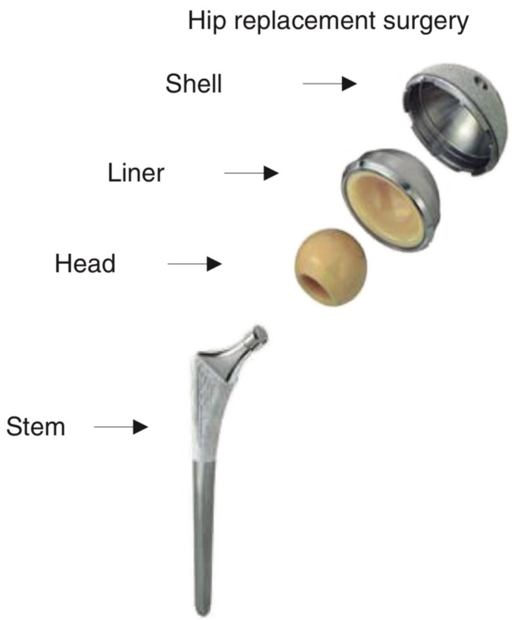
# Give Humans Advice, and You Feed Them for a Day

## Balancing Efficiency & Learning in Algorithmic Recommendations

Park Sinchaisri
(UC Berkeley)

with Philippe Blaettchen (SMU)

# Learning is Hard & Costly




New device

= **+32.4%**

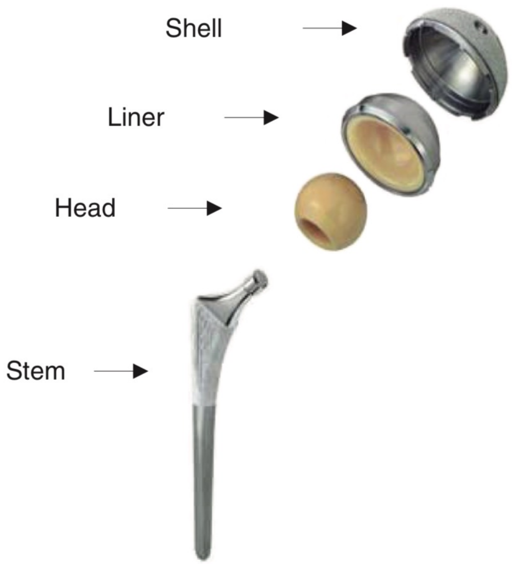
surgery duration

- Ramdas et al. 2018

Also – Tucker et al 2002, Ibanez et al 2017, Gurvich et al 2019, Bloom et al 2020, Bavafa & Jonasson 2021, ...

# Learning is Hard & Costly

Hip replacement surgery




New device

$= +32.4\%$

surgery duration

- Ramdas et al. 2018

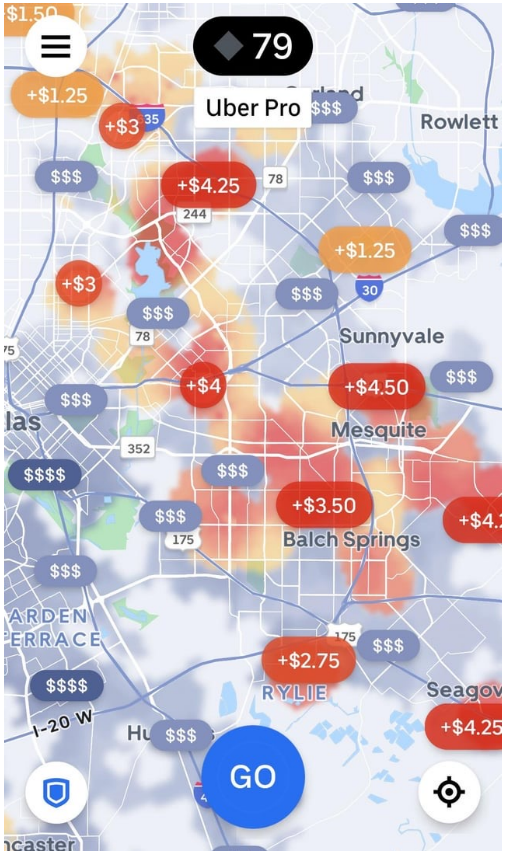
r/uberdrivers · Posted by u/kanyda 8 years ago

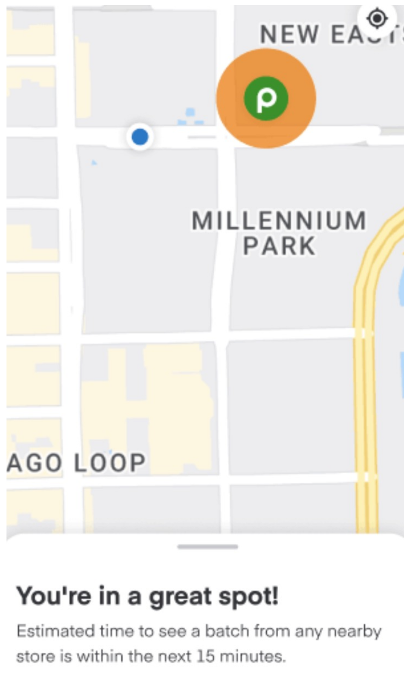
**First day report**

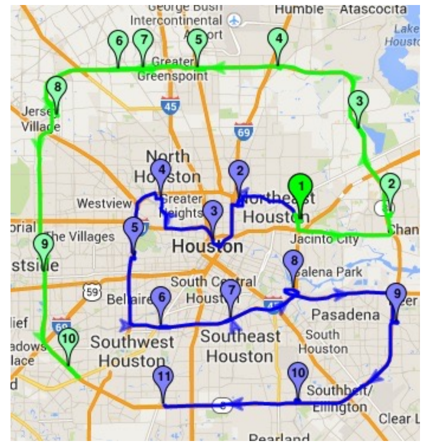
First night: 5 hours, no riders. I think I need to change my strategy.

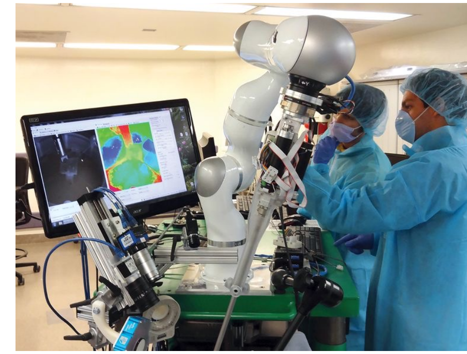
Also – Tucker et al 2002, Ibanez et al 2017, Gurvich et al 2019, Bloom et al 2020, Bavafa & Jonasson 2021, ...

# AI Can Help...

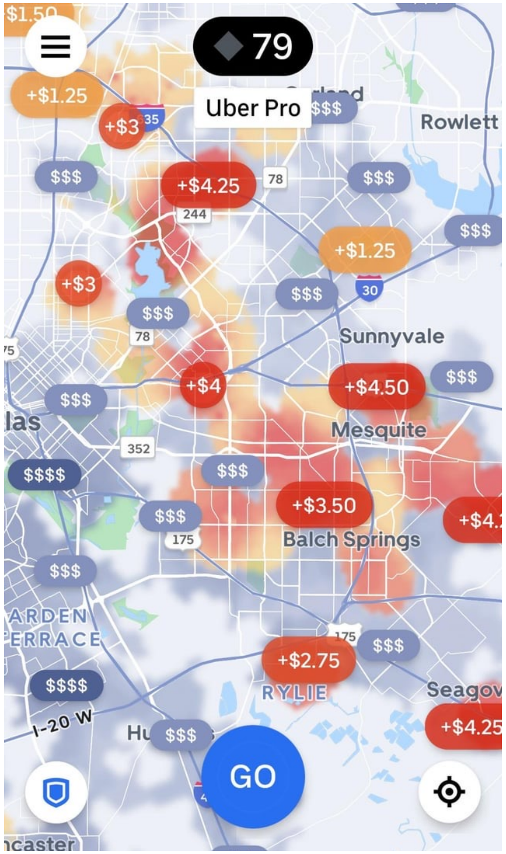

79
Uber Pro
+$1.25
+$3
+$4.25
$$$
+$1.25
+$3
$$$
+$4
+$4.50
Rowlett
Sunnyvale
Mesquite
$$$$
+$3.50
Balch Springs
+$2.75
RYLIE
Seagov
+$4.25
GO

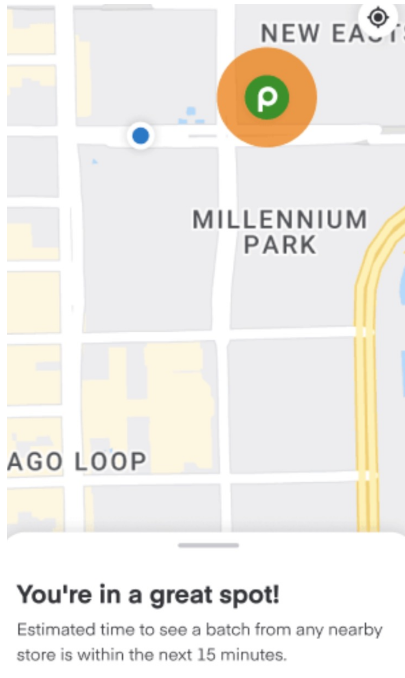

MILLENNIUM PARK
AGO LOOP
You're in a great spot!
Estimated time to see a batch from any nearby store is within the next 15 minutes.

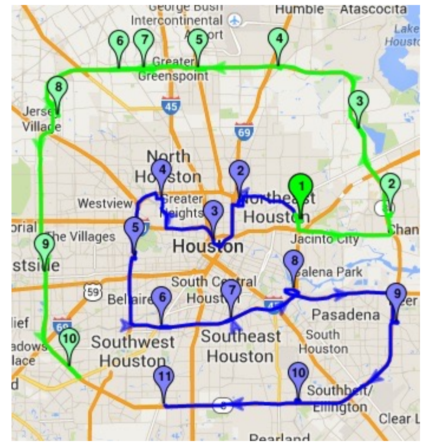

Humble
Atascocita
Greenspoint
North Houston
Westview
Houston
The Villages
Jacinto City
Galena Park
South Central Houston
Bellaire
Pasadena
South Houston
Southwest Houston
Southeast Houston
Clear L
Pearland

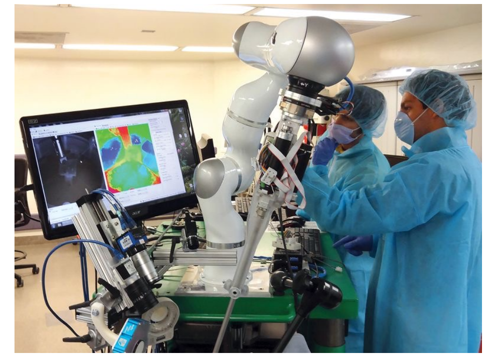


# AI Can Help…



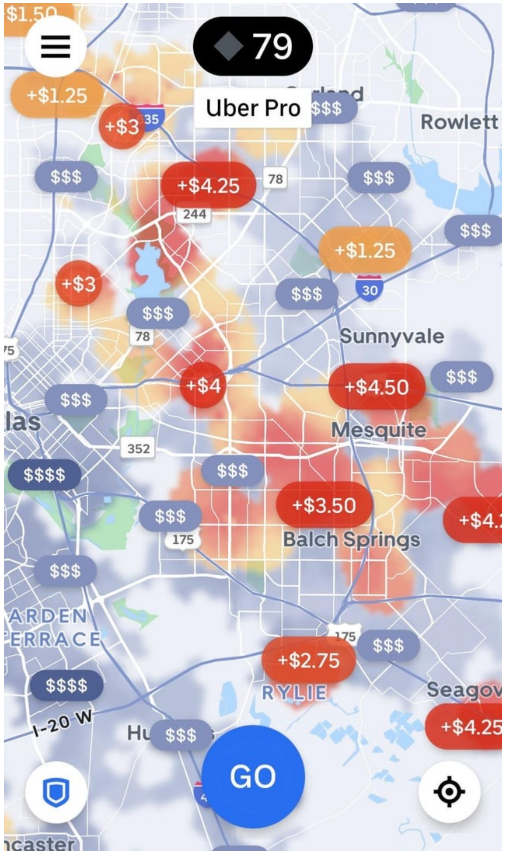



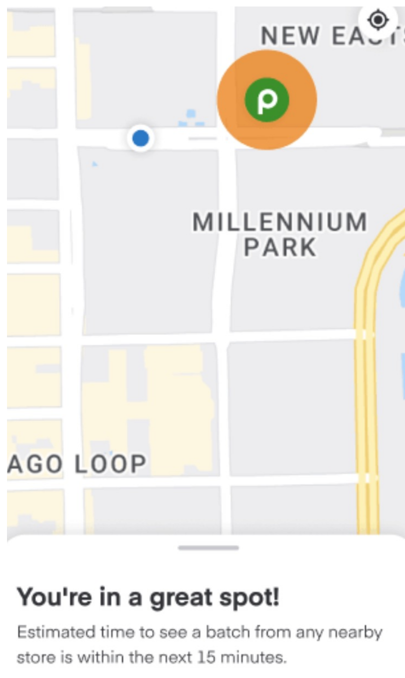



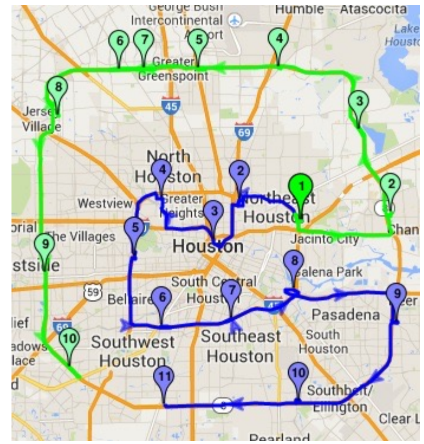
Usually, workers still have discretion over the actual action taken

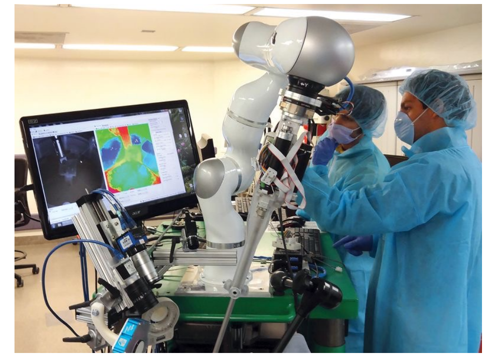
# …But Humans Deviate




# AI Can Help...



Usually, workers still have discretion over the actual action taken

Radiologists prioritize similar tasks and those they expect to complete faster. They exercise more discretion as they accumulate experience.

- Ibanez et al 2018

Dietvorst, Simmons & Massey 2015
Castelo, Bos & Lehmann 2019
Sun, Zhang, Hu & van Mieghem 2022
Balakrishnan, Ferreira & Tong 2022
Bastani, Bastani & Sinchaisri 2025

# ...But Humans Deviate

# ...Or Humans Get Worse

## 'Automation Addiction': Are Pilots Forgetting How to Fly?

Is auto-pilot weakening response time to emergency situations?

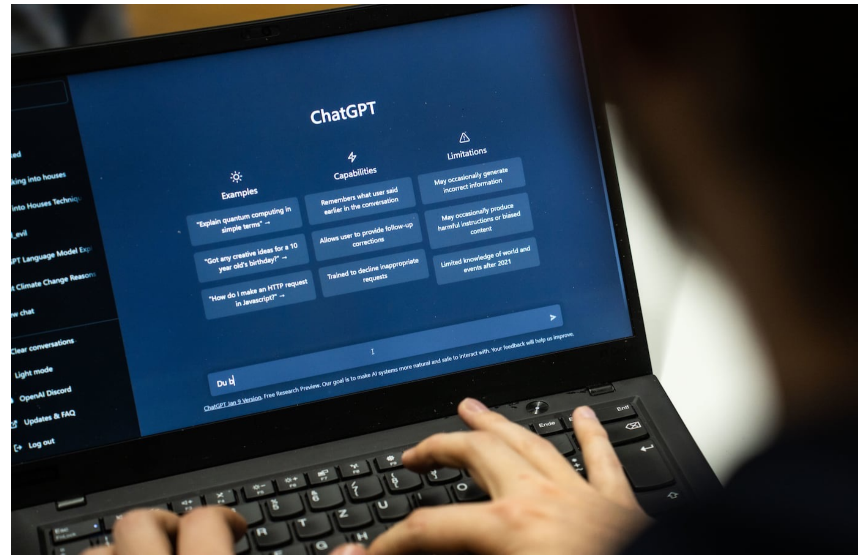
By ABC News
August 31, 2011, 3:39 AM

# ...Or Humans Get Worse

## 'Automation Addiction': Are Pilots Forgetting How to Fly?

Is auto-pilot weakening response time to emergency situations?

By ABC News
August 31, 2011, 3:39 AM




NEW YORK POST

BUSINESS

## ChatGPT outage live updates: College students spiraling during finals

By Taylor Herzlich, Ariel Zilber, Kaydi Pelletier and Isabella Bernabeo
Updated June 10, 2025, 12:51 p.m. ET

10 Comments

# How to Help Humans Learn to
# Make Better Sequential Decisions

# How to Help Humans Learn to
# Make Better Sequential Decisions

"Driving Game"

How Humans Make Sequential Decisions?

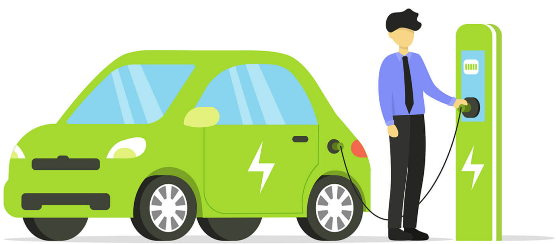
# How to Help Humans Learn to
# Make Better Sequential Decisions

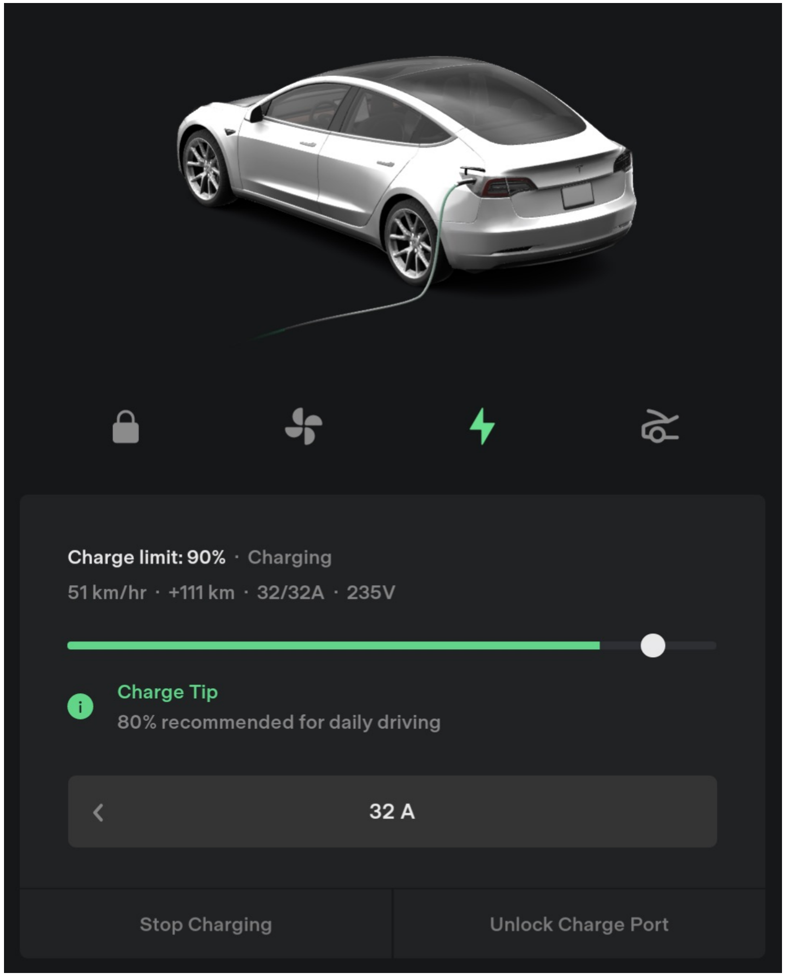
AI Tips

"Driving Game"

How Humans Make Sequential Decisions?

How Do Tips Shape Learning & Behavior Over Time?

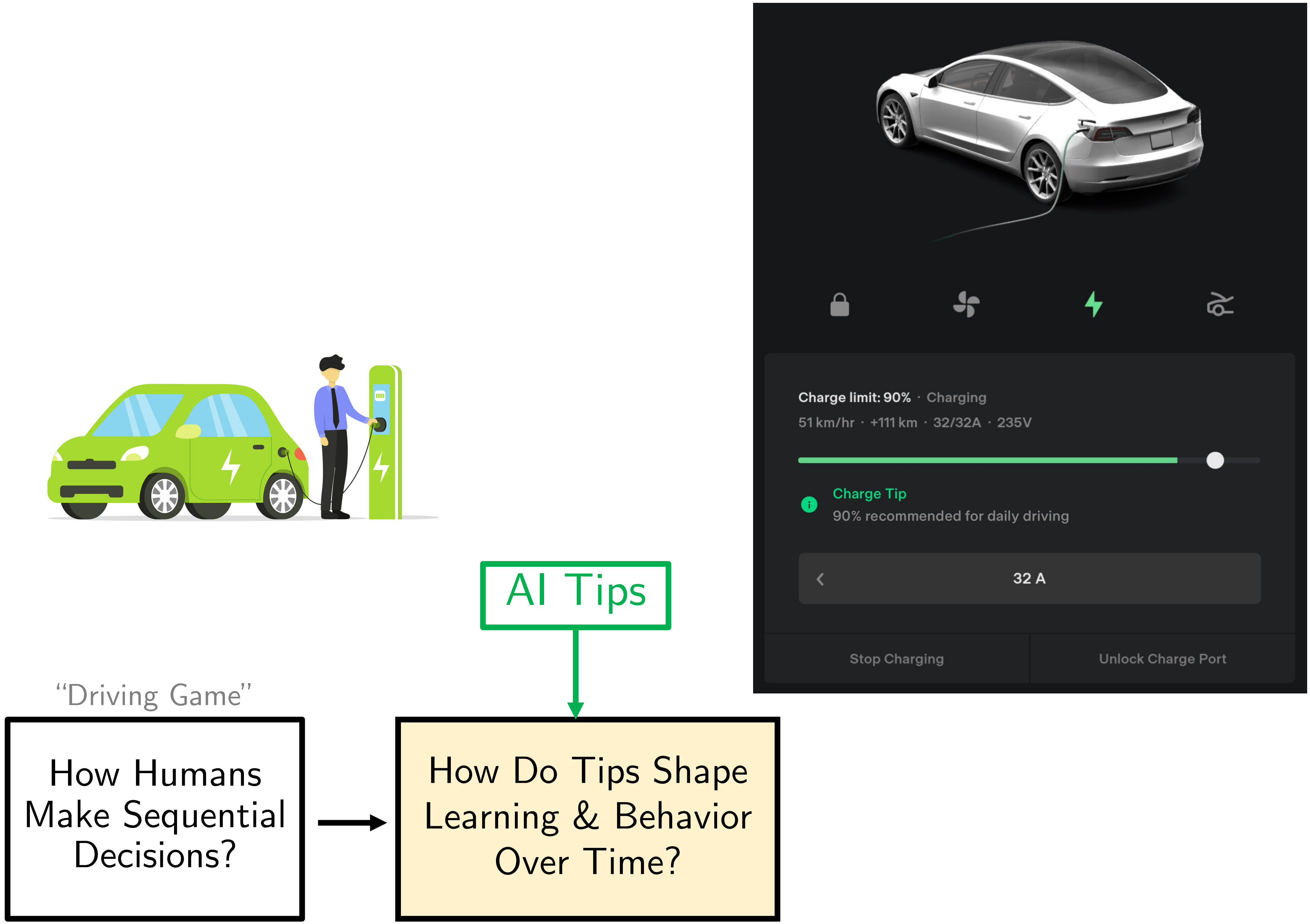




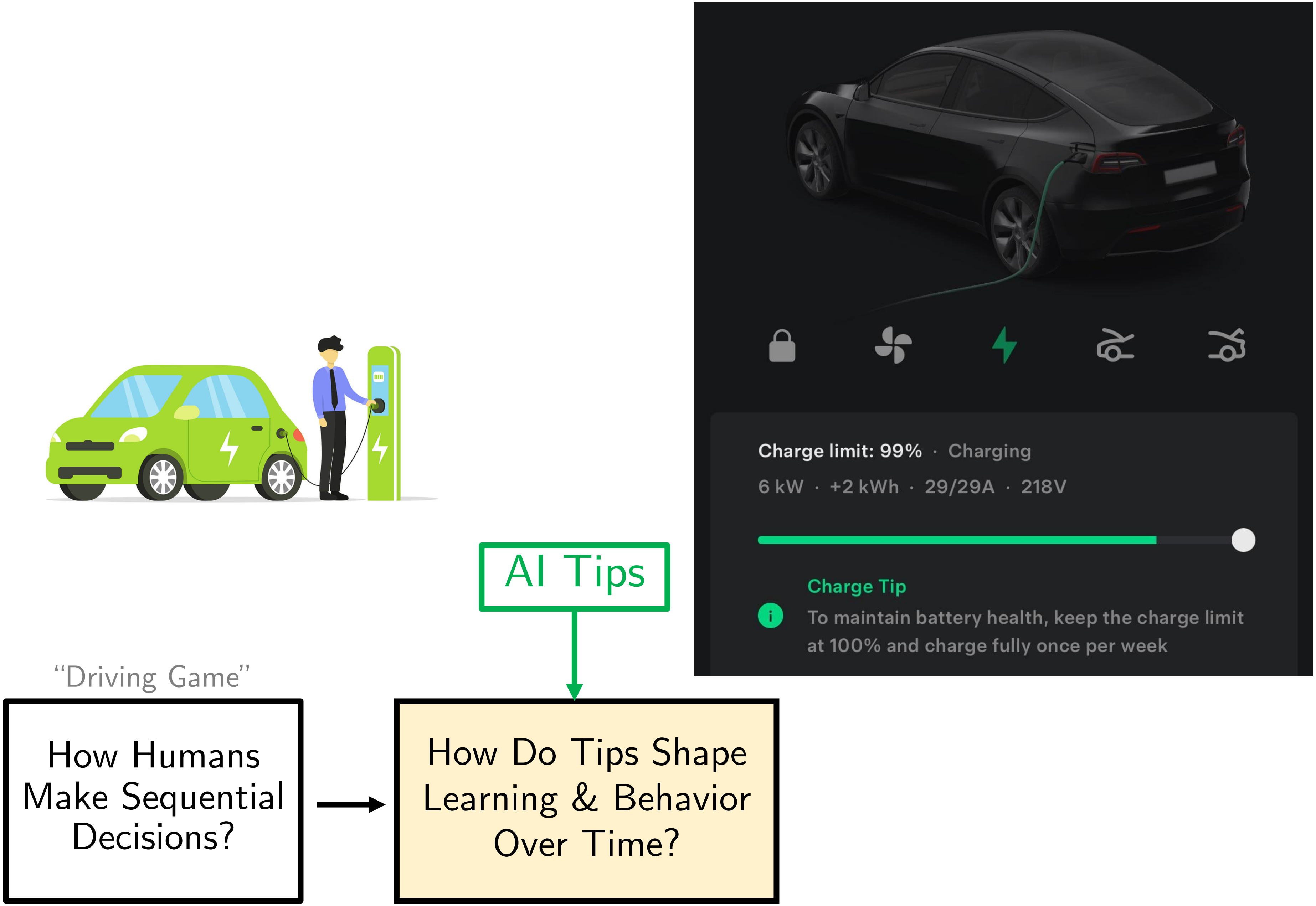

AI Tips

"Driving Game"

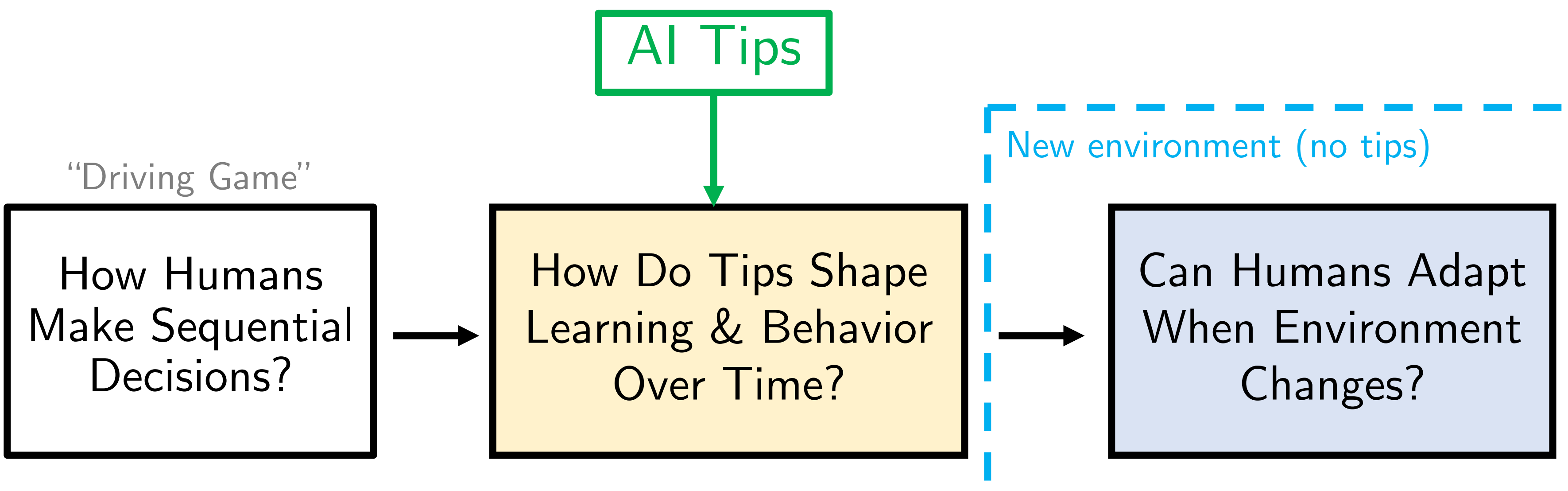
How Humans Make Sequential Decisions? → How Do Tips Shape Learning & Behavior Over Time?



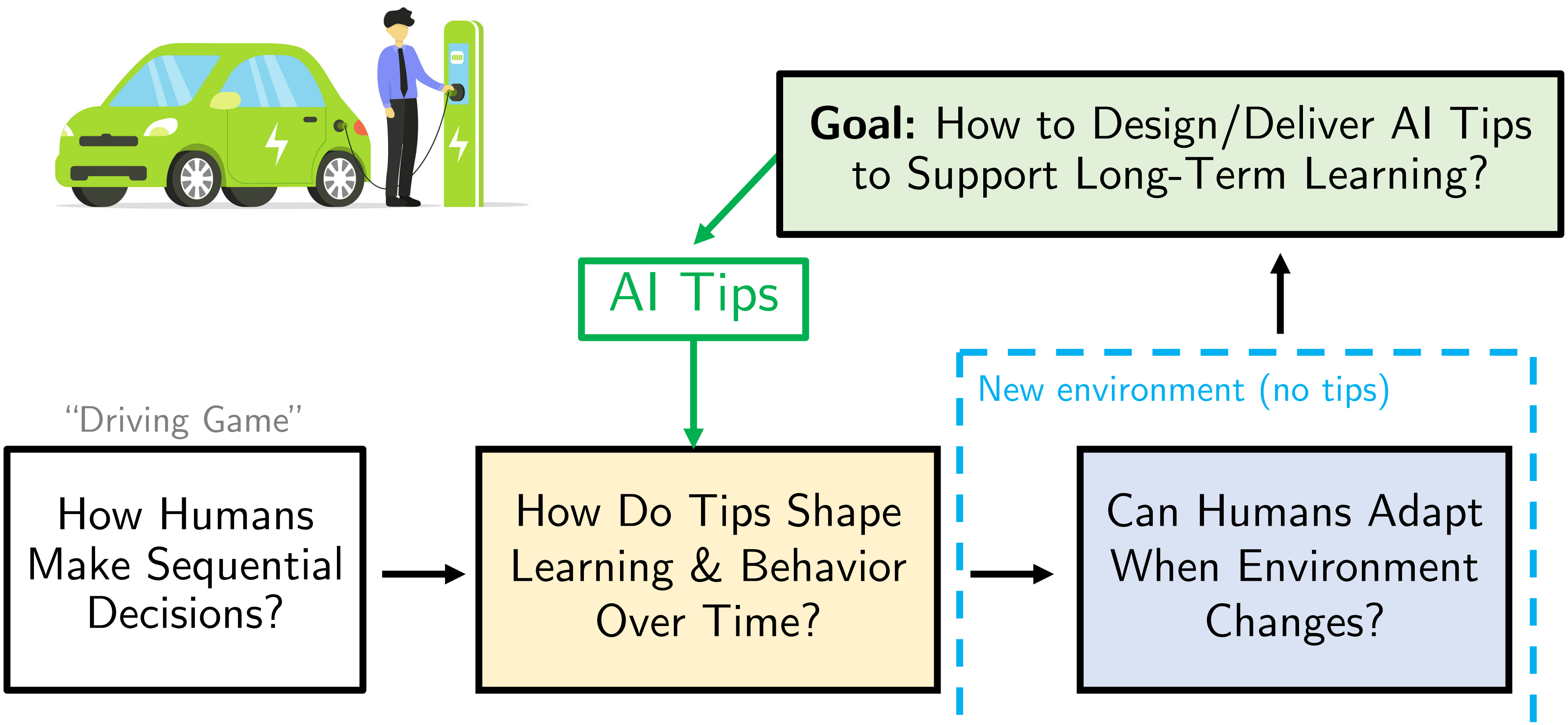
Charge limit: 90% · Charging
51 km/hr · +111 km · 32/32A · 235V
Charge Tip
90% recommended for daily driving
32 A
Stop Charging
Unlock Charge Port
AI Tips
"Driving Game"
How Humans Make Sequential Decisions?
How Do Tips Shape Learning & Behavior Over Time?


Charge limit: 99% · Charging
6 kW · +2 kWh · 29/29A · 218V
Charge Tip
To maintain battery health, keep the charge limit at 100% and charge fully once per week
AI Tips
"Driving Game"
How Humans Make Sequential Decisions?
How Do Tips Shape Learning & Behavior Over Time?

# How to Help Humans Learn to **Make Better Sequential Decisions** Even After Advice is Gone?

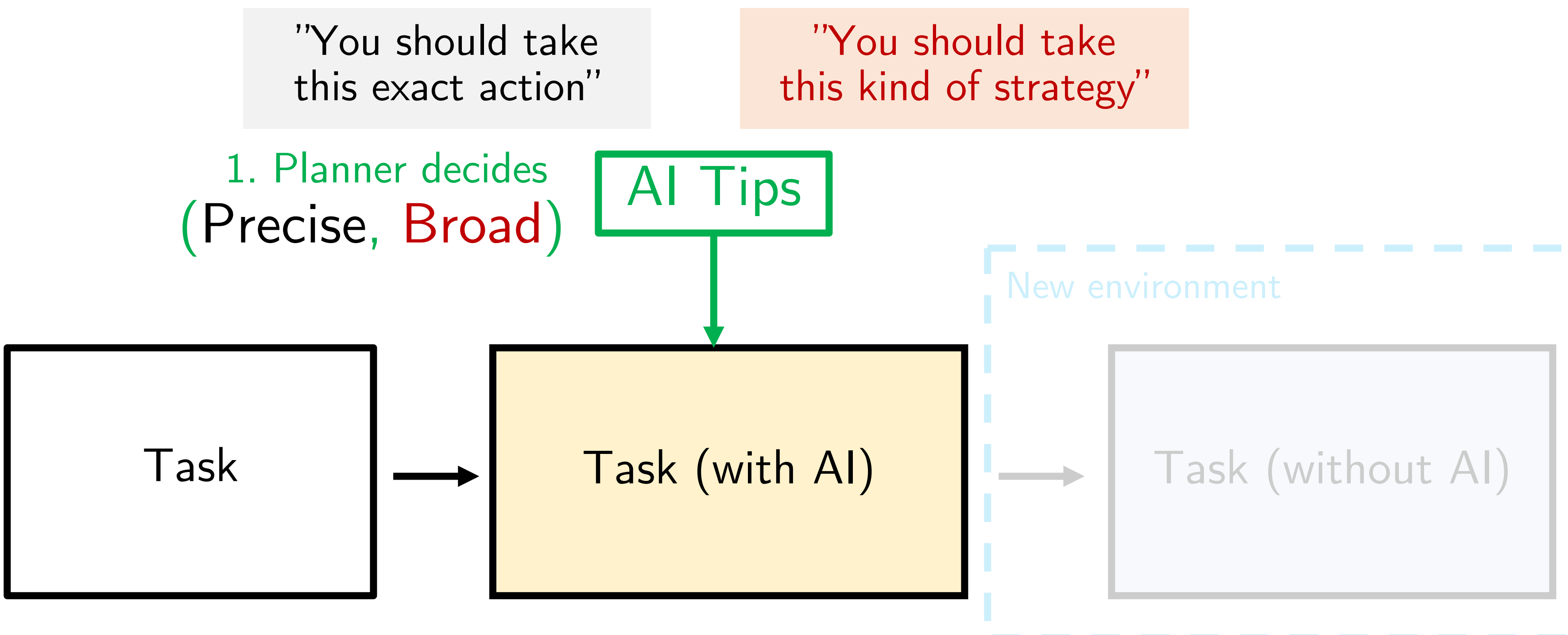


# How to Help Humans Learn to Make Better Sequential Decisions Even After Advice is Gone?



(Super Brief)

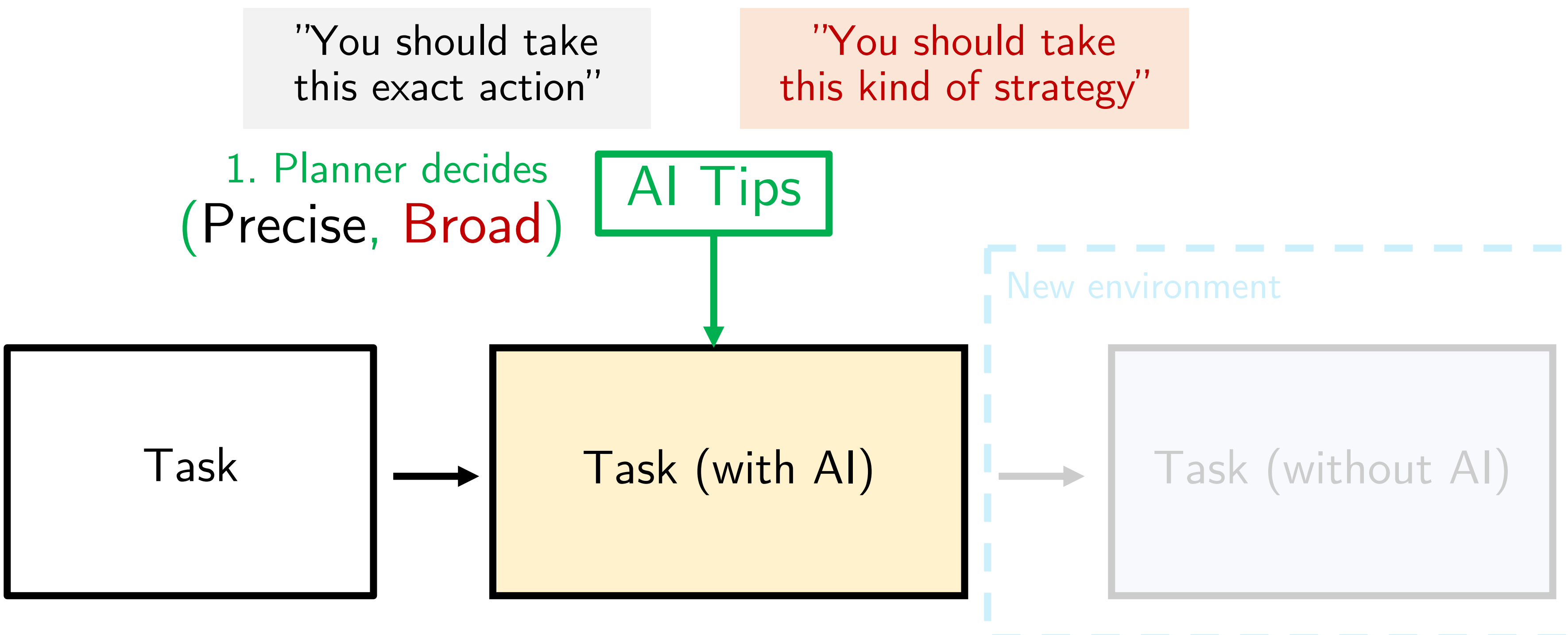
# Model Overview



(Super Brief)

# Model Overview

2. Worker decides effort $e_1 \in [0, 1]$ with cost $c(e_1) = \frac{k}{2}e_1^2$ with $k > 0$.



(Super Brief)

# Model Overview

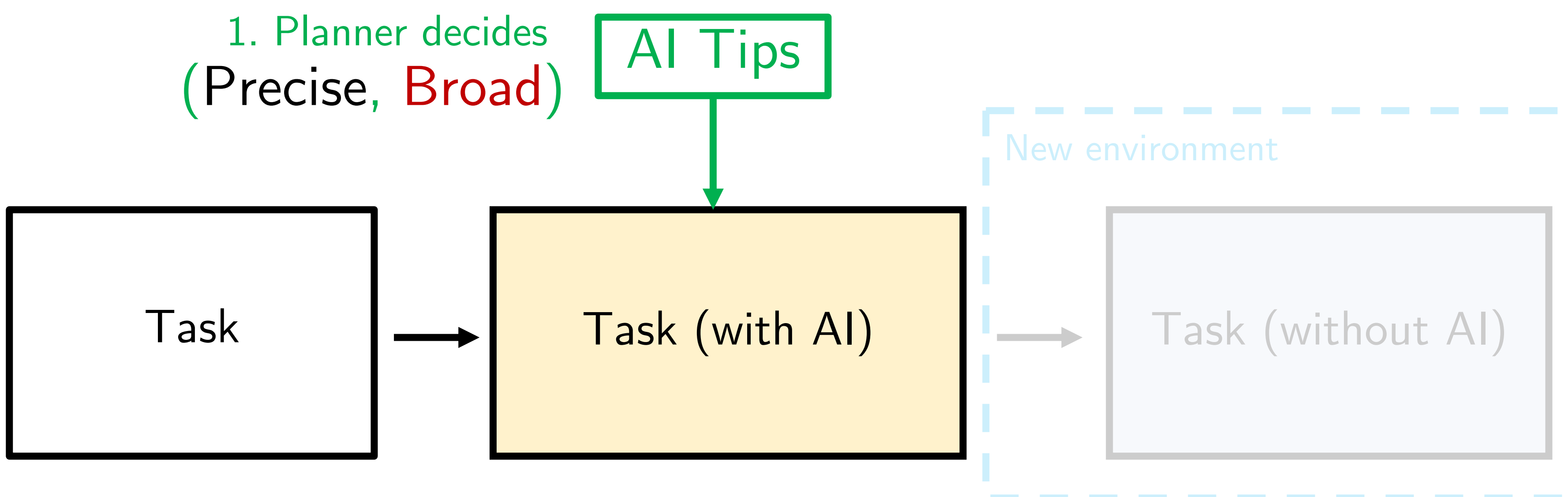
2. Worker decides effort $e_1 \in [0, 1]$ with cost $c(e_1) = \frac{k}{2} e_1^2$ with $k > 0$.
3. Worker follows the advice with probability: $\pi_a^1(e_1) = \alpha_a + \beta_a e_1$

$$0 \le \alpha_b < \alpha_p < 1$$

Precise tip is easier to follow

$$0 \le \beta_p < \beta_b < 1$$

Broad tip provides strategic insight, better converting efforts into results



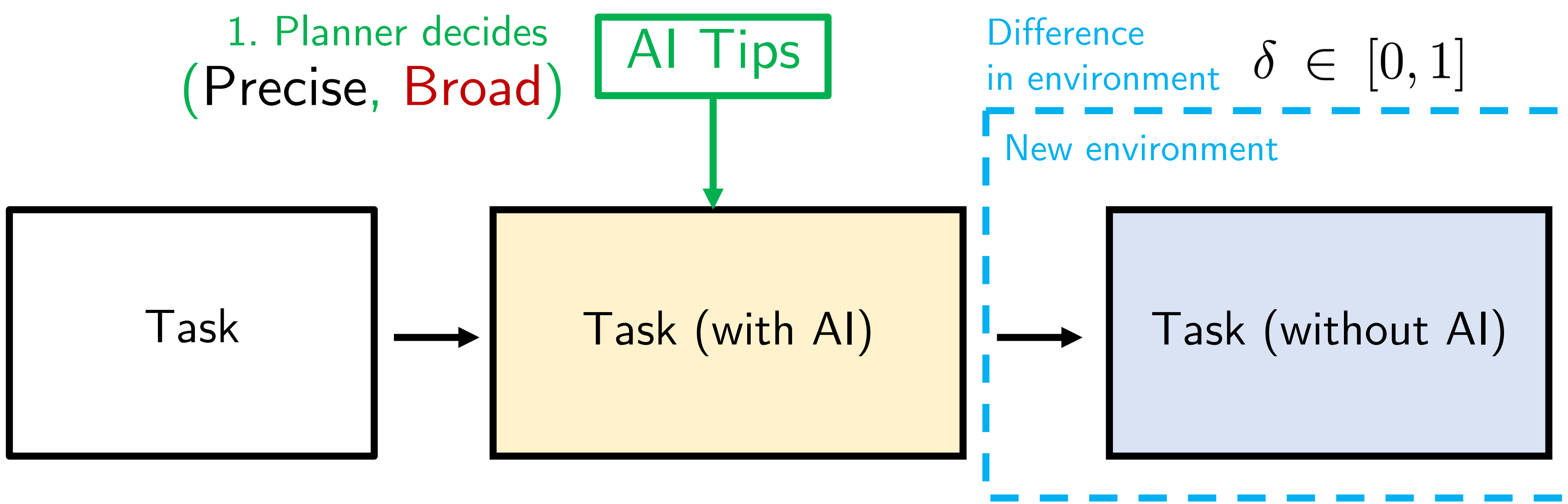
(Super Brief)

# Model Overview

2. Worker decides effort $e_1 \in [0,1]$ with cost $c(e_1) = \frac{k}{2}e_1^2$ with $k > 0$.

3. Worker follows the advice with probability: $\pi_a^1(e_1) = \alpha_a + \beta_a e_1$

4. Workers decides effort $e_2 \in [0,1]$ at cost $c(e_2) = \frac{k}{2}e_2^2$



(Super Brief)

# Model Overview

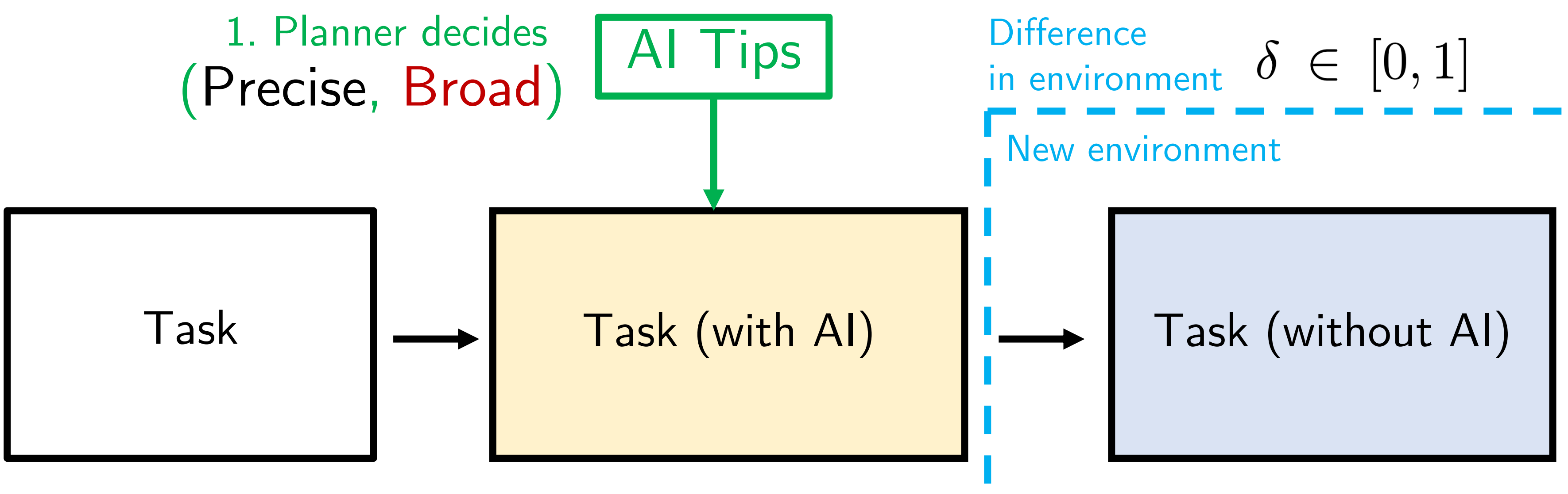
2. Worker decides effort $e_1 \in [0,1]$ with cost $c(e_1) = \frac{k}{2}e_1^2$ with $k > 0$.

3. Worker follows the advice with probability: $\pi_a^1(e_1) = \alpha_a + \beta_a e_1$

4. Workers decides effort $e_2 \in [0,1]$ at cost $c(e_2) = \frac{k}{2}e_2^2$

5. Worker selects best action with $\pi_a^2(e_1, e_2) = e_2\left[\lambda + (1-\delta)\,\omega_a e_1\right]$

Broad tip is more transferrable $\omega_b > \omega_p$



(Super Brief)

# Model Results

**1.** Broad tip is optimal for the planner if doing well in the new environment is more important than doing well in the current environment.
For high $\delta$, the difference needs to be more pronounced.

(Super Brief)

# Model Results

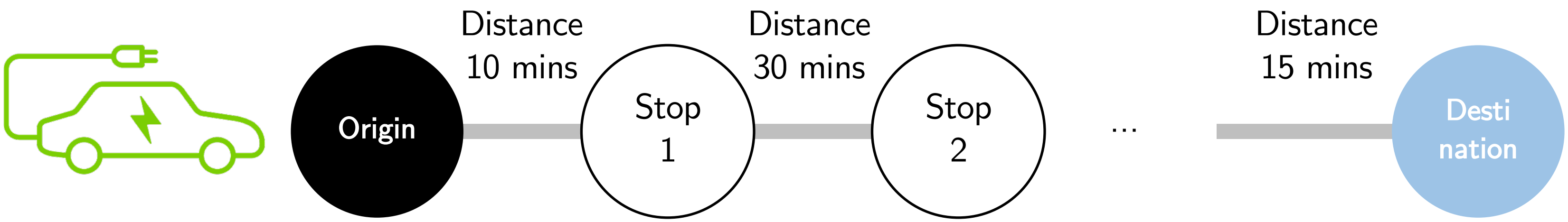
**1.** Broad tip is optimal for the planner if doing well in the new environment is more important than doing well in the current environment.
For high $\delta$, the difference needs to be more pronounced.

**2.** When the task in the initial environment is a sequential task (MDP), learning benefits of broad over precise increase with time horizon.

Pre-registered at https://aspredicted.org/7LP_J9G

🚨 Our task is *not* designed to perfectly simulate real-world battery mechanics

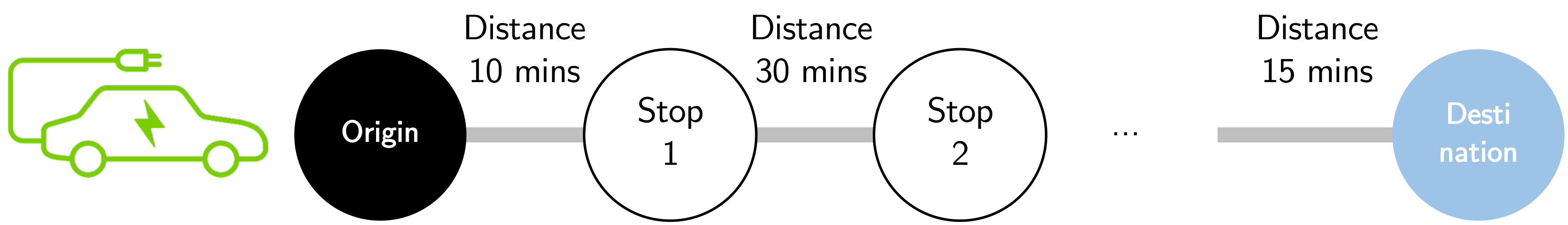
# Driving Game



Pre-registered at https://aspredicted.org/7LP_J9G

🚨 Our task is *not* designed to perfectly simulate real-world battery mechanics

# Driving Game

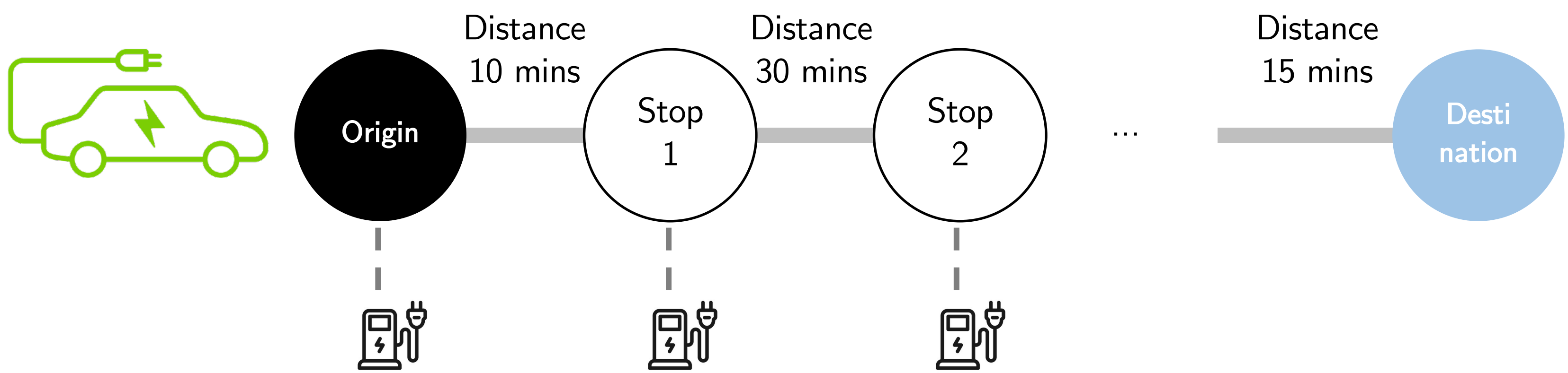



**Goal:** Get to destination as fast as possible **(1% charge can travel 1 minute)**

Pre-registered at https://aspredicted.org/7LP_J9G

Our task is *not* designed to perfectly simulate real-world battery mechanics

# Driving Game



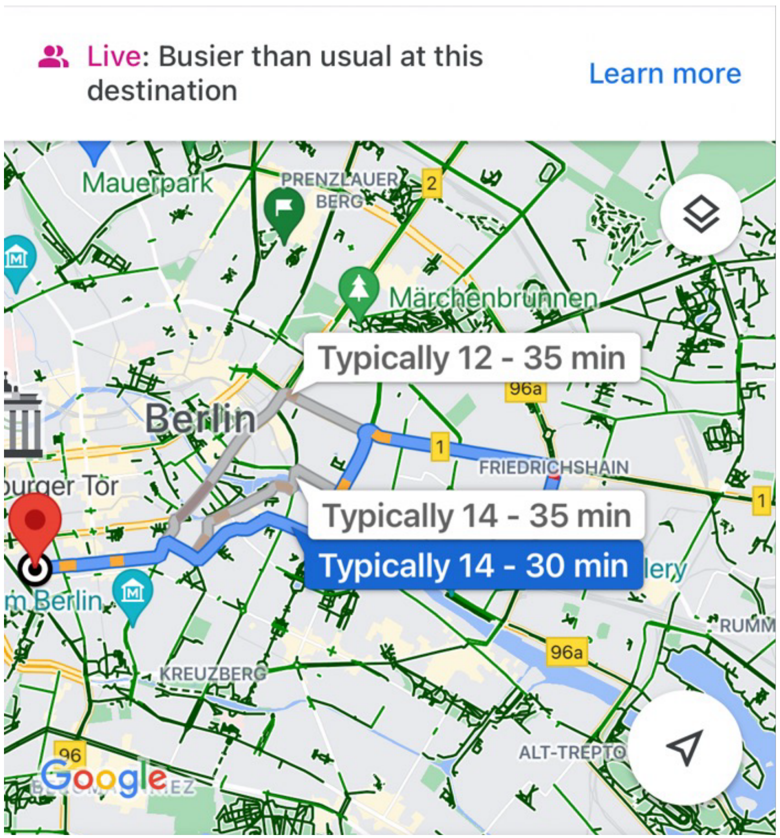

**Goal:** Get to destination as fast as possible **(1% charge can travel 1 minute)**

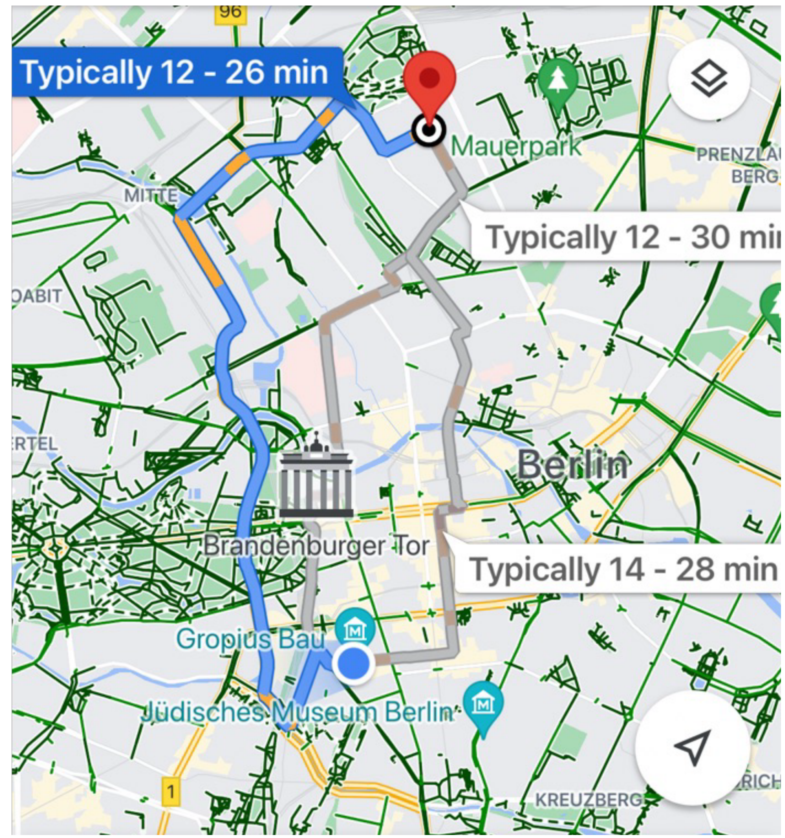
**Challenges:** Even without traffic, the amount of initial charge won't be enough

Pre-registered at https://aspredicted.org/7LP_J9G

🚨 Our task is *not* designed to perfectly simulate real-world battery mechanics

# Driving Game

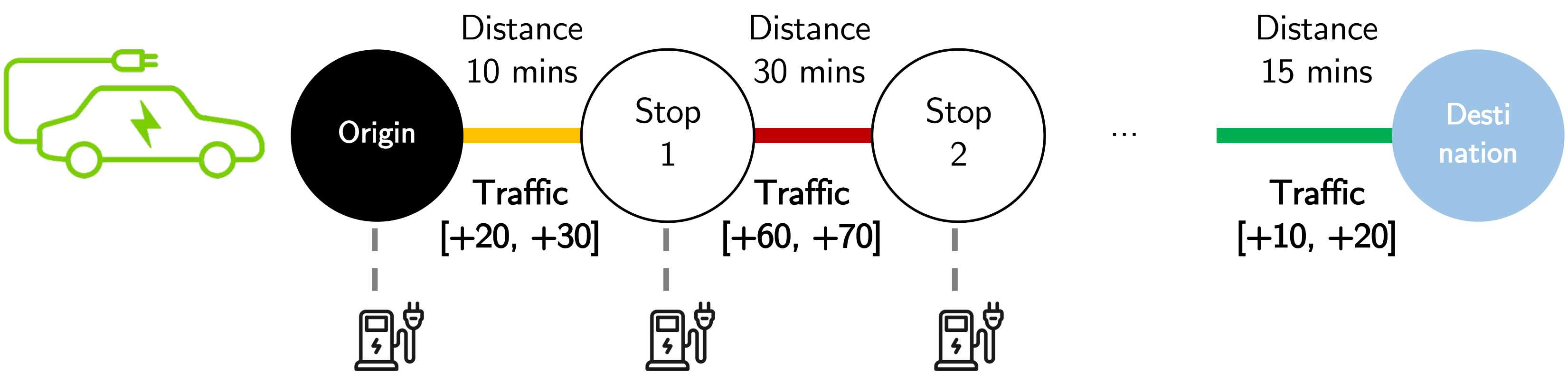


Pre-registered at https://aspredicted.org/7LP_J9G

Our task is *not* designed to perfectly simulate real-world battery mechanics

# Driving Game




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

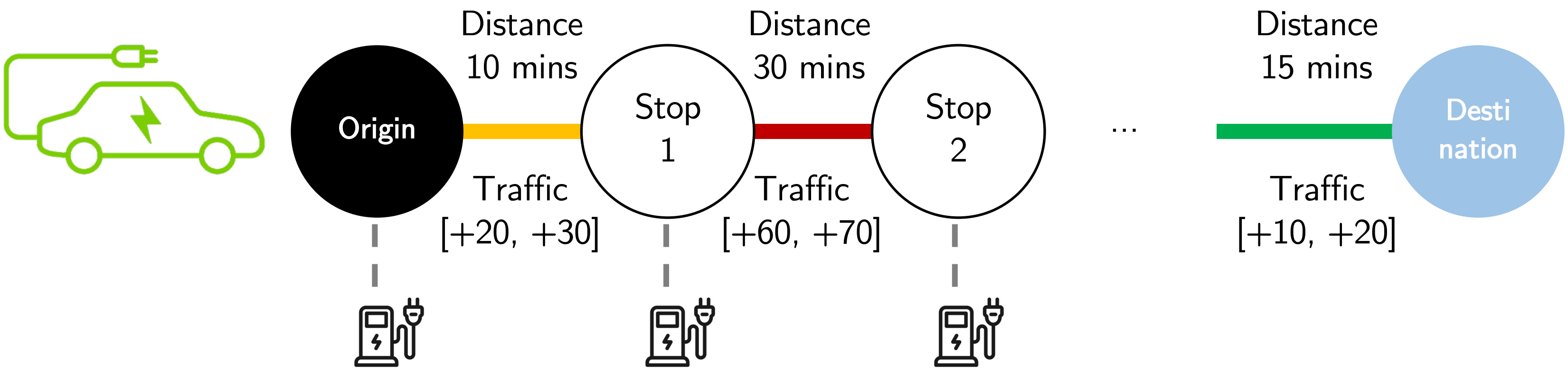
**Challenges:** Even without traffic, the amount of initial charge won't be enough

+ **Uncertain traffic**, uniformly distributed within the specified range

Pre-registered at https://aspredicted.org/7LP_J9G

🚨 Our task is *not* designed to perfectly simulate real-world battery mechanics

# Driving 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

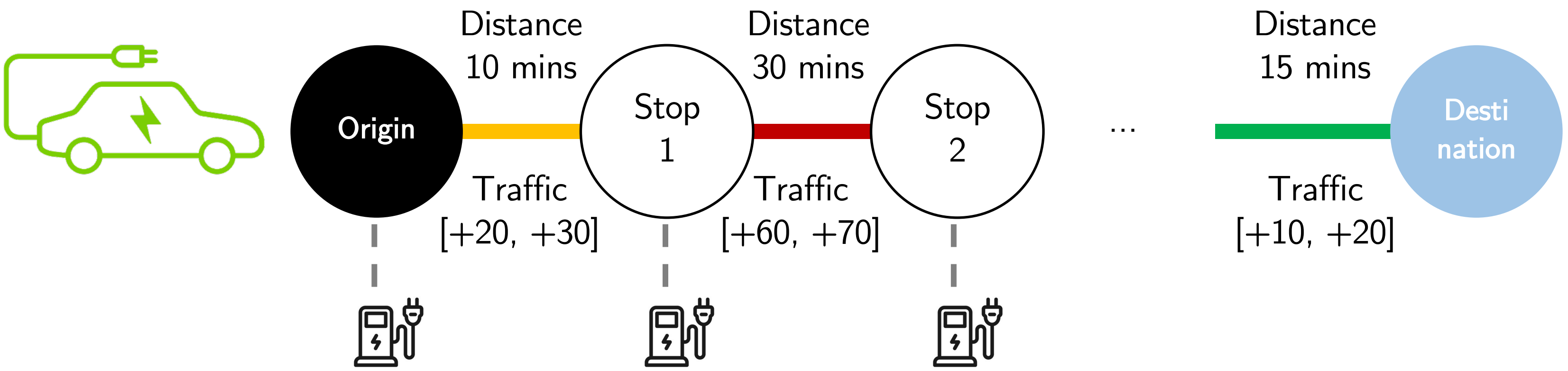
+ **Uncertain traffic**, uniformly distributed within the specified range

+ Cost to exit to charge **(+30 mins 🕑)** + **nonlinear charging time** 🕑

Pre-registered at https://aspredicted.org/7LP_J9G

🚨 Our task is *not* designed to perfectly simulate real-world battery mechanics

# Driving 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 🕒)** + **nonlinear charging time** 🕒
+ Penalty if running out of charge in the middle **(+300 mins 💀)**

Pre-registered at https://aspredicted.org/7LP_J9G

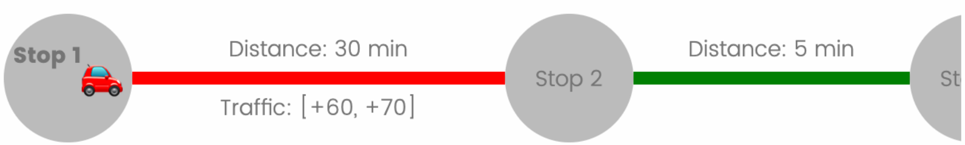
# Driving Game

**Current Stop:** 1

**Elapsed Trip Time:** 33 minutes

**Current Charge Level:** 67%

**Map**



Pre-registered at https://aspredicted.org/7LP_J9G

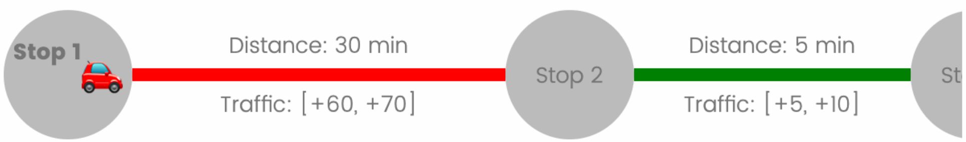
# Driving Game

**Current Stop:** 1

**Elapsed Trip Time:** 33 minutes

**Current Charge Level:** 67%

**Forward-looking behavior**



Pre-registered at https://aspredicted.org/7LP_J9G

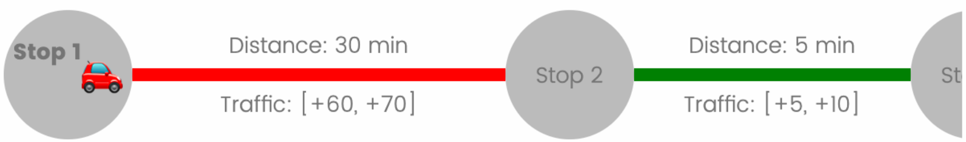
# Driving Game

**Current Stop:** 1

**Elapsed Trip Time:** 33 minutes

**Current Charge Level:** 67%

**Forward-looking behavior**

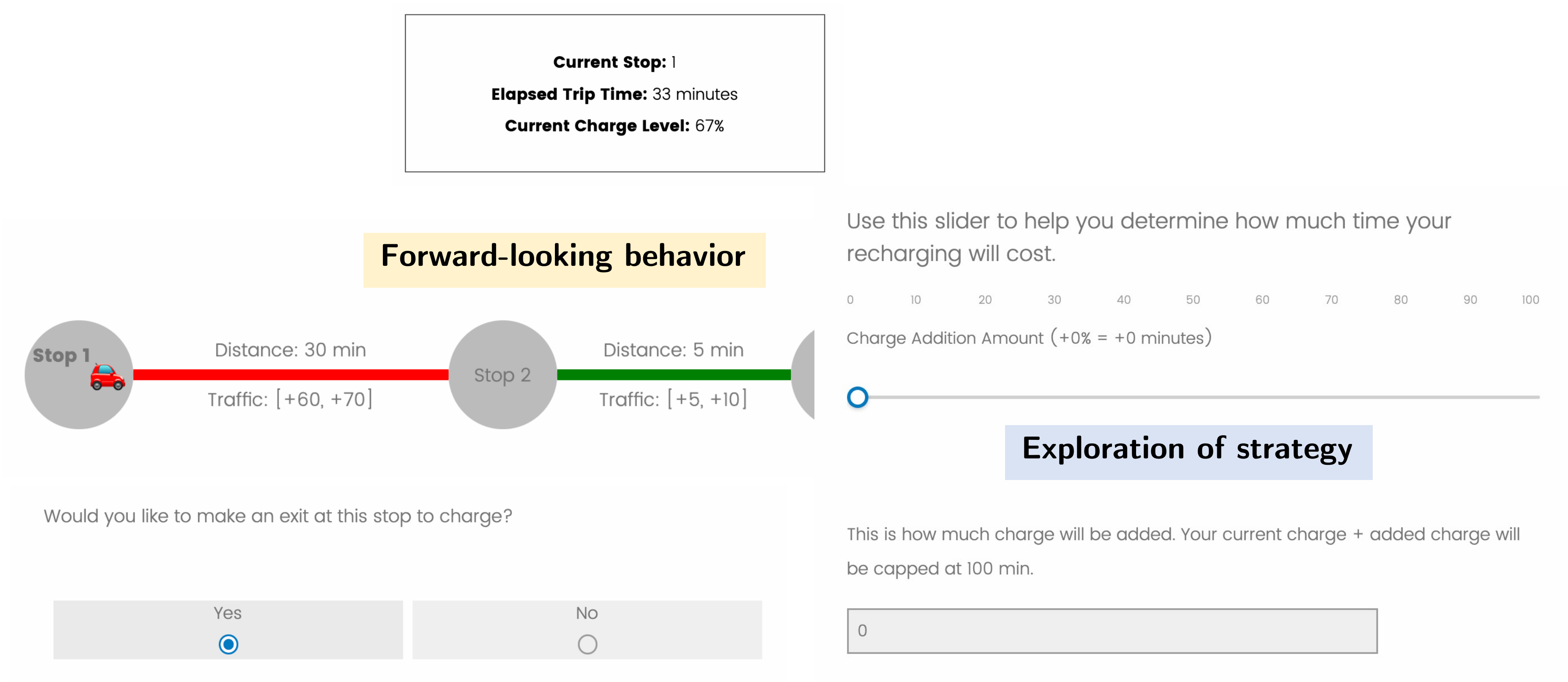



Would you like to make an exit at this stop to charge?

| Yes | No |
|---|---|
| ◉ | ○ |

Pre-registered at https://aspredicted.org/7LP_J9G

# Driving Game



Pre-registered at https://aspredicted.org/7LP_J9G

# Driving Game

**Current Stop:** 1
**Elapsed Trip Time:** 33 minutes
**Current Charge Level:** 67%

**Forward-looking behavior**


Would you like to make an exit at this stop to charge?

Yes (selected) | No

Use this slider to help you determine how much time your recharging will cost.

0 10 20 30 40 50 60 70 80 90 100

Charge Addition Amount (+15% = +80 minutes)

15

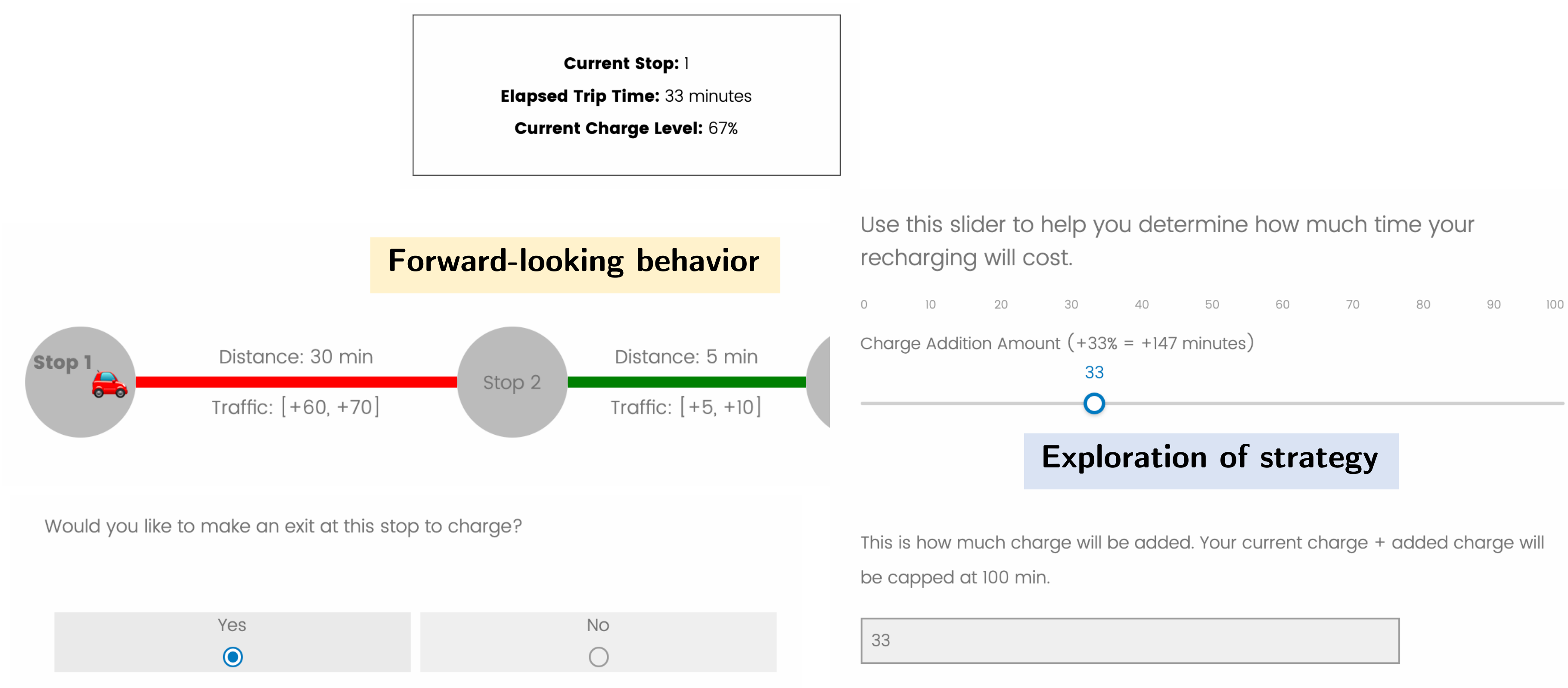
**Exploration of strategy**

This is how much charge will be added. Your current charge + added charge will be capped at 100 min.

15

Pre-registered at https://aspredicted.org/7LP_J9G

# Driving Game



Pre-registered at https://aspredicted.org/7LP_J9G

# Driving Game

Feedback after each round

**Previous Stop Summary**

**Distance:** 10
**Actual Traffic:** 23 min (from expected [+20, +30] min)
**Duration:** 33 min

**Current Stop:** 1
**Elapsed Trip Time:** 33 minutes
**Current Charge Level:** 67%

**Forward-looking behavior**


Would you like to make an exit at this stop to charge?

Yes ◉ No ○

Use this slider to help you determine how much time your recharging will cost.

0 10 20 30 40 50 60 70 80 90 100

Charge Addition Amount (+33% = +147 minutes)

33

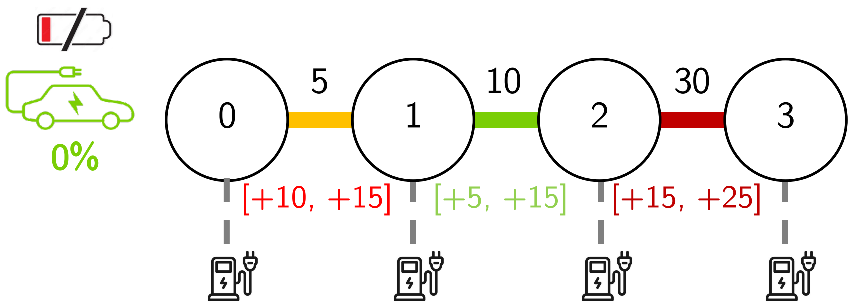
**Exploration of strategy**

This is how much charge will be added. Your current charge + added charge will be capped at 100 min.

33

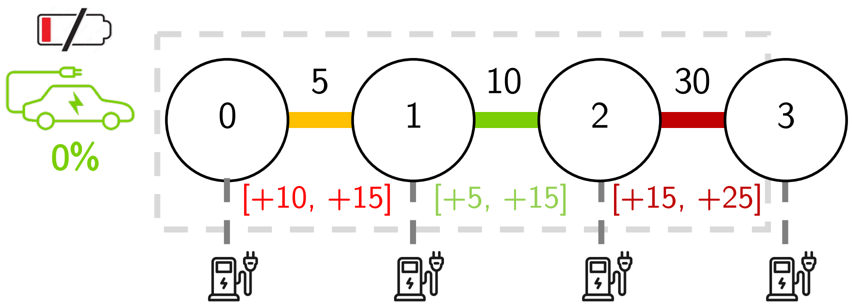
Pre-registered at https://aspredicted.org/7LP_J9G

# Design

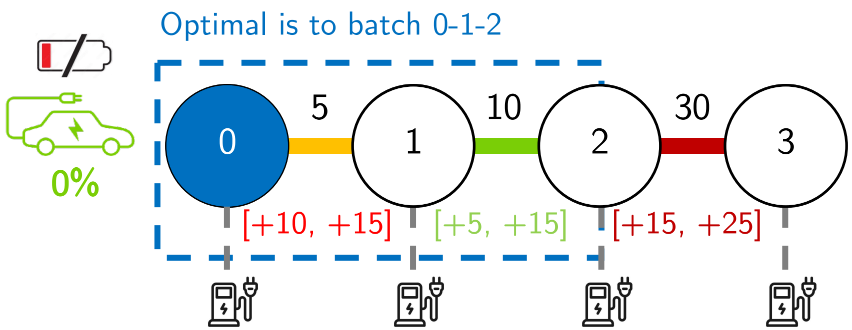


Pre-registered at https://aspredicted.org/7LP_J9G

# Design To Batch, or Not to Batch



# Design To Batch, or Not to Batch

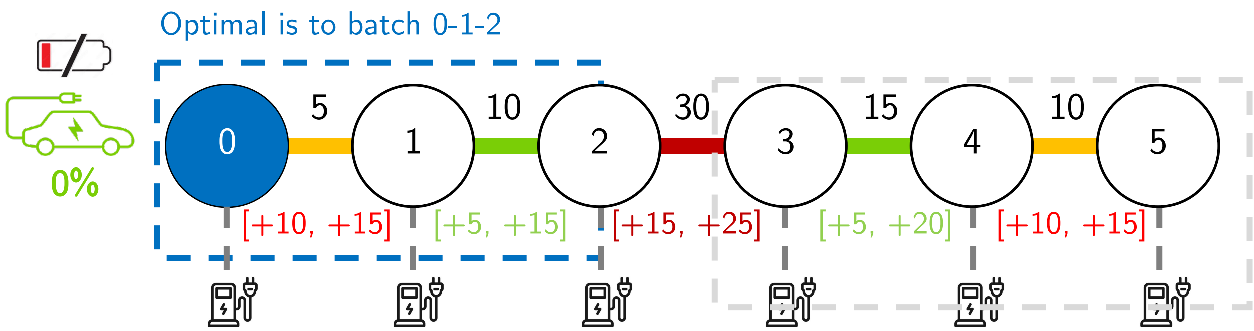


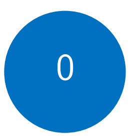

0

Optimal = **"batch"** required charges for the next two stops (0 → 2) rather than just 0 → 1 or further batch 0 → 3.

Pre-registered at https://aspredicted.org/7LP_J9G

# Design To Batch, or Not to Batch

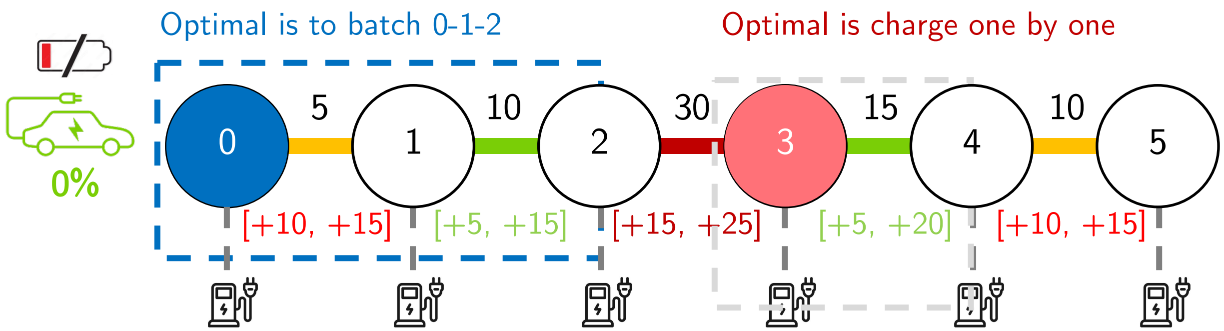


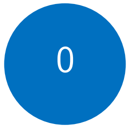




Optimal = **"batch"** required charges
for the next two stops (0 → 2)
rather than just 0 → 1 or
further batch 0 → 3.

Pre-registered at https://aspredicted.org/7LP_J9G

# Design To Batch, or Not to Batch



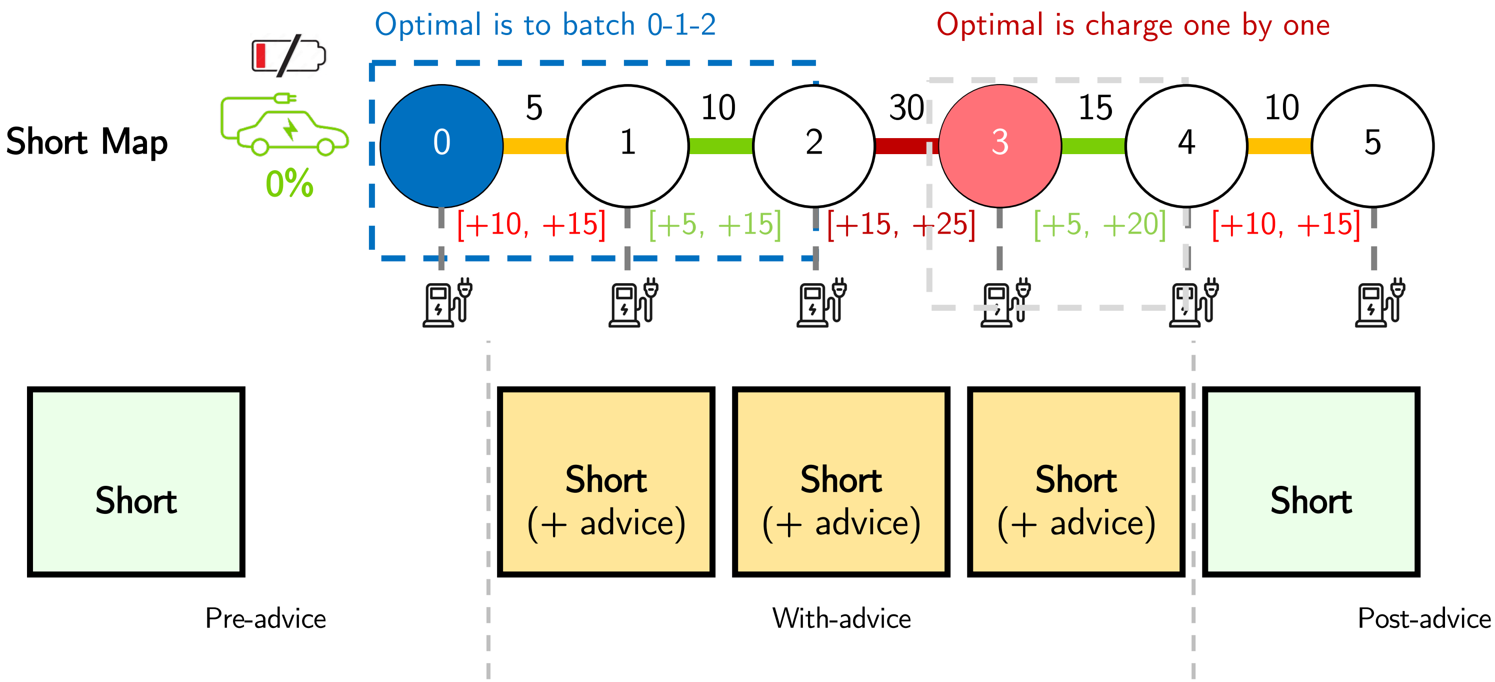




Optimal = "batch" required charges for the next two stops (0 → 2) rather than just 0 → 1 or further batch 0 → 3.


Optimal ="split" = only charge for the next stop (3 → 4) rather than batch 3 → 5.

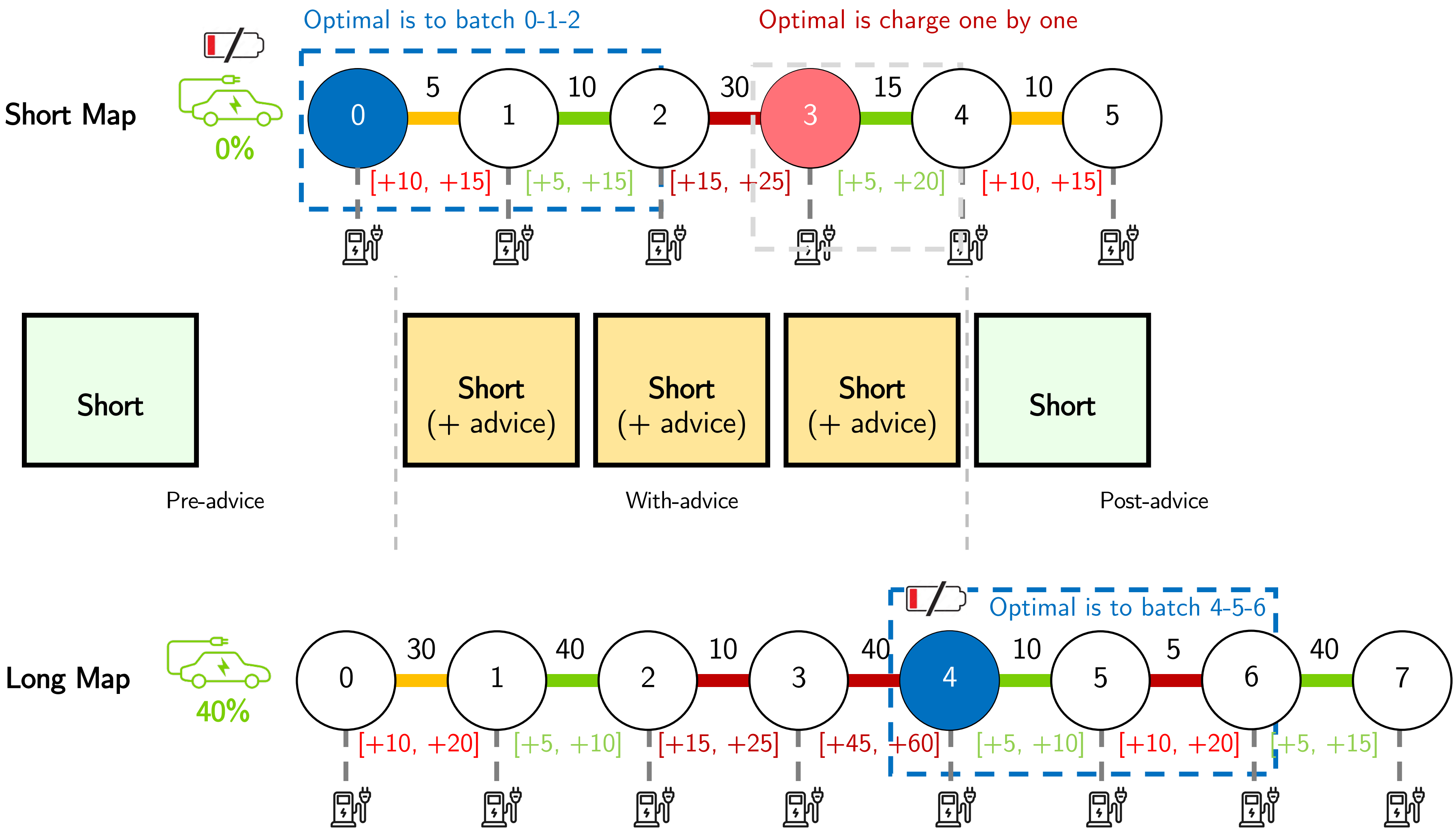
Pre-registered at https://aspredicted.org/7LP_J9G

# Study 1 Learning in One Environment



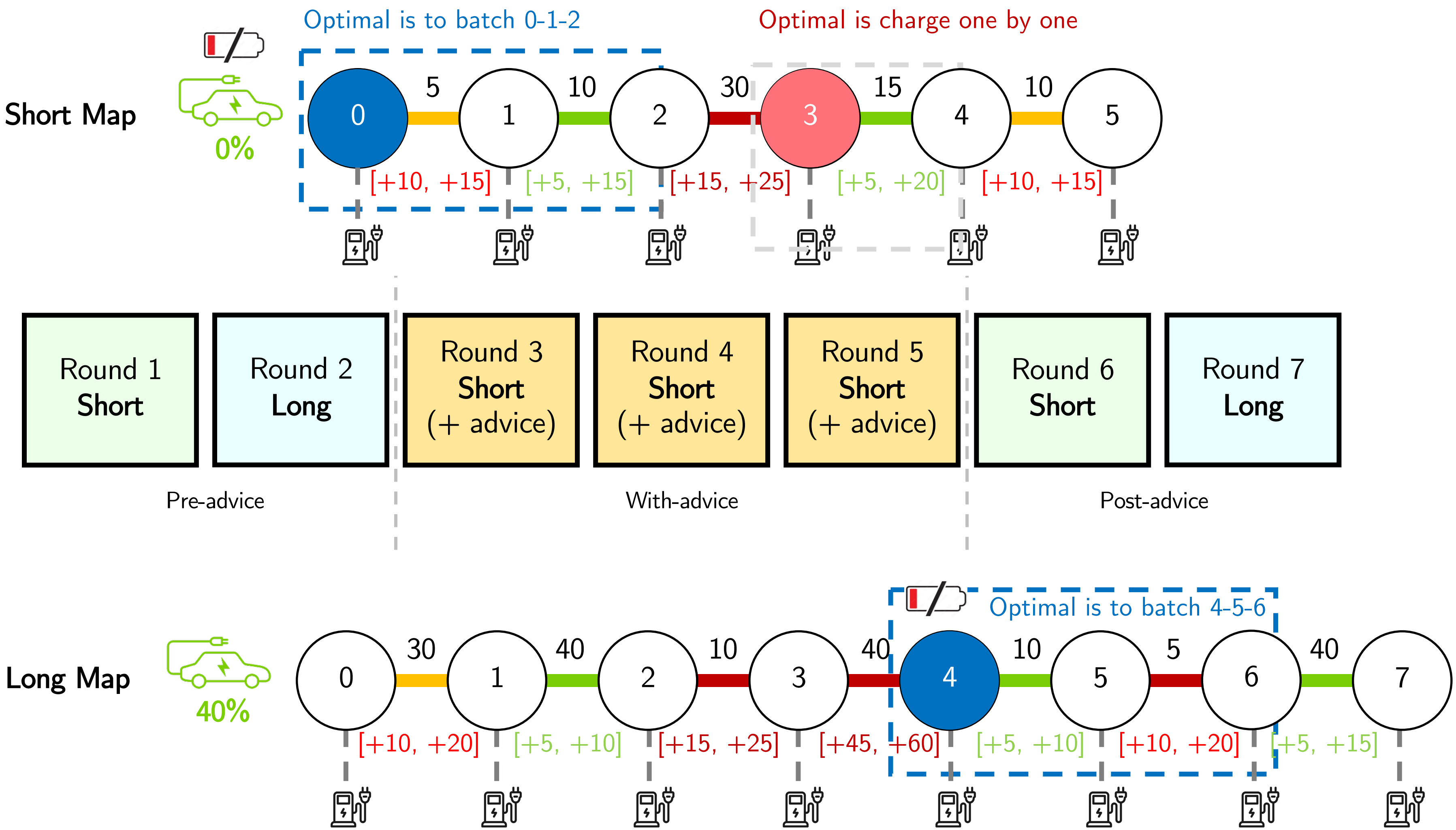
Pre-registered at https://aspredicted.org/7LP_J9G

# Study 1 + New Environment



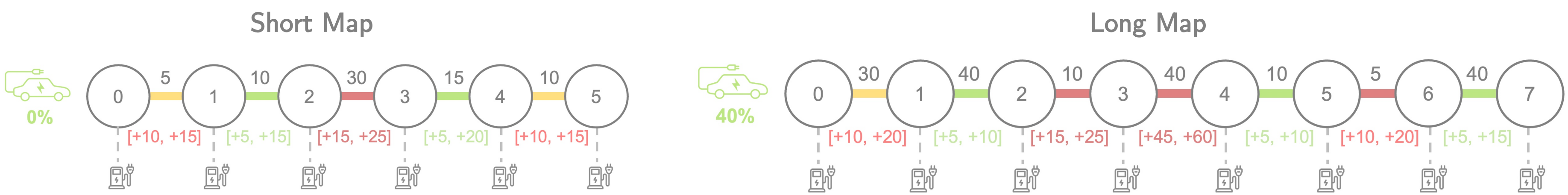
Pre-registered at https://aspredicted.org/7LP_J9G

# Study 1 + New Environment



Pre-registered at https://aspredicted.org/7LP_J9G

# Study 1 Treatment Conditions




| Round 1 Short | Round 2 Long | Round 3 Short (+ advice) | Round 4 Short (+ advice) | Round 5 Short (+ advice) | Round 6 Short | Round 7 Long |
|---|---|---|---|---|---|---|

Pre-advice | With-advice | Post-advice

You should **charge** X%

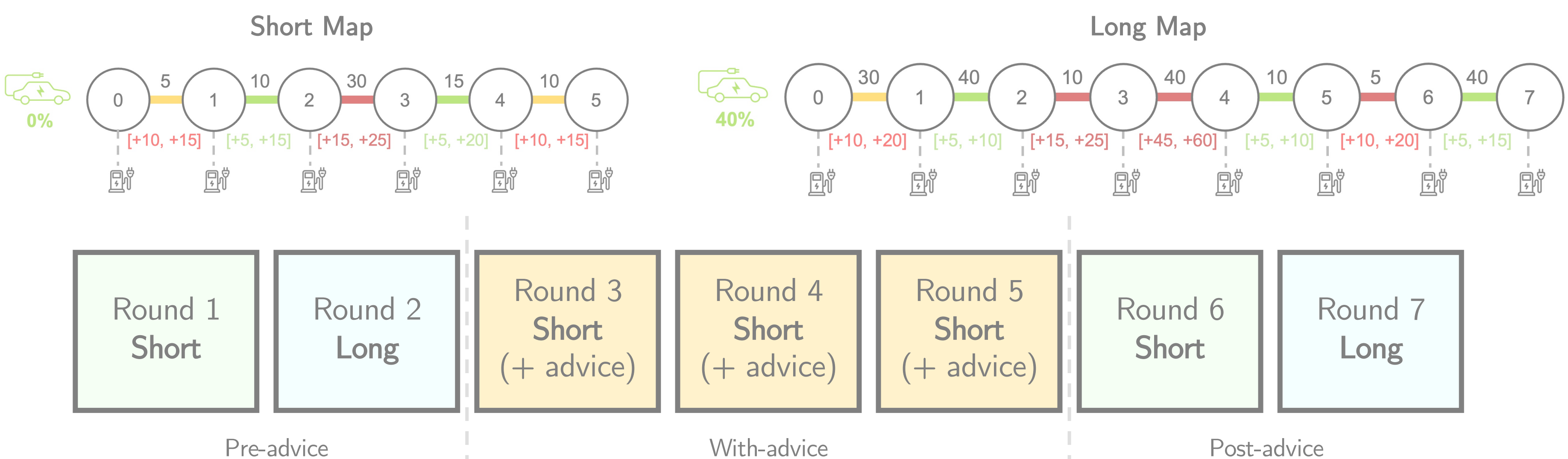
You should **charge just enough for** this segment + the next one

**2**

**precise** / **broad**
advice

Pre-registered at https://aspredicted.org/7LP_J9G

# Study 1 Treatment Conditions




You should **charge** X%

You should **charge just enough for** this segment + the next one

# 2

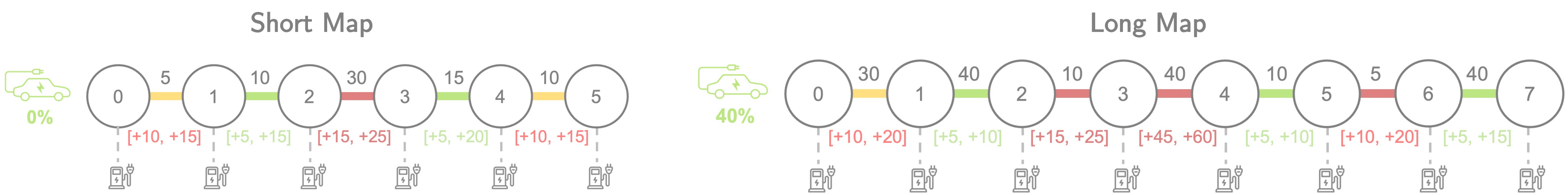
**precise** / **broad**
advice

Paramedics experiencing one (two) prior critical incident(s) spend 2.6% (7.5%) more time completing their tasks (Bavafa & Jonasson 2020)

Workers facing uncertain workload: higher workload → use AI advice more (Snyder et al 2023)

Pre-registered at https://aspredicted.org/7LP_J9G

# Study 1 Treatment Conditions




| Round 1 Short | Round 2 Long | Round 3 Short (+ advice) | Round 4 Short (+ advice) | Round 5 Short (+ advice) | Round 6 Short | Round 7 Long |
|---|---|---|---|---|---|---|

Pre-advice | With-advice | Post-advice

You should **charge** X%

You should **charge just enough for** this segment + the next one

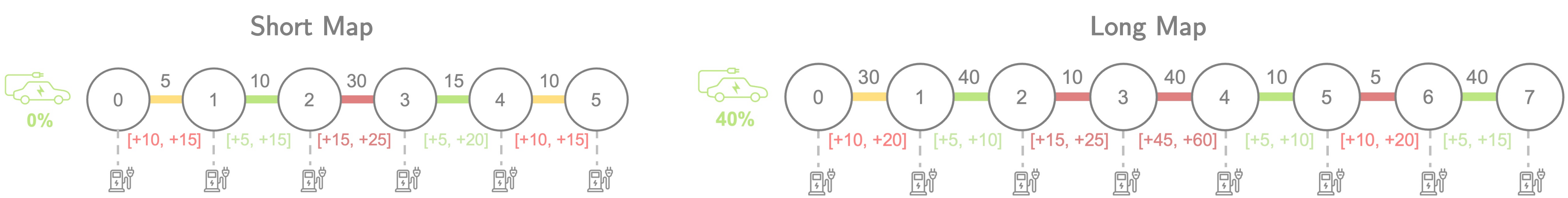
**2** precise / broad advice × **2** centered / skewed realized traffic

Paramedics experiencing one (two) prior critical incident(s) spend 2.6% (7.5%) more time completing their tasks (Bavafa & Jonasson 2020)

Workers facing uncertain workload: higher workload → use AI advice more (Snyder et al 2023)

Pre-registered at https://aspredicted.org/7LP_J9G

# Study 1 Treatment Conditions




| Round 1 Short | Round 2 Long | Round 3 Short (+ advice) | Round 4 Short (+ advice) | Round 5 Short (+ advice) | Round 6 Short | Round 7 Long |
|---|---|---|---|---|---|---|
| Pre-advice | | With-advice | | | Post-advice | |

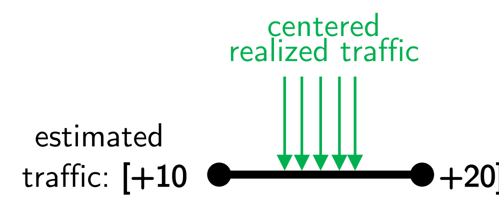
You should **charge** X%

You should **charge just enough for** this segment + the next one

**2** x **2**

**precise** / **broad** advice

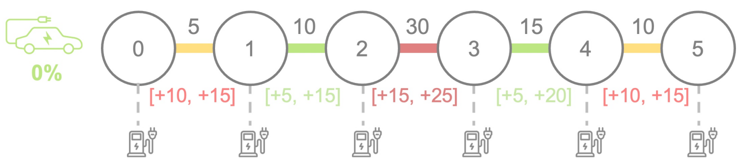
**centered** / **skewed** realized traffic



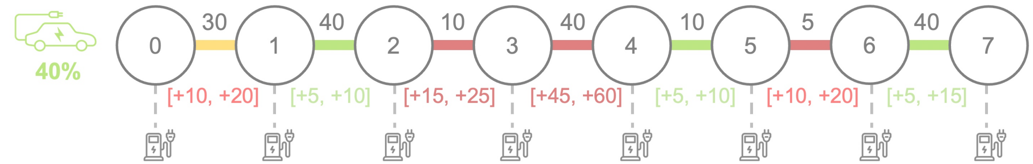
Pre-registered at https://aspredicted.org/7LP_J9G

# Study 1 Treatment Conditions

Short Map




Long Map




| Round 1 Short | Round 2 Long | Round 3 Short (+ advice) | Round 4 Short (+ advice) | Round 5 Short (+ advice) | Round 6 Short | Round 7 Long |
|---|---|---|---|---|---|---|

Pre-advice | With-advice | Post-advice

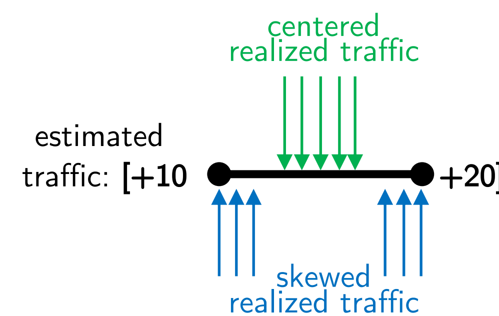
You should **charge** X%

You should **charge just enough for** this segment + the next one

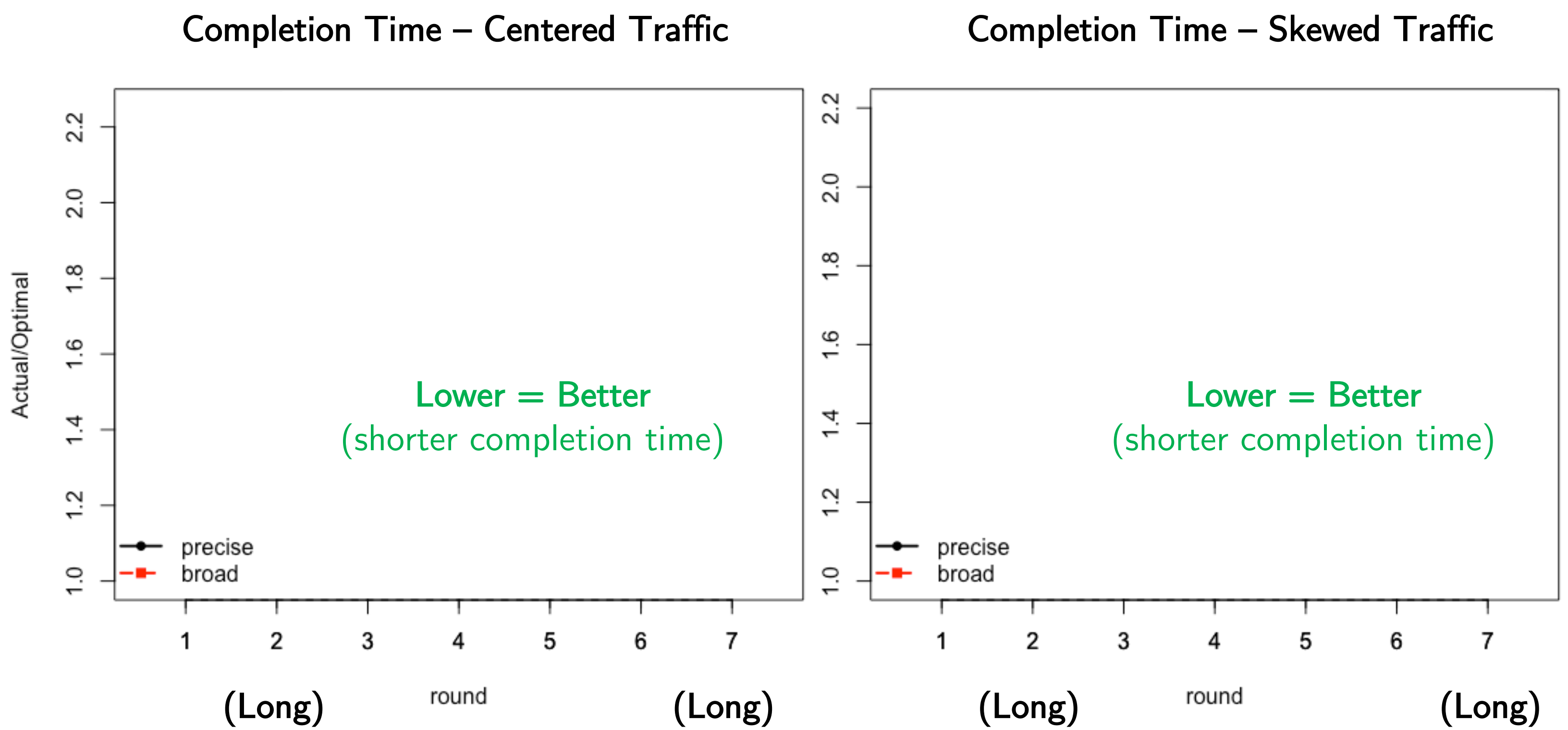
**2** x **2**

**precise** / **broad** advice

**centered** / **skewed** realized traffic



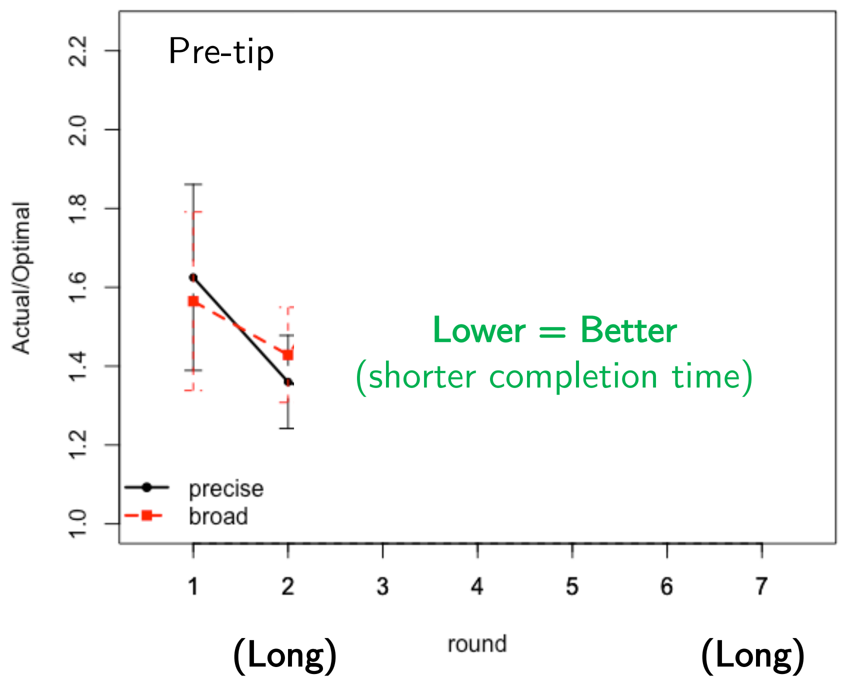
# Study 1: Results Performance Across Rounds



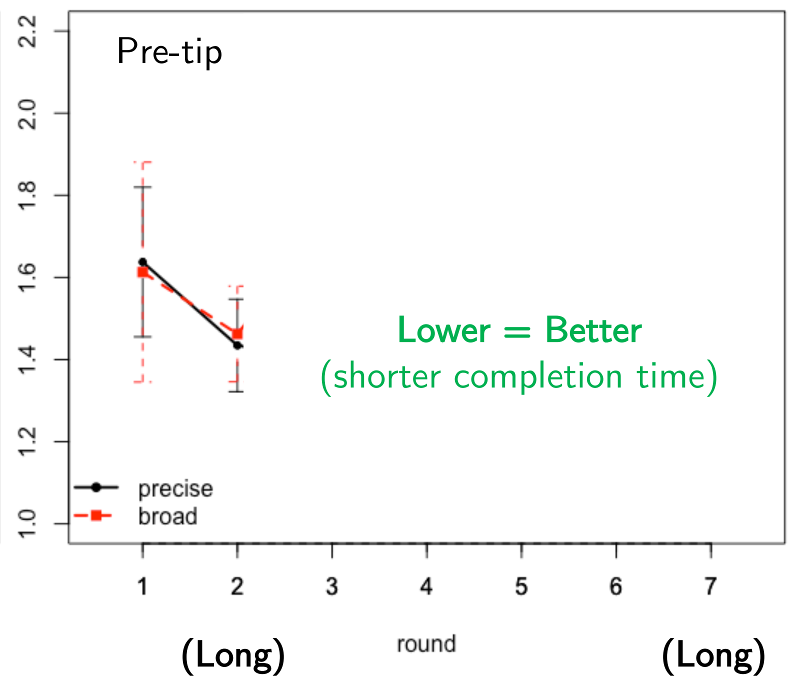

Amazon Mechanical Turk, N = 102, 3,978 decision points

# Study 1:
# Results

Completion Time – Centered Traffic



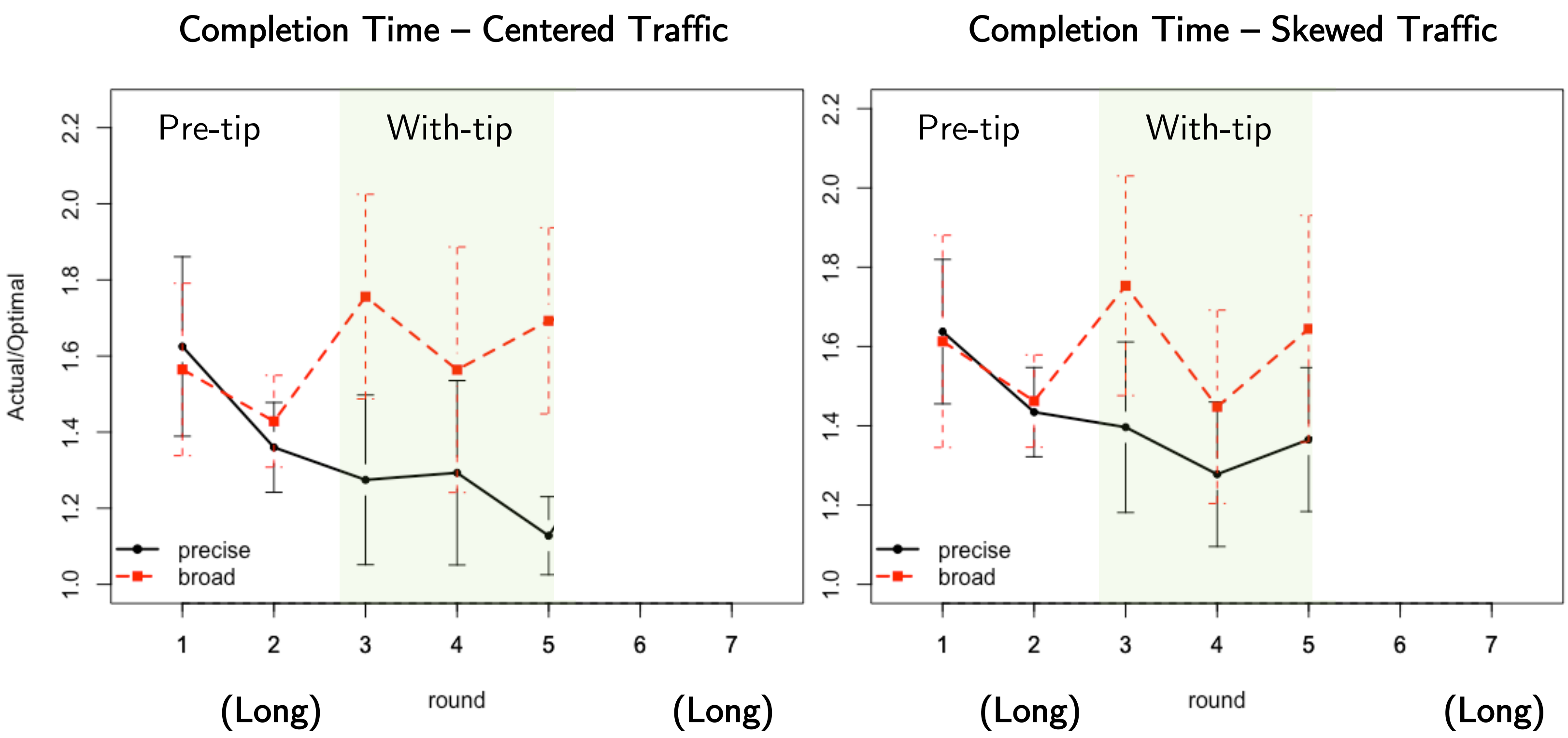

Completion Time – Skewed Traffic




Amazon Mechanical Turk, N = 102, 3,978 decision points

**Study 1:**

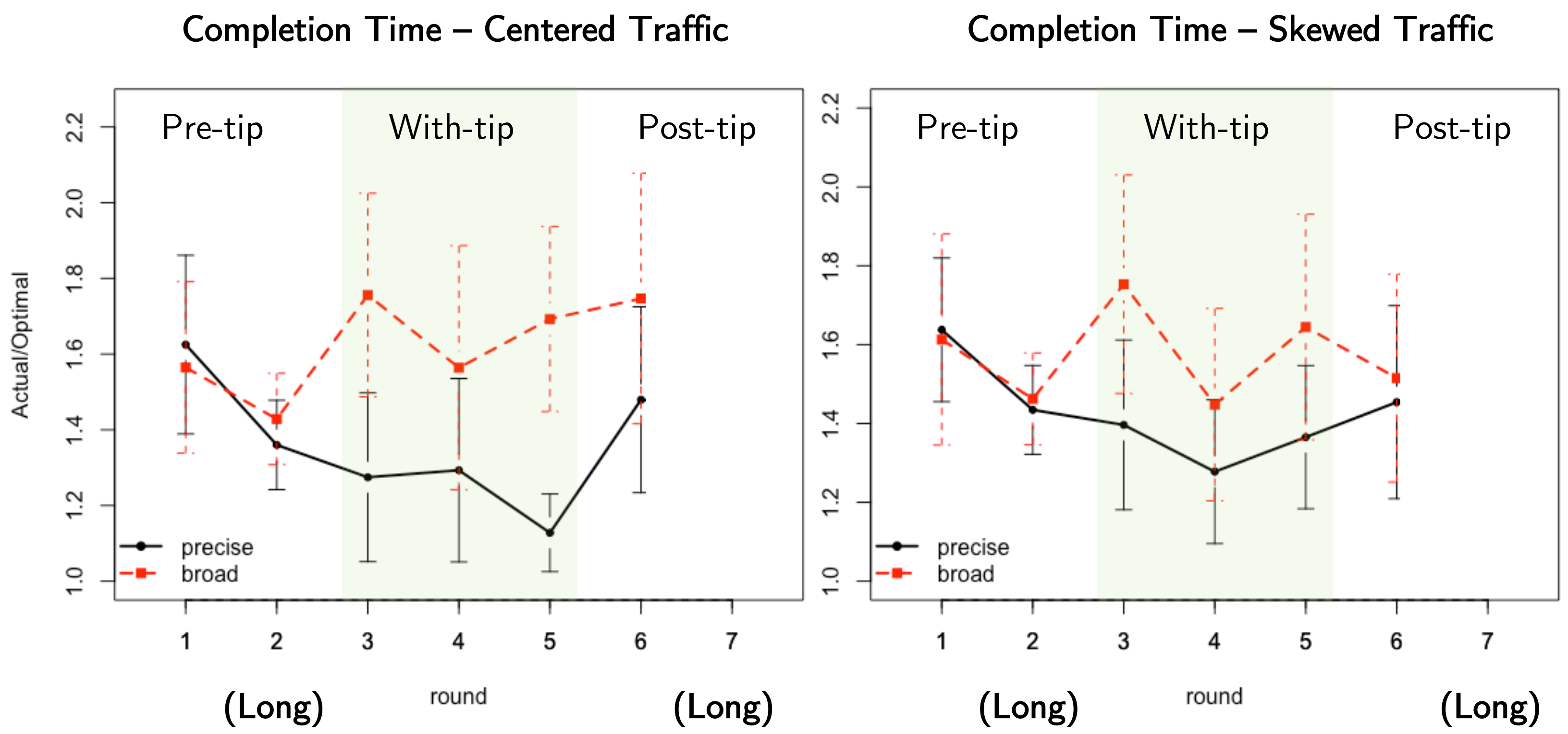
# Results Precise Tip Works Instantly




Amazon Mechanical Turk, N = 102, 3,978 decision points

# Study 1:

# Results Performance Dips Post-Tip

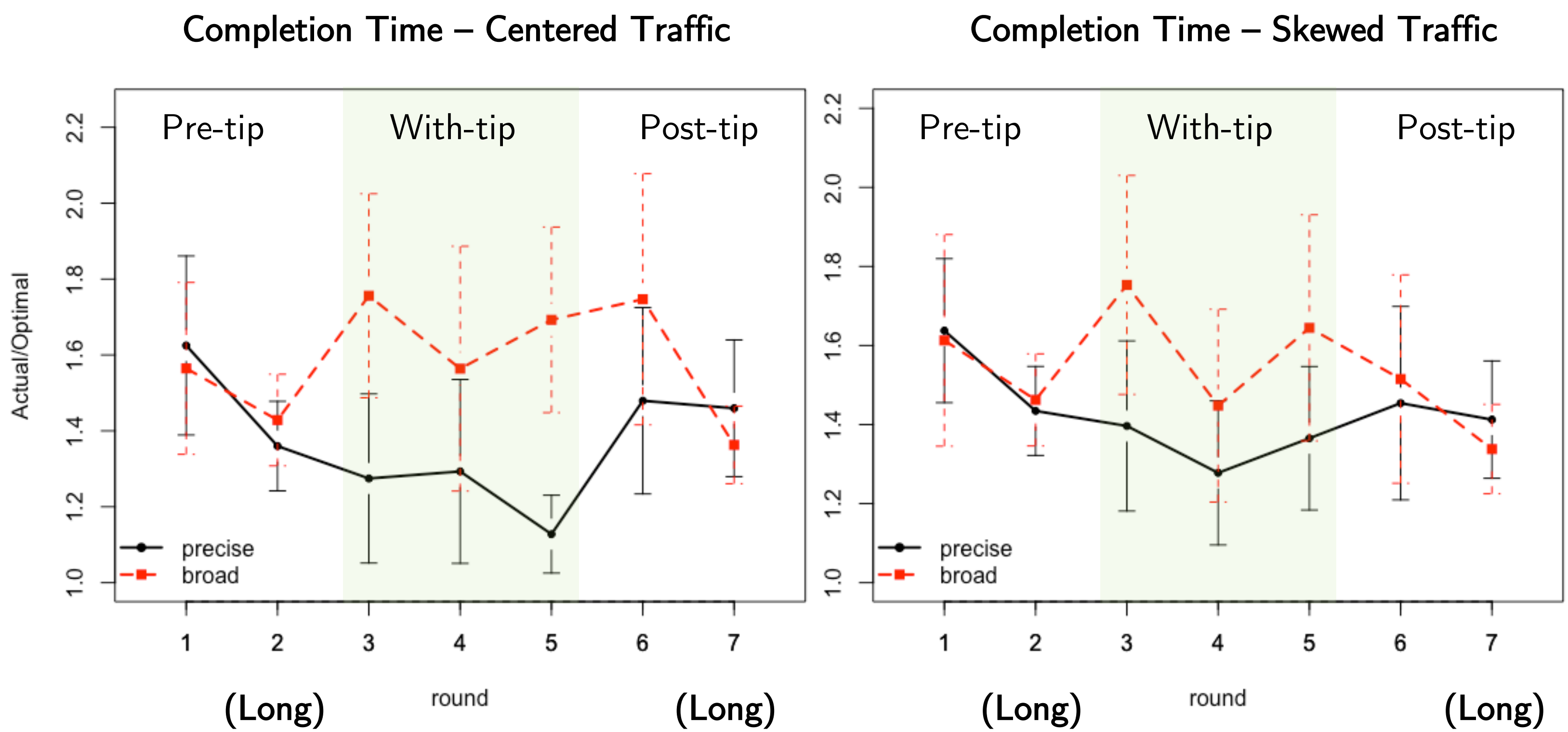



Amazon Mechanical Turk, N = 102, 3,978 decision points

# Study 1:

## Results

## Broad Tip Seems to Help with New Environment




Amazon Mechanical Turk, N = 102, 3,978 decision points

# Study 1: Results

## Broad Tip Seems to Help with New Environment

AI Meets the Classroom:
When Does ChatGPT Harm Learning?

Matthias Lehmann,[1] Philipp B. Cornelius,[2] Fabian J. Sting[1,2]
[1]University of Cologne, [2]Rotterdam School of Management, Erasmus University

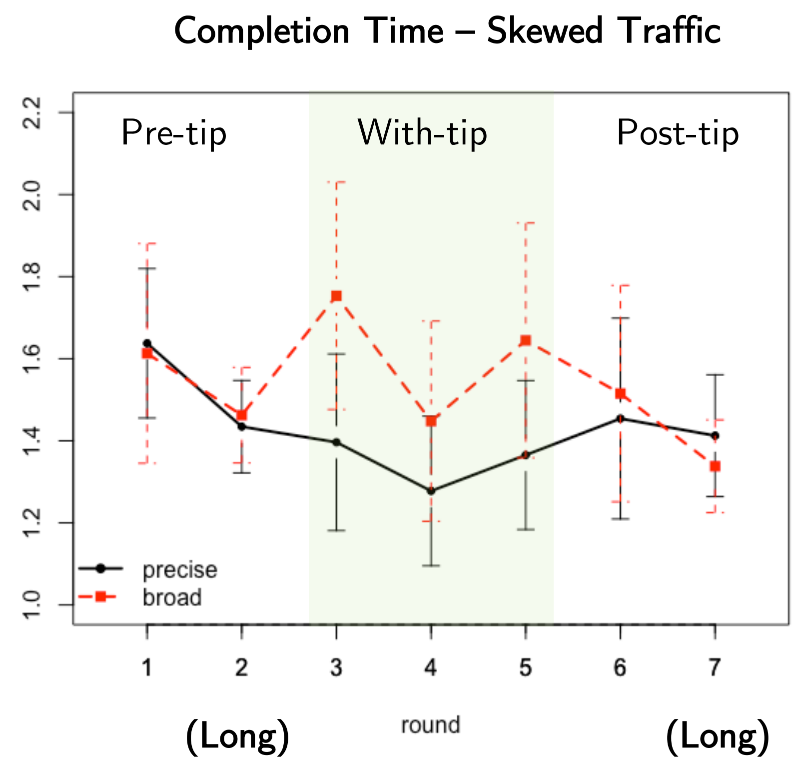
RESEARCH ARTICLE | ECONOMIC SCIENCES

**Generative AI without guardrails can harm learning: Evidence from high school mathematics**

Hamsa Bastani, Osbert Bastani, Alp Sungu, +2, and Rei Mariman Authors Info & Affiliations

Edited by Emma Brunskill, Stanford University, Stanford, CA; received November 3, 2024; accepted May 5, 2025 by Editorial Board Member Mark Granovetter

June 25, 2025 | 122 (26) e2422633122 | https://doi.org/10.1073/pnas.2422633122

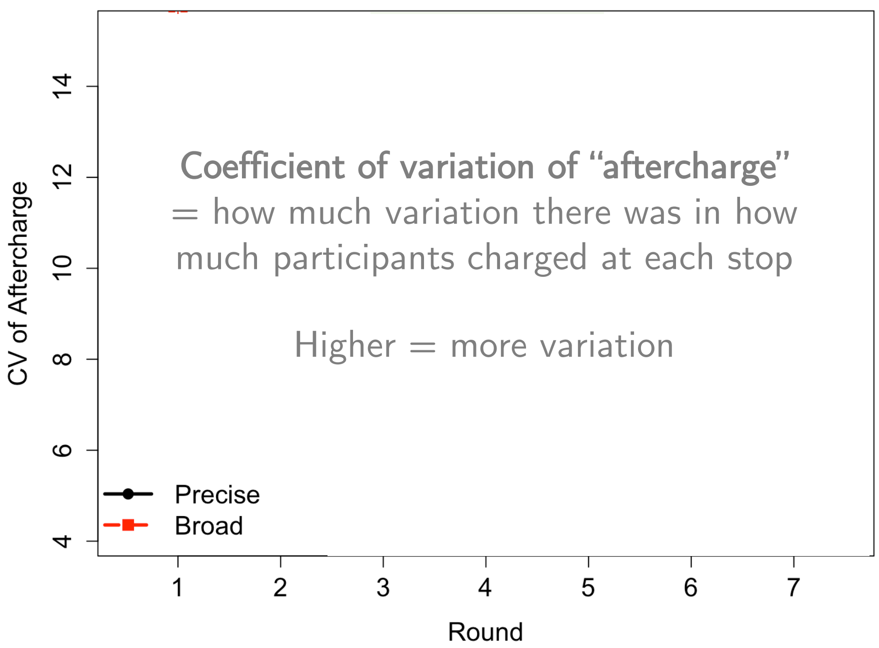



Amazon Mechanical Turk, N = 102, 3,978 decision points

**Study 1:**

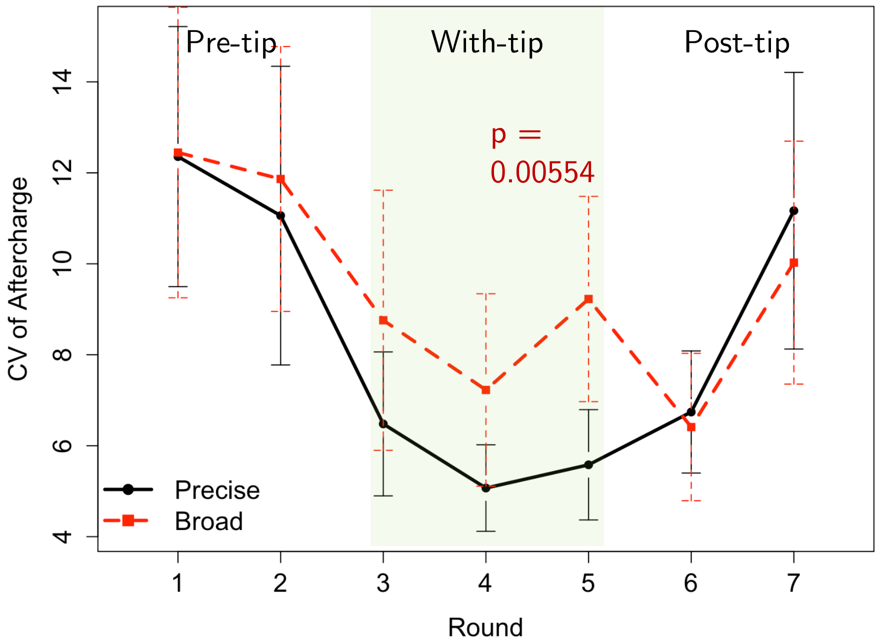
# Mechanism

## Strategy Exploration



# Study 1:

# Mechanism Strategy Exploration




Broad tip: Greater variation in charging decisions
→ continued experimentation and flexible adjustment

# Quantifying Human Strategy

- Suppose $\mu^*$ is the policy maximizing $V^\psi$ under the reward function $r_s^h$ of the **human decision-maker**. We assume that $r_{s_t}^h(a_t) = \sum_{j=1}^k \boldsymbol{\theta_j} \phi_j(s_t, a_t)$

various reward components

# Quantifying Human Strategy

- Suppose $\mu^*$ is the policy maximizing $V^\psi$ under the reward function $r_s^h$ of the **human decision-maker**. We assume that $r_{s_t}^h(a_t) = \sum_{j=1}^{k} \boldsymbol{\theta_j} \phi_j(s_t, a_t)$

various reward components

- Meanwhile, $\nu^*$ is the policy maximizing $V^\psi$ under the reward function $r_s^d$ intended by the designer (our tip).

# Quantifying Human Strategy

- Suppose $\mu^*$ is the policy maximizing $V^{\psi}$ under the reward function $r_s^h$ of the **human decision-maker**. We assume that $r_{s_t}^h(a_t) = \sum_{j=1}^{k} \boldsymbol{\theta}_j \phi_j(s_t, a_t)$

various reward components

- Meanwhile, $\nu^*$ is the policy maximizing $V^{\psi}$ under the reward function $r_s^d$ intended by the designer (our tip).

- Without tips, the decision-maker follows $\mu^*$.
- With tips, the decision-maker follows:

$$\psi^* = \begin{cases} \nu^*, & \text{with prob.}\, \pi \\ \mu^*, & \text{with prob.}\, 1 - \pi \end{cases}$$

Inverse Reinforcement Learning

# Quantifying Human Strategy

- Human reward function: $r^h_{s_t}(a_t) = \sum_{j=1}^{k} \theta_j \phi_j(s_t, a_t)$

Inverse Reinforcement Learning

# Quantifying Human Strategy

- Human reward function: $r_{s_t}^h(a_t) = \sum_{j=1}^{k} \theta_j \phi_j(s_t, a_t)$
- Based on topic modeling of participant's textual feedback, we derive several components of the reward function, e.g.,

Inverse Reinforcement Learning

# Quantifying Human Strategy

- Human reward function: $r_{s_t}^{h}(a_t) = \sum_{j=1}^{k} \theta_j \phi_j(s_t, a_t)$
- Based on topic modeling of participant's textual feedback, we derive several components of the reward function, e.g.,

  - $\phi_t =$ **Time taken at current stop** (including charging & emergency)

"Take the **total amount of time needed** to get to the next point after including worst traffic scenario…"

# Inverse Reinforcement Learning

# Quantifying Human Strategy

- Human reward function: $r_{s_t}^{h}(a_t) = \sum_{j=1}^{k} \theta_j \phi_j(s_t, a_t)$
- Based on topic modeling of participant's textual feedback, we derive several components of the reward function, e.g.,

  - $\phi_t =$ **Time taken at current stop** (including charging & emergency)

    "Take the **total amount of time needed** to get to the next point after including worst traffic scenario..."

  - $\phi_s =$ **Simplicity of action** (e.g., high for 0, max, following tip)

    "I first just tried to **make sure my charge was 100%.** I then tried to make sure I had just enough charge for the next stop."

Inverse Reinforcement Learning

# Quantifying Human Strategy

- Human reward function: $r^h_{s_t}(a_t) = \sum_{j=1}^{k} \theta_j \phi_j(s_t, a_t)$
- Based on topic modeling of participant's textual feedback, we derive several components of the reward function, e.g.,

  - $\phi_t$ = **Time taken at current stop** (including charging & emergency)

    "Take the **total amount of time needed** to get to the next point after including worst traffic scenario…"

  - $\phi_s$ = **Simplicity of action** (e.g., high for 0, max, following tip)

    "I first just tried to **make sure my charge was 100%.** I then tried to make sure I had just enough charge for the next stop."

  - $\phi_r$ = **Risk exposure** (decreasing with likelihood of reaching next stop)

    "Play it safe!"

Inverse Reinforcement Learning

# Quantifying Human Strategy

- Human reward function: $r_{s_t}^{h}(a_t) = \sum_{j=1}^{k} \theta_j \phi_j(s_t, a_t)$
- Based on topic modeling of participant's textual feedback, we derive several components of the reward function, e.g.,

  - $\phi_t$ = **Time taken at current stop** (including charging & emergency)

    "Take the **total amount of time needed** to get to the next point after including worst traffic scenario…"

  - $\phi_s$ = **Simplicity of action** (e.g., high for 0, max, following tip)

    "I first just tried to **make sure my charge was 100%.** I then tried to make sure I had just enough charge for the next stop."

  - $\phi_r$ = **Risk exposure** (decreasing with likelihood of reaching next stop)

    "Play it safe!"

  - $\phi_b$ = **Batch** (charge sufficiently for multiple stops)

    "I would probably charge all the way on stop … so you don't have to charge on the last stop"

Inverse Reinforcement Learning

# Estimating Human Strategy

Challenge: Each subject has their own weights and compliance probabilities, but we only observe a few rounds per subject/condition.

Inverse Reinforcement Learning

# Estimating Human Strategy

Challenge: Each subject has their own weights and compliance probabilities, but we only observe a few rounds per subject/condition.

**The proposed model:**

$$\boldsymbol{\theta}_i = \boldsymbol{\theta}_0 + \underbrace{\boldsymbol{\Delta}_s}_{\text{scenario-specific shift (e.g., pre/with/post, precise vs broad)}} + \underbrace{\boldsymbol{\Delta}_i}_{\text{individual's shift}}$$

$$\pi_i = \frac{1}{1 + e^{-(\eta_s + \eta_i)}} \quad \text{(or 0 when } s \text{ is not a tip-scenario)}$$

Inverse Reinforcement Learning

# Estimating Human Strategy

Challenge: Each subject has their own weights and compliance probabilities, but we only observe a few rounds per subject/condition.

**The proposed model:**

scenario-specific shift (e.g., pre/with/post, precise vs broad)

individual's shift

$$\boldsymbol{\theta}_i = \boldsymbol{\theta}_0 + \boldsymbol{\Delta}_s + \boldsymbol{\Delta}_i$$

$$\pi_i = \frac{1}{1 + e^{-(\eta_s + \eta_i)}}$$ (or 0 when $s$ is not a tip-scenario)

Mean of the scenario shift = average for a given (tip)-scenario, across subjects and even experiments.

Inverse Reinforcement Learning

# Estimating Human Strategy

Challenge: Each subject has their own weights and compliance probabilities, but we only observe a few rounds per subject/condition.

**The proposed model:**

scenario-specific shift (e.g., pre/with/post, precise vs broad)

individual's shift

$$\boldsymbol{\theta}_i = \boldsymbol{\theta}_0 + \boldsymbol{\Delta}_s + \boldsymbol{\Delta}_i$$

$$\pi_i = \frac{1}{1 + e^{-(\eta_s + \eta_i)}}$$ (or 0 when $s$ is not a tip-scenario)

Mean of the scenario shift = average for a given (tip)-scenario, across subjects and even experiments.

**Estimation:** We use Stochastic Variational Inference (a Bayesian approach)

This enables us to efficiently estimate posterior distributions, rather than point estimates only

# How Tips Affect Strategy

1. We estimate $\Delta_s$, where $s \in$ {pre, with(*type*), post(*type*)}

# How Tips Affect Strategy

1. We estimate $\Delta_s$, where $s \in$\{pre, with(*type*), post(*type*)\}

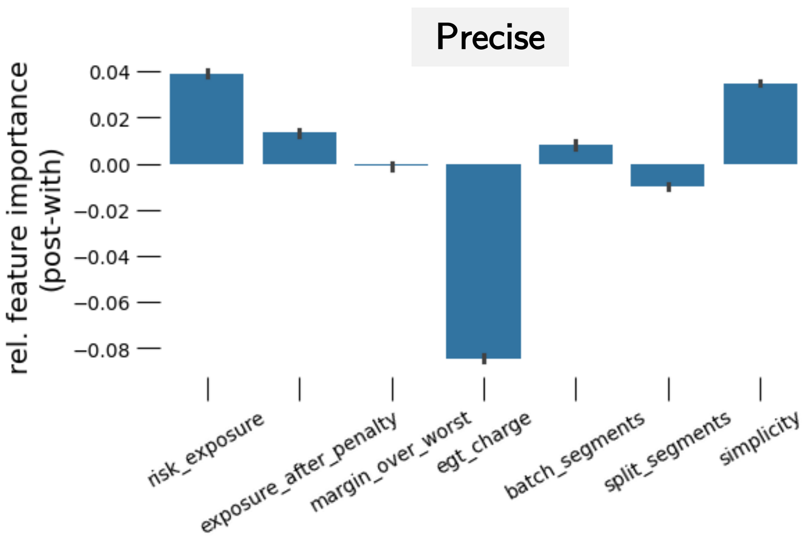
2. We can then characterize the impact of tips by looking at how the vector changes, e.g., how $\Delta_{with(broad)} \rightarrow \Delta_{post(broad)}$ or $\Delta_{with(precise)} \rightarrow \Delta_{post(precise)}$

= Shifts in reward function's weights
after learning with tips
and now the tips are gone.

# How Tips Affect Strategy

1. We estimate $\Delta_s$, where $s \in${pre, with(*type*), post(*type*)}

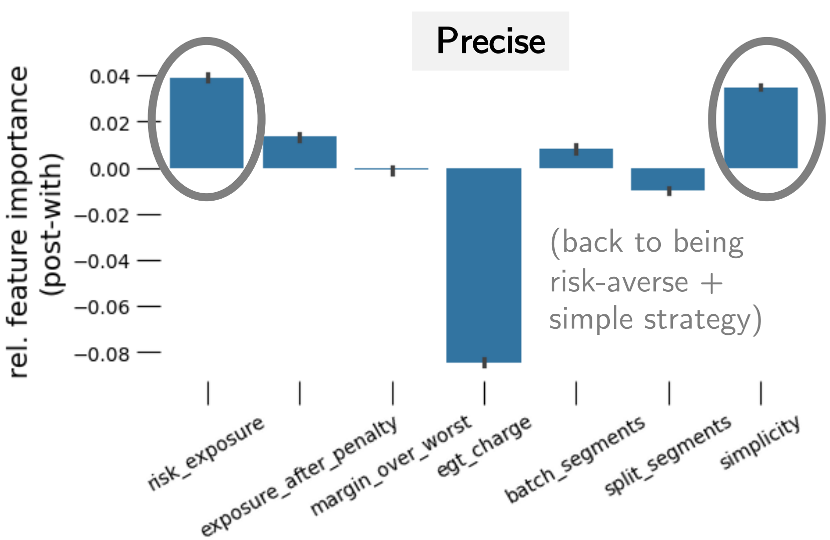
2. We can then characterize the impact of tips by looking at how the vector changes, e.g., how $\Delta_{with(broad)} \rightarrow \Delta_{post(broad)}$ or $\Delta_{with(precise)} \rightarrow \Delta_{post(precise)}$



# How Tips Affect Strategy

1. We estimate $\varDelta_s$, where $s \in$ {pre, with(*type*), post(*type*)}

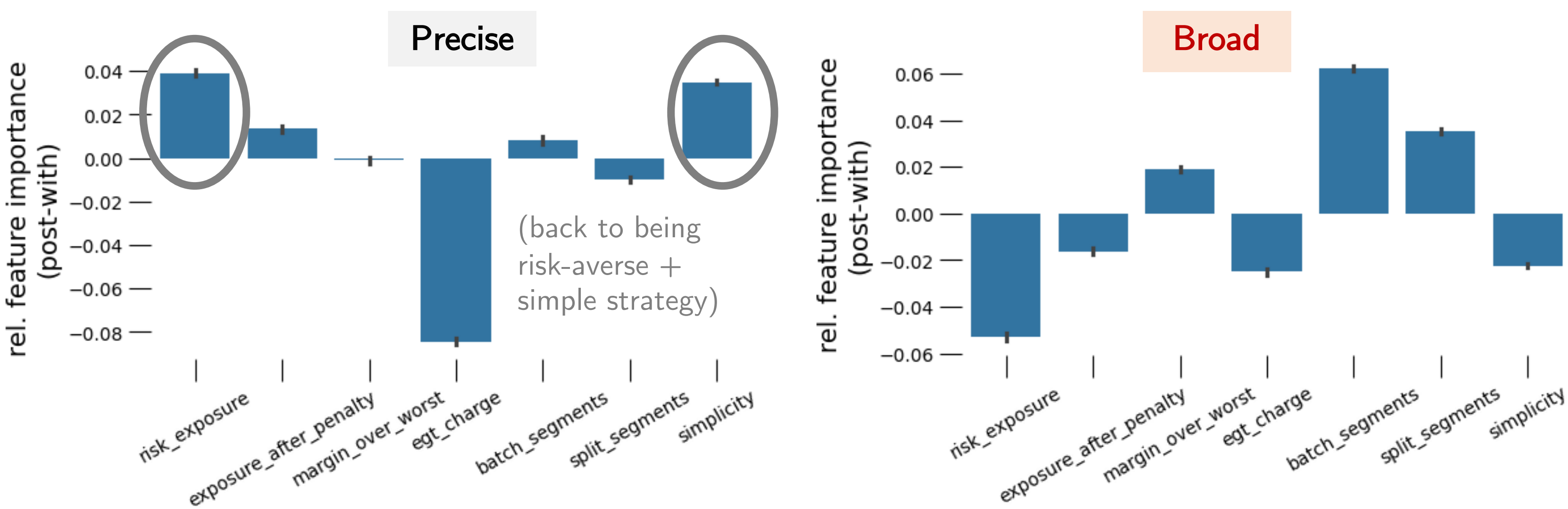
2. We can then characterize the impact of tips by looking at how the vector changes, e.g., how $\varDelta_{with(broad)} \rightarrow \varDelta_{post(broad)}$ or $\varDelta_{with(precise)} \rightarrow \varDelta_{post(precise)}$



# How Tips Affect Strategy

1. We estimate $\Delta_s$, where $s \in${pre, with(*type*), post(*type*)}

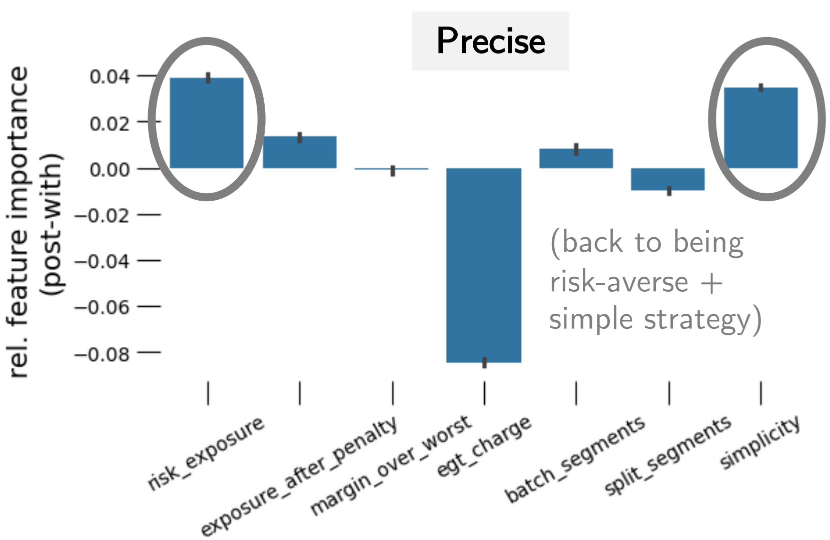
2. We can then characterize the impact of tips by looking at how the vector changes, e.g., how $\Delta_{with(broad)} \rightarrow \Delta_{post(broad)}$ or $\Delta_{with(precise)} \rightarrow \Delta_{post(precise)}$

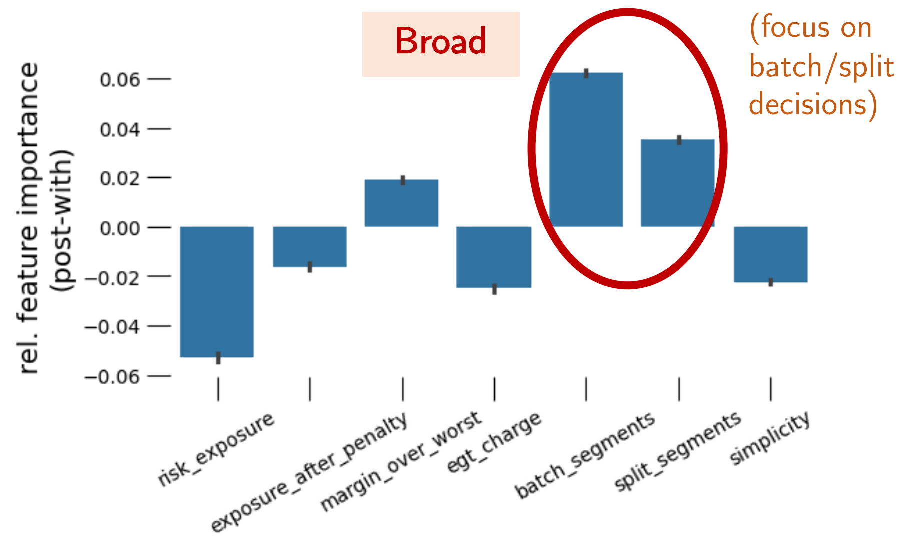


# How Tips Affect Strategy

1. We estimate $\varDelta_s$, where $s$ ∈{pre, with(*type*), post(*type*)}

2. We can then characterize the impact of tips by looking at how the vector changes, e.g., how $\varDelta_{with(broad)} \rightarrow \varDelta_{post(broad)}$ or $\varDelta_{with(precise)} \rightarrow \varDelta_{post(precise)}$

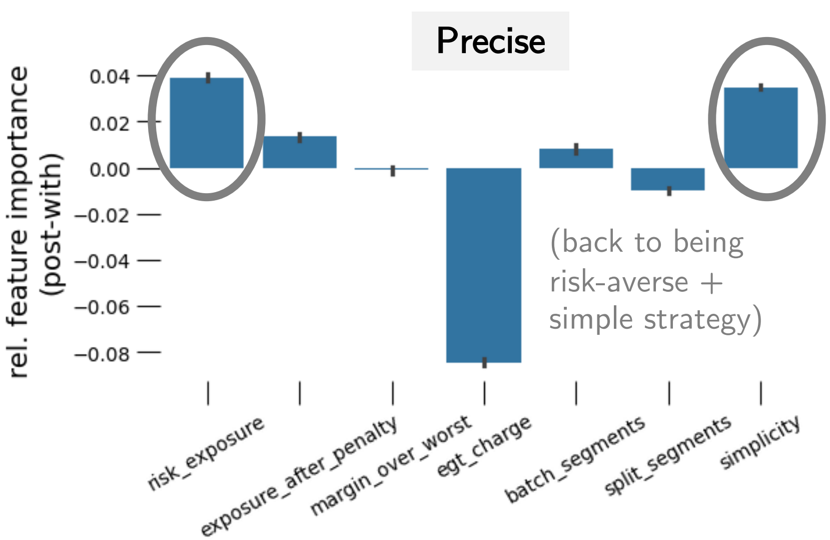


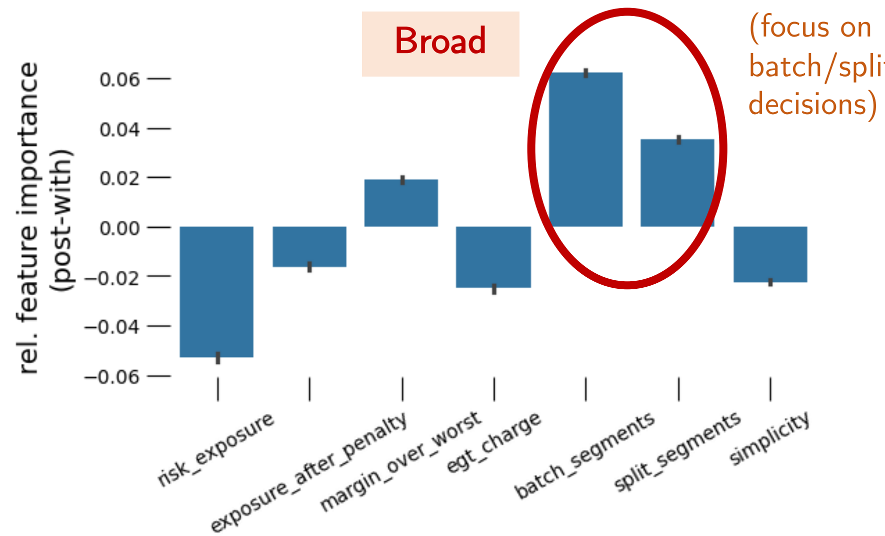



# How Tips Affect Strategy

1. We estimate $\Delta_s$, where $s \in$ {pre, with(*type*), post(*type*)}

2. We can then characterize the impact of tips by looking at how the vector changes, e.g., how $\Delta_{with(broad)} \rightarrow \Delta_{post(broad)}$ or $\Delta_{with(precise)} \rightarrow \Delta_{post(precise)}$




Broad tip nudges humans to internalize the nonlinearity of charging costs: helping them move beyond default preferences for simplicity and safety, and toward more reward-optimal strategies.

Pre-registered at https://aspredicted.org/23C_WBQ



(Super Brief)

# Study 2

| | Pre-advice | | With-advice | | Post-advice | | |
|---|---|---|---|---|---|---|---|
| Sequence A | Round 1 Short | Round 2 Short | Round 3 Short (+advice) | Round 4 Short (+advice) | Round 5 Short | Round 6 Short II | Round 7 Short II |
| Sequence B | Round 1 Short | Round 2 Short | Round 3 Short (+advice) | Round 4 Short (+advice) | Round 5 Long | Round 6 Long | Round 7 Long |

Experimental conditions

**2** map sequences x **3** advice precision x **2** rationale?

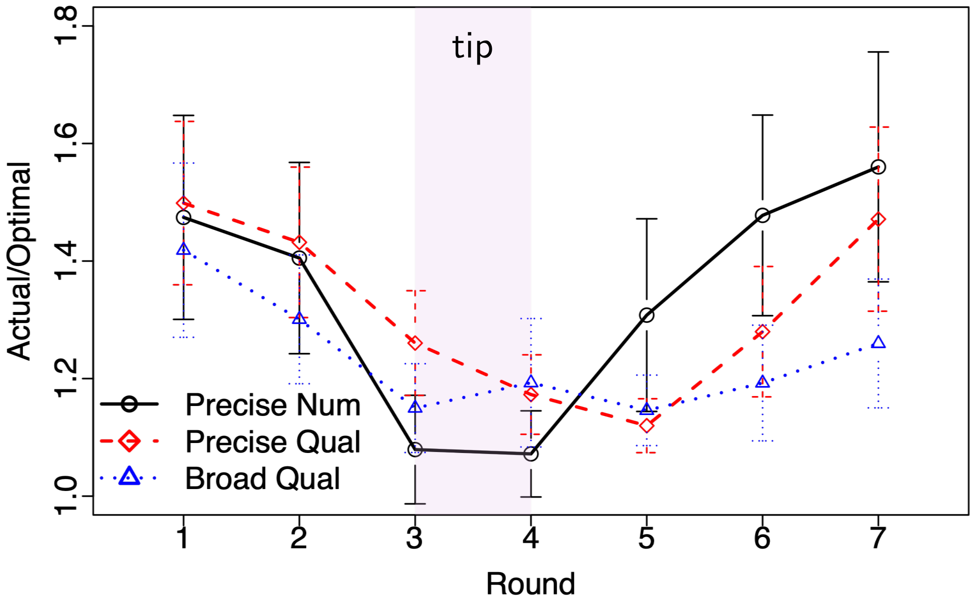
+ **Mid**
= Broad + "assuming worst case traffic")

**Rationale:**
"Looking ahead, if you need to charge but the sum of charges required for next segments <50%, charging is fast, so you should charge enough for these segments in one stop."

**Study 2:**

# Results Broad's Success Replicated! 😊

😊 Familiar new map, no rationale

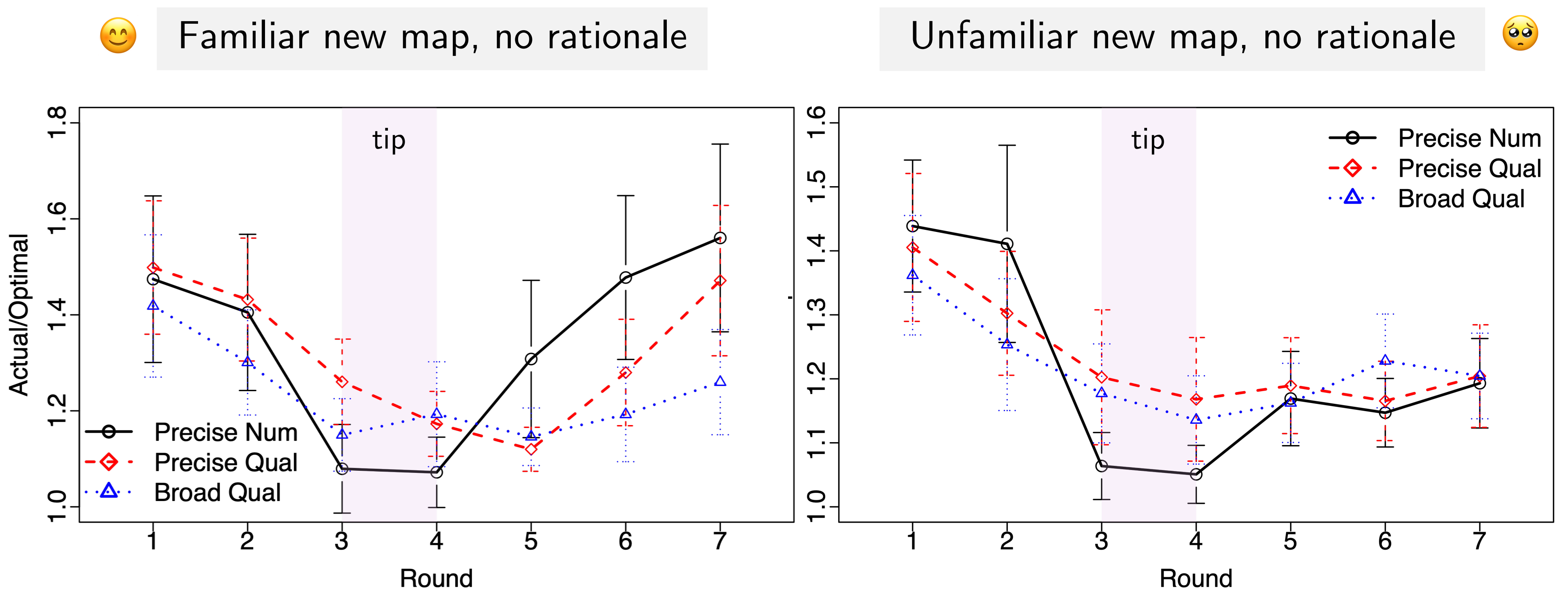



Prolific, N = 400

# Study 2:

## Results

Broad's Success Replicated! 😊
…Not When Things Change A Lot 🥺

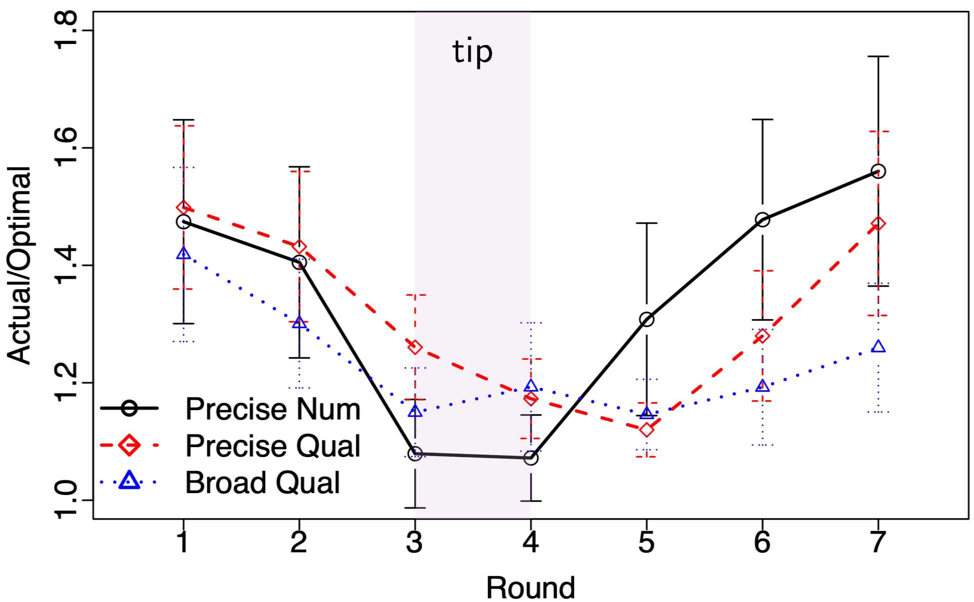



Prolific, N = 400

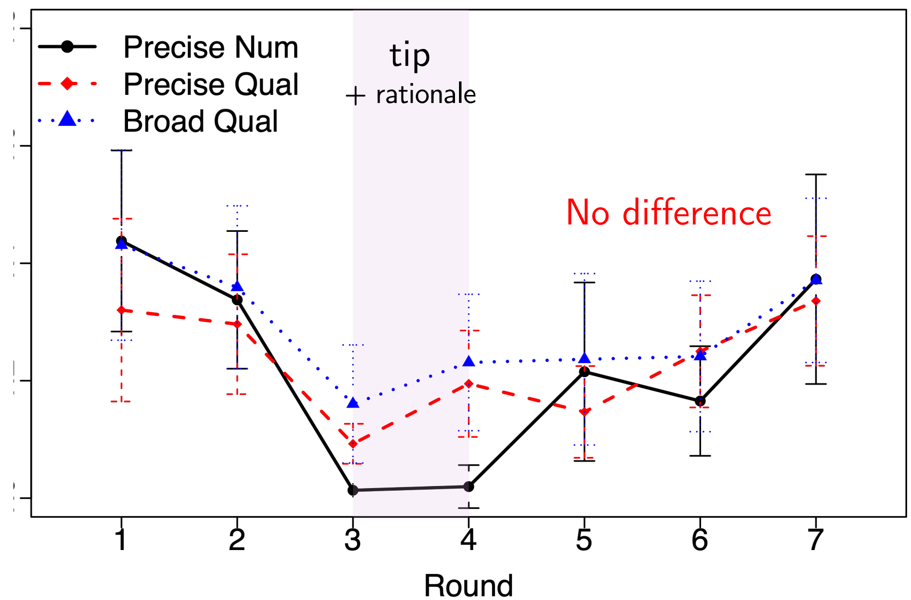
# Study 2:
# Results

**Rationale:**
"Looking ahead, if you need to charge but the sum of charges required for next segments <50%, charging is fast, so you should charge enough for these segments in one stop."

Familiar new map, no rationale




Familiar new map + rationale




Prolific, N = 400

# Summary

**Goal:** How to Design/Deliver AI Tips to Support Long-Term Learning?

AI Tips

New environment (no tips)

"Driving Game"

How Humans Make Sequential Decisions? → How Do Tips Shape Learning & Behavior Over Time? → Can Humans Adapt When Environment Changes?

# Summary

**Goal:** How to Design/Deliver AI Tips to Support Long-Term Learning?

AI Tips

Precise tips improve short-run efficiency, but without explanation, they can limit learning and adaptability.

Broad tips promote strategic exploration and long-term learning, but only when users can infer the rationale themselves

"Driving Game"

New environment (no tips)

How Humans Make Sequential Decisions? → How Do Tips Shape Learning & Behavior Over Time? → Can Humans Adapt When Environment Changes?

# Summary

**Inverse RL:** We recover each participant's hidden reward weights.
- With precise, people drift back to simple, safe heuristics.
- Broad permanently increase the weight on sophisticated strategies.

**Goal:** How to Design/Deliver AI Tips to Support Long-Term Learning?

AI Tips

Precise tips improve short-run efficiency, but without explanation, they can limit learning and adaptability.

Broad tips promote strategic exploration and long-term learning, but only when users can infer the rationale themselves

"Driving Game"

How Humans Make Sequential Decisions? → How Do Tips Shape Learning & Behavior Over Time? → Can Humans Adapt When Environment Changes?

New environment (no tips)

# Summary + Thank You!

**Inverse RL:** We recover each participant's hidden reward weights.
- With precise, people drift back to simple, safe heuristics.
- Broad permanently increase the weight on sophisticated strategies.

**Goal:** How to Design/Deliver AI Tips to Support Long-Term Learning?

AI Tips

Precise tips improve short-run efficiency, but without explanation, they can limit learning and adaptability.

Broad tips promote strategic exploration and long-term learning, but only when users can infer the rationale themselves

New environment (no tips)

"Driving Game"

How Humans Make Sequential Decisions? → How Do Tips Shape Learning & Behavior Over Time? → Can Humans Adapt When Environment Changes?

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

**Philippe Blaettchen**
Singapore Management University

**Park Sinchaisri**
Berkeley Haas
parksinchaisri@berkeley.edu

# Mechanism

# Interlude: Validating the Approach

1. Synthetic data:
   - We **randomly generate** model **parameters**
   - Then, in the **existing trajectories**, we **replace the actions** taken by actions drawn from the policy suggested by our model
   - This shows that our estimation procedure can **identify parameters accurately**

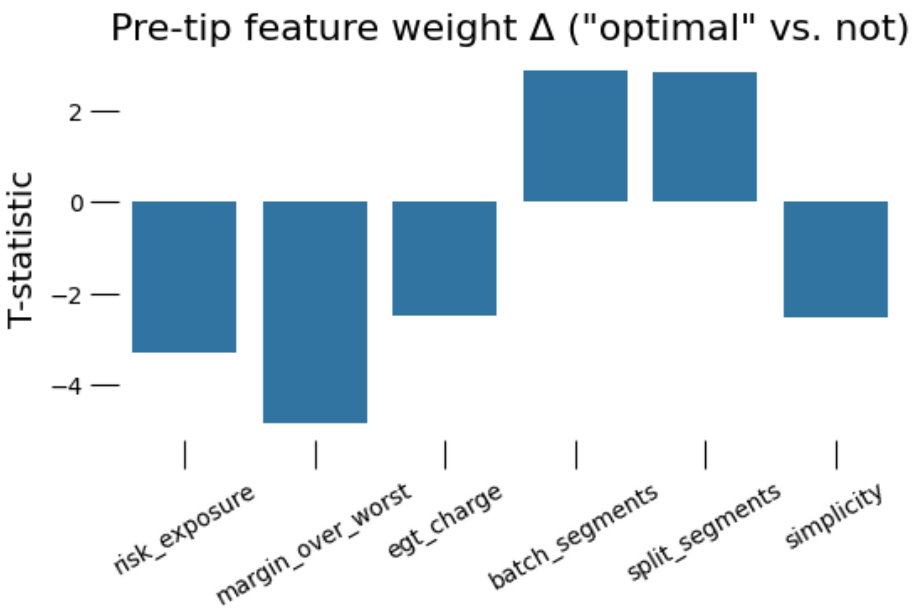
| | $corr(\Delta_s, \widehat{\Delta}_s)$ | $corr(\Delta_i, \widehat{\Delta}_i)$ | $corr(\pi_s, \pi_s)$ | $corr(\pi_i, \pi_i)$ |
|---|---|---|---|---|
| Avg. | 0.96 | 0.88 | 0.99 | 0.94 |
| Std. | 0.03 | 0.06 | / | / |

2. Real data, standard checks:
   - **Log-likelihoods**: Test set LL on real data (-13.41) in line with training LL on real data (-13.03) and test set LL on synthetic data (-11.81)
   - **Posterior-predictive checks**: Actual action is "as far" from estimated probabilities as action generated from model (Brier score of observations = 36.60, Avg. Brier score of simulations = 38.35, p = 0.93)

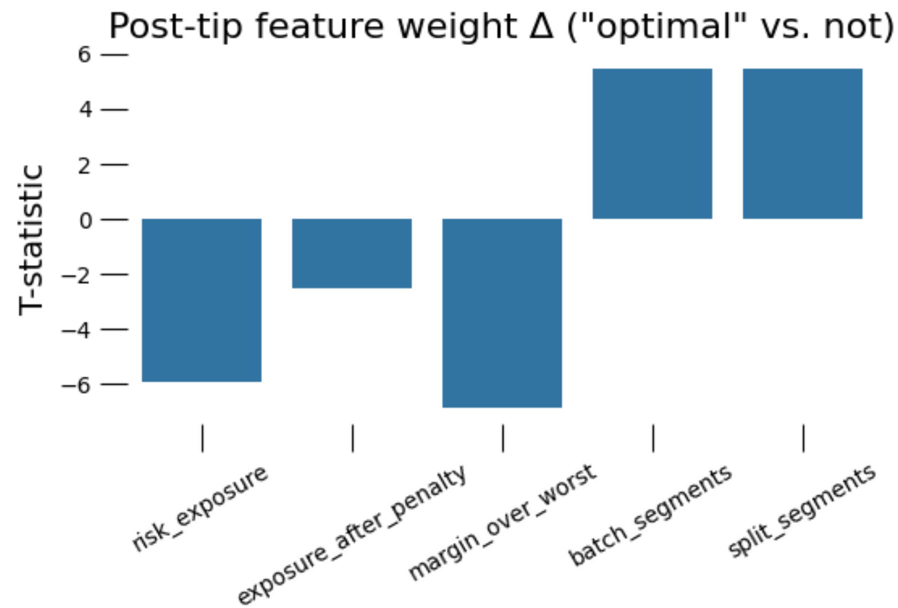
# Mechanism

# Interlude: Validating the Approach

3. Real data, consistency with qualitative insights:
   - We estimate our model on the actual data, then **compare the participant-specific shifts** based on the qualitatively assigned groups (sequence clusters)
   - We compare the individuals in the **"Optimal"/"learning" clusters to others**




Only showing results with p<0.05




Only showing results with p<0.05

**Compliance** (logit) of "Optimal" vs. other clusters: T-stat=11.47 (p-value < 0.01)

# Study 2:

# Mechanism How Tips Affect Strategy

Recall that in our IRL approach, we use a scenario-specific shift: $\boldsymbol{\Delta}_s$, respectively $\eta_s$

Here, we adjust $\boldsymbol{\Delta}_s$ and $\eta_s$, to take into consideration explanations: $s \in${pre, with(*type,reveal*), post(*type,reveal*)}:

- Broad+reveal leads to similar compliance as precise
- Feature-changes (with-tip → post-tip) are "in-between" what we observe from broad alone and precise alone

→ Useful in capturing the **immediate benefits of precise tips** and some of the **long-term benefits of broad tips**