Active Inference for Decentralised Traffic Coordination: A Multi-Agent Route-Choice Demonstration

Xue Yao[†1][0000−0002−6575−8758], Dmitry Bagaev[†2][0000−0001−9655−7986], Yanan Xin[1][0000−0003−3866−821X], Bert de Vries[2][0000−0003−0839−174X], Hans van Lint[1][0000−0003−1493−6750], and Serge P. Hoogendoorn[1][0000−0002−1579−1939]

[1] Department of Transport & Planning, Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands
x.yao-3@tudelft.nl

[2] Department of Electrical Engineering, Eindhoven University of Technology, Groene Loper 3, 5612 AP Eindhoven, The Netherlands
d.v.bagaev@tue.nl

Abstract. Traffic systems are fundamentally decentralised and partially observable, where network-level outcomes emerge from the local decisions of travellers under uncertain congestion. Existing traffic management pipelines separate estimation, prediction, routing, and control into disconnected objectives and treat uncertainty as an external disturbance rather than an intrinsic component of decision-making. This paper investigates whether traffic management can instead be formulated as a unified inference problem within the Active Inference (AIF) framework. We argue that three structural challenges of cooperative traffic systems naturally align with the mechanisms of AIF, including uncertainty-aware action selection, belief-guided adaptation, and decentralised coordination through shared environmental dynamics. We demonstrate this in a route-choice network in which AIF travel agents update their beliefs and adapt their route choices by minimising Expected Free Energy, under both stationary and non-stationary traffic conditions (recurrent congestion, sustained capacity disruptions, and stochastic disruptions). Results show that AIF agents learn latent congestion conditions under partial observability and adapt more robustly in changing network conditions. Compared with multinomial logit (MNL) route choice and reinforcement-style day-to-day (RL-DTD) learning, AIF produces more stable traffic patterns while naturally generating interpretable belief dynamics and exploratory behaviour. These findings highlight the potential of Active Inference as a unified framework for uncertainty-aware traffic coordination and decentralised collective inference.

Keywords: Active Inference · Collective adaptation · Decentralised traffic coordination · Probabilistic Machine Learning · Route choice under uncertainty.

[†] These authors contributed equally to this work.

1 Introduction

Traffic systems are characterised by decentralised decision-making and uncertainty. Macroscopic traffic patterns emerge from repeated local decisions made by travellers and increasingly automated agents operating with incomplete information about future traffic conditions. Effective traffic management therefore requires continuous estimation, prediction, learning, routing, and control across multiple interacting actors and time scales. Despite substantial advances in intelligent transportation systems, these functions are typically addressed through separate optimisation, control, or learning frameworks, resulting in fragmented system decision-making architectures and limited integration of uncertainty across decision layers.

Most existing traffic management approaches are grounded in optimisation and equilibrium principles, where travellers respond to known or estimated network costs [19,23]. Model Predictive Control (MPC) and reinforcement learning approaches have further improved adaptive traffic control capabilities [1, 26]. However, traffic coordination under uncertainty remains challenging because future network conditions are only partially observable. Travellers must therefore adapt decisions based on incomplete and noisy information. This raises a more fundamental question: can network-level traffic dynamics emerge from repeated local probabilistic inference under partial observability, rather than solely from optimisation or equilibrium principles?

Active Inference (AIF) provides a probabilistic framework for modelling adaptive behaviour under uncertainty through variational Bayesian inference and Expected Free Energy minimisation [8, 9, 22]. By integrating belief updating, uncertainty-aware action selection, and information-seeking exploration within a single formulation, AIF offers a principled perspective on adaptive decision-making in decentralised systems. These properties make AIF a natural framework for addressing this question [7].

Prior applications of AIF in the traffic domain have focused on single-vehicle driving behaviour [7, 14], pairwise road-user interactions [18], and adaptive traffic signal control [20], which addresses infrastructure-side actuation rather than traveller-side route choice. This paper investigates whether macroscopic traffic coordination can emerge from repeated local Active Inference by decentralised travellers operating under partial observability. Using a route-choice demonstration, we: (i) establish a unified Active Inference formulation that maps route choice under partial observability to probabilistic inference over latent traffic conditions; (ii) show that repeated local inference is sufficient to generate coordinated network-level traffic dynamics without centralised optimisation; (iii) demonstrate that explicit belief updating and uncertainty-aware decision-making enable robust adaptation under both stationary and non-stationary traffic conditions. The study demonstrates how Active Inference simultaneously explains individual route adaptation and the emergence of coordinated network-level traffic dynamics within a unified probabilistic framework.

2 Motivating Active Inference in Traffic Systems

Uncertainty is intrinsic to traffic systems. Traffic systems operate under substantial uncertainty arising from demand fluctuations, incidents, weather conditions, sensor failures, and behavioural variability. Travellers therefore choose routes with incomplete knowledge of future congestion, and uncertainty itself shapes their exploration, adaptation, and behavioural diversity. Effective coordination consequently requires explicitly representing and responding to uncertainty, not merely accounting for expected travel costs. Classical traffic assignment models, however, generally assume deterministic or equilibrium travel costs [19], MPC-based control incorporates uncertainty primarily through additive disturbance models and short-horizon prediction corrections [11], and reinforcement learning optimises expected rewards under fixed training distributions [26]. In each case, actions follow from expected system states, leaving uncertainty a limited role in the decision itself.

Traffic management is fragmented across objectives. Estimation, prediction, routing, learning, and control must operate simultaneously, yet are commonly addressed by separate modules with distinct objectives, assumptions, and representations: filtering or data assimilation for state estimation, statistical or machine learning models for prediction, optimisation or equilibrium formulations for route choice, and Model Predictive Control or reinforcement learning for control [1,21]. Although effective within individual tasks, such methods coordinate poorly because each optimises its own objective through task-specific interfaces. Meanwhile, vehicle-to-everything communication, autonomous vehicles, and distributed sensing are shifting traffic management from centralised control towards distributed adaptation, creating a growing need for frameworks that can coordinate many interacting agents under uncertainty.

Decentralised coordination remains unresolved. Traffic performance emerges from the collective behaviour of many independent decision-makers. Centralised optimisation becomes increasingly difficult as network size and behavioural complexity grow, while purely decentralised adaptation can produce inefficient outcomes because locally rational decisions need not align with system-level objectives, as illustrated by Wardrop equilibria and Braess-type phenomena [17, 23]. Existing approaches typically rely on either centralised coordination mechanisms or externally designed optimisation and reward structures to align individual and system-level objectives, so achieving robust, scalable coordination among decentralised agents under uncertain and evolving traffic conditions remains an open challenge.

Active Inference offers a promising unified framework for decentralised traffic coordination. By treating perception, learning, prediction, and action selection as coupled inference processes under a shared generative model, it addresses these three challenges within a single formulation, and through Expected Free Energy both goal-directed behaviour and information-seeking exploration emerge from the same objective.

3 Methodology

We consider a route-choice problem within the Active Inference (AIF) framework. AIF agents maintain probabilistic beliefs about latent traffic conditions, update these beliefs from experienced travel times, and select routes by minimising Expected Free Energy.

3.1 Multi-agent route-choice setting

Travellers choose among n competing routes between a single origin–destination pair. Time evolves over two interacting scales: a day-to-day learning [4] index $d \in \{1, \dots, D\}$, and a within-day departure interval $t \in \{1, \dots, T\}$ for traffic evolution and dynamic demand. Each traveller i has a fixed departure interval $t^{(i)}$ that remains unchanged throughout the simulation, so the day-to-day adaptation task concerns only route choice: on each day d, traveller i selects a route $a_d^{(i)} \in \{r_1, \dots, r_n\}$. After completing the trip, the traveller only observes information $y_d^{(i)}$ on the chosen route, while congestion conditions on unchosen routes remain unobserved. Within the Active Inference framework, each traveller is represented by an agent that maintains a latent state $x_d^{(i)}$ describing its beliefs about network congestion.

3.2 Probabilistic generative model of agents

We model each agent's interaction with the network as a partially observed Markov process with route-choice actions as causes of future congestion evolution [5, 15]. Let $x_{1:D}^{(i)}$, $y_{1:D}^{(i)}$, and $a_{1:D}^{(i)}$ denote the trajectories of hidden congestion states, travel-time observations, and route-choice actions over the simulation horizon. The generative model of a single agent i factorises as

$$p(y_{1:D}, x_{0:D}, a_{1:D}) = p(x_0) \prod_{d=1}^{D} p(y_d \mid x_d)\; p(x_d \mid x_{d-1}, a_d)\; p(a_d), \tag{1}$$

where for brevity we dropped the superscript (i) and x_0 represents agent's prior beliefs before entering the network. The generative model comprises: (i) the observation likelihood $p(y_d \mid x_d)$, linking experienced travel times to latent states; (ii) the transition kernel $p(x_d \mid x_{d-1}, a_d)$, encoding the agent's belief about how its actions affect the network; and (iii) the prior over actions $p(a_d)$ (policy prior). The equivalent factor graph is given in Appendix A.

The agent i does not directly observe the latent congestion state x_d and instead maintains an approximate posterior belief

$$q(x_d^{(i)}) \approx p(x_d^{(i)} \mid y_{1:d}^{(i)}, a_{1:d}^{(i)}), \tag{2}$$

in the form of a probability distribution which is updated after each trip using realised travel-time observations accumulated through repeated interactions with the traffic environment. Route choice (§3.5) is therefore formulated as probabilistic inference under uncertainty rather than deterministic optimisation under perfect information.

3.3 Dynamic traffic environment and observations

Traffic conditions evolve through a dynamic queueing environment [6]. The store-and-forward model represents each route as a queue whose length evolves according to the difference between inflow and bottleneck discharge capacity. Vehicles entering a route are temporarily stored before leaving at the downstream bottleneck, providing a simple approximation of congestion propagation and travel-time evolution. Each route $r \in \{1, 2, \ldots, n\}$ is characterised by its free-flow travel time F_r, bottleneck capacity C_r, route inflow $Q_r(d,t)$, and queue length $L_r(d,t)$. The within-day departure demand $D(t)$ follows a shifted-sine profile resembling typical afternoon traffic conditions with a peak at $t = T/2$.

$$D(t) = D_{\min} + (D_{\max} - D_{\min}) \sin^2(\pi t/T), \tag{3}$$

Route inflows are determined by the collective route choices of all agents,

$$Q_r(d,t) = P_r(d,t) D(t), \tag{4}$$

where $P_r(d,t)$ denotes the proportion of agents selecting route r at departure interval t on day d.

Queue lengths then evolve as [12]

$$L_r(d, t+1) = \max\left(0,\, L_r(d,t) + (Q_r(d, t - N_r) - C_r)\Delta t\right), \tag{5}$$

where Δt is the difference between two subsequent time intervals and $N_r = F_r/\Delta t$ represents the route propagation delay. The realised travel time on route r is then given by

$$TT_r(d,t) = F_r + 60\, L_r(d, t + N_r)/C_r. \tag{6}$$

The latent state introduced in §3.1 is therefore represented by the evolving traffic conditions across all available routes at the agent's departure interval $t^{(i)}$.

$$x_{d,k}^{(i)} = \left(L_{r_k}(d,\, t^{(i)} + N_{r_k}),\, C_{r_k}(d, t^{(i)}),\, F_{r_k}(d, t^{(i)})\right), \;\; k = 1:n \tag{7}$$

After completing the trip on the chosen route $a_d^{(i)}$, agent i receives two noisy observations: the realised travel time and a noisy measurement of the encountered queue length,

$$y_{d,\mathrm{TT}}^{(i)} = TT_{a_d^{(i)}}(d, t^{(i)}) + \tau_d, \;\; \tau_d \sim \mathcal{N}(0, \sigma_{\mathrm{obs}}^2), \tag{8}$$

$$y_{d,L}^{(i)} = L_{a_d^{(i)}}(d,\, t^{(i)} + N_{a_d^{(i)}}) + \eta_d, \;\; \eta_d \sim \mathcal{N}(0, \sigma_{L,\mathrm{obs}}^2), \tag{9}$$

where σ_{obs} and $\sigma_{L,\mathrm{obs}}$ are configurable network parameters. The queue-length signal $y_{d,L}$ models direct sensing of congestion at the start of the drive (e.g. via navigation applications, variable-message signs, or visual inspection).

Combined with travel-time observations, it allows the latent variables (F_r, C_r, L_r) of the generative model to be identified despite observing only the chosen route each day.

3.4 Perception via Variational Free Energy

The traffic model and the Active Inference model form a closed perception–action loop: travellers infer latent route conditions from experienced travel times, while their route choices determine the aggregate flows and congestion that generate subsequent observations. Given the observed history of travel-time $\hat{y}$ and route-choice actions $\hat{a}$, agent perception and belief updating are formulated as Variational Free Energy minimisation

$$F[q] \triangleq \underbrace{\int q(x) \log\big(q(x)/p(x \mid \hat{y}, \hat{a})\big)\, \mathrm{d}x}_{\mathrm{KL}[q(x)\|p(x|\hat{y},\hat{a})]} - \underbrace{\log p(\hat{y} \mid \hat{a})}_{\text{log evidence}}, \tag{10}$$

where we again omitted the superscript (i) for brevity.

Agents update their beliefs by minimising $F[q]$, which drives the posterior q toward the true posterior implied by the generative model in eq. (1). Congestion estimation, belief updating, and uncertainty reduction thus become different aspects of approximate Bayesian inference on the same probabilistic model p.

3.5 Action selection via Expected Free Energy

Prior to departure at interval t on day d, agents evaluate candidate actions by predicting future congestion conditions and expected travel-time observations under each route alternative given their current beliefs q. Action selection is formulated through the Expected Free Energy (EFE) [10, 13],

$$G(a_d) = \underbrace{D_{\mathrm{KL}}\left[q(y_d \mid a_d) \| \tilde{p}(y_d \mid a_d)\right]}_{\text{risk}} + \underbrace{\mathbb{E}_{q(x_d \mid a_d)}\left[H[p(y_d \mid x_d)]\right]}_{\text{ambiguity}}, \tag{11}$$

where $\tilde{p}(y_d \mid a_d)$ denotes the preference distribution over preferred travel outcomes. Since agents naturally prefer low travel times, we define

$$\tilde{p}(y_d \mid a_d) = \mathcal{N}(y_d \mid \mu_{F_{a_d}}, \sigma^2_{\text{pref}}), \tag{12}$$

where $\mu_{F_{a_d}}$ is the agent's posterior mean of the free-flow travel time on the chosen route and σ_{pref} represents tolerance for delay. The risk term therefore penalises expected deviations from free-flow conditions, while the ambiguity term favours actions expected to reduce uncertainty about latent states.

Route-choice probabilities are then obtained through a softmax policy over Expected Free Energy $p(a_d^{(i)}) \propto \exp(-\gamma G(a_d^{(i)}))$, where γ denotes the policy precision parameter. Exploratory route choice therefore emerges from inference itself, without externally imposed exploration heuristics.

4 Experiments and Results

The experiments evaluate the proposed Active Inference framework under a range of traffic conditions, including recurrent congestion and network disruptions.

4.1 Experimental setup

The simulation considers a classic two-route corridor based on the motorway network between the Valburg and Velperbroek junctions in the Netherlands [12], as illustrated in Figure 1. Route A represents a shorter and higher-capacity motorway corridor, while Route B represents a longer but lower-capacity alternative route. The route characteristics and traffic parameters are derived from this real-world corridor, where free-flow travel times and bottleneck capacities are specified as $F_A = 16$ min, $C_A = 4300$ veh/h, and $F_B = 18$ min, $C_B = 2300$ veh/h. Because Route B possesses a consistently lower bottleneck capacity, its congestion level becomes highly sensitive to redirected demand. This asymmetry creates a decentralised coordination problem. Detailed implementation settings of the simulation environment are provided in Appendix B.

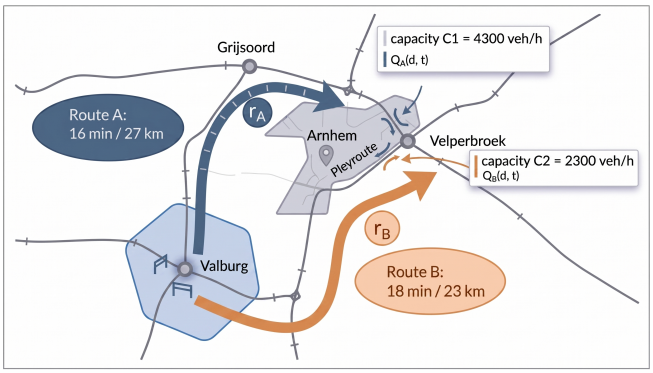


Fig. 1: Route-choice demonstration based on the Valburg–Velperbroek motorway corridor [12].

4.2 Belief learning under partial observability

We first examined whether Active Inference agents can infer latent states from noisy and delayed travel-time observations gathered in a disruption-free environment. Our verification confirms that they can, through sequential probabilistic inference based on local observations; detailed results are reported in Appendix C.

The within-day learning profiles in Figure 2 illustrate how individual learning processes collectively shape network-level traffic dynamics. During the early learning phase, most agents favour the shorter Route A, resulting in substantial congestion on Route A while Route B remains underutilised. As experience accumulates, agents progressively shift toward Route B, particularly during peak periods, reducing congestion on Route A by redistributing traffic demand across the network. The travel-time belief profiles in Figure 2a show that AIF agents

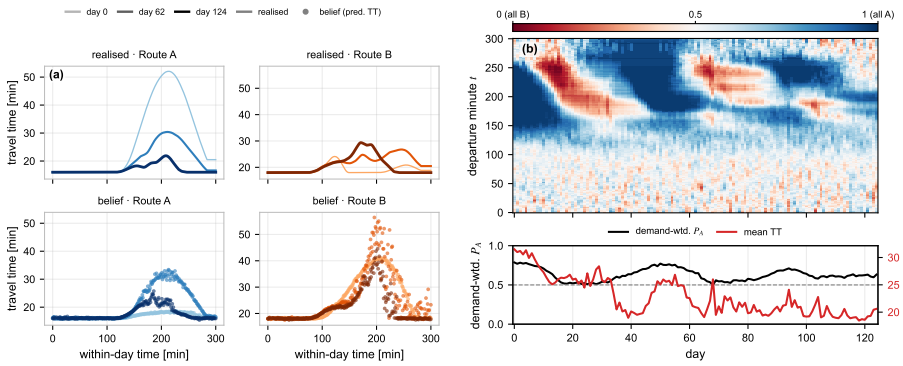


Fig. 2: Baseline learning profiles: (a) Within-day realised travel-time (top) and beliefs on predictive travel time (bottom). (b) Route-choice split (top), demand-weighted P_A vs average travel time (bottom).

successfully learn the underlying congestion patterns, although they tend to overestimate travel times on Route B and underestimate them on Route A, indicating greater sensitivity to low-capacity routes. The route-choice heatmap in Figure 2b reveals clear time-dependent adaptation across departure intervals, with peak-period agents adjusting more strongly than off-peak agents.

Overall, the results reveal a dynamically stable traffic allocation in which decentralised agent interactions continuously redistribute congestion across the network. Congestion patterns remain dynamically balanced through ongoing belief updating and route adaptation, rather than settling into a static allocation, consistent with stochastic-process formulations of day-to-day assignment [24].

4.3 Adaptation under disruptions

We next investigate whether Active Inference agents can detect and adapt to network disruptions through uncertainty-aware probabilistic inference. Two representative disruption scenarios are considered: (i) road construction causing a sustained capacity drop and (ii) stochastic incidents causing within-day congestion spikes. In the road-construction scenario, the capacity of Route A is reduced to $0.5C_A$ for 20 consecutive days before being restored to its original value. The capacity drop immediately increases queue lengths and travel times on Route A, after which agents gradually shift towards Route B, leading to increased congestion on Route B as well (Figure 3a,c). Despite this behavioural adaptation, the average system cost remains increased throughout the construction period due to the reduced capacity of the main route (Figure 3b,d). Once the construction ends, system costs decrease and the route share of Route A gradually recovers as traffic redistributes across the network.

The earliest departure intervals in Figure 3c remain near $P_A \approx 0.5$ rather than collapsing onto the faster Route A, because the preference of eq. (12) is centred

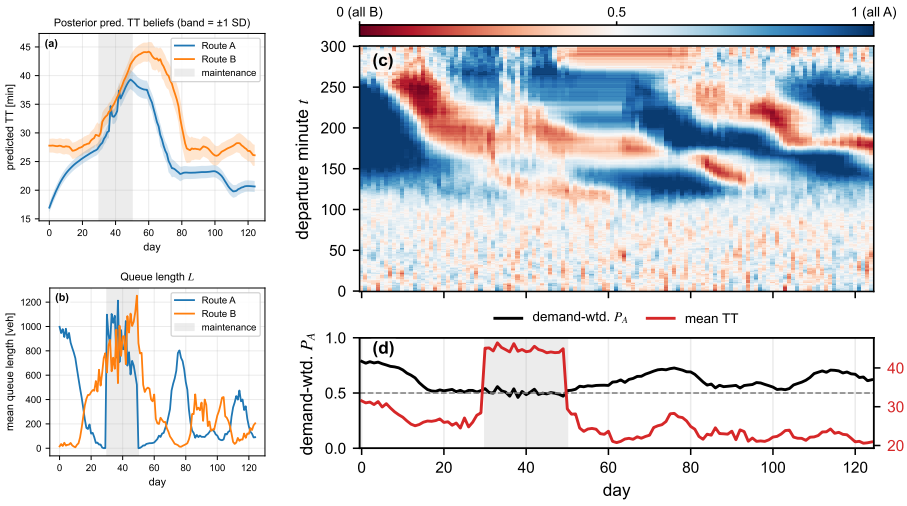


Fig. 3: Road construction: (a) Posterior predictive travel-time beliefs. (b) Daily average queue length. (c) Route-A share P_A. (d) Demand-weighted P_A vs average travel time.

on each route's own free-flow time and the risk term therefore scores congestion delay rather than absolute travel time. Before the demand peak both routes run at free flow and thus attain their preferred outcome, leaving the Expected Free Energy almost indifferent between them, an indifference the stochastic policy realises as an even split (Appendix B).

Figure 4 demonstrates the individual learning processes underlying the observed traffic adaptation. During off-peak departure periods, the AIF agents receive limited information about capacity and therefore show weaker inference of the capacity change. As a result, travel times on both routes remain close to their free-flow values. In contrast, during peak-hour departure periods, when capacity becomes a critical determinant of congestion, AIF agents successfully learn the capacity drop and adjust their route choices according to the inferred queue conditions, albeit with some delay. Interestingly, the representative agent departing after the peak period develops a lower capacity belief after the construction (column b). This occurs because queues are almost absent there, so a lower inferred capacity still explains the observations adequately.

The second disruption scenario considers stochastic within-day traffic incidents. For each route $r \in \{A, B\}$ and day d, incidents occur probabilistically with route-specific probabilities $p_A = 0.15$ and $p_B = 0.05$. When an incident occurs, the corresponding route capacity is temporarily reduced by half for 30 minutes. Incident start times are sampled according to the within-day demand distribution, making incidents more likely during peak-demand periods. Figure 5a shows that stochastic incidents generally increase agents' predictive travel-time beliefs. The

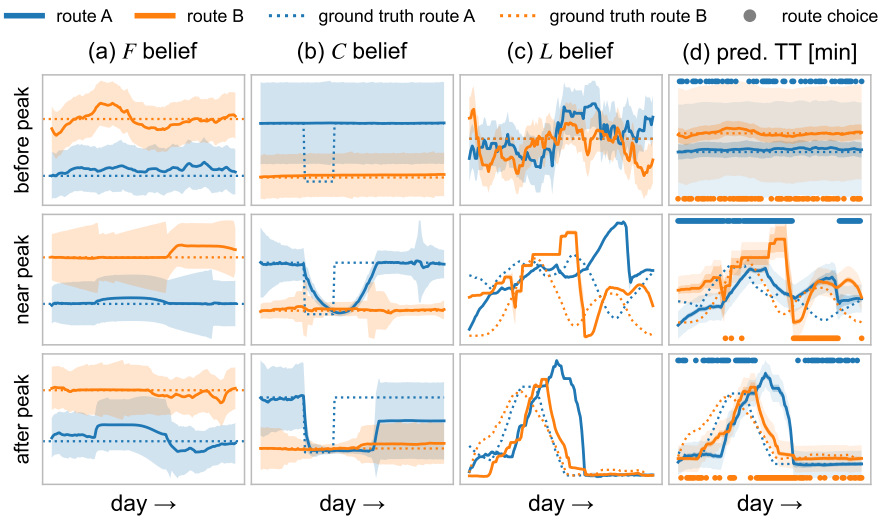


Fig. 4: Road construction: per-agent inference for three representative agents departing before, near, and after peak hours.

realised travel-time distributions in Figure 5b exhibit heavier tails on Route A and lighter tails on Route B compared with the baseline case. This is because Route A serves as the primary route and is therefore more sensitive to temporary capacity disruptions. Figures 5c and d show that incidents cause occasional spikes in daily average travel time. Route shares also exhibit small fluctuations; however, the overall day-to-day route-choice pattern remains similar to the baseline case.

Overall, the disruption scenarios demonstrate that, despite delayed and partial observations, Active Inference agents adjust their beliefs and route choices to both persistent and stochastic disturbances without requiring explicit knowledge of the underlying disruptions.

4.4 Performance comparison with benchmarks

To evaluate the proposed Active Inference (AIF) framework, we compare it with two widely used route-choice models: a multinomial logit (MNL) model [2] based on stochastic utility maximisation, and a reinforcement-style day-to-day learning model (RL-DTD) [25] based on experience-driven adaptation. In contrast, AIF formulates route adaptation as probabilistic belief updating and Expected Free Energy minimisation under uncertainty and partial observability. A brief review of benchmarks, conceptual comparison of the three frameworks, and implementation details are provided in Appendix D.1.

Figure 6 compares the performance of AIF, MNL, and RL-DTD models in terms of route shares and total system cost. Under baseline conditions (column a), both AIF and RL-DTD achieve substantially lower system costs than under MNL

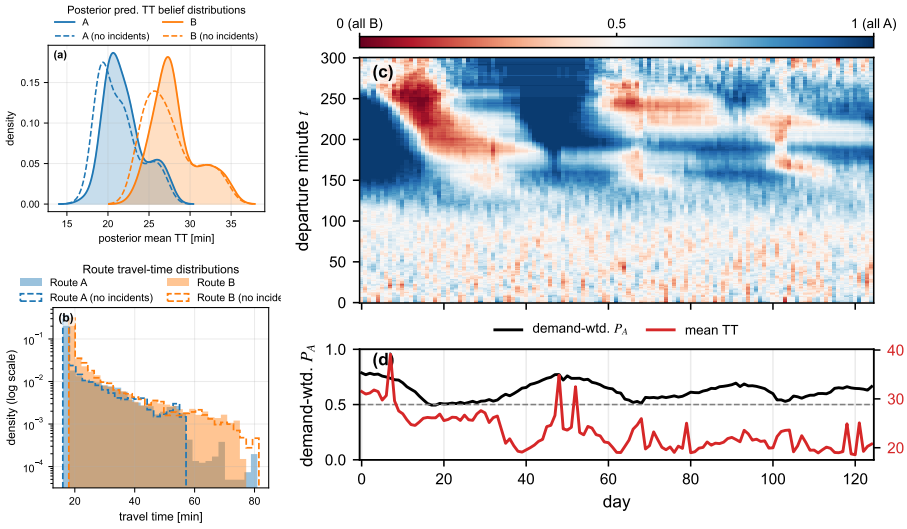


Fig. 5: Stochastic incidents: (a) Posterior predictive average travel-time distributions. (b) Realised travel-time distributions. (c) Route-A share P_A. (d) Demand-weighted P_A vs average travel time.

by adapting route shares and maintaining balanced queue lengths across the two routes. The differences become more pronounced under disruptions. During road construction (column b), RL-DTD responds aggressively to the capacity reduction, producing large oscillations in route shares and queue lengths. Although the average system cost remains lower than under MNL, the resulting traffic dynamics are unstable. Instead, AIF adapts to the disruption with substantially smoother adjustments. Similar behaviour is observed under stochastic incidents (column c), where AIF exhibits the smallest performance fluctuations despite stochastic capacity disruptions. The detailed comparison of the route congestion in each scenario is provided in Appendix D.2.

Overall, the benchmark comparison suggests the advantage of AIF lies in its inferential mechanism. By maintaining and updating probabilistic beliefs over latent congestion states, AIF continuously guides route choices using inferred network conditions rather than only recent observations, resulting in more stable route allocation, lower system costs, and smoother adaptation under both stationary and non-stationary traffic conditions.

5 Conclusion

This paper investigated whether decentralised traffic coordination can be formulated as a unified inference problem within the Active Inference (AIF) framework. Using a route-choice demonstration, we showed that AIF agents infer latent congestion conditions from partial observations, jointly optimise expected

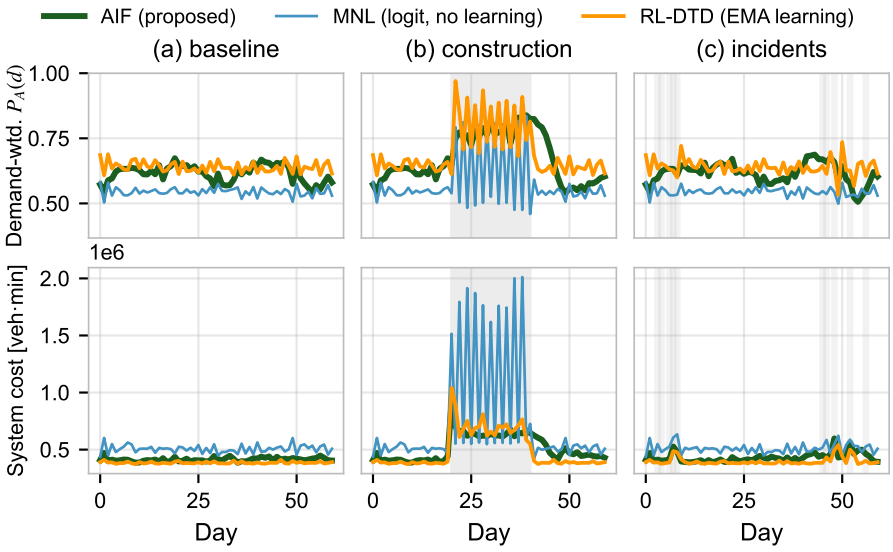


Fig. 6: Performance comparison of AIF, MNL, and RL-DTD across three scenarios: (a) baseline, (b) road construction, and (c) stochastic incidents: demand-weighted daily route share P_A (top); daily total system cost (bottom). Shaded bands mark the disruption periods, over which the three policies differ most in the stability of their route allocation. Route congestion is compared separately in Figure 11.

travel performance and uncertainty reduction through a single inference process, and adapt robustly to both persistent and stochastic disruptions. Compared with utility-based and reinforcement-learning benchmarks, AIF produced more stable traffic adaptation while maintaining interpretable probabilistic beliefs about underlying traffic conditions.

The present study serves as a conceptual demonstration of Active Inference in a simple route-choice network. Future research will extend the framework to larger networks with multiple origin–destination pairs and richer traveller heterogeneity, including habit formation in human route choice, which an explicit habit term in policy selection would capture alongside Expected Free Energy [16]. Asymmetric preference distributions, representing travellers' sensitivity to delay more faithfully than the Gaussian adopted here, are a further natural extension (Appendix B). More broadly, the results suggest that Active Inference provides a promising foundation for uncertainty-aware decentralised traffic coordination.

Supplementary Materials

The source code and an interactive simulator are provided as supplementary material.

- **Source code:** https://github.com/bvdmitri/iwai-2026-marimo-experiments. The repository contains the full simulator, the inference engine, the environment model, and the marimo notebooks that reproduce every figure in this paper, together with a headless smoke-test harness.
- **Interactive experiments:** https://iwai2026.experiments.bvdmitri.me/. The notebooks are served as a read-only web application in which every hyperparameter is exposed as a slider and the simulation re-runs live, so that readers can reproduce the reported regimes and probe parameter sensitivity without a local installation.

Acknowledgments. This research was conducted as part of the AI Learning Initiative for Multi-modal Traffic and Transportation (AiMTT). The authors gratefully acknowledge the support of the AiMTT consortium and its partners from academia, industry, and government for fostering interdisciplinary collaboration and knowledge exchange. The AiMTT project is (partly) financed by the National Growth Fund programme AI Coalition for the Netherlands (AIC4NL). Funded by the European Union — NextGenerationEU.

A Generative Model as a Forney-Style Factor Graph

The generative model of (1) admits an equivalent graphical representation as a Forney-style factor graph (FFG). In an FFG, edges represent variables and nodes represent the factors of the joint density; an equality (=) node duplicates a single variable that is shared by several factors. Figure 7 shows the FFG of (1) unrolled over three consecutive day slices, making explicit how route-choice actions enter as causes of congestion evolution and how each latent state connects the congestion-evolution model and the observation model.

Transition factor f. The transition factor encodes the agent's internal model of how route-choice actions influence future congestion dynamics,

$$f \;\equiv\; p(x_d \mid x_{d-1}, a_d) \;=\; \mathcal{N}\big(x_d \,\big|\, \mu_f(x_{d-1}, a_d),\, Q\big), \tag{13}$$

acting on the per-route latent triple $x_{d,k} = (L_{r_k}, C_{r_k}, F_{r_k})$ of §3.3. The mean μ_f is given by the store-and-forward queue dynamics of eq. (1)'s parameterisation: the queue component L_{a_d} of the chosen route a_d evolves according to the queue update $L_r(d, t{+}1) = \max(0,\, L_r(d,t) + (Q_r(d, t{-}N_r) - C_r)\Delta t)$, where the action a_d selects which route inflow Q_r is driven, while the near-static free-flow time F_{r_k} and capacity C_{r_k} are propagated by a slow random walk. The process covariance Q realises this slow drift; in the implementation it corresponds to the per-route, staleness-dependent drift inflation σ^2_{drift} of Appendix B (eq. (16)), so that latents the agent has not observed recently are allowed to diffuse.

Observation factor g. The observation factor maps the latent state of the chosen route to the two noisy signals the agent receives after completing the trip,

$$g \;\equiv\; p(y_d \mid x_d) \;=\; \mathcal{N}\Big(y_d \;\Big|\; \big[\, TT_{a_d}(x_d),\; L_{a_d}(x_d) \,\big],\; R\Big), \tag{14}$$

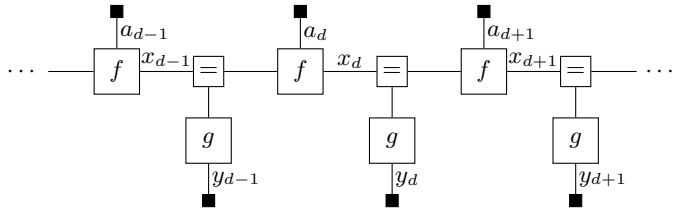


Fig. 7: Forney-style factor graph of the agent generative model (1) over three day slices. The transition factor $f \equiv p(x_d \mid x_{d-1}, a_d)$ represents congestion-state evolution conditioned on the route-choice action, while the observation factor $g \equiv p(y_d \mid x_d)$ represents the travel-time and queue-length observations generated from the latent state. The same time-invariant factors f and g are instantiated at every day slice, so the day index is carried by the variables (edges) rather than by the factors. Edges are variables; equality nodes (=) duplicate the hidden state x_d across the transition into the next day slice and the same-day observation factor. Filled terminals on the action edges denote the action priors $p(a_d)$; filled terminals on the observation edges denote the realised observations y_d.

where the noise covariance is $R = \mathrm{diag}(\sigma^2_{\mathrm{obs}}, \sigma^2_{L,\mathrm{obs}})$, the travel-time mean is $TT_{a_d} = F_{a_d} + 60\, L_{a_d}/C_{a_d}$, and the queue-length mean is L_{a_d}, matching the realised observations $y_{d,\mathrm{TT}}$ and $y_{d,L}$ of §3.3. Crucially, g reads only the chosen route a_d: the latents of the unchosen routes do not enter the day-d likelihood, which is precisely the partial observability of the route-choice setting. This is why the equality node in Figure 7 forwards the full state x_d to the next day while the observation factor depends only on the chosen-route component.

B Simulation Implementation Details

This appendix documents the numerical implementation of the proposed Active Inference framework. While the conceptual model is described in Section 3, this appendix provides the inference equations, implementation details, and hyperparameter settings used in the experiments.

B.1 Closed-form belief update

The variational perception of §3.4 is implemented as a closed-form, deterministic Gaussian smoother, so no stochastic variational inference, Monte-Carlo sampling, or inference-side pseudo-randomness is used. Each agent maintains a separate multivariate-Gaussian posterior per route over the latent triple of §3.3, $z_r = (F_r, C_r, L_r) \in \mathbb{R}^3$, parameterised by a mean vector and a lower-triangular Cholesky factor; the two routes' posteriors are independent, since each route's latents enter only its own arm of the likelihood.

Rather than a one-step sequential filter, the posterior is re-fitted each day over a *rolling window of the last W days*. All three latents are held fixed within

the window, so the same L_r explains all W queue observations; this couples the travel-time channel y_{TT} and the queue channel y_L over the whole window and yields tight joint identification of the latent variables (F_r, C_r, L_r) from data that, on any single day, only weakly constrains the pair (C_r, L_r).

The window fit is an *iterated linearised* (Laplace / extended-Kalman) update. The travel-time forward map $h(F, C, L) = F + 60\,L/C$ is linearised about the current posterior mean with the analytic Jacobian

$$\nabla h \;=\; \Big[\,1,\;\; -\tfrac{60\,\mu_L}{\mu_C^2},\;\; \tfrac{60}{\mu_C}\,\Big], \tag{15}$$

while the queue channel y_L is already linear in L. The W daily observations are folded into the prior by sequential rank-one Kalman updates, and the linearisation point is refreshed by repeating the whole pass n_{Laplace} times (Table 1). Because the update is closed-form, repeated runs with identical seeds are bitwise-reproducible.

Between-window prior and unobserved routes. At each window boundary the smoother carries forward the previous window's posterior *means* for F_r, C_r, L_r but resets the covariance to the cohort prior, so the posterior never collapses and the agent relearns a new regime within $\sim W$ days after a change. The reset covariance is then inflated per route according to how stale the agent's last observation of that route is,

$$\sigma^2_{\mathrm{eff}}(i,r) \;=\; \sigma^2_{\mathrm{cohort}} \;+\; n_{\mathrm{stale}}(i,r)\,\sigma^2_{\mathrm{drift}}, \tag{16}$$

applied separately to F, C and L with the drift scales of Table 1, where $n_{\mathrm{stale}}(i,r)$ is the number of days since agent i last chose route r. This both models potential changes in the environment during periods without observations and feeds the ambiguity term of the EFE, encouraging re-exploration of neglected routes. For a route absent from the window the Kalman gain is zero everywhere and the posterior equals the inflated prior exactly. Optionally, the carried-forward mean of an unobserved route is reverted toward the cohort prior mean on a back-loaded (quadratic) schedule with timescale $N = \texttt{mean_revert_days}$,

$$t = \min\!\big(n_{\mathrm{stale}}/N,\,1\big), \quad f = t^2, \quad \mu_{\mathrm{eff}} = (1-f)\,\mu_{\mathrm{carry}} + f\,\mu_{\mathrm{prior}}, \tag{17}$$

which keeps the estimate near its learnt value for most of the window and then accelerates to the prior, preventing permanent anchoring to a stale congested estimate. While the window is still filling ($d < W$) the smoother does not fire: agents act from the cohort prior and observations are merely accumulated.

B.2 Closed-form action selection

The EFE of §3.5 is also evaluated in closed form. The predictive travel-time moments under the current posterior $\mathcal{N}(\mu, \Sigma)$ over (F, C, L) are obtained from the same first-order expansion as eq. (15),

$$\mu_q(a) = h(\mu_F, \mu_C, \mu_L), \qquad \sigma^2_q(a) = \nabla h^{\top}\,\Sigma\,\nabla h. \tag{18}$$

The risk term is then the analytic Gaussian Kullback–Leibler divergence between the predicted-observation distribution and the free-flow preference of eq. (12),

$$\mathrm{risk}(a) = D_{\mathrm{KL}}\Big[\mathcal{N}\big(\mu_q(a),\, \sigma_q^2(a) + \sigma_{\mathrm{obs}}^2\big) \,\big\|\, \mathcal{N}\big(\mu_{F_a},\, \sigma_{\mathrm{pref}}^2\big)\Big], \tag{19}$$

and the ambiguity term is realised through its equivalent epistemic value, the expected information gain about the latent state,

$$\mathrm{info}(a) = \tfrac{1}{2}\log\Big(1 + \sigma_q^2(a)/\sigma_{\mathrm{obs}}^2\Big). \tag{20}$$

Route-choice probabilities follow the softmax of $G(a) = w_R\,\mathrm{risk}(a) - w_I\,\mathrm{info}(a)$ as in §3.5, with weights $w_R = w_I = 1$ in all experiments. The preferred travel time on route a is the agent's own posterior mean free-flow time μ_{F_a}, so the risk penalises expected *congestion delay* symmetrically across routes, without any hard-coded route preference.

On the choice of Gaussian preferences. The preference of eq. (12) was selected for analytic tractability rather than behavioural realism. Being conjugate to the linearised predictive density of eq. (18), it renders the risk term the closed-form Gaussian divergence of eq. (19), so that the entire perception–action loop reduces to deterministic linear algebra and remains affordable when beliefs and policies are re-evaluated for every agent, route, and day. Sampling-based estimators of the Expected Free Energy would admit arbitrary preference families, but in our exploratory implementations their cost was prohibitive at the population sizes and horizons used here. The specification has two behavioural consequences worth noting. It is symmetric about its setpoint, so arriving earlier than free-flow is penalised as strongly as an equal delay; and, more consequentially, the setpoint is route-specific, so each alternative is compared against itself rather than against a common reference. The latter is what leaves agents indifferent between routes whenever both run at free flow, as observed at the earliest departure intervals in Figure 3c: referring all alternatives to a single reference instead, such as the fastest believed free-flow time in the network, removes that degeneracy and favours the faster route. A monotonically decreasing preference over absolute travel time, for instance a shifted-exponential, Gamma, or truncated-Gaussian density, would capture travellers' asymmetric sensitivity to delay more faithfully. Since the observation channel is scalar, such preferences can be scored by one-dimensional quadrature while leaving the closed-form smoother intact, which makes them the most direct extension of the present model.

B.3 Aggregation and numerical safeguards

The route-choice proportion $P_r(d,t)$ that drives the inflow $Q_r = P_r D$ (§3.3) is the empirical fraction of agents choosing route r in each departure interval, smoothed with a centred rolling mean (width 13) and clipped to $[0.02, 0.98]$ to keep both route inflows strictly positive when nearly all agents select one route. The forward map applies the guards $\max(C, 100)$ and $\max(L, 0)$ so the

linearisation in eq. (15) stays bounded when the Gaussian posterior places mass on infeasible negative latents. Each simulation draws four independent random streams (agent construction, route choice, observation noise, demand noise) from a single seed sequence, and the first `burn_in` days are discarded before any statistic is recorded.

B.4 Experimental configuration

The within-day departure demand follows a shifted-sine profile, as shown in Figure 8, which resembles typical afternoon traffic conditions with a peak at $t = \frac{T}{2}$. Agents are assumed to share the following priors: free-flow times receive high-confidence priors (low standard deviation representing time-invariant network property), capacities receive medium-confidence priors (medium standard deviation representing potential disruptions), and queue lengths receive broad priors (high standard deviation indicating that they vary strongly with daily demand and collective route choice). Table 1 lists the values used throughout the experiments. Unless stated otherwise the baseline and disruption scenarios use a window $W = 30$ days over 125 recorded days; the benchmark comparison uses $W = 14$ over 60 days in order to make plots more readable.

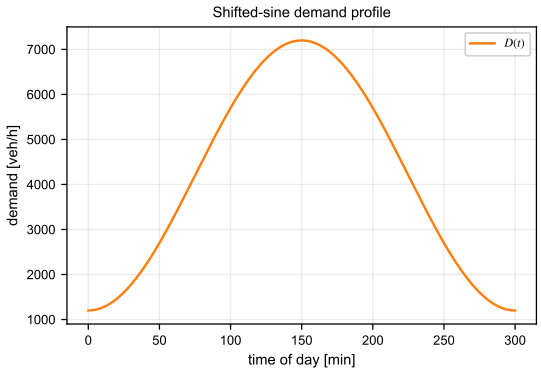


Fig. 8: The within-day departure demand follows a shifted-sine profile resembling typical afternoon traffic conditions with a peak at $t = \frac{T}{2}$.

B.5 Code availability

The full simulator and the interactive marimo notebooks that reproduce every figure in this paper are released as open source at https://github.com/bvdmitri/iwai-2026-marimo-experiments. The repository contains the inference engine, the environment model, and a headless smoke-test harness, together with notebooks for the demand profile, baseline learning, the disruption scenarios, and the benchmark comparison.

Table 1: Simulation and Active Inference hyperparameters used in the experiments (route A / route B where they differ).

Parameter	Description	Value
Network (cf. §3.3)		
F_A, F_B	Free-flow travel time	16, 18 min
C_A, C_B	Bottleneck capacity	4300, 2300 veh/h
N_A, N_B	Propagation delay $N_r = F_r/\Delta t$	16, 18 steps
Demand and timing (cf. demand profile, §3.3)		
$D_{\min}$, $D_{\max}$	Min / max within-day demand	1200, 7200 veh/h
T	Within-day intervals	300
Δt	Time-step (grid size $K = 301$)	1 min
Simulation		
N_{ag}	Agents per cohort	2000
D	Recorded days (benchmarks: 60)	125
	Burn-in days (discarded)	30
	Master random seed	42
	Route-share smoothing window	13
Priors over (F_r, C_r, L_r)		
μ_{F_A}, μ_{F_B}	Prior mean free-flow time	16, 18 min
σ_F	Prior SD free-flow time (strong)	1 min
μ_{C_A}, μ_{C_B}	Prior mean capacity	4300, 2300 veh/h
$\sigma_{C_A}, \sigma_{C_B}$	Prior SD capacity (medium)	1500, 800 veh/h
μ_{L_A}, μ_{L_B}	Prior mean queue length	100, 100 veh
σ_L	Prior SD queue length (broad)	200 veh
Observation (cf. y_{TT}, y_L, §3.3)		
σ_{obs}	Travel-time observation-noise SD	5 min
$\sigma_{L,\text{obs}}$	Queue observation-noise SD	30 veh
	Realised TT noise (environment)	1 min
Between-window drift (eq. (16))		
σ_F^{drift}	Free-flow drift SD	0.2 min/day
σ_C^{drift}	Capacity drift SD	150 veh/h/day
σ_L^{drift}	Queue drift SD	10 veh/day
N	Mean-reversion timescale (eq. (17))	60 days
Action selection (cf. §3.5)		
σ_{pref}	Delay tolerance (preference SD)	4 min
γ	Policy precision	1
w_R, w_I	Risk / info-gain weights	1, 1
Inference		
W	Rolling-window length (benchmarks: 14)	30 days
n_{Laplace}	Iterated-linearisation passes	3

B.6 Interactive online simulator

To make the experiments fully explorable, the notebooks are also served as a read-only interactive web application. Every hyperparameter in Table 1 is exposed as a slider, and the simulation re-runs live so that readers can reproduce the reported regimes and probe parameter sensitivity without a local installation, see Figure 9. The deployed simulator is available at https://iwai2026.experiments.bvdmitri.me/.

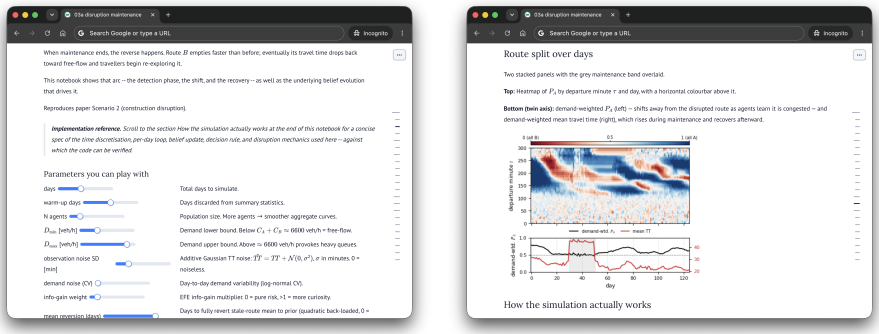


Fig. 9: The interactive online simulator. Sliders on the left reconfigure the network, demand, prior, observation, and inference parameters of Table 1; the panels on the right re-run and update the route queues, choice shares, and per-agent latent-state estimates in real time.

C Inference in Disruption-free Environment

Figure 10 illustrates the day-to-day learning dynamics of three representative AIF agents departing before, near, and after the peak. The inferred free-flow travel time (column a) and route capacity (column b) remain stable across days, reflecting their near-static nature and the comparatively confident priors placed on them. In contrast, queue-length beliefs (column c) vary substantially over time and exhibit a lag relative to the realised queues, reflecting the delayed and partial congestion observations available to each agent. Fluctuations are strongest for the peak-hour agent, who experiences the largest and most variable queues. The predictive travel-time beliefs (column d) inherit this pattern because travel time is dominated by the queue component. Consequently, day-to-day adaptation is driven primarily by inference of queue dynamics rather than the free-flow time or capacity estimates.

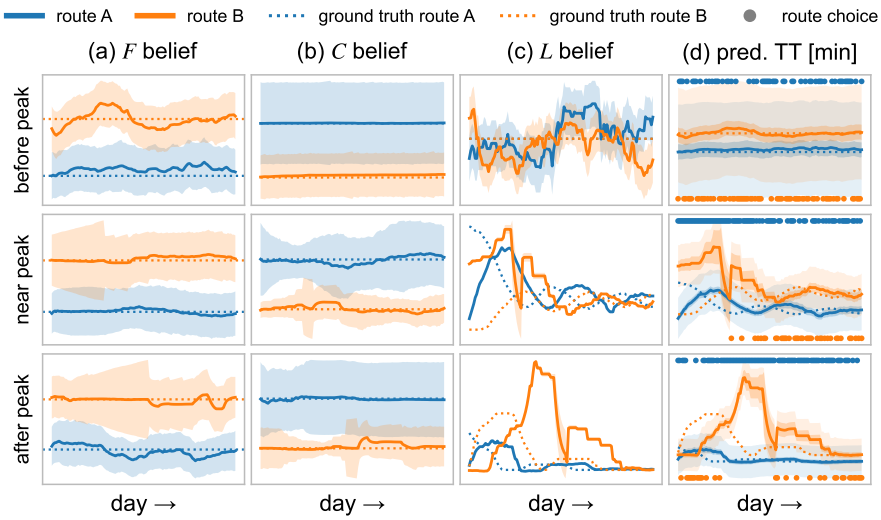


Fig. 10: Per-agent inference for three representative agents departing before, near, and after peak hours: (a) free-flow time F, (b) capacity C, (c) queue length L, and (d) predictive travel time (dark band: $\pm\sigma_{\text{post}}$, light band: observation noise, dots: daily route choice).

D Benchmark Comparison

D.1 Review of benchmark models

The benchmarks chosen in this study represent two common behavioural paradigms in transportation systems: stochastic utility-based choice and experience-driven learning. Empirical studies show both mechanisms shape real route-choice behaviour [3].

Multinomial logit (MNL) model The first benchmark is a multinomial logit (MNL) model [2], which assumes that agents probabilistically choose routes according to their perceived travel costs. Let $c_r(d)$ denote the expected travel time of route r on day d. The probability of choosing route r is

$$p_r(d) = \frac{\exp[-\beta c_r(d)]}{\sum_{k\in\mathcal{R}} \exp[-\beta c_k(d)]}, \tag{21}$$

where $\beta > 0$ is the sensitivity parameter and $\mathcal{R}$ denotes the set of available routes. Travellers are assumed to have access to expected route costs but do not explicitly represent uncertainty or infer latent traffic conditions.

Reinforcement-style day-to-day learning model (RL-DTD) The second benchmark is a reinforcement-style day-to-day learning model (RL-DTD) [25], which assumes that travellers update perceived route costs based on experienced travel times. Let $\hat{c}_r(d)$ denote the perceived cost of route r. After observing travel time $T_r(d)$, beliefs are updated according to

$$\hat{c}_r(d+1) = (1-\alpha)\hat{c}_r(d) + \alpha T_r(d), \tag{22}$$

where $\alpha \in [0,1]$ is the learning rate.

Route-choice probabilities are then determined using a logit rule,

$$p_r(d) = \frac{\exp[-\beta \hat{c}_r(d)]}{\sum_{k\in\mathcal{R}} \exp[-\beta \hat{c}_k(d)]}. \tag{23}$$

A conceptual comparison of the three route-choice frameworks of the benchmark models is provided in Table 2.

Table 2: Description of benchmark route-choice models.

Framework	Decision principle	Probabilistic	Learning	Uncertainty representation
MNL [2]	Utility maximisation	✓	×	×
RL-DTD [25]	Reward adaptation	✓	✓	×
AIF (proposed)	Free-energy minimisation	✓	✓	✓

D.2 Performance comparison with benchmarks

This subsection compares congestion patterns across the three route-choice models using the daily accumulated queue lengths on Routes A and B. As shown in Figure 11, the MNL model produces an inefficient route split, with excessive queues on Route B while Route A remains only lightly congested. This imbalance in queue formation is reflected in the queue ratio and helps explain the consistently higher system costs observed in §4.4. Both learning-based approaches reduce the persistent queue imbalance observed under MNL by adapting to experienced congestion. However, RL-DTD exhibits larger day-to-day fluctuations in queue formation and occasionally produces sharp congestion spikes on certain routes, especially during the construction and incident scenarios (Figure 11b–c). These spikes suggest that, without explicitly accounting for uncertainty, RL-DTD struggles to maintain stable adaptation under changing traffic conditions. In contrast, AIF maintains a more balanced and stable distribution of congestion across both routes in all scenarios, avoiding large congestion spikes even under disruptions. This suggests that explicit uncertainty representation and the information-gain component of Expected Free Energy in AIF enable more flexible adaptation to changing traffic conditions than benchmarks.



Fig. 11: Performance comparison of AIF, MNL, and RL-DTD: Route queues.

References

1. Belletti, F., Haziza, D., Gomes, G., Bayen, A.M.: Expert level control of ramp metering based on multi-task deep reinforcement learning. IEEE Transactions on Intelligent Transportation Systems **19**(4), 1198–1207 (2017). https://doi.org/10.1109/TITS.2017.2725912
2. Ben-Akiva, M.E., Lerman, S.R.: Discrete choice analysis: theory and application to travel demand, vol. 9. MIT press (1985)
3. Bogers, E.A.I., Viti, F., Hoogendoorn, S.P.: Joint modeling of advanced travel information service, habit, and learning impacts on route choice by laboratory simulator experiments. Transportation Research Record **1926**(1), 189–197 (2005). https://doi.org/10.1177/0361198105192600122
4. Cantarella, G.E., Cascetta, E.: Dynamic processes and equilibrium in transportation networks: towards a unifying theory. Transportation Science **29**(4), 305–329 (1995). https://doi.org/10.1287/trsc.29.4.305
5. Da Costa, L., Parr, T., Sajid, N., Veselic, S., Neacsu, V., Friston, K.: Active inference on discrete state-spaces: A synthesis. Journal of Mathematical Psychology **99**, 102447 (2020). https://doi.org/10.1016/j.jmp.2020.102447

6. Daganzo, C.F.: The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. Transportation research part B: methodological **28**(4), 269–287 (1994)
7. Engström, J., Wei, R., McDonald, A.D., Garcia, A., O'Kelly, M., Johnson, L.: Resolving uncertainty on the fly: modeling adaptive driving behavior as active inference. Frontiers in Neurorobotics **18**, 1341750 (2024). https://doi.org/10.3389/fnbot.2024.1341750
8. Friston, K.: The free-energy principle: a unified brain theory? Nature Reviews Neuroscience **11**(2), 127–138 (2010). https://doi.org/10.1038/nrn2787
9. Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., Pezzulo, G.: Active inference: a process theory. Neural computation **29**(1), 1–49 (2017). https://doi.org/10.1162/NECO_a_00912
10. Friston, K., Rigoli, F., Ognibene, D., Mathys, C., FitzGerald, T., Pezzulo, G.: Active inference and epistemic value. Cognitive neuroscience **6**(4), 187–214 (2015). https://doi.org/10.1080/17588928.2015.1020053
11. Hegyi, A., De Schutter, B., Hellendoorn, H.: Model predictive control for optimal coordination of ramp metering and variable speed limits. Transportation Research Part C: Emerging Technologies **13**(3), 185–209 (2005). https://doi.org/10.1016/j.trc.2005.04.004
12. Hoogendoorn, S.P.: Optimal control of dynamic route information panels. IFAC Proceedings Volumes **30**(8), 399–404 (1997). https://doi.org/10.1016/S1474-6670(17)43856-0
13. Millidge, B., Tschantz, A., Buckley, C.L.: Whence the expected free energy? Neural Computation **33**(2), 447–482 (2021). https://doi.org/10.1162/neco_a_01354
14. Nozari, S., Krayani, A., Marin, P., Marcenaro, L., Martín Gómez, D., Regazzoni, C.: Exploring action-oriented models via active inference for autonomous vehicles. EURASIP Journal on Advances in Signal Processing **2024**, 92 (2024). https://doi.org/10.1186/s13634-024-01173-9
15. Parr, T., Friston, K.J.: Generalised free energy and active inference. Biological cybernetics **113**(5), 495–513 (2019). https://doi.org/10.1007/s00422-019-00805-w
16. Parr, T., Holmes, E., Friston, K.J., Pezzulo, G.: Cognitive effort and active inference. Neuropsychologia **184**, 108562 (2023). https://doi.org/10.1016/j.neuropsychologia.2023.108562
17. Roughgarden, T.: Selfish routing and the price of anarchy. MIT press (2005)
18. Schumann, J.F., Engström, J., Wei, R., Liu, S.Y., Kober, J., Zgonnikov, A.: Resolving space-sharing conflicts in road user interactions through uncertainty reduction: An active inference-based computational model. arXiv preprint arXiv:2604.19838 (2026). https://doi.org/10.48550/arXiv.2604.19838
19. Sheffi, Y.: Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods. Prentice Hall, Englewood Cliffs, NJ (1985)
20. Toth, D., Ambroladze, G., Sundberg, E., Beikmohammadi, A., Lapkovskis, A.: Active inference for adaptive traffic signal control in noisy nonstationary iot environments. arXiv preprint arXiv:2606.13698 (2026)
21. Treiber, M., Kesting, A.: Traffic flow dynamics: data, models and simulation. Springer Science & Business Media (2012). https://doi.org/10.1007/978-3-642-32460-4
22. de Vries, B.: Active inference for physical ai agents–an engineering perspective. arXiv preprint arXiv:2603.20927 (2026). https://doi.org/10.48550/arXiv.2603.20927
23. Wardrop, J.G.: Road paper. some theoretical aspects of road traffic research. Proceedings of the institution of civil engineers **1**(3), 325–362 (1952). https://doi.org/10.1680/ipeds.1952.11259

24. Watling, D., Hazelton, M.L.: The dynamics and equilibria of day-to-day assignment models. Networks and Spatial Economics **3**(3), 349–370 (2003). https://doi.org/10.1023/A:1025398302560
25. Wei, F., Ma, S., Jia, N.: A day-to-day route choice model based on reinforcement learning. Mathematical Problems in Engineering **2014**(1), 646548 (2014). https://doi.org/10.1155/2014/646548
26. Wei, H., Xu, N., Zhang, H., Zheng, G., Zang, X., Chen, C., Zhang, W., Zhu, Y., Xu, K., Li, Z.: Colight: Learning network-level cooperation for traffic signal control. In: Proceedings of the 28th ACM international conference on information and knowledge management. pp. 1913–1922 (2019). https://doi.org/10.1145/3357384.3357902