

# Decentralized Traffic Flow Optimization Through Intrinsic Motivation

July 29, 2024

## Abstract

Traffic congestion has long been an ubiquitous problem that is exacerbating with the rapid growth of megacities. In this proof-of-concept work we study *intrinsic motivation*, implemented via the empowerment principle, to control autonomous car behavior to improve traffic flow. In standard models of traffic dynamics, self-organized traffic jams emerge spontaneously from the individual behavior of cars, affecting traffic over long distances. Our novel car behavior strategy improves traffic flow while still being decentralized and using only locally available information without explicit coordination. Decentralization is essential for various reasons, not least to be able to absorb robustly substantial levels of uncertainty.

Our scenario is based on the well-established traffic dynamics model, the *Nagel-Schreckenberg cellular automaton*. In a fraction of the cars in this model, we substitute the default behavior by empowerment, our intrinsic motivation-based method. This proposed model significantly improves overall traffic flow, mitigates congestion, and reduces the average traffic jam time.

## 1 Introduction

Traffic congestion poses a significant challenge, with its impact increasingly felt on the global economy. Among the various solutions for traffic management, centralized approaches [1–3] are typically restricted to areas with high-quality infrastructure, enabling near real-time connectivity to central servers and neighboring vehicles. This is a limiting assumption, in terms of location and of the number of vehicles.

We therefore assert that any viable scalable solution must be decentralized, eliminating the dependence on a central hub or protocol between

vehicles, and must be capable of withstanding significant levels of uncertainty and noise. This opens the door to artificial intelligence approaches that hold promise in making traffic more efficient.

A priori, the drivers' individual goal would be to both safely and quickly arrive at their destination. However, there is no obvious externally-defined objective for artificial agents which would simultaneously aim to achieve the collective good of maintaining high-throughput traffic flow and mitigate the congestion level. Our proposal is to use a suitable intrinsic motivation measure to control the individual cars, which lead to the emergence of global collective behaviour with less traffic jams.

*Intrinsic motivation* has been increasingly established as a class of artificial agent behavior strategies that do not rely on a specific reward [4–9], but, inspired by biological agents, are driven by objectives internal to the agents or via their interaction with the environment [10–18]. Empowerment constitutes a paradigm dual to that of training artificial agents by externally-provided objectives in form of rewards or costs which require expert domain knowledge.

*Empowerment* measures the potential influence of the agent through its actions on its future; in practice, this may be only partially observable, so observing this influence is limited by the agent's sensors, reflecting the agent's limited environmental knowledge. Empowerment has been demonstrated in a broad range of domains including locomotion, stabilization, tool use, and others. For our best knowledge it has not been explored for decentralized traffic flow optimization, which is one of the main contributions of this work.

The present work is a proof-of-concept study of autonomous cars driven by empowerment to improve traffic flow in the collective. We employ the well-established, but highly useful abstracted cellular automaton model for traffic, the *Nagel-Schreckenberg* model [19], which we will refer to as NaSch. It is widely used as a standard model of traffic dynamics, and it captures effects of traffic self-organization, such as the spontaneous emergence of traffic jams affecting traffic over long distances. So, the default behavior is modeled by the NaSch model, our intrinsic motivation is realized via empowerment.

Our method operates in fully decentralized manner and the autonomous cars use only locally available information, without an explicit coordination or communication protocol even between neighbouring drivers. Such decentralization is essential for any practical scalable solution that ensures drivers' privacy as well as the viability of the method when communication between cars is impaired, e.g., under network communication failures or lag and without relying on compatible protocols.

We address this by replacing, for randomly selected cars, the standard NaSch rules of the car behaviour with empowerment-based rules, which operates on the individual level, as exemplified at Figure 1.

In the autonomous car, empowerment implements the drive to locally maintain or increase the car’s degrees of freedom as much as possible (in terms of local ”freedom of operation”, e.g. moving into states with more options to accelerate or brake) [16]. It will turn out that this, at the same time, manages to affect the general good favourably.

We want to emphasize that our goal is not to surpass specific state-of-the-art algorithms or to compare against a diverse set of road optimization models. We acknowledge the importance of such comparisons for full application in traffic scenarios, which we defer to future work. Our work represents the first attempt, to our knowledge, to utilize intrinsic motivation for decentralized traffic optimization. We believe that demonstrating its efficacy in such a diverse scenario is a significant and opens new research directions in hybrid models of intelligent traffic control. Our proposed

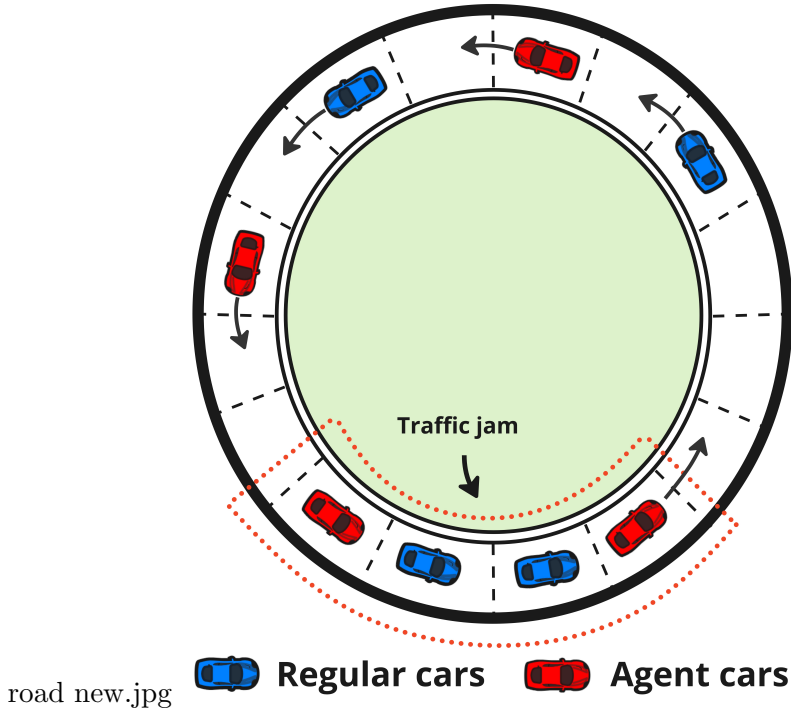


Figure 1: Intrinsically motivated artificial agents in traffic with regular cars for decentralized mitigation of traffic jams.

model demonstrates substantial improvements in traffic flow in most cases with respect to the unmodified original system, even when only a fraction of the cars has been modified to be driven by empowerment.

## 2 Prior Work

We now review existing methodologies for traffic regulation and/or congestion alleviation, specifically in light of the unique characteristics of the present work, namely: i) level of decentralization and inter-vehicle communication, ii) inherent uncertainty concerning traffic state owing to the restricted (local) information accessible to each car, and iii) the expert domain knowledge typically necessary to formulate effective reward functions for artificial agents.

The recent work by [20] on traffic flow control emphasizes the importance of decentralized autonomous control and has, to some extent, similar objectives to the current work. The model defined three types of state spaces, *minimal*, *radar* and *aggregate* differing in their information granularity. However, [20] admit a certain level of centralization and communication such as requiring global information about "the number of vehicles in the bottleneck", even in the *minimal state*, (cf., Section 3.3 in [20]). Two other state representations, *radar* and *aggregate*, require GPS information and the average car speed at various road segments, respectively. Also, the global reward function must be shared between all the agents in the network, which again requires centralization and network access.

In [21], decentralized control of traffic is achieved by explicit communication between cars via network. Specifically, when an Inter-Vehicle Communication (IVC) equipped car enters a traffic jam, a corresponding message containing jam related information (position, time) is constructed and broadcasted. However, if there is significant distance between IVC equipped cars or interference due to variations in driving speeds, the communication would become unreliable. In [22], a decentralized traffic control model is proposed where agents are distributed across the road and observe specific areas. Agents provide control commands like desired headway and lane choices to Autonomous Vehicles (AVs) within their respective regions. The decentralized control is achieved by allowing agents' observational spaces to overlap while keeping their controlling spaces non-overlapping, facilitating an understanding of other agents' behavior without direct communication. However, communication between agents and AVs is still necessary in this model. In [23], a congestion-aware collaborative automatic cruise control is



proposed, which requires vehicle-to-everything technology (V2X) for inter-vehicle communication.

In contrast to the existing works, we propose a fully decentralized approach for traffic regulation without any type of inter-vehicle communication, with a substantial uncertainty about the overall traffic state, and without manually designed reward function. The method is demonstrated as a proof-of-concept in a standardized and widely used traffic model, the NagelSchreckenberg (NaSch) cellular automaton model [19, 24–26]. It consists of equipping a selected subset of cars with an autonomous control strategy based on intrinsic motivations, concretely, empowerment [10, 16, 27].

### 3 Preliminaries

#### 3.1 Traffic Model

The standard NagelSchreckenberg (NaSch) model [19] which we use here is a stochastic particle hopping model used to simulate traffic flow. It matches real-life traffic patterns to a remarkable extent and predicts traffic flow phenomena with remarkable accuracy. The model captures complex phenomena observed in real traffic, including traffic jams, and it is an important tool for the analysis of traffic flow [24, 28].

In the standard NaSch model one considers a road divided into  $L$  cells and typically with periodic boundary conditions as in Figure 1. Each cell can either be occupied by one of  $N$  cars or left empty. The car density on this road is thus given by  $\rho = \frac{N}{L}$ .

**Traffic Dynamics.** In the standard NaSch model, each car has not only a location, but a positive velocity along the direction of movement. These are discrete integers ranging from  $v_{\min} = 0$ , indicating an idle car, to a predefined maximum speed  $v_{\max} > 0$  that a car can move at. The velocity of a car  $i \in \{1 \dots N\}$  in the cell  $x_i$  and at time  $t$  is denoted by  $v(x_i, t)$ .

The traffic dynamics in the NaSch model is defined by four update rules,  $R_1, \dots, R_4$ , which are applied sequentially to all the cars at each time step,  $t \in [1, \dots, T]$ , with  $T$  denoting the total number of updates (the simulation time):

$R_1$  (**Accelerate**): Increase the velocity  $v(x_i, t)$  of the car in cell  $x_i$  by 1 if it is less than the maximum velocity  $v_{\max}$ .

$$v = \min(v + 1, v_{\max}) \tag{1}$$

$R_2$  (**Brake to avoid collisions**): If the velocity  $v(x_i, t)$  of a car at cell  $x_i$  is greater than the distance,  $\Delta \doteq x_k - x_i$ , to the next car at cell  $x_k$  (i.e. if acting upon it would cause a collision with the front car), then decrease  $v(x_i, t)$  to  $\Delta - 1$ .

$$v = \begin{cases} \Delta - 1, & \text{if } v \geq \Delta \\ v, & \text{otherwise} \end{cases} \quad (2)$$

$R_3$  (**Random Brake**): Reduce the velocity  $v(x_i, t)$  of a car by 1 with a probability  $0 < p_{brake} < 1$ .

$$v = \begin{cases} v - 1, & \text{if } v > 0 \text{ \& } \mathbf{rand}() < p_{brake} \\ v, & \text{otherwise} \end{cases} \quad (3)$$

$R_4$  (**Move**): Advance each car by  $v(x_i, t)$  cells according to its velocity.

These simple rules capture plausible driver behavior, such as the desire to drive as fast as possible, while acknowledging that acceleration is limited (*Rule 1*), the inclination to avoid collisions (*Rule 2*) and the influence of stochasticity due to various factors such as inconsistent human driving behavior, weather or road conditions, as well as other external influences (*Rule 3*). Finally, (*Rule 4*) describes the forward movement of the vehicles.

Traffic simulated by these rules shows several essential features of real traffic, including regions of free traffic flow, traffic congestion, and phase transitions taking place between these two modes of movements. The validity of the core model is well-established in [29, 30]. Notably, these rules not only directly govern the behavior of individual cars, but rather give raise to intricate collective phenomena such as long-range traffic waves.

**Fundamental diagram of traffic flow.** The main quantity in the NaSch model is the traffic flow,  $j$ . It measures how many cars pass through a given spot in a given time [19]. The expected traffic flow,  $j$ , can be estimated for a finite number of simulation steps,  $T$ , and a finite number of cells,  $L$  with  $N$  number of cars, as follows:

$$j \approx \frac{1}{L} \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T v(x_i, t). \quad (4)$$

Figure 2 illustrates the dependence of the traffic flow as a function of car density, also known as the *fundamental diagram of traffic flow*. This diagram exhibits essential characteristics of traffic flow, namely a steep increase for

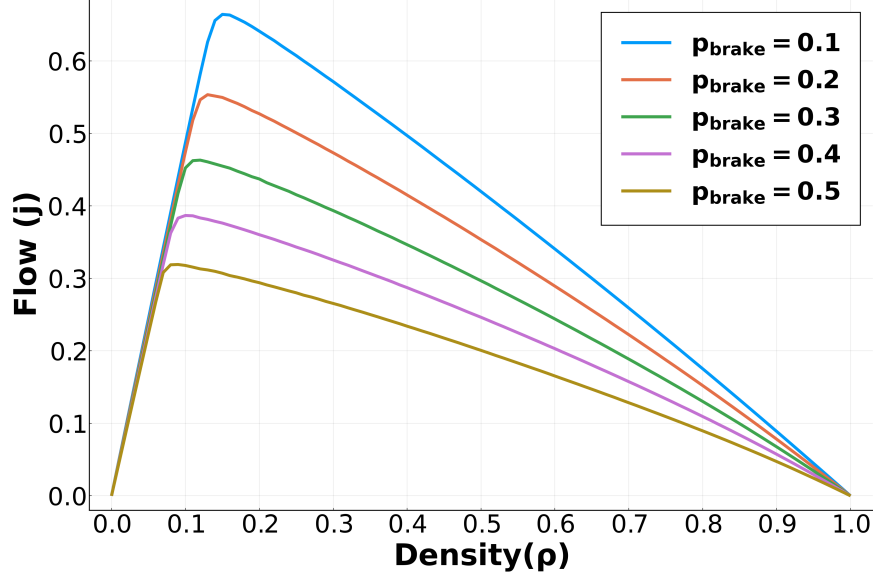


Figure 2: **Fundamental diagram** of expected traffic flow for different values of the braking probability  $p_{brake}$ .

low traffic densities, representing *free-flowing traffic*. The flow increases linearly as cars can travel at full speed without any forced deceleration. Subsequently, we see at a specific density, (the *critical density*  $\rho_c$ ), an abrupt transition to a regime,  $\rho > \rho_c$ , where average flow decreases due to reduced space, increased braking and emergence of self-organized traffic jams. This diagram is commonly used for the analysis of the traffic properties and the peak of the diagram indicates the maximally reachable traffic flow; the higher the curve lies in the fundamental diagram, the better the flow achieved<sup>1</sup>.

The objective of the current work is to develop a fully decentralized method that enables to increase the traffic flow over a broad range of car densities on the fundamental diagram by adding a small fraction of agents driven by intrinsic motivation to the traffic. These agents neither communicate with each other nor do they coordinate by explicit protocols. Instead, they follow a variation of the NaSch rules, enhanced by the *empowerment* measure.

As we will show, this is sufficient to globally improve traffic, which is

<sup>1</sup>The simulation is conducted on a road with a length,  $L$ , of 100,000 cells,  $v_{max} = 5$  and  $T = 5000$ . To ensure accurate data collection, the initial 1000 time steps are disregarded to allow transient effects to stabilize. Subsequently, data is collected at every 5<sup>th</sup> time step to reduce correlations between consecutive time steps.

characterized by a new *augmented fundamental diagram* wherein the peak is shifted upwards and towards the right.

### 3.2 Intrinsic Motivation by Empowerment

We propose to cast the problem of mitigation of traffic congestion as a problem of achieving *traffic flow viability and self-maintenance* for the individual car agent: traffic jams decrease this viability and harm autonomous flow maintenance for that care. Concretely, we suggest empowerment as a suitable candidate for the intrinsic objective for the mitigation of traffic congestion. It has been demonstrated by various studies to provide artificial agents with the desired properties of viability and drives towards self-maintenance [10,31–33], remarkably often coinciding with behaviors induced by explicit hand-designed reward functions.

Formally, one characterizes empowerment as the *channel capacity* between a potential sequence of agent actions and the (possibly limited) agent’s sensor in the future after executing this sequence [34]. Empowerment is defined for a given time horizon of actions. Concretely, for a  $n$ -step action sequence  $A^n$ , define  $n$ -step empowerment as<sup>2</sup>:

$$\mathcal{E}^n(s_t) = \max_{p(A_t^n | s_t)} I[S_{t+n}; A_t^n | s_t], \quad (5)$$

where  $I[S_{t+n}; A_t^n | s_t]$  is the mutual information between the distribution of final states,  $S_{t+n}$ , and that of the action sequence,  $A_t^n$ , conditioned on the specific current state,  $s_t$ . Stated differently, it is the channel capacity between  $A_t^n$  and  $S_{t+n}$ . We emphasize that  $A_t^n$  is a distribution over *potential* actions, not the action actually taken by the agent — the latter is selected to move into a state with maximum such channel capacity, i.e. one maximizing the term from (5).

The essential properties of empowerment are summarized as follows. When the agent’s set of actions only lead to outcomes that are similar or closely aligned, the agent’s capacity to exert influence is constrained, reflected through low values of empowerment. Conversely, when the available action spectrum results in a wide range of potential distinguishable states, the agent’s ability to shape its environment is considered stronger; this corresponds to larger empowerment values. In our traffic scenario, when a car agent is trapped closely behind e.g. a slow-moving car, and there are not

---

<sup>2</sup>We denote random variables by capital letters, and their particular realisations by small letters.

many different state outcomes the current agent can impose, empowerment will be low [16] and vice versa.

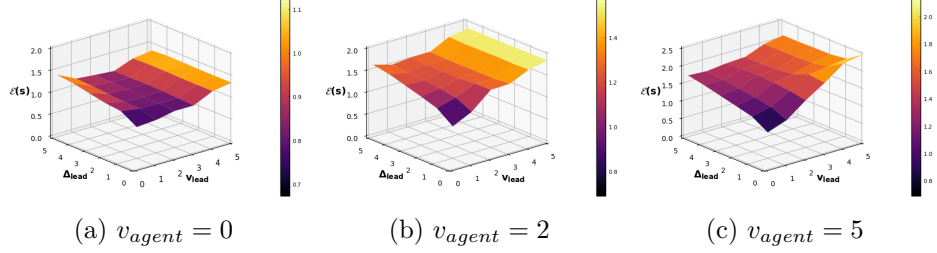


Figure 3: 3-step empowerment,  $\mathcal{E}^3(s)$ , in bits for  $p_{brake} = 0.2$  and  $\rho = 0.2$

## 4 Proposed Method

Here we present our novel approach to employ empowerment within the context of the NaSch model and to investigate its effect on the traffic dynamics. We equip certain cars with non-default decision-making capabilities throughout the simulation. These agent cars are not bound by NaSch rules; instead, they select velocities based on maximizing their empowerment. We hypothesize that when the cars aim towards states with higher empowerment, this enhances the viability of traffic and decreases traffic jams.

Starting in a state  $s_{t-1}$  (we describe the state in more detail below), each agent computes (6) for the possible states  $s_t$  it can reach, subject to NaSch dynamics, and then takes the action that moves it into the state with the highest value of,  $\mathcal{E}^n(s_t)$ :

$$\mathcal{E}^n(s_t) = \max_{p(a_t^n | s_t)} I[S_{t+n}; A_t^n | s_t] \quad (6)$$

subject to NaSch dynamics.

Given that a particular car's velocity on the road is primarily influenced by its leading car's velocity, we focus solely on the local information with respect to the leading car. The state  $s$  of an agent thus encompasses only locally observable information such as the distance between the agent and its leading car and the velocity of the leading car  $s = (\Delta_{lead}, v_{lead})$ . With this local information, agents strive to make decisions, particularly (choosing velocity, with the greedy goal of stepping into a state of as high empowerment as possible.

We consider choosing a velocity as an action  $a$ , and a sequence of actions  $a^n$  representing consecutive actions taken by such an agent over  $n$  time steps. Agents have the capacity to increase their velocity by 1, while they can decrease it by any arbitrary amount. Consequently, the one-step actions available to the agent are  $a_t \in \{0, \dots, \min(v_{agent}(t) + 1, v_{max})\}$ . Furthermore, we assume that each agent has a full model of the system rules, including the probability of braking ( $p_{brake}$ ) and the maximally permissible velocity ( $v_{max}$ ) on the road. Additionally, agents assume their neighboring cars follow traditional NaSch update rules. However, the empowered agents still operate under constraints such as the ability to increase their velocity only by one unit or having to decelerate to avoid collisions.

The empowerment horizon, the duration for which the potential future action sequences are probed is a crucial parameter of the simulation which we set at the outset and which remains constant throughout the simulation. For each experiment, the overall braking probability  $p_{brake}$  and density  $\rho$  is specified. Given this, the transitional probability of velocities for normal cars  $p(v_{t+1}|v_t)$ <sup>3</sup> is calculated. To estimate empowerment, agents need to generate the distribution of destination states ( $s_{t+n}$ ) for every available action sequence ( $a^n$ ) — we call this distribution as state transitional channel  $p(s_{t+n}|a^n, s_t)$ . With the help of  $p(v_{t+1}|v_t)$ , agents can construct  $p(s_{t+n}|a^n, s_t)$  discussed in detail in Appendix 6.1. Once  $p(s_{t+n}|a^n, s_t)$  is constructed, the empowerment of a particular state is calculated using the Blahut-Arimoto (BA) algorithm [35] for the channel  $p(s_{t+n}|a^n, s_t)$ .

Figure 3 illustrates the interplay between  $\Delta_{lead}$ ,  $v_{lead}$ , and  $v_{agent}$  in influencing the empowerment of a state  $\mathcal{E}(s)$ . The analysis indicates that larger distances to the lead car  $\Delta_{lead}$  generally improve the agent’s empowerment. However, the influence of faster velocity of the lead car  $v_{lead}$  is stronger, and this effect is more pronounced when the agent’s car itself moves faster. This observation aligns logically with the expectation that when agent is trapped behind a slow moving car, there are not many different state outcomes the agent can impose, because whatever action it takes it has to reduce the velocity to avoid collisions, resulting in low empowerment values. However, if agent is behind the fast moving car, it would have more options to accelerate and can end up in different states resulting in high empowerment values.

The main goal of the agent is to find an action that results in more empowered state. To achieve this, it assesses how empowered it could be

---

<sup>3</sup> The transitional probability of velocities for normal cars  $p(v_{t+1}|v_t)$  have been sampled independently for a given  $p_{brake}$  and vehicle density  $\rho$ , on a lane comprising 10,000 cells over the course of  $10^6$  time steps.

in various possible future situations if it takes a particular action  $a_t$  at a certain time  $t$ . There are different ways to approach this, but for the present purpose, the agent considers the probability of each outcome for the given one step action  $p(s_t|s_{t-1}, a_{t-1})$  and averages the empowerment across the outcomes  $s_t$  for a given action  $a_{t-1}$ .

$$\langle \mathcal{E} \rangle(s_{t-1}, a_{t-1}) \doteq \sum_{s_t} p(s_t|s_{t-1}, a_{t-1}) \mathcal{E}^n(s_t) \quad (7)$$

For simplicity, we omit the current state in the notation, denoting empowerment of the current state,  $s_{t-1}$ , and the action,  $a_{t-1}$ , by  $\langle \mathcal{E} \rangle(a)$ . Agents always chooses the action  $a_{t-1}$  maximizing the expected empowerment in (7). If multiple actions have same expected empowerment, the action is chosen randomly with equal probability.

---

**Algorithm 1** Expected Empowerment

---

**Input:** velocity of an agent ( $v_{agent}$ ), initial state ( $s_{t-1}$ ),  $p(v_{t+1}|v_t)$ , and  $v_{max}$

**Parameter:** Empowerment horizon,  $n$

**Output:**  $\langle \mathcal{E} \rangle(a)$

- 1: Initialize 1-step actions  $a_{t-1} \in \{0, \dots, \min(v_{agent} + 1, v_{max})\}$
  - 2: Compute  $p(s_t|s_{t-1}, a_{t-1})$
  - 3: **for**  $s$  in  $s_t$  **do**
  - 4:   Get  $n$ -step action sequences  $a^n$
  - 5:   Compute channel  $p(s_{t+n}|a^n, s)$
  - 6:   Compute Empowerment,  $\mathcal{E}^n(s)$ , by Eq. (5)
  - 7: **end for**
  - 8:  $\langle \mathcal{E} \rangle(a) \leftarrow \sum_{s_t} p(s_t|s_{t-1}, a_{t-1}) \mathcal{E}^n(s_t)$
  - 9: **return**  $\langle \mathcal{E} \rangle(a)$
- 

**Traffic Flow with Empowerment:** In the final step, we outline the procedure for calculating traffic flow in the presence of empowered agents. We begin by randomly initializing a road of length  $L$ , populated with a given number of cars  $N$ . These cars will have velocities in the range  $v \in \{0 : v_{max}\}$ . Among these cars, certain cars are considered empowered agents. Their positions are recorded in the vector  $\widehat{X}$ , while normal cars' positions are recorded in vector  $\widehat{X}_n$ . These data structures are then used for subsequent empowerment calculations. For each agent in  $\widehat{X}$  we will independently compute the expected empowerment  $\langle \mathcal{E} \rangle(a)$  and asynchronously carry out the action that maximizes  $\langle \mathcal{E} \rangle(a)$ . After the actions are chosen, agents verify whether their chosen actions result in any collisions. If so, agents adjust their actions to

---

**Algorithm 2** Traffic Flow with Empowerment

---

**Input:** position of agents ( $\widehat{X}$ ), position of regular cars ( $\widehat{X}_n$ ), road length ( $L$ ), simulation time ( $T$ ).

**Parameter:** density,  $\rho$ , braking probability  $p_{brake}$

**Output:** Flow  $j$

```
1: Let  $t = 1$ .
2: while  $t \leq T$  do
3:   for  $x$  in  $\widehat{X}$  do
4:      $v(x, t + 1) = \underset{a}{\operatorname{argmax}} \langle \mathcal{E} \rangle(a)$ 
5:      $v(x, t + 1) \leftarrow$  NaSch Rule 2
6:   end for
7:   for  $x_n$  in  $\widehat{X}_n$  do
8:      $v(x_n, t + 1) \leftarrow$  NaSch Rules 1-3
9:   end for
10:  Displace cars by NaSch Rule 4
11: end while
12:  $j \approx \frac{1}{L} \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T v(x_i, t)$ 
13: return  $j$ 
```

---

$\Delta - 1$  (NaSch rule  $R_2$ ). All the normal cars in  $\widehat{X}_n$  will follow the NaSch update rules ( $R_1, R_2, R_3$ ). Subsequently, all cars, including agents, will follow rule  $R_4$  and move forward based on their velocities. The positions all cars will be correspondingly updated and stored for the subsequent time step computations. This process is repeated for  $T$  time steps. Upon completion of the simulation, the traffic flow will be evaluated using the approximation provided in (4).

## 5 Experiments

We now study the influence of the empowered agents on traffic dynamics and especially its important global signatures, such as traffic flow and average traffic jam time. The full code repository is provided at [36].

In our investigation, we emphasize that agents rely solely on local information for decision-making. We conduct experiments with varying ratios of empowered agents relative to the total number of vehicles on the road. Figure 4 presents the fundamental diagram illustrating expected traffic flow across different densities, considering braking probabilities of  $p_{brake} = 0.2$  and  $0.5$ , alongside different empowerment planning horizons.



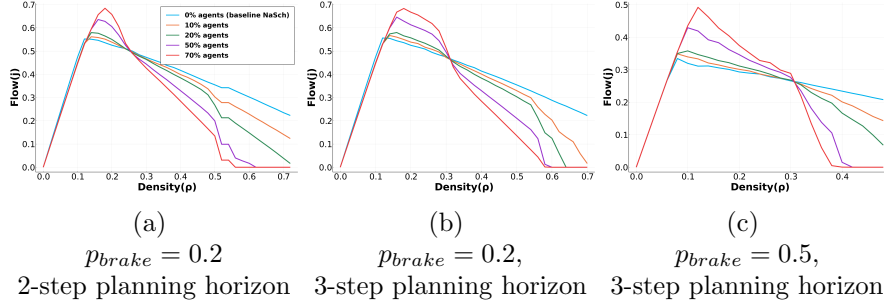


Figure 4: Comparison of traffic flow with varying ratios of empowered agents, 0% to 70%, across different vehicle densities on the x axis, considering different  $p_{brake}$  values and empowerment planning horizons. It is evident that an increase in the planning horizon expands the range of densities for which the model remains effective. As expected, with an increase in  $p_{brake}$  value, the flow decreases, but the range of densities for which the model works and the effect of empowerment remain consistent. The analysis is done on a road of  $L = 1000$ ,  $v_{max} = 5$  with  $T = 5000$ .

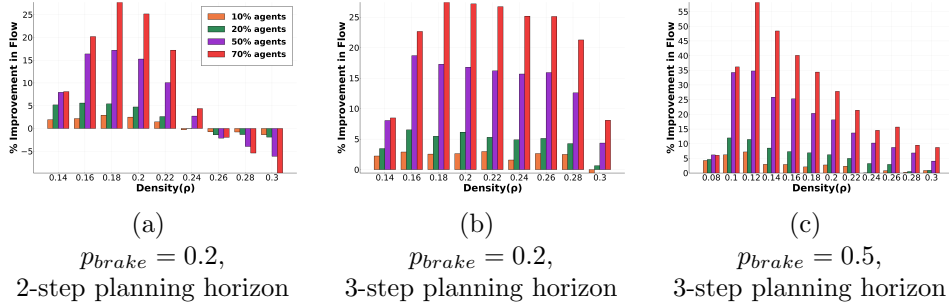


Figure 5: Comparison of the percentage (%) improvement in traffic flow for densities beyond critical density ( $\rho_c$ ) and those within the model's effective range, under varying ratios of empowered agents, different  $p_{brake}$  values, and different empowerment planning horizons.

Remarkably, as depicted in Figure 5, the inclusion of empowered agents yields substantial improvements in traffic flow, reaching approximately 27% for  $p_{brake} = 0.2$  and approximately 58% for  $p_{brake} = 0.5$ , compared to scenarios devoid of empowered agents. These enhancements are observed across a range of densities, extending beyond the critical density ( $\rho_c$ ).

As the number of agents on the road increases, we observe a corresponding increase in the maximum attainable traffic flow. Additionally, the critical

density ( $\rho_c$ ) shifts rightward, indicating that empowered agents prolong the transition from a free-flowing state to the onset of traffic congestion when compared to the standard NaSch model.

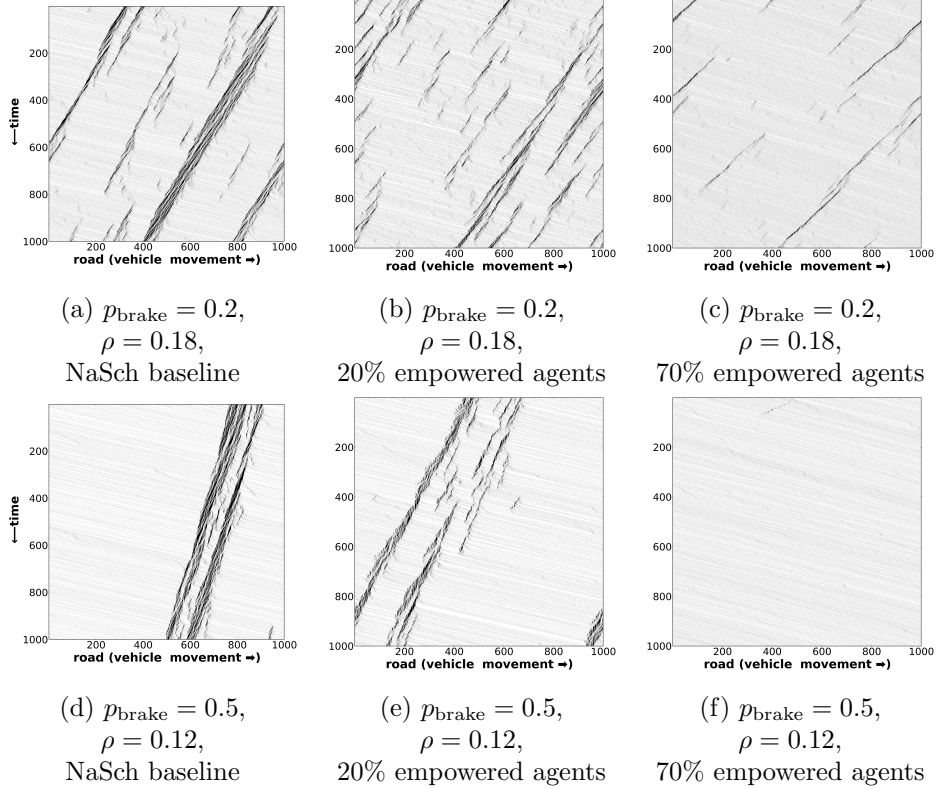


Figure 6: Spatio-temporal view of vehicle movement over time with varying  $p_{\text{brake}}$  values at densities corresponding to peak flow improvement achieved by our model with a 3-step planning horizon. Each black dot represents a vehicle moving from left to right, without distinguishing between NaSch and empowered cars. The black streaks depict traffic jam waves propagating in reverse through the traffic. Increasing the number of empowered vehicles on the road results in less dense traffic jams and quicker dissipation of traffic congestion.

The impact of empowerment is inherently local, constrained by the planning horizon. As depicted in Figure 4, extending the empowerment horizon expands the range of densities where our model maintains effectiveness, illustrating its potential to mitigate traffic congestion across broader scenarios.

The slight reduction of flow in low-traffic density (before *critical density* ( $\rho_c$ )) can be intuitively explained by the fact that a competent driver never operates at the limits of their vehicle’s performance, but leaves themselves options to react. As a consequence, they won’t drive in a maximally aggressive fashion.

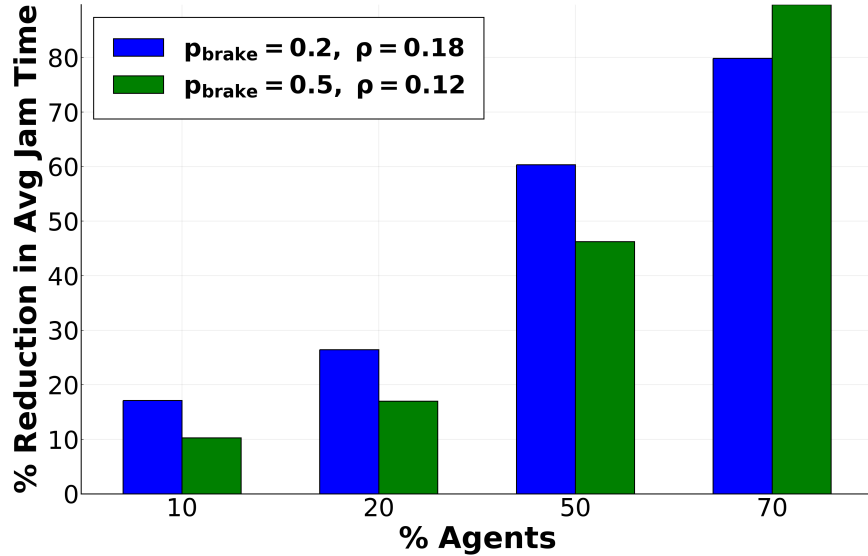


Figure 7: Percentage(%) reduction in average traffic jam time for different ratios of empowered agents compared to a baseline model with varying  $p_{\text{brake}}$  values at densities corresponding to peak flow improvement achieved by our model with a 3-step planning horizon.

For a deeper investigation into the factors driving improved flow in intermediate densities, we present the spatiotemporal diagram of vehicle movement in Figure 6. This figure offers a comprehensive view of vehicle dynamics over time, specifically focusing on densities where our model achieved peak flow improvement with a 3-step planning horizon, across different  $p_{\text{brake}}$  values. The horizontal axis depicts the road, which repeats periodically at the edges, hosting both normal cars and agents. The vertical axis indicates the time (flowing from top to bottom). The cars are represented as black dots in space and time (no distinction is made between NaSch and empowered cars). The black streaks in the diagram now indicate traffic jam waves which, as NaSch remarkably captures, flow backwards through the traffic.

The comparison of the diagrams shows that setups without agents (Fig-

ure 6a, 6d) contain more concentrated and prolonged traffic jams indicating more number of cars trapped in jammed states for longer durations. On the other hand, scenarios with empowered agents (Figure 6b, 6c 6e, 6f) result in shorter traffic jams involving fewer cars that are distributed more thinly across the road. This observation is corroborated by the significant decrease (upto  $\approx 85\%$ ) in average jam time as depicted in Figure 7, corresponding to an increase in the percentage of empowered agents. The computation of average traffic jam time in our study is similar to the methodology outlined in [28] and is discussed in detail in Appendix 6.2.

The effectiveness of empowered agents in mitigating congestion can be attributed to their adaptive braking behavior, which adjusts based on their current speed. Unlike conventional cars that brake with a fixed probability  $p_{brake}$  regardless of their speed (update *Rule 3*), empowered agents refrain from braking when traveling at low speeds, particularly when stationary ( $v = 0$ ). The constant braking of regular cars, especially when stationary, contributes to the delay in jam dissolution. As long as there is sufficient leeway in the density, the tendency of the empowered agents is to maintain their own freedom of operation appears to buffer off the fluctuations induced by the regular randomly braking drivers.

## 6 Conclusion and Future Work

We introduced and examined a decentralized strategy aimed at optimizing traffic flow. Our approach diverges substantially from traditional decentralized approaches that often require the meticulous design of reward functions tailored to the application, and instead uses a generic model for intrinsic motivation, namely *empowerment*, to produce local behavior that improves global traffic flow.

Over a range of traffic densities, empowered cars can substantially improve global traffic flow despite genuinely local decision-making requiring no joint protocols and only basic traffic rule coordination. As shown at Figure 4 (a) and (b), increasing the empowerment horizon to 3-steps extends the range of densities where empowerment improves flow. This extended range is not sensitive to the value of braking probability, as seen at Figure 4 (b) and (c).

The approach eliminates the need for explicit reward function crafting and demonstrates another promising application of generic intrinsic motivation approaches to address specific problems even in multiagent scenarios.

As a cellular automaton, the traditional NaSch model discretizes roads

and velocities. Studying a continuous limit of the NaSch model [37] will require the generalization of empowerment towards the continuum and may reveal completely novel phenomena which we defer to future work.

NaSch has been a highly successful model for the study of typical traffic phenomena; crossings and many other extensions have been introduced [38–40]. For full deployment, it will be important to scale the method developed here not only towards the continuum, but towards crossings, various types of roads, and much more.

Apart from the continuous limit, it would be of high interest to consider a hybrid model, where decentralized control between the cars (either human or agents) is additionally globally coordinated by e.g., intelligent traffic lights and other elements. This requires a hierarchical approach interleaving centralized and decentralized control at different levels of the hierarchy.

## APPENDIX

### 6.1 Estimation of State Transition Channel

To calculate empowerment, agents are required to generate the distribution of destination states ( $s_{t+n}$ ), for each available action sequence ( $a^n$ ). The agents possess knowledge of the transitional probability of the velocity of normal cars  $p(v_{t+1}|v_t)$ . For each action  $a$  in the action sequence, agents can sample the  $v_{\text{lead}}(t+1)$  (future velocity of their leading car) from  $p(v_{t+1}|v_t)$ . Subsequently, they can estimate the distance in next time step using  $\Delta_{\text{lead}}(t+1) = \Delta_{\text{lead}}(t) + v_{\text{lead}}(t+1) - a$ . This process provides the agent with the future sample state  $s_{t+1} = (\Delta_{\text{lead}}(t+1), v_{\text{lead}}(t+1))$ . Agents iterate through this procedure for each action  $a$  in the available action sequence  $a^n$ , constructing the distribution  $p(s_{t+n}|a^n, s_t)$ .

### 6.2 Average Traffic Jam Time calculation

In our computation of average traffic jam time, we followed a methodology closely aligned with the approach outlined in [28]. The definition of jam time pertains to the number of time steps during which cars remain stationary. Specifically, cars with a velocity of  $v = 0$  are identified as stationary vehicles within our traffic model.

## References

- [1] T. Yamashita, K. Izumi, and K. Kurumatani, “Car navigation with route information sharing for improvement of traffic efficiency,” in *Proceedings. The 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No. 04TH8749)*. IEEE, 2004, pp. 465–470.
- [2] Y. Chen, M. G. Bell, and K. Bogenberger, “Reliable pretrip multipath planning and dynamic adaptation for a centralized road navigation system,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 1, pp. 14–20, 2007.
- [3] J. L. Zambrano-Martinez, C. T. Calafate, D. Soler, L.-G. Lemus-Zúñiga, J.-C. Cano, P. Manzoni, and T. Gayraud, “A centralized route-management solution for autonomous vehicles in urban areas,” *Electronics*, vol. 8, no. 7, p. 722, 2019.
- [4] H. J. Charlesworth and M. S. Turner, “Intrinsically motivated collective motion,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 31, pp. 15 362–15 367, 2019.
- [5] C. Guckelsberger, C. Salge, and S. Colton, “Intrinsically motivated general companion npcs via coupled empowerment maximisation,” in *2016 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 2016, pp. 1–8.
- [6] T. D. Kulkarni, K. Narasimhan, A. Saeedi, and J. Tenenbaum, “Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation,” in *Advances in Neural Information Processing Systems, NeurIPS*, 2016.
- [7] J. Choi, A. Sharma, H. Lee, S. Levine, and S. S. Gu, “Variational empowerment as representation learning for goal-conditioned reinforcement learning,” *International Conference on Machine Learning*, pp. 1953–1963, 2021.
- [8] S. Mohamed and D. J. Rezende, “Variational information maximisation for intrinsically motivated reinforcement learning,” *Advances in neural information processing systems*, pp. 2125–2133, 2015.
- [9] K. Gregor, D. J. Rezende, and D. Wierstra, “Variational intrinsic control,” *International Conference on Learning Representations (ICLR), Workshop Track Proceedings*, 2017.

- [10] R. Zhao, P. Abbeel, and S. Tiomkin, “Efficient empowerment estimation for unsupervised stabilization,” *International Conference Learning Representations*, 2020.
- [11] A. S. Klyubin, D. Polani, and C. L. Nehaniv, “Empowerment: A universal agent-centric measure of control,” in *2005 IEEE Congress on Evolutionary Computation*, vol. 1. IEEE, 2005, pp. 128–135.
- [12] C. Salge, C. Glackin, and D. Polani, “Approximation of empowerment in the continuous domain,” *Advances in Complex Systems*, vol. 16, no. 02n03, p. 1250079, 2013.
- [13] T. Anthony, D. Polani, and C. L. Nehaniv, “General self-motivation and strategy identification: Case studies based on sokoban and pac-man,” *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 6, no. 1, pp. 1–17, 2014.
- [14] N. C. Volpi and D. Polani, “Goal-directed empowerment: combining intrinsic motivation and task-oriented behaviour,” *IEEE Transactions on Cognitive and Developmental Systems*, 2020.
- [15] C. Salge, C. Glackin, and D. Polani, “Changing the environment based on empowerment as intrinsic motivation,” *Entropy*, vol. 16, no. 5, pp. 2789–2819, 2014.
- [16] —, “Empowerment an introduction,” *Guided Self-Organization: Inception*, pp. 67–114, 2014.
- [17] D. Polani, “Information: currency of life?” *HFSP journal*, vol. 3, no. 5, pp. 307–316, 2009.
- [18] A. S. Klyubin, D. Polani, and C. L. Nehaniv, “Empowerment: A universal agent-centric measure of control,” in *2005 ieee congress on evolutionary computation*, vol. 1. IEEE, 2005, pp. 128–135.
- [19] K. Nagel and M. Schreckenberg, “A cellular automaton model for free-way traffic,” *Journal de physique I*, vol. 2, no. 12, pp. 2221–2229, 1992.
- [20] E. Vinitzky, N. Lichtlé, K. Parvate, and A. Bayen, “Optimizing mixed autonomy traffic flow with decentralized autonomous vehicles and multi-agent reinforcement learning,” *ACM Transactions on Cyber-Physical Systems*, vol. 7, no. 2, pp. 1–22, 2023.

- [21] A. Kesting, M. Schönhof, S. Lämmner, M. Treiber, and D. Helbing, “Decentralized approaches to adaptive traffic control,” in *Managing complexity: insights, concepts, applications*. Springer, 2008, pp. 189–199.
- [22] H. Maske, T. Chu, and U. Kalabić, “Large-scale traffic control using autonomous vehicles and decentralized deep reinforcement learning,” in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2019, pp. 3816–3821.
- [23] T. Kim and K. Jerath, “Congestion-aware cooperative adaptive cruise control for mitigation of self-organized traffic jams,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 6621–6632, 2021.
- [24] K. Nagel and M. Paczuski, “Emergent traffic jams,” *Physical Review E*, vol. 51, no. 4, p. 2909, 1995.
- [25] R. Barlovic, L. Santen, A. Schadschneider, and M. Schreckenberg, “Metastable states in cellular automata for traffic flow,” *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 5, pp. 793–800, 1998.
- [26] W. Staffeldt and A. K. Hartmann, “Rare-event properties of the nagel-schreckenberg model,” *Physical Review E*, vol. 100, no. 6, p. 062301, 2019.
- [27] T. Jung, D. Polani, and P. Stone, “Empowerment for continuous agent-environment systems,” *Adaptive Behavior*, vol. 19, no. 1, pp. 16–39, 2011.
- [28] K. Nagel, “Life times of simulated traffic jams,” *International Journal of Modern Physics C*, vol. 5, no. 03, pp. 567–580, 1994.
- [29] S. Lübeck, M. Schreckenberg, and K. Usadel, “Density fluctuations and phase transition in the nagel-schreckenberg traffic flow model,” *Physical Review E*, vol. 57, no. 1, p. 1171, 1998.
- [30] L. Roters, S. Lübeck, and K. Usadel, “Critical behavior of a traffic flow model,” *Physical Review E*, vol. 59, no. 3, p. 2672, 1999.
- [31] A. Sharma, S. Gu, S. Levine, V. Kumar, and K. Hausman, “Dynamics-aware unsupervised discovery of skills,” in *International*



*Conference on Learning Representations*, 2020. [Online]. Available: <https://openreview.net/forum?id=HJgLZR4KvH>

- [32] C. Salge and D. Polani, “Empowerment as replacement for the three laws of robotics,” *Frontiers in Robotics and AI*, p. 25, 2017.
- [33] Y. Du, S. Tiomkin, E. Kiciman, D. Polani, P. Abbeel, and A. Dragan, “Ave: Assistance via empowerment,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 4560–4571, 2020.
- [34] A. S. Klyubin, D. Polani, and C. L. Nehaniv, “All else being equal be empowered,” in *European Conference on Artificial Life*. Springer, 2005, pp. 744–753.
- [35] R. Blahut, “Computation of channel capacity and rate-distortion functions,” *IEEE transactions on Information Theory*, vol. 18, no. 4, pp. 460–473, 1972.
- [36] “Anonymous code repository for reproduction of the results and further experiments,” 2024. [Online]. Available: <https://anonymous.4open.science/r/Dcentralized-traffic-flow-optimization-CFBC/>
- [37] S. Krauß, P. Wagner, and C. Gawron, “Continuous limit of the nagel-schreckenberg model,” *Physical Review E*, vol. 54, no. 4, p. 3707, 1996.
- [38] N. Lakouari, O. Oubram, R. Marzoug, H. Ez-Zahraouy, J. Velásquez-Aguilar, and L. Cisneros-Villalobos, “Simulation study of traffic circle intersection with traffic lights,” *International Journal of Modern Physics C*, vol. 29, no. 07, p. 1850062, 2018.
- [39] R. Jiang, M. Hu, B. Jia, R. Wang, and Q. Wu, “The effects of reaction delay in the nagel-schreckenberg traffic flow model,” *The European Physical Journal B*, vol. 54, pp. 267–273, 2006.
- [40] M. M. Pedersen and P. T. Ruhoff, “Entry ramps in the nagel-schreckenberg model,” *Physical Review E*, vol. 65, no. 5, p. 056705, 2002.