Strategic Route Choice in Navigation Apps: Testing for Level-K Thinking

Jared Joseph Donohue

Columbia University
New York, NY, USA
jjd2203@columbia.edu

Abstract

Navigation apps like Google Maps guide millions of daily driving decisions by recommending the fastest route, but these suggestions can create congestion if too many users follow them simultaneously. This study investigates whether users anticipate such effects by strategically deviating from the recommended route, in line with Level-K reasoning models. Using a lab-based randomized controlled trial with Columbia students, we estimate the distribution of Level-0, Level-1, and Level-2 reasoning in route choice and examine whether visibly adaptive algorithms promote higher-order strategic thinking. We also analyze what other stated contextual factors increase strategic responses. Findings can be used to improve the design of routing platforms to account for both human-human and human-algorithm strategic interactions, with the goal of reducing app-induced traffic congestion.

Table of Contents

Introduction

Navigation app users have noticed that suggestions labeled as "faster" can fail to deliver actual time savings. This disconnect raises a broader question: how do users interpret routing recommendations, especially when others may follow the same guidance?

> *"Even when typical faster routes are offered (e.g., 4 minutes faster), when I choose to take those routes my ETA doesn't change, indicating the route wasn't actually correct."*
>
> — Google Maps user complaint [1]

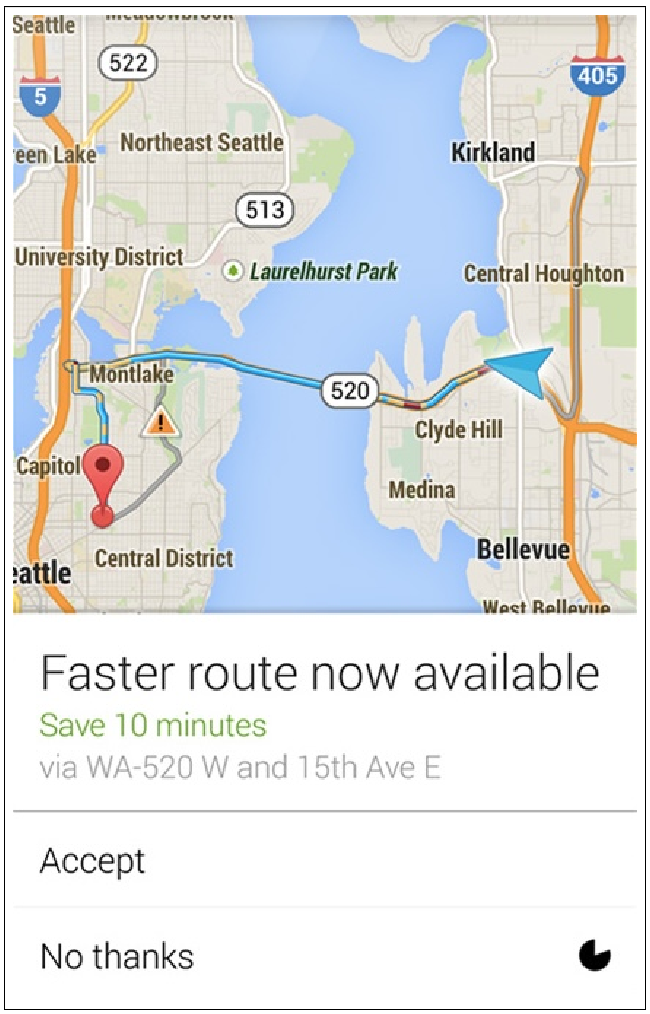

Figure 1: Google Maps dynamic re-routing interface [2]

Literature Review

Strategic User Behavior on Navigation Apps

Navigation apps such as Google Maps and Waze are now used by a majority of drivers, with surveys reporting that roughly 72% of users follow app-suggested routes almost all the time [3]. This widespread reliance can create herding effects, where many drivers converge on the same path, causing unexpected congestion in neighborhoods and side streets [4]. Community pushback has included attempts to block apps or introduce fake traffic reports to deter rerouting. These cases underscore that app influence extends beyond individual convenience to broader urban traffic dynamics.

Guin (2021) highlights that while smartphone applications often improve individual travel efficiency, they can unintentionally increase congestion by guiding multiple drivers onto the same routes [3]. This coordination failure poses a major challenge in shared routing environments.

Additionally, Mayr (2023) shows that the way apps frame routing information can meaningfully influence user decisions, indicating that some drivers anticipate others' behavior and adjust accordingly [5]. This aligns with behavioral game-theoretic models such as Level-K reasoning, where users vary in how deeply they consider others' likely actions.

Jiang (2014) applies evolutionary game theory to similar routing problems and shows that drivers adapt based on congestion patterns over time [6]. Furthermore, Vosough (2024) finds that some users are willing to accept longer personal travel times for social benefit when presented with clear incentives or messaging about shared outcomes [7]. These findings suggest that user behavior is heterogeneous, and that strategic deviation from app guidance is both possible and context-dependent.

Evidence of such behavior also appears in observational and experimental work. Razo and Gao (2013) find that some drivers plan routes by anticipating future congestion from others' choices [8]. Similarly, Ben-Elia and Shiftan (2010) show that repeated exposure to traffic outcomes encourages users to adapt their behavior over time [9]. These studies highlight that app users do not always follow guidance blindly, and that strategic learning can emerge in practice. Kroller (2021) further supports this view by documenting how users form expectations about the effects of app-based congestion and adjust their behavior accordingly [10].

In summary, prior work confirms that routing apps shape behavior at scale, but user responses are heterogeneous and sensitive to framing and prior experience. Some users follow the app without question, while others act strategically based on beliefs about traffic and the decisions of fellow drivers. Level-K reasoning provides a structured way to interpret these differences, with limited empirical evidence for the distribution of Level-0, Level-1, and Level-2 users in route selection behavior.

How do Routing Algorithms Work?

Routing algorithms used in platforms like Google Maps and Waze rely on real-time and historical traffic data to recommend the route with the lowest expected travel time. These systems continuously update suggestions using location data from millions of devices, combining predictive analytics with live feedback on congestion levels [11].

At its foundation, Google Maps models the road network as a mathematical graph, where intersections are nodes and roads are edges connecting them. To calculate the optimal path through this network, it relies on two powerful algorithms: Dijkstra's Algorithm and the A* Algorithm. Dijkstra's algorithm systematically explores the lowest-cost paths to find the shortest route from a starting point to all other nodes, while A* improves efficiency by using a heuristic to prioritize paths likely to reach the destination faster. These algorithms, combined with real-time traffic data, power features like Google Maps' "faster route available," which dynamically reroutes users as conditions change during their journey [12].

As navigation technology continues to evolve, we can expect Google Maps to further refine its routing algorithms and threshold criteria, potentially incorporating more sophisticated user behavior models like Level-K distributions to enhance the accuracy and usefulness of its "faster route available" suggestions.

Theoretical Framework

Level-K Thinking

Level-K theory says that in strategic interactions, people reason in hierarchical steps or "levels," with each level anticipating and responding to the predicted strategies of lower-level thinkers. A Level-0 player chooses actions in a simple, non-strategic way, often by following defaults or salient cues. A Level-1 player best responds to the behavior of a Level-0 player. A Level-2 player best responds to a Level-1 player, and so on. Each level represents one additional step of reasoning about others' actions.

In the context of navigation apps:

- **Level-0** users follow the app's recommended "fastest" route without considering how others might behave.
- **Level-1** users expect that many others will follow the same recommendation, leading to congestion, and choose an alternative route.
- **Level-2** users anticipate that Level-1 users will avoid the recommended route, so they return to it, expecting it to be relatively clear.

Traditional Level-K models are typically applied to games with only human players. They do not account for adaptive systems that respond to behavior in real time. Navigation apps, however, continuously update their recommendations based on user movements and traffic conditions. This creates a feedback loop between user decisions and algorithmic updates.

To capture this dynamic, we model the navigation algorithm as a strategic actor. The algorithm is assumed to be fully rational and capable of predicting the aggregate behavior of users based on the distribution of Level-K types. Although it does not observe any user's type directly, it adjusts its guidance to optimize expected outcomes given how users are likely to respond.

By explicitly incorporating the algorithm as a player in the strategic environment, this paper extends the Level-K framework to include human-algorithm interaction. This approach enables us to study how real-time route recommendations and user reasoning co-evolve. It also highlights new forms of strategic complexity that emerge when decisions are shaped by both other users and an adaptive system.

We hypothesize that visibly changing travel times in the dynamic condition act as a behavioral cue that prompts participants to question the stability of the recommendation, thereby nudging them to reason more deeply about others' choices and potential congestion. This mechanism is expected to elevate participants from Level-0 to Level-1 or Level-2 reasoning.

Nash Equilibrium

We define the Nash equilibrium of the route choice game as the outcome in which no user can reduce their travel time by unilaterally switching routes. Suppose there are two available routes, R_1 and R_2, each with a fixed base travel time and a capacity constraint. A total of N users each choose one route. Travel time on each route increases with congestion, modeled as a penalty that grows linearly with the load-to-capacity ratio.

Let the travel time on route R_i be defined as:

$$T_i(n_i) = t_i + \alpha \cdot \frac{n_i}{C_i}$$

where:

- t_i is the base travel time on route R_i,
- n_i is the number of users on route R_i,
- C_i is the capacity of route R_i,
- α is a congestion sensitivity parameter.

A Nash equilibrium occurs when no individual driver can improve their travel time by switching routes, assuming all other drivers' choices remain fixed. If both routes are used in equilibrium, then travel times must be equal:

$$T_1(n_1^*) = T_2(n_2^*) \quad \text{and} \quad n_1^* + n_2^* = N$$

This equilibrium model provides a theoretical benchmark against which we evaluate participant behavior in our experiment. If users are fully rational and possess complete information, we would expect their choices to converge to the equilibrium distribution. In our experimental setup with fixed values for t_1, t_2, C_1, and C_2, the equilibrium predicts a specific allocation (n_1^*, n_2^*) across the two routes.

To quantify deviations from rational coordination, we define a "Distance from Equilibrium" metric:

$$D = |n_1^{\text{observed}} - n_1^*| + |n_2^{\text{observed}} - n_2^*|$$

This metric captures how far observed behavior deviates from the Nash equilibrium benchmark, which represents fully rational coordination. We compute this distance for each round and compare averages across experimental groups to test whether adaptive routing algorithms foster faster convergence to efficient distributions.

These deviations serve as a quantitative foundation for our analysis (Table 4), where we examine how closely each group's behavior aligns with equilibrium predictions. Systematic gaps offer evidence of bounded strategic reasoning consistent with Level-K thinking. For example, persistent overuse of a single route despite congestion may reflect limited depth of reasoning or failure to anticipate others' behavior.

Experimental Design

We will implement a lab-based randomized controlled trial (RCT) to learn if participants exposed to adaptive routing algorithms demonstrate higher average reasoning levels (Level-1 or Level-2) compared to those interacting with static algorithms. Participants will be randomly assigned to one of two groups that vary the responsiveness of the routing algorithm interface.

Experiment Group	**Description**
Group 1: Static Algorithm (Baseline)	Route recommendations do not change in response to simulated user behavior. Used to estimate the baseline distribution of Level-K reasoning under stable conditions.
Group 2: Dynamic Algorithm (Treatment)	Route time projections adapt in real time to simulate user choices, allowing us to test whether visible algorithm responsiveness increases strategic deviations.

Table 1: Experimental Design: Static vs. Dynamic Routing Algorithm

This design enables a direct comparison of how adaptive algorithms influence users' strategic reasoning in route choice. This design also enables us to easily analyze the distribution of inferred K-levels and participants' written explanations to uncover additional influences on behavior. These qualitative responses provide insight into factors such as risk aversion, distrust of the app, prior experience, and awareness of others' choices, all of which may affect the depth of strategic thinking.

To ensure sufficient statistical power, we target a sample size that allows us to detect a minimum detectable effect (MDE) of approximately 10 percentage points in the share of Level-1 or Level-2 participants between groups, assuming baseline strategic reasoning of 15%. Power calculations indicate that a total sample of at least 160 participants (80 per group) is required to detect this difference with 80% power at a conservative significance level of $p < 0.01$. This sample size enables us to confidently evaluate `Hypothesis 2`.

Despite efforts to simulate real-world navigation conditions, we acknowledge that a lab experiment inherently lacks the contextual stakes of actual driving, such as real-time pressure, multitasking, or safety concerns. While controlled environments improve internal validity, they may underestimate or misrepresent how users balance strategic reasoning with cognitive load. Future work should examine whether similar reasoning patterns emerge in naturalistic driving contexts or through mobile A/B testing in deployed apps.

Navigation App Interface

Experiment participants will interact with a simulated navigation app interface modeled on real-world platforms like Google Maps. The interface displays two route options, each with estimated travel times and a visual map indicating congestion levels.

In each round, participants face a choice between two routes: one with a base travel time of X minutes and the other with $X + Y$ minutes (e.g. 20 and 25 minutes, respectively). To model congestion effects, travel time increases by 2 minutes for every additional five participants selecting the same route. Each participant completes 10 sequential route choices, with randomized initial positions to reduce learning effects and order bias. In the dynamic algorithm condition, travel time estimates are refreshed every 5 seconds to simulate real-time re-routing based on traffic flow.

The experiment will be implemented as a web-based application using the oTree framework. Participant interactions, including route choices, timestamps, and written justifications, are recorded in a PostgreSQL database.

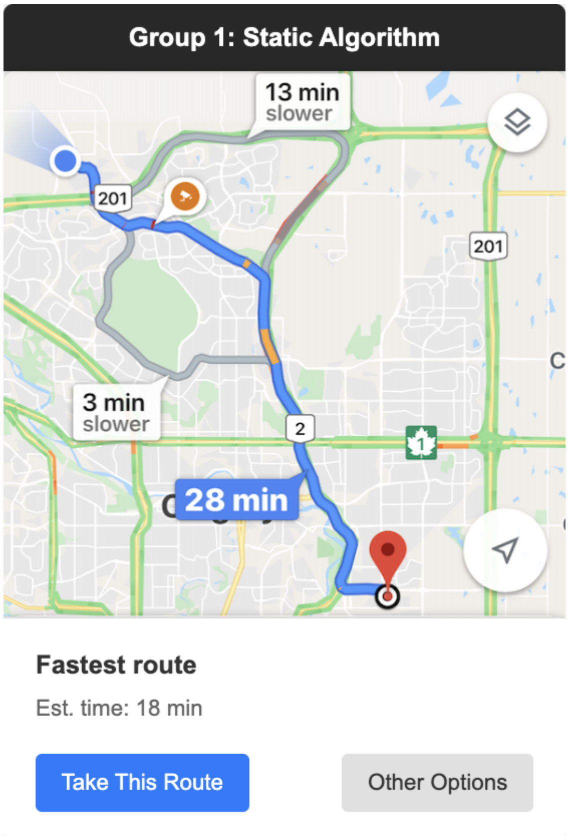

Figure 2: Group 1 (Static Algorithm) Map Interface

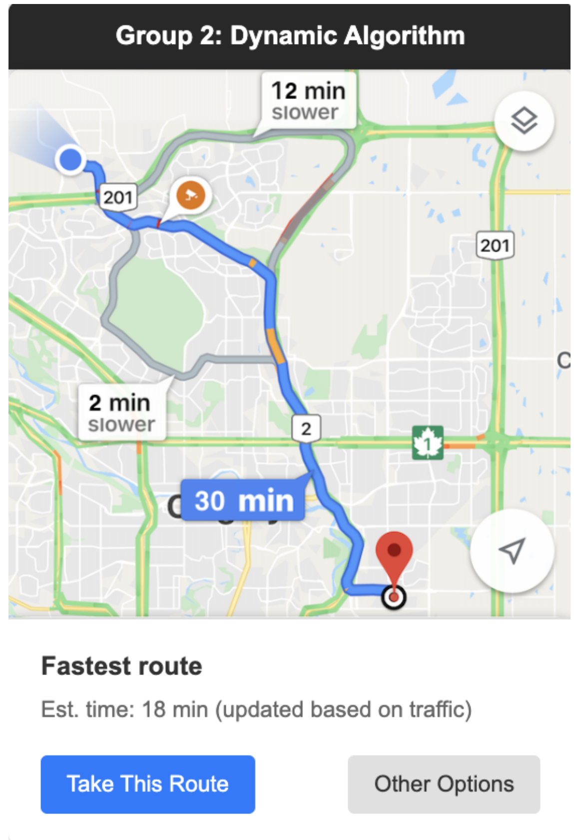

Figure 3: Group 2 (Dynamic Algorithm) Map Interface

Map interfaces are based on actual road networks but feature anonymized landmarks to eliminate geographic bias. Travel time estimates are computed using empirical congestion functions adapted from transportation planning models, grounding the experiment in plausible behavioral dynamics.

Research Questions and Hypotheses

This study investigates how users of navigation apps reason strategically when choosing routes, particularly in response to algorithmic design. We address the following research questions:

1. **RQ1:** What is the distribution of Level-K reasoning ($K = 0, 1, 2$) among navigation app users when making route choices?
2. **RQ2:** Does exposure to visibly adaptive routing algorithms increase the prevalence of higher-level strategic reasoning compared to static algorithms?

3. **RQ3:** What factors do users most commonly cite when explaining their decisions to deviate from app-recommended routes?

These questions motivate three empirically testable hypotheses:

1. **Hypothesis 1 (H1, related to RQ1):** The majority of participants will exhibit Level-0 reasoning, with smaller proportions displaying Level-1 and Level-2 reasoning. This distribution reflects patterns found in prior Level-K studies in non-algorithmic domains.

2. **Hypothesis 2 (H2, related to RQ2):** Participants exposed to adaptive routing algorithms will demonstrate higher average reasoning levels (Level-1 or Level-2) than those interacting with static algorithms.

3. **Hypothesis 3 (H3, related to RQ3):** The most frequently cited justification for deviating from the recommended route will be the anticipation of congestion caused by other users.

Each hypothesis is designed to be explicitly falsifiable. H1 would be rejected if higher-level reasoning dominates participant behavior. H2 requires that users meaningfully respond to algorithmic adaptation. H3 depends on consistent qualitative themes emerging from coded responses. All hypothesized directions, outcome measures, and effect sizes will be pre-registered to enhance transparency and guard against post-hoc theorizing. We use a conservative threshold of $p < 0.01$ to assess statistical significance and reduce the likelihood of false positives.

Key Metrics

We use four key metrics to evaluate our hypotheses:

- **Route Choice ("fastest" or "alternate")**: The primary behavioral outcome is whether a participant follows or deviates from the app's recommended route. This binary decision reflects compliance versus strategic deviation and provides a first-stage test of **H1** and **H2**. These choices also serve as inputs for inferring Level-K reasoning levels.

- **Self-Reported Reasoning (Text)**: Participants' open-ended justifications will be analyzed to validate inferred reasoning levels and uncover motivations. This qualitative data supports both **H1** and **H2**, and is central to testing **H3**—specifically, whether deviations stem from concerns about others' choices or congestion effects.

- **Inferred K-Level of Route Choice (K = 0, 1, 2)**: Each decision will be classified as Level-0, Level-1, or Level-2 based on observed route choice and textual explanation. This metric enables direct estimation of the reasoning distribution and supports comparisons across conditions to test **H1** and **H2**.

- **Distance from Equilibrium (D)**: This quantitative measure captures how far each group's observed route shares deviate from Nash equilibrium predictions. By comparing D across rounds and groups, we assess whether adaptive routing promotes more efficient coordination. This metric supports tests of **H2** and offers indirect evidence for bounded strategic reasoning consistent with Level-K theory.

To calculate **Inferred K-Level of Route Choice**, we will use a logistic regression model that predicts the probability of being classified as Level-0, Level-1, or Level-2 based on each route choice and accompanying explanation. Each participant is assigned the K-level with the highest predicted probability. These individual-level classifications are then aggregated to estimate the overall distribution of strategic reasoning and to compare treatment conditions. While rule-based classification would be a simpler approach, it lacks flexibility in handling ambiguous or mixed-strategy responses and does not scale effectively for group-level inference. To train the model, we will manually label a subset of data based on participants' written justifications and route choices. Labels are assigned using a predefined coding scheme aligned with Level-K definitions and reviewed for consistency. While the model outputs probability distributions over K-levels, we assign the classification with the highest probability. We will conduct basic validation checks, including inter-rater agreement and cross-validation, to ensure model reliability.

Together, these metrics provide the evidence needed to evaluate all three hypotheses. By combining coded behavioral outcomes with textual analysis, we ensure a comprehensive understanding of how and why participants deviate from default recommendations.

Control Variables and Confounding Factors

To improve internal validity and isolate the causal impact of algorithm design on user behavior, we will control for some key individual-level factors that may influence route choice:

- **Navigation app experience:** How often the participant uses navigation apps in everyday driving.
- **Navigation platforms used:** The specific apps participants typically rely on (e.g., Google Maps, Apple Maps, Waze).
- **Driving experience:** The total number of years participants have been driving.
- **Route familiarity and preference:** Whether participants are familiar with the test route or have a preferred route based on factors other than travel time (e.g., scenic value, habit).

To account for these factors, we use a pre-experiment questionnaire followed by stratified randomization to balance these characteristics across treatment groups. This approach helps reduce bias from individual differences and ensures that observed effects are attributable to the experimental manipulations rather than prior habits or preferences.

Incentive Structure

Each participant receives a base participation payment. To align user behavior with realistic decision-making, we offer performance-based monetary incentives. Simulated travel times are stochastic and congestion-sensitive: when many users select the same route, its travel time increases probabilistically. Participants earn higher bonuses by minimizing travel time over the session, rewarding both foresight and adaptability. This structure encourages participants to internalize coordination problems and engage in strategic reasoning when beneficial.

Analysis Plan

Our proposed analysis strategy integrates both quantitative and qualitative data to evaluate the three pre-registered hypotheses.

To test **H1**, we report the distribution of inferred K-levels across the full sample. A predominance of Level-0 classifications would support the hypothesis, while a higher share of Level-1 or Level-2 behavior would falsify it.

To test **H2**, we compare K-level distributions between treatment groups using a chi-squared test of independence. This test evaluates whether the distribution of reasoning types is statistically independent of algorithm adaptiveness.

	Observed (%)			**Relative Change (%)**		
Group	Level-0	Level-1	Level-2	Level-0	Level-1	Level-2
Group 1 (Static)	83.6	12.9	3.5	–	–	–
Group 2 (Dynamic)	78.2	13.3	8.5	−6.5	+3.1	+142.9
p-value from chi-squared test of independence: **0.007**						

Table 2: Example Distribution of Inferred K-Level Reasoning by Group

H3 is evaluated using the Self-Reported Reasoning (Text) metric. If the most common justification for deviation references anticipated congestion caused by other users, the hypothesis is supported. Open-ended responses will be thematically coded using a predefined rubric derived from Level-K reasoning types.

K-Level	Typical Explanation	Interpretation / Reasoning Logic
Level-0	*"I followed the app because it said it was fastest."*	Non-strategic. Accepts the app's recommendation at face value without considering others' behavior. Assumes the app is optimal by default.
Level-1	*"I thought the recommended route would be crowded since everyone would take it."*	Strategic response to anticipated congestion. Assumes others are Level-0 and avoids the obvious route as a best response.
Level-2	*"Others will try to avoid the recommended route, so it might actually be clear."*	Second-order reasoning. Assumes others are Level-1 and best responds by returning to the recommended route.

Table 3: Anticipated Participant Explanations by Inferred K-Level

Finally, we will compare observed route choices to theoretical predictions from the Nash equilibrium model introduced earlier. This benchmark reflects fully rational coordination. Deviations from equilibrium provide additional evidence of bounded strategic reasoning consistent with Level-K theory.

Group	Observed Share on Route 1 (%)	Nash Predicted Share (%)	Deviation (%)
Group 1 (Static)	62.5	50.0	+12.5
Group 2 (Dynamic)	52.0	50.0	+2.0

Table 4: Example Comparison of Observed Route Choices vs. Nash Equilibrium Predictions

Conclusion, Implications, and Future Work

This research investigates whether users of navigation apps exhibit Level-K reasoning in route choices, and whether dynamic algorithm designs encourage deeper strategic thinking. By estimating the distribution of Level-0, Level-1, and Level-2 users in a controlled lab setting, we contribute new empirical evidence to the Level-K literature while extending it to human-algorithm interaction. Modeling the navigation platform as a rational player reveals how system design can shape user behavior, particularly in environments with real-time feedback.

Practical Implications

The proposed study has direct implications for the design of navigation platforms and transportation policy. If adaptive interfaces increase higher-order reasoning, developers could implement crowd-aware messages or selectively deploy real-time updates to nudge users toward socially efficient routing. Personalized framing based on inferred reasoning levels may also reduce congestion externalities by discouraging herd behavior. From a policy perspective, understanding the cognitive mechanisms behind route selection can inform infrastructure decisions aligned with broader goals such as congestion pricing, traffic equity, and environmental sustainability.

Future Work

Future research should extend these lab-based insights to field settings. An online randomized experiment on a navigation app could test whether strategic framing, such as *"Many drivers are accepting this route"*, influences deviation behavior and reasoning levels at scale. In-app metrics would allow for passive tracking of route choices, compliance rates, and congestion outcomes.

Longitudinal studies could examine whether reasoning levels evolve with repeated exposure to adaptive guidance, and whether users learn to anticipate algorithmic behavior over time. Additional work might explore heterogeneity in treatment effects by user type. For example, whether urban drivers respond differently than rural users to strategic cues.

Ultimately, this research aims to inform the improved design of routing platforms to account for both human-human and human-algorithm strategic interactions, with the goal of reducing app-induced traffic congestion.

References

[1] Google Maps Support Community, "Offered faster route options are mathematically impossible." `https://support.google.com/maps/thread/226593428/offered-faster-route-options-are-mathematically-impossible`, 2024. Accessed: 2025-05-04.

[2] Android Community, "Google maps dynamic rerouting helps users outsmart traffic," 2014. Accessed: 2025-05-09.

[3] A. Guin, "The impact of smartphone applications on trip routing," tech. rep., Bureau of Transportation Statistics, 2021.

[4] Y. Cabannes, "Waze and the battle for the neighborhood: How apps create new urban traffic patterns," *Urban Planning and Transport Research*, vol. 6, no. 1, pp. 75–91, 2018. Case studies on rerouting effects of navigation apps.

[5] E. Mayr, "Designing mobile application messages to impact route choice: A survey and simulation study," *National Institutes of Health (NIH)*, 2023.

[6] R. Jiang, "A study of driver's route choice behavior based on evolutionary game theory," *National Institutes of Health (NIH)*, 2014.

[7] S. Vosough, "Achieving social routing via navigation apps: User acceptance of travel time sacrifice," *Transportation Research*, 2024.

[8] M. Razo and S. Gao, "Strategic route choice behavior with real-time information: An experimental approach," *Transportation Research Part C: Emerging Technologies*, vol. 27, pp. 49–61, 2013.

[9] E. Ben-Elia and Y. Shiftan, "Which road do i take? a learning-based model of route-choice behavior with real-time information," *Transportation Research Part A: Policy and Practice*, vol. 44, no. 4, pp. 249–264, 2010.

[10] M. Kroller, "Driver expectations toward strategic routing," *Journal of Transportation Behavior*, 2021.

[11] Y. Zhang, Z. Liu, and W. Xu, "Real-time routing in traffic navigation systems: A review of algorithms and challenges," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 1, pp. 1–15, 2023.

[12] J. Palifka, "Ever wondered how google maps finds the shortest route? the secret revealed," 2023. Accessed: 2025-05-08.