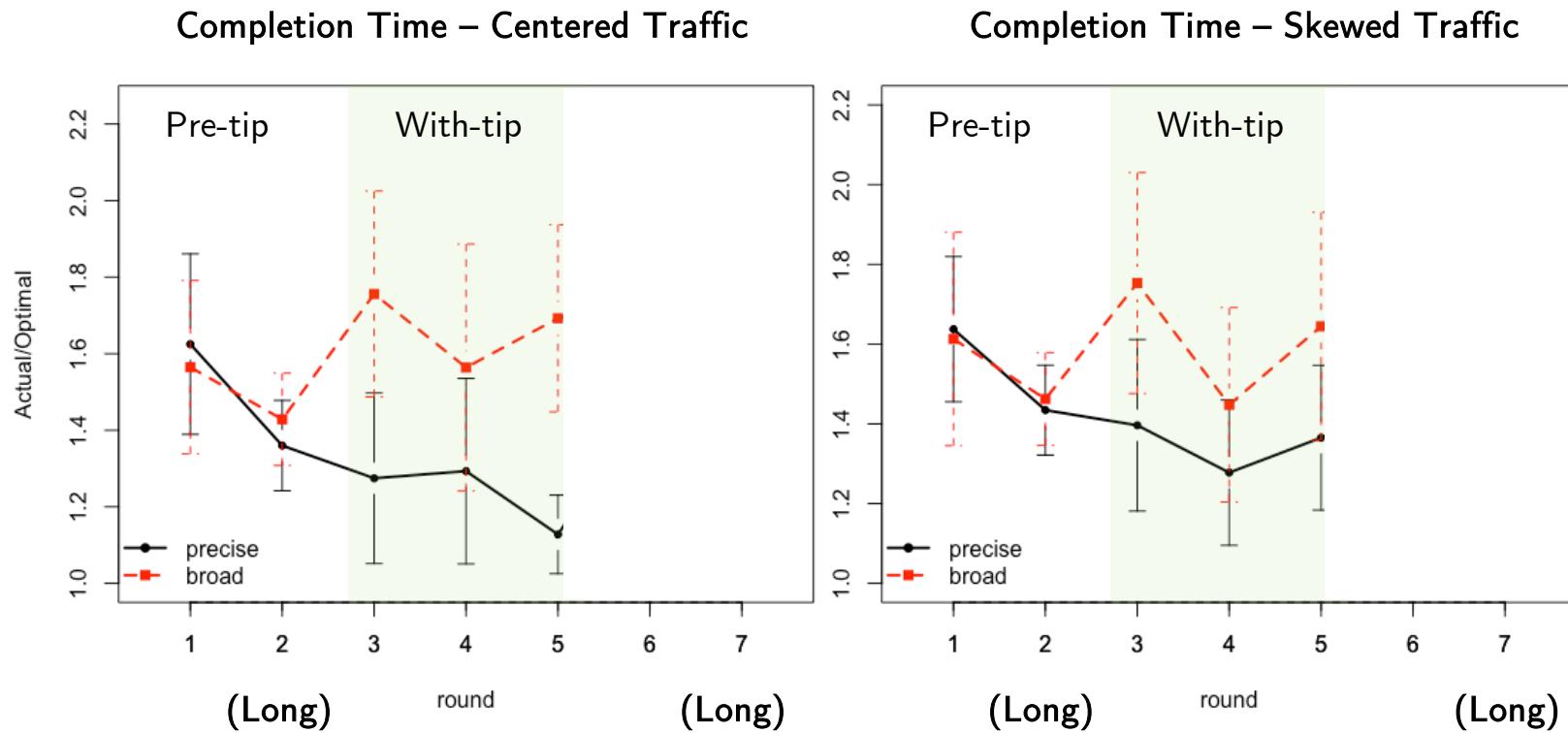
Completion Time – Skewed Traffic



# Study 2B:

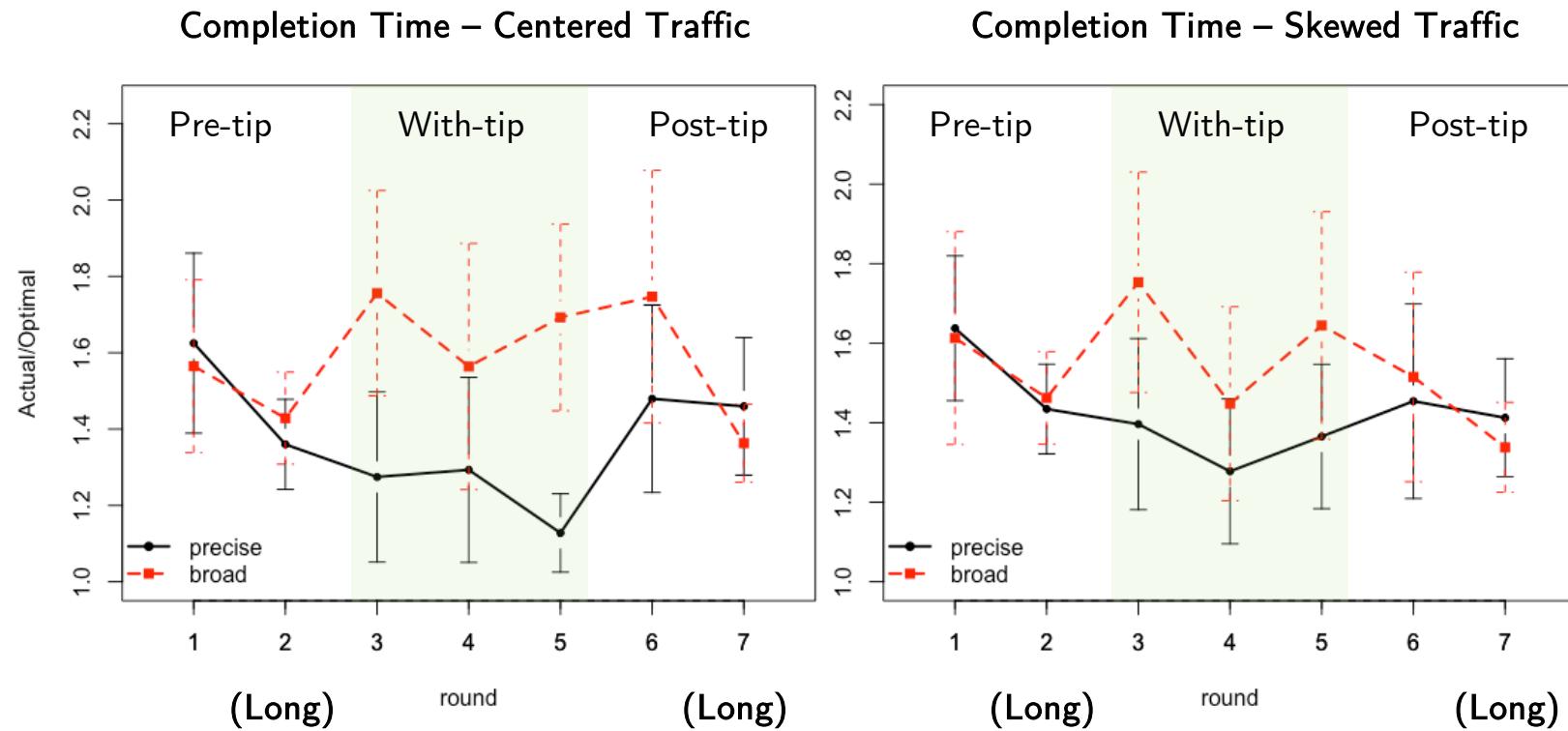
# Results

## Precise Tip Worked, Again



# Study 2B:

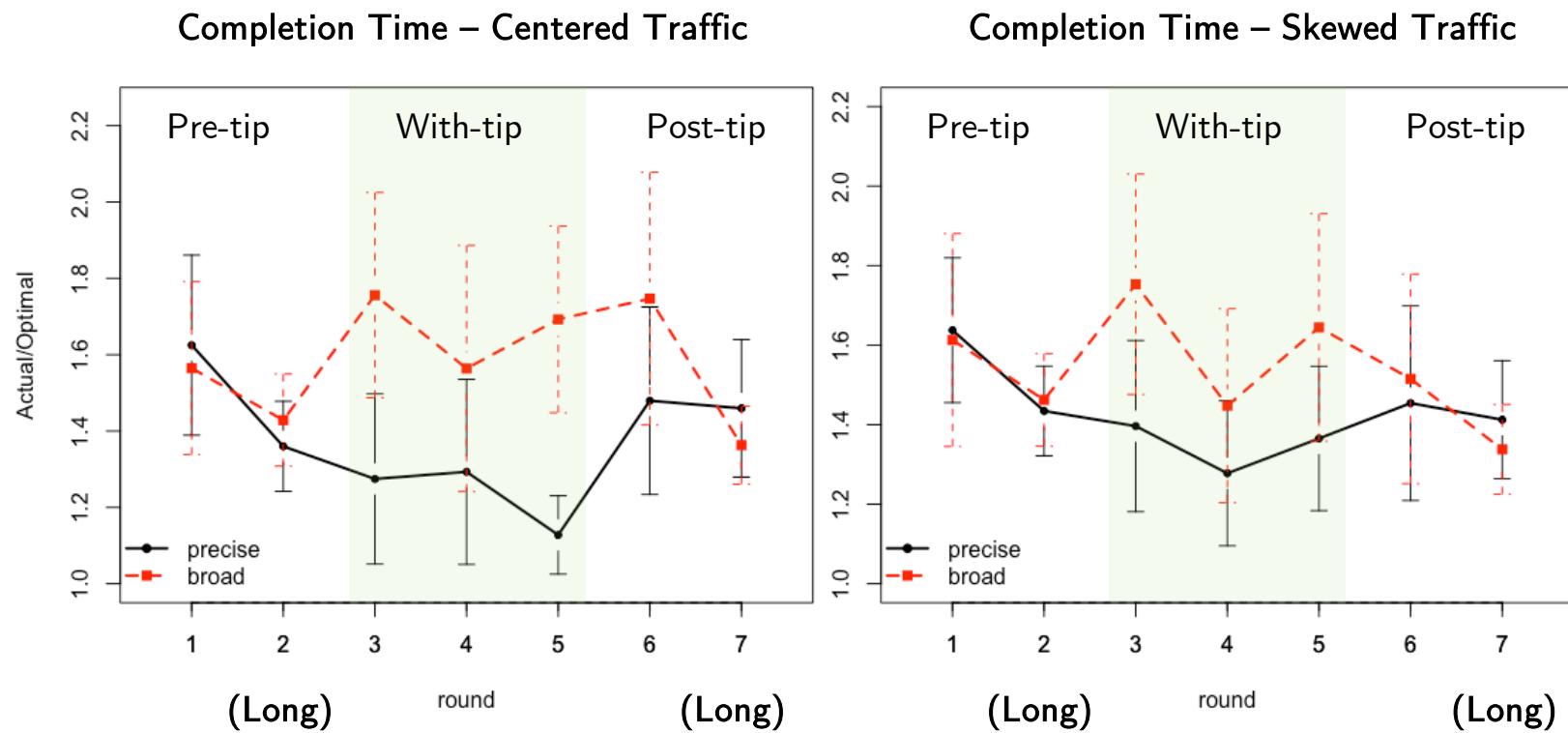
# Results



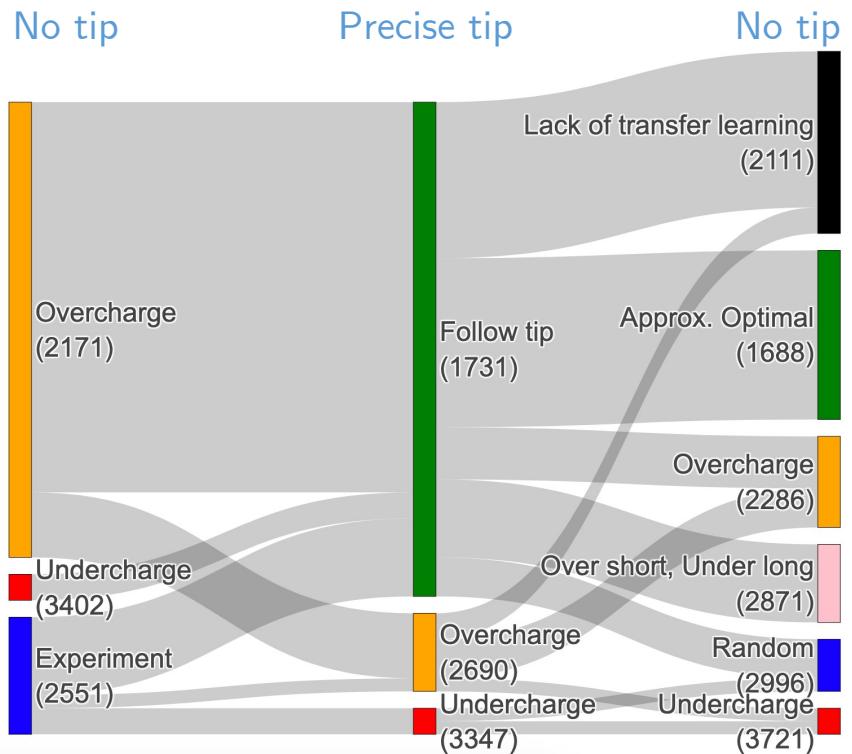
# Study 2B:

# Results

Broad Tip Seemed to Help with New Environment



# Study 2B: Results Long-Term Learning from Tip



34% stay with  
optimal strategy afterwards

# Study 2B:

# Results

## Long-Term Learning from Tip

No tip

Precise tip

No tip

Overcharge  
(2171)

Undercharge  
(3402)  
Experiment  
(2551)

Follow tip  
(1731)

Overcharge  
(2690)  
Undercharge  
(3347)

Approx. Optimal  
(1688)

Overcharge  
(2286)

Over short, Under long  
(2871)

Random  
(2996)  
Undercharge  
(3721)

Lack of transfer learning  
(2111)

No tip

Broad tip

No tip

Overcharge  
(2118)

Experiment  
(2525)

Undercharge  
(3723)

Follow tip  
(2054)

Overcharge  
(2667)

Undercharge  
(3657)

Approx. Optimal  
(1761)

Overcharge  
(2313)

Random  
(2650)

Lack of transfer learning  
(2164)

Over short, Under long  
(2871)

Undercharge  
(3222)

34% stay with  
optimal strategy afterwards

56% stay with  
optimal strategy afterwards

# Summary

ML framework to leverage behavioral trace data to infer simple tips that help humans



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

Performance/compliance tradeoff

with Hamsa Bastani & Osbert Bastani  
*Minor Revision @ Management Science*



(Available at: [bit.ly/tipspaper](https://bit.ly/tipspaper))

Pre-tip behavior classification helps predict With- and Post-tip behavior

Precise tips are useful in the short term, but broad tips can help with long-term learning/new environment

with Philippe Blaettchen, *Work in Progress*



# Summary

ML framework to leverage behavioral trace data to infer simple tips that help humans



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

Performance/compliance tradeoff

with Hamsa Bastani & Osbert Bastani  
*Minor Revision @ Management Science*



(Available at: [bit.ly/tipspaper](https://bit.ly/tipspaper))

Pre-tip behavior classification helps predict With- and Post-tip behavior

Precise tips are useful in the short term, but broad tips can help with long-term learning/new environment

with Philippe Blaettchen, *Work in Progress*



**Thank you! Feedback (+ tips) very welcome!**

# **Appendix**

# Study 1: Results Learning Beyond Tips

## Structure of Optimal Policy

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

Algorithm      Baseline

The diagram illustrates the structure of an optimal policy for two roles: Sous-Chef and Server. The policy is organized into three columns: Chop, Cook, and Plate. The Sous-Chef row shows values 3, 2, and 2 respectively. The Server row shows values 1, 2, and 2 respectively. The word 'times' is placed to the right of each row, indicating the frequency of each action. Below the table, two arrows point upwards from the words 'Algorithm' and 'Baseline' to the sous-chef and server rows respectively.

# Study 1:

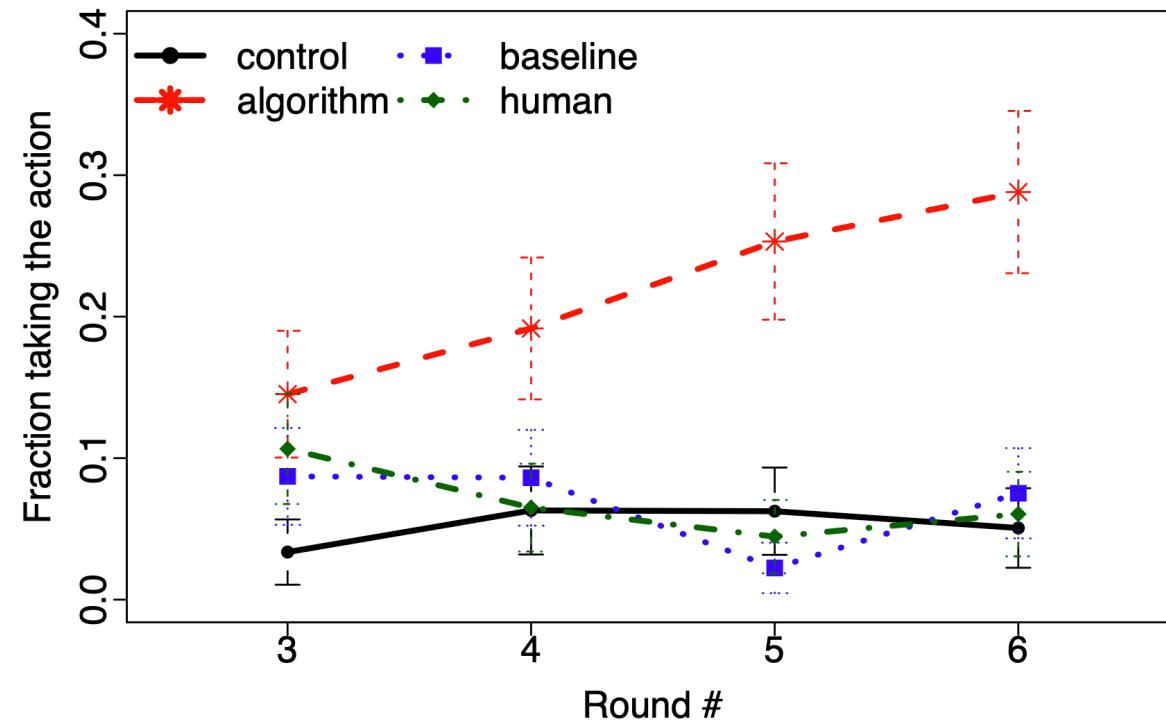
# Results Learning Beyond Tips

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

Sous-Chef  
chops 3 times



Part of optimal  
policy but not stated  
in any of the tips

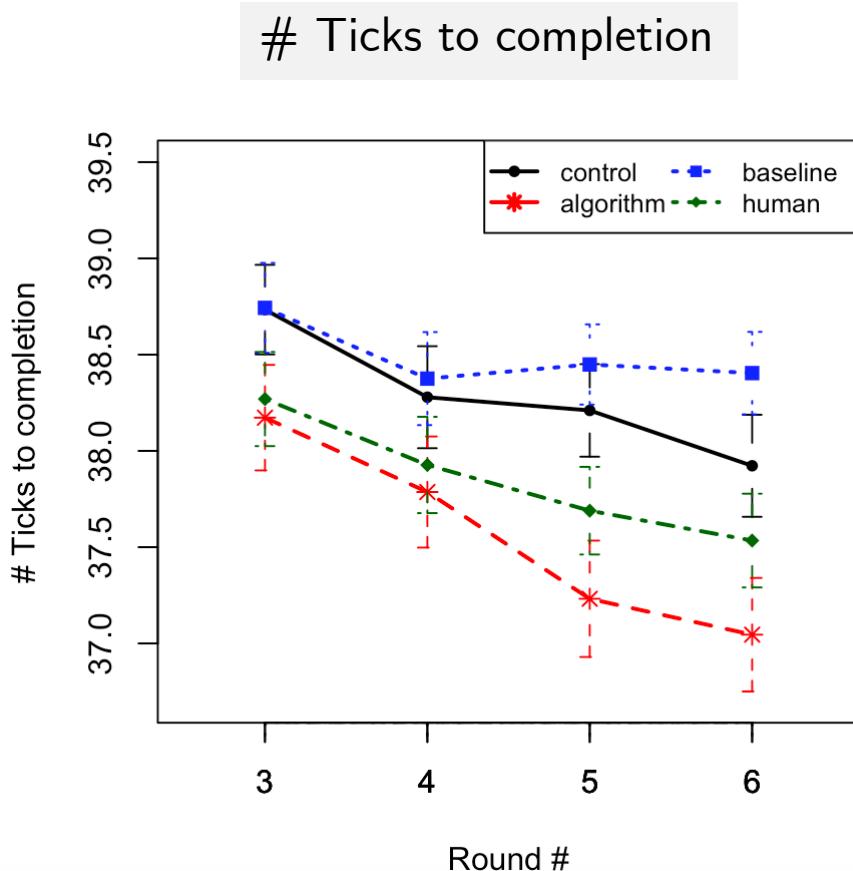


Amazon Mechanical Turk, N = 1,011  
mean age 34.9, 60% female

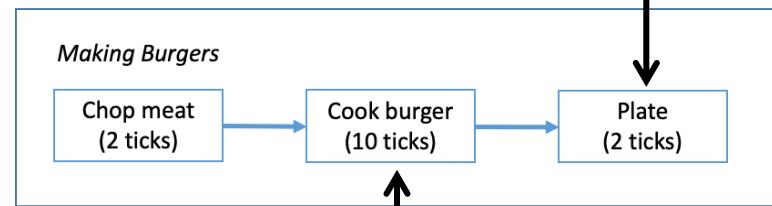
# Study 1:

# Results

## Good Tip = Consequential



Baseline Sous-Chef should plate twice



Algorithm Server should cook twice

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

# Study 1: 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.

# Study 1:

# 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%
≥ 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	Baseline			Human
		“Chef shouldn’t plate”	“Chef chops once”	“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	Baseline			Human
		“Server cooks twice”	“Sous-chef plates twice”	“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%	

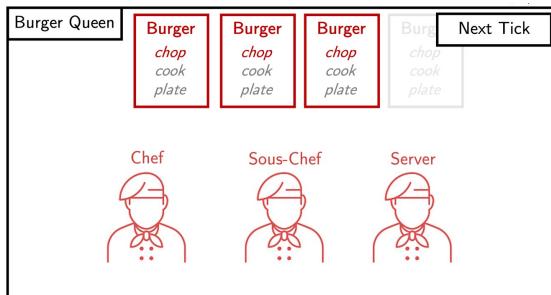
Table 8: Participants’ coded feedback on the provided tips (disrupted configuration).

# User Study Design

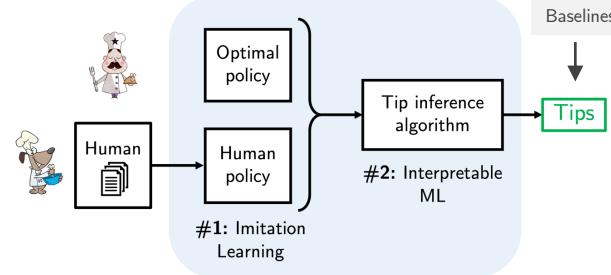
N = 1400

## Phase I

N = 200



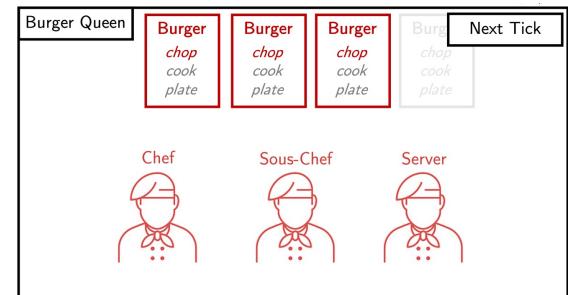
Gather trace data



Tip inference

## Phase II

Tip: [randomly assigned tip here]



Tip evaluation

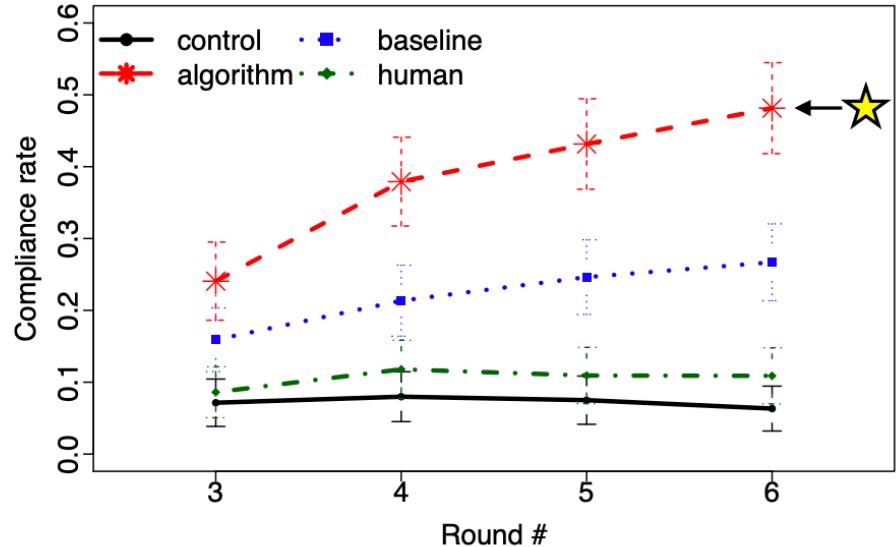
## Environment

Normal

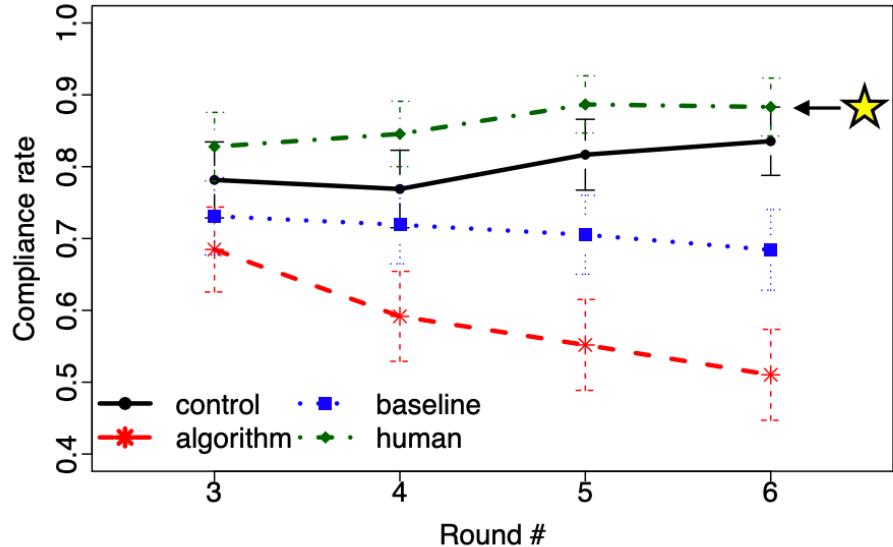


Disrupted

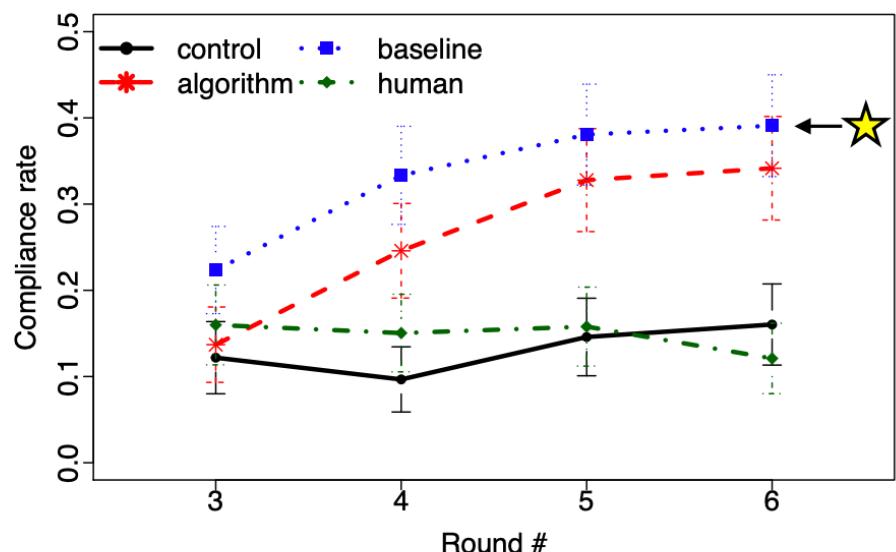




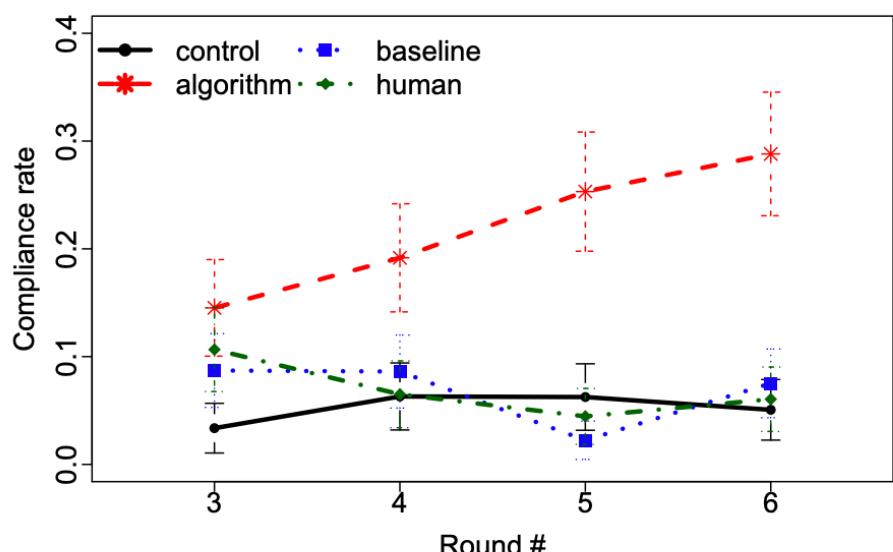
(a) Algorithm Tip: “Server cooks twice”



(b) Human Tip: “Server cooks once”

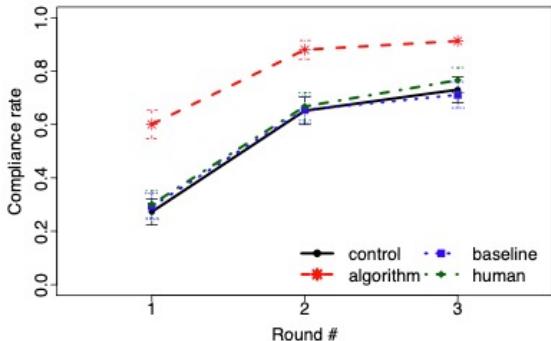


(c) Baseline Tip: “Sous-chef plates twice”

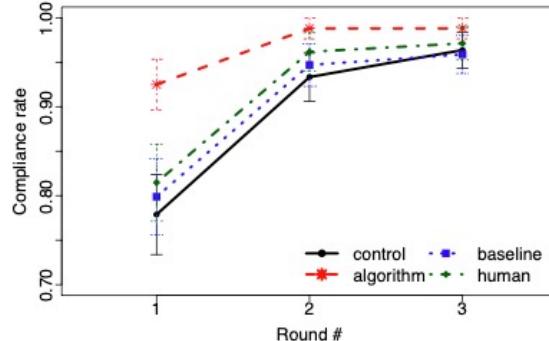


(d) Unshown Tip: “Server chops once”

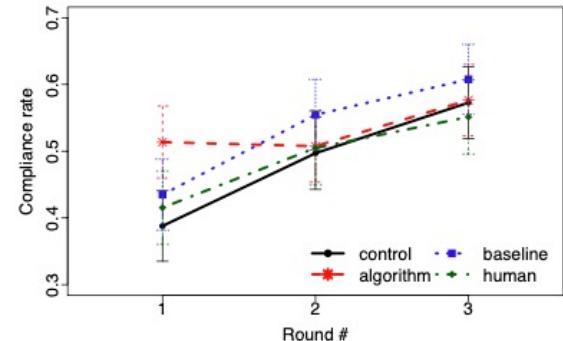
# Learning Beyond Tips



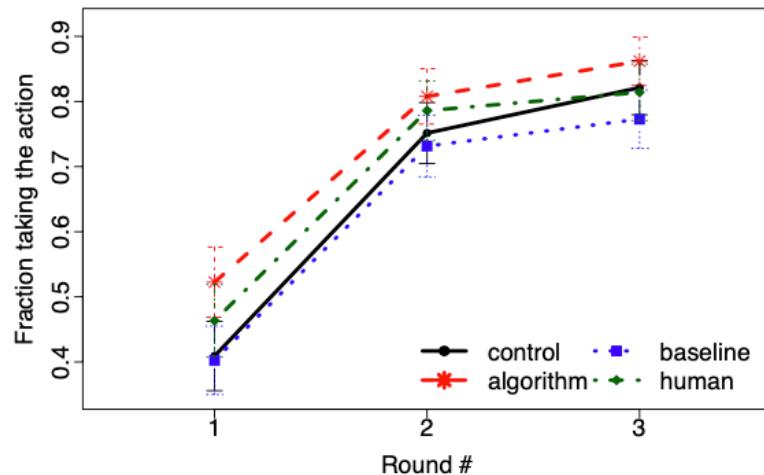
(a) Algorithm: “Chef shouldn’t plate”



(b) Human: “Leave some idle”



(c) Baseline: “Chef chops once”



(a) Fully-staffed: “Server shouldn’t cook”

## Study 2:

**What If Decisions Are More Granular?**

**Also, What If Things Are Uncertain?**



Sequential

To arrive fast, min travel + charging times

Complex

Non-linear charging, fixed vs variable cost

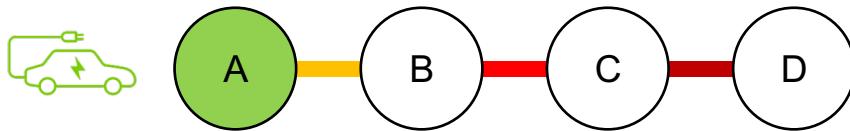
Uncertain

Uncertain traffic, risk-preferences

Continuous

Anything from 0-100%,  
all-or-nothing/charge to full suboptimal

# Classifying Decisions



“Reasonable” levels:  
pos. probability of  
reaching next stop

“Batch 1”



“Reasonable” levels:  
pos. probability of  
reaching two stops

“Batch 2”



“Reasonable” levels:  
pos. probability of  
reaching three stops

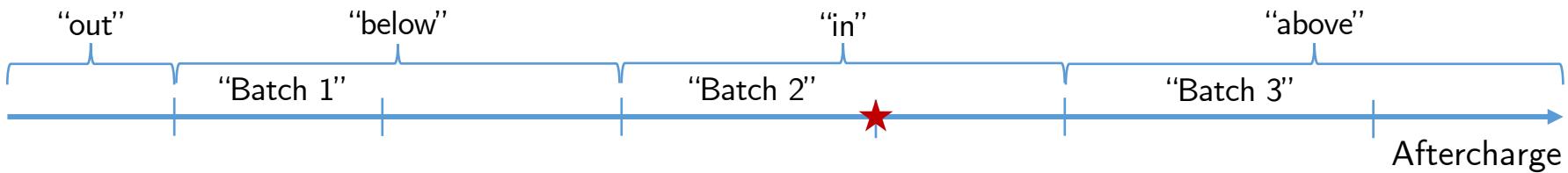
“Batch 3”



Aftercharge

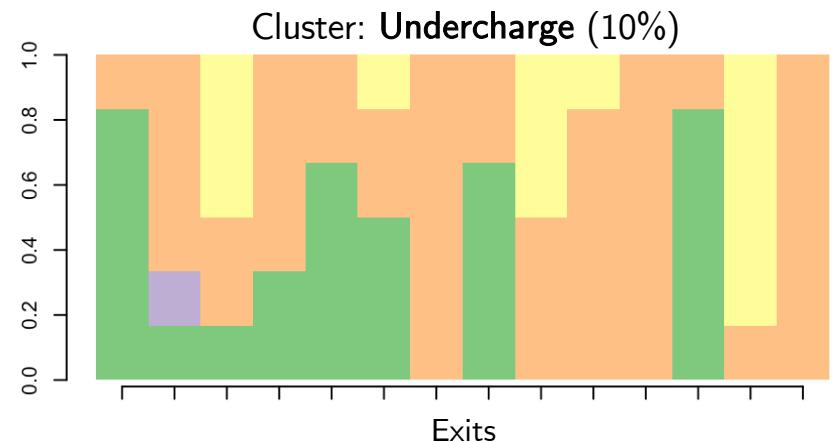
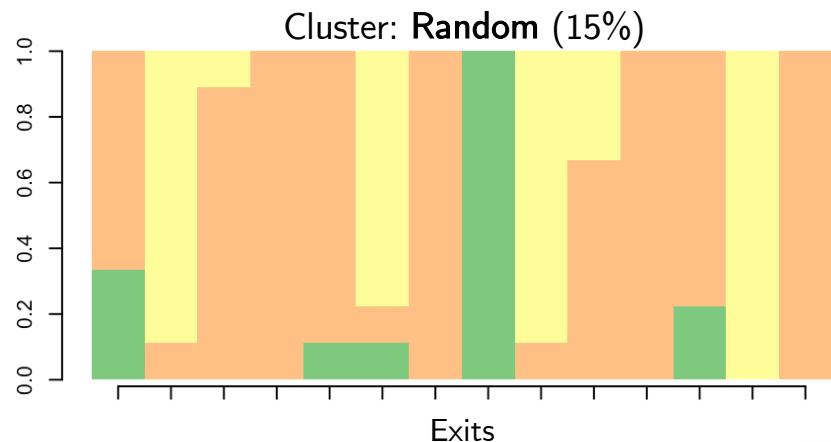
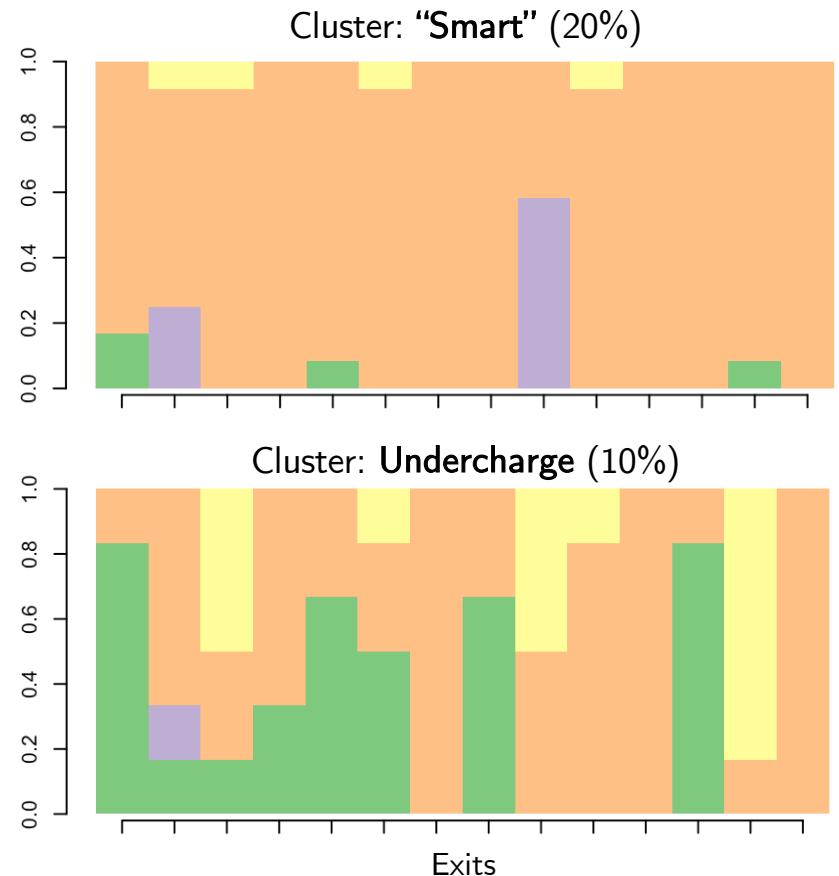
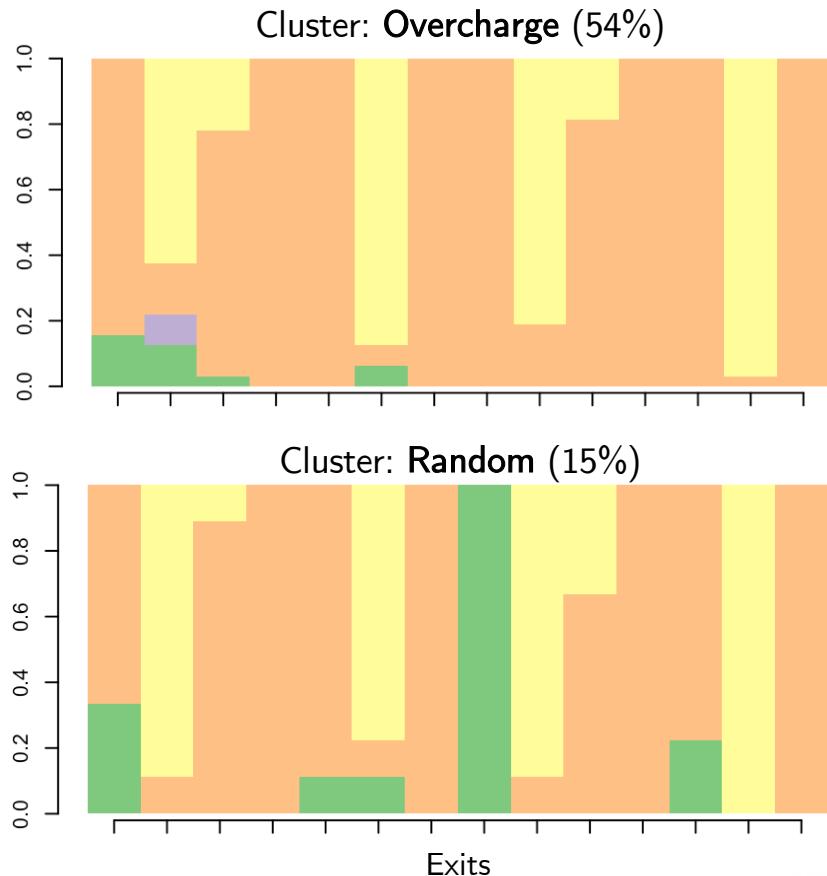
Too little charge for next “batch”,  
but too much for current...  
...but **most charging here**  
(risk aversion, limited look-ahead)  
→ unnecessarily conservative

Hierarchical model of decision-making:  
1. Choice of broad interval  
2. Choice of charge within interval



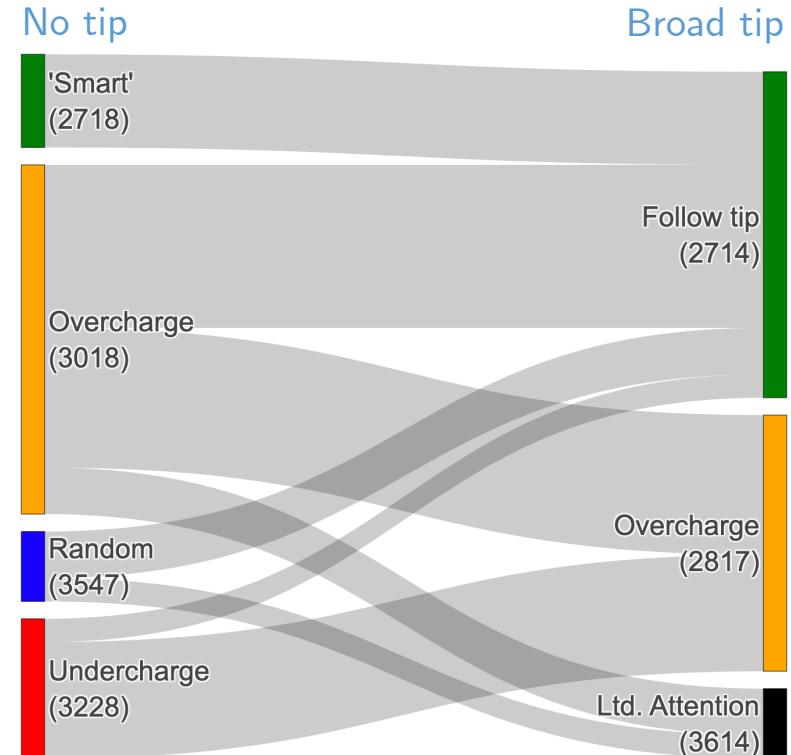
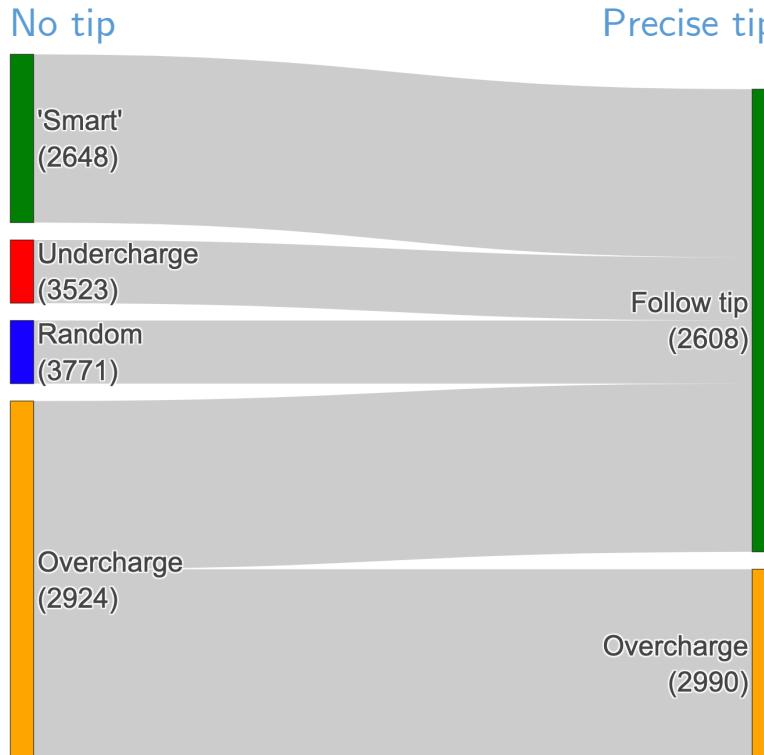
→ Consistent choice of intervals as an indicator of strategic behavior

# Study 2A: Results Clustering Pre-Tip Behaviors



- Out
- Below
- In = optimal
- Above

# Study 2A: Results Predicting Response to Tip



- Advice followed exactly or more conservatively
- Precise advice followed more
- “Smart” behavior is good indicator of “converting” imprecise tips