ERC Starting Grant 2022
Research proposal [Part B2]
(not evaluated in Step 1)

COeXISTENCE

Imagine an urban mobility system where intelligent machines (e.g. connected autonomous vehicles *CAVs*) collect and share data to optimise their actions in real-time using machine learning (ML). Undoubtedly, this offers relief for a congested urban transport system which can be more efficient, generate fewer externalities (CO_2, NO_x, etc.) and pursue tight sustainability goals. Yet impacts of this technological revolution on human mobility remains largely unknown.

Urban mobility is commonly interpreted as a social game for limited resources (the capacity of transport network), converging towards user equilibrium (Wardrop 1952). Remarkably, lately AI/ML machines outperformed humans not only in simple, abstract games (L. Wang *et al.* 2020) but also in arcade games (Kaiser *et al.* 2019), action games (Vinyals *et al.* 2019) and collaborative games (Jaderberg *et al.* 2019). It seems likely that machines will outperform humans also in the urban mobility games. What's more, introducing a class of more efficient (AI-driven) players into the system might negatively impact others (humans). The more successful machine intelligence the more likely it is to identify an aggressive complex strategy, beyond the reach of naturally behaving humans. In the game for limited resources this will generate externalities, not only to humans but presumably also to the system, potentially harming the environment and violating sustainability goals.

I examined four urban mobility games in which intelligent, connected machines of the near future are likely to outperform spontaneous, self-organising, yet suboptimal human mobility. The broad term **machines** rather than referring to the automation of physical activities in transport (e.g. automated driving), denotes here cases where algorithms start making urban mobility *decisions* traditionally belonging to humans, like:

- route-choice: *which path do I take to reach my destination?*
- adaptation: *how do I improve my decisions based on experience and information?*
- mode-choice: *shall I travel by tram or car?*
- pricing: *how to charge customers for mobility services?*
- operations: *how to optimally operate the fleet of vehicles?*

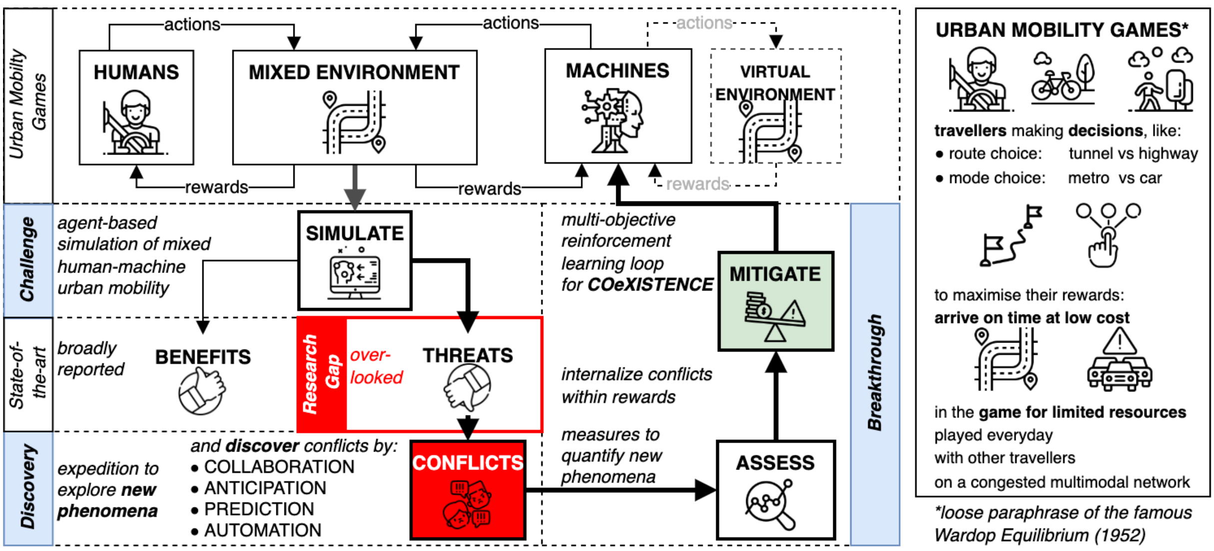


Figure 1: Synthesis: The *Urban Mobility Game* (top) will not be played by humans alone, intelligent machines – better informed, more precise in calculations and predictions, connected and collaborative, trained within the virtual environment - are entering into the game for limited resources (explained on the right and on detailed on fig.2). Studies so far demonstrated how the game benefits from machine intelligence, while potential threats are overlooked. *COeXISTENCE* embarks on the expedition to discover conflicts between humans and machines in the urban mobility games. Within the simulation environment this project will investigate when machine advantages (*collaboration, anticipation, prediction and automation*) yield conflicts. Mitigated with a novel reinforcement learning loop where machines learn to simultaneously optimise their own objectives and avoid conflicts - ultimately leading to the synergy of human-machine *COeXISTENCE*.

Section a) State-of-the-art and objectives

I came up with a series of plausible case scenarios, where machines pursuing their objectives negatively impact the traditional urban mobility of humans. The *scenario of interest* is the: Machine-dominated system, where (collective) decisions of machine intelligence improve individual performance (e.g. shorter travel times of CAVs), yet at the cost of humans, now facing e.g. longer delays, greater monetary costs or being nudged to change spontaneous travel habits into the ones desired by the system.

COeXISTENCE **objective** is to understand if such scenarios are feasible and (if they are) whether the negative impact can be mitigated. To this end, I *aim* to verify the two main hypotheses:

- **H1** intelligent machines in urban mobility games will learn to win **at the cost of humans**.
- **H2** yet, intelligent machines can learn to simultaneously reach their own goals and **mitigate conflicts**.

I propose an interdisciplinary research programme in which I will apply multi-agent simulations of urban mobility to reproduce human behaviour (travel demand models) and machine intelligence (reinforcement learning - *RL*) and let them play the game for limited resources (i.e., finite capacity of transportation system). In a diverse virtual simulation environment I will teach the RL agents to win with humans and observe outcomes of the game. A broad and deep experimental scheme will aim at revealing cases where machines benefit at the cost of humans (now facing e.g. longer travel times). By moving beyond the obvious benefits, I will focus on potential threats to understand if the key advantages of the machines: collaboration, anticipation, prediction and automation actually yield conflicts. Results will be assessed with proposed measures to quantify discovered phenomena. When conflicts are discovered and quantified they may be understood and - with the proposed novel multi-objective reinforcement loop - mitigated.

The research programme is designed to reach the following *objectives*:

A **SIMULATE** urban mobility games played between humans and machines in the virtual environment.
B **DISCOVER** conflicts arising along with introducing machine intelligence into the games.
C **ASSESS** the impact of machines with proposed conflict measures.
D **MITIGATE** conflicts with the proposed novel reinforcement loop where agents simultaneously learn to both maximise rewards and mitigate conflicts.

Significance Society seems to be not aware of the conflicts that may arise when machines enter urban mobility games. With already demonstrated capabilities of machines to win complex games, it is likely they can win in a variety of urban mobility games (as illustrated in fig.2. This, however, was not reported yet. There were no simulation studies that show how machines' strategic behaviour in pursuing their objective functions is impacting humans. Consequently, this potential negative impact is missing solid scientific underpinning. Neglecting conflicts not only puts us at risk of deploying potentially malicious AI into our cities but also prevents us from unleashing the full AI potential. Without introducing conflicts into the picture, algorithms will stop learning when reaching their (inherently selfish) objectives, instead of digging deeper, seeking for trade-off and synergy of *COeXISTENCE*.

Notably, cities are one of the last resorts for our natural, spontaneous interactions. In the digitalised world we still can naturally traverse urban space and digital technology remains only optional. Though it seems that the unavoidable success of ML/AI in urban mobility can lead to the scenario in which machines outperform us in decision making and gradually takeover control over urban mobility - making cities optimal rather than bursting, making mobility coordinated and efficient rather than spontaneous, asking humans to move to the back-seat and passively observe optimised flow of autonomous robots. Undoubtedly tempting vision: a smooth sci-fi-like flow of automated vehicles orchestrated by the central AI, internally handling the issues of efficiency (delays and congestion) and externalities (CO_2, NOx emissions, noise, accidents, etc.). Yet resembling a controllable logistic system of parcel deliveries, rather than a self-organising, diverse social system of urban mobility as we know it. <u>Such paradigm shift may dramatically change not only mobility but the entire urban landscapes of future cities</u>. Surprisingly, the demands of modern societies may be quite opposite: the 15-minute city where mobility demand is supplied with active modes (walking or cycling) via inclusive shared-spaces full of spontaneous human interactions (Moreno *et al.* 2021).

This calls for a cautious and rigorous analysis to which I want to contribute by answering the following broad questions:

- What is the future of urban mobility once intelligent machines coexist with humans?
- Will machines improve efficiency for all parties involved, or they operations need to be regulated to ensure inclusiveness, equity and synergy?

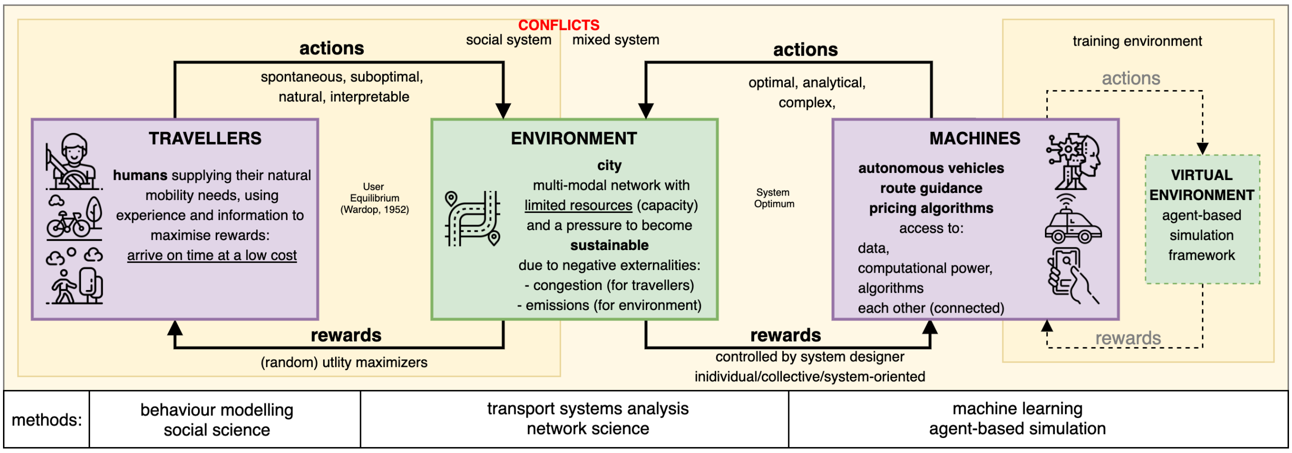


Figure 2: Playing urban mobility games with machines: Traditional, natural urban mobility (left), represented as a game, where agents (travellers) make actions (when and how to travel) and learn how the environment rewards them. As long as agents have equal access to information and cognitive capacities, the game remains in fair *user-equilibrium*. Yet when machines – better informed, more precise in calculations and predictions, connected and collaborative, trained within the virtual environment - enter the game for common resources (right) humans are likely to be outperformed and dominated. **Such conflicts, however, remain to be simulated, discovered, assessed and mitigated** - the objectives of the project.

Machines will shift the paradigm of urban mobility games - yet we still need to understand how. In particular *we still do not know* whether:

a) the intelligent machines are ready to play the games and learn to outperform us,
b) can this be simulated with available software and hardware,
c) is that true that effective agents while maximising their rewards will negatively impact others,
d) can we control this negative impact and mitigate it,
e) is there a place for synergy with the machines?

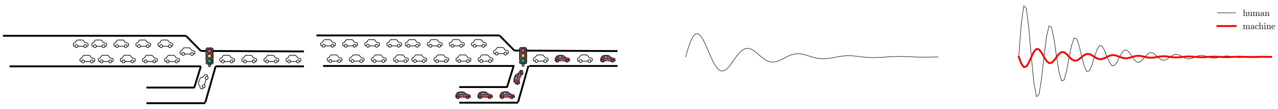
Recent findings, theoretical case studies introduced below and preliminary results suggest above, but this still has to be demonstrated and reported. Though, providing methodologically solid experimental answers is *challenging* and requires an interdisciplinary research programme spanned between deep reinforcement learning, urban mobility, agent-based models, human behaviour and adaptive complex systems. All simultaneously integrated within the already complex social system of urban mobility with non-deterministic behaviour and feedback loops.

Discovering human-machine conflicts and demonstrating them with reproducible realistic scenarios will be *ground-breaking* and will shift the paradigm of introducing AI into urban mobility from enthusiastic to cautious and methodologically solid. Consequently, the new learning paradigm where the machines playing urban mobility games care not only about their own goals but also about their external impact can be a *breakthrough* - paving a path to the synergy of humans and machines in urban mobility.

Urban mobility games Let's introduce the following four urban mobility games in which introducing machine intelligence may lead to conflicts with humans: a) the route choice game, where machines may win by collaboration, d) the day-to-day adaptation game, where machines may win by anticipation, c) the dynamic pricing game, where machines may win by prediction, and d) the repositioning game, where machines may win by automation.

With the strategic, adaptive behaviour of travellers and limited resources (finite capacity of the transportation networks), our commuting can be represented as a *game* (see fig.2). Played among humans to maximise rewards: arrive at the destination in a short time, at a low cost, with high comfort and reliability (Djavadian and Chow 2017). Despite selfish behaviour, in practice, the system typically converges to a stable state (Iryo and Watling 2019), where each agent is satisfied with her/his decisions. Empirical observations support the theoretical assumptions that travellers converge to the so-called User Equilibrium, Nash equilibrium paraphrased by (Wardrop 1952) in the context of road networks.

Since agents in the systems are humans, they are hardly controllable and act according to their (our) natural behaviour. Arguably, this game is rarely treated seriously by the players, who naturally rationalise their (our) behaviour and adapt to changes. While formally there exist a global system optimum (where total travel costs are minimal), it is never reached by utility maximising individual agents, who pay the so called price-of-anarchy for their selfish behaviour (Iryo and Watling 2019).



(a) The route-choice game: Network bottleneck (highway narrowed to a single-lane). Under user equilibrium (left) vehicles queue from the west, while the south inlet is hardly used. Yet when CAVs (red) start making **collaborative** routing decisions (right) they successfully *cheat* adaptive signal control and gain priority from the south inlet. Yielding *conflict by collaboration*: CAVs reduce their waiting times at the cost of longer queue for human-driven vehicles.

(b) The day-to-day adaptation game: Travellers adapt after a network disruption. Social system (left) where rational humans adjust their decisions stabilises smoothly after few days (Iryo and Watling 2019). CAVs learn to anticipate this process and benefit from it (right), presumably at the cost of humans (adapting now longer with stronger oscillations), yielding *conflict by anticipation*.

Figure 3: Pre-identified conflict scenarios

Those games are fair as long as all agents are equal (Cats and West 2020), e.g. have the same access to the information, cognitive capacity, experience - all likely to be violated when intelligent machines start playing the game with us. Machines hold advantages to outperform humans in the above decision-making processes. Humans are bound to rational behaviour, imperfect knowledge, own experiences, taste heterogeneity and limited cognitive capacity. While intelligent machines:

- are designed to behave optimally, i.e. use all the data and computational power to make optimal decisions;
- can **collaborate**, i.e. share information and cooperatively reach synergy;
- may understand human behaviour: **predict** it and **anticipate** our decisions;
- are **automated** and thus controllable by design;

which makes them likely to induce the conflicts in the following games:

1. The route choice game where autonomous vehicles **collaborate** to reduce their travel times at the cost of greater delays for human-driven vehicles. This game starts from the stable state, where agents are already in equilibrium - satisfied with their route choices, i.e. they would travel the next day exactly like they travel today (Kucharski and Gentile 2019a). Then, as in a plausible scenarios for the future, some human players are replaced with autonomous vehicles. As long as they are digital twins of humans and will act the same, their impact on the system is neutral. However, machines have some advantages that can be used to improve their actions. First, they treat the game seriously, unlike humans who just naturally commute, machines are designed to reach their objective functions with all the allocated computational power (which is increasingly accessible). Second, travellers can only learn by experience (typically day-to-day), while machines can explore thousands of strategies in the virtual training environment. With those advantages, CAVs are likely to somehow increase their payoffs, but still they are bounded within the user equilibrium. To break out of this they need to cooperate.

Cooperative strategies among humans are rarely observed in the urban mobility games, we do not share information, intentions and we have no direct incentives for collaboration. Contrary to the machines which, instead of being bounded by their behaviour, are controllable by the design (e.g. via formulation of the rewards in the learning process). Autonomous vehicles of the future will be connected. They will be able to share information about their location, destination and, possibly, their planned actions. When in mass, they may plan **collaboratively**, which opens a huge space for their joint actions. In particular, machines may reach critical mass to gain priority in the adaptive traffic signal control (already present in our networks) and benefit from being on the prioritised flow (see fig.3a). Notably, in the collaborative setting, the game is not zero-sum anymore, as collaborative agents may reach synergy and utilize limited resources more effectively (J. Wang, Peeta, and X. He 2019).

Consequently, when intelligent machines **collaborate** they are likely not only to outperform but also negatively impact humans in this game. Specifically when the stake is high: among the two competing CAVs providers the one offering provably shorter travel times is likely to attract a greater share of this valuable market.

2. The day-to-day adaptation game. When the time dimension is added to the previous game, another opportunity opens up for the machines. When humans face a new situation, e.g. when a new road is opened, or a metro line is closed, we first have to understand how the new system works and then iteratively adapt to the new situation (Kucharski and Gentile 2019a). This is a natural, behavioural process, both formalised (Iryo and Watling 2019) and empirically observed (Cats and West 2020). Often stated as an evolutionary game where travellers strategically swap their routes (H. Ye, Xiao, and H. Yang 2018).

Crucially, unless the system reaches equilibrium it is typically ineffectively loaded, congested in some parts and empty in another (Kucharski and Gentile 2014). When machines correctly **anticipate** the adaptation process they may learn how to benefit from it (fig. 3b). Moreover, when machine intelligence identifies that strategy of interrupting the adaptation process is beneficial, they will exploit it - presumably preventing the system to equilibrate at all. Imagine playing a rock-paper-scissors game with a machine correctly anticipating your strategies (L. Wang *et al.* 2020): when you decide to drive the highway - it is congested and the tunnel is empty; when you decide to take the tunnel - it is congested and the highway is empty.

3. The dynamic pricing game. In the above two games, the dynamics are driven by human travel behaviour and machines benefit by correctly predicting it. Similarly, the mobility service providers rely their actions on human behaviour: they play the game of offering the maximal acceptable prices. In the standard setting they set a fixed price which yields both a profit margin and acceptable market share.

With growing volumes of behavioural data available to estimate more accurate discrete choice models, rich individual behavioural profiles of travellers may be used. In particular, when a predictive algorithm correctly understands a decision-making process at the individual level it may start offering the maximal prices accepted by the individual traveller (M. K. Chen 2016; Cohen *et al.* 2016). The negative consequences of that can be twofold. First, travellers pay more, since they are offered a maximal acceptable price (R. Lu, Hong, and X. Zhang 2018). While in the typical daily commute the negative impact is limited (since attractive alternative modes upper-bound the price), in the atypical, unexpected cases it may become substantial. Imagine a case when e.g. the last bus departed and an urgent meeting starts soon. An algorithm equipped with a rich behavioural model may properly understand the circumstances and be tempted to test the limits of our price sensitivity. Without ethical guidance, such dynamic pricing mechanisms will simply exploit such cases to maximise profits (L. Chen, Mislove, and Wilson 2015).

Moreover, AI-driven predictive models may be devastating for public transport systems. The biggest mobility platforms (`Uber, Lyft, DiDi`) are so far hugely subsidised by their stakeholders, waiting for future profits (Carvalho 2020). With such huge stakes, aggressive market strategies may be explored. With precise behavioural predictions made by mobility providers, publicly subsidised public transport may start losing market shares. Potentially triggering a negative feedback loop, know as the Mohring effect (Fielbaum, Jara-Diaz, and Gschwender 2020) where public transport first loses its customers, then cut unprofitable services which, in turn, will reduce a demand - forming a vicious circle. Can it be intentionally triggered by the mobility provider, offering unprofitable services to increase market share?

4. The repositioning game. While introducing machines into the previous games negatively impacted travellers, the results of this game may be negative for the suppliers. Mobility platforms introduced a new paradigm, where instead of centrally planned operations, the fleet of independent drivers make a series of decisions on how to serve incoming requests (Kucharski and Cats 2020a). Similar to other platform revolutions such setting can be beneficial not only for the drivers (having now a source of income), but also for the clients (V. Kumar, Lahiri, and Dogan 2018), which - thanks to decentralisation and collective behaviour of individual drivers - may enjoy a resilient and competitive system (Weyl 2010). It translates to better accessibility, coverage and reliability of the system, mainly due to the competitive co-evolution of individual suppliers (drivers) (Ashkrof, Almeida Correia, Cats, and Arem 2020). However, in the near future, distributed, collective social systems are likely to be replaced with more-efficient AI-driven systems. A thorough assessment of how such a shift will impact the system performance is missing.

For instance, in the repositioning game, where drivers aim to understand the future demand and strategically reposition to be the closest one when the new trip is requested. In the game either the individual drivers build their experience or a central dispatcher repositions a fleet of vehicles based on system-wide demand predictions. Recent findings on drivers behaviour (Ashkrof, Almeida Correia, Cats, and Arem 2020) demonstrate that drivers' decisions are sub-optimal (Ashkrof, Correia, Cats, and Arem 2021) and can be easily improved using data-driven strategies, both in the centralised (Turan, Pedarsani, and Alizadeh 2020) and decentralised (Xi *et al.* 2021) settings (as reflected also in our preliminary studies - fig. 4).

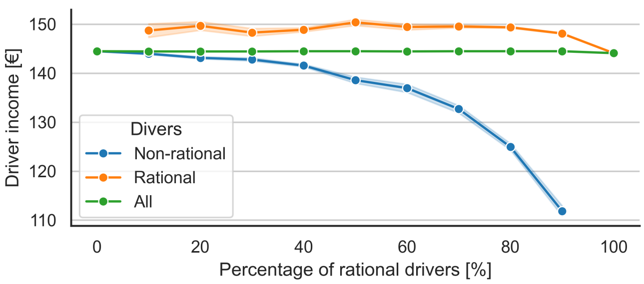


Figure 4: Preliminary results of **conflict by automation:** average incomes when using AI-driven guidance (orange) are greater than those of human drivers (acting behaviourally - blue). Master thesis experimental results of F. Ghasemi with my agent-based two-sided mobility platform simulator `https://github.com/RafalKucharskiPK/MaaSSim`, one of the workhorses of this project.

Yet, what is not clear are the downsides of switching from the distributed two-sided platform into a centrally dispatched one. By reproducing this game I want to understand wider consequences - are there some quantifiable advantages of human drivers overlooked when focusing on system-wide efficiency? If discovered, they shall be exploited to improve the system and justify the labour costs of drivers operating on the mobility platforms worldwide.

Conflicts The above four cases of games are played every day among travellers, service providers and drivers. The game is perceived as fair and travellers are in some kind of equilibrium. While being far from satisfied with their morning commute, they also do not feel cheated. Yet introducing machine intelligence may yield conflicts, machine advantages may be realised at the cost of humans. AI-driven machine operations can make them travel longer (in the 1^{st} and 2^{nd} game), pay more (3^{rd}) or lose jobs and reduce system equity and accessibility (4^{th}). At the same time, they promise significant improvements - better utilisation of scarce resources (capacity) with greater efficiency (shorter travel times) and lower externalities (emissions). That yields a second-degree dilemma between the negative impact on individuals and positive system-wide impact.

State-of-the-art

The above scenarios of our interest are more than just a futuristic sci-fi vision, quite contrary, recent breakthroughs across multiple fields make it increasingly realistic and feasible. This project leverages on recent advances across multiple fields, which on one hand make them feasible, on the other novel and challenging.

The landscape of urban mobility is rapidly changing, with ubiquitous information, sensors, traffic control, route guidance and real-time information and, last but not least, connected and autonomous vehicles (*CAVs* (Narayanan, Chaniotakis, and Antoniou 2020)). Notably, the business strategy of two-sided mobility (like Uber and Didi (S. Li, Tavafoghi, Poolla, and Varaiya 2019)) seems to rely on quick adoption of CAVs, catalysing the ongoing platform revolution, already present in our cities via shared mobility (city bikes), micro-mobility (e-scooters) or mobility-on-demand (flexible transit lines).

In parallel, advances in models and simulation frameworks allow to understand the increasing complexity of urban mobility, also in presence of mobility platforms (Kucharski and Cats 2020a) and autonomous vehicles (C. Wu *et al.* 2017). Agent-based models, detailing interaction of multiple travellers traversing multimodal transport networks in time and space to reach their destinations (Zhuge *et al.* 2021) are fuelled with travel behavioural models relying on increasingly detailed discrete-choice models formulations (Alonso-González *et al.* 2021) and increasing volumes of empirical data (Yap, Cats, and Arem 2020).

Below I synthesise state-of-the-art in the crucial fields to this interdisciplinary project: a) deep reinforcement learning (which makes machines capable to play and win urban mobility games with humans) b) human travel behaviour (now better understood and reproducible in simulations), c) AI-driven methods improving urban mobility performance (with enormous potential, but overlooked issues of transparency and ethics) and, finally, d) state-of-the-art software (simulators, models, deep learning frameworks) and hardware.

COeXISTENCE research programme building blocks
- deep reinforcement learning
- human travel behaviour
- AI-driven methods in urban mobility
- agent-based simulation of urban mobility, deep learning frameworks

a) Deep reinforcement learning Core enabler for this research is deep learning, in particular deep reinforcement learning (*RL*) which proved successful in a variety of individual, as well as collaborative tasks. Intelligent machines learning via reinforcement lately outperformed humans not only in the simple, abstract games (L. Wang *et al.* 2020), arcade games (Mnih *et al.* 2015), action games (Vinyals *et al.* 2019) and Go (Silver *et al.* 2017), but also in the cooperative games (Stooke *et al.* 2021).

To win urban mobility games a very specific setting of RL is needed. Challenging, since: a) instead of single-agent multiple agents learn simultaneously (`MARL`), b) urban mobility games are collaborative and to become effective AI agents need to cooperate, c) the environment is stochastic, mainly due to non-deterministic behaviour of travellers and finally d) the proposed mitigation framework adds another dimension to reward functions, making it multi-criteria learning. Altogether making our problems challenging, both in theory (e.g. convergence criteria of classic algorithms, like Q-learning, do not hold true) and in practice (number of learning epochs substantially increases). Luckily, many of the above challenges were already addressed.

To advocate for readiness of reproducing human-machine urban mobility games, let's refer to the *AlphaStar* (Vinyals *et al.* 2019) lately famous for reaching a GrandMaster level in the challenging real-time strategy game of StarCraft. It shares a number of similarities with our games, in particular: it uses reinforcement loops for learning, humans play with machines for limited resources and optimal strategy requires reaction to the opponent moves, knowledge is imperfect and many assumptions are based on experience and actual realisation may be different, the human behaviour (games with humans) is reproduced in a virtual environment to first learn how to play and then how to win. Arguably, winning urban mobility games can be easier: first, since professional StarCraft players take it seriously and employ all their cognitive resources and experience to win, it is hardly true for urban commuters making daily urban mobility choices. This makes opponents in our games weaker, yet also less predictable (due to big variance among heterogeneous, suboptimal travellers). Secondly, AlphaStar was successful in the huge search spaces of 10^{26}, while our players will often have just a few choices: tunnel vs. highway in the adaptation game, or accept vs. reject in the dynamic pricing game. Thirdly, AlphaStar adopted a fair strategy and handicapped machines (delaying their actions), one can fairly assume that companies introducing new technologies will use all available resources to win the competitive market.

Nonetheless, urban mobility games can be more challenging - since they are cooperative. StarCraft still can be seen as cooperative, since one central entity controls the actions of multiple units (soldiers). Which is often seen as a practical solution to convergence in the coordination problems, e.g. when CAVs are centrally routed towards a System Optimum (J. Wang, Peeta, and X. He 2019), or upper-level guided via shared knowledge (Bazzan 2019)). Similarly to our mitigation approach, where each agent receives feedback not only on his individual rewards but also on the system performance. Fortunately, the DeepMind team delivered a breakthrough also to the cooperative games. To win the cooperative Capture The Flag (played in the real-time action game of Quake III Arena) they applied a two-tier optimization process on a population of independent RL agents each learning and acting independently to cooperate and compete with other agents (Jaderberg *et al.* 2019). Similar, Multi-Agent RL (*MARL*) where multiple agents operate jointly, is applied e.g. in the autonomous robots, which (supported by recent algorithms (Alonso-Mora, Beardsley, and Siegwart 2018; Alonso-Mora, S. Baker, and Rus 2017; Z. Sun, T. Huang, and P. Zhang 2020)) learn to effectively collaborate. Via evolutionary algorithms, collective systems explore various strategies and evolve towards efficient strategies, which despite computationally challenging seem to become mature and applicable (Teixeira, d'Orey, and Kokkinogenis 2020; Wolpert and Tumer 2002; Lowe *et al.* 2017).

Solutions proposed in *COeXISTENCE* will be highly inspired by architectures winning in StarCraft and Capture the Flag (notably, W. Czarnecki, the main contributor to both those breakthroughs - cover stories in Science and Nature - is our lab's alumnus). AlphaStar applies complex architecture of various methods and solutions, apart from reinforcement learning there are various neural networks, supervised learning and multi-agent learning all hyper-parameterized to reach specific goals. Due to this lack of generalisation, successful implementation requires a substantial amount of trial and error. Luckily, while AlphaStar is IP protected at DeepMind, the vibrant ML community contributes strongly to the development ecosystem, pushing ML towards open-source, transparent, reproducible with high-level frameworks (like `Tensorflow`, `pytorch`, `RLlib`). Thus Setting up a simulation environment, despite being highly challenging looks feasible for all the games.

b) Human Travel Behaviour Machines will struggle to outperform human actions without a thorough understanding of our behaviour. Accurate behavioural models will play multiple roles in this project. First, as a background to the machine's operations in our simulation environments - realistic and empirically solid representation of how humans act and react. Secondly, conflicts will arise when machines correctly predict our behaviour. To leverage on this knowledge, machines need to first learn how do we behave, then predict it and finally rely their strategies on our behaviour. Lastly, to properly assess conflicts we need to understand how humans react, i.e. apply travel behavioural models to quantify the impact on individual travel experiences.

The picture of heterogeneous travel behaviour is becoming more detailed. Seemingly random non-deterministic human behaviour is now better observed and understood with more detailed and meaningful discrete choice models (Alonso-González *et al.* 2021; Ashkrof, Almeida Correia, Cats, and Arem 2020; Yap, Cats, and Arem 2020), including also adaptive learning behaviour (Blake *et al.* 2020) (useful to model adaptations in the day-to-day game). A solid theoretical understanding of both optimal (Iryo and Watling 2019) and behavioural (Djavadian and Chow 2017) day-to-day adaptive process, allows to reproduce how travellers adjust to the new situations (Cats and West 2020). In parallel, deep learning algorithms successfully anticipate human actions in a game-theoretical setting (L. Wang *et al.* 2020).

Human behaviour is better understood from data (Kosinski, D. Stillwell, and Graepel 2013), predicted (Stachl *et al.* 2020), controlled (Matz, Kosinski, Nave, and D. J. Stillwell 2017) and learned with RL from empirical data (X. Zhao, Yan, A. Yu, and Van Hentenryck 2020), e.g. for the case on-line shopping recommendation systems (Shi *et al.* 2019).

According to (L. Chen, Mislove, and Wilson 2015) 40% of brands use AI to tailor pricing and promotions in real time (Ezrachi 2017). The lack of transparency has led to concerns about whether mobility providers like Uber abuse behavioural data e.g. to implement personalised pricing leveraging on inferred propensity to pay more to travel a certain route at a certain time of day and charge more for that route (Mahdawi 2018).

One of the crucial aspects in adoption of CAVs relates to travel behaviour. (Pawlak, Polak, and Sivakumar 2017) advocates that travel time (the core negative component of travel (dis)utility) in presence of CAVs becomes a positive component. The premise is that if we do not have to operate a vehicle, it becomes our workplace - where the driver can both be entertained and productive. Which is used as an argument to propose a system optimum, reduce the so-called price-of-anarchy (J. Wang, Peeta, and X. He 2019) and non-coordinated selfish behaviour of agents (Bazzan 2019). Which, while appropriate for controllable systems with an aggregated pay-offs (e.g. logistic systems), seems doubtful for inherently individualistic system of urban mobility ("*I do not care if total travel time in the system is reduced if I arrive at destination 10 minutes later*"). I hypothesise that enforcing system optima in the social systems is a potential root of human-machine conflicts (Mariotte *et al.* 2021).

Finally, human behaviour is reflected in methods to solve complex social dilemmas which can be aided with multi-criteria analyses (like climate change (Hopkins *et al.* 2021)), game-theory inspired algorithms (like collective risk social dilemma (M. Kumar and Dutt 2020)) or negotiated team formation (Bachrach *et al.* 2020) (to reach goals of sustainability (Hopkins *et al.* 2021)). In (Teixeira, d'Orey, and Kokkinogenis 2020) a game-theoretical collective auctioning and bargaining induces cooperative behaviour among autonomous vehicles taking into account the individual preferences of the auction participants. On the same note, the multi-objective RL (MORL) on which I build my mitigation framework, is successfully deployed with multiple agents' goals, which allows to explore Pareto-optimality (X. Lin *et al.* 2019) or learn policies over multiple competing objectives (R. Yang, X. Sun, and Narasimhan 2019). I will exploit such approaches in conflict assessment and enforcing their mitigation.

c) Optimising transport systems through AI The enormous potential to improve the efficiency of the system has been demonstrated with a variety of technologies both already present in our daily commute, or ready to be implemented. Starting with control algorithms for traffic (C. Chen *et al.* 2020) and public transport (Moreira-Matias *et al.* 2016), via optimal route-guidance (Aradi 2020) up to trip planning (Koushik, Manoj, and Nezamuddin 2020). Benefits of Connected Autonomous Vehicles (Narayanan, Chaniotakis, and Antoniou 2020) were demonstrated on traffic flow (Van Arem, Van Driel, and Visser 2006), intersection capacity (Y. J. Zhang, Malikopoulos, and Cassandras 2016), traffic safety (L. Ye and Yamamoto 2019). The positive impact was demonstrated also at the network level (when AI is the decision-maker in the urban mobility game) via: cooperation (Chremos, Beaver, and Malikopoulos 2020; Klein and Ben-Elia 2016; Teixeira, d'Orey, and Kokkinogenis 2020), centralised optimization (Levin 2017) and deep learning (Zhou, Song, Z. Zhao, and T. Liu 2020). Recent studies of mixed environment demonstrated that well-designed algorithms improve system performance not only locally (C. Wu, Kreidieh, Vinitsky, and Bayen 2017) but also at the network-scale (J. Wang, Peeta, and X. He 2019; Lazar, Bıyık, Sadigh, and Pedarsani 2021) (relevant for our route-choice and day-to-day games).

In the context of mobility platforms, several optimisation methods solutions were proposed, e.g. for the repositioning problem. Both centralised algorithms (Turan, Pedarsani, and Alizadeh 2020) and deep RL frameworks (Xi *et al.* 2021; Holler *et al.* 2019; Shou and Di 2020) were demonstrated to increase profits of service providers, reflected also in my preliminary studies (fig. 4).

Surprisingly, within overwhelming enthusiasm, criticism of introducing AI into the urban mobility game is scarce, both theoretical scenarios and experimental results are missing. Whereas in other fields issues of *ethics, transparency, explainability and controllability* are in the spotlight - AI in mobility overlooks them. Conflicts are broadly explored at the micro-level of mobility (e.g. traffic flow stability, signal control, lane-changing and merging) both among automated vehicles (G. Lu, Nie, X. Liu, and D. Li 2019; C. Liu, C.-W. Lin, Shiraishi, and Tomizuka 2017; Stern *et al.* 2018) (including the famous ethical dilemma of CAVs tragically deciding to save pedestrian or the driver (Bonnefon, Shariff, and Rahwan 2016)) and between humans and machines (C. Huang, C. Lv, Naghdy, and Du 2020; C. Wu, Kreidieh, Vinitsky, and Bayen 2017). While, at the micro-level, apart from selfish also more social, altruistic behaviours were shown applicable for CAVs (Schwarting *et al.* 2019), at the network level, efficiency seems to be the only component of reward formulations and the externalities remain out of scope. Moreover, authors demonstrating CAVs' potential often report humans as *suboptimal* and *selfish* (Lazar, Bıyık, Sadigh, and Pedarsani 2021), or objects nudged towards the behaviour desired by the system (Levin 2017) - opposed to machines - fully compliant objects allowing to reach system optimum (J. Wang, Peeta, and X. He 2019).

With this project, I aim to contribute to the general discussion, already active in multiple other fields, on *trustworthy AI* (Floridi 2019) with issues of explainability (Adadi and Berrada 2018), fairness (Bird *et al.* 2020), inclusiveness and transparency (Gunning *et al.* 2019).

d) Software and hardware Increasing details of urban mobility picture require more advanced and robust simulation frameworks. Simulation of urban mobility is a mature and developed field, with reliable frameworks to represent: traffic flow (`SUMO` (C. Wu *et al.* 2017)), two-sided mobility platforms (our in-house `MaaSSim` (Kucharski and Cats 2020a)), demand patterns (`MATsim` (Zhuge *et al.* 2021)). The detail level needed in this project calls for an agent-based simulation (Balac and Hörl 2021), which explicitly handles individual travel behaviour and its variability, complex interactions with network and other agents.

Similarly, open-source toolkits successfully support RL development, training and benchmarking (like `Gym` from `openAI` (Stooke *et al.* 2021), `Acme` from Google (Hoffman *et al.* 2020), or `melting pot` from `DeepMind` (Leibo *et al.* 2021)). Benchmark scenarios are available for cooperative games (Lowe *et al.* 2017), traffic flow (Ault and Sharon 2021) and CAVs (Vinitsky *et al.* 2018)). Thanks to either RL-dedicated cloud servers (like AWS or Azure) or physical high-efficiency computational grids (like nVIDIA DGX Tensor Core GPU) with user-friendly job management (like `slurm`) challenging deep learning tasks and complex gird-searches on stochastic environments are increasingly feasible, also for this project.

Research gaps: While the positive impact of handing urban mobility decisions to machine intelligence is broadly reported - potential threats to human mobility remain unknown. Researchers focused on demonstrating the capabilities of proposed technologies, while downsides were out of the spotlight. I argue they are real (theoretically plausible and observable in preliminary findings) and of profound significance for future urban mobility. If overlooked, society risks ending up in a machine-dominated system not meeting our standards. While recent advances across several fields laid solid foundations making it feasible to describe, understand and simulate scenarios of interest, multiple gaps remain to be filled with at each of four stages of the research programme.

Simulation frameworks are ready to simulate the complexity of modern urban mobility systems, yet they were rarely applied to simulate the mixed environment, where human and machine agents learn simultaneously. Importantly: urban mobility games with effectively cooperating machines were not introduced (crucial for the first two games); the adaptive process was described, observed and predicted but not exploited and controlled by machines for own benefits (crucial for the adaptation game); the individual travel behaviour learning is presumably done by the biggest mobility providers, but studies of its impact are missing (crucial for the dynamic pricing game); the benefits of the two-sided distributed market are broadly reported, but consequences of abandoning the two-sided decentralised market to the centrally optimised one is missing (crucial for the repositioning game).

Despite rapid advancement in reinforcement learning, a framework where agents effectively learn from interactions with complex multi-agent environments of urban mobility games played jointly with non-deterministic humans was not proposed before. Many components of such a framework were proposed, but integrating them is an open challenge. So far, a mixed human-machine framework to simulate urban mobility games is missing. This refers both to setting the realistic simulation environment and parameterising it to ensure that machines effectively learn optimal strategies. Since urban mobility games were not introduced into reinforcement learning, there are no benchmarks and environments on the most popular RL communities like `OpenAI Gym`. Such contribution to the community is needed to facilitate advances, both to identify new conflicts and reproducible settings under which they emerge.

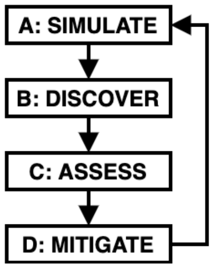
Consequently, the conflicts were not discovered in such virtual environments. Machine intelligence in urban mobility is typically simulated to demonstrate the capabilities of the technology at hand, whereas conflicts may emerge in some specific settings only. Discovering them requires a broad experimental schema in the multi-dimensional stochastic search space, which was not done before and poses a computational and theoretical challenge.

While a variety of KPIs is used to quantify the performance of the transport systems, we do not know how to measure conflicts. Presumably, the classic welfare, utility and passenger-hours will be sufficient, but this has to be demonstrated. Moreover, we ask if machines impact freedom of our mobility and spontaneity, which is hardly quantifiable - there are no measures of unconstrained mobility, thus we do not know how machines coordinated optimal actions may limit this. This is challenging and we will exploit notions of accessibility (Cats, Kucharski, Danda, and Yap 2021), equity (G. Lu, Nie, X. Liu, and D. Li 2019) or price of anarchy (Fielbaum, Kucharski, Cats, and Alonso-Mora 2021).

With such measures at hand we may try to internalise them, this was done mainly with analytical solutions (where individual agents are centrally controlled to reach global optimum) yet not in the distributed mixed environment. While some multi-objective RL frameworks were proposed they are still in their infancy and no established methods are available. Multi-agent RL in the cooperative setting is so far solved only for single-criteria, adding another dimension remains an open challenge.

Section b. Methodology

State-of-the-art on one hand made reaching the objectives of this project feasible, on the other leaves a substantial amount of challenges. To reach the project's objectives I set up the following four-step **research plan** in which:



First, I create virtual environments where humans and agents play urban mobility games,

where I experimentally discover conflicts between machines and humans,

which are assessed with the proposed measures so that

they can be integrated within agents learning process to mitigate conflicts.

With this project, I will integrate and extend existing simulation frameworks, behavioural models and machine learning algorithms to reproduce the complexity of future urban mobility. I will create virtual environments for human-machine interactions and conduct experiments aimed to discover new phenomena. Conflicts, observable via quantifiable measures will be assessed from various perspectives: humans, machines and their operators, system and policymakers. Finally, the negative impact of machines will be internalised within agents learning loops. Novel multi-objective reinforcement learning framework will enforce not only the efficiency of machines, but also synergic human-machine *COeXISTENCE*. This challenging, interdisciplinary work programme not only fills the gaps in state-of-the-art but also will lead to a breakthrough when a new phenomenon is experimentally discovered and successfully mitigated.

A: SIMULATE

where machines interact with humans within the mixed urban mobility simulation environment.

The project will be executed in the virtual environment of urban mobility - the computational workhorse of this project. Fed with realistic models of urban mobility and accurate representations of travel behaviour. I selected three open-source state-of-the-art urban mobility platforms capable to reproduce the four games: a) `FLOW` (developed at Berkeley (C. Wu *et al.* 2017)), to simulate traffic flow with RL-learning and reproduce machine's *collaboration*), b) `MATSim` (developed at ETH Zurich and TU Berlin (Zhuge *et al.* 2021)) to simulate day-to-day adaptations of travel behaviour with replanning module and reproduce machine's *anticipation*), c) `MaaSSim` (developed by me and prof. Oded Cats at TU Delft (Kucharski and Cats 2020a)) two-sided mobility platform simulator to reproduce machine's *prediction* - via dynamic pricing module - and machine's *automation* - via fleet management modules.

All of the above promising starting points to reproduce our urban mobility games. To become operational each of them needs to be introduced as a RL pipeline, where machines are implemented as intelligent agents learning how to maximise rewards received from the simulation environment (e.g. how to arrive on time or maximise profits). I will rely on state-of-the-art RL toolkits (like *Gym* from *openAI* (Stooke *et al.* 2021), *Acme* from *Google* (Hoffman *et al.* 2020), or *melting pot* from *deepmind* (Leibo *et al.* 2021)).

While standard PC is enough to simulate urban mobility, deep reinforcement learning on practical problems requires dedicated computational infrastructure (C. Wu *et al.* 2017). Whereas dynamic urban mobility scenarios typically converge in hundreds of iterations (Kucharski and Gentile 2019b) and humans adapt to the new situation in the matter of weeks (Cats and West 2020), complex RL tasks may require millions of epochs to find stable policies (Vinyals *et al.* 2019). Moreover, RL tasks are suited for GPU processing, while synchronous agent-based simulations of urban mobility rely on the CPU, which requires a dedicated computational framework. Thus, I will simulate experiments on the dedicated computational grid, using know-how and existing infrastructure in my lab GMUM, equipped with cutting edge `nVidia AI` servers (`DGX-1`). I will extend it with RL dedicated frameworks (4 x GPU A30, 256GB RAM, 128 core AMD) with strong both CPU (to simulate urban mobility environment) and GPU (for reinforcement learning).

Outcome: Dedicated simulation environment in which humans play urban mobility games with machines. Reproducible for various random seeds, input (e.g. road network, demand structure, etc.) and parameters (e.g. learning rate, value of time, pricing strategy). Wrapped in the pipeline where multiple scenarios and replications may be run in parallel, leveraging on powerful computational servers. The output of simulations will be stored in a readable format, to facilitate analysis of a great number of simulations.

B: DISCOVER

Broad and deep expedition searching for conflicts

Four separate expeditions will set out to discover conflicts in four urban mobility games. Each of them comprises multiple scenarios in the huge search space of network structures, demand patterns, behavioural parameters and reinforcement learning settings. In general, I aim to verify hypothesis **H1** and understand if the conflicts can actually arise as machines become more successful and learn to effectively exploit the environment. I expect hypothesis **H1** to be true, which means that the conflicts are triggered by machine success: as they become more effective and collaborative they will exploit the environment more aggressively and conflict with humans.

The outcome common to the four expeditions will be reproducible scenarios, where the conflict arise in a formally defined game played in a given environment (e.g. transport network), among a given population of humans and machines acting according to a given behavioural parameters and learning algorithms. Stored on a public repository to be reproduced and analysed, both within this project and beyond.

B.1. The route choice game

Objective: Demonstrate **conflict by collaboration** where machine's cooperative actions have a negative impact on humans - now facing longer travel times.

This experiment starts with an agent-based dynamic user equilibrium on a road network, where travellers iteratively adjust their route choices until they stabilise at their individually optimal routes. This part poses no challenge. I plan to use the classic `SUMO` (Krajzewicz 2010) to obtain the so-called dynamic user equilibrium, which I will use as a fixed background and gradually replace humans with autonomous vehicles. They will have the same demand (origin, destination and departure times) but instead of behavioural adaptation (Kucharski and Gentile 2019b) they will employ deep reinforcement learning.

I plan to follow an already established framework of `FLOW` where mixed human-machine operations are simulated in the microscopic scale of traffic flow (C. Wu *et al.* 2017). I can thus rely on working solutions of software architecture, communication between environment (`SUMO`) and reinforcement learning agents (`RLlib`) and their parameterisations. I will introduce the route-choice game by defining agents' possible actions (route choice set), reward formulation (utility) and learning process (discount rate). I will implement adaptive traffic control in `SUMO` since I presume tricking it will be a successful strategy for collaborative machines. When agents already effectively learn how to maximise rewards I introduce collaboration among them.

While the above steps seem to be straightforward, the collaboration poses a challenge. I will explore the variety of available collaborative MARL algorithms as an inspiration (Vinyals *et al.* 2019; Bazzan 2019; Lowe *et al.* 2017; R. Yang, X. Sun, and Narasimhan 2019) to identify how to enforce collaboration. By experimenting with communication protocols between agents, various pricing strategies and coordinated actions I will induce collaboration emergence. Starting with tractable toy-network examples (like the Braess paradox (Wolpert and Tumer 2002), where human suboptimal behaviour shall be easily tricked with collaborating machines) I will gradually move towards real-world examples.

B.2 The day-to-day adaptation game

Objective: Demonstrate **conflict by anticipation**, where machines learn our future behaviour and benefit from it - at our cost.

Like in the previous experiment, I will start with a stable traffic user equilibrium. Now I trigger a new game by dis-equilibrating the system. I close a crucial road network and force travellers to find a new equilibrium using the activity-based `MATSim` model (Balmer *et al.* 2008). Similarly to the previous game, I will gradually replace humans with intelligent machines that will learn to maximise their rewards first individually and then collectively. Here, I plan to employ the evolutionary process where consecutive generations of agents evolve towards more efficient populations (similar to e.g. (Vinyals *et al.* 2019)). First, they shall gain skills of collaboration, then start anticipating human behaviour and - finally - intentionally controlling the adaptation process for the sake of their own benefits. I presume the early stages to be somehow neutral for humans and conflicts to emerge when humans cannot equilibrate, tricked by evolved machines who successfully keep the system in an unstable state from which they benefit. Formally, collaborative actions of machines would introduce a white noise into reinforcement signals received by humans which will make their learning gradient flat (Iryo and Watling 2019), leaving humans pointless about the future directions.

B.3 The dynamic pricing game.

Objective: Demonstrate **conflict by prediction**, where mobility providers abuse individual behavioural profiles of travellers to maximise their own profits.

While the two previous games are played in the road traffic, this one is played in the multi-modal setting. Using my own MaaSSim (Kucharski and Cats 2020a) I will simulate the multi-modal scenario, where travellers select between mobility platform (like Uber) and public transport. I will use recently estimated discrete choice models for the ride-hailing options (Alonso-González *et al.* 2021; Ashkrof, Almeida Correia, Cats, and Arem

2020) to represent travellers' price sensitivity and heterogeneity. In the classic setting, the platform sets the price to maximise its profits - for which optimum can be easily found.

Here, I will assume that each individual is different and the platform operator needs to learn this heterogeneity. I will simulate a day-to-day pricing game with real-world demand patterns (e.g. for Amsterdam (Arentze, Hofman, Mourik, and Timmermans 2000)) on the synthetic, yet realistic population. The platform will learn the behaviour of individuals by testing if they accept price offers in various cases (to work, for leisure, time-critical etc.) or opt-out (select public transport or another platform instead) (Mahdawi 2018). Eventually, AI-driven operators will learn a balance between trip fares and market share (which drop when fares are high) where profits are maximised (Ezrachi 2017).

Yet, more interesting is the long planning horizon, presumably triggering the aggressive strategy, where initially unprofitable operations pay off later - when the critical mass is reached. Such strategies are common for high-tech start-ups fuelled with investors money, including mobility platforms (V. Kumar, Lahiri, and Dogan 2018). I want to simulate such scenarios and see how it impacts subsidised public transport, in particular, if AI-driven business decisions may effectively trigger a vicious circle of the so-called Mohring effect (Fielbaum, Jara-Diaz, and Gschwender 2020).

B.4. The repositioning game

Objective: Discover unrevealed advantages of sharing economy - before mobility platforms shift towards centrally dispatched automated mobility.

I use our two-sided mobility simulator (Kucharski and Cats 2020a) to reproduce how the fleet of vehicles serve the travel demand requested by travellers. Travellers request a trip and select the best among available platforms (comparing their waiting times and costs) while drivers (and platforms) start the game of repositioning - to predict the demand and be the first to serve incoming requests.

To simulate how drivers learn to reposition I will apply behavioural models estimated with empirical data (Ashkrof, Almeida Correia, Cats, and Arem 2020) and compare it with state-of-the-art centralised methods (Holler *et al.* 2019; Shou and Di 2020) (reproducible thanks to explicit algorithms).

I will simulate a variety of real-world demand patterns (e.g. from NYC (Fielbaum, Kronmueller, and Alonso-Mora 2021)), imperfectly predicted by machines due to their non-deterministic nature. I presume human drivers hold some advantages that I want to reveal. For instance, the efficiency-oriented dispatcher may restrict services to high-income zones of greater demand, leaving out outskirts and low-income zones - as we observed in our analysis of 4 million Uber trips (Cats, Kucharski, Danda, and Yap 2021).

In this game, unlike the previous ones, the experiments are designed to identify cases where human performance outperforms machines. While human drivers seem to be in a hopeless situation (easily outperformed by data-driven centralised fleet management) it is challenging to anyhow identify the added value of human operations to the system performance (equity, resilience and accessibility) and shift the paradigm of centralisation, securing not only workplaces but also added-value of sharing platform economy.

C: ASSESS

Objective: Quantify the impact of discovered conflicts on machines on humans, system and policymakers with proposed measures.

assessment
MACHINES - objective functions - rewards
HUMANS - travel time and cost - utility
POLICYMAKERS - sustainability - emmissions

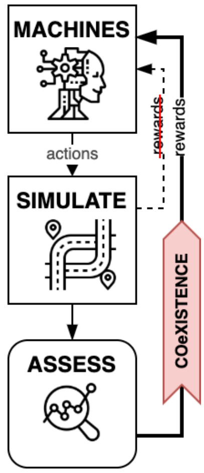
To conclude that I actually discovered conflicts proper measures are needed. Sensitive to the actual negative impact - both globally (on the system) and, more importantly, individually (on the travellers). The former is observable via aggregated, system-wide measures (e.g. total welfare) the latter is composed of utilities faced by individuals (travel time, cost, comfort, delay, reliability, etc.). While the perspective of machines is directly reflected in their objective functions, the measure covering angles of humans (rational decision-makers) and policymakers (focusing on the system-wide efficiency and sustainability) is missing. In the social systems individuals do not cancel out (Yap, Cats, and Arem 2020), thus I advocate, to measure conflicts at the strictly individual level rather than globally. Individual measures allow to understand variances, equity, price of anarchy and other factors crucial for the overall satisfaction. Ideally, covering also non-linearity (according e.g. to the Prospect Theory) and indifference band (where minor changes are insignificant). Non-linearity shall also facilitate convergence when conflicts are internalised in the next phase.

Proposing a complete set of measures is non-trivial, not only it has to reflect the actual negative impact on human individuals, but also be internalised in the reward function in the next stage of this project. Some measures will be common across the games and some will be game-specific. In all the games the system performance can be measured with the total vehicle-hours and -kilometres, passenger-hours and -kilometres. But conflict measures will vary across the games and be expressed with: longer travel times (first and second game), greater travel costs (third game), reduced accessibility and reliability (fourth game).

D: MITIGATE

Objective: To internalise the negative impact of machine operations. Now along with reaching their own goals, machines learn how to mitigate conflicts and reach *COeXISTENCE*.

Thanks to discovering the conflicts and assessing them I may propose conflict internalisation. All of the introduced games incorporate a learning loop, where agents adjust strategies focusing on their own goals. Their behaviour, despite a successful is inevitably selfish, both at the individual and collaborative levels. In the game for limited resources, it is likely to generate externalities - measured with KPIs proposed in the previous step. Not only to humans but also to the system, presumably harming the environment and violating sustainability goals.



To overcome this, I propose the following generic framework to reach *COeXISTENCE* between machines and humans in urban mobility. It operates along with learning (see figure to the right) so that agents learn to simultaneously maximise rewards and minimise the negative impact on humans. Now, along with rewards, the environment outputs also negative impacts, which can become an internal part of the reward. In such a setting the *COeXISTENCE* is not designed and controlled (which may not even be possible). Quite the contrary: similar to machine intelligence it is learned. Conflicts will enter either as a hard constrain (strictly forbidden) or as another dimension in multi-criteria analysis. The latter seems more promising as it not only allows for trade-off analysis of Pareto-optimal solutions but also to reach synergy. It allows the design flexibility when agents may learn first selfishly and then altruistically, or vice versa - depending on the learning trajectories. Importantly, such framework may be extended to cover other externalities (emissions, sustainability) to align the three perspectives simultaneously: humans, machines and policymakers. Notably, since it operates outside of the core algorithm, it allows to control even IP-protected and black-box technologies. Surprisingly, such an approach simplifies the learning convergence (Bazzan 2019) (which is a central issue in MARL) as a central level bias orchestrating non-coordinated, selfish agents. Learning trajectories may be set either to avoid conflicts at any times or (with sufficiently big computational budget) to explore strategies where initial negative impact is ultimately mitigated.

With such a solution (built on the recent promising advances in multi-criteria RL (X. Lin *et al.* 2019), hierarchical multi-agent learning (Bazzan 2019), and cooperative games (Jaderberg *et al.* 2019)) I aim to verify the hypothesis **H2** and see if internalising conflicts in the rewards can mitigate them.

Execution Part **A** is technical - where I create tools to reproduce a variety of conflicts, **B** is experimental - where I use them to discover new phenomena, **C** is methodological - where I propose measures to quantify and understand new discovery, **D** is creative - where I propose original solutions to mitigate conflicts. For each of four games there will be five *milestones*:

1. when I integrate humans and machines in a virtual environment,
2. when agents successfully learn to play the respective urban mobility game,
3. when I identify the conflict setting under which machine actions negatively impact humans,
4. when I internalise conflict measure in the learning and manage to mitigate it and
5. when conflicts are not only mitigated but also the synergy is reached.

In the experimental part (**B**) I plan four separate expeditions to discover conflicts in the four pre-identified mobility games. For efficiency, they will set off from a common ground (**A** - sharing simulation frameworks, RL pipelines, agent definitions, input data on networks and demand, etc.). Expeditions will run on shared computational resources, optimised for RL with a complex environment of urban mobility. Common input and output interfaces will facilitate grid-searches and assessment. The big datasets of experimental results will be synthesised in common formats and aggregated for assessments (**C**). Proposed mitigation frameworks (**D**), tailored and validated for each conflict separately will share a common methodological concept of multi-objective reinforcement learning.

The team will execute this plan asynchronously, gradually building simulation capabilities and deploying experiments. I presume the majority of tasks to be cooperative rather than stand-alone, to facilitate team working I plan to apply agile management techniques (like `scrum`). Starting from less challenging settings (presumably the single-agent dynamic pricing game) towards the more challenging (collaborative multi-agent setting of the route-choice games). I will first experiment with minimal settings (toy-networks and famous paradoxes) gradually moving towards real-world scenarios (big cities and actual demand patterns e.g. of Amsterdam

(Kucharski and Cats 2020b)). The interdisciplinarity at various levels requires team-working of researchers with various competences. My diverse experiences will allow me to manage this challenging project. In particular, to reproduce the first two games, I leverage on my PhD experiences in dynamic user equilibrium (Kucharski and Gentile 2019b), for the second two games I will exploit my PostDoc experiences with agent-based models of two-sided mobility platforms (Kucharski and Cats 2020a), for game-theory context I can refer to my theoretical study of ride-pooling (Fielbaum, Kucharski, Cats, and Alonso-Mora 2021), for big-data analysis I can rely on my industry experience (`NorthGravity`) and for deep learning my experiences (Cantelmo, Kucharski, and Antoniou 2020) are supported with the excellent ecosystem of the lab where I tenure (GMUM). The team with expertise in machine-learning, urban mobility and complex systems (PhDs and a PostDoc) will be led by me and backed up with urban mobility (prof. Oded Cats) and deep learning (prof. Jacek Tabor) experts.

High risk - high gain Reaching project objectives, highly significant for the future of our cities is equally challenging. It requires a broad interdisciplinary expedition into the unknown, where one can easily get lost or come back with nothing. I anticipate a series of technical hurdles in this complex research plan which, however, are mitigated with the back-up of my lab (in deep learning, reinforcement learning and computational hardware), my experience (in urban mobility, simulation and software) and well-tailored complementary skills of the interdisciplinary team to be hired. This way the simulation of urban mobility, stochasticity of human behaviour, demanding computations of reinforcement learning, convergence issues in non-smooth complex adaptive systems or rapid advances across many fields are more challenges than risks.

The true risks lay in verifying the two hypotheses of the project: **H1** discovering unknown phenomena and **H2** mitigating them - here the stake and challenges are the biggest.

plan in case of failing to:	
DISCOVER	**MITIGATE**
focus on synergy, rather than mitigation	report and understand discovered conflicts

Discovering and reproducing conflicts is far from granted. It requires a carefully tailored simulation framework, precisely reproduced urban mobility games and effectively learning, collective machines. Despite relying on solid theoretical and practical background the tailored experiments, I may fail to discover conflicts - even on the minimal examples of the famous paradoxes, with extensive directed grid-searches. In such, unlikely, case <u>I will resort to the positive side of the project: the synergy of COeXISTENCE</u>. I will use the loop designed to internalise negative impact and now focus on learning how intelligent machines may help us reach sustainability goals and improve performance for humans. Such research will be equally significant to the original plan.

Similarly, *mitigating conflicts is far from granted*. I may fail to effectively internalise conflicts, with a risk that in the already complex, multi-agent collaborative stochastic environment it may be impossible for agents to learn multiple objectives. In such a case, <u>I will resort to newly discovered conflicts and their assessment</u> - which will anyhow shift the paradigm and set the new ground for designing future urban mobility in a mixed human-machine environment.

Impact and conclusion We are at the eve of the revolution in urban mobility. Technology shall be ready anytime and will enter the market straightaway. Mobility providers will start implementing technology that remains largely a black-box. Benefits are highlighted, while potential threats are overlooked. New phenomena, discovered in quantifiable simulation experiments within *COeXISTENCE*, shall bring a scientific breakthrough giving us a methodologically solid framework to assess a variety of technologies before they enter the market. With this we can not only anticipate conflicts and mitigate them but (within the same framework) explore limits and unleash a full synergic potential responsibly.

Since both AI and sustainable urban mobility are hot public topics, findings shall resonate beyond the scientific community. This is desired, since wide social participation, based on solid scientific foundations is crucial before democratic societies will decide on the future of their cities. With a set of reproducible benchmarks, *COeXISTENCE* will set a new standard, shifting the paradigm by introducing a new dimension to the picture. Proposed AI solutions will be assessed not only based on how they perform but also how they affect others, including us - humans. I hope to set a standard to a) verify proposed ML solutions on our benchmarks b) create reproducible benchmarks for any newly identified conflicts.

The proposed RL framework to mitigate conflicts is novel and extends existing approaches where multiple agents collaborate and compete for limited resources with the multi-objective rewards. Being generically applicable beyond the direct scope of this project e.g. when we trade-off between carbon trail and cost-effectiveness in the economy, space-quality and accessibility in the urban planning, or optimal and natural behaviour in the social sciences. Proposed framework will allow a robust, generic control framework for black-box AI - contributing towards trustworthy AI. Even with agents policies hidden behind black-box and IP protected algorithms, third-party (users, policymakers, society) may assess it before allowing it to the markets.

References

Me and my collaborators appear in bold.

[1] J. G. Wardrop, “Road paper. some theoretical aspects of road traffic research.,” *Proceedings of the institution of civil engineers*, vol. 1, no. 3, pp. 325–362, 1952.

[2] L. Wang, W. Huang, Y. Li, J. Evans, and S. He, “Multi-ai competing and winning against humans in iterated rock-paper-scissors game,” *Scientific Reports*, vol. 10, no. 1, pp. 1–8, 2020.

[3] L. Kaiser, M. Babaeizadeh, P. Milos, B. Osinski, R. H. Campbell, K. Czechowski, D. Erhan, C. Finn, P. Kozakowski, S. Levine, *et al.*, “Model-based reinforcement learning for atari,” *arXiv preprint arXiv:1903.00374*, 2019.

[4] O. Vinyals, I. Babuschkin, W. M. Czarnecki, M. Mathieu, A. Dudzik, J. Chung, D. H. Choi, R. Powell, T. Ewalds, and P. Georgiev, “Grandmaster level in starcraft ii using multi-agent reinforcement learning,” *Nature*, vol. 575, no. 7782, pp. 350–354, 2019.

[5] M. Jaderberg, W. M. Czarnecki, I. Dunning, L. Marris, G. Lever, A. G. Castaneda, C. Beattie, N. C. Rabinowitz, A. S. Morcos, A. Ruderman, *et al.*, “Human-level performance in 3d multiplayer games with population-based reinforcement learning,” *Science*, vol. 364, no. 6443, pp. 859–865, 2019.

[6] C. Moreno, Z. Allam, D. Chabaud, C. Gall, and F. Pratlong, “Introducing the "15-minute city": Sustainability, resilience and place identity in future post-pandemic cities,” *Smart Cities*, vol. 4, no. 1, pp. 93–111, 2021, ISSN: 2624-6511.

[7] S. Djavadian and J. Y. Chow, “An agent-based day-to-day adjustment process for modeling ‘mobility as a service’with a two-sided flexible transport market,” *Transportation research part B: methodological*, vol. 104, pp. 36–57, 2017.

[8] T. Iryo and D. Watling, “Properties of equilibria in transport problems with complex interactions between users,” *Transportation Research Part B: Methodological*, vol. 126, pp. 87–114, 2019.

[9] **O. Cats** and J. West, “Learning and adaptation in dynamic transit assignment models for congested networks,” *Transportation Research Record*, vol. 2674, no. 1, pp. 113–124, 2020.

[10] **R. Kucharski** and **G. Gentile**, “Simulation of rerouting phenomena in dynamic traffic assignment with the information comply model,” *Transportation Research Part B: Methodological*, vol. 126, pp. 414–441, 2019.

[11] J. Wang, S. Peeta, and X. He, “Multiclass traffic assignment model for mixed traffic flow of human-driven vehicles and connected and autonomous vehicles,” *Transportation Research Part B: Methodological*, vol. 126, pp. 139–168, 2019.

[12] H. Ye, F. Xiao, and H. Yang, “Exploration of day-to-day route choice models by a virtual experiment,” *Transportation Research Part C: Emerging Technologies*, vol. 94, pp. 220–235, 2018.

[13] **R. Kucharski** and **G. Gentile**, “Indirect observation of rerouting phenomena in traffic networks–case study of warsaw bridges,” *Archives of Transport*, vol. 32, 2014.

[14] M. K. Chen, “Dynamic pricing in a labor market: Surge pricing and flexible work on the uber platform,” in *Proceedings of the 2016 ACM Conference on Economics and Computation*, ser. EC ’16, Maastricht, The Netherlands: Association for Computing Machinery, 2016, p. 455, ISBN: 9781450339360.

[15] P. Cohen, R. Hahn, J. Hall, S. Levitt, and R. Metcalfe, “Using big data to estimate consumer surplus: The case of uber,” National Bureau of Economic Research, Tech. Rep., 2016.

[16] R. Lu, S. H. Hong, and X. Zhang, “A dynamic pricing demand response algorithm for smart grid: Reinforcement learning approach,” *Applied Energy*, vol. 220, pp. 220–230, 2018.

[17] L. Chen, A. Mislove, and C. Wilson, “Peeking beneath the hood of uber,” in *Proceedings of the 2015 internet measurement conference*, 2015, pp. 495–508.

[18] M. L. N. Carvalho, “Uber, a path for profitability or a market misperception?–advanced technologiesgroup and other technology programs,” Ph.D. dissertation, 2020.

[19] **A. Fielbaum**, S. Jara-Diaz, and A. Gschwender, “Beyond the mohring effect: Scale economies induced by transit lines structures design,” *Economics of Transportation*, vol. 22, p. 100 163, 2020.

[20] **R. Kucharski** and **O. Cats**, “Maassim–agent-based two-sided mobility platform simulator,” *arXiv preprint arXiv:2011.12827*, 2020.

[21] V. Kumar, A. Lahiri, and O. B. Dogan, “A strategic framework for a profitable business model in the sharing economy,” *Industrial Marketing Management*, vol. 69, pp. 147–160, 2018.

[22] E. G. Weyl, “A price theory of multi-sided platforms,” *American Economic Review*, vol. 100, no. 4, pp. 1642–72, 2010.

[23] P. Ashkrof, G. H. de Almeida Correia, **O. Cats**, and B. van Arem, “Understanding ride-sourcing drivers’ behaviour and preferences: Insights from focus groups analysis,” *Research in Transportation Business & Management*, vol. 37, p. 100 516, 2020.

[24] P. Ashkrof, G. H. d. A. Correia, **O. Cats**, and B. van Arem, “Ride acceptance behaviour of ride-sourcing drivers,” *arXiv preprint arXiv:2107.07864*, 2021.

[25] B. Turan, R. Pedarsani, and M. Alizadeh, “Dynamic pricing and fleet management for electric autonomous mobility on demand systems,” *Transportation Research Part C: Emerging Technologies*, vol. 121, p. 102 829, 2020.

[26] J. Xi, F. Zhu, Y. Chen, Y. Lv, C. Tan, and F. Wang, “Ddrl: A decentralized deep reinforcement learning method for vehicle repositioning,” in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, IEEE, 2021, pp. 3984–3989.

[27] S. Narayanan, E. Chaniotakis, and **C. Antoniou**, "Shared autonomous vehicle services: A comprehensive review," *Transportation Research Part C: Emerging Technologies*, vol. 111, pp. 255–293, 2020.
[28] S. Li, H. Tavafoghi, K. Poolla, and P. Varaiya, "Regulating tncs: Should uber and lyft set their own rules?" *Transportation Research Part B: Methodological*, vol. 129, pp. 193–225, 2019.
[29] C. Wu, A. Kreidieh, K. Parvate, E. Vinitsky, and A. M. Bayen, "Flow: Architecture and benchmarking for reinforcement learning in traffic control," *arXiv preprint arXiv:1710.05465*, vol. 10, 2017.
[30] C. Zhuge, M. Bithell, C. Shao, X. Li, and J. Gao, "An improvement in matsim computing time for large-scale travel behaviour microsimulation," *Transportation*, vol. 48, no. 1, pp. 193–214, 2021.
[31] M. J. Alonso-González, **O. Cats**, N. van Oort, S. Hoogendoorn-Lanser, and S. Hoogendoorn, "What are the determinants of the willingness to share rides in pooled on-demand services?" *Transportation*, vol. 48, no. 4, pp. 1733–1765, 2021.
[32] **M. Yap**, **O. Cats**, and B. van Arem, "Crowding valuation in urban tram and bus transportation based on smart card data," *Transportmetrica A: Transport Science*, vol. 16, no. 1, pp. 23–42, 2020.
[33] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, *et al.*, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.
[34] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, and A. Bolton, "Mastering the game of go without human knowledge," *nature*, vol. 550, no. 7676, pp. 354–359, 2017.
[35] A. Stooke, A. Mahajan, C. Barros, C. Deck, J. Bauer, J. Sygnowski, M. Trebacz, M. Jaderberg, M. Mathieu, *et al.*, "Open-ended learning leads to generally capable agents," *arXiv preprint arXiv:2107.12808*, 2021.
[36] A. L. Bazzan, "Aligning individual and collective welfare in complex socio-technical systems by combining metaheuristics and reinforcement learning," *Engineering Applications of Artificial Intelligence*, vol. 79, pp. 23–33, 2019.
[37] **J. Alonso-Mora**, P. Beardsley, and R. Siegwart, "Cooperative collision avoidance for nonholonomic robots," *IEEE Transactions on Robotics*, vol. 34, no. 2, pp. 404–420, 2018.
[38] **J. Alonso-Mora**, S. Baker, and D. Rus, "Multi-robot formation control and object transport in dynamic environments via constrained optimization," *The International Journal of Robotics Research*, vol. 36, no. 9, pp. 1000–1021, 2017.
[39] Z. Sun, T. Huang, and P. Zhang, "Cooperative decision-making for mixed traffic: A ramp merging example," *Transportation research part C: emerging technologies*, vol. 120, p. 102 764, 2020.
[40] M. Teixeira, P. M. d'Orey, and Z. Kokkinogenis, "Simulating collective decision-making for autonomous vehicles coordination enabled by vehicular networks: A computational social choice perspective," *Simulation Modelling Practice and Theory*, vol. 98, p. 101 983, 2020.
[41] D. H. Wolpert and K. Tumer, "Collective intelligence, data routing and braess' paradox," *Journal of Artificial Intelligence Research*, vol. 16, pp. 359–387, 2002.
[42] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," *arXiv preprint arXiv:1706.02275*, 2017.
[43] M. R. Blake, **S. Dubey**, J. Swait, E. Lancsar, and P. Ghijben, "An integrated modelling approach examining the influence of goals, habit and learning on choice using visual attention data," *Journal of Business Research*, vol. 117, pp. 44–57, 2020.
[44] M. Kosinski, D. Stillwell, and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," *Proceedings of the national academy of sciences*, vol. 110, no. 15, pp. 5802–5805, 2013.
[45] C. Stachl, Q. Au, R. Schoedel, S. D. Gosling, G. M. Harari, D. Buschek, S. T. Völkel, T. Schuwerk, M. Oldemeier, T. Ullmann, *et al.*, "Predicting personality from patterns of behavior collected with smartphones," *Proceedings of the National Academy of Sciences*, vol. 117, no. 30, pp. 17 680–17 687, 2020.
[46] S. C. Matz, M. Kosinski, G. Nave, and D. J. Stillwell, "Psychological targeting as an effective approach to digital mass persuasion," *Proceedings of the national academy of sciences*, vol. 114, no. 48, pp. 12 714–12 719, 2017.
[47] X. Zhao, X. Yan, A. Yu, and P. Van Hentenryck, "Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models," *Travel behaviour and society*, vol. 20, pp. 22–35, 2020.
[48] J.-C. Shi, Y. Yu, Q. Da, S.-Y. Chen, and A.-X. Zeng, "Virtual-taobao: Virtualizing real-world online retail environment for reinforcement learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 4902–4909.
[49] A. Ezrachi, *Virtual competition*. Harvard University Press, 2017.
[50] A. Mahdawi. "Is your friend getting a cheaper uber fare than you are?" (2018), [Online]. Available: `https://www.theguardian.com/commentisfree/2018/apr/13/uber-lyft-prices-personalized-data` (visited on 12/12/2021).
[51] J. Pawlak, J. W. Polak, and A. Sivakumar, "A framework for joint modelling of activity choice, duration, and productivity while travelling," *Transportation Research Part B: Methodological*, vol. 106, pp. 153–172, 2017.
[52] G. Mariotte, L. Leclercq, H. G. Ramirez, J. Krug, and C. Becarie, "Assessing traveler compliance with the social optimum: A stated preference study," *Travel behaviour and society*, vol. 23, pp. 177–191, 2021.
[53] S. R. Hopkins, S. H. Sokolow, J. C. Buck, G. A. De Leo, I. J. Jones, L. H. Kwong, C. LeBoa, A. J. Lund, A. J. MacDonald, and N. Nova, "How to identify win–win interventions that benefit human health and conservation," *Nature Sustainability*, vol. 4, no. 4, pp. 298–304, 2021.

[54] M. Kumar and V. Dutt, "Understanding decisions in collective risk social dilemma games using reinforcement learning," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 12, no. 4, pp. 824–840, 2020.

[55] Y. Bachrach, R. Everett, E. Hughes, A. Lazaridou, J. Z. Leibo, M. Lanctot, M. Johanson, W. M. Czarnecki, and T. Graepel, "Negotiating team formation using deep reinforcement learning," *Artificial Intelligence*, vol. 288, p. 103 356, 2020.

[56] X. Lin, H.-L. Zhen, Z. Li, Q.-F. Zhang, and S. Kwong, "Pareto multi-task learning," *Advances in neural information processing systems*, vol. 32, pp. 12 060–12 070, 2019.

[57] R. Yang, X. Sun, and K. Narasimhan, "A generalized algorithm for multi-objective reinforcement learning and policy adaptation," *arXiv preprint arXiv:1908.08342*, 2019.

[58] C. Chen, H. Wei, N. Xu, G. Zheng, M. Yang, Y. Xiong, K. Xu, and Z. Li, "Toward a thousand lights: Decentralized deep reinforcement learning for large-scale traffic signal control," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, 2020, pp. 3414–3421.

[59] L. Moreira-Matias, **O. Cats**, J. Gama, J. Mendes-Moreira, and J. F. De Sousa, "An online learning approach to eliminate bus bunching in real-time," *Applied Soft Computing*, vol. 47, pp. 460–482, 2016.

[60] S. Aradi, "Survey of deep reinforcement learning for motion planning of autonomous vehicles," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[61] A. N. Koushik, M. Manoj, and N. Nezamuddin, "Machine learning applications in activity-travel behaviour research: A review," *Transport Reviews*, vol. 40, no. 3, pp. 288–311, 2020.

[62] B. Van Arem, C. J. Van Driel, and R. Visser, "The impact of cooperative adaptive cruise control on traffic-flow characteristics," *IEEE Transactions on intelligent transportation systems*, vol. 7, no. 4, pp. 429–436, 2006.

[63] Y. J. Zhang, A. A. Malikopoulos, and C. G. Cassandras, "Optimal control and coordination of connected and automated vehicles at urban traffic intersections," in *2016 American Control Conference (ACC)*, IEEE, 2016, pp. 6227–6232.

[64] L. Ye and T. Yamamoto, "Evaluating the impact of connected and autonomous vehicles on traffic safety," *Physica A: Statistical Mechanics and its Applications*, vol. 526, p. 121 009, 2019.

[65] I. V. Chremos, L. E. Beaver, and A. A. Malikopoulos, "A game-theoretic analysis of the social impact of connected and automated vehicles," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 2020, pp. 1–6.

[66] I. Klein and E. Ben-Elia, "Emergence of cooperation in congested road networks using ict and future and emerging technologies: A game-based review," *Transportation Research Part C: Emerging Technologies*, vol. 72, pp. 10–28, 2016.

[67] M. W. Levin, "Congestion-aware system optimal route choice for shared autonomous vehicles," *Transportation Research Part C: Emerging Technologies*, vol. 82, pp. 229–247, 2017.

[68] B. Zhou, Q. Song, Z. Zhao, and T. Liu, "A reinforcement learning scheme for the equilibrium of the in-vehicle route choice problem based on congestion game," *Applied Mathematics and Computation*, vol. 371, p. 124 895, 2020.

[69] C. Wu, A. Kreidieh, E. Vinitsky, and A. M. Bayen, "Emergent behaviors in mixed-autonomy traffic," in *Conference on Robot Learning*, PMLR, 2017, pp. 398–407.

[70] D. A. Lazar, E. Bıyık, D. Sadigh, and R. Pedarsani, "Learning how to dynamically route autonomous vehicles on shared roads," *Transportation research part C: emerging technologies*, vol. 130, p. 103 258, 2021.

[71] J. Holler, R. Vuorio, Z. Qin, X. Tang, Y. Jiao, T. Jin, S. Singh, C. Wang, and J. Ye, "Deep reinforcement learning for multi-driver vehicle dispatching and repositioning problem," in *2019 IEEE International Conference on Data Mining (ICDM)*, IEEE, 2019, pp. 1090–1095.

[72] Z. Shou and X. Di, "Reward design for driver repositioning using multi-agent reinforcement learning," *Transportation research part C: emerging technologies*, vol. 119, p. 102 738, 2020.

[73] G. Lu, Y. Nie, X. Liu, and D. Li, "Trajectory-based traffic management inside an autonomous vehicle zone," *Transportation Research Part B: Methodological*, vol. 120, pp. 76–98, 2019, ISSN: 0191-2615.

[74] C. Liu, C.-W. Lin, S. Shiraishi, and M. Tomizuka, "Distributed conflict resolution for connected autonomous vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 1, pp. 18–29, 2017.

[75] R. E. Stern, S. Cui, M. L. Delle Monache, R. Bhadani, M. Bunting, M. Churchill, N. Hamilton, H. Pohlmann, F. Wu, B. Piccoli, *et al.*, "Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments," *Transportation Research Part C: Emerging Technologies*, vol. 89, pp. 205–221, 2018.

[76] J.-F. Bonnefon, A. Shariff, and I. Rahwan, "The social dilemma of autonomous vehicles," *Science*, vol. 352, no. 6293, pp. 1573–1576, 2016.

[77] C. Huang, C. Lv, F. Naghdy, and H. Du, "Reference-free approach for mitigating human–machine conflicts in shared control of automated vehicles," *IET Control Theory & Applications*, vol. 14, no. 18, pp. 2752–2763, Oct. 2020.

[78] W. Schwarting, A. Pierson, **J. Alonso-Mora**, S. Karaman, and D. Rus, "Social behavior for autonomous vehicles," *Proceedings of the National Academy of Sciences*, vol. 116, no. 50, pp. 24 972–24 978, 2019.

[79] L. Floridi, "Establishing the rules for building trustworthy ai," *Nature Machine Intelligence*, vol. 1, no. 6, pp. 261–262, 2019.

[80] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (xai)," *IEEE access*, vol. 6, pp. 52 138–52 160, 2018.

[81] S. Bird, M. Dudik, R. Edgar, B. Horn, R. Lutz, V. Milan, M. Sameki, H. Wallach, and K. Walker, "Fairlearn: A toolkit for assessing and improving fairness in ai," *Microsoft, Tech. Rep. MSR-TR-2020-32*, 2020.

[82] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G.-Z. Yang, "Xai—explainable artificial intelligence," *Science Robotics*, vol. 4, no. 37, 2019.

[83] M. Balac and S. Hörl, "Simulation of intermodal shared mobility in the san francisco bay area using matsim," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, IEEE, 2021, pp. 3278–3283.

[84] M. Hoffman *et al.*, "Acme: A research framework for distributed reinforcement learning," *arXiv preprint arXiv:2006.00979*, 2020.

[85] J. Z. Leibo, E. Duéñez-Guzmán, A. S. Vezhnevets, J. P. Agapiou, P. Sunehag, R. Koster, J. Matyas, C. Beattie, I. Mordatch, and T. Graepel, "Scalable evaluation of multi-agent reinforcement learning with melting pot," PMLR, 2021.

[86] J. Ault and G. Sharon, "Reinforcement learning benchmarks for traffic signal control," 2021.

[87] E. Vinitsky, A. Kreidieh, L. Le Flem, N. Kheterpal, K. Jang, C. Wu, F. Wu, R. Liaw, E. Liang, and A. M. Bayen, "Benchmarks for reinforcement learning in mixed-autonomy traffic," in *Conference on robot learning*, PMLR, 2018, pp. 399–409.

[88] **O. Cats**, **R. Kucharski**, S. R. Danda, and **M. Yap**, "Beyond the dichotomy: How ride-hailing competes with and complements public transport," *arXiv preprint arXiv:2104.04208*, 2021.

[89] **A. Fielbaum**, **R. Kucharski**, **O. Cats**, and **J. Alonso-Mora**, "How to split the costs among travellers sharing a ride? aligning system's optimum with users' equilibrium," *European Journal of Operational Research Follow journal*, 2021.

[90] **R. Kucharski** and **G. Gentile**, "Simulation of rerouting phenomena in dynamic traffic assignment with the information comply model," *Transportation Research Part B: Methodological*, vol. 126, pp. 414–441, 2019.

[91] D. Krajzewicz, "Traffic simulation with sumo–simulation of urban mobility," in *Fundamentals of traffic simulation*, Springer, 2010, pp. 269–293.

[92] M. Balmer, K. Meister, M. Rieser, K. Nagel, and K. W. Axhausen, "Agent-based simulation of travel demand: Structure and computational performance of matsim-t," *Arbeitsberichte Verkehrs-und Raumplanung*, vol. 504, 2008.

[93] T. Arentze, F. Hofman, H. van Mourik, and H. Timmermans, "Albatross: Multiagent, rule-based model of activity pattern decisions," *Transportation Research Record*, vol. 1706, no. 1, pp. 136–144, 2000.

[94] **A. Fielbaum**, M. Kronmueller, and **J. Alonso-Mora**, "Anticipatory routing methods for an on-demand ridepooling mobility system," *Transportation*, pp. 1–42, 2021.

[95] **R. Kucharski** and **O. Cats**, "Exact matching of attractive shared rides (exmas) for system-wide strategic evaluations," *Transportation Research Part B: Methodological*, vol. 139, pp. 285–310, 2020, ISSN: 0191-2615.

[96] G. Cantelmo, **R. Kucharski**, and **C. Antoniou**, "Low-dimensional model for bike-sharing demand forecasting that explicitly accounts for weather data," *Transportation Research Record*, vol. 2674, no. 8, pp. 132–144, 2020.