To Spend or to Gain: Online Learning in Repeated Karma Auctions

Damien Berriaud
ETH Zürich
Zürich, Switzerland
dberriaud@ethz.ch

Ezzat Elokda
ETH Zürich
Zürich, Switzerland
elokdae@ethz.ch

Devansh Jalota
Stanford University
Stanford, California, USA
djalota@stanford.edu

Emilio Frazzoli
ETH Zürich
Zürich, Switzerland
efrazzoli@ethz.ch

Marco Pavone
Stanford University
Stanford, California, USA
pavone@stanford.edu

Florian Dörfler
ETH Zürich
Zürich, Switzerland
dorfler@ethz.ch

ABSTRACT

Recent years have seen a surge of artificial currency-based mechanisms in contexts where monetary instruments are deemed unfair or inappropriate, e.g., in allocating food donations to food banks, course seats to students, and, more recently, even for traffic congestion management. Yet the applicability of these mechanisms remains limited in repeated auction settings, as it is challenging for users to learn how to bid an artificial currency that has no value outside the auctions. Indeed, users must jointly learn the value of the currency in addition to how to spend it optimally. Moreover, in the prominent class of *karma mechanisms*, in which artificial *karma payments* are redistributed to users at each time step, users do not only *spend karma* to obtain public resources but also *gain karma* for yielding them. For this novel class of karma auctions, we propose an *adaptive karma pacing strategy* that learns to bid optimally, and show that this strategy a) is asymptotically optimal for a single user bidding against competing bids drawn from a stationary distribution; b) leads to convergent learning dynamics when all users adopt it; and c) constitutes an approximate Nash equilibrium as the number of users grows. Our results require a novel analysis in comparison to adaptive pacing strategies in monetary auctions, since we depart from the classical assumption that the currency has known value outside the auctions, and consider that the currency is both spent *and* gained through the redistribution of payments.

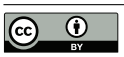
CCS CONCEPTS

• **Theory of computation → Convergence and learning in games**.

KEYWORDS

Online learning; Artificial currency; Karma economy; Repeated auctions; Budget-constrained auctions; Adaptive pacing.

ACM Reference Format:
Damien Berriaud, Ezzat Elokda, Devansh Jalota, Emilio Frazzoli, Marco Pavone, and Florian Dörfler. 2025. To Spend or to Gain: Online Learning in Repeated Karma Auctions. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 9 pages.



Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

1 INTRODUCTION

In public resource allocation contexts, the use of monetary instruments to regulate resource consumption is often deemed inequitable (e.g., to manage traffic congestion [2, 10, 15, 40]), inappropriate (e.g., for organ and food donations [29, 33, 39] or course allocations [11]), or simply undesired (e.g., for peer-to-peer file sharing [17, 42] or babysitting services [28]). As a consequence, significant attention has been devoted to the study of non-monetary mechanism design [36, 38], which is known to be challenging due to interpersonal comparability [34] and the lack of a general instrument to manipulate incentives [20, 37].

However, a number of mechanisms have seen some recent success in jointly achieving the objectives of fairness, efficiency, and strategy-proofness when resources are allocated *repeatedly over time* [6, 8, 21–24]. The core principle of these mechanisms is to restrict the number of times the resource can be consumed and let the users trade off when it is most beneficial for them to do so. To achieve these goals, many of these mechanisms employ *artificial currencies* [8, 13, 21–23, 28, 33], which involves issuing a budget of non-tradable credits or currency to users which they may use to repeatedly bid for resources. In artificial currency mechanisms, users, who may have time-varying and stochastic valuations for the resources, must be strategic in their bidding to not deplete the budget too quickly, and to spare currency for periods when they have the highest valuation for the resources. Thus, artificial currencies serve the dual purpose of monitoring resource consumption and providing a means for users to express their time-varying preferences, resulting in fair and efficient allocations over time.

The literature on artificial currency mechanisms for repeated resource allocation can be broadly categorized in two classes. In the first class, artificial currency is issued at the beginning of a finite episode only to be spent during the episode [8, 19, 21, 22]. In the second class, artificial currency is issued at the beginning of the episode but can also be gained throughout it, typically by means of peer-to-peer exchanges [13, 17, 28, 42] or by redistributing the payments collected in each time step [14, 33]. Some works have referred to this class of artificial currencies as *karma* [13, 42, 43]: when users yield resources to others they gain karma, and when instead they consume resources they lose karma. In comparison to

mechanisms relying on initial endowments of artificial currency only, karma mechanisms are particularly suited to settings in which resources are allocated indefinitely with no finite horizon in sight.

In both classes of artificial currency mechanisms, the focus thus far has been on analyzing equilibrium properties, including existence [13], strategy-proofness [21, 23], and robustness [8, 22]. However, few works have considered the problem of *learning how to optimally bid artificial currency in repeated auction settings*, and whether such a learning procedure converges to a Nash equilibrium. This problem holds both significant importance and challenge. The importance is two-fold: on one hand, the equilibrium-based analysis of previous studies is only meaningful if an equilibrium is reached; on the other, devising simple learning rules that align with users' self-interest is crucial to implement these mechanisms in practice. The challenge stems from the fact that, unlike traditional monetary instruments, artificial currency does not have any value outside the resource allocation context for which it has been issued. Therefore, users must jointly learn the value of the currency as well as how to spend it optimally. Moreover, the possibility of gaining currency in the class of karma mechanisms leads to new challenges that are particular to these novel mechanisms. In the related setting of online Fisher markets, [19] proposes a distributed algorithm to compute online market equilibria but does not analyze convergence to Nash equilibria, and does not consider currency gains.

The problem of learning how to bid optimally in *monetary* auctions is a classical one [30, 32]. This problem has gained recent traction in the context of repeated, budget-constrained auctions [3, 5, 7, 12, 18, 31, 44], most famously to automate the bidding in multi-period online ad campaigns. We draw inspiration from these works, and in particular [5], to derive *adaptive pacing strategies* in artificial currency-based auctions. The first class of artificial currency auctions in which users are issued an initial endowment of currency only is most closely related to works considering *value maximization* [4, 18, 31]. In these works, monetary payment costs are not included in the users' optimization objective, but these monetary costs still enter the optimization explicitly in constraints, either on individual bids [18] or on the total expenditure (also known as *return of investment constraints*) [31]. In contrast, the cost of spending currency must be learned and does not appear in the optimization objective nor constraints in artificial currency auctions.

Moreover, the second class of karma auctions in which karma is gained throughout the auction campaign leads to new strategic opportunities that are not typically considered in monetary settings. For instance, even in a second-price auction users have an incentive to bid non-truthfully to maximize the karma gained upon losing. Furthermore, the preservation of total karma held by the users in this class of auctions leads to challenges in the simultaneous adoption of classical adaptive pacing strategies, since it becomes impossible for all users to simultaneously deplete their budgets.

1.1 Contribution

In this paper, we adopt techniques from adaptive pacing in budget-constrained monetary auctions [5] to the two aforementioned classes of artificial currency mechanisms. The first class of artificial currency mechanisms in which users are issued an initial endowment of currency only is addressed in the full version of the paper [9]. In the following, we instead focus on the second and more challenging class of karma mechanisms. We specifically consider mechanisms in which the karma payments are redistributed uniformly in each time step. We derive an *adaptive karma pacing* strategy for these mechanisms, and show that: a) adaptive karma pacing is asymptotically optimal for a single user bidding against competing bids drawn from a stationary distribution; b) when all users adopt adaptive karma pacing, the expected dynamics converge asymptotically to a unique stationary point; and c) adaptive karma pacing constitutes an approximate Nash equilibrium under the additional assumption that there is a large number of parallel auctions for which the matching probability of any two particular users decays asymptotically to zero. Performing an asymptotic analysis is natural for karma mechanisms as they can be infinitely repeated. The novel technical challenges of the karma-based setting in comparison to the monetary setting, and how they are tackled in our paper, are summarized as follows:

- The possibility of gaining karma through redistribution leads to non-truthfulness of the second-price auction, and a complex dependency of the karma budget dynamics on the user's bid. This requires relaxing the primal problem's budget constraint and using non-perfect gradient information in the associated relaxed dual problem;
- The preservation of karma due to redistribution leads to several complications requiring a novel adaptive karma pacing strategy and analysis. It is impossible for all agents to simultaneously deplete their budgets in this setting, which requires removing the target expenditure rate from the multiplier updates. This leads to non-uniqueness of the optimal stationary multipliers, requiring further strategy modifications to converge to a unique multiplier profile.

As a consequence of these modifications to standard adaptive pacing, our paper performs a novel regret analysis in order to extend previous performance guarantees to the karma-based setting.

The remainder of the paper is organized as follows. In Section 2 we introduce the problem formulation including key definitions and notations. We derive our adaptive karma pacing strategy in Section 3. Our main results are then included in Section 4 which establishes performance guarantees for this strategy. Finally, Section 5 discusses the results, shedding light on the key assumptions made and providing directions for future work. In the full version of the paper [9], we additionally include a detailed literature review, results for artificial currency mechanisms with no redistribution, supplementary numerical experiments, and detailed proofs.

2 PROBLEM SETUP

This section introduces the setting studied in the paper, including notations and important definitions.

2.1 Notation

We denote by $[N]$ the set $\{1, \dots, N\}$, by $\mathbb{1}\{\cdot\}$ the indicator function, by $(\cdot)^+$ the function $x \mapsto \max\{x, 0\}$, and by $P_{[a,b]}(\cdot)$ the projection $x \mapsto \min\{\max\{x, a\}, b\}$. Scalars x are distinguished from vectors $\boldsymbol{x} = (x_i)_{i \in I}$, for some index set I, through the use of boldface. If x is a scalar, then $\underline{x}$ (respectively, $\overline{x}$) is a lower bound (respectively, upper bound) of x. If $\boldsymbol{x}$ is a vector, then $\underline{x} = \min_{i \in I} x_i$ (respectively,

$\overline{x} = \max_{i \in I} x_i$). Finally, for the vector $\boldsymbol{x}$ and an index $i \in I$, the vector $\boldsymbol{x_{-i}} = (x_j)_{j \in I,\, i \neq j}$ is constructed by dropping component i.

2.2 Setting

We study a general class of repeated resource allocation problems in which a limited number of resources must be repeatedly allocated to a population $\mathcal{N} = [N]$ of agents. For the sake of presentation, we instantiate this class of problems using a stylized morning commute setting [1, 41], which is schematically illustrated in Figure 1. At discrete time steps $t \in \mathbb{N}$ (e.g., days), the agents seek to commute from the suburb to the city center using one of two roads. The *general purpose road* is subject to congestion, while access to the *priority road* is limited to its free-flow capacity of $\gamma \in [N-1]$ agents per time step. Traveling on the general purpose road takes unit time, while traveling on the priority road takes a shorter time $0 \leq 1 - \Delta < 1$. This model can be interpreted as an abstraction of a multi-lane highway with a governed express lane and an un-governed, congested general purpose lane. At each time step $t \in \mathbb{N}$, each agent $i \in \mathcal{N}$ is associated with a private *valuation of time* $v_{i,t} \in [0,1]$ drawn independently across time from fixed, exogenous distributions $\mathcal{V}_i$. The valuations represent the agents' time-varying sensitivities to travel delays, e.g., because they have flexible schedules on some days but must be punctual on other days, and are normalized to the interval $[0,1]$ without loss of generality. We denote by $\boldsymbol{v_t} = (v_{i,t})_{i \in \mathcal{N}}$ the vector of agents' valuations at time t, which are distributed according to $\boldsymbol{\mathcal{V}} = \prod_{i \in \mathcal{N}} \mathcal{V}_i$ with support over $[0,1]^N$. As is common in the literature [5, 12, 21], we assume that the valuation distribution $\boldsymbol{\mathcal{V}}$ is absolutely continuous with bounded density $\nu : [0,1]^N \mapsto \left[\underline{\nu}, \overline{\nu}\right]^N \subset \mathbb{R}^N_{>0}$.

Access to the priority road is governed by means of an artificial currency called *karma*. Each agent $i \in \mathcal{N}$ is endowed with an initial karma budget $k_{i,1} \in \mathbb{R}_+$. Then, at each time step $t \in \mathbb{N}$, each agent places a sealed bid $b_{i,t} \in \mathbb{R}_+$ smaller than its current budget $k_{i,t} \in \mathbb{R}_+$. The γ-highest bidders, referred to as the 'auction winners', are granted access to the priority road, and must pay $p_t^{\gamma+1} := \gamma + 1^{\text{th}}\text{-}\max\{b_{i,t}\}_{i=1}^N$ in karma. The $N-\gamma$ remaining agents, referred to as the 'auction losers', must instead take the general purpose road and make no payments. The price $p_t^{\gamma+1}$ is set by the highest bid among the auction losers, i.e., it corresponds to the second price auction if $\gamma = 1$. After payments are settled, they are redistributed uniformly to the agents such that each agent gains $g_{i,t} = \gamma\, p_t^{\gamma+1}/N$ units of karma in the next time step. Notice that under this redistribution scheme, agents that access the priority road have a net decrease in karma, while those using the general purpose road have a net increase in karma. Meanwhile, the total amount of karma in the system set by the initial endowments $\boldsymbol{k_1}$ is preserved over time.

Let $\boldsymbol{b_{-i,t}} = (b_{j,t})_{j \neq i}$ be the bid profile of agents other than i, and $d_{i,t}^{\gamma} = \gamma^{\text{th}}\text{-}\max_{j:j\neq i}\{b_{j,t}\}$ be the associated *competing bid*, since agent i must bid higher than $d_{i,t}^{\gamma}$ to be among the auction winners. We assume that ties in bids do not occur (as is common in the literature [5]; in practice ties could be settled randomly). Let $x_{i,t} = \mathbb{1}\{b_{i,t} > d_{i,t}^{\gamma}\} \in \{0,1\}$ indicate whether agent i is an auction winner at time t. Then the agent suffers a cost $c_{i,t} = v_{i,t}\,(1 - x_{i,t}\Delta)$ and pays $z_{i,t} = x_{i,t}\, d_{i,t}^{\gamma}$ at that time. Their budget for the next time step is hence determined by $k_{i,t+1} = k_{i,t} - z_{i,t} + g_{i,t}$.

We consider rational agents that aim to minimize their expected total cost over a time horizon T. At time t, the information available to agent i to formulate its bid is the *history* $\mathcal{H}_{i,t}^T = \left\{T, (v_{i,s}, k_{i,s}, b_{i,s}, z_{i,s}, c_{i,s})_{s=1}^{t-1}, v_{i,t}, k_{i,t}\right\}$. A bidding strategy $\beta_i^T \in \mathcal{B}^T$ for agent i is thus a sequence of functions $\beta_i^T = (\beta_{i,1}^T, \ldots, \beta_{i,T}^T)$, where $\beta_{i,t}^T$ maps the history $\mathcal{H}_{i,t}^T$ to a probability distribution over the set of feasible bids $[0, k_{i,t}]$. A profile of strategies for all agents is accordingly denoted by $\boldsymbol{\beta}^T = \left(\beta_i^T\right)_{i \in \mathcal{N}}$. For a fixed agent i following strategy β_i^T, it will be convenient to define three notions of expected total cost, given by

$$C_i^{\beta_i^T}(\boldsymbol{v_i}, \boldsymbol{d_i}) = \mathop{\mathbb{E}}_{\boldsymbol{b_i} \sim \beta_i^T}\left[\sum_{t=1}^T c_{i,t}\right] = \mathop{\mathbb{E}}_{\boldsymbol{b_i} \sim \beta_i^T}\left[\sum_{t=1}^T v_{i,t}\left(1 - \mathbb{1}\left\{b_{i,t} > d_{i,t}^{\gamma}\right\}\Delta\right)\right], \tag{1}$$

$$C_i^{\beta_i^T} = \mathop{\mathbb{E}}_{\boldsymbol{v_i} \sim \prod_{t=1}^T \mathcal{V}_i,\ \boldsymbol{d_i} \sim \prod_{t=1}^T \mathcal{D}_i,\ \boldsymbol{b_i} \sim \beta_i^T}\left[C_i^{\beta_i^T}(\boldsymbol{v_i}, \boldsymbol{d_i})\right], \tag{2}$$

$$C_i^{\boldsymbol{\beta}^T} = \mathop{\mathbb{E}}_{\boldsymbol{v} \sim \prod_{t=1}^T \boldsymbol{\mathcal{V}},\ \boldsymbol{b_{-i}} \sim \boldsymbol{\beta}_{-i}^T}\left[C_i^{\beta_i^T}(\boldsymbol{v_i}, \boldsymbol{d_i}(\boldsymbol{b_{-i}}))\right]. \tag{3}$$

Equation (1) defines the *sample path cost* $C_i^{\beta_i^T}\left(\boldsymbol{v_i}, \boldsymbol{d_i^{\gamma}}\right)$ for a fixed realization of valuations $\boldsymbol{v_i} = (v_{i,t})_{t \in [T]}$ and competing bids $\boldsymbol{d_i} := \left(d_{i,t}^{\gamma}, d_{i,t}^{\gamma+1}\right)_{t \in [T]}$. Equation (2) defines the *stationary competition cost* $C_i^{\beta_i^T}$, in which the competing bids $d_{i,t}$ are assumed to be drawn independently across time from a stationary distribution $\mathcal{D}_i$. Finally, Equation (3) defines the *strategic competition cost* $C_i^{\boldsymbol{\beta}^T}$, which is agent i's expected total cost when all agents follow strategy profile $\boldsymbol{\beta}^T$. Finally, we define the infinite series of strategy profiles $\boldsymbol{\beta} = \left(\boldsymbol{\beta}^T\right)_{T \in \mathbb{N}}$ and say that strategy profile $\boldsymbol{\beta}$ constitutes an *approximate Nash equilibrium* if its strategic competition cost satisfies, for all agents $i \in \mathcal{N}$,

$$\lim_{T \to \infty} \frac{1}{T}\left(C_i^{\boldsymbol{\beta}^T} - \inf_{\tilde{\beta}_i^T \in \mathcal{B}^T} C_i^{\tilde{\beta}_i^T, \boldsymbol{\beta}_{-i}^T}\right) = 0. \tag{4}$$

Equation (4) implies that under strategy profile $\boldsymbol{\beta}$, no single agent i can asymptotically improve its expected average cost per time step by unilaterally deviating to a strategy $\tilde{\beta}_i^T \neq \beta_i^T$.

3 DERIVATION OF ADAPTIVE KARMA PACING

In the remainder of the paper, our main goal is to devise a bidding strategy that constitutes an approximate Nash equilibrium when all agents follow it. As a first step towards this goal, this section derives a candidate optimal bidding strategy using an *online dual gradient ascent scheme*. This classical optimization technique has gained recent traction in the context of budget-constrained auctions [5] and other related problems [7, 25–27]. To elucidate our bidding strategy, we first introduce the optimization problem of a single agent $i \in \mathcal{N}$ who has the *benefit of hindsight*, i.e., who can make optimal bidding decisions with prior knowledge of the future realizations of

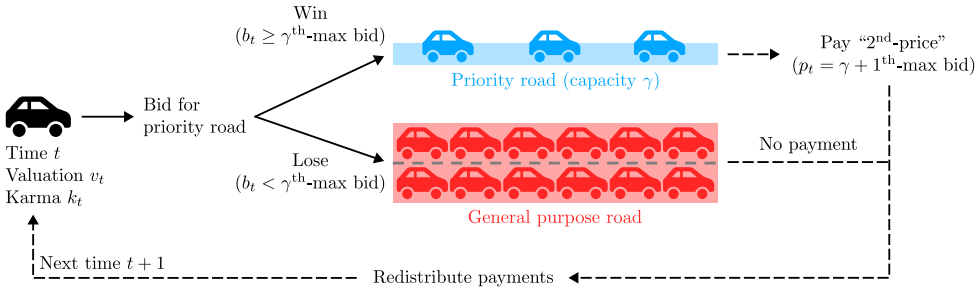


Figure 1: Schematic representation of repeated resource allocation using karma.

valuations $\boldsymbol{v_i}$ and competing bids $\boldsymbol{d_i}$. Thus, the optimal cost of this problem serves as a theoretical benchmark for the lowest cost that the agent can hope to achieve. Then, since in practice the agent only observes the stochastic valuations and competing bids online as the auctions progress, we introduce a candidate bidding strategy, based on online gradient ascent, to approximate the agent's optimal bidding strategy with the benefit of hindsight.

Optimal Cost with the Benefit of Hindsight. We construct hereafter a lower bound on the optimal cost with the benefit of hindsight. For a fixed realization of valuations $\boldsymbol{v_i}$ and competing bids $\boldsymbol{d_i}$, agent i's optimal cost with the benefit of hindsight is given by the following optimization problem

$$\begin{aligned} C_i^H(\boldsymbol{v_i}, \boldsymbol{d_i}) = \min_{\boldsymbol{b_i} \in \mathbb{R}_+^T} \sum_{t=1}^{T} v_{i,t} \left(1 - \mathbb{1}\left\{b_{i,t} > d_{i,t}^\gamma\right\} \Delta\right), \text{ such that} \\ \sum_{t=1}^{s} \mathbb{1}\left\{b_{i,t} > d_{i,t}^\gamma\right\} d_{i,t}^\gamma \leq \rho_i T + \sum_{t=1}^{s-1} g_{i,t}\left(b_{i,t}, d_{i,t}\right), \ \forall s \in [T], \end{aligned} \tag{5}$$

where, following the standard literature [5], we define $\rho_i = k_{i,1}/T$ as the *target expenditure rate*, i.e., the average expenditure per time step that would fully deplete the initial budget by the end of the time horizon if no karma was gained. Notice that Problem (5) bears significant complexity in comparison to its counterpart in the standard setting with no budget gains. First, the possibility of gaining karma requires a budget constraint for each time step $s \in [T]$ instead of only one at the end of the horizon. Second, the outcomes $x_{i,t}$ cannot be used directly as decision variables since the gains $g_{i,t}$ depend non-trivially on the bids $b_{i,t}$, as given by

$$g_{i,t}\left(b_{i,t}, d_{i,t}\right) = \frac{\gamma}{N} p_t^{\gamma+1}\left(b_{i,t}, d_{i,t}\right) = \frac{\gamma}{N} \begin{cases} d_{i,t}^\gamma, & d_{i,t}^\gamma < b_{i,t}, \\ b_{i,t}, & d_{i,t}^{\gamma+1} < b_{i,t} \leq d_{i,t}^\gamma, \\ d_{i,t}^{\gamma+1}, & b_{i,t} \leq d_{i,t}^{\gamma+1}. \end{cases} \tag{6}$$

Finally, notice that with respect to the standard setting, Problem (5) has an additional dependency on $d_{i,t}^{\gamma+1}$, i.e., the $\gamma+1^{\text{th}}$-highest competing bid, c.f. Equation (6). If agent i is among the auction winners, the price $p_t^{\gamma+1}$ and thereby the gain $g_{i,t}$ is determined by $d_{i,t}^\gamma$; if the agent is among the auction losers and is not the *price setter*, the gain is determined by $d_{i,t}^{\gamma+1}$; and if the agent is among the auction losers but sets the price the gain is determined by its own bid $b_{i,t}$. To address the complexity of Problem (5), we perform a relaxation that forms a lower bound on the optimal cost with the benefit of hindsight, given by

$$\begin{aligned} C_i^H(\boldsymbol{v_i}, \boldsymbol{d_i}) \geq \underline{C}_i^H(\boldsymbol{v_i}, \boldsymbol{d_i^\gamma}) = \min_{\boldsymbol{x_i} \in \{0,1\}^T} \sum_{t=1}^{T} v_{i,t}(1 - x_{i,t}\Delta), \\ \text{s.t.} \sum_{t=1}^{T} \left(x_{i,t} - \frac{\gamma}{N}\right) d_{i,t}^\gamma \leq \rho_i T. \end{aligned} \tag{7}$$

This lower bound is obtained by a) allowing temporary negative balances of karma as long as it is non-negative at the end of the horizon T; and b) eliminating the dependency of $g_{i,t}$ on $b_{i,t}$ and $d_{i,t}^{\gamma+1}$ by assuming that when the agent is among the auction losers, it always manages to be the price setter and impose the maximum gain $\frac{\gamma}{N} d_{i,t}^\gamma$. The Lagrangian dual problem associated with Problem (7) is

$$\begin{aligned} &C_i^H(\boldsymbol{v_i}, \boldsymbol{d_i}) \geq \delta_i^H(\boldsymbol{v_i}, \boldsymbol{d_i^\gamma}) \\ &:= \sup_{\mu_i \geq 0} \min_{\boldsymbol{x_i} \in \{0,1\}^T} \sum_{t=1}^{T} x_{i,t}\left(\mu_i d_{i,t}^\gamma - \Delta v_{i,t}\right) + v_{i,t} - \mu_i\left(\rho_i + \frac{\gamma}{N} d_{i,t}^\gamma\right) \end{aligned} \tag{8}$$

The relaxation in Problem (7) results in dual Problem (8) with similar structure to its counterpart in the standard setting with no budget gains. The main difference is that the target expenditure rate ρ_i is replaced by the time-varying term $\rho_i + \frac{\gamma}{N} d_{i,t}^\gamma$. This is intuitive as the agent now aims to deplete both its initial budget as well as the gains it receives. For a fixed multiplier $\mu_i \geq 0$, the inner minimum in (8) is obtained by winning all auctions satisfying $\Delta v_{i,t} > \mu_i d_{i,t}^\gamma$. This can be achieved by bidding $b_{i,t} = \Delta v_{i,t}/\mu_i$, yielding

$$\begin{aligned} \delta_i^H(\boldsymbol{v_i}, \boldsymbol{d_i^\gamma}) &= \sup_{\mu_i \geq 0} \sum_{t=1}^{T} v_{i,t} - \mu_i\left(\rho_i + \frac{\gamma}{N} d_{i,t}^\gamma\right) - \left(\Delta v_{i,t} - \mu_i d_{i,t}^\gamma\right)^+ \\ &:= \sup_{\mu_i \geq 0} \sum_{t=1}^{T} \delta_{i,t}^H(v_{i,t}, d_{i,t}^\gamma, \mu_i). \end{aligned} \tag{9}$$

Adaptive Karma Pacing. We perform a stochastic gradient ascent scheme in order to approximately solve the relaxed dual Problem (9) using online observations. Namely, the agent considers a candidate optimal multiplier $\mu_{i,t}$ and places its bid accordingly with $b_{i,t} = \Delta v_{i,t}/\mu_{i,t}$. In an ideal case, it would then update $\mu_{i,t+1}$ using the subgradient given by $\frac{\partial \delta_{i,t}^H}{\partial \mu_{i,t}}(v_{i,t}, d_{i,t}^\gamma, \mu_{i,t}) = z_{i,t} - \rho_i - \frac{\gamma}{N} d_{i,t}^\gamma$. However, adopting this subgradient raises two issues. First, the term $\frac{\gamma}{N} d_{i,t}^\gamma$ is only observed by auction winners, hence we use the observed gain $g_{i,t}$ as a proxy instead. Second, if all agents are to adopt this subgradient, it is impossible for them to simultaneously track ρ_i and fully deplete their karma by the end of the horizon, as the total amount of karma in the system is preserved. For this reason, we will omit the term ρ_i, such that each agent attempts to match

its expenditures to its gains. This yields the *adaptive karma pacing strategy*, which is denoted by K and summarized in Algorithm 1.

ALGORITHM 1: Adaptive Karma Pacing K

Input: Time horizon T, initial budget $k_{i,1} > 0$, multiplier bounds $\overline{\mu} > \underline{\mu} > 0$, gradient step size $\epsilon > 0$.
Initialize: Initial multiplier $\mu_{i,1} \in [\underline{\mu}, \overline{\mu}]$.
for $t = 1, \dots, T$ **do**
(1) Observe the realized valuation $v_{i,t}$ and place bid
$$b_{i,t} = \min\left\{\frac{\Delta v_{i,t}}{P_{[\underline{\mu},\overline{\mu}]}(\mu_{i,t})}, k_{i,t}\right\}; \quad (K\text{-}b)$$
(2) Observe the expenditure $z_{i,t}$ and the gain $g_{i,t}$. Update the multiplier $\mu_{i,t+1} = \mu_{i,t} + \epsilon(z_{i,t} - g_{i,t})$, as well as the karma budget $k_{i,t+1} = k_{i,t} - z_{i,t} + g_{i,t}$. $(K\text{-}\mu)$
end

The term 'adaptive karma pacing' is in line with the literature on budget-constrained monetary auctions [5, 16], but instead of trying to *pace* the budget depletion rate to match the target rate ρ_i, strategy K attempts to match the time-varying expenses $z_{i,t}$ with the gains $g_{i,t}$. Indeed, karma losses $z_{i,t} - g_{i,t} > 0$ increase $\mu_{i,t+1}$, effectively reducing future bids; and vice versa for $z_{i,t} - g_{i,t} < 0$. Another important novelty in strategy K is that the denominator in the bid (K-b) is $\mu_{i,t}$ instead of $1 + \mu_{i,t}$, as common in the standard monetary setting. This is a consequence of the fact that the valuation in karma is not known a-priori; and could lead to a rapid depletion of the budget if $\mu_{i,t}$ becomes small during the learning process even for a short transient period. For this reason, it is necessary to introduce the lower bound $\underline{\mu}$ in Algorithm 1, which we note is unlike adaptive pacing algorithms in the monetary setting [5]. Moreover, the projection of multiplier μ_i on $[\underline{\mu}, \overline{\mu}]$ now occurs in the bid (K-b) instead of the multiplier update (K-μ). The importance of this technical difference will be discussed in Section 4.2.

4 ANALYSIS OF ADAPTIVE KARMA PACING

In this section, we analyze the previously derived adaptive karma pacing strategy K, with the main goal of establishing that it constitutes an approximate Nash equilibrium when adopted by all agents. To achieve this goal, this section proceeds as follows. In section 4.1, we establish that this strategy is asymptotically optimal for a single agent bidding against competing bids drawn from a stationary distribution (Theorem 4.1). Section 4.2 then establishes that the learning dynamics converge to a unique stationary point when all agents follow strategy K (Theorem 4.2). Finally, Section 4.3 combines these results to achieve the main goal of proving that the strategy constitutes an approximate Nash equilibrium under suitable conditions (Theorem 4.3).

4.1 Asymptotic Optimality under Stationary Competition

In this section, we establish that strategy K is *asymptotically optimal in a stationary competition setting*, where a single agent i bids against competing bids $\boldsymbol{d_i} = \left(d_{i,t}^{\gamma}, d_{i,t}^{\gamma+1}\right)_{t\in[T]}$ drawn independently across time from a fixed distribution $\mathcal{D}_i$. This section, as well as Sections 4.2 and 4.3, are organized as follows. We first state the required definitions and the novel assumptions needed in the karma-based setting[1]. We then present a brief statement of the main result of the section (Theorem 4.1 in this case)[2], and focus our discussion on the differences to the standard setting with no budget gains.

To state the main result of this section, we must first define the *expected dual objective*, *expected gain*, *expected expenditure*, and *expected karma loss* when agent i follows strategy K. For a fixed multiplier $\mu_i > 0$, these quantities are respectively given by

$$\Psi_i^0(\mu_i) = \mathop{\mathbb{E}}_{v_i,d_i}\left[v_i - \mu_i g_i - (\Delta v_i - \mu_i d_i^{\gamma})^+\right], \quad G_i(\mu_i) = \mathop{\mathbb{E}}_{v_i,d_i}\left[g_i\right],$$
$$Z_i(\mu_i) = \mathop{\mathbb{E}}_{v_i,d_i}\left[d_i^{\gamma}\mathbb{1}\{\Delta v_i > \mu_i d_i^{\gamma}\}\right], \quad L_i(\mu_i) = Z_i(\mu_i) - G_i(\mu_i), \quad (10)$$

where the expectation is with respect to the stationary distributions $\mathcal{V}_i$ and $\mathcal{D}_i$. We make two observations on the definition of the expected dual objective. First, in line with the multiplier update (K-μ), the dual objective Ψ_i^0 that strategy K aims to maximize artificially considers a target expenditure rate of zero. Notice however that ρ_i appears in the expected optimal dual objective $\Psi_i^H(\mu_i) = \mathbb{E}_{v_i,d_i}\left[\delta_i^H\left(v_i, d_i^{\gamma}, \mu_i\right)\right]$ of Problem (9). For this reason, we require the initial budget to grow sublinearly with the time horizon, so as to control the difference between Ψ_i^0 and Ψ_i^H.

Assumption 1 (Initial Budget $k_{i,1}(T)$). *The initial budget $k_{i,1}$ is a function of T satisfying $\lim_{T\to\infty} k_{i,1}(T)/T = 0$.*

Second, Equation (10) is defined in terms of the actual gain g_i rather than the maximum possible gain $\frac{\gamma}{N} d_i^{\gamma}$ used in the relaxed problems (7)–(9). This discrepancy implies yet another gap between Ψ_i^0 and Ψ_i^H, since if agent i is not among the auction winners, it can gain more by bidding $d_i^{\gamma+1} < b_i \le d_i^{\gamma}$. For this reason, it is convenient to define the *residual gain* $\hat{\varepsilon} = \frac{\gamma}{N}\mathbb{E}_{d_i}\left[d_i^{\gamma} - d_i^{\gamma+1}\right]$, that is, the expected maximum additional gain that agent i can get by becoming the price setter.

We moreover denote by $\mu_i^{\star 0} > 0$ the *stationary multiplier* that satisfies $L_i\left(\mu_i^{\star 0}\right) = 0$ and causes the expected expenditure to equal the expected gain. Finally, we define the notion of *hitting time* as

$$\mathcal{T}_i = \min\left\{\mathcal{T}_i^k, \mathcal{T}_i^{\underline{\mu}}, \mathcal{T}_i^{\overline{\mu}}\right\}, \quad \text{where } \mathcal{T}_i^k = \mathop{\text{argmax}}_{t\in[T]}\left\{\forall s \in [t], k_{i,s} \ge \Delta/\underline{\mu}\right\},$$
$$\mathcal{T}_i^{\underline{\mu}} = \mathop{\text{argmax}}_{t\in[T]}\left\{\forall s \in [t], \mu_{i,s} \ge \underline{\mu}\right\}, \quad \mathcal{T}_i^{\overline{\mu}} = \mathop{\text{argmax}}_{t\in[T]}\left\{\forall s \in [t], \mu_{i,s} \le \overline{\mu}\right\}. \quad (11)$$

This is the latest time step which guarantees that $b_{i,t} = \Delta v_{i,t}/\mu_{i,t}$ in (K-b) for any valuation $v_{i,t} \in [0, 1]$. By definition, the hitting time is a stricter notion than the *budget depletion time* $\mathcal{T}_i^k$ used in the standard setting with no budget gains. This modification is needed since the projection occurs in the bid (K-b) instead of the multiplier update (K-μ) in strategy K, and in turn requires the following additional assumption to establish our result.

Assumption 2 (Control of Hitting Time). *The following holds:*
2.1 Distribution $\mathcal{V}_i$ has support in $\left[\underline{v_i}, 1\right]$, where $0 < \underline{v_i} < 1$;
2.2 Distribution $\mathcal{D}_i$ has support in $\left[\underline{d_i}, \overline{d_i}\right]^2$, where $0 < \underline{d_i} < \overline{d_i}$.

We are now ready to state the main result regarding the asymptotic optimality of strategy K under stationary competition.

[1] Standard assumptions in the literature, as well as minor technical assumptions, are included in the full version of the paper [9], Appendix B and D, respectively.
[2] Full theorem statements are included in the full version of the paper [9], Appendix D.

Theorem 4.1 (Asymptotic Optimality under Stationary Competition). *There exists a constant $C \in \mathbb{R}_+$ such that the average expected regret of an agent $i \in \mathcal{N}$ for following strategy K in the stationary competition setting satisfies*

$$\frac{1}{T} \mathop{\mathbb{E}}_{\boldsymbol{v_i}, \boldsymbol{d_i}} \left[C_i^K(\boldsymbol{v_i}, \boldsymbol{d_i}) - C_i^H(\boldsymbol{v_i}, \boldsymbol{d_i}) \right] \leq C \left(\epsilon + \frac{1}{\epsilon T} + \frac{k_{i,1}}{T} + \frac{\mathbb{E}_{\boldsymbol{v_i}, \boldsymbol{d_i}} [T - \mathcal{T}_i]}{T} + \hat{\varepsilon} \right)$$

Moreover, for suitably chosen parameters, strategy K asymptotically converges to an $O(\hat{\varepsilon})$-neighborhood of the optimal expected cost with the benefit of hindsight, i.e.,

$$\lim_{T \to \infty} \frac{1}{T} \mathop{\mathbb{E}}_{\boldsymbol{v_i}, \boldsymbol{d_i}} \left[C_i^K(\boldsymbol{v_i}, \boldsymbol{d_i}) - C_i^H(\boldsymbol{v_i}, \boldsymbol{d_i}) \right] = O(\hat{\varepsilon}).$$

The full statement of Theorem 4.1 with the required technical assumptions, as well as the detailed proof of the theorem, are included in the full version of the paper [9], and we provide here a sketch of the proof. The average expected optimal cost with the benefit of hindsight is lower-bounded by the maximum of the expected dual objective $\Psi_i^H(\mu_i^{\star H})$ derived from the Lagrangian dual problem in Equation (9). Meanwhile, the average stationary competition cost of strategy K is upper-bounded in terms of $\Psi_i^0(\mu_i^{\star 0})$ through a Taylor expansion in $\mu_i^{\star 0}$. Controlling this upper bound requires to control a) the hitting time such that the budget is depleted, or the multiplier leaves its bounds, only towards the end of the horizon T, as assumed in the definition of Ψ_i^0; and b) the expected distance of the multiplier iterates to the optimal multiplier $\mu_{i,t} - \mu_i^{\star 0}$ such that $\mu_{i,t}$ converges asymptotically to $\mu_i^{\star 0}$. These two objectives are achieved with a suitable choice of the gradient step size ϵ. Finally, we bound the difference of dual objectives $\Psi_i^0(\mu_i^{\star 0}) - \Psi_i^H(\mu_i^{\star H})$ in terms of the target expenditure rate ρ_i and the residual gain $\hat{\varepsilon}$.

This last step constitutes a fundamental difference to the standard setting with no budget gains. Indeed, Theorem 4.1 does not establish an asymptotic average regret of zero but rather in the order of the residual gain $\hat{\varepsilon}$. The intuition for the $O(\hat{\varepsilon})$ term is that, without the benefit of hindsight, agent i will always regret not setting the price to the a priori unknown maximum d_i^γ, instead of $d_i^{\gamma+1}$, when losing the auction. However, in cases where $d_i^\gamma - d_i^{\gamma+1}$ correspond to the distance between two adjacent independent samples from a common distribution $\mathcal{D}_{-i}$, the residual gain $\hat{\varepsilon}$ diminishes as the number of samples, or equivalently $N-1$, grows. For a large number of agents N, the residual gain $\hat{\varepsilon}$ is hence expected to be modest.

Another important difference to the standard setting with no budget gains is that the asymptotic guarantee of Theorem 4.1 requires the initial budget $k_{i,1}$ to grow sublinearly rather than linearly with respect to the time horizon T, c.f. Assumption 1. This effectively ensures that the target expenditure rate ρ_i tends to zero and that strategy K maximizes the correct expected dual objective Ψ_i^0. However, the initial budget $k_{i,1}$ cannot be kept constant. We show in the proof of Theorem 4.1 that the multiplier under strategy K is bounded at all time steps $t \in [T]$ by $\mu_{i,t} \leq \mu_{i,1} + \epsilon k_{i,1}$. Since, as in the standard setting with no budget gains, it is required that ϵ_i diminishes to zero asymptotically, $\mu_{i,t}$ would not be able to converge to any value greater than the initial value $\mu_{i,1}$ with a constant $k_{i,1}$. We refer to this phenomenon as the *vanishing box problem*.

Finally, the proof of Theorem 4.1 requires establishing the sublinear growth rate of $\mathbb{E}_{\boldsymbol{v_i}, \boldsymbol{d_i}} [T - \mathcal{T}_i]$ with respect to T, which is harder to obtain since the hitting time $\mathcal{T}_i$ defined in Equation (11) is a stricter notion than the budget depletion time used in the setting with no budget gains. This difficulty is addressed by Assumption 2. By introducing a strictly positive minimum valuation $\underline{v_i}$, Assumption 2.1 ensures that agent i always wins when the multiplier is close to $\underline{\mu}$, whereas the strictly positive minimum competing bid $\underline{d_i}$ introduced in Assumption 2.2 ensures that the agent always loses when the multiplier is close to $\overline{\mu}$. In practice, Assumption 2 implies that agents always have a need to participate in the auction. A numerical validation of Theorem 4.1 is included in Figure 2a.

4.2 Convergence under Simultaneous Learning

In this section, we take the next step towards our main goal of establishing that strategy K constitutes an approximate Nash equilibrium when adopted by all agents. Namely, we establish that the learning dynamics *converge in the simultaneous learning setting* in which all agents follow strategy K, denoted by joint strategy profile $\boldsymbol{K}$. The exact notion of convergence considered is presented in the main result of the section, Theorem 4.2.

Before stating this result, we first adapt our previous definitions to the multi-agent setting. Let $\boldsymbol{\mu_t} \in \mathbb{R}_+^N$ be the *multiplier profile* stacking the multipliers $\mu_{i,t}$ of all agents $i \in \mathcal{N}$. We extend the *expected dual objective*, *expected gain*, *expected expenditure* and the *expected loss* respectively as

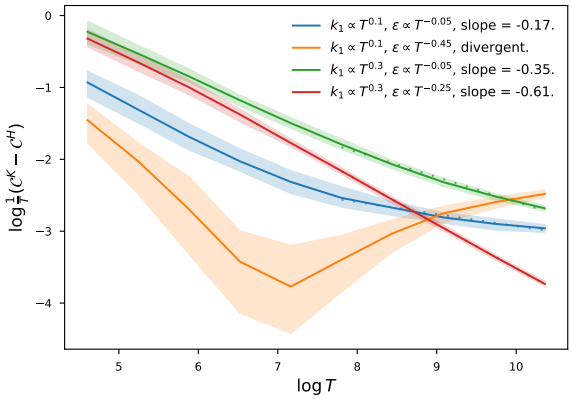
$$\Psi_i^0(\boldsymbol{\mu}) = \mathop{\mathbb{E}}_{\boldsymbol{v}} \left[v_i - \mu_i g_i - (\Delta v_i - \mu_i d_i^\gamma)^+ \right], \quad G_i(\boldsymbol{\mu}) = \mathop{\mathbb{E}}_{\boldsymbol{v}} [g_i],$$
$$Z_i(\boldsymbol{\mu}) = \mathop{\mathbb{E}}_{\boldsymbol{v}} \left[d_i^\gamma \mathbb{1}\{\Delta v_i > \mu_i d_i^\gamma\} \right], \quad L_i(\boldsymbol{\mu}) = Z_i(\boldsymbol{\mu}) - G_i(\boldsymbol{\mu}),$$

Compared to (10), the expectation is now with respect to the profile of valuations $\boldsymbol{v} \sim \boldsymbol{\mathcal{V}}$. Indeed, both the competing bid $d_i^\gamma = \gamma^{\text{th}}\text{-}\max_{j:j\neq i}\{\Delta v_j/\mu_j\}$ and the auction price $p^{\gamma+1} = \gamma{+}1^{\text{th}}\text{-}\max_j\{\Delta v_j/\mu_j\}$ are functions of $\boldsymbol{v}$ and $\boldsymbol{\mu}$. We will aim to show that $\boldsymbol{\mu_t}$ converges to a *stationary multiplier profile*, which is a multiplier profile $\boldsymbol{\mu^{\star 0}} \in \mathbb{R}_{>0}^N$ satisfying $L_i(\boldsymbol{\mu^{\star 0}}) = 0$ for all agents $i \in \mathcal{N}$. This multiplier profile is stationary in the sense that in expectation, update rule (K-μ) will yield $\mu_{i,t+1}^\star = \mu_{i,t}^\star$ for all agents i, since the expected expenditures $Z_i(\boldsymbol{\mu^{\star 0}})$ are equal to the expected gain $G_i(\boldsymbol{\mu^{\star 0}})$.

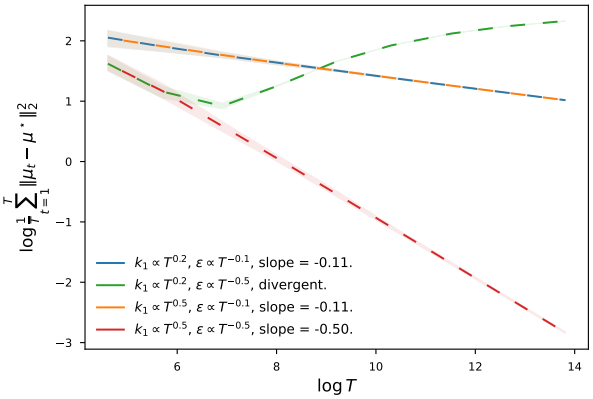
Notice that stationary multipliers $\boldsymbol{\mu^{\star 0}}$ are numerous: if multiplier $\boldsymbol{\mu^{\star 0}}$ is stationary, so is $\eta\boldsymbol{\mu^{\star 0}}$ for all $\eta > 0$, as the expenditures Z_i and gains G_i are equally scaled by $1/\eta$. This property of $\boldsymbol{\mu^{\star 0}}$ is novel to the karma setting, as with no budget gains a unique scale is fixed by the target expenditure rates ρ_i. To fix the unique $\boldsymbol{\mu^{\star 0}}$ that strategy profile $\boldsymbol{K}$ converges to, the projection is moved from the multiplier update (K-μ) to the bid (K-b) in strategy K, as compared to standard adaptive pacing. This modification, combined with a shared gradient step size ϵ for all agents, implies the following

$$\sum_{i\in\mathcal{N}} \mu_{i,t+1} = \sum_{i\in\mathcal{N}} \mu_{i,t} + \epsilon \sum_{i\in\mathcal{N}} (z_{i,t} - g_{i,t}) \overset{(a)}{=} \sum_{i\in\mathcal{N}} \mu_{i,t},$$
$$\text{hence } \boldsymbol{\mu_t} \in \boldsymbol{H_{\mu_1}} = \left\{ \boldsymbol{\mu} \in \mathbb{R}^N \,\middle|\, \sum_{i\in\mathcal{N}} (\mu_i - \mu_{i,1}) = 0 \right\}. \tag{12}$$

Property (12) anchors the scale of $\boldsymbol{\mu^{\star 0}}$ that is feasible under strategy profile $\boldsymbol{K}$ to the initial multiplier profile $\boldsymbol{\mu_1}$, and implies that the *average multiplier* $\mu_m = \sum_{i\in\mathcal{N}} \mu_{i,1}/N$ is preserved over time.



(a) Theorem 4.1: Stationary competition.



(b) Theorem 4.2: Simultaneous learning.

Figure 2: Numerical validation of Theorems 4.1 and 4.2. Figure 2a shows the convergence of costs to the minimum with the benefit of hindsight, while Figure 2b shows the convergence of multipliers under simultaneous learning. Both figures are based on 100 simulations per parameter combination, with the mean and the 95% confidence interval shown.

Notice that Property (12) would not hold if the projection was in the multiplier update (K-μ): intuitively, a projection there would cause agent i to 'forget' part of the history of expenses and gains, and affect the convergence of the whole population as a consequence by shifting the hyperplane of feasible profiles.

In the standard setting with no budget gains, it is common to assume strong monotonicity of the expected expenditure $\boldsymbol{Z}$, c.f. [5] which shows that it is implied by a *diagonal strict concavity* condition [35], and that it always holds in symmetric settings. In our karma setting with budget gains, the natural extension is to assume strong monotonicity of the expected loss $\boldsymbol{L}$ which effectively replaces the expected expenditure $\boldsymbol{Z}$. This adaptation is however not straightforward, as the multiple zeros of $\boldsymbol{L}$ on $\boldsymbol{U} := \prod_{i \in \mathcal{N}} \left(\underline{\mu}, \overline{\mu}\right)$ would immediately break the property. Instead, we use Property (12) and restrict our monotonicity requirement to multipliers lying in the hyperplane $\boldsymbol{H}_{\boldsymbol{\mu}_1}$.

Assumption 3 (Monotonicity). *The expected loss $\boldsymbol{L}$ is λ-strongly monotone over $\boldsymbol{U} \cap \boldsymbol{H}_{\boldsymbol{\mu}_1}$ with parameter $\lambda > 0$, i.e., for all $\boldsymbol{\mu}, \boldsymbol{\mu'} \in \boldsymbol{U} \cap \boldsymbol{H}_{\boldsymbol{\mu}_1}$, it holds that $(\boldsymbol{\mu} - \boldsymbol{\mu'})^\top (\boldsymbol{L}(\boldsymbol{\mu}) - \boldsymbol{L}(\boldsymbol{\mu'})) \leq -\lambda \|\boldsymbol{\mu} - \boldsymbol{\mu'}\|_2^2$.*

Assumption 3 ensures that the stationary multiplier profile $\boldsymbol{\mu}^{\star \mathbf{0}}$ is unique up to a multiplicative constant if it exists. With these preliminaries, we are ready to state the main result of this section regarding the asymptotic convergence of strategy profile $\boldsymbol{K}$, both with respect to the multiplier profile iterates $\boldsymbol{\mu}_t$ and the strategic competition costs $C_i^{\boldsymbol{K}}$ of all agents $i \in \mathcal{N}$ defined in Equation (3).

Theorem 4.2 (Convergence under Simultaneous Learning). *There exist constants C_1 and $C_2 \in \mathbb{R}_+$ such that the average expected distance to the stationary multiplier profile $\boldsymbol{\mu}^{\star \mathbf{0}} \in \boldsymbol{H}_{\boldsymbol{\mu}_1}$ and the strategic competition cost of strategy profile $\boldsymbol{K}$ for any agent $i \in \mathcal{N}$ satisfy respectively*

$$\frac{1}{T} \sum_{t=1}^{T} \mathop{\mathbb{E}}_{\boldsymbol{v}} \left[\left\| \boldsymbol{\mu}_t - \boldsymbol{\mu}^{\star \mathbf{0}} \right\|_2^2 \right] \leq C_1 N \left(\epsilon + \frac{1}{\epsilon T} + \frac{\mathbb{E}_{\boldsymbol{v}} \left[T - \underline{\mathcal{T}} \right]}{T} \right),$$

$$\frac{1}{T} C_i^{\boldsymbol{K}} - \Psi_i^0(\boldsymbol{\mu}^{\star \mathbf{0}}) \leq C_2 \left(N \left(\epsilon^{1/2} + \frac{1}{\epsilon T} \right) + \frac{\mathbb{E}_{\boldsymbol{v}} \left[T - \underline{\mathcal{T}} \right]}{T} \right).$$

Moreover, for suitably chosen parameters of strategy profile $\boldsymbol{K}$, the multiplier profile $\boldsymbol{\mu}_t$ converges in expectation to the stationary profile $\boldsymbol{\mu}^{\star \mathbf{0}}$, and the average strategic competition cost $C_i^{\boldsymbol{K}}$ converges to the expected dual objective $\Psi_i^0(\boldsymbol{\mu}^{\star \mathbf{0}})$ for all agents $i \in \mathcal{N}$, i.e., $\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{\boldsymbol{v}} \left[\left\| \boldsymbol{\mu}_t - \boldsymbol{\mu}^{\star \mathbf{0}} \right\|_2^2 \right] = \lim_{T \to \infty} \frac{1}{T} C_i^{\boldsymbol{K}} - \Psi_i^0(\boldsymbol{\mu}^{\star \mathbf{0}}) = 0.$

The full statement of Theorem 4.2 and the detailed proof are included in the full version of the paper [9], and mostly follow similar arguments as Theorem 4.1. The main step from the first to the second bound is to show that the profiles of expected dual objectives $\boldsymbol{\Psi}^{\mathbf{0}}$ and expected losses $\boldsymbol{L}$ are Lipschitz continuous in $\boldsymbol{\mu}$ on the compact set $\boldsymbol{U}$, which can be guaranteed by the absolute continuity of the valuations. As before, the main difficulty in comparison to the standard setting with no budget gains lies in ensuring that the expectation $\mathbb{E}_{\boldsymbol{v}} \left[T - \underline{\mathcal{T}} \right]$ grows sublinearly with respect to T. This challenge is addressed analogously as in Section 4.1: Assumption 2.1 deterministically guarantees that $\mathcal{T}_i^{\overline{\mu}} = T$ as any agent close to $\overline{\mu}$ will lose the auction and transition away from $\overline{\mu}$. A similar deterministic guarantee cannot be derived at the lower bound $\underline{\mu}$, however, due to the preservation of the average multiplier μ_m. For this reason, we impose a probabilistic condition on the lower bound $\underline{\mu}$, which is discussed further in Section 5. A numerical validation of Theorem 4.2 is included in Figure 2b.

4.3 Approximate Nash Equilibrium in Parallel Auctions

In this section, we finally combine the results of the previous two sections to achieve the main goal of establishing that the profile of adaptive karma pacing strategies $\boldsymbol{K}$ constitutes an approximate Nash equilibrium under suitable conditions.

Notice that one cannot immediately conclude that strategy profile $\boldsymbol{K}$ constitutes an approximate Nash equilibrium despite the previously established asymptotic guarantee on the strategic competition costs. Namely, Theorem 4.2 ensures that the multiplier profile converges asymptotically to $\boldsymbol{\mu}^{\star \mathbf{0}}$ under strategy profile $\boldsymbol{K}$.

Therefore, agent i's distribution of competing bids becomes stationary, and we showed in the proof of Theorem 4.1 that $\Psi_i^0(\boldsymbol{\mu}^{\star\mathbf{0}})$ lower bounds any *stationary competition cost*. However, agent i could potentially improve its cost by unilaterally deviating to a strategy $\beta_i \neq K$ that causes non-convergence of $\boldsymbol{\mu_t}$ and violation of the stationary competition assumption. For this reason, following [5], we consider an extension of our setting with multiple *parallel auctions* causing the effect of any single agent on the multipliers of others to become negligible as the number of agents grows.

Parallel Auctions. Let there be $M \geq 1$ auctions that are held in parallel at each time step $t \in [T]$. The number M could represent different priority roads, or the same road accessed at different times of the day, or a combination thereof. Each agent $i \in \mathcal{N}$ participates in one auction $m_{i,t} \in [M]$ per time step, where $m_{i,t}$ is drawn independently across agents and time from a fixed distribution $\boldsymbol{\pi_i} = (\pi_{i,m})_{m\in[M]}$; each $\pi_{i,m}$ denotes the probability for agent i to participate in auction m. We adapt the definition of competing bids accordingly as $d_{i,t}^{\gamma} = \gamma^{\text{th}}\text{-}\max_{j:j\neq i}\{\mathbb{1}\{m_{j,t} = m_{i,t}\}b_{j,t}\}$. Finally, we consider that the aggregate payment of all M auctions gets redistributed uniformly, leading to karma gains $g_{i,t} = \frac{\gamma}{N}\sum_{m\in[M]} p_{m,t}^{\gamma+1}$, where $p_{m,t}^{\gamma+1}$ is the price of auction m defined as $p_{i,t}^{\gamma+1} = \gamma + 1^{\text{th}}\text{-}\max_{i\in\mathcal{N}}\{\mathbb{1}\{m_{i,t} = m\}b_{i,t}\}$. This aggregate redistribution scheme is advantageous over redistributing the payment of each auction among its agents, since the aggregation restricts the influence of a single agent over the gains, and thereby the multipliers, of others. The distributions $(\boldsymbol{\pi_i})_{i\in\mathcal{N}}$ yield *matching probabilities* $\boldsymbol{a_i} = (a_{i,j})_{j\neq i}$, where $a_{i,j} = \mathbb{P}\{m_j = m_i\}$ denotes the probability that agent j is matched in the same auction as agent i. It is straightforward to show that the previous Theorems 4.1 and 4.2 also hold in the extended parallel auction setting.

Theorem 4.3 (Approximate Nash Equilibrium). *There exists a constant $C \in \mathbb{R}_+$ such that each agent $i \in \mathcal{N}$ can decrease its average strategic competition cost by deviating from strategy K to any strategy $\beta_i^T \in \mathcal{B}^T$ by at most*

$$\frac{1}{T}\left(C_i^K - C_i^{\beta_i^T,\boldsymbol{K_{-i}}}\right) \leq C\left(\left(\|\boldsymbol{a_i}\|_2 + \frac{M\gamma}{N}\right)\left(\sqrt{N\epsilon}\left(1 + \frac{1}{\epsilon^{3/2}T}\right) + \|\boldsymbol{a_i}\|_2 + \frac{\gamma}{\sqrt{N}}\right) + \left(\frac{\gamma}{N} + \frac{\overline{k_1}}{T}\right) + \left(\frac{\overline{k_1}}{T} + \frac{M\gamma}{N}\right)\frac{\mathbb{E}_{\boldsymbol{v},\boldsymbol{m}}\left[T - \underline{\mathcal{T}}\right]}{T}\right)$$

Moreover, for suitably chosen parameters, strategy profile $\boldsymbol{K}$ constitutes an approximate Nash equilibrium, i.e., it holds for all agents $i \in \mathcal{N}$ that $\lim_{T,N,M\to\infty} \frac{1}{T}\left(C_i^K - \inf_{\beta_i^T\in\mathcal{B}^T} C_i^{\beta_i^T,\boldsymbol{K_{-i}}}\right) = 0$.

The full statement of Theorem 4.3 with the detailed proof is included in the full version of the paper [9]. The proof involves lower-bounding the average expected cost under strategy β_i in terms of the optimal expected dual objective $\Psi_i^0(\boldsymbol{\mu}^{\star\mathbf{0}})$, and showing that asymptotically agent i cannot affect the multiplier profile of the other agents which converges to $\boldsymbol{\mu}^{\star\mathbf{0}}$. Competition hence becomes stationary, for which strategy K is optimal. While the proof follows a similar structure as in the standard setting with no budget gains, notice however that the bound in Theorem 4.3 is substantially different from its counterpart with no budget gains and requires adapting the technical assumptions in order to establish the asymptotic guarantee.

This achieves the main goal of our analysis. For the class of karma mechanisms with redistribution of payments, we have devised the simple adaptive karma pacing strategy K and provided conditions in which it constitutes an approximate Nash equilibrium.

5 DISCUSSION

In this paper, we devised a learning strategy, called *adaptive karma pacing*, that learns to bid optimally in karma mechanisms in which payments are redistributed in every time step. This simple strategy constitutes an approximate Nash equilibrium in large populations, and can hence be effectively employed to provide decision support, which is an important step toward the practical implementation of these mechanisms.

Discussion of assumptions. Our main results require a number of technical assumptions which, we argue, are not highly restrictive. These assumptions can be categorized as follows. The assumptions on *valuation and competing bid distributions*[3] are mild continuity and differentiability assumptions that are common in the literature, including in the standard monetary setting [5]. On the other hand, assumptions requiring to *vary parameters asymptotically*[4] are less natural to interpret in practice since typically the time horizon T and number of agents N are fixed by the setting. For this reason, we provided bounds in all our theorems that give finite time and population guarantees. Assumptions needed to *control the hitting time*[5] arise from our proof technique seeking to deterministically guarantee that the multipliers $\mu_{i,t}$ will never reach their bounds $\underline{\mu}$ and $\overline{\mu}$. In the full version of the paper [9], we perform numerical experiments verifying that the hitting time quickly approaches the end of the time horizon as assumed. Finally, the assumptions on *input parameters* of the adaptive karma pacing strategy[6] can be satisfied by design and/or tuning. Generally, setting the initial multiplier $\mu_{i,1}$ close to the center of a sufficiently low $\underline{\mu}$ and a sufficiently high $\overline{\mu}$, and using a sufficiently small gradient step size ϵ, suffices to satisfy these assumptions.

CREDIT AUTHOR STATEMENT

Damien Berriaud: Formal Analysis, Investigation, Writing – Original Draft, Software. **Ezzat Elokda**: Conceptualization, Methodology, Writing – Review & Editing, Visualization. **Devansh Jalota**: Conceptualization, Methodology, Writing – Review & Editing. **Emilio Frazzoli**: Supervision, Funding Acquisition. **Marco Pavone**: Supervision, Funding Acquisition. **Florian Dörfler**: Supervision, Funding Acquisition.

ACKNOWLEDGMENTS

Research supported by NCCR Automation, a National Centre of Competence in Research, funded by the Swiss National Science Foundation (Grant number 180545), and the Stanford Interdisciplinary Graduate Fellowship.

[3] The absolute continuity of valuations; and Assumption 14 in the full version [9].
[4] Assumptions 1 and 4; and 12 in the full version [9].
[5] Assumptions 2; and 9, 11, 13 in the full version [9].
[6] Assumptions 8, 10 in the full version [9].

REFERENCES

[1] Richard Arnott, André De Palma, and Robin Lindsey. 1990. Economics of a bottleneck. *Journal of urban economics* 27, 1 (1990), 111–130.

[2] Richard Arnott, André De Palma, and Robin Lindsey. 1994. The welfare effects of congestion tolls with heterogeneous commuters. *Journal of Transport Economics and Policy* 28 (1994), 139–161.

[3] Santiago R Balseiro, Omar Besbes, and Gabriel Y Weintraub. 2015. Repeated auctions with budgets in ad exchanges: Approximations and design. *Management Science* 61, 4 (2015), 864–884.

[4] Santiago R. Balseiro, Yuan Deng, Jieming Mao, Vahab S. Mirrokni, and Song Zuo. 2021. The Landscape of Auto-bidding Auctions: Value versus Utility Maximization. In *Proceedings of the 22nd ACM Conference on Economics and Computation* (Budapest, Hungary) *(EC '21)*. Association for Computing Machinery, New York, NY, USA, 132–133. https://doi.org/10.1145/3465456.3467607

[5] Santiago R Balseiro and Yonatan Gur. 2019. Learning in repeated auctions with budgets: Regret minimization and equilibrium. *Management Science* 65, 9 (2019), 3952–3968.

[6] Santiago R Balseiro, Huseyin Gurkan, and Peng Sun. 2019. Multiagent mechanism design without money. *Operations Research* 67, 5 (2019), 1417–1436.

[7] Santiago R Balseiro, Haihao Lu, and Vahab Mirrokni. 2023. The best of many worlds: Dual mirror descent for online allocation problems. *Operations Research* 71, 1 (2023), 101–119.

[8] Siddhartha Banerjee, Giannis Fikioris, and Eva Tardos. 2023. Robust Pseudo-Markets for Reusable Public Resources. In *Proceedings of the 24th ACM Conference on Economics and Computation* (London, United Kingdom) *(EC '23)*. Association for Computing Machinery, New York, NY, USA, 241. https://doi.org/10.1145/3580507.3597723

[9] Damien Berriaud, Ezzat Elokda, Devansh Jalota, Emilio Frazzoli, Marco Pavone, and Florian Dörfler. 2025. To Spend or to Gain: Online Learning in Repeated Karma Auctions. arXiv:2403.04057 [cs.GT] https://arxiv.org/abs/2403.04057

[10] Devi K Brands, Erik T Verhoef, Jasper Knockaert, and Paul R Koster. 2020. Tradable permits to manage urban mobility: market design and experimental implementation. *Transportation Research Part A: Policy and Practice* 137 (2020), 34–46.

[11] Eric Budish and Estelle Cantillon. 2012. The multi-unit assignment problem: Theory and evidence from course allocation at Harvard. *American Economic Review* 102, 5 (2012), 2237–71.

[12] Matteo Castiglioni, Andrea Celli, and Christian Kroer. 2022. Online Learning with Knapsacks: the Best of Both Worlds. In *Proceedings of the 39th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 162)*, Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (Eds.). PMLR, 2767–2783. https://proceedings.mlr.press/v162/castiglioni22a.html

[13] Ezzat Elokda, Saverio Bolognani, Andrea Censi, Florian Dörfler, and Emilio Frazzoli. 2023. A self-contained karma economy for the dynamic allocation of common resources. *Dynamic Games and Applications* (2023), 1–33.

[14] Ezzat Elokda, Carlo Cenedese, Kenan Zhang, Andrea Censi, John Lygeros, Emilio Frazzoli, and Florian Dörfler. 2024. CARMA: Fair and Efficient Bottleneck Congestion Management via Nontradable Karma Credits. *Transportation Science* (09 2024). https://doi.org/10.1287/trsc.2023.0323

[15] Andrew W Evans. 1992. Road congestion pricing: when is it a good policy? *Journal of transport economics and policy* (1992), 213–243.

[16] Huang Fang, Nick Harvey, Victor Portella, and Michael Friedlander. 2020. Online mirror descent and dual averaging: keeping pace in the dynamic case. In *Proceedings of the 37th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 119)*, Hal Daumé III and Aarti Singh (Eds.). PMLR, 3008–3017. https://proceedings.mlr.press/v119/fang20a.html

[17] Eric J. Friedman, Joseph Y. Halpern, and Ian Kash. 2006. Efficiency and nash equilibria in a scrip system for P2P networks. In *Proceedings of the 7th ACM Conference on Electronic Commerce* (Ann Arbor, Michigan, USA) *(EC '06)*. Association for Computing Machinery, New York, NY, USA, 140–149. https://doi.org/10.1145/1134707.1134723

[18] Jason Gaitonde, Yingkai Li, Bar Light, Brendan Lucier, and Aleksandrs Slivkins. 2023. Budget Pacing in Repeated Auctions: Regret and Efficiency Without Convergence. In *14th Innovations in Theoretical Computer Science Conference (ITCS 2023) (Leibniz International Proceedings in Informatics (LIPIcs), Vol. 251)*, Yael Tauman Kalai (Ed.). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 52:1–52:1. https://doi.org/10.4230/LIPIcs.ITCS.2023.52

[19] Yuan Gao, Alex Peysakhovich, and Christian Kroer. 2021. Online market equilibrium with application to fair division. *Advances in Neural Information Processing Systems* 34 (2021), 27305–27318.

[20] Allan Gibbard. 1973. Manipulation of voting schemes: a general result. *Econometrica* 41, 4 (1973), 587–601.

[21] Artur Gorokh, Siddhartha Banerjee, and Krishnamurthy Iyer. 2021. From monetary to nonmonetary mechanism design via artificial currencies. *Mathematics of Operations Research* 46, 3 (2021), 835–855.

[22] Artur Gorokh, Siddhartha Banerjee, and Krishnamurthy Iyer. 2021. The Remarkable Robustness of the Repeated Fisher Market. In *Proceedings of the 22nd ACM Conference on Economics and Computation* (Budapest, Hungary) *(EC '21)*. Association for Computing Machinery, New York, NY, USA, 562. https://doi.org/10.1145/3465456.3467560

[23] Mingyu Guo, Vincent Conitzer, and Daniel M. Reeves. 2009. Competitive Repeated Allocation without Payments. In *Proceedings of the 5th International Workshop on Internet and Network Economics* (Rome, Italy) *(WINE '09)*. Springer-Verlag, Berlin, Heidelberg, 244–255. https://doi.org/10.1007/978-3-642-10841-9_23

[24] Yingni Guo and Johannes Hörner. 2020. *Dynamic Allocation without Money*. TSE Working Papers 20-1133. Toulouse School of Economics (TSE). https://ideas.repec.org/p/tse/wpaper/124604.html

[25] Elad Hazan. 2023. Introduction to Online Convex Optimization. arXiv:1909.05207 [cs.LG]

[26] Devansh Jalota, Karthik Gopalakrishnan, Navid Azizan, Ramesh Johari, and Marco Pavone. 2023. Online Learning for Traffic Routing under Unknown Preferences. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics (Proceedings of Machine Learning Research, Vol. 206)*, Francisco Ruiz, Jennifer Dy, and Jan-Willem van de Meent (Eds.). PMLR, 3210–3229. https://proceedings.mlr.press/v206/jalota23a.html

[27] Devansh Jalota and Yinyu Ye. 2023. Stochastic Online Fisher Markets: Static Pricing Limits and Adaptive Enhancements. arXiv:2205.00825 [cs.GT]

[28] Kris Johnson, David Simchi-Levi, and Peng Sun. 2014. Analyzing scrip systems. *Operations Research* 62, 3 (2014), 524–534.

[29] Jaehong Kim, Mengling Li, and Menghan Xu. 2021. Organ donation with vouchers. *Journal of Economic Theory* 191 (2021), 105–159.

[30] V. Krishna. 2009. *Auction Theory*. Academic Press, Burlington MA.

[31] Brendan Lucier, Sarath Pattathil, Aleksandrs Slivkins, and Mengxiao Zhang. 2024. Autobidders with Budget and ROI Constraints: Efficiency, Regret, and Pacing Dynamics. In *Proceedings of Thirty Seventh Conference on Learning Theory (Proceedings of Machine Learning Research, Vol. 247)*, Shipra Agrawal and Aaron Roth (Eds.). PMLR, 3642–3643. https://proceedings.mlr.press/v247/lucier24a.html

[32] Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay V Vazirani. 2007. *Algorithmic Game Theory*. Cambridge University Press, Cambridge.

[33] Canice Prendergast. 2022. The allocation of food to food banks. *Journal of Political Economy* 130, 8 (2022), 1993–2017.

[34] Kevin WS Roberts. 1980. Interpersonal comparability and social choice theory. *The Review of Economic Studies* 47, 2 (1980), 421–439.

[35] J. B. Rosen. 1965. Existence and Uniqueness of Equilibrium Points for Concave N-Person Games. *Econometrica* 33, 3 (1965), 520–534. https://onlinelibrary.wiley.com/doi/abs/0012-9682(196507)33:3<520:EAUOEP>2.0.CO;2-2

[36] Alvin E Roth and Marilda A Oliveira Sotomayor. 1990. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Number 18 in Econometric Society Monographs. Cambridge University Press.

[37] Mark Allen Satterthwaite. 1975. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory* 10, 2 (1975), 187–217.

[38] James Schummer and Rakesh V. Vohra. 2007. Mechanism Design without Money. In *Algorithmic Game Theory*, Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V.Editors Vazirani (Eds.). Cambridge University Press, 243–266.

[39] Tayfun Sönmez, M Utku Ünver, and M Bumin Yenmez. 2020. Incentivized kidney exchange. *American Economic Review* 110, 7 (2020), 2198–2224.

[40] Brian D Taylor and Rebecca Kalauskas. 2010. Addressing equity in political debates over road pricing: Lessons from recent projects. *Transportation research record* 2187, 1 (2010), 44–52.

[41] William S Vickrey. 1969. Congestion theory and transport investment. *The American Economic Review* 59, 2 (1969), 251–260.

[42] Vivek Vishnumurthy, Sangeeth Chandrakumar, and Emin Gun Sirer. 2003. Karma: A secure economic framework for peer-to-peer resource sharing. In *Workshop on Economics of Peer-to-peer Systems*.

[43] Midhul Vuppalapati, Giannis Fikioris, Rachit Agarwal, Asaf Cidon, Anurag Khandelwal, and Éva Tardos. 2023. Karma: Resource Allocation for Dynamic Demands. In *17th USENIX Symposium on Operating Systems Design and Implementation (OSDI 23)*.

[44] Qian Wang, Zongjun Yang, Xiaotie Deng, and Yuqing Kong. 2023. Learning to bid in repeated first-price auctions with budgets. In *International Conference on Machine Learning*. PMLR, 36494–36513.