Studies on Robust Social Influence Mechanisms

Incentives for Efficient Network Routing in Uncertain Settings

Philip N. Brown and Jason R. Marden

Many of today's engineered systems are tightly interconnected with their users, and in many cases, system performance depends greatly on user behavior [1]. As a result, the traditional lines between engineering and the social sciences are becoming increasingly blurred, and analytical tools such as game theory are finding new applications in engineering [2], [3]. It is often insufficient to judge an engineered system on its technical merits alone, since strategic user behavior can lead to unpredictable and undesirable results [4]. Of particular importance to this article are socially-integrated engineering problems in which users' strategic behavior has a significant impact on overall system performance. These types of systems appear in a variety of contexts in theory and practice: transportation networks [5], ridesharing applications [6], [7], supply-chain management [8], cloud computing [9], and electric power grids [10] are immediate examples. A common problem in these settings is that individual users' incentives may not be aligned with the objectives of the central planner. Thus, in addition to the merely-technical challenges they pose, an engineer may need to consider methods of influencing individual user behavior to effect positive change on aggregate system performance [11]. These behavior-influencing mechanisms often take the form of offering users a tradeoff between quality-of-service and monetary incentive.

Fundamentally, any behavior-influencing mechanism requires information about the underlying system and the user population who are to be influenced, as depicted in Figure 1. For example, if a system planner desires to price a network resource to encourage efficient network usage, it may be desirable to characterize the sensitivity of the user population to pricing. If this information is difficult to gather or altogether unavailable, the planner may need to rely on crude estimates of user price-sensitivities, and the pricing design must take this uncertainty into account. At worst, a misunderstanding of informational dependencies can lead to "perverse incentives," or incentives that exacerbate the very problems they were intended to solve.

Here, a theory of "robust social influence" is an attractive goal: how can behavior-influencing mechanisms be designed so that they are robust to a variety of mischaracterizations or variations in models of social behavior? Some natural questions in this context include

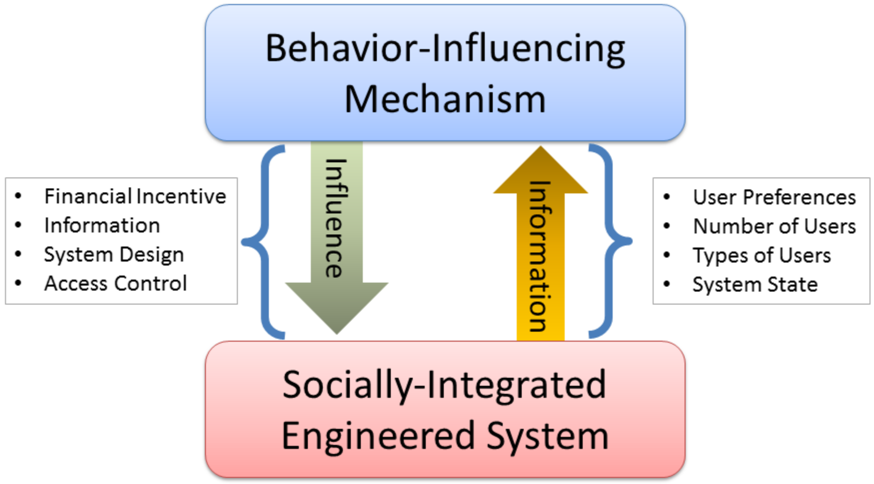

Figure 1. Relationship between social system and behavior-influencing mechanism. In order to exert influence (whether via financial incentives, access control, system design, or providing specific information to users), a behavior-influencing mechanism typically requires some information about the social system to be influenced. Recent research has investigated the relationship between the quality of this information and the resulting effectiveness of the influence mechanism. This article highlights recent efforts to characterize this informational dependence.

1) How robust are existing behavior-influencing methodologies to system mischaracterizations?
2) How "close" do behavior models need to approximate true behavior for an influencing mechanism to provide good performance?
3) How can perverse incentives be systematically avoided?

This article investigates the concept of robust social influence with regard to a well-studied model of static network traffic routing in which drivers need to be routed across a congestion-sensitive network. In this model, the focus is typically on analyzing the quality of static routing equilibria; thus, the results discussed in this article are complementary to recent research on dynamic models of distributed routing [12]–[14]. It is known that if individual drivers make their own routing decisions to minimize their own experienced delays, overall network congestion can be considerably higher than if a central planner had the ability to explicitly direct traffic [15]. Accordingly, there has been a great deal of research on the application of road tolls for the purpose of influencing drivers to make routing choices that result in globally-optimal routing [16]–[21]. In much of this literature, a toll-designer is assumed to have a detailed characterization of the system: network topology, link congestion characteristics, and user demand structure are commonly assumed to be perfectly known quantities.

If road tolls are designed to incentivize good performance for one instance of a routing problem, and then some detail of the routing problem changes (for example, a link is "removed" by a traffic accident or natural disaster), it would be desirable for the original tolls to incentivize good performance on the changed routing problem as well. That is, we would like to know if the performance guarantees provided by the original tolls are *robust* to changes or mischaracterizations in the underlying details of the system (e.g., network structure, traffic rate, user demands). Here, we distinguish between different degrees of robustness: a taxation mechanism is *strongly robust* to mischaracterizations of some system parameter if it incentivizes optimal behavior for variations of that parameter, whereas a mechanism is *weakly robust* if it merely incentivizes behavior that is no worse than the un-influenced behavior.

Existing research has demonstrated that if a tax-designer has an accurate and detailed characterization of the routing problem, it is possible to design simple road tolls which incentivize optimal routing [22], [23]. These are termed *fixed tolls*, since they are simple constant functions of traffic flow. However, the robustness of fixed tolls has not been investigated. Another well-studied taxation mechanism is that of *marginal-cost tolls*, in which each network link is assigned a flow-varying tax that is specifically designed to penalize inefficient congestion. Marginal-cost tolls are known to incentivize optimal network routing in the special case that all network users are homogeneous in tax-sensitivity (i.e., they all value time equally) [24], [25]. These tolls are notable since they are known to be strongly robust to variations of network structure for homogeneous users. In [26], it is shown for a small class of routing problems that marginal-cost tolls are weakly robust to variations of user tax-sensitivities, but the robustness of marginal-cost tolls for general routing problems remains an open question.

The contribution of this article is to show that fixed tolls and marginal-cost tolls are not robust to a variety of mischaracterizations of routing problems and user populations. Specifically, this article's theoretical contributions are

- Proposition 4: Fixed tolls fail to be strongly robust to variations in network latency functions.
- Theorem 5: Fixed tolls cannot even be weakly robust unless they depend on some global information about network structure.
- Proposition 6: Marginal-cost tolls can create perverse incentives in general routing problems if the tax-sensitivities of users are unknown.

In all of these results, the lack of robustness is shown using very simple problem instances. For example, Theorem 5 is proved using networks with only two or three links. Since these taxation mechanisms fail on very simple networks, it seems reasonable to assume that larger networks could make things even worse; though more work is required to verify this analytically.

This article investigates the specific case of equilibria in static network routing problems, but note that at its core, taxation in routing problems is simply a matter of offering users a financial incentive to accept reduced quality-of-service. These tradeoffs between money and quality-of-service are not specific to routing problems, but appear in many contexts, including the aforementioned settings of ridesharing systems, supply chains, cloud computing, and smart grids. It may be that the approaches used in this article will bear fruit in these other diverse settings as well.

Traffic Routing Model and Toll Robustness

A classical example of an engineered system whose performance depends heavily on the choices of its users is that of a transportation network; this is captured in the literature by a problem known as a "non-atomic routing game." The basic problem setup is this: there is a group of travelers who need to be routed through a congestion-sensitive network in a way that minimizes the users' average travel time. It is typically straightforward to compute an optimal routing profile (also called a *network flow*), but implementing a particular flow would require that a central planner had the ability to force every driver to take a specified route. Unfortunately, it is well-known that if each driver chooses his route in order to individually minimize his own travel time, the resulting aggregate behavior can be substantially less efficient than the centrally-computed optimal flow [27].

Since the system planner cannot direct traffic explicitly, he must use some indirect means of influencing users to behave optimally. Various methods which have been studied include financial incentives [28], providing drivers with specific information [29], access control [30], and socially-conscious network design [31]. All of these operate by indirectly modifying the drivers' preferences over their available routes. Following a formal model introduction, this section provides examples of the inefficiency resulting from self-interested behavior, and shows that simple attempts to influence user behavior can have surprisingly negative results.

Model and Performance Metrics

Consider a network routing problem in which a mass of r units of traffic needs to be routed across a directed, acyclic network (V, E), which consists of a vertex set V and edge set $E \subseteq (V \times V)$. We write $\mathcal{P} \subset 2^E$ to denote the common set of *paths* available to traffic, where each path $p \in \mathcal{P}$ consists of a set of edges connecting the source to the destination. We write f_p

to denote the mass of traffic using path p. A *feasible flow* $f \in \mathbb{R}^{|\mathcal{P}|}$ is an assignment of traffic to various paths such that $\sum_{p\in\mathcal{P}} f_p = r$ and for every path p, $f_p \geq 0$. Note that in this model, demand is inelastic: all travelers must arrive at their destination. For more information about a model variation which allows drivers to choose not to travel, see "Elastic Traffic."

Given a flow f, the flow on edge e is given by $f_e = \sum_{p:e\in p} f_p$. Note that f always denotes a vector of path flows; when referring to an individual edge flow, the symbol f_e is always used with an unambiguous e subscript. To characterize transit delay as a function of traffic flow, each edge $e \in E$ is associated with a specific latency function $\ell_e : [0, r] \to [0, \infty)$. We adopt the standard assumptions that latency functions are nondecreasing, continuously differentiable, and convex: these assumptions model the natural phenomenon that adding traffic to a link provides gracefully-decreasing marginal benefits. Note that latency functions are anonymous: all users affect network delay equally. The system-planner's hope is to minimize the *total latency*, which is a measure of aggregate network delay, expressed by the formula

$$\mathcal{L}(f) \triangleq \sum_{e\in E} f_e \cdot \ell_e(f_e) = \sum_{p\in\mathcal{P}} f_p \cdot \ell_p(f), \tag{1}$$

where $\ell_p(f) = \sum_{e\in p} \ell_e(f_e)$ denotes the latency on path p. We denote the flow that minimizes the total latency by

$$f^* \in \underset{f \text{ is feasible}}{\operatorname{argmin}} \ \mathcal{L}(f). \tag{2}$$

Due to the convexity of each ℓ_e, $\mathcal{L}(f^*)$ is unique. Finally, a *routing game* is given by the tuple $G = \{V, E, \mathcal{P}, \{\ell_e\}_{e\in E}, r\}$. We write the set of all such routing problems as $\mathcal{G}$.

An inherent challenge in this setting is that users' self-interested routing choices may lead to very different aggregate flows than those which minimize the total latency. One popular way to model the effects of users' self-interested choices is that of the Nash Flow (also called Wardrop Equilibrium). A Nash flow $f^{\rm nf}$ is a routing profile in which no individual user can change routes and decrease his or her latency; put differently, in a Nash flow every user's route choice is individually-optimal with respect to the choices of other users. Formally, in a Nash flow $f^{\rm nf}$, for any two paths p and p', if $f_p^{\rm nf} > 0$, then $\ell_p(f^{\rm nf}) \leq \ell_{p'}(f^{\rm nf})$. That is, if any user is choosing path p, then the latency experienced on path p can be no greater than the latency of any other path p'. This fact implies that in a Nash flow, every path with positive traffic has equal latency:

$$\text{If } f^{\rm nf} \text{ is a Nash flow and } f_p^{\rm nf}, f_{p'}^{\rm nf} > 0 \text{ for some paths } p,\ p' \text{ then } \ell_p\left(f^{\rm nf}\right) = \ell_{p'}\left(f^{\rm nf}\right). \tag{3}$$

In this setting, Nash flows always exist and are essentially unique: that is, for a given network, any two Nash flows f and f' satisfy $\mathcal{L}(f) = \mathcal{L}(f')$. [1].

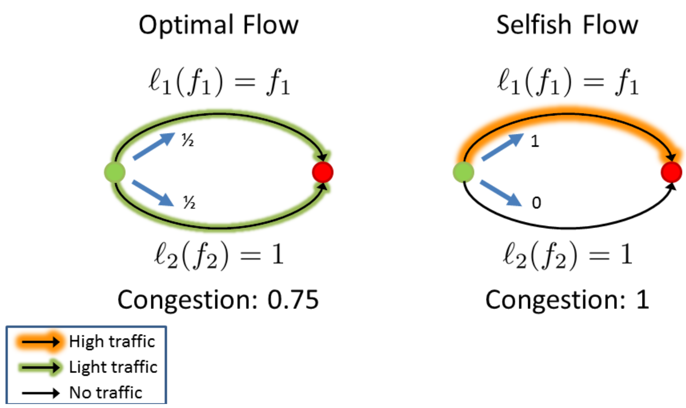

Figure 2. Pigou's Network, illustrating the negative effects of selfish behavior. In this network, drivers can choose between the upper, congestion-sensitive link and the lower, constant-latency link. The image on the left depicts a congestion-minimizing routing profile in which the traffic is split evenly between the two links. However, in this optimal flow, agents on the lower link experience a latency of 1, and (individually) could decrease their travel time by switching to the upper link. Unfortunately, this self-interested behavior can have negative consequences for system performance. The image on the right depicts a routing profile arising when every driver chooses the path with lowest delay; here, drivers have crowded on to the upper link, degrading its performance. A central problem is that no driver has an incentive to choose the lower, more efficient path.

Pigou's Example: the Inefficiency of Self-Interested Behavior

The first example, depicted in Figure 2, illustrates the basic problem that travelers' individual self-interested choices can lead to over-congested network flows. This illustration was first published in 1920 by the economist Arthur Pigou, and remains a centerpiece of work in this area [15]. The setting is as follows: there is a simple two-link network in which 1 unit of travelers can choose between a linear-latency congestion-sensitive link and a constant-latency link. The first link offers a faster journey, provided that it is not chosen by too many users.

The optimal flow on this network as shown on the left in Figure 2 is to split the traffic evenly between the two links, so that a mass of $1/2$ experiences a latency of $1/2$ on the top link, and the remaining traffic experiences a latency of 1 on the lower link, giving a total latency of $\mathcal{L}(f) = (1/2)^2 + 1/2 = 0.75$. However, this requires half the drivers to choose quite a long route; any individual driver has a compelling incentive to switch to the upper link and arrive at her destination in half the time. Unfortunately, if all drivers choose the route with the lowest latency,

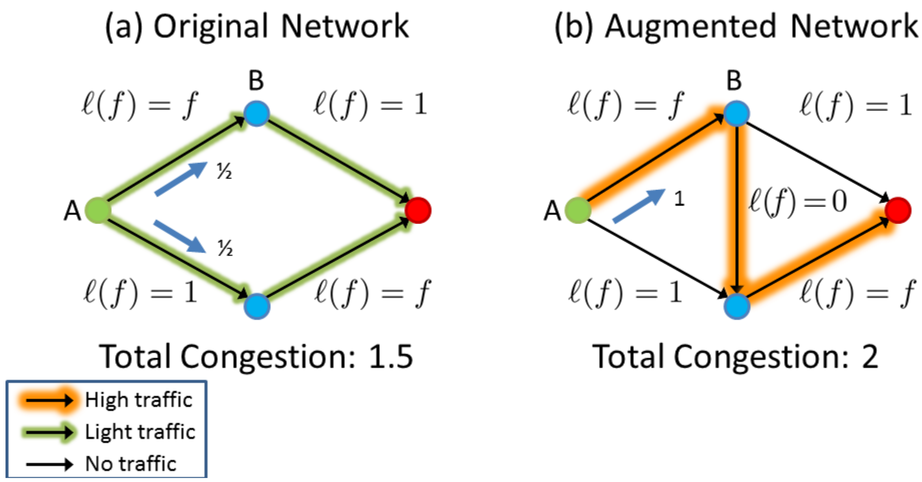

Figure 3. Braess's network, depicting an unintended consequence of attempting to influence social behavior. The image on the left in (a) depicts a transportation network and its associated Nash flow. A well-meaning traffic engineer, hoping to improve network congestion, adds a new link to the network connecting the two intermediate nodes. Despite the fact that this link's cost is zero, its addition to the network leads to the setting on the right in (b). Any user at node B can take the new link without increasing his cost, but in doing so, he increases the cost of the lower path, which in turn leads to more users at node A choosing the upper path. The ultimate effect of augmenting the network with a zero-cost link is that every driver's travel time increases by 33%.

they will all crowd on to the upper link and establish a Nash flow as depicted on the right in Figure 2. The reader can verify that $(1, 0)$ is indeed a Nash flow, since in this configuration, both edges have an equal latency of 1, so no user can change routes and decrease their latency. In this Nash flow, the network's total congestion is 1, a factor of $4/3$ greater than the optimal total congestion.

Braess's Paradox: the Unintended Consequences of Naïve Influence

A second canonical example known as Braess's Paradox (first noted by Dietrich Braess [4]) illustrates that seemingly-innocuous attempts to influence user behavior can lead to unexpected and perverse consequences. Consider the network depicted in Figure 3(a); traffic can choose between two paths, each routing through its own intermediate node. As-is, the total congestion on the network is 1.5, since half the traffic uses the upper path and half uses the lower path.

Suppose now that the system planner adds a single zero-cost link to the network connecting the two intermediate nodes to one another, as depicted in Figure 3(b). Now, under the old flow in which users split evenly, any user at node (B) would prefer to take the new zero-cost link rather than continue on the upper path. This increases the lower path's congestion, causing more users at (A) to choose the upper path, but those users in turn will choose the new zero-cost link once they arrive at node (B). Ultimately, equilibrium is reached at the routing profile depicted in Figure 3(b), with a corresponding total congestion of 2. Here, this behavior-influencing mechanism (augmenting the network with a zero-cost link) backfired and caused a dramatic increase in total congestion.

Price of Anarchy: Worst-Case Equilibria

A common approach to quantifying the performance loss resulting from self-interested behavior is to divide the total latency of a Nash flow with that of an optimal flow. In each case above, this ratio was $4/3$ – the Nash flow in each instance was 33% worse than the corresponding optimal flow. This ratio between equilibrium and optimal costs has been extensively explored in the literature and is known as the *price of anarchy* [32]. It is typically evaluated in worst-case over classes of games; formally, for a game G in a class of games $\mathcal{G}$, writing $\mathcal{L}^{\rm nf}(G)$ to denote the Nash flow total latency for game G and $\mathcal{L}^{*}(G)$ to denote the respective optimal total latency, the price of anarchy of $\mathcal{G}$ is defined as

$$\mathrm{PoA}(\mathcal{G}) = \max_{G \in \mathcal{G}} \frac{\mathcal{L}^{\rm nf}(G)}{\mathcal{L}^{*}(G)}. \tag{4}$$

For example, if $\mathcal{G}$ is defined as the class of routing games where all edges have linear-affine latency functions, it is known that $\mathrm{PoA}(\mathcal{G}) = 4/3$ [33]. Thus, both Pigou's example and Braess's paradox are worst-case examples of linear-latency routing games. However, for networks with higher-degree polynomial latency functions, the price of anarchy can be arbitrarily high. This can be seen in Pigou's network by replacing the linear cost function on the first link with a polynomial cost function of $\ell_1(f_1) = (f_1)^d$; the price of anarchy is unbounded in d [34]. The price of anarchy has been evaluated for many different classes of games, with applications as diverse as network resource allocation [35], distributed control [36], and more. For more details, see "Computing Price of Anarchy Bounds."

Robust Social Coordination Using Tolls

It was shown above how self-interested behavior can lead to poor system performance; in both examples, the fundamental problem was that in an optimal flow, users could decrease their own cost only by imposing a greater cost on those around them. If users were altruistic, willing to accept a personal degradation of service for the sake of the greater good, they might be expected to adopt this type of socially-optimal configuration. One way to influence users to choose this configuration is to charge tolls on over-congested links, hoping to increase their costs enough that a sufficient number of users will avoid them. The tolls are put in place essentially to induce an “artificial altruism” in the population; in fact, there are many parallels between the literature on altruism in congestion games and the literature on financial incentives in congestion games [37]. Of course, the toll-designer must take care in choosing the tolls. It has already been seen in Braess’s Paradox that seemingly-innocuous approaches can have unexpected consequences, and the toll-designer must be certain not to fall into a similar trap. If tolls are too high on a particular edge, it may be that too many users will avoid that edge; if tolls are not properly balanced throughout the network, uneven and inefficient flow distributions could arise.

Tolling Model

Formally, we write $\tau_e(f_e)$ to denote the (possibly flow-varying) toll on edge e; assigning tolls to network edges modifies the costs experienced by users on those edges, inducing a new game with new associated equilibria. The system-planner’s goal is to levy tolls which induce the network’s optimal flow as an equilibrium of the tolled game. To model network users’ response to tolls, the user population is represented by the interval $[0, r]$ and a *sensitivity* function $s : [0, r] \to \mathbb{R}_+$ that assigns a sensitivity value s_x to each user $x \in [0, r]$. This sensitivity can be interpreted as the reciprocal of the user’s value-of-time: given a flow f, the cost experienced by user $x \in [0, r]$ using path $\tilde{p} \in \mathcal{P}$ is simply the sum of the latencies and sensitivity-weighted tolls on that path, given by

$$J_x(f) = \sum_{e \in \tilde{p}} \left[\ell_e(f_e) + s_x \tau_e(f_e) \right]. \tag{5}$$

Intuitively, a user with a large s_x value is willing to use a high-latency route to avoid tolls, while a user with a low s_x value is willing to pay a high toll to use a low-latency route. The heterogeneous game specification now includes the population’s sensitivity function as well as other parameters: $G = \{V, E, \mathcal{P}, \{\ell_e\}_{e \in E}, r, s\}$.

Given a set of edge tolls, a Nash flow $f^{\rm nf}$ is defined in a similar way as before, with the exception that now the Nash condition must be checked for each user. That is, $f^{\rm nf}$ is a flow in which each user is individually choosing the lowest-cost route available, given the choices of other users. As before, Nash flows in the tolled setting always exist and are known to be essentially unique [38]. Formally, for each user $x \in [0, r]$,

$$J_x\left(f^{\rm nf}\right) = \min_{p \in \mathcal{P}} \left\{ \sum_{e \in p} \left[\ell_e\left(f_e^{\rm nf}\right) + s_x \tau_e\left(f_e^{\rm nf}\right) \right] \right\}. \tag{6}$$

Finally, the system-level costs are measured in the same way as before with the total latency:

$$\mathcal{L}(f) = \sum_{e \in E} f_e \ell_e(f_e) = \sum_{p \in \mathcal{P}} f_p \ell_p(f).$$

Note that tolls are not included in the system cost – the system planner is still only attempting to minimize aggregate delay.

The robustness of taxation mechanisms

It is the goal of robust social influence design to levy tolls that incentivize desirable behavior irrespective of changes or mischaracterizations of the underlying system. Figure 4 depicts several types of system changes which could potentially create problems for taxation methodologies. In these diagrams, the tolls were designed for the nominal system on the left, but after the respective change, these same tolls are effectively being applied to different networks than that for which they were designed. The hope is that the tolls designed for the original system provide comparable performance guarantees on the "new" systems; to this end, we will ask if each of several common taxation methodologies is robust to variation in parameters such as:

1) Network Changes: If tolls are designed to incentivize efficient flows for a particular network, and the network undergoes some change, do the original tolls still incentivize efficient flows for the new, changed network?
2) Traffic Rate: Do tolls designed for one traffic rate (i.e., one value of r) still incentivize efficient flows if the rate changes?
3) Demand Structure: If some users have access to different paths than others, how does this impact the design of the correct tolls?

To investigate the robustness of a particular tolling strategy to variations of a parameter, a system-planner can design tolls for a specific system realization, hold the tolls constant, and study the effect on Nash total latency of varying the parameter in question. One way to model perturbations of games is to define a correspondence $\Gamma_G(\cdot)$ that returns games similar to some nominal game G that have been perturbed in the argument of Γ_G. For example, $\Gamma_G(\{\ell, r\})$ represents the set of routing games that differ from G only in their latency functions ℓ and total traffic rate r.

To formally discuss tolling strategies, let a *taxation mechanism* $\boldsymbol{\tau}$ be a mapping from games to edge taxation functions; thus, we write $\boldsymbol{\tau}(G)$ to denote the edge taxes that $\boldsymbol{\tau}$ assigns to game G. We write $\mathcal{L}^{\rm nf}(G^*, \boldsymbol{\tau}(G))$ to mean the Nash flow total latency for some (possibly different) game G^* induced by tolls generated by $\boldsymbol{\tau}$ for the nominal game G. A taxation mechanism $\boldsymbol{\tau}$ is said to be *strongly robust on* G if $\boldsymbol{\tau}$ incentivizes optimal flows for all allowable perturbations of G.

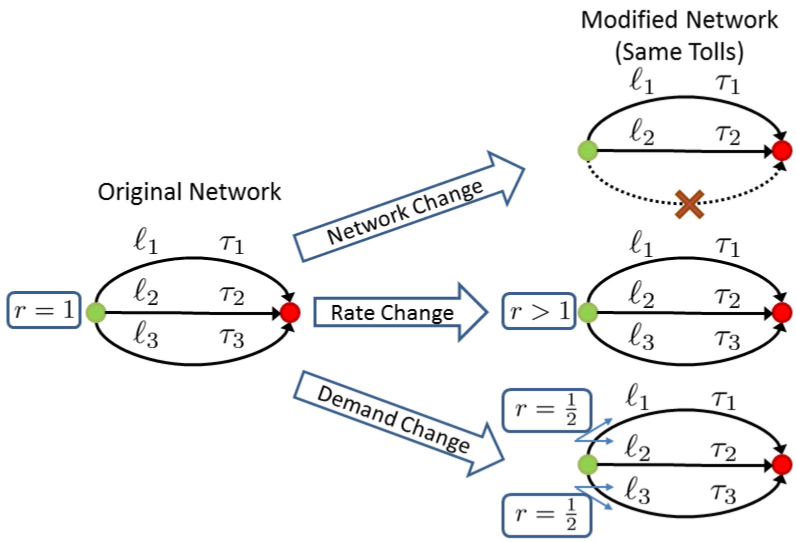

Figure 4. Diagram depicting possible network changes a designer may consider in robustness analysis. Suppose tolls τ_1, τ_2, andτ_3 are designed for the network on the left. To analyze the robustness of this toll design, the designer may subject the routing problem to various hypothetical changes: topology (deleting link 3), traffic rate (increasing r above its original value), and demand structure (restricting half the traffic to the upper two links, the other half of the traffic to the lower two links), while keeping the tolls designed for the original network. If the tolls are able to incentivize efficient behavior on the "new" networks despite having been designed for the original network, they are called robust.

That is, the strong robustness of $\boldsymbol{\tau}$ to variations in game parameters $X \subseteq \{V, E, \mathcal{P}, \{\ell_e\}_{e\in E}, r, s\}$ implies that

$$\mathcal{L}^{\rm nf}(G^*, \boldsymbol{\tau}(G)) = \mathcal{L}^*(G^*) \text{ for all } G^* \in \Gamma_G(X). \tag{7}$$

We likewise say that $\boldsymbol{\tau}$ is strongly robust on a larger class of games $\mathcal{G}$ if (7) holds for all $G \in \mathcal{G}$.

This may be too strong a condition in some settings (though strongly-robust taxation mechanisms do exist; see Theorem 2), so a taxation mechanism is said to be *weakly robust on* G if it never incentivizes Nash flows on perturbed networks that are *worse* than the un-tolled flows. That is, writing $\mathcal{L}^{\rm nf}(G, \emptyset)$ to denote the Nash flow total latency on G without tolls, the weak robustness of $\boldsymbol{\tau}$ on G to parameters specified by X implies that

$$\mathcal{L}^{\rm nf}(G^*, \boldsymbol{\tau}(G)) \leq \mathcal{L}^{\rm nf}(G^*, \emptyset) \text{ for all } G^* \in \Gamma_G(X). \tag{8}$$

Again, we say that $\boldsymbol{\tau}$ is weakly robust on a larger class of games $\mathcal{G}$ if (8) holds for all $G \in \mathcal{G}$. Put differently, if tolls are weakly robust, they will not create perverse incentives. Thus, by the definitions presented in this section, assigning a toll of 0 to every link is always a weakly robust taxation mechanism, as it certainly cannot make Nash flows worse.

Survey of Existing Taxation Methodologies

Fixed tolls for designers with detailed information

A simple way to apply tolls for social coordination would simply be to charge all users of each link a fixed price. Tolls of this form are known as "fixed" tolls, since the tolling function on each edge is a constant function of edge flow. To see fixed tolls in action, consider again Pigou's Example in Figure 2. Since the upper link is over-congested, it seems natural that charging the proper toll on that link would influence some users to deviate to the lower link. If all users have a tax-sensitivity equal to 1, one set of edge tolls that enforces the optimal flow for Pigou's example is simply $\tau_1(f_1) = 0.5$, and $\tau_2(f_2) = 0$. Under these tolls, the optimal flow of $(1/2, 1/2)$ is a Nash flow, since in it all users experience a cost of 1. If users are heterogeneous, an optimal fixed toll can still easily be found by charging the price that would cause the most sensitive half of the users to deviate to the lower link.

Similarly, in Braess's Paradox (see Figure 3), one way to enforce optimal flows for unit-sensitivity homogeneous users is to charge a toll of 1 on the center zero-cost link. This effectively removes the pathological link from the network, returning the network to its original optimal topology as in Figure 3(a).

This fixed-tolling approach has been studied in general, and it is known that fixed tolls can be computed to enforce *any* feasible flow, provided that the system planner has a complete characterization of the system: network topology, user demand profile, latency functions, and user sensitivities [22], [23].

Theorem 1 (Fleischer et al., 2004 [22]; Karakostas et al., 2004 [23]): For any routing game, fixed tolls can be computed which incentivize any feasible flow.

In Figure 5(a), this fixed-toll approach is depicted as a block diagram; note in particular that tolls are computed by a single, centralized optimization problem that takes as inputs all system variables and outputs a list of edge tolls.

Marginal-cost tolls: strongly robust to network variations

To motivate the concept of marginal-cost tolls, note that in traffic routing, an agent's total cost can be viewed as being two-fold: the first component is the agent's own experienced delay,

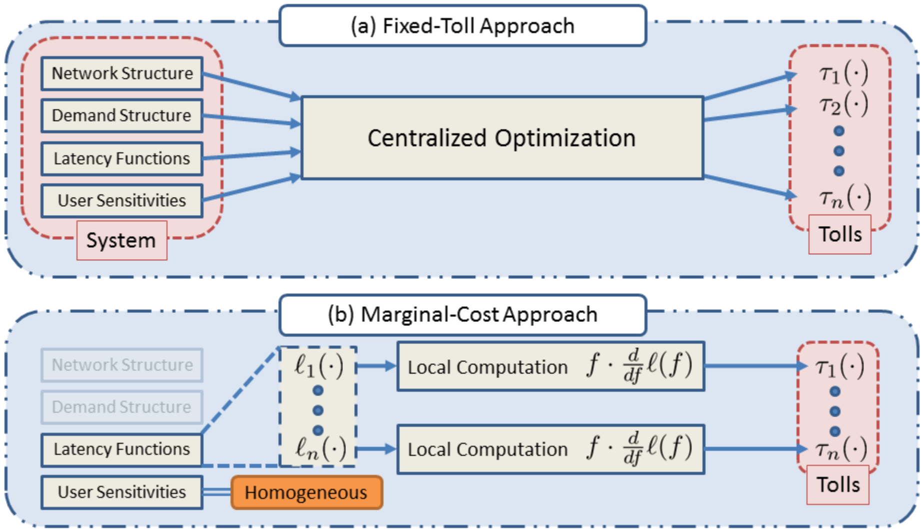

Figure 5. Depiction of informational dependencies of various taxation mechanisms. Subfigure (a) depicts the optimization strategy and informational dependencies of the fixed-toll mechanisms of [22], [23], in which a network's optimal fixed tolls are calculated as the output of a single centralized optimization problem that takes a complete system specification as input. In contrast, (b) is a corresponding diagram for the marginal-cost toll approach of [24], [25]. First, note that network structure and user demands are not required for the computation of optimal marginal-cost tolls. Second, note that the tolling function for any given edge e is computed locally – $\tau_e(\cdot)$ depends only on edge e's latency function $\ell_e(\cdot)$ and traffic flow f_e. This implies that the efficiency guarantees provided by marginal-cost tolls are robust to variations in network and demand structure simply by merit of their functional form.

the second is the delay that the agent's presence imposes on others. A marginal-cost toll explicitly charges each agent for his imposition on other agents; in economic language, marginal-cost tolls internalize the agent's negative externalities [24], [25]. The marginal-cost taxation mechanism $\boldsymbol{\tau}^{\mathrm{mc}}$ assigns tolls to each edge e given by

$$\tau_e^{\mathrm{mc}}(f_e) = f_e \cdot \frac{d}{df_e}\ell_e(f_e). \tag{9}$$

Note that each edge's toll depends only on that local edge's congestion properties and traffic flow; global information regarding traffic rate and network topology is not used.

It is well-known that for homogeneous user populations, charging marginal-cost tolls enforces exactly-optimal network flows [24], [25]:

Theorem 2 (Beckmann, McGuire, Winsten, 1956 [24]): For unit-sensitivity homogeneous populations, the marginal-cost taxation mechanism $\boldsymbol{\tau}^{\mathrm{mc}}$ incentivizes optimal flows on all networks:

$$\mathcal{L}^{\mathrm{nf}}(G, \boldsymbol{\tau}^{\mathrm{mc}}(G)) = \mathcal{L}^{*}(G). \tag{10}$$

Since these tolls can be implemented with no global information, by construction they are strongly robust to variations of network topology, user demand structure, and overall traffic rate.

This can be seen for Pigou's and Braess's networks in Figures 2 and 3, where marginal-cost tolls prescribe tolling functions of $\tau(f) = f$ for each of the linear-cost edges. For homogeneous users, the linear-cost edges have a resulting effective cost of $2f$, which incentivizes the desired optimal flows. See Figure 5(b) for a depiction of the informational dependencies and localized optimization structure of marginal-cost tolls in comparison with that of the fixed-toll methods of [22], [23].

Large universal tolls: instance-agnostic and weakly robust

Recent interest in robust social influence has prompted the development of more-sophisticated taxation mechanisms which confer robustness guarantees by design. One of these is investigated for heterogeneous users in [5], where the authors prove the following theorem:

Theorem 3 (Brown and Marden, 2015 [5]): Consider any routing problem G. If for some $S_{\mathrm{L}} > 0$, the heterogeneous tax-sensitivities of every agent $x \in [0, r]$ satisfy $s_x \geq S_{\mathrm{L}}$, and the tolling function on each edge is given by

$$\tau_e^{\mathrm{u}}(f_e) = \kappa \left(\ell_e(f_e) + f_e \cdot \frac{d}{df_e} \ell_e(f_e) \right), \tag{11}$$

then it will be true that

$$\lim_{\kappa \to \infty} \mathcal{L}^{\mathrm{nf}}\left(G, \boldsymbol{\tau}^{\mathrm{u}}(G)\right) = \mathcal{L}^{*}(G). \tag{12}$$

Note that the tolling functions in (11) are network-agnostic in that they do not depend on any instance-specific information other than latency functions; tolling functions can be designed with no knowledge of network topology, demand profile or distribution of user sensitivities, and computed locally at each edge using only local information. Thus, this shows that there is no network so pathological that nothing can be done about it with instance-agnostic tolls. In the language of robustness, (12) shows that these Universal Tolls are at least weakly robust to heterogeneous sensitivities for large enough κ, and in the large-κ limit, they get arbitrarily "close" to strong robustness.

Alternative Mechanisms

A wide variety of other mechanisms have been studied for purposes of robust social influence. One such mechanism is that of the Stackelberg routing mechanism [39]. In this setting, the game progresses sequentially: first, the system planner directly controls some fraction of traffic and assigns it to feasible paths; second, the remaining fraction of traffic routes selfishly. This mechanism can be viewed as a best-case scenario when it comes to influencing social behavior, and in some cases the price of anarchy guaranteed by Stackelberg routing can be shown to be a lower bound for the price of anarchy of a taxation mechanism [30].

Other possibilities include augmenting the universal tolls of 3 with an ability to communicate limited information between the links. In many control applications, allowing limited communication between controllers can greatly improve system performance [40]; it is possible that allowing limited information sharing between different link taxation functions could help compensate for poor system characterizations.

It is also possible to consider the budget-balanced case in which the system planner collects no net revenue, rather simultaneously returning the tolls collected on some links as subsidies on other links. This is studied in [41]; it is possible that a setting like this could yield some intriguing robustness benefits. For more information, see "Budget-Balanced Tolls."

The robustness of bounded tolls

Theorem 3 proves that near-optimality can be achieved with large tolls, but it gives no indication as to how high tolls must be. There may be settings where this high-toll approach would not be possible for technical or political reasons, either of which could impose an implicit upper limit on allowable tolling functions. Accordingly, there has been research on the robustness of tolls which respect an upper bound for parallel networks with affine cost functions [5]. Other work on bounded tolls, not explicitly focusing on robustness, can be found in [42].

Price-discrimination and robustness

The effect of the universal tolls of Theorem 3 is to obscure the differences between users by charging high tolls which wash out the effect of the nominal latency functions. However, a conceptually simple way of accomplishing the same thing would be for the tolls to depend on the identity and sensitivity of the payer. If each user with sensitivity s_x were charged a marginal-cost toll (9) divided by s_x, every user would experience costs exactly equal to those induced for a unit-sensitivity homogeneous population by marginal-cost tolls, which would incentivize optimal flows. This type of toll applies a principle known by economists as *perfect price-discrimination*, sometimes also called 1st-degree price discrimination. Perfect price-discrimination sidesteps the issue of individual price-sensitivities by charging each agent precisely the "correct" price for that agent. Of course, putting such a scheme into practice would be a formidable task, since it would require not only that the tax-designer *be able* to charge each agent an individualized price, but furthermore *know* the correct price for each agent. In view of this challenge, the authors of [43] consider the possibility of coarse price-discrimination, and suggest that such a scheme could indeed yield significant efficiency gains.

Centralized Adaptive Tolls for Unknown Latency Functions

The foregoing discussion has assumed that the system-planner knows the edge latency functions perfectly, regardless of other uncertainties. What if the opposite were true, and a system-planner were given a network and demand profile, but did not have any information regarding the latency functions? This situation is considered in [44]. The authors assume that the population is homogeneous, with unit-sensitivity to tolls, that the demand profile is constant, and that the system-planner has a desired target flow that he wishes to enforce. They develop an iterative algorithm whereby a system planner levies tolls on the network, records the resulting Nash flow, updates the tolls accordingly, and so on – and it is proved that Nash flows resulting from this sequence of tolls converge to the planner's target flow. Interesting robustness questions in this context could be posed in a dynamic framework: does this algorithm converge quickly enough to ensure efficient behavior if the underlying system is changing rapidly?

Contributions: The Non-Robustness of Fixed and Marginal-Cost Tolls

A Study on the Robustness of Fixed Tolls

As discussed above, fixed tolls can enforce any feasible flow if chosen properly. How robust are they to variations in a system's underlying parameters? Here, we investigate the robustness of fixed tolls to variations of three parameters: latency functions, overall traffic rate, and general network changes. All of this analysis of fixed tolls will be in the simplified context of homogeneous unit-sensitivity populations; thus, any negative results here can only be worsened by extending to the broader heterogeneous case.

A brief summary of these results is as follows: Proposition 4 shows that unsurprisingly, fixed tolls cannot be strongly robust to variations of latency functions. That is, if a network's latency functions are unknown, it is not possible to guarantee perfectly-optimal flows. However, Proposition 4 leaves open the possibility of weak robustness along this dimension.

Second, we exhibit a simple setting in which a particular fixed toll incentivizes arbitrarily-inefficient flows when the traffic rate has been mischaracterized. Unlike Proposition 4, this second example does not immediately imply any general conclusions; it is possible that some nonzero fixed tolls will be found which are at least weakly-robust to rate variations. Nonetheless, it represents a cautionary tale, suggesting that much care must be taken with fixed tolls if the traffic rate is unknown.

Finally, Theorem 5 shows that fixed tolls must incorporate some information about overall network structure (including information about latency functions) if they are even to be weakly-robust. Here, the approach is to show that for any fixed-toll mechanism in which all edge taxes depend only on local edge properties, networks can be constructed in which the tolled total latency is strictly worse than the un-tolled latency. This means that to systematically avoid perverse incentives when applying fixed tolls to a routing problem, a system planner must allow the tolls to depend on global information.

Fixed Tolls Cannot Be Strongly Robust

To begin, we ask if fixed tolls can ever be strongly robust to changes in latency functions. One could imagine sudden changes to latency functions arising as a result of traffic accidents, weather, or natural disasters; strong robustness to these would imply that optimal performance would be incentivized regardless of the severity of the disturbance. We have the following easy fact:

Proposition 4: Fixed tolls are not strongly robust on $\mathcal{G}$ to changes in latency functions.

Put differently, to guarantee optimal routing in general, fixed tolls require detailed characterizations of a network's latency functions.

Proof: Here, the network topology (V, E), path set $\mathcal{P}$, and total traffic rate r are held constant while latency functions $\{\ell\}$ are varied. Note from (7) that to disprove strong robustness on general networks $\mathcal{G}$, it suffices to exhibit two networks G and G^* that differ only in their latency functions for which $\mathcal{L}^{\rm nf}(G, \boldsymbol{\tau}^{\rm ft}(G)) = \mathcal{L}^*(G)$ and $\mathcal{L}^{\rm nf}(G^*, \boldsymbol{\tau}^{\rm ft}(G)) > \mathcal{L}^*(G^*)$. To achieve this, consider a perturbed version of Pigou's network. In Pigou's network, a fixed toll of $1/2$ on the upper congestible link incentivized optimal routing. This fixed toll was computed for a network whose lower link latency function was given precisely by $\ell_2(f_2) = 1$; what if, instead of 1, the lower latency function was some unknown constant b? To be precise, let G_b represent Pigou's network with $\ell_2(f_2) = b$ so that G_1 represents the nominal Pigou network. To ascertain whether fixed tolls can be strongly robust on Pigou's network, consider the quantity $\mathcal{L}^{\rm nf}(G_b, \boldsymbol{\tau}^{\rm ft}(G_1))$ which represents the total latency on perturbed network G_b resulting from tolls computed for nominal G_1. If fixed tolls $\boldsymbol{\tau}^{\rm ft}$ are strongly robust, then for each b, it will be true that $\mathcal{L}^{\rm nf}(G_b, \boldsymbol{\tau}^{\rm ft}(G_1)) = \mathcal{L}^*(G_b)$. Thus, the robustness of $\boldsymbol{\tau}^{\rm ft}$ can be checked by varying b in the following price-of-anarchy-like expression:

$$\mathrm{PoA}(b) \triangleq \frac{\mathcal{L}^{\rm nf}\left(G_b, \boldsymbol{\tau}^{\rm ft}(G_1)\right)}{\mathcal{L}^*(G_b)}. \tag{13}$$

In Figure 6, $\mathrm{PoA}(b)$ is plotted as b varies between 0 and 1.5 for two sets of tolls: the solid curve corresponds to tolls computed for $b = 1$, and the dashed curve corresponds to tolls equal to 0. Note that the fixed toll only incentivizes perfectly efficient behavior exactly at $b = 1$ (that is, $\mathrm{PoA}(1) = 1$). For all other values, $\mathrm{PoA}(b) > 1$, which means that tolls lead to worse-than-optimal total latencies for $b \neq 1$, or $\mathcal{L}^{\rm nf}(G_b, \boldsymbol{\tau}^{\rm ft}(G_1)) > \mathcal{L}^*(G_b)$, showing that these tolls are not strongly robust. Note that here we only checked the single fixed toll $\tau_1 = 1/2$, but it is easy to show that on this network, this toll is equivalent in all respects to any set of tolls for which $\tau_1 - \tau_2 = 1/2$, so assuming that $\tau_2 = 0$ is without loss of generality. Put differently, the optimal fixed tolls for G_1 are essentially unique, so we have shown that there can exist no fixed-toll taxation mechanism on G (and by extension on $\mathcal{G}$) that is robust to variations of latency functions. ■

Fixed Tolls and Rate-Dependence

We next investigate the robustness of fixed tolls along a different dimension: overall traffic rate. Even if fixed tolls are not robust to latency function changes, perhaps they can be robust to variations of the traffic rate. To this end, we return to the Braess's Paradox network of Figure 3. Now, suppose that r, the total amount of traffic on the network, is not fixed at 1, but can take any value between 0 and 1. Let G_r represent the Braess's Paradox network with r units of traffic, so the canonical version is simply given by G_1. Let $\boldsymbol{\tau}^{\rm Braess}(G_1)$ simply be the toll design proposed earlier: a single fixed toll of 1 on the center zero-latency link. Note that unlike Pigou's example, this toll is far from unique – but since it incentivizes optimal flows on G_1, it is an adequate starting point.

As for Pigou, the robustness of $\boldsymbol{\tau}^{\rm Braess}$ can be checked by varying r in the following price-of-anarchy-like expression:

$$\mathrm{PoA}(r) \triangleq \frac{\mathcal{L}^{\rm nf}\left(G_r, \boldsymbol{\tau}^{\rm Braess}(G_1)\right)}{\mathcal{L}^*(G_r)}. \tag{14}$$

It has already been shown that for $r = 1$, the optimal flow has no traffic on the center zero-cost link, so the proposed fixed toll achieves the goal of enforcing optimal flows. On the other hand, if $r \leq 0.5$, the optimal flow is to send *all* the traffic on the center zero-cost link, so that no traffic uses the constant-latency links. Unfortunately, the fixed toll on the center link is

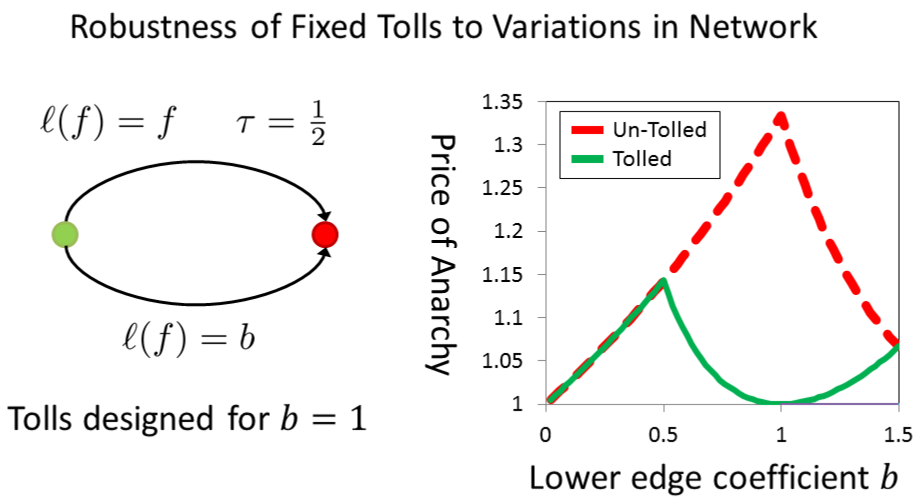

Figure 6. Pigou's Network: fixed tolls applied to the "wrong" network. Depicted here is an analysis of the robustness of fixed tolls to variations in latency functions. On the left, a Pigou-style network has a fixed toll of $1/2$ charged to the upper link; this toll incentivizes optimal behavior when the lower latency function satisfies $\ell(f) = 1$. To investigate the robustness of fixed tolls to network variations, the toll is held constant while the lower-link latency function b is allowed to vary between 0 and 1.5. For each value of b, the total latency of tolled and un-tolled Nash flows as well as the optimal total latency for that particular value of b are recorded. Finally, the price of anarchy curves are generated by dividing the Nash latencies by the respective optimal latency for each b. Note that the tolled price of anarchy is only 1 when $b = 1$; that is, this fixed toll only incentivizes optimal behavior on the specific network for which it was designed. The fact that the toll does not incentivize optimal behavior for all networks proves Proposition 4, stating that fixed tolls are not strongly robust to latency function variations.

still boldly incentivizing all users to avoid the now-optimal center link. In Figure 7, the price of anarchy from (14) is plotted as a function of r with and without the fixed toll on the center link. Note that the tolled curve in Figure 7 increases rapidly as r approaches 0, quickly driving the price of anarchy above the un-tolled maximum of $4/3$; by setting r low enough, the price of anarchy in this instance can be made arbitrarily high.

Here, despite the unbounded price of anarchy, this does *not* immediately imply that fixed tolls are not weakly robust to rate changes, merely that *this particular toll* is not weakly robust. This example does demonstrate that great care must be taken with fixed tolls when the total traffic rate is varying or unknown, because fixed tolls designed for one demand profile can cause arbitrarily poor performance under a different demand profile.

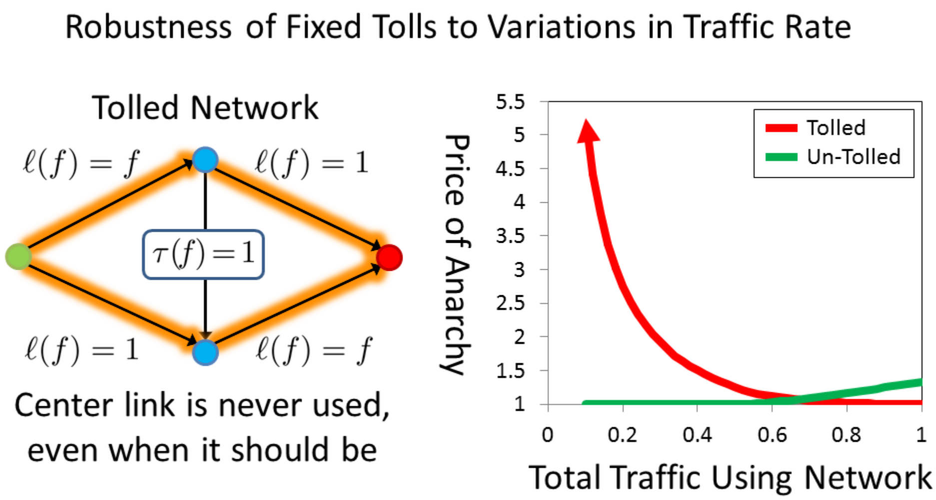

Figure 7. Braess' network: perverse incentives piled on top of unintended consequences. Following on the classical Braess's paradox (see Figure 3), this figure depicts an attempt to redeem the pathological network augmentation with a simple fixed toll. If a toll of 1 is levied on the link connecting the two intermediate nodes, it is simple to show that no traffic will ever use that link. Unfortunately, for low traffic rates (particularly when the total mass of traffic is less than 0.5), it is actually *optimal* for traffic to use the center link, but the rigid fixed toll prevents this. On the right is plotted the price of anarchy as a function of the total traffic rate; note that under the influence of tolls, the price of anarchy can become arbitrarily large as traffic approaches 0 – despite the fact that with no tolls, the price of anarchy can never exceed $4/3$.

Weakly Robust Fixed Tolls Must Depend on Network Structure

A fundamental problem with the fixed tolls applied to an uncertain Pigou network (as in Figure 6) was that the correct toll on the upper edge depended on the latency function of the lower edge; if the lower latency function was unknown, there was no way to compute an optimal toll on the upper edge, so fixed tolls could not be strongly robust. This prompts the question: could weakly robust fixed tolls be designed by letting the tolling function for edge e depend only on the local latency function ℓ_e? Such a taxation mechanism is called *network-agnostic*, prescribing tolling functions to edges without knowledge of how exactly the edges will be connected. Thus, a network-agnostic taxation mechanism $\boldsymbol{\tau}^{\mathrm{na}}$ is simply a mapping from latency functions to tolling functions so $\tau_e(f_e)$ (the tolling function on edge e) is given by $\boldsymbol{\tau}^{\mathrm{na}}(\ell_e)$. If a taxation mechanism is network-agnostic, then each edge toll depends only on local information, so any efficiency guarantees are automatically robust to changes in network structure.

Considering network-agnostic tolls, the robustness notion of "designing tolls for one network and applying them to another" must be slightly refined: as networks are varied, we assume that the tolling function τ_e has access to information *only* about the local latency function ℓ_e, but has no information about the location of e in the network or about the other latency functions in the network. One way to view this is that the toll-designer pre-commits to a tolling function for each possible latency function without specific knowledge of which latency functions will appear in the final realization. Thus, the notation of (8) can be simplified by writing $\mathcal{L}^{\rm nf}(G, \boldsymbol{\tau})$ to mean the total latency of a Nash flow on network G resulting from the tolls generated by taxation mechanism $\boldsymbol{\tau}$, and $\mathcal{L}^{\rm nf}(G, \emptyset)$ to mean the total latency of a Nash flow on G with no tolls. This leads to the main result about the lack of robustness of fixed tolls:

Theorem 5: The only nonnegative network-agnostic fixed tolls that are weakly robust to network variations satisfy

$$\tau_e = 0 \tag{15}$$

for all possible network edges.

Theorem 5 shows that positive fixed tolls must in general require some global information about network structure (e.g., network topology or latency functions) in order to ensure that they do not cause harm, even for the simple setting of homogeneous populations. Theorem 5 is proved with a series of simple example networks. This shows that even on simple networks, fixed tolls lack robustness, suggesting that complex networks could exhibit even more severe pathologies.

Proof: Let $\boldsymbol{\tau}^{\rm naft}$ be a network-agnostic taxation mechanism such that for any network G, $\mathcal{L}^{\rm nf}(G, \boldsymbol{\tau}^{\rm naft}) \leq \mathcal{L}^{\rm nf}(G, \emptyset)$. Write the toll assigned to an edge with latency function ℓ as $\boldsymbol{\tau}^{\rm naft}(\ell)$.

First, consider the network shown in Figure 8(a) in which the latency functions satisfy $\ell_1 + \ell_2 = \ell_3$. The flow $(1/2, 1/2)$ is both a Nash flow and an optimal flow; the only tolls which will always support this must satisfy $\boldsymbol{\tau}^{\rm naft}(\ell_1) + \boldsymbol{\tau}^{\rm naft}(\ell_2) = \boldsymbol{\tau}^{\rm naft}(\ell_3)$. This is the first condition on $\boldsymbol{\tau}^{\rm naft}$:

$$\ell_1 + \ell_2 = \ell_3 \quad \Longrightarrow \quad \boldsymbol{\tau}^{\rm naft}(\ell_1) + \boldsymbol{\tau}^{\rm naft}(\ell_2) = \boldsymbol{\tau}^{\rm naft}(\ell_3). \tag{16}$$

Next, consider the network shown in Figure 8(b), a two-link parallel network with degree-d monomial cost functions $\ell_1(f_1) = \alpha(f_1)^d$ and $\ell_2(f_2) = \beta(f_2)^d$. It can be shown that the optimal flow on this network is equal to the untolled Nash flow for any $\alpha > 0, \beta > 0$, and $d \geq 1$. Thus, the tolls on each link must be equal; otherwise, the tolled flow will have a strictly higher total latency than the un-tolled flow. That is, all monomials of the same degree must be charged the same toll, regardless of the scale of the latency function: $\boldsymbol{\tau}^{\rm naft}\left(\alpha f^d\right) = \boldsymbol{\tau}^{\rm naft}\left(\beta f^d\right)$. In

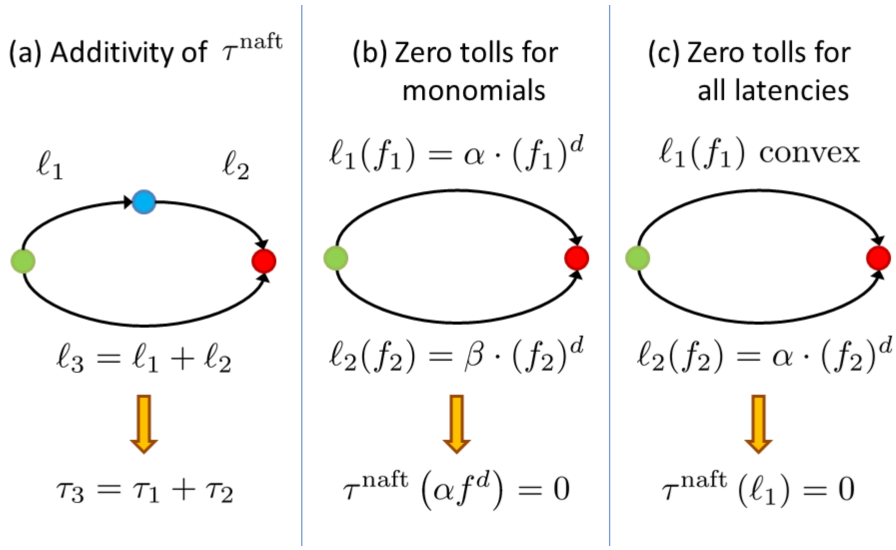

Figure 8. Networks used to prove Theorem 5, showing that any network-agnostic fixed-toll taxation mechanism must charge taxes of 0 on every link. The network in (a) is used to show that network-agnostic fixed-toll taxation mechanism $\boldsymbol{\tau}^{\text{naft}}$ is additive in latency functions. Next, (b) is used to show in conjunction with (a) that monomial cost functions of any degree must be assigned a zero toll. Finally, in (c), for any general latency function ℓ_1, it is shown how to create a polynomial latency function ℓ_2 which is more congestible at equilibrium than ℓ_1, so that a positive toll charged on ℓ_1 would increase the total latency of the network. Thus it is shown that network-agnostic fixed tolls must charge 0 on every edge or they risk increasing the total latency compared to un-tolled networks.

particular, letting $\beta = 2\alpha$ and appealing to (16), it holds that that $2\boldsymbol{\tau}^{\text{naft}}\left(\alpha f^d\right) = \boldsymbol{\tau}^{\text{naft}}\left(2\alpha f^d\right)$, which implies the second condition on $\boldsymbol{\tau}^{\text{naft}}$:

$$\text{for all } \alpha > 0 \text{ and } d \geq 1, \quad \boldsymbol{\tau}^{\text{naft}}\left(\alpha f^d\right) = 0. \tag{17}$$

Using this fact regarding polynomials, positive fixed tolls on any other latency function can now be ruled out. Refer to Figure 8(c), another two-link parallel network. Given an arbitrary convex latency function ℓ (with flow derivative ℓ') on the upper link, it is possible to design a polynomial latency function for the lower link $\ell_2(f_2) = \alpha(f_2)^d$ to show that $\boldsymbol{\tau}^{\text{naft}}(\ell) = 0$. Let the polynomial degree satisfy $d > \frac{\ell(1/2)}{2\ell'(1/2)}$, and coefficient satisfy $\alpha = 2^d \ell(1/2)$. Then by design, the un-tolled Nash flow on the network is $(1/2, 1/2)$, and at this flow, shifting any positive mass of traffic from the upper link to the lower link strictly increases the total latency on the network. To avoid this, the toll on the upper link must be zero: $\boldsymbol{\tau}^{\text{naft}}(\ell) = 0$. Since ℓ is an arbitrary convex latency function, the theorem is proved. ■

Theorem 5 (regarding the lack of weak robustness of fixed tolls) carefully specified that the fixed tolls in question be nonnegative – and this prompts the question of negative tolls, i.e., subsidies. It turns out that for the special case of linear-latency networks, strongly robust network-agnostic fixed *subsidies* do exist. For a linear latency function of the form $\ell_e(f_e) = a_e f_e + b_e$, the corresponding network-agnostic fixed subsidy (represented as a negative toll) is given by

$$\tau_e^{\mathrm{subsidy}} = -\frac{b_e}{2}. \tag{18}$$

By simply paying users half the constant-term cost on each link, optimal flows can be incentivized as Nash flows. As a side note, these subsidies are a special case of the "variable price schemes" of [25] with $\bar{\eta} = 1$.

To show the optimality of these subsidies, the cost functions resulting from these subsidies can be related to the cost functions resulting from standard marginal-cost tolls τ^{mc} (see (9)), given in this case by $\tau_e^{\mathrm{mc}} = a_e f_e$. Recall that for homogeneous users, marginal-cost tolls are known to induce optimal Nash flows. Under the homogeneous-user model, the cost functions resulting from the network agnostic fixed subsidies are

$$J_e^{\mathrm{subsidy}}(f_e) = \underbrace{a_e f_e + b_e}_{\ell_e} - \underbrace{\frac{b_e}{2}}_{\tau_e} = a_e f_e + \frac{b_e}{2}, \tag{19}$$

while the cost functions resulting from marginal-cost tolls are

$$J_e^{\mathrm{mc}}(f_e) = 2a_e f_e + b_e. \tag{20}$$

Since these two cost functions are related by a constant multiplicative factor for all agents (i.e., $J^{\mathrm{mc}} = 2J^{\mathrm{subsidy}}$), they induce the same optimal Nash flows, and these subsidies inherit the strong robustness of marginal-cost tolls. While these strongly robust network-agnostic fixed subsidies are theoretically appealing, it is not clear that the concept generalizes beyond linear cost functions or homogeneous users.

Are marginal-cost tolls robust for heterogeneous users?

The strong robustness guarantees of marginal-cost tolls have been proved by [24], [25] in the setting of homogeneous known-sensitivity users; do these guarantees carry over to the more detailed heterogeneous model? As a first step towards studying robust taxation mechanisms for heterogeneous users, the performance of "off-the-shelf" marginal-cost tolls is investigated on simple price-sensitive settings.

In Pigou's network, the marginal-cost taxation mechanism $\boldsymbol{\tau}^{\mathrm{mc}}$ assigns a flow-varying toll of $\tau_1^{\mathrm{mc}}(f_1) = f_1$ to the upper link; for homogeneous users, this incentivizes optimal routing. What if the users' sensitivities remained homogeneous, but took on some unknown value other than 1? We can employ a parallel argument to that seen for fixed tolls in Proposition 4: Write G_s to denote Pigou's network in which all users have sensitivity s. To ascertain whether marginal-cost tolls can be strongly robust to sensitivity variations, consider the quantity $\mathcal{L}^{\rm nf}(G_s, \boldsymbol{\tau}^{\rm mc}(G_1))$, which represents the total latency on the perturbed population in G_s resulting from marginal-cost tolls computed for unit-sensitivity G_1. If marginal-cost tolls $\tau^{\rm mc}(G_1)$ are strongly robust, then for each s, it will be true that $\mathcal{L}^{\rm nf}(G_s, \boldsymbol{\tau}^{\rm mc}(G_1)) = \mathcal{L}^*(G_s)$. Since the optimal flow on a network does not depend on the user sensitivities, this is simply equivalent to writing $\mathcal{L}^{\rm nf}(G_s, \boldsymbol{\tau}^{\rm mc}(G_1)) = 0.75$.

In Figure 9, $\mathcal{L}^{\rm nf}(G_s, \boldsymbol{\tau}^{\rm mc}(G_1))$ is plotted as s varies between 0 and 2. Note that the marginal-cost toll only incentivizes perfectly efficient behavior exactly at $s = 1$. For all other values, the total latency is strictly greater than the optimal 0.75, showing that these tolls are not strongly robust. This implies that when considering user price-sensitivity, marginal-cost tolls are not strongly robust to sensitivity variations, even in the simplified setting of homogeneous users on Pigou's network.

Perverse Marginal-Cost Tolls For Heterogeneous Users

Here, we ask a similar question of marginal-cost tolls in the heterogeneous model to that asked for fixed tolls in the homogeneous model: Despite lacking strong robustness to user sensitivities, is it at least possible to show that marginal-cost tolls are weakly robust? Unfortunately, in general settings, it is possible to show that even marginal-cost tolls can incentivize flows that are strictly worse than their un-tolled counterparts, thus lacking even weak robustness to heterogeneity. Consider the network and demand profile in Figure 10. This is essentially a three-link network; a population of mass 0.5 has access only to the upper two links, and a population of mass 1 has access only to the lower two links. The optimal flow has all users from the upper population using the center (congestible) link and all users from the lower population using the lower link. In the unique un-tolled Nash flow, half the traffic from the lower link has shifted to the center link so as to equalize the latencies of those two links.

Note that the un-tolled version bears a strong resemblance to Pigou's example if the upper link is ignored: self-interested users from the lower source have over-congested the center link and degraded its performance. To see how tolls can make things even worse, suppose that the upper population is toll-sensitive (with sensitivities equal to 1), but the lower population is not (i.e., they have sensitivities close to 0). The insensitivity of the lower population effectively fixes the flow on the center link at 1, since this is the flow that equalizes the latencies of the lower two links. That is, regardless of what the upper population does, the center link flow will always be 1. Thus, marginal-cost tolls on this network serve only to force the upper population to choose

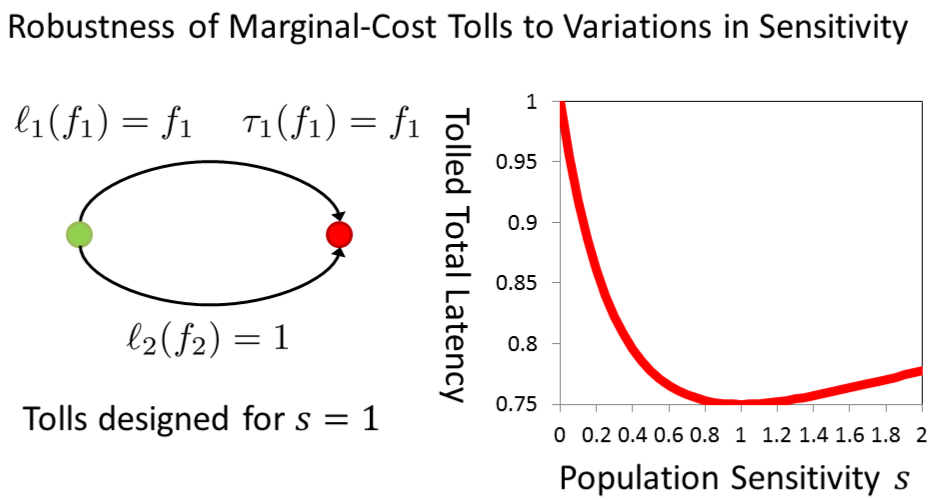

Figure 9. A setting in which marginal-cost tolls are shown not to be strongly robust to variations of user sensitivity. Depicted is the canonical Pigou Network with marginal-cost toll $\tau^{\mathrm{mc}}(f) = f$ assigned to the top link. If marginal-cost tolls were strongly robust to variations of user sensitivities, they would incentivize optimal flows for every user sensitivity profile. To check this, tolls are chosen without *a priori* knowledge of the user sensitivities, and then the population's homogeneous sensitivity is swept from 0 to 2 and the resulting total latency is plotted. It is evident from the plot that the optimal total latency is only obtained when the user sensitivities are exactly 1; all other sensitivity values incentivize some suboptimal total latency $\mathcal{L}(f) > 0.75$. Thus, marginal-cost tolls are not strongly robust to variations of user sensitivity.

the upper link, resulting in the flow depicted on the right in Figure 10. This pathological flow has a total latency of 1.75, which corresponds to a price of anarchy of 1.4. This is greater than the $4/3$ guaranteed as a worst-case on linear-latency networks, demonstrating that marginal-cost tolls lack even weak robustness for heterogeneous populations. In this case, instead of incentivizing altruism, marginal-cost tolls simply amplified the selfishness-induced inefficiency that was already present. This allows us to state the following fact:

Proposition 6: Marginal-cost tolls (9) are not weakly-robust on $\mathcal{G}$ to variations of user sensitivities.

Proof: Let G denote the routing game in Figure 10. Per the above analysis, $\boldsymbol{\tau}^{\mathrm{mc}}$ induces a Nash total latency of $\mathcal{L}^{\mathrm{nf}}(G, \boldsymbol{\tau}^{\mathrm{mc}}(G)) = 1.75$, but the un-tolled total latency is $\mathcal{L}^{\mathrm{nf}}(G, \emptyset) = 1.5$; since the tolled total latency is strictly worse, $\boldsymbol{\tau}^{\mathrm{mc}}$ cannot be weakly robust to heterogeneous user sensitivities. ■

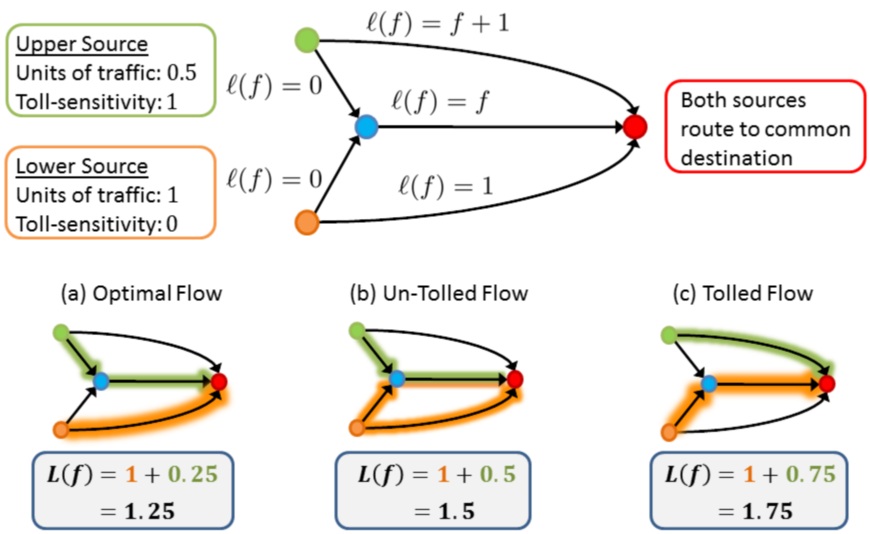

Figure 10. A network demonstrating that marginal-cost tolls are not weakly robust to user heterogeneity. This figure depicts a simple two-source network in which 0.5 units of traffic route from the upper (green) source, and 1 unit of traffic routes from the lower (orange) source. If traffic from the upper source trades off time and money equally (i.e., $s \equiv 1$), but traffic from the lower source cares only about time (i.e., $s \equiv 0$), marginal-cost tolls result in a price of anarchy of 1.4 on this network. The optimal flow here requires all of the traffic from the lower source to use the lower, constant-latency link. However, only the traffic from the upper source responds to tolls; when marginal-cost tolls are levied, all of the upper-source (green traffic) moves to the inefficient upper path, and the lower-source (orange) traffic moves to replace it on the middle path, as depicted on the right.

As a side note, the most extreme pathology in this example arises when the lower population has a sensitivity of 0, but perversities still arise for any low, positive sensitivity. That is, the poor performance in this example can still occur when every agent has a nonzero price-sensitivity.

Conclusion

Using Pigou's example, it was shown in Proposition 4 that fixed tolls can never be strongly robust to network variations; using Braess's Paradox, it was suggested that fixed tolls are not robust to rate variations either. Subsequently, Theorem 5 showed that fixed tolls cannot even be weakly robust if they do not depend on global network information. Finally, Proposition 6 showed on the new example depicted in Figure 10 that marginal-cost tolls cannot even be weakly robust to heterogeneous price sensitivities in general. All of these results were shown using very simple networks; increasing the complexity of networks would likely only degrade worst-case performance.

There are many open questions in this domain. For instance, marginal-cost tolls do not require knowledge of the other latency functions in the network, but (9) makes it clear that the tolling function on edge e strongly depends on the specific latency function ℓ_e. It remains an open question how mischaracterizations of local latency functions impact the congestion-minimizing guarantees of marginal-cost tolls. One possible approach for studying this could be to restrict attention to linear latency functions of the form $\ell_e(f_e) = a_e f_e + b_e$, and assume that the true coefficients a_e and b_e were only known to exist in some range $[\bar{a} - \delta, \bar{a} + \delta]$ and $[\bar{b} - \delta, \bar{b} + \delta]$ for some nominal $\bar{a}$ and $\bar{b}$. It seems reasonable to conjecture that if $\delta > 0$, marginal-cost tolls cease to be strongly robust, but it would be interesting to ask in what settings they remain weakly robust for homogeneous users.

The examples and results in this article demonstrate some of the diverse challenges in social coordination in uncertain environments, and they may leave the reader with the unsettling question "can we guarantee *anything* without knowing *everything?*" As the previous examples have shown, the answers to this question are complex and nuanced; ultimately, what is needed is a characterization of the fundamental relationship between the amount of information a taxation methodology requires, the sophistication of the mechanism, and the efficiency guarantees it can provide under uncertainty. Fixed tolls assign relatively "unsophisticated" tolling functions (being merely constant functions of flow), but appear to require a great deal of information to enforce optimal flows. Thus, they apparently sacrifice robustness for simplicity. Marginal-cost tolls enforce optimal flows with a greater degree of robustness, doing so by allowing more-sophisticated flow-varying tolling functions. However, Proposition 6 shows that they are not so sophisticated that they can prevent all pathologies from arising when agents are heterogeneous.

Throughout this article, the exclusive focus has been on static equilibrium analysis from a worst-case perspective. Robustness as defined in (7) and (8) is essentially a worst-case metric; if a single pair of routing problems exhibit a pathology in response to some taxation mechanism, that is sufficient to disprove robustness of the taxation mechanism. However, a similar analysis could be done from a stochastic perspective: a corresponding notion of robustness could focus on avoiding pathologies in expectation over random networks drawn from some appropriate distribution.

Research on robust social influence has conceptual applications in many areas beyond routing games; any setting in which social systems are enmeshed with engineered systems seems fertile ground for similar analysis to that surveyed in this article.

References

[1] J. Wardrop, "Some Theoretical Aspects of Road Traffic Research," in *Proc. Inst. Civ. Eng. Part II, Vol. 1, No. 36*, pp. 352–362, 1952.

[2] L. Pavel, *Game Theory for Control of Optical Networks*. Springer Science & Business Media, 2012.

[3] J. R. Marden and J. S. Shamma, "Game Theory and Distributed Control," in *Handb. Game Theory Vol. 4* (H. Young and S. Zamir, eds.), Elsevier Science, 2014.

[4] D. Braess, A. Nagurney, and T. Wakolbinger, "On a Paradox of Traffic Planning," *Transp. Sci.*, vol. 39, pp. 446–450, nov 2005.

[5] P. N. Brown and J. R. Marden, "Optimal Mechanisms for Robust Coordination in Congestion Games," in *54th IEEE Conf. Decis. Control*, (Osaka, Japan), pp. 2283–2288, 2015.

[6] A. Kleiner, B. Nebel, and V. A. Ziparo, "A mechanism for dynamic ride sharing based on parallel auctions," in *Int. Jt. Conf. Artif. Intell.*, pp. 266–272, 2011.

[7] S. Banerjee, R. Johari, and C. Riquelme, "Pricing in Ride-Sharing Platforms : A Queueing-Theoretic Approach," in *16th ACM Conf. Econ. Comput.*, pp. 639–639, 2015.

[8] M. Esmaeili, M.-B. Aryanezhad, and P. Zeephongsekul, "A game theory approach in sellerbuyer supply chain," *Eur. J. Oper. Res.*, vol. 195, no. 2, pp. 442–448, 2009.

[9] K. Tsakalozos, H. Kllapi, E. Sitaridi, M. Roussopoulos, D. Paparas, and A. Delis, "Flexible use of cloud resources through profit maximization and price discrimination," in *Proc. Int. Conf. Data Eng.*, pp. 75–86, 2011.

[10] C. Silva, "Application of Mechanism Design to Electric Power Markets (Republished)," *IEEE Trans. Power Syst.*, vol. 16, no. 4, pp. 862–869, 2001.

[11] M. Gairing, "Covering games: Approximation through non-cooperation," in *Internet Econ.*, pp. 184–195, 2009.

[12] G. Como, K. Savla, D. Acemoglu, and M. A. Dahleh, "Robust Distributed Routing in Dynamical Networks Part I : Locally Responsive Policies and Weak Resilience," *IEEE Trans. Automat. Contr.*, vol. 58, no. 2, pp. 317–332, 2013.

[13] G. Como, K. Savla, D. Acemoglu, and M. A. Dahleh, "Robust Distributed Routing in Dynamical Networks Part II : Strong Resilience , Equilibrium Selection and Cascaded Failures," *IEEE Trans. Automat. Contr.*, vol. 58, no. 2, pp. 333–348, 2013.

[14] G. Como, E. Lovisari, and K. Savla, "Throughput Optimality and Overload Behavior of Dynamical Flow Networks Under Monotone Distributed Routing," *IEEE Trans. Control Netw. Syst.*, vol. 2, no. 1, pp. 57–67, 2015.

[15] A. C. Pigou, *The Economics of Welfare*. Macmillan, New York, 1920.

[16] M. J. Smith, "The Marginal Cost Taxation of a Transportation Network," *Transp. Res. Part B Methodol.*, vol. 13, no. 3, pp. 237–242, 1979.

[17] S. Morrison, "A Survey of Road Pricing," *Transp. Res. Part A Gen.*, no. 2, pp. 87–97, 1986.

[18] R. B. Dial, "Network-Optimized Road Pricing: Part I: A Parable and a Model," *Oper. Res.*, vol. 47, pp. 54–64, jan 1999.

[19] R. Cole, Y. Dodis, and T. Roughgarden, "How Much Can Taxes Help Selfish Routing?," *J. Comput. Syst. Sci.*, vol. 72, pp. 444–467, may 2006.

[20] L. Engelson and P. Lindberg, "Congestion pricing of road networks with users having different time values," *Appl. Optim.*, vol. 101, pp. 81–104, 2006.

[21] G. Christodoulou, K. Mehlhorn, and E. Pyrga, "Improving the price of Anarchy for Selfish Routing via coordination mechanisms," *Algorithmica*, vol. 69, no. 3, pp. 619–640, 2014.

[22] L. Fleischer, K. Jain, and M. Mahdian, "Tolls for Heterogeneous Selfish Users in Multicommodity Networks and Generalized Congestion Games," in *Proc. 45th IEEE Symp. Found. Comput. Sci.*, (Rome, Italy), pp. 277–285, 2004.

[23] G. Karakostas and S. Kolliopoulos, "Edge pricing of multicommodity networks for heterogeneous selfish users," in *Proc. 45th IEEE Symp. Found. Comput. Sci.*, (Rome, Italy), pp. 268–276, 2004.

[24] M. Beckmann, C. McGuire, and C. B. Winsten, "Studies in the Economics of Transportation," 1956.

[25] W. Sandholm, "Evolutionary Implementation and Congestion Pricing," *Rev. Econ. Stud.*, vol. 69, no. 3, pp. 667–689, 2002.

[26] P. N. Brown and J. R. Marden, "The Robustness of Marginal-Cost Taxes in Affine Congestion Games," *Trans. Autom. Control*, to be published.

[27] T. Roughgarden, *Selfish Routing and the Price of Anarchy*. MIT Press, 2005.

[28] W. Sandholm, "Negative Externalities and Evolutionary Implementation," *Rev. Econ. Stud.*, vol. 72, pp. 885–915, jul 2005.

[29] R. Arnott, A. de Palma, and R. Lindsey, "Does providing information to drivers reduce traffic congestion?," *Transp. Res. Part A Gen.*, vol. 25, no. 5, pp. 309–318, 1991.

[30] V. Bonifaci, T. Harks, and G. Shaefer, "The Impact of Stackelberg Routing in General Networks," *Math. Oper. Res.*, vol. 32, no. 2, pp. 1–17, 2010.

[31] H.-L. Chen, T. Roughgarden, and G. Valiant, "Designing Networks with Good Equilibria," in *Proc. 19th ACM SIAM Symp. Discret. Algorithms*, pp. 854–863, 2008.

[32] C. Papadimitriou, "Algorithms, Games, and the Internet," in *Proc. 28th Int. Colloq. Autom. Lang. Program.*, 2001.

[33] T. Roughgarden, "How Bad Is Selfish Routing?," *J. ACM*, vol. 49, no. 2, pp. 236–259, 2002.

[34] T. Roughgarden, "The price of anarchy is independent of the network topology," *J. Comput. Syst. Sci.*, vol. 67, pp. 341–364, sep 2003.

[35] R. Johari and J. N. Tsitsiklis, "Efficiency Loss in a Network Resource Allocation Game," *Math. Oper. Res.*, vol. 29, pp. 407–435, aug 2004.

[36] G. Piliouras, E. Nikolova, and J. S. Shamma, "Risk sensitivity of price of anarchy under uncertainty," in *Proc. 14th ACM Conf. Electron. Commer.*, vol. 9, (New York, New York, USA), pp. 715–732, 2013.

[37] P.-a. Chen, B. De Keijzer, D. Kempe, and G. Shaefer, "Altruism and Its Impact on the Price of Anarchy," *ACM Trans. Econ. Comput.*, vol. 2, no. 4, 2014.

[38] R. Cole, Y. Dodis, and T. Roughgarden, "Pricing network edges for heterogeneous selfish users," in *Proc. 35th ACM symp. Theory Comput.*, (New York, New York, USA), pp. 521–530, 2003.

[39] T. Roughgarden, "Stackelberg Scheduling Strategies," 2004.

[40] J. R. Marden, "The Role of Information in Multiagent Coordination," in *53rd IEEE Conf. Decis. Control.*, pp. 445 – 450, 2014.

[41] I. Arieli, "Transfer Implementation in Congestion Games," *Dyn. Games Appl.*, vol. 5, no. 2, pp. 228–238, 2015.

[42] V. Bonifaci, M. Salek, and G. Schäfer, "Efficiency of restricted tolls in non-atomic network routing games," in *Algorithmic Game Theory*, vol. 6982 LNCS, pp. 302–313, 2011.

[43] P. N. Brown and J. R. Marden, "A Study on Price-Discrimination for Robust Social Coordination," in *Am. Control Conf.*, (Boston, MA, USA), 2016.

[44] U. Bhaskar, K. Ligett, L. Schulman, and C. Swamy, "Achieving Target Equilibria in Network Routing Games without Knowing the Latency Functions," in *Proc. IEEE Symp. Found. Comput. Sci.*, (Philadelphia, PA, USA), pp. 31–40, 2014.

[45] G. Perakis and G. Roels, "The Price of Anarchy in Supply Chains: Quantifying the Efficiency of Price-Only Contracts," *Manage. Sci.*, vol. 53, pp. 1249–1268, aug 2007.

[46] R. P. Leme and É. Tardos, "Pure and Bayes-Nash price of anarchy for generalized second price auction," in *Annu. IEEE Symp. Found. Comput. Sci. FOCS*, pp. 735–744, 2010.

[47] B. Yolken and N. Bambos, "Game based capacity allocation for utility computing environments," *Telecommun. Syst.*, vol. 47, pp. 165–181, may 2010.

[48] V. Syrgkanis and E. Tardos, "Composable and Efficient Mechanisms," in *Symp. Theory Comput.*, pp. 211–220, 2013.

[49] R. Hoeksma and M. Uetz, "The price of anarchy for minsum related machine scheduling," in *Approx. Online Algorithms*, pp. 261–273, Springer, 2011.

[50] T. Roughgarden and F. Schoppmann, "Local smoothness and the price of anarchy in splittable congestion games," *J. Econ. Theory*, vol. 1, pp. 1–26, apr 2014.

Author Information

Philip N. Brown is a PhD student in the Department of Electrical and Computer Engineering at the University of California, Santa Barbara. Philip received the BS with highest honor in Electrical Engineering in 2007 from Georgia Tech. Following this, Philip designed control systems and process technology for the biodiesel industry with SunsOil, LLC. He then received the MS in Electrical Engineering in 2015 from the University of Colorado at Boulder under the supervision of Jason R. Marden, where he received the University of Colorado Chancellor's Fellowship. Philip's research interests include characterizing and influencing user behavior in socio-technical systems.

Jason Marden is an Assistant Professor in the Department of Electrical and Computer Engineering at the University of California, Santa Barbara. Jason received the BS in Mechanical Engineering in 2001 from UCLA, and the PhD in Mechanical Engineering in 2007, also from UCLA, under the supervision of Jeff S. Shamma, where he was awarded the Outstanding Graduating PhD Student in Mechanical Engineering. After graduating from UCLA, he served as a junior fellow in the Social and Information Sciences Laboratory at the California Institute of Technology until 2010, and then as an Assistant Professor at the University of Colorado until 2015. Jason is a recipient of an ONR Young Investigator Award (2015), NSF Career Award (2014), AFOSR Young Investigator Award (2012), SIAM CST Best Sicon Paper Award (2015), and the American Automatic Control Council Donald P. Eckman Award (2012). Jason's research interests focus on game theoretic methods for the control of distributed multiagent systems.

Sidebar 1: Computing Price of Anarchy Bounds

The concept known as "price of anarchy," a measure of the inefficiency of selfish behavior, was first rigorously studied in the area of routing games in [33], and has found far-reaching applications in fields as diverse as supply-chain management [45], auction theory [46], telecommunication systems [47], and others. Price of anarchy is defined generally as follows: In a class of games $\mathcal{G}$, suppose a k-player game $G \in \mathcal{G}$ has outcomes denoted by x and some global cost function to be minimized $C(x) = \sum_{i=1}^{k} C_i(x)$. The price of anarchy is defined as

$$\mathrm{PoA}(\mathcal{G}) \triangleq \max_{G \in \mathcal{G}} \frac{C(x^{\mathrm{ne}})}{\min_{x} C(x)}, \tag{S1}$$

where x^{ne} is a Nash equilibrium. That is, price of anarchy is a worst-case measure over all games in $\mathcal{G}$ of the performance degradation due to selfish behavior.

There is no universal technique for computing exact worst-case equilibria for an arbitrary game, but there do exist methods which can simplify the process of upper-bounding the worst-case performance of equilibria. One such method is known as (λ, μ)-*smoothness*. The approach is as follows: Suppose it can be shown for games in $\mathcal{G}$ that the following is true for some $\lambda > 0$ and $\mu < 1$ and all outcomes x and x^*:

$$\sum_{i=1}^{k} C_i\left(x_i^*, x_{-i}\right) \leq \lambda C(x^*) + \mu C(x), \tag{S2}$$

where the notation (x_i^*, x_{-i}) indicates that player i is playing according to action profile x^*, and all players different from i are playing according to action profile x. In this case, the following bound holds on the price of anarchy of $\mathcal{G}$:

$$\mathrm{PoA}(\mathcal{G}) \leq \frac{\lambda}{1-\mu}. \tag{S3}$$

Thus, computing an upper bound on the price of anarchy reduces to finding the (λ, μ) parameters which make the inequality in (S3) as tight as possible. While this may seem a daunting task, it has proved useful in computing price of anarchy bounds in many settings, including mechanism design [48], machine scheduling [49], and a variety of network routing problems [50]. In the case of affine-cost routing games, it can be shown that $\lambda = 1$ and $\mu = 1/4$, implying a price of anarchy of $4/3$.

Sidebar 2: Elastic Traffic

The results in this article consider only the case of inelastic traffic; that is, all drivers have been required to arrive at their destination, regardless of cost. A more detailed model may consider the elastic demand case, assuming that each user has a cost threshold; if the total cost of travel exceeds this threshold, he will simply decide to "stay home." It turns out that marginal-cost tolls preserve their robustness characteristics in the elastic model.

Under an elastic demand model, the system-planner's goal is to maximize a linear combination of aggregate travel benefit minus congestion costs; this objective is called *social welfare* maximization. For simplicity, consider the case of parallel networks.

We model the user population as the interval $[0, r]$, and define a continuous nonincreasing value-of-travel function $v : [0, r] \to \mathbb{R}$ that assigns each user $x \in [0, r]$ a cost threshold. Assuming that v is nonincreasing is equivalent to assuming without loss of generality that the user population is sorted in decreasing order of value-of-travel. Given a flow f, the utility of user $x \in [0, r]$ on edge e is given by

$$U_e^f(x) = v(x) - (\ell_e(f_e) + \tau_e(f_e)) . \tag{S4}$$

If this quantity is negative for all edges in the network, user x will choose to stay home, as the total cost of commuting is higher than the user's value-of-travel. It is easily shown that in equilibrium, all users using the network have higher values than the users who stay home; write x^* to denote the lowest-value user who chooses to travel. The planner's objective, the social welfare W, is defined by the following:

$$W(x^*, f) = \int_0^{x^*} v(t)dt - \mathcal{L}(f), \tag{S5}$$

simply the difference between the total value obtained by users who choose to commute and the total latency of the resulting flow.

In [25], the author shows that in this elastic case, marginal-cost tolls maximize the social welfare for any valuation profile v without requiring that the system planner know anything about the specific valuation profile. In the language of robustness, a theorem from [25] can be rephrased as

Theorem 7 (Sandholm, 2002 [25]): Marginal-cost tolls are strongly robust to variations in user elasticity profiles v.

Thus, in addition to being robust to all the parameters previously discussed, marginal-cost toll are *also* robust to populations with unknown elasticity.

What if the user population is heterogeneous in price-sensitivity? To investigate the robustness of marginal-cost tolls to variations in elasticity for heterogeneous populations, we adopt the simple setting of Pigou's network (see Figure 2) and study the performance of a variety of scaled marginal-cost tolls for a specific family of value-of-travel functions. The value-of-travel functions in question satisfy

$$v(x) = \bar{v} - x. \tag{S6}$$

Tolls of κf_1 are applied to the congestible link for various values of $\kappa \in [0, 1]$. For each $\bar{v} \in [1, 2]$ and $\kappa \in [0, 1]$, the worst-case social welfare $W^{\max}(\bar{v}, \kappa)$ is found by searching over the space of (possibly heterogeneous) populations with price sensitivities between $S_{\mathrm{L}} = 0.1$ and $S_{\mathrm{U}} = 10$. This worst-case social welfare is then normalized by the optimal social welfare for the corresponding $\bar{v}$, and plotted in Figure S1.

In Figure S1, the blue dots correspond to the optimal choice of κ for any $\bar{v}$. Note that the optimal κ depends on $\bar{v}$, and that furthermore, no combination of κ and $\bar{v}$ achieves the optimal social welfare (that is, a normalized social welfare of 1); this implies that scaled marginal-cost tolls are not strongly robust to mischaracterizations of elasticity in a heterogeneous-user model. However, the question of weak robustness is still open; this study was for a very simple class of networks and elasticity profiles and it may be that increasingly complex settings will further demonstrate that these tolls are not even weakly robust.

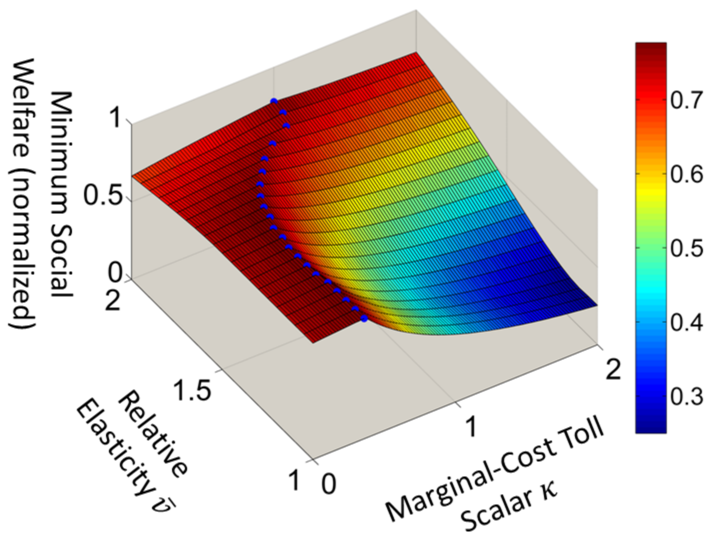

Figure S1. Empirical results for finding the optimal robust tolls for elastic traffic. Depicted are the results of simulating scaled marginal-cost tolls applied to Pigou's Network (see Figure 2) when traffic is both elastic *and* price-sensitive. The elasticity is modeled by value-of-travel function $v(x) = \bar{v} - x$; if the minimum cost of travel on the network exceeds $v(x)$, then user x will stay home. The right horizontal axis represents a scalar κ on marginal-cost tolls and the left horizontal axis represents the relative elasticity of the population $\bar{v}$ (high values of $\bar{v}$ correspond to highly-inelastic traffic). The surface of the plot is the minimum social welfare (divided by the optimal social welfare) on the network resulting from a given toll scalar and elasticity parameter. The key thing to note on this plot is that the optimal choice of toll scalar κ (denoted on the plot by small blue circles) depends on the elasticity parameter $\bar{v}$ (that is, tolls depend on the value-of-travel functions); this is not true in the homogeneous-user model of [25] in which marginal-cost tolls optimize social welfare for all elasticity profiles. This indicates that if users are heterogeneous, marginal-cost tolls are no longer robust to mischaracterizations of a population's elasticity.

Sidebar 3: Budget-Balanced Tolls

Recent work has investigated the application of budget-balanced tolls to a network routing problem. Here, the planner has no net revenue but rather returns all tolls collected on some links as subsidies on the remaining links. This is an attractive approach when it is desirable to include the tolls paid as part of the system cost, since the net tolls paid will always be zero. Budget-balance is a condition which could lead to hopelessly complicated tolling functions for large networks if not properly computed: since everything paid on one edge must be returned to users on other edges, in principle the toll charged on one edge could actually be a function of the flows on *all* other edges. Fortunately, [41] provides an algorithm for computing budget-balanced tolls for any network which enforce optimal flows for homogeneous populations, and each edge's toll is a function of only that edge's flow.

The algorithm from [41] for optimal budget-balanced tolls prescribes a family of tolls for Pigou's Example (refer to Figure 2 for this example) parameterized by a number $q \geq 0$:

$$\tau_1(f_1) = q\,(2 - 1/f_1) + 1/4 \tag{S7}$$

$$\tau_2(f_2) = q\,(2 - 1/f_2) + 1/4\,(1 - 1/f_2)\,. \tag{S8}$$

The q parameter represents the degree to which the upper-link-toll is flow-varying. To see that these are in fact budget-balanced, it is necessary to distinguish between the *toll charged* on an edge and the *revenue collected* from that edge: the revenue collected is given by the toll multiplied by the flow. Combining this with the fact that $f_2 = 1 - f_1$, it is simple to verify that the edge revenues are equal and opposite, or $f_1 \cdot \tau_1(f_1) = -f_2 \cdot \tau_2(f_2)$, for all feasible flows.

The simplest case has $q = 0$ (so that the upper link toll is simply $1/4$), but any nonnegative value of q induces optimal flows for homogeneous unit-sensitivity populations. As a preliminary investigation into the robustness of these tolls to mischaracterizations of user sensitivity, Nash flows are computed as a function of homogeneous sensitivity for various values of q. This plot is shown in Figure S2; naturally, when the sensitivity parameter is equal to 1, the price of anarchy is equal to 1, since this is the sensitivity for which the tolls were designed. Note that for $q = 1$, the price of anarchy is within 1% of optimal for a large range of sensitivities. Furthermore, larger values of q yield a lower price of anarchy for all sensitivity values plotted, suggesting that strongly flow-varying tolls may play an important role in sensitivity-robustness.

For a sense of why a budget-balance constraint in this particular case may confer sensitivity-robustness, consider carefully the tolling functions in (S7) and (S8). Note that near the optimal flow of $(1/2, 1/2)$, both tolling functions are rather gently flow-varying, but that near either extreme point of $(0, 1)$ or $(1, 0)$, users on the low-flow link are being paid extremely large subsidies. This is a simple consequence of a budget-balance constraint on a two-link network: a large number of people paying a small toll on one road implies a small number of people receiving a large subsidy on the other road. From a robustness standpoint, this effect serves to buffer the damage that extremely-high or extremely-low sensitivity agents can cause. Of course, this is far from a comprehensive study on the robustness of budget-balanced tolls, and the apparent robustness seen here could simply be a artifact of the simplified setting of Pigou's network.

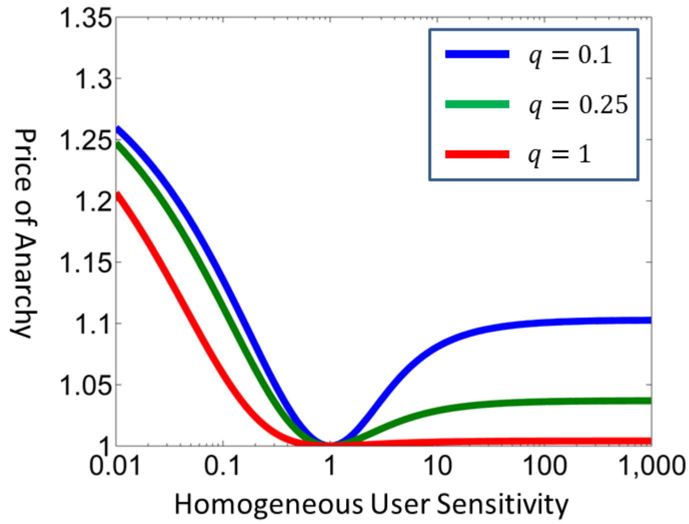

Figure S2. Empirical study on the sensitivity-robustness of budget-balanced tolls. Budget-balanced tolls were computed for a unit-sensitivity homogeneous population on Pigou's network (refer to Figure 2) by the procedure prescribed in [41]; the q parameter describes to what extent the tolling function on edge 1 is flow-varying. Once computed, the population's price sensitivity was varied from 0.01 to 1000 and the resulting price of anarchy was plotted. For $q = 1$, the best case, the price of anarchy is within 1% of optimal for all sensitivities plotted greater than about 0.33, indicating a high level of robustness to mischaracterizations of user sensitivity. Note that for all sensitivities plotted, larger values of q correspond to lower values for the price of anarchy; indicating that strongly flow-varying tolls may play an important role in robustness.