Social Coordination in Unknown Price-Sensitive Populations

Philip N. Brown[1] and Jason R. Marden[2]

Abstract— **In this paper, we investigate the relationship between uncertainty and a designer's ability to influence social behavior. Pigovian taxes are a common approach to social coordination. However, guaranteeing efficient behavior typically requires that the system designer has complete knowledge of the user population's sensitivity to taxation. In this paper, we explore the effect of relaxing this requirement in the context of congestion games with affine costs. Focusing on the class of scaled Pigovian taxes, we derive the optimal tolling scheme that minimizes the worst-case efficiency loss under uncertainty in user sensitivity. Furthermore, we derive explicit bounds which highlight how the level of uncertainty in sensitivity degrades performance.**

I. INTRODUCTION

As engineered systems and their users become increasingly interconnected, engineers must consider not only mere system design, but also efficient system utilization [1]. It is widely known that uninfluenced social systems often exhibit suboptimal behavior. This is a well-researched topic in many fields, including electric power markets [2], communication spectrum sharing [3], and traffic congestion [4]. Accordingly, a central design challenge is to influence social behavior through admissible mechanisms in these systems to promote efficient outcomes.

To analyze the role of social coordination in engineered systems, we model the behavior of a system's users as a game between the users, in which each user chooses his actions to maximize his personal benefit. Equilibrium behavior established by every user acting in this way is known as a Nash equilibrium. It is widely known that Nash equilibria need not be efficient in any global sense. It is common to quantify the efficiency losses inherent in Nash equilibria by a metric known as the "price of anarchy" of a system, defined as the worst-possible ratio between the efficiency of a Nash equilibrium behavior and that of an optimal behavior [5], [6]. The price of anarchy has been analyzed for a variety of engineered systems [7], [8], [9]. This leads to the central question: how may social behavior be influenced to reduce the price of anarchy?

A common methodology for social coordination is known as a Pigovian tax [10]. Under a Pigovian tax, users are charged a tax equal to the negative effect they have on other users; thus, each user is incentivized to choose actions that maximize the public benefit. In a general sense, Pigovian taxes are attractive since they guarantee that Nash equilibrium behavior exactly equals system optimal behavior. This type of taxation strategy has been suggested as a solution to industrial pollution [11], harmful speculative trading in financial markets [12], and road congestion [13], to name a few.

To illustrate a social coordination setting in which Pigovian taxes apply, we turn to the problem of road congestion. It is well-known that Nash equilibria in routing games are often inefficient [14], and a central research focus has been to influence routing behavior using tolls to promote efficient routing. In the congestion setting, a Pigovian tax is known as a marginal-cost toll or congestion toll; each driver is charged a toll equivalent to the congestion caused by an additional unit of traffic [15], [16], [17]. An inherent challenge is that congestion is experienced in units of time, but tolls are charged in units of money. If the designer lacks perfect information regarding drivers' price-sensitivity, it will not in general be possible for the designer to levy a perfect toll.

Recent research has sidestepped this issue by fixing all system variables and deriving a tolling strategy that optimizes a fixed instance of the system [18]. It has been shown that under a general class of congestion games, this type of fixed tolling can optimize Nash Equilibria [19].

While these results represent successes, they rely strongly on a complete characterization of the network structure, user capabilities, and price sensitivities. Accordingly, it may be that small variations in any of these parameters could lead to highly suboptimal behavior; uncertainty in any of these parameters may dramatically reduce a designer's ability to influence social behavior and could in general lead to a degradation in efficiency from the uninfluenced case. Little research has been done to characterize the effect of this uncertainty on the ability to coordinate a social system.

With these limitations in mind, we initiate a study on the relationship between information and a designer's ability to influence social behavior. Specifically, we investigate uncertain user price-sensitivity for a routing problem in a class of congestion games: the set of network routing games such that each edge in a network possesses a congestion function that is affine in the number of edge users. Congestion games of this type have been extensively studied in the existing literature [20], [21], [14].

For this class of congestion games, we develop a distributed tolling scheme that increases routing efficiency for any bounded uncertainty of the user population's price-sensitivity. We show that under certain common utilization conditions, provided that the population's price-sensitivity is

This research was supported by AFOSR grants #FA9550-09-1-0538 and #FA9550-12-1-0359 and by ONR grant #N00014-09-1-0751.

P. N. Brown is a graduate research assistant with the Department of Electrical, Computer, and Energy Engineering, University of Colorado, Boulder, CO 80309, philip.brown@colorado.edu.

J. R. Marden is with the Department of Electrical, Computer, and Energy Engineering, University of Colorado, Boulder, CO 80309, jason.marden@colorado.edu. Corresponding author.

range-restricted, scaled marginal-cost tolls always increase efficiency. Furthermore, we show that the unique marginal-cost toll scale factor that minimizes the price of anarchy is the geometric mean of the upper and lower bounds of the population's price sensitivity. That is, our toll on a road is a function *only* of that road's congestion properties and the ratio of the population's sensitivity bounds – it is independent of the traffic flow rate and the network topology.

We also provide a tight bound on the price of anarchy for games using our specified tolling methodology. This price of anarchy depends only on the ratio of the user population's sensitivity bounds; it is independent of network cost functions, network topology, and overall traffic rate. Thus, for a broad class of games, we concisely characterize the relationship between information and efficiency.

II. MODEL

A. Network Definition

Consider a network (V, E), which consists of a vertex set V and edge set $E \subseteq (V \times V)$. Each edge $e \in E$ is associated with a latency function of the affine form

$$l_e(f_e) = d_e f_e + c_e, \tag{1}$$

where f_e designates the flow on edge e.

In this paper, we will focus on characterizing the efficiency of flows induced by individuals' strategic behavior. Specifically, we focus on the setting in which there is a population of users seeking to traverse the network from a common source to a common destination. Consequently, each user chooses its path from a common set of paths $\mathcal{P} = \{p_1, p_2, \ldots, p_n\}$ where each path $p \in \mathcal{P}$ consists of a set of edges, i.e., $\mathcal{P} \subseteq 2^E$. We define the total mass of the population to be $r \geq 0$. Using these definitions, let a *game* G be defined by the tuple $G = (V, E, \mathcal{P}, \{l_e\}_{e\in E}, r)$, and let $\mathcal{G}$ represent the class of all such games G.

B. Network Flows

For a game G, the set of feasible flows $f(G) = (f_{p_1}, f_{p_2}, \ldots, f_{p_n})$ is characterized by

$$F(G) = \left\{ \{f_p\}_{p\in\mathcal{P}} : f_p \geq 0 \text{ and } \sum_{p\in\mathcal{P}} f_p = r \right\},$$

where f_p designates the flow on path p. For a given flow $f(G) \in F(G)$, the latency $l_p(f(G))$ of any path $p \in \mathcal{P}$ is the sum of the latencies of edges in that path:

$$l_p(f(G)) = \sum_{e\in p} l_e(f_e), \tag{2}$$

where f_e is the flow on edge e given $f(G)$, i.e.,

$$f_e = \sum_{p\in\mathcal{P}:e\in p} f_p. \tag{3}$$

Thus, the total latency on the network is the flow-weighted sum of all path latencies:

$$L(f(G)) = \sum_{p\in\mathcal{P}} f_p \cdot l_p(f_p). \tag{4}$$

Users are concerned with minimizing their latencies, so the cost J_p to a user of any path $p \in \mathcal{P}$ is the path latency:

$$J_p(f(G)) = \sum_{e\in p} l_e(f_e). \tag{5}$$

We focus in particular on the setting in which each user chooses the path with the lowest cost. A routing obtained in this way is known as a Nash flow [14], which corresponds to a feasible flow $f^{\rm ne}(G) \in F(G)$ such that for all $p_i, p_j \in \mathcal{P}$

$$f^{\rm ne}_{p_i} > 0, f^{\rm ne}_{p_j} > 0 \implies J_{p_i}(f^{\rm ne}(G)) = J_{p_j}(f^{\rm ne}(G)), \tag{6}$$

$$f^{\rm ne}_{p_i} > 0, f^{\rm ne}_{p_j} = 0 \implies J_{p_i}(f^{\rm ne}(G)) \leq J_{p_j}(f^{\rm ne}(G)). \tag{7}$$

We define an *optimal* flow on G as

$$f^{\rm opt}(G) = \inf_{f(G)\in F(G)} L(f(G)).$$

Then we define the price of anarchy of $\mathcal{G}$ as

$${\rm PoA}(\mathcal{G}) = \sup_{G\in\mathcal{G}} \frac{L(f^{\rm ne}(G))}{L(f^{\rm opt}(G))}. \tag{8}$$

C. Tolling for Known Homogeneous Populations

The network owner, wishing to enforce optimal routing, charges a toll $\tau_e(f_e)$ on each link in the network, adding to the link user's cost. A Pigovian toll, also known as a marginal-cost toll, is of the form $\tau_e(f_e) = d_e f_e$. In general, these marginal-cost tolls induce optimal routing. Homogeneity of users is a common assumption when analyzing worst-case behavior [21]; it is a simple approach that can often be used to bound the behavior of heterogeneous populations. The clear benefit of tolling functions of the Pigovian form is that they can be assigned in a distributed manner; each edge toll is a function only of the congestion properties of that edge. For latency functions of the form in (1), Pigovian tolls modify (5) to give a path cost of

$$J_p(f(G)) = \sum_{e\in p} \left[2d_e(f_e) + c_e\right]. \tag{9}$$

In [14], it is shown that cost functions of this form always induce optimal behavior.

D. Tolling for Unknown Homogeneous Populations

In real systems, it is often true that users exhibit price-sensitivities that are unknown or that vary from day to day. We model this price sensitivity with a parameter $\beta \in [\beta_{\rm L}, \beta_{\rm U}]$. Intuitively, high β implies high aversion to toll. Incorporating price sensitivity, we represent the cost of a path generally as

$$J_p(f(G)) = \sum_{e\in p} \left[l_e(f_e) + \beta\tau_e(f_e)\right]. \tag{10}$$

In the case when β is known precisely, an optimal marginal-cost toll is of the form $\tau_e(f_e) = d_e f_e/\beta$; with this toll, (10) clearly simplifies to (9). We desire to exploit the distributed nature of marginal-cost tolls, so we investigate tolling functions of the form $\tau_e(f_e) = \mu d_e f_e$, where $\mu \geq 0$ is a common factor for every edge, and scales with our estimate of the true β of the population. We call tolls of this form *scaled marginal-cost tolls*. These tolls further modify (5) to give the final form of our path cost function

$$J_p(f(G)) = \sum_{e\in p} \left[(1+\beta\mu)d_e f_e + c_e\right]. \tag{11}$$

These modified cost functions induce a new Nash flow which we denote $f^{\mathrm{ne}}(\mu; G, \beta)$. For simplicity of notation, we define the total latency of this Nash flow under scaled marginal-cost tolls as

$$L^{\mathrm{ne}}(\mu; G, \beta) = L(f^{\mathrm{ne}}(\mu; G, \beta)), \tag{12}$$

Where $L(f(\cdot))$ is defined as in (4). Note that μ is our control parameter; G and β represent the system and population we are attempting to control.

Therefore, given a game G, we seek to characterize the scaled marginal-cost tolls which minimize the worst-case total latency for G given a range of user price sensitivities. That is, we seek to characterize solutions μ^* to the optimization problem

$$L^{\mathrm{ne}}(\mu^*; G, \beta) = \inf_{\mu\geq 0}\left(\sup_{\beta\in[\beta_{\mathrm{L}},\beta_{\mathrm{U}}]} \left(L^{\mathrm{ne}}(\mu; G, \beta)\right)\right). \tag{13}$$

III. MAIN RESULT

We show that under reasonable utilization constraints, the solution to the optimization problem defined in (13) is independent of the underlying game G.

Definition 3.1: A game $G = (V, E, \mathcal{P}, \{l_e\}_{e\in E}, r)$ is *fully-utilized* if $\forall \bar{r} \geq r$, $\forall p_i \in \mathcal{P}$, $f^{\mathrm{ne}}_{p_i}(\mu = 0; G, \beta) > 0$. That is, all paths are in use in an un-tolled Nash flow, and increasing r will not cause any path flow to decrease to zero. In Section V-A we show that this definition applies to a broad class of games.

Theorem 3.2: Consider any game G that satisfies Definition 3.1. Let

$$\mu^* = \frac{1}{\sqrt{\beta_{\mathrm{L}}\beta_{\mathrm{U}}}}. \tag{14}$$

Then μ^* satisfies

$$L^{\mathrm{ne}}(\mu^*; G, \beta) = \inf_{\mu\geq 0}\left(\sup_{\beta\in[\beta_{\mathrm{L}},\beta_{\mathrm{U}}]} \left(L^{\mathrm{ne}}(\mu; G, \beta)\right)\right).$$

Furthermore, let $\mathcal{G}^*$ represent the class of all games G satisfying Definition 3.1. For $\beta \in [\beta_{\mathrm{L}}, \beta_{\mathrm{U}}]$,

$$\mathrm{PoA}\left(\mu^*; \mathcal{G}^*, \beta\right) = \frac{4\left(1+\sqrt{\beta_{\mathrm{L}}/\beta_{\mathrm{U}}}+\beta_{\mathrm{L}}/\beta_{\mathrm{U}}\right)}{3\left(1+\sqrt{\beta_{\mathrm{L}}/\beta_{\mathrm{U}}}\right)^2}. \tag{15}$$

Thus, μ^* ensures that the worst-case total latency of a game is limited to a minimum for any range-restricted β.

Note that when $\beta_{\mathrm{L}} = 0$, meaning that it is not possible to influence behavior with tolls, we obtain the well-known price of anarchy of $4/3$ derived in [21]. On the other hand, if $\beta_{\mathrm{L}} = \beta_{\mathrm{U}}$, meaning that the price-sensitivity of all users is known precisely, we obtain a price of anarchy of 1, implying that complete information allows complete control.

Note also that μ^* depends only on β_{L} and β_{U}; it is independent of G. This allows us to define a distributed rule for our edge tolls so that we can set each edge's toll without knowledge of the overall network topology, and guarantees a reduction in total latency from the un-tolled network.

IV. PROOF OF RESULTS

The proof proceeds as follows:

1) In Lemma 4.1 we show that in games satisfying Definition 3.1, the flow on any path can be decomposed as the sum of a term depending on r and a term depending on μ.
2) In Lemma 4.2 we show that for such games, Lemma 4.1 implies that scaled marginal-cost tolls always decrease latency by a quantity that is independent of the traffic rate r.
3) Lastly, we show that the properties proved in Lemma 4.2 lead to a characterization of the unique marginal-cost toll scale factor that minimizes worst-case total latency for uncertain sensitivity. We then derive the price of anarchy under our optimal marginal-cost tolls.

For simplicity, we will henceforth express $f^{\mathrm{ne}}_p(\mu; G, \beta)$ as merely f^{ne}_p when the dependence on μ, G, and β is clear.

Our proof utilizes the following definition: for any $p_i, p_j \in \mathcal{P}$,

$$d_{ij} = \begin{cases} \sum_{e\in p_i\cap p_j} d_e, & p_i \cap p_j \neq \emptyset \\ 0, & p_i \cap p_j = \emptyset \end{cases}. \tag{16}$$

Intuitively, d_{ij} is the amount that $f^{\mathrm{ne}}_{p_i}$ affects $J_{p_j}(f^{\mathrm{ne}})$. Note that $d_{ij} = d_{ji}$.

Lemma 4.1: For any game $G = (V, E, \mathcal{P}, \{l_e\}_{e\in E}, r)$ satisfying Definition 3.1, for any $\mu \geq 0$, the Nash flow on each path $p_i \in \mathcal{P}$ is of the form

$$f^{\mathrm{ne}}_{p_i} = rR_i + \frac{1}{1+\mu\beta}M_i, \tag{17}$$

where the coefficients $R_i \in \mathbb{R}$ and $M_i \in \mathbb{R}$ have no dependence on μ, r, or β, and satisfy the following equations:

$$\sum_{p_i\in\mathcal{P}} R_i = 1, \tag{18}$$

$$\sum_{p_i\in\mathcal{P}} M_i = 0, \tag{19}$$

$$\sum_{p_k\in\mathcal{P}} R_k d_{ik} = \sum_{p_k\in\mathcal{P}} R_k d_{jk}, \tag{20}$$

$$\sum_{p_k\in\mathcal{P}} M_k d_{ik} + \sum_{e\in p_i} c_e = \sum_{p_k\in\mathcal{P}} M_k d_{jk} + \sum_{e\in p_j} c_e, \tag{21}$$

where $p_i, p_j, p_k \in \mathcal{P}$.

Proof: First, we prove the decomposition of the path flows. Then we show that due to the properties induced by Definition 3.1, the coefficients R_i and M_i satisfy (18)-(21).

Combining (11) and (3), the cost of path p_i is

$$J_{p_i}(f^{\mathrm{ne}}) = \sum_{e\in p_i}\left((1+\beta\mu)d_e\left(\sum_{p_k: e\in p_k} f^{\mathrm{ne}}_{p_k}\right) + c_e\right). \tag{22}$$

Define $\bar{d}_e(p)$ in the following way:

$$\bar{d}_e(p) = \begin{cases} d_e, & e \in p \\ 0, & e \notin p. \end{cases}$$

Now, (22) can be written as

$$\begin{aligned} J_{p_i}(f^{\text{ne}}) &= (1+\beta\mu) \sum_{e \in p_i} \sum_{p_k \in \mathcal{P}} \bar{d}_e(p_i) f^{\text{ne}}_{p_k} + \sum_{p_i \in \mathcal{P}} c_e \\ &= (1+\beta\mu) \sum_{p_k \in \mathcal{P}} f^{\text{ne}}_{p_k} \sum_{e \in p_i} \bar{d}_e(p_i) + \sum_{p_i \in \mathcal{P}} c_e \\ &= (1+\beta\mu) \sum_{p_k \in \mathcal{P}} f^{\text{ne}}_{p_k} d_{ik} + \sum_{e \in p_i} c_e. \end{aligned} \tag{23}$$

Since f^{ne} is a Nash flow and every path flow is positive, by (6) all path costs are equal, so $\forall p_i, p_j \in \mathcal{P}$,

$$\sum_{p_k \in \mathcal{P}} f^{\text{ne}}_{p_k} (d_{ik} - d_{jk}) = \frac{1}{1+\beta\mu} \left[\sum_{e \in p_j} c_e - \sum_{e \in p_i} c_e \right]. \tag{24}$$

Thus, each $\{i, j\}$ defines an equation of the form $\sum_{p_k \in \mathcal{P}} f^{\text{ne}}_{p_k} a_{ij} = b_{ij}$. We take any distinct[1] $|\mathcal{P}| - 1$ of these equations and combine them with $\sum_{p_k \in \mathcal{P}} f^{\text{ne}}_{p_k} = r$ to obtain a $|\mathcal{P}|$-dimensional system of linear equations, which can be described by the following matrix equation:

$$A f^{\text{ne}} = b.^2 \tag{25}$$

A solution to (25) represents a Nash flow for (G, β, μ), and in [22] it is shown Nash flows always exist for congestion games of the type considered in this paper. Since b is of the form

$$b = \begin{bmatrix} b_1/(1+\beta\mu) \\ \vdots \\ b_{|\mathcal{P}|-1}/(1+\beta\mu) \\ r \end{bmatrix}, \tag{26}$$

Any solution f^{ne} must be a linear combination of $1/(1+\beta\mu)$ and r.[3] That is, $\forall p_i \in \mathcal{P}$, $f^{\text{ne}}_{p_i}$ can be represented as shown in (17). Note that $\{R_i\}$ and $\{M_i\}$ have no dependence on μ, r, or β.

Clearly, since G satisfies Definition 3.1, R_i must be nonnegative for every path $p_i \in \mathcal{P}$. Note also that $\forall \mu > 0$, $f^{\text{ne}}_{p_i}(\mu; G, \beta) > 0$ since $f^{\text{ne}}_{p_i}(0; G, \beta) > 0$. That is, (17) is robust to increases in tolls.

Next, we derive two important properties of R_i and M_i. Substituting (17) into $\sum_{p_i \in \mathcal{P}} f^{\text{ne}}_{p_i} = r$, we obtain

$$r \left(\sum_{p_i \in \mathcal{P}} R_i - 1 \right) + \frac{1}{1+\mu\beta} \left(\sum_{p_i \in \mathcal{P}} M_i \right) = 0. \tag{27}$$

[1] Note that $\{i, j\}$ is *not* distinct from $\{j, i\}$.

[2] With abuse of notation, here we represent the tuple of path flows as a column vector.

[3] To see this, perform Gaussian elimination on the augmented matrix $[A\ b]$ to solve for f. Note that matrix row operations only involve sums of elements of b, never products.

Since (27) must hold for any arbitrarily large r and for all $\mu \geq 0$, we obtain the proof of (18) and (19):

$$\sum_{p_i \in \mathcal{P}} R_i = 1,$$
$$\sum_{p_i \in \mathcal{P}} M_i = 0.$$

Now, substituting (17) into (24) and simplifying, we obtain

$$r \sum_{p_k \in \mathcal{P}} R_k (d_{ik} - d_{jk}) + \frac{1}{1+\mu\beta} \left[\sum_{p_k \in \mathcal{P}} M_k (d_{ik} - d_{jk}) + \sum_{e \in p_j} c_e - \sum_{e \in p_i} c_e \right] = 0. \tag{28}$$

Similarly to above, since (28) must hold for any arbitrarily large r and for all $\mu \geq 0$, we obtain the proof of (20) and (21):

$$\sum_{p_k \in \mathcal{P}} R_k d_{ik} = \sum_{p_k \in \mathcal{P}} R_k d_{ij},$$
$$\sum_{p_k \in \mathcal{P}} M_k d_{ik} + \sum_{e \in p_i} c_e = \sum_{p_k \in \mathcal{P}} M_k d_{jk} + \sum_{e \in p_j} c_e,$$

Which completes the proof. ■

Lemma 4.2: For any game $G = (V, E, \mathcal{P}, \{l_e\}_{e \in E}, r)$ satisfying Definition 3.1, there exists a function $L_r(r)$ depending only on r, and a constant $K_\mu \geq 0$, such that

$$L^{\text{ne}}(\mu; G, \beta) = L_r(r) - \frac{\mu\beta}{(1+\mu\beta)^2} K_\mu. \tag{29}$$

Here, we show for total latency a similar quality to that expressed for flows in Lemma 4.1; namely, rate and tolls have independent effects on the total latency. This is a crucially-important property of games satisfying Definition 3.1, since it allows us to define tolling rules without knowledge of r.

Proof: First, we show the decomposition of the total latency of a Nash flow. We then show that the total latency of every Nash flow is of the specific form shown in (29).

Substitute (17) into (4) and simplify to obtain

$$L^{\text{ne}}(\mu; G, \beta) = L_r(r) + \frac{r}{1+\mu\beta} \zeta + \eta, \tag{30}$$

where

$$L_r(r) = \sum_{p_i \in \mathcal{P}} \left[r^2 \sum_{p_j \in \mathcal{P}} d_{ij} R_i R_j + r \sum_{e \in p_i} c_e R_i \right], \tag{31}$$

$$\begin{aligned} \zeta &= \sum_{p_i \in \mathcal{P}} \sum_{p_j \in \mathcal{P}} d_{ij} (R_j M_i + R_i M_j), \\ \eta &= \sum_{p_i \in \mathcal{P}} \left[\sum_{p_j \in \mathcal{P}} \frac{d_{ij}}{(1+\mu\beta)^2} M_i M_j + \sum_{e \in p_i} c_e \frac{1}{1+\mu\beta} M_i \right]. \end{aligned} \tag{32}$$

Note that the second term in (30) depends on both r and μ. The following argument shows that $\zeta = 0$. By factoring and reversing sum order, we obtain

$$\zeta = \sum_{p_i\in\mathcal{P}} M_i \sum_{p_j\in\mathcal{P}} R_j d_{ij} + \sum_{p_j\in\mathcal{P}} M_j \sum_{p_i\in\mathcal{P}} R_i d_{ij}.$$

Choose any $p_k \in \mathcal{P}$. By Lemma 4.1, $\forall p_i, p_k \in \mathcal{P}$, $\sum_{p_j\in\mathcal{P}} R_j d_{ij} = \sum_{p_j\in\mathcal{P}} R_j d_{kj}$, and analogously, $\forall p_j, p_k \in \mathcal{P}$, $\sum_{p_i\in\mathcal{P}} R_i d_{ij} = \sum_{p_i\in\mathcal{P}} R_i d_{ik}$. Thus, ζ can be written as

$$\zeta = \left(\sum_{p_i\in\mathcal{P}} M_i\right) \sum_{p_j\in\mathcal{P}} R_j d_{kj} + \left(\sum_{p_j\in\mathcal{P}} M_j\right) \sum_{p_i\in\mathcal{P}} R_i d_{ik}.$$

By Lemma 4.1, we know that each term in parentheses equals 0, so $\zeta = 0$. This allows us to express the total latency as the sum of a function dependent on r, and a function dependent on μ and β. Thus, we see here the decoupling of the effects of rate and tolls.

Consider now the $1/(1+\mu\beta)^2$ term of (32) and denote it K_μ:

$$K_\mu = \sum_{p_i\in\mathcal{P}} M_i \sum_{p_j\in\mathcal{P}} d_{ij} M_j. \tag{33}$$

By (21) in Lemma 4.1, $\forall p_i, p_k \in \mathcal{P}$, we know that

$$\sum_{p_j\in\mathcal{P}} d_{ij} M_j = \sum_{p_j\in\mathcal{P}} d_{kj} M_j + \sum_{e\in p_k} c_e - \sum_{e\in p_i} c_e.$$

Accordingly, we can write K_μ as

$$\left(\sum_{p_i\in\mathcal{P}} M_i\right)\left[\sum_{p_j\in\mathcal{P}} d_{jk} M_j + \sum_{e\in p_k} c_e\right] - \sum_{p_i\in\mathcal{P}} M_i \sum_{e\in p_i} c_e.$$

By Lemma 4.1 we know that $\sum_{p_i\in\mathcal{P}} M_i = 0$. Therefore K_μ is simply

$$K_\mu = -\sum_{p_i\in\mathcal{P}} \sum_{e\in i} c_e M_i.$$

We can then express (30) as

$$\begin{aligned} L^{\mathrm{ne}}(\mu; G, \beta) &= L_r(r) + \left(\frac{1}{(1+\mu\beta)^2} - \frac{1}{(1+\mu\beta)}\right) K_\mu \\ &= L_r(r) - \frac{\mu\beta}{(1+\mu\beta)^2} K_\mu. \end{aligned} \tag{34}$$

To complete the proof of the lemma, we show that $K_\mu \geq 0$. Equation (34) clearly shows that

$$L^{\mathrm{ne}}(0; G, \beta) = L_r(r). \tag{35}$$

That is, $L_r(r)$ represents the total latency of the un-tolled system. Equations (34) and (35) show that for a given game G, if some toll decreases latency, then all tolls must decrease latency. Since $\mu = 1/\beta$ always results in optimal routing[4] (i.e., a reduction in latency), it must be true that for any G, $K_\mu \geq 0$. Thus, scaled marginal-cost tolls always decrease total latency. This property of the total latency of a Nash flow lets us concisely quantify the effect of tolls, and leads directly to our proof of the optimal marginal-cost toll scale factor. ■

[4]See (9) and (11) in Section II.

Proof of Theorem 3.2. Consider the partial derivative of (29) with respect to β:

$$\begin{aligned} \frac{\partial L^{\mathrm{ne}}(\mu; G, \beta)}{\partial \beta} &= -\frac{\partial}{\partial \beta}\left(\frac{\mu\beta}{(1+\mu\beta)^2} \cdot K_\mu\right) \\ &= \left(\frac{\mu(\mu\beta - 1)}{(1+\mu\beta)^3}\right) K_\mu. \end{aligned} \tag{36}$$

From (36), it can be seen that $L^{\mathrm{ne}}(\mu; G, \beta)$ has a minimum at $\beta = 1/\mu$, so the end-points of the β-range give the worst possible latencies:

$$\sup_{\beta\in[\beta_{\mathrm{L}},\beta_{\mathrm{U}}]} \{L^{\mathrm{ne}}(\mu; G, \beta)\} = \max\{L^{\mathrm{ne}}(\mu; G, \beta_{\mathrm{L}}), L^{\mathrm{ne}}(\mu; G, \beta_{\mathrm{U}})\}.$$

Consider now the partial derivative of $L^{\mathrm{ne}}(\mu; G, \beta)$ with respect to μ:

$$\frac{\partial L^{\mathrm{ne}}(\mu; G, \beta)}{\partial \mu} = \left(\frac{\beta(\mu\beta - 1)}{(1+\mu\beta)^3}\right) K_\mu. \tag{37}$$

Clearly, for any $\beta > 0$, $L^{\mathrm{ne}}(\mu; G, \beta)$ has a minimum at $\mu = 1/\beta$. Furthermore, since $L^{\mathrm{ne}}(1/\beta; G, \beta) = L(f^{\mathrm{opt}}(G))$, it must be true that

$$L^{\mathrm{ne}}(1/\beta_{\mathrm{L}}; G, \beta_{\mathrm{L}}) = L^{\mathrm{ne}}(1/\beta_{\mathrm{U}}; G, \beta_{\mathrm{U}}).$$

Thus, since $L^{\mathrm{ne}}(\mu; G, \beta)$ is continuous in μ, there must exist some $\mu^* \in [1/\beta_{\mathrm{U}}, 1/\beta_{\mathrm{L}}]$ such that

$$L^{\mathrm{ne}}(\mu^*; G, \beta_{\mathrm{L}}) = L^{\mathrm{ne}}(\mu^*; G, \beta_{\mathrm{U}}).$$

It can easily be verified from (29) that

$$\mu^* = \frac{1}{\sqrt{\beta_{\mathrm{L}}\beta_{\mathrm{U}}}}. \tag{38}$$

Now we bound the inefficiency of a game G under tolls as defined in (38). Since we know that an un-tolled latency can never be more than $4/3$ times an optimal latency, from (29) we can write

$$\begin{aligned} \frac{L^{\mathrm{ne}}(0; G, \beta)}{L^{\mathrm{opt}}(G)} &= \frac{L_r(r)}{L_r(r) - \frac{1}{4}K_\mu} \\ &\leq \frac{4}{3}. \end{aligned} \tag{39}$$

This allows us to compute a bound on K_μ:

$$K_\mu \leq L_r(r). \tag{40}$$

It follows algebraically that for μ^* as defined in (38), $\beta \in [\beta_{\mathrm{L}}, \beta_{\mathrm{U}}]$, and $\mathcal{G}^*$ as defined in the statement of the theorem,

$$\mathrm{PoA}(\mu^*; \mathcal{G}^*, \beta) = \frac{4\left(1 + \sqrt{\beta_{\mathrm{L}}/\beta_{\mathrm{U}}} + \beta_{\mathrm{L}}/\beta_{\mathrm{U}}\right)}{3\left(1 + \sqrt{\beta_{\mathrm{L}}/\beta_{\mathrm{U}}}\right)^2}.$$

We show that this bound is tight in Section V-B with Pigou's Example. ■

V. DISCUSSION

A. A Note on Full Utilization

To see that Definition 3.1 applies to a broad class of networks, consider a simple network of n parallel links, with latency functions of the form specified in (1). Due to the simplicity of the network, $\mathcal{P} = E$. Let

$$\bar{G} = (V, E, \mathcal{P}, \{l_e\}_{e \in E}, \bar{r}).$$

Suppose that every path p_i carries positive rate $f^{\mathrm{ne}}_{p_i}(\bar{G})$ in a Nash flow. That is, all links have equal cost. Now, let

$$G = (V, E, \mathcal{P}, \{l_e\}_{e \in E}, r),$$

where $r > \bar{r}$. That is, our only change from $\bar{G}$ to G is that we increased the traffic rate for G. By increasing the overall traffic rate, there must be some path p_j such that $f^{\mathrm{ne}}_{p_j}(G) > f^{\mathrm{ne}}_{p_j}(\bar{G})$. Thus,

$$J_{p_j}(f^{\mathrm{ne}}_{p_j}(G)) \geq J_{p_j}(f^{\mathrm{ne}}_{p_j}(\bar{G})).$$

This implies that for any path $p_i \neq p_j$,

$$J_{p_i}(f^{\mathrm{ne}}_{p_i}(G)) \geq J_{p_j}(f^{\mathrm{ne}}_{p_j}(\bar{G})) = J_{p_i}(f^{\mathrm{ne}}_{p_i}(\bar{G})).$$

Since cost functions are nondecreasing, this implies that $f^{\mathrm{ne}}_{p_i}(G) \geq f^{\mathrm{ne}}_{p_i}(\bar{G})$. That is, no path flow decreased; thus, all networks of n parallel links such that all links carry positive flow satisfy Definition 3.1. There are many ways to extend this class of networks; for example, simple arguments can show that any combination of parallel networks in series with other parallel networks also satisfies Definition 3.1.

B. Pigou's Example

To demonstrate these results, let us turn to a simple canonical network routing problem known as Pigou's Example [20], illustrated in Figure 1. In this example, traffic has access to two routes: the first, a costly constant-latency route; the second, a congestion-sensitive route. Let $r = 1$. Without tolls, all traffic prefers route 2 since no user can improve her latency by using route 1, and the total latency $L^{\mathrm{ne}}(0; G, \beta) = 1$. [5]

However, consider the optimal routing: the traffic splits evenly between the routes so that $f_{p_1} = f_{p_2} = 0.5$; now half the population experiences a latency of 0.5 and half 1, so the total latency is $L(f^{\mathrm{opt}}(G)) = 3/4$, a 25% improvement.

We know that charging a toll equal to f_2 (corresponding to $\mu = 1/\beta$; see (9) and (11) in Section II) on route 2 will enforce this optimal flow. Now assume that all users share some value of $\beta \in [\beta_{\mathrm{L}}, \beta_{\mathrm{U}}]$. Consider for example the case in which $\beta = \beta_{\mathrm{L}}$. If our toll is f_2/β_{L}, the routing will still be optimal. However, for $\beta > \beta_{\mathrm{L}}$, a disproportionate number of users will divert to route 1.

[5] The astute reader will notice that this example does not strictly satisfy Definition 3.1. This will not adversely impact our analysis, because we may consider the case in which $r = 1+\epsilon$ for $\epsilon > 0$. Then $f_{p_1} = \epsilon$ and $f_{p_2} = 1$, so Definition 3.1 is satisfied. We take the limit as $\epsilon \to 0$, and all analysis proceeds as before.

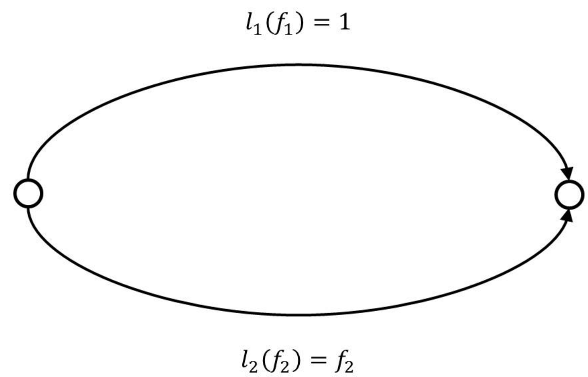

Fig. 1. Pigou's Example. Network used in Section V-B to demonstrate the price of anarchy for scaled marginal-cost tolls and uncertain user populations.

β	μ	τ_2	f_{p_1}	f_{p_2}	$L^{\mathrm{ne}}(\mu : G, \beta)$
$\beta_{\mathrm{L}} = 1$	1	f_{p_2}	0.5	0.5	0.75
$\beta_{\mathrm{U}} = 9$	1	$9f_{p_2}$	0.9	0.1	0.91
$\beta_{\mathrm{L}} = 1$	$1/3$	$(1/3)f_{p_2}$	0.75	0.25	0.8125
$\beta_{\mathrm{U}} = 9$	$1/3$	$3f_{p_2}$	0.25	0.75	0.8125

TABLE I
TOLLS AND TOTAL LATENCY IN PIGOU'S EXAMPLE

Now, the Nash flow induced by $\mu > 0$ is such that $f_{p_1} = \beta/(\mu+\beta)$ and $f_{p_2} = \mu/(\mu+\beta)$. Thus the total latency is

$$\begin{aligned} L^{\mathrm{ne}}(\mu; G, \beta) &= \frac{\beta}{(\mu+\beta)} + \left(\frac{\mu}{\mu+\beta}\right)^2 \\ &= \frac{(\mu)^2 + \beta(\mu+\beta)}{(\mu+\beta)^2}. \end{aligned} \tag{41}$$

It is easy to verify that for $\mu = \mu^* = (\beta_{\mathrm{L}}\beta_{\mathrm{U}})^{-1/2}$ as prescribed by Theorem 3.2 and $\beta = \beta_{\mathrm{L}}$, the total latency is

$$L^{\mathrm{ne}}(\mu^*; G, \beta_{\mathrm{L}}) = \frac{\left(1 + \sqrt{\beta_{\mathrm{L}}/\beta_{\mathrm{U}}} + \beta_{\mathrm{L}}/\beta_{\mathrm{U}}\right)}{\left(1 + \sqrt{\beta_{\mathrm{L}}/\beta_{\mathrm{U}}}\right)^2}. \tag{42}$$

Since $L(f^{\mathrm{opt}}(G)) = 3/4$, the result in (42) proves the tightness of the price of anarchy specified in Theorem 3.2.

Table I presents some numerical examples. In particular, if $\beta \in [1, 9]$, then for $\mu = \mu^* = 1/3$, the total latency will never be worse than 0.8125, nearly a 19% improvement over the un-tolled case.

C. Braess Topology

To illustrate a network that may not satisfy Definition 3.1, we turn to a second canonical example: that of Braess's Paradox [23], [14]. The topology in Braess's Paradox has been used to show that adding links to a graph can increase the cost of a Nash flow; we use it to show that increasing the total rate of flow can cause an individual path flow to decrease.

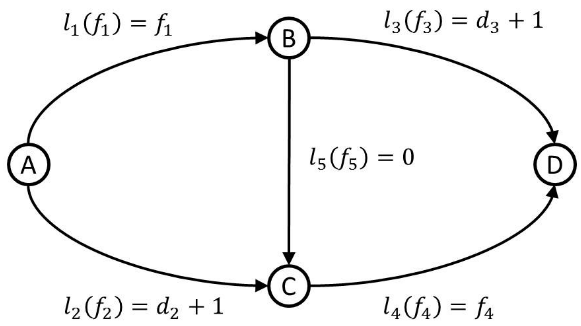

Fig. 2. Braess's Topology. For $d_2 = d_3 = 0$, no game using this topology satisfies Definition 3.1. However, for $d_2 = 1$, $d_3 = 2$, games using this topology can satisfy Definition 3.1.

Variant	R_1	R_2	R_3
1	1	-1	1
2	$1/3$	$1/6$	$1/2$

TABLE II
VALUES OF R_i FOR BRAESS TOPOLOGY

Consider the network in Figure 2. Note that in the canonical Braess's Paradox network, $d_2 = d_3 = 0$. We consider two variants of the network: in Variant 1, $d_2 = d_3 = 0$; in Variant 2, $d_2 = 1$, $d_3 = 2$. Traffic routes from node A to node D, and has a choice of three paths. Define $p_1 = \{e_1, e_3\}$, $p_2 = \{e_1, e_5, e_4\}$, and $p_3 = \{e_2, e_4\}$. The values of R_i (refer to Lemma 4.1 in Section IV) for each variant appear in Table II.

For Variant 1, $R_2 = -1$. This means that increasing the total traffic rate will actually *decrease* the amount of traffic using path 2, which is to say edge 5. Thus this network never satisfies Definition 3.1, and our optimal toll analysis is not necessarily valid.

For Variant 2, the topology is unchanged, but the modified latency functions have changed routing behavior dramatically. Now $R_i \geq 0$ for every path, so all path flows are increasing in r, and thus if all paths are in use, the game satisfies Definition 3.1. Thus, scaled marginal-cost tolls can never degrade network performance.

D. Price of Anarchy

Our expression for the price of anarchy in Theorem 3.2 succinctly characterizes the relationship between efficiency and information about the user population. It illustrates the intuitive principle that with perfect information, a designer can achieve perfect efficiency. Figure 3 shows how the price of anarchy varies with $\beta_{\mathrm{L}}/\beta_{\mathrm{U}}$. Note that for any value of this ratio, the price of anarchy is less than $4/3$; that is, less than that of an un-tolled network.

Though imperfect information does degrade efficiency, the figure shows that high levels of efficiency are maintained even for a relatively high level of uncertainty. Note that even

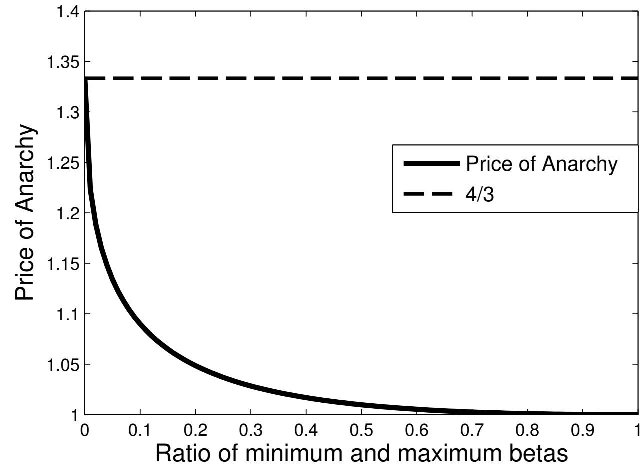

Fig. 3. Price of anarchy for games using our scaled marginal-cost tolling rule, as a function of the precision with which we know the user population's price-sensitivity.

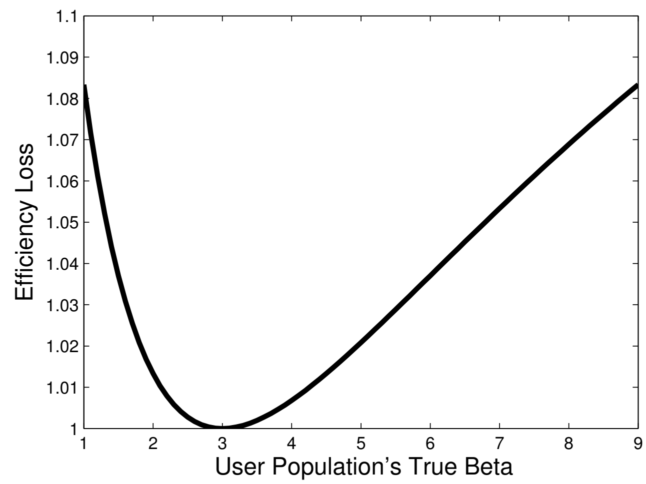

Fig. 4. Efficiency Loss in Pigou's Example with respect to true beta for $\mu = 1/3$ and $r = 1$. Note that the efficiency loss is equal at the endpoints of the β-range, and perfect efficiency is achieved at the geometric mean of the endpoints.

when β is only known within a factor of 10, the efficiency loss is less than 10%.

Our price of anarchy result is encouraging for two reasons. First, this worst-case analysis was obtained from assuming that the *entire* population lies at one of the extremes of the range. This assumption, though important to establish the bound, is likely to be unrealistic for most systems. Any β closer to our geometric mean tolling parameter will yield even better performance than the price of anarchy suggests. This is illustrated in Figure 4; the figure was generated using values from Pigou's Example in (see Section V-B) with $r = 1$, $\beta_{\mathrm{L}} = 1$, $\beta_{\mathrm{U}} = 9$, and $\mu = 1/3$. The horizontal axis represents the true β of the population; thus it is easy to see that values close to the geometric mean allow only very low efficiency loss.

Second, the analysis contains a strong assumption of homogeneity; all users are required to have identical values of β. This is also unlikely in real systems. Consider a population whose β-values are distributed over some range. More rigorous analysis is needed, but our simulations and intuition have suggested that if a greater mass of users are closer to the geometric mean than to the endpoints of the range, tolls would achieve considerably improved performance.

VI. CONCLUSIONS

Effective design for social coordination requires information about a system's user population; we have provided initial results on the relationship between the quality of this information and the efficiency of Nash equilibria resulting from a tolling scheme. In our model, imperfect information caused a bounded degradation of efficiency, but relatively high efficiency was maintained even for high levels of uncertainty. Future research will focus on extending this result to broader classes of games. In particular, we hope to eliminate the requirement that games satisfy Definition 3.1 and develop a tolling rule applicable to all affine-cost routing games.

In the model addressed in this paper, we have shown that even under uncertainty, it is possible to achieve significant gains in the efficiency of social behavior with a simple distributed taxation scheme. Our model strongly assumes homogeneity of the population's price sensitivities; as research progresses, we hope to relax this assumption. Homogeneity is rare in practice, and incorporating the effects of heterogeneity into our models will be a crucial step. In particular, uncertain heterogeneous populations may be well-suited to heterogeneous tolls or a level of price discrimination. A crucial question is: what gains in efficiency are possible if we charge different tolls for different users?

This work has found the optimal toll among a specific class of tolling functions: Pigovian marginal-cost tolls scaled by a uniform factor. It remains to be shown how our tolls compare with tolls of some other form. Are these tolls the best of any possible toll? A more comprehensive review of general taxation functions for uncertain populations is required to answer this question.

REFERENCES

[1] G. E. Bolton and A. Ockenfels, "Behavioral economic engineering," *Journal of Economic Psychology*, vol. 33, pp. 665–676, June 2012.

[2] C. Silva, "Application of mechanism design to electric power markets (Republished)," *IEEE Transactions on Power Systems*, vol. 16, no. 4, pp. 862–869, 2001.

[3] D. Niyato and E. Hossain, "Competitive Pricing for Spectrum Sharing in Cognitive Radio Networks : Dynamic Game , Inefficiency of Nash Equilibrium , and Collusion," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, pp. 192–202, 2008.

[4] S. Morrison, "A survey of road pricing," *Transportation Research Part A: General*, no. 2, pp. 87–97, 1986.

[5] C. Papadimitriou, "Algorithms , Games , and the Internet," in *Proc. of the 28th International Colloquium on Automata, Languages and Programming*, 2001.

[6] E. Koutsoupias and C. Papadimitriou, "Worst-case equilibria," *Computer Science Review*, vol. 3, pp. 65–69, May 2009.

[7] R. Johari and J. N. Tsitsiklis, "Efficiency Loss in a Network Resource Allocation Game," *Mathematics of Operations Research*, vol. 29, pp. 407–435, Aug. 2004.

[8] G. Perakis and G. Roels, "The Price of Anarchy in Supply Chains: Quantifying the Efficiency of Price-Only Contracts," *Management Science*, vol. 53, pp. 1249–1268, Aug. 2007.

[9] H. Youn, M. Gastner, and H. Jeong, "Price of Anarchy in Transportation Networks: Efficiency and Optimality Control," *Physical Review Letters*, vol. 101, p. 128701, Sept. 2008.

[10] N. G. Mankiw, "Smart Taxes: An Open Invitation to Join the Pigou Club," *Eastern Economic Journal*, vol. 35, no. 1, pp. 14–23, 2009.

[11] W. Baumol, "On taxation and the control of externalities," *The American Economic Review*, vol. 62, no. 3, pp. 307–322, 1972.

[12] J. Stiglitz, "Using tax policy to curb speculative short-term trading," *Journal of Financial Services Research*, vol. 115, no. May, pp. 101–115, 1989.

[13] I. W. H. Parry, M. Walls, and W. Harrington, "Automobile externalities and policies," *SSRN Electronic Journal*, 2007.

[14] T. Roughgarden, *Selfish routing and the price of anarchy*. MIT Press, 2005.

[15] R. Cole, Y. Dodis, and T. Roughgarden, "How much can taxes help selfish routing?," *Journal of Computer and System Sciences*, vol. 72, pp. 444–467, May 2006.

[16] H. Yang and H.-J. Huang, "Principle of marginal-cost pricing: how does it work in a general road network?," *Transportation Research Part A: Policy and Practice*, vol. 32, pp. 45–54, Jan. 1998.

[17] M. J. Smith, "The Marginal Cost Taxation of a Transportation Network," *Transportation Research Part B: Methodological*, vol. 13, no. 3, pp. 237–242, 1979.

[18] R. Cole, Y. Dodis, and T. Roughgarden, "Pricing network edges for heterogeneous selfish users," *Proc. of the thirty-fifth ACM symp. on Theory of computing - STOC '03*, p. 521, 2003.

[19] L. Fleischer, K. Jain, and M. Mahdian, "Tolls for heterogeneous selfish users in multicommodity networks and generalized congestion games," *45th Annu. IEEE Symp. on Foundations of Computer Science*, pp. 277–285, 2004.

[20] A. C. Pigou, *The Economics of Welfare*. 1920.

[21] T. Roughgarden, "How Bad Is Selfish Routing?," *Journal of the ACM JACM (2002)*, vol. 49, no. 2, pp. 236–259, 2002.

[22] J. Wardrop, "Some theoretical aspects of road traffic research," in *Proc. of the Institution of Civil Engineers, Part II, Vol. 1, No. 36*, pp. 352–362, 1952.

[23] D. Braess, "Uber ein Paradoxon aus der Verkehrsplanung," *Unternehmensforschung*, pp. 258–268, 1968.