RAISE the Bar: Restriction of Action Spaces for Improved Social Welfare and Equity in Traffic Management

Michael Oesterle
University of Mannheim
Mannheim, Germany
michael.oesterle@uni-mannheim.de

Tim Grams
University of Mannheim
Mannheim, Germany
tim.nico.grams@uni-mannheim.de

Christian Bartelt
University of Mannheim
Mannheim, Germany
christian.bartelt@uni-mannheim.de

Heiner Stuckenschmidt
University of Mannheim
Mannheim, Germany
heiner.stuckenschmidt@uni-mannheim.de

ABSTRACT

Restriction-based governance has recently been proposed as an alternative to reward shaping for achieving system-level goals in competitive multi-agent systems. In this work, we apply these two approaches to the domain of traffic management, specifically investigating their efficacy and fairness. Our results show that edge restrictions in congested traffic networks are superior to dynamic pricing with regard to equity (i.e., equal treatment of agents) while achieving comparable travel-time improvements. We argue that the former metric, as an adequate proxy for fairness, can be crucial for the quality and acceptance of a governance scheme, particularly when human agents are affected.

KEYWORDS

Multi-Agent System; Governance; Restriction; Fairness; Equity; Traffic Management

ACM Reference Format:
Michael Oesterle, Tim Grams, Christian Bartelt, and Heiner Stuckenschmidt. 2024. RAISE the Bar: Restriction of Action Spaces for Improved Social Welfare and Equity in Traffic Management. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 9 pages.

1 INTRODUCTION

In competitive Multi-Agent Systems (MAS), the selfish strategies of the participating agents (i.e., strategies that maximize the agent's utility) often deviate from the socially optimal solution, which maximizes the social welfare[1]. This discrepancy is a defining trait of the class of *Social Dilemmas* (or *Collective Action Problems*) [26, 49]. It emerges across diverse application areas, with traffic flow optimization [4] being a notable example: Agents leveraging heuristics or machine learning to identify the shortest routes on a directed weighted graph might inadvertently reduce the social welfare significantly below the optimal value [14]. For affine latency functions on graph edges, this *price of anarchy* can be as high as 33% [37] and can rise indefinitely for non-linear latency functions [16].

However, social welfare is not the only benchmark for optimality in a MAS. Other objectives, like fairness or even goals unrelated to agents, can influence the design and functioning of such a system. Within the traffic context, this might manifest as a utility function for traffic authorities or state administration, aiming to curtail road erosion, decrease noise, or increase toll revenue, for instance. To emphasize the universality of such objectives, we use the term *Governance Utility*. A MAS then becomes a *Governance Dilemma* when a stable, joint strategy fails to attain the maximum governance utility. Contrasting with a social dilemma, this broader definition captures scenarios where the governance objective is not a straightforward function of agent objectives[2].

Let us take a closer look at fairness: The prevalent definition of social welfare does not distinguish between two agents achieving equal utility ($u_1 = u_2 = x$) and one agent achieving $u_1 = 2x$ while the other gets $u_2 = 0$. To prevent such disparities, the governance utility could include a metric that diminishes with increased inequality between agents (e.g., variance, entropy, or Gini coefficient). Illustratively, consider a traffic junction where each car's utility inversely corresponds to its waiting time. From a social welfare viewpoint, a scenario where a hundred cars each wait five seconds is identical to one where a single car waits 500 seconds while all others proceed immediately. However, the latter scenario, being inherently inequitable, would be deemed suboptimal. The term *equity* formalizes this intuition as "the absence of unfair, avoidable or remediable differences among groups of people" [27]; we use this as a well-defined operationalization of the subjective word "fairness".

In this work, we investigate two governance paradigms—action-space shaping and reward shaping—in terms of efficacy (i.e., improvement in social welfare) and equity (i.e., equitable treatment). Specifically, we compare dynamic action-space limitations, as recently proposed by [24], with dynamic marginal tolling [41]. The traffic management domain serves as an apt backdrop to explore and compare the effect of both governance schemes.

[1] We adopt the widely accepted *utilitarian* meaning of the term social welfare, which refers to the aggregate utility of all agents in a system [39].

[2] *Governance* in this context is an entity with both an objective function and the power to interact with the MAS in order to maximize this objective. In particular, *restriction-based governance* acts by imposing action space restrictions on the agents at run-time.

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

Remark 1. *We will not focus on finding optimal restrictions for a given traffic network. There exist numerous complexity and inapproximability results for such systems, one being [36]'s proof that barring $\mathcal{P} = \mathcal{NP}$, no algorithm can achieve an approximation ratio $< \frac{n}{2}$ to find optimal edge constraints in a congested network with n nodes. We will therefore use existing knowledge about suitable restrictions and evaluate their impact relative to a recognized reward-shaping technique.*

2 RELATED WORK

Overall, we are concerned with influencing a MAS to enhance the governance utility of stable results; any actions guiding the system towards this objective are termed as "governance". The interaction loop of the standard Markov Game model [40] and the more general Partially Observable Stochastic Game (POSG) [32] allow for governance intervention at four different points or a combination of them (see Figure 1):

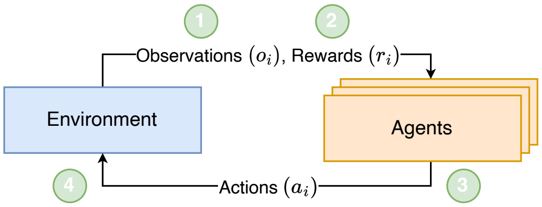


Figure 1: Partially Observable Stochastic Game (POSG) and potential governance interventions

(1) *Observation*: The governance can change what agents observe, given the environmental state
(2) *Reward*: The governance can change what rewards agents receive for a given state/action pair
(3) *Action space*: The governance can change what actions agents can take
(4) *Transition*: The governance can change what the next environmental state is, given the current state and joint action

While interventions at the observation and transition level are rather artificial and thus have very limited applicability in real-world systems, interventions with respect to rewards and action spaces are common in the literature. *Reward shaping* [18] is by far the most widely used paradigm; the rewards can either be defined by an external entity, as in the game-theoretically grounded Vickrey-Clarke-Groves (VCG) mechanisms [21] and the related marginal cost pricing [47], or emerge from the agents themselves as the ability for cooperation, as defined by Normative Systems [1, 7, 8, 20]. Depending on the definition, normative systems cover both reward shaping and action space shaping: Norms can be categorized as *soft* or *hard* [42], where the former type means that violations are punished with negative rewards, while the latter simply prohibits actions that violate the norm.

Dynamic action-space restrictions have been explored by [31] using search-based optimization; other techniques are Reinforcement Learning (RL) over a small discrete action space in POSGs [22], and tree search over one-dimensional continuous spaces for Normal-Form Games [24]. In general, it is not possible to predict how (if at all) an action-space restriction influences the equilibria of a MAS; however, [24] have shown that, to increase the minimum governance utility of a stable joint strategy, at least one of the actions which were taken by the agents at the worst unrestricted equilibrium must be removed from the action space.

Braess' Paradox (described in more detail in Section 4.2) is an example of social welfare improvement through restriction of multi-agent systems, and it has been extensively studied from the perspectives of network design, graph theory, game theory, and others after it had been first described by [6]. Most early (and some later) work focuses on the original four-node network structure, examining criteria for the occurrence of the paradox in terms of latency functions and traffic rate [29, 30, 50]. As a second focus area, [37] show that the *price of anarchy*, defined as the ratio between the equilibrium and the social optimum and therefore an upper bound for the improvement achievable through edge restrictions, is $\leq \frac{4}{3}$ for affine latency functions, regardless of the underlying graph. For general latency functions, particularly for polynomials of arbitrarily high degree, [16] demonstrate that Braess' Paradox can be arbitrarily severe, and [36] prove various inapproximability and hardness results for the problem of identifying the edges causing the paradox. A third line of research deals with the occurrence of Braess' Paradox in random and real-world networks: [44] derive a likelihood of 50% via a non-constructive proof, and [48] argue that for large Erdős-Rényi graphs [10] with certain assumptions on edge density and latency functions, the paradox occurs with high probability *for some traffic rate*. At the same time, [11, 29] provide evidence suggesting that Braess' Paradox is less likely to occur with randomly chosen traffic rates compared to adversarial rates.

Single-commodity networks (i.e., there is exactly one source-sink pair) with constant traffic rate allow for a static solution, while changing demand and multiple commodities can require adaptive strategies (see Section 4.2). The multi-commodity case of Braess' Paradox, albeit with constant traffic rates, has been examined in [9, 16, 36, 37], and it turns out that the worst-case behavior of congested networks can be much worse than in a single-commodity scenario.

While most of the research on Braess' Paradox models the traffic flow on a macroscopic level, it can also be shown to occur in microscopic [3, 5] and mesoscopic [28] models[3]. In the present work, we use a microscopic perspective to trace the impact of various governance mechanisms on individual agents.

The notion of fairness in algorithmic decision-making [2, 19] has become an important metric next to traditional performance indicators like reward, loss or prediction error, particularly for mechanisms whose output directly affects people. While *fairness by unawareness* [13] emphasizes the importance of a fair process, this does not imply a fair (i.e., balanced) outcome. The above-mentioned concept of *equity*, defined by the World Health Organization as "the absence of unfair, avoidable or remediable differences among groups of people" and widely used in the context of health [33, 38], can

[3] In the context of traffic modeling, *microscopic* refers to modeling individual vehicles and their behavior, such as acceleration, lane changes, and interactions at a small scale. A *mesoscopic* model involves a broader perspective, focusing on groups of vehicles or traffic flow as a whole, typically without modeling individual vehicle-level details. Finally, *macroscopic* means an even broader perspective, considering average characteristics of traffic such as flow rates, density, and speed over larger sections of a roadway or an entire network.

help judge the fairness of an algorithm or a governance mechanism by evaluating the impact the mechanism has on different groups.

In this context, [17] pursue fairness between agents by promoting *prosocial norms* in dynamic systems, while [45] propose compliance with pre-defined ethical frameworks as the driver of fair treatment. However, both approaches require agents to possess bespoke capabilities for compliance with norms or ethical rules. This strong requirement is not necessary for restriction-based governance, where agents are assumed to have their usual objective function.

3 THEORY

3.1 Multi-Agent System

We model a multi-agent system as a POSG, i.e., a tuple

$$(I, \mathcal{S}, \boldsymbol{O}, \boldsymbol{\sigma}, \boldsymbol{A}, \boldsymbol{r}, \delta) ,$$

where I is the set of agents, $\mathcal{S}$ is the set of environmental states, $\boldsymbol{O} = (O_i)_{i \in I}$ are the sets of observations for each agent[4], $\sigma_i : \mathcal{S} \to O_i$ is agent i's observation function, $\boldsymbol{A} = (A_i)_{i \in I}$ are the agents' action spaces, $r_i : \mathcal{S} \times \boldsymbol{A} \times \mathcal{S} \to \mathbb{R}$ is agent i's reward function, and $\delta : \mathcal{S} \times \boldsymbol{A} \to \Delta_{\mathcal{S}}$ is the stochastic transition function[5].

The agents' action policies (which are not part of the formal POSG model) are defined as $\pi_i : O_i \to \Delta_{A_i}$. The interaction between agents and environment in this model can be succinctly described by the *evolution formula*

$$s_{t+1} = \delta(s_t, \boldsymbol{\pi}^{(t)}(\boldsymbol{\sigma}(s_t))) , \tag{1}$$

where the output of the stochastic functions is sampled according to the respective distribution[6].

Remark 2. *In reality, the full flexibility of a POSG is often not exploited; common reductions are statelessness (i.e., $|\mathcal{S}| = 1$), determinism (i.e., $\forall s \in \mathcal{S}, \boldsymbol{a} \in \boldsymbol{A}\ \exists s' \in \mathcal{S} : \mathbb{P}[\delta(s, \boldsymbol{a}) = s'] = 1$), or uniformity (i.e., $A_i = A_j\ \forall i, j \in I$). We will use uniformity throughout the paper, and statelessness for some illustrative examples in Section 4.1.*

3.2 Stable action policies

In the following, let $\boldsymbol{\Pi}$ denote the space of all joint action policies for a given MAS.

3.2.1 Best response. Let $\boldsymbol{\pi} \in \boldsymbol{\Pi}$ be a joint action policy. Then

$$\mathcal{B}_i(\boldsymbol{\pi}_{-i}) := \operatorname*{arg\,max}_{\pi \in \Pi} \mathbb{E}_{(a, \boldsymbol{a}_{-i})}[r_i] \subseteq \Pi_i$$

denotes the set of all *best responses* (BRs) of player i to the other players' given action policies $\boldsymbol{\pi}_{-i}$[7].

3.2.2 Nash Equilibrium. A joint action policy $\boldsymbol{\pi} \in \boldsymbol{\Pi}$ is a *Nash Equilibrium* (NE) if each individual action policy in $\boldsymbol{\pi}$ is a best response to the other players' policies. $\mathcal{N}$ denotes the set of all NE:

$$\mathcal{N} := \{\boldsymbol{\pi} \in \boldsymbol{\Pi} :\ \pi_i \in \mathcal{B}_i(\boldsymbol{\pi}_{-i})\,\forall i\} .$$

3.3 Governance

Similar to the agent reward functions, the *governance utility function* $\mathfrak{u}$ is defined as $\mathfrak{u} : \mathcal{S} \times \boldsymbol{A} \times \mathcal{S} \to \mathbb{R}$. This function defines how the governance interacts with the MAS.

The Nash Equilibria of a MAS can yield different values for the governance utility. Arguably, the governance has the goal of achieving a *guaranteed* high utility $\mathfrak{u}$, so we need to consider the minimum governance utility of a NE. We call this quantity the *Minimum Equilibrium Governance Utility* (MEGU). The MEGU of a governed MAS is defined as

$$\mathfrak{m} := \min_{\boldsymbol{\pi} \in \mathcal{N}} \mathfrak{u}(\boldsymbol{\pi}) .$$

3.3.1 Reward shaping. The POSG model defined in Section 3.1 is flexible enough to accommodate a governance which acts by changing the reward functions $\boldsymbol{r}$ in order to maximize its utility. Therefore, there is no need for an extension of the model.

3.3.2 Action-space shaping. To incorporate the possibility of action-space restrictions, we extend the POSG model (Section 3.1) by adding a set of restriction functions $\boldsymbol{\rho} = (\rho_i)_{i \in I}$ with $\rho_i : \mathcal{S} \to 2^{A_i}$ that are applied to each agent, respectively[8]. Accordingly, agents' action policies are now defined as $\pi_i : O_i \times 2^{A_i} \to \Delta_{A_i}$ with the requirement that $\operatorname{supp}(\pi_i(s, R)) \subseteq R$ for any restriction $R \subseteq A_i$[9]. This requirement ensures that any action not in R (i.e., forbidden actions) is taken with probability zero[10].

The evolution formula from Equation (1) thus becomes

$$s_{t+1} = \delta\left(s_t, \boldsymbol{\pi}^{(t)}\left(\boldsymbol{\sigma}(s_t), \boldsymbol{\rho}^{(t)}(s_t)\right)\right) .$$

4 THE EFFECT OF RESTRICTIONS

4.1 Static restrictions

The most simple and extensively studied social dilemmas are two-player, two-action matrix games such as the Prisoner's Dilemma [34], Stag Hunt [43] and the Chicken Game [35]. Clearly, a restriction-based governance is inappropriate in such cases: Forbidding even a single action would result in fully prescribed strategies, simplifying agent behavior to the point of triviality.

Let us therefore consider a two-player, three-action matrix game with symmetric payoffs, and assume the governance utility $\mathfrak{u}$ to be the social welfare. The following examples offer an initial insight into the possible impacts of restrictions:

Example 4.1. Given the game payoffs in Figure 2a, where both players possess the action space $A = \{a, b, c\}$, the unique (pure) NE in an unrestricted scenario is the joint strategy (c, c) with agent utilities $u_1(c, c) = u_2(c, c) = 1$ and $\mathfrak{u}(c, c) = 2$. The unique social optimum (SO) is (b, b) with $\mathfrak{u}(b, b) = 4$.

By excluding action c for both players, we can align the unique NE with the SO at (b, b). As a result, the action space restriction increases the MEGU from 2 to 4.

[4] By convention, we use bold face for vectors or sequences of variables.

[5] Δ_X denotes the set of probability distributions over a (finite or infinite) set X, i.e., the set of all functions $p : X \to [0, 1]$ with $\sum_{x \in X} p(x) = 1$.

[6] The use of a time step t as a subscript or superscript indicates that a variable or function evolves over time.

[7] For a vector $\boldsymbol{x} \in \mathcal{X}^n$, let $\boldsymbol{x}_{-i} := (x_1, ..., x_{i-1}, x_{i+1}, ..., x_n)$ denote the vector obtained by removing x_i. By convention, we also use the concatenation $\boldsymbol{x} = (x_i, \boldsymbol{x}_{-i})$.

[8] We denote with $2^S := \{S' : S' \subseteq S\}$ the power set (i.e., the set of all subsets) of an arbitrary set S, both finite and infinite.

[9] For a real-valued function $f : X \to \mathbb{R}$, $\operatorname{supp}(f) := \{x \in X : f(x) \neq 0\}$ denotes the *support* of f.

[10] In other words, we assume restrictions to be *hard constraints* in the terminology of [42] (see Section 2).

		Player 2		
		a	b	c
Player 1	a	0, 0	0, 1	0, 0
	b	1, 0	2, 2	0, 3
	c	0, 0	3, 0	1, 1

(a) Restricting action c increases the MEGU.

		Player 2		
		a	b	c
Player 1	a	1, 1	1, 0	0, 0
	b	0, 1	2, 2	3, 6
	c	0, 0	6, 3	4, 4

(b) Restricting actions never increases the MEGU.

Figure 2: Payoff matrices for exemplary matrix games where action-space restrictions have different effects. Following conventional notation, the first number in each cell is the utility of Player 1, and the second one of Player 2.

Example 4.2. Conversely, if the payoffs are given by Figure 2b, the unique unrestricted NE is (c, c), where $u_1(c, c) = u_2(c, c) = 4$ and $\mathfrak{u}(c, c) = 8$. The SOs are (b, c) and (c, b), both with $\mathfrak{u}(b, c) = \mathfrak{u}(c, b) = 9$.

If we were to eliminate action c, both the NE and the SO become unattainable. The new NE evolves to (b, b), with utilities $u_1(c, c) = u_2(c, c) = 2$ and $\mathfrak{u}(c, c) = 4$. Even though this NE now matches the *new* SO, its governance utility is less than its original value.

Example 4.3. Revisiting the payoffs from Figure 2a, but this time restricting action a, we observe no governance effect since this limitation does not impact either the SO (b, b) or the NE (c, c).

4.2 Dynamic restrictions

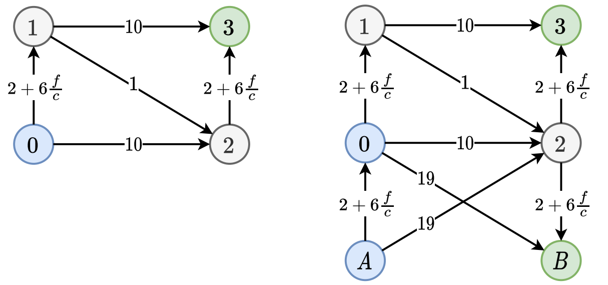
Braess' Paradox, depicted in Figure 3a, is a frequently referenced illustration of the effectiveness of restrictions in stateful MAS[11]. In the present work, we use edge latency functions of the form

$$l(f) = a + b\left(\frac{f}{c}\right)^d \tag{2}$$

with parameters $a > 0$, $b \geq 0$, and $c, d \geq 1$. This model is based on the Bureau of Public Roads' (BPR) proposal [25] and is a common choice in literature. The latency functions suggested by Braess, $l(f) = 0$ and $l(f) = f$, do not fit this model since their free-flow time $l(0)$ is zero. Hence, we use a slightly modified version of the paradox that retains its essential characteristics but ensures $l_e(0) > 0 \; \forall e \in E$.

The original network has a well-known static solution for the problem of finding the best restriction: Closing the edge $(1, 2)$ is socially optimal. Thus, a one-off analysis might suggest a permanent road closure, deeming dynamic restrictions redundant. However, when this network is superposed with a similar structure, the restriction's effect becomes demand-dependent:

Example 4.4. Consider the traffic network shown in Figure 3b. When all (or most) agents travel from node 0 to node 3, the optimal restriction is to close road $(1, 2)$; however, when demand between A and B dominates, closing both $(0, 2)$ and $(1, 2)$ is optimal[12]. It is therefore not possible to find the optimal restriction without taking into account the real-time behavior of the agents, and adapting the governance policy when the behavior changes.



(a) Original Braess Paradox **(b) Double Braess Paradox**

Figure 3: Each edge of these traffic networks has a latency function $l_e(f)$, indicating the length (i.e., travel time) of the edge for a given flow $f \in \mathbb{N}_0$ (c is a capacity parameter). For agents traveling from node 0 to 3, social welfare is increased from −17 to −15 by closing the road between nodes 1 and 2. If two Braess Paradoxes are combined (as shown on the right), different road closures result in the optimal social welfare, depending on the dominating demand.

Table 1: Travel times at equilibrium for the Double Braess Paradox (optimal values underlined for each demand pattern).

	(0, 2), (1, 2)	(0, 2), ~~(1, 2)~~	(0, 2), ~~(1, 2)~~	~~(0, 2)~~, ~~(1, 2)~~
(0, 3)	17	$\underline{15}$	17	18
(A, B)	25	26	25	$\underline{24}$

Example 4.4 illustrates the importance of dynamic restrictions, which rely on real-time MAS observations. This perspective contrasts with much of the existing Braess Paradox research and related problems, where a static flow is the basis for determining the optimal edge subset.

4.3 Limitations

It is evident that action space restrictions cannot improve the social optimum, as the maximum is taken from a strictly smaller joint strategy space. As for stable strategies, [24] show that the MESU can only rise if the relevant NE is eliminated, i.e., if at least one action present in the lowest-welfare equilibrium is restricted.

Applying restrictions to individual agents can achieve the SO by dictating specific actions for all agents, thereby removing their decision-making autonomy. However, transforming a MAS into a single-agent optimization problem by centralizing all decisions is typically not feasible, for reasons ranging from ethical and legal concerns to issues of resilience and scalability.

Viewing restrictions through a fairness lens (as discussed in Section 6.2), it becomes apparent that restrictions often need to be *uniform*, that is, $A_i = A_j := A$ and $\rho_i = \rho_j := \rho; \forall i, j \in I$. This ensures all agents are treated equitably and have the same actions

[11] The formal definition of graph and latency functions is provided in Section 5.

[12] See Appendix B in the supplementary material for more details about this setup.

available at each time step. However, the uniformity of restrictions might reduce the SO, as observed in Example 4.2.

5 EXPERIMENTS

In this section, we use a microscopic multi-step traffic model (see Appendix A.1) to simulate agent behavior across a number of networks, both with and without governance. We analyze the impact of governance mechanisms, as outlined in Section 3.3, on agent behavior and overall system outcomes by varying parameters such as traffic rate, latency functions, demand, and *value of money* (see Section 5.3).

Remark 3. *As our focus is on the observed interaction between agents, their environment, and the governance, there is only one equilibrium: The joint strategy which is achieved experimentally after a sufficient number of steps. This simplifies the MEGU concept from Section 3.3, which is based on a larger set of NE. Our experiments indicate sufficient convergence for us to consider joint strategies as stable after a few thousand time steps.*

5.1 Traffic model

Let $G = (V, E, \boldsymbol{l})$ be a directed graph with a BPR latency function $l_e : \mathbb{N}_0 \to \mathbb{R}^+$ as in Equation (2) for each edge $e \in E$. These functions map a flow value (i.e., the number of agents currently using the edge) to a latency value, which indicates the number of time steps required to traverse the edge (see Figure 3). Each agent i has a starting node s_i, a current position $p_i \in [0, 1]$ along an edge $e_i \in E$ and a designated destination node d_i[13]. At any time step, the flow f_e of an edge e is defined as $f_e = |\{i \in I : e_i = e\}|$, and the corresponding latency on e is $l_e(f_e)$. The graph, together with the tuple (p_i, e_i, d_i) for all agents, represents the system's current state.

An agent i can only decide its next move (i.e., select its next edge) upon reaching a node, specifically when $p_i = 1$. Therefore, the agent needs to observe only its current node v_i, destination node d_i, and the current latency values of all edges. As described in Section 3.3.2, the set $A' \in 2^E$ of currently permissible actions is also provided as an input.

5.2 Δ-tolling

Δ-tolling, introduced by [41], is a dynamic reward-shaping strategy for congested networks. It updates per-edge tolls based on the difference between free-flow time and actual latency. This method has been proven to be equivalent to marginal-cost tolling for BPR latency functions, ensuring optimal system performance. Its adaptability, scalability, and straightforward implementation make it a suitable benchmark, representing the reward-shaping paradigm of multi-agent governance.

The toll update in Δ-tolling is expressed as

$$\tau_e^{(t)} := R\left[d_e \cdot (l_e(f_e) - l_e(0))\right] + (1 - R)\tau_e^{(t-1)} , \qquad (3)$$

where d is the exponent of the latency function (see Equation (2)), and R is a responsiveness parameter. Intuitively, Δ-tolling assigns a toll to each edge in proportion to its congestion, while using exponential smoothing to prevent abrupt updates.

5.3 Setup

We evaluate the travel time of agents in unrestricted traffic ("Base"), edge restrictions ("Restriction") and edge tolls ("Tolling") across two distinct network types, as depicted in Figure 4:

Random Erdős-Rényi ($G_{n,p}$) graphs Braess' Paradox was shown to occur very likely on random graphs as the number of nodes tends to infinity [48], but this result comes with some caveats; most notably, the edge likelihood and the traffic rate need to be chosen carefully in order to generate the paradox. We have thus selected a graph (Figure 4a) with $n = 59$ via random search where restricting a single edge improves social welfare[14]. More details and graphs are shown in Appendix C, along with the results of the analysis.

Generalized Braess graphs The family $B_n, n \in \mathbb{N}^+$ (see Figure 3) generalizes the original network and the "Double Braess" structure from Figure 3b. It allows to route n different commodities (i.e., source/sink pairs) on the graph B_n, each commodity with a different optimal restriction.

For the Braess-based graphs, the optimal edges to close for improved latency are already known; as previously mentioned in Section 1, our aim is not to provide new results on the Braess Paradox's occurrence or detection.

Agents in the MAS employ a simple shortest-path algorithm to select the optimal edge upon reaching an intersection. This approach is well-defined for both unrestricted and restricted scenarios. However, for the Δ-tolling scenario, a relative weighting between travel time and tolls is necessary: Each agent $i \in I$ has a *value of money* $v_i \in \mathbb{R}_0^+$, which defines the time-equivalent worth of one unit of tolls. In other words, the agent minimizes $\sum_{e \in P} \left(l_e(f_e) + v_i t_e\right)$ over all paths P from its current position to its target node and selects one of the optimal paths. To see the effect of the value of money on the agents' behavior and treatment, we assign to each agent some v_i, uniformly drawn from the set $\{0, 1, 2, 5\}$.

Remark 4. *As highlighted by [29] and [36], traffic rate plays a pivotal role in the occurrence and severity of Braess' Paradox. In relation to this parameter, our choices are:*

$G_{n,p}$ *Using random search, we found a traffic rate of $f = 56$ to be adequate for the selected graph.*

Braess *A rather generic traffic rate of $f = \frac{5}{2}c$ for edge capacity c is sufficient*[15].

5.4 Performance metrics

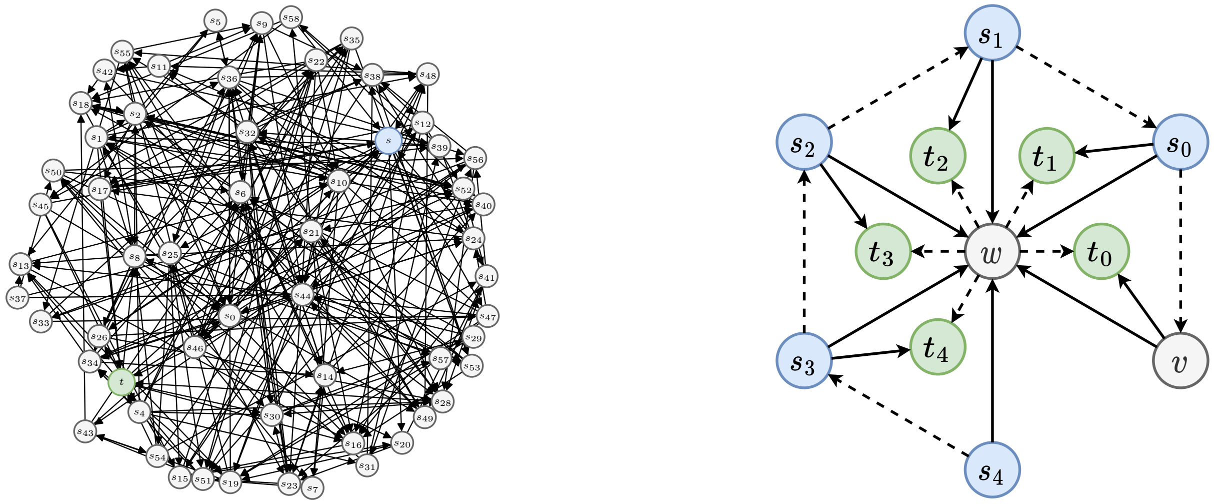
The *mean travel time* of all agents (representing social welfare) is presented as the main performance indicator for the respective governance methods. As a measure for equity, we examine the *correlation between travel time and value of money* for all agents. Specifically, we assess the slope of the closest linear regression between these two variables[16]. This evaluation is vital as it probes

[13] In this context, each source/sink pair that is used by an agent as starting and destination nodes is called a *commodity*.

[14] However, efficacy compared to Δ-tolling was not considered in the selection process, nor was the fairness metric defined in Section 5.4.

[15] Intuitively, this can be explained by the fact that the routes on these graphs consist of either 3 or 4 edges, two of which are dominating in latency. Therefore, cars are randomly distributed along a route whose length is ≈ $\frac{5}{2}$ times the latency of a "long" edge.

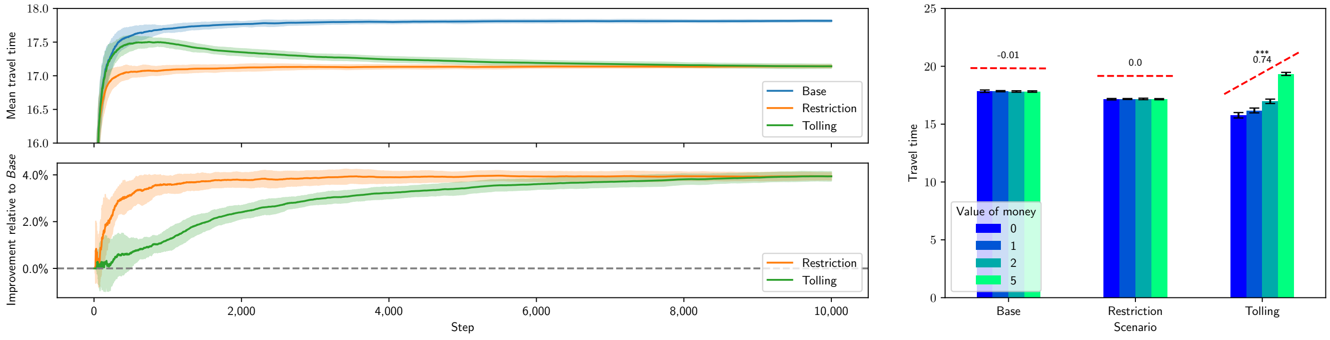
[16] A correlation coefficient like Pearson or Spearman is not suitable since it only measures the strength of the connection, but not its direction; in particular, if all agents



(a) $G_{n,p}$ graph with Braess' Paradox. For this graph, we have chosen $n = 59$ and $p = 0.07$, in accordance with [48]'s assumption that $p = \Omega(n^{-1/2+\varepsilon})$ for some $\varepsilon > 0$.

(b) Generalized Braess graph B_4. Solid lines denote constant-latency edges, while dashed lines are edges with affine latency functions (details and the construction schema for B_n are provided in Appendix D).

Figure 4: Network structures used for the experiments. Start nodes are marked blue, while target nodes are green. For each graph, we measure travel times, improvement and fairness for unrestricted traffic, edge restrictions and Δ- tolling.



(a) Travel time (i.e., social welfare) and improvement

(b) Fairness

Figure 5: Results of the $G_{n,p}$ graph experiment. First, we measure travel times and improvement for the *Base*, *Restriction*, and *Tolling* scenarios as described in Section 5.3. In addition, we investigate the dependency of the travel time on the value that an agent assigns to money compared to time.

the governance mechanism's *fairness*, i.e., the treatment equality towards agents from different groups.

5.5 Reproducing the experiments

The supplementary material, including a code package with all experiments as Jupyter Notebook files, as well as an appendix with more details about the experimental setup, have been published at https://github.com/michoest/aamas-2024. Using this code and the seeds listed in Appendix E, the reported results can be fully reproduced. Different seeds, of course, can give slightly different results, but will confirm that all claims are robust with respect to randomization.

have the same travel time (i.e., the mechanism is perfectly fair), these coefficients are not zero, but one.)

5.6 Results

5.6.1 Random Erdős-Rényi ($G_{n,p}$) graphs. Figure 5 shows the performance metrics for ten independent runs on the graph from Figure 4a (the mean is drawn as a solid line, while the shaded area denotes the standard deviation of the results). With respect to travel times, both *Restriction* and *Tolling* outperform *Base* by approximately 3%. The fairness metric, however, shows a substantial difference between the governance paradigms: While agents with different values of money are treated largely equally in the *Base* and *Restriction* scenarios, their travel times differ significantly ($p \leq 0.01$) for *Tolling*, such that agents with higher value of money v_i have longer travel times.

5.6.2 Generalized Braess Graphs B_n. The performance metrics for the graphs B_n with the single commodity (s_n, t_n) are shown in Figure 6 for $n \in [0, 20]$ and five independent runs. Similarly to the results on the $G_{n,p}$ graph, both *Restriction* and *Tolling* improve the *Base* case, and this time *Restriction* (using optimal road closures) outperforms *Tolling* by a few percentage points. Regarding fairness, Figure 6b displays the regression lines' slopes (see Figure 5b for a single graph) in an aggregated way for all graphs B_n. This fairness metric shows that *Base* and *Restriction* are nearly unbiased across all commodities. For *Tolling*, the correlation between value of money and travel time in this particular setup shifts from disadvantaging high values of money for small values of n to disadvantaging low values of money for larger n.

6 DISCUSSION

6.1 Efficacy

The declared objective of the governance is maximizing social welfare, i.e., minimizing the travel time of all agents. To this end, both approaches succeed in improving the status quo (the unrestricted *Base* scenario), and the improvements are comparable in magnitude.

The results in Figures 5 and 6 show the mean travel time, but not the mean *total cost*, i.e., the weighted combination of travel time and tolls which reveals the additional reward that the toll-based governance has to invest. As can be seen in Figure 7 for the generalized Braess graphs, including the tolls results in a much lower efficacy of the reward-shaping scheme. This can be an additional argument for restrictions, where no monetary rewards are involved.

Remark 5. *Marginal-cost tolling, by definition, targets edges where heightened demand results in latency spikes. However, in situations such as Braess' Paradox, it is the "constant-low-latency edges" that need to be closed to attain optimal flow. Such edges, by definition of the tolling scheme, can never be tolled. As a result, the tolling strategy must reduce demand for these roads by imposing higher tolls on connecting roads. This culminates in a "proxy-tolling" outcome where certain high-demand edges remain toll-free, while others bear the brunt with exorbitant tolls.*

6.2 Fairness

Reward-shaping strategies inherently differentiate between agents based on how additional rewards influence them. In essence, an agent with a minimal value of money might remain largely unaffected by rewards or penalties, while others might drastically alter their behavior. Given that rewards and penalties often manifest as monetary values, this can inadvertently compromise fairness, especially in scenarios where an agent's wealth should not dictate their actions. Our experiments reveal that while Δ-tolling effectively reduces average travel time, it simultaneously introduces significant variance. Notably, the relationship between the value of money and travel time isn't straight-forward but varies with the network structure. Restriction-based governance, in contrast, offers comparable travel-time efficiency but ensures a distribution of travel times which is close to equity.

6.3 Resource restrictions

To conclude the discussion of restriction-based governance and reward shaping, we outline *resource restriction* as a hybrid governance paradigm, combining elements of restriction-based and reward-based governance:

Restrictions, as they have been defined so far, directly limit the action spaces of the agents, thereby generating new equilibria. Another way restrictions can be used in multi-agent systems is to restrict *resources* in order to change incentives. The restrictions therefore serve to *indirectly* shape the rewards by encouraging or discouraging actions which relate to the restricted resources. In Appendix F of the supplementary material, we outline a parking management scenario based on existing work for dynamic pricing [15] and show that closing some of the parking spaces can increase social welfare. Without restricting any *actions*, the restriction of resources (in this case, parking spaces) affects how the agents valuate their options, which, in turn, steers their behavior in the desired direction.

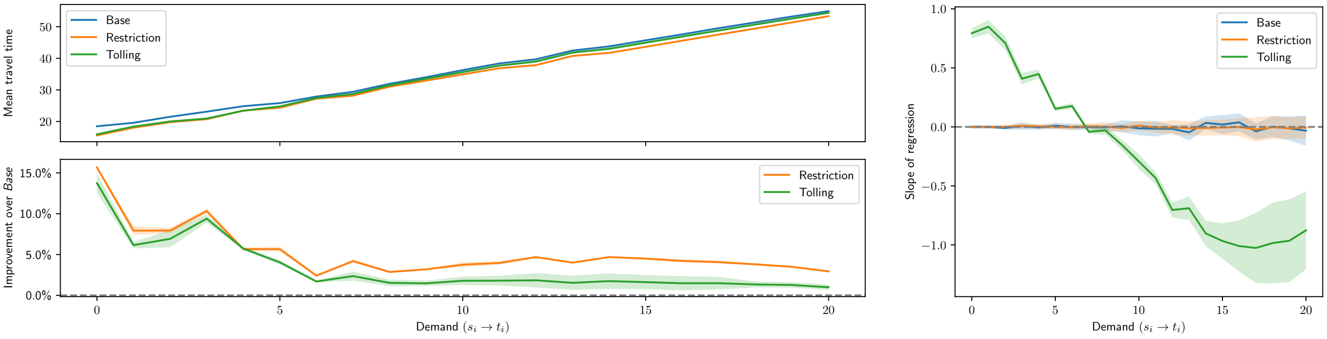
In contrast to "pure" reward shaping approaches, this method avoids monetary incentives and therefore maintains some of the mentioned advantages of action-space shaping. On the other hand, it does not lend itself to direct calculation of equilibrium strategies without explicit knowledge of the connection between resource restriction and corresponding changes in agent utility.

7 CONCLUSION

Action-space restrictions seem to be inferior to reward shaping at first glance, as they only allow a binary distinction between allowed and forbidden actions (similar to a reward of 0 and $-\infty$ for choosing an action, respectively). However, as we have shown in the present work, the actual comparison is more complex, and restrictions come with a considerable advantage regarding fairness.

The study of action-space restrictions as a means of governing multi-agent systems is far from exhausted: Only recently have common multi-agent learning environments like PettingZoo [46] been equipped with governance capabilities [23], and there is still scarce consideration of restrictions for Reinforcement Learning algorithms (first steps are described in [12]). It has been shown that finding optimal restrictions for dynamic systems can be hard, but the dependency of their effect on the (a priori unknown) behavior of the agents makes real-time adaptation and optimization necessary.

Despite these challenges, its unique way of interacting with agents and environment makes action-space shaping a valuable tool for governance entities, both in abstract game-theoretic settings and in real-world systems. We want to emphasize that the acceptance



(a) Social welfare (i.e., mean travel time)

(b) Fairness

Figure 6: Results of the generalized Braess graph experiment, showing the performance metrics in relation to the demand pattern. Both *Restriction* and *Tolling* improve the *Base* case, but the fairness measure is very different: In *Tolling*, the value of money has a major influence on the travel time.

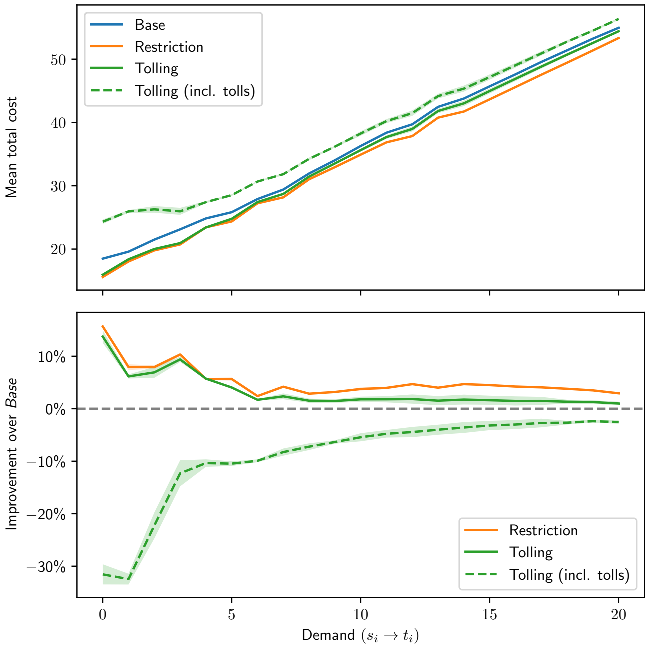


Figure 7: Total cost of travel (including tolls) for the generalized Braess graphs.

of governance mechanisms, be it reward shaping or restrictions, crucially depends on their (perceived and objective) fairness. With respect to this condition, restriction-based governance, together with equity considerations, has the potential to substantially push the applicability of governance schemes for systems consisting of human or artificial agents.

ACKNOWLEDGMENTS

This work was supported by the German Federal Ministry for Economic Affairs and Climate Action (BMWK), and partially supported by the Volkswagen Foundation under the grant "Consequences of Artificial Intelligence for Urban Societies (CAIUS)".

REFERENCES

[1] Giulia Andrighetto, Guido Governatori, Pablo Noriega, and Leon van der Torre (Eds.). 2013. *Normative Multi-Agent Systems.* Dagstuhl Follow-Ups, Saarbrücken, Germany.

[2] Solon Barocas, Moritz Hardt, and Arvind Narayanan. 2019. Fairness and Machine Learning: Limitations and Opportunities. http://www.fairmlbook.org.

[3] Ana L.C. Bazzan and Franziska Klügl. 2005. Case studies on the Braess Paradox: Simulating route recommendation and learning in abstract and microscopic models. *Transportation Research Part C: Emerging Technologies* 13, 4 (2005), 299–319.

[4] Martin J. Beckmann, C. B. McGuire, and C. B. Winsten. 1955. *Studies in the Economics of Transportation.* RAND Corporation, Santa Monica, CA.

[5] Stefan Bittihn and Andreas Schadschneider. 2021. The effect of modern traffic information on Braess' paradox. *Physica A: Statistical Mechanics and its Applications* 571 (2021).

[6] Dietrich Braess. 1968. Über ein Paradoxon aus der Verkehrsplanung. *Unternehmensforschung* 12 (1968), 258–268.

[7] Amit Chopra, Leendert van der Torre, and Harko Verhagen (Eds.). 2018. *Handbook of Normative Multiagent Systems.* College Publications, Milton Keynes, UK.

[8] Davide Dell'Anna, Mehdi Dastani, and Fabiano Dalpiaz. 2020. Runtime Revision of Sanctions in Normative Multi-Agent Systems. *Autonomous Agents and Multi-Agent Systems* 34, 2 (2020), 54.

[9] Kord Eickmeyer and Ken-ichi Kawarabayashi. 2013. Approximating Multi Commodity Network Design on Graphs of Bounded Pathwidth and Bounded Degree. In *Algorithmic Game Theory*, Berthold Vöcking (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 134–145.

[10] Paul Erdös and Alfréd Rényi. 1959. *On Random Graphs I.*

[11] E.J. Friedman. 2004. Genericity and congestion control in selfish routing. In *2004 43rd IEEE Conference on Decision and Control (CDC) (IEEE Cat. No.04CH37601)*, Vol. 5. IEEE, Nassau, Bahamas, 4667–4672.

[12] Tim Grams. 2023. *Dynamic interval restrictions on action spaces in deep reinforcement learning for obstacle avoidance.* Master's thesis. University of Mannheim. https://arxiv.org/abs/2306.08008

[13] Nina Grgic-Hlaca, Muhammad Bilal Zafar, Krishna P. Gummadi, and Adrian Weller. 2016. The Case for Process Fairness in Learning: Feature Selection for Fair Decision Making. In *Proceedings of the Symposium on Machine Learning and the Law at the 29th Conference on Neural Information Processing Systems (NIPS 2016).* NIPS, Barcelona, Spain.

[14] Mary Sissons Joshi, Vijay Joshi, and Roger Lamb. 2005. The Prisoners' Dilemma and City-Centre Traffic. *Oxford Economic Papers* 57, 1 (2005), 70–89.

[15] Jakob Kappenberger, Kilian Theil, and Heiner Stuckenschmidt. 2022. Evaluating The Impact Of AI-Based Priced Parking With Social Simulation. In *Social Informatics: 13th International Conference, SocInfo 2022, Glasgow, UK, October 19–21, 2022, Proceedings* (Glasgow, United Kingdom). Springer-Verlag, Berlin, Heidelberg, 54–75.

[16] Henry Lin, Tim Roughgarden, Éva Tardos, and Asher Walkover. 2011. Stronger Bounds on Braess's Paradox and the Maximum Latency of Selfish Routing. *SIAM J. Discrete Math.* 25 (2011), 1667–1686.

[17] Mehdi Mashayekhi, Nirav Ajmeri, George F. List, and Munindar P. Singh. 2022. Prosocial Norm Emergence in Multi-agent Systems. *ACM Trans. Auton. Adapt. Syst.* 17, 1–2 (2022), 1–24.

[18] Maja J Mataric. 1994. Reward Functions for Accelerated Learning. In *Machine Learning Proceedings 1994*, William W. Cohen and Haym Hirsh (Eds.). Morgan Kaufmann, San Francisco (CA).

[19] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A Survey on Bias and Fairness in Machine Learning. *ACM Comput. Surv.* 54, 6 (2021), 1–35.

[20] Andreasa Morris-Martin, Marina De Vos, and Julian Padget. 2021. A Norm Emergence Framework for Normative MAS – Position Paper. In *Coordination, Organizations, Institutions, Norms, and Ethics for Governance of Multi-Agent Systems XIII*, Andrea Aler Tubella, Stephen Cranefield, Christopher Frantz, Felipe Meneguzzi, and Wamberto Vasconcelos (Eds.). Springer International Publishing, Cham, 156–174.

[21] Noam Nisan and Amir Ronen. 2004. Computationally Feasible VCG Mechanisms. *Journal of Artificial Intelligence Research* 29 (2004), 242–252.

[22] Michael Oesterle, Christian Bartelt, Stefan Lüdtke, and Heiner Stuckenschmidt. 2022. Self-learning Governance of Black-Box Multi-Agent Systems. In *Coordination, Organizations, Institutions, Norms, and Ethics for Governance of Multi-Agent Systems XV - International Workshop, COINE 2022, Revised Selected Papers (Lecture Notes in Computer Science, Vol. 13549)*, Nirav Ajmeri, Andreasa Morris-Martin, and Bastin Tony Roy Savarimuthu (Eds.). Springer, Cham, 73–91.

[23] Michael Oesterle and Tim Grams. 2024. DRAMA at the PettingZoo: Dynamically Restricted Action Spaces for Multi-Agent Reinforcement Learning Frameworks. (2024). To appear in: Proceedings of the 57th Hawaii International Conference on System Sciences (HICSS).

[24] Michael Oesterle and Guni Sharon. 2023. Socially Optimal Non-discriminatory Restrictions for Continuous-Action Games. *Proceedings of the AAAI Conference on Artificial Intelligence* 37, 10 (2023), 11638–11646.

[25] United States. Bureau of Public Roads. 1964. *Traffic Assignment Manual for Application with a Large, High Speed Computer.* U. S. Department of Commerce, Bureau of Public Roads, Office of Planning, Urban Planning Division, Washington, D.C.

[26] Mancur Olson. 1965. *The logic of collective action: public goods and the theory of groups.* Number 124 in Harvard economic studies. Harvard Univ. Press, Cambridge, Mass.

[27] World Health Organization. 2023. Health equity. https://www.who.int/health-topics/health-equity

[28] Marco Pala, Hermann Sellier, B. Hackens, Frederico Martins, Vincent Bayot, and Serge Huant. 2012. A new transport phenomenon in nanostructures: A mesoscopic analog of the Braess paradox encountered in road networks. *Nanoscale research letters* 7 (2012), 472–472.

[29] Eric I. Pas and Shari L. Principio. 1997. Braess' paradox: Some new insights. *Transportation Research Part B: Methodological* 31, 3 (1997), 265–276.

[30] Claude M. Penchina. 1997. Braess paradox: Maximum penalty in a minimal critical network. *Transportation Research Part A: Policy and Practice* 31, 5 (1997), 379–388.

[31] Michael Pernpeintner, Christian Bartelt, and Heiner Stuckenschmidt. 2021. Governing Black-Box Agents in Competitive Multi-Agent Systems. In *Multi-Agent Systems: 18th European Conference, EUMAS 2021, Virtual Event, June 28–29, 2021, Revised Selected Papers.* Springer-Verlag, Berlin, Heidelberg, 19–36.

[32] Martin L. Puterman. 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming* (1st ed.). John Wiley & Sons, Inc., USA.

[33] Alvin Rajkomar, Michaela Hardt, Michael Howell, Greg Corrado, and Marshall Chin. 2018. Ensuring Fairness in Machine Learning to Advance Health Equity. *Annals of Internal Medicine* 169 (2018), 866–872.

[34] A. Rapoport and A.M. Chammah. 1965. *Prisoner's Dilemma: A Study in Conflict and Cooperation.* University of Michigan Press, Michigan.

[35] Anatol Rapoport and Albert M. Chammah. 1966. The Game of Chicken. *American Behavioral Scientist* 10, 3 (1966), 10–28.

[36] Tim Roughgarden. 2006. On the severity of Braess's Paradox: Designing networks for selfish users is hard. *J. Comput. Syst. Sci.* 72 (2006), 922–953.

[37] Tim Roughgarden and Éva Tardos. 2002. How Bad is Selfish Routing? *J. ACM* 49, 2 (2002), 236–259.

[38] L Rychetnik, Michael Frommer, P Hawe, and Alan Shiell. 2002. Criteria for evaluating evidence on public health interventions. *Journal of epidemiology and community health* 56 (2002), 119–127.

[39] Amartya Sen. 1973. Welfare Economics, Utilitarianism, and Equity. In *On Economic Inequality.* Oxford University Press, Oxford, UK.

[40] Lloyd S. Shapley. 1953. Stochastic Games. *Proceedings of the National Academy of Sciences* 39 (1953), 1095–1100.

[41] Guni Sharon, Josiah P. Hanna, Tarun Rambha, Michael W. Levin, Michael Albert, Stephen D. Boyles, and Peter Stone. 2017. Real-Time Adaptive Tolling Scheme for Optimized Social Welfare in Traffic Networks. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems* (São Paulo, Brazil) *(AAMAS '17).* International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 828–836.

[42] Yoav Shoham and Moshe Tennenholtz. 1995. On social laws for artificial agent societies: off-line design. *Artificial Intelligence* 73, 1 (1995), 231–252.

[43] Brian Skyrms. 2003. *The Stag Hunt and the Evolution of Social Structure.* Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781139165228

[44] Richard Steinberg and Willard I. Zangwill. 1983. The Prevalence of Braess' Paradox. *Transportation Science* 17, 3 (1983), 301–318.

[45] Justin Svegliato, Samer B. Nashed, and Shlomo Zilberstein. 2021. Ethically Compliant Sequential Decision Making. *Proceedings of the AAAI Conference on Artificial Intelligence* 35, 13 (2021), 11657–11665.

[46] Jordan Terry, Benjamin Black, Nathaniel Grammel, Mario Jayakumar, Ananth Hari, Ryan Sullivan, Luis S Santos, Clemens Dieffendahl, Caroline Horsch, Rodrigo Perez-Vicente, et al. 2021. Pettingzoo: Gym for multi-agent reinforcement learning. *Advances in Neural Information Processing Systems* 34 (2021), 15032–15043.

[47] Ralph Turvey. 1969. Marginal Cost. *The Economic Journal* 79, 314 (1969), 282–299.

[48] Gregory Valiant and Tim Roughgarden. 2010. Braess's Paradox in large random graphs. *Random Struct. Algorithms* 37 (2010), 495–515.

[49] Paul A.M. Van Lange, Jeff Joireman, Craig D. Parks, and Eric Van Dijk. 2013. The psychology of social dilemmas: A review. *Organizational Behavior and Human Decision Processes* 120, 2 (2013), 125–141.

[50] Vadim E. Zverovich and Erel Avineri. 2012. Braess' Paradox in a Generalised Traffic Network. *ArXiv* abs/1207.3251 (2012), 114–138.