Negative Externalities and Evolutionary Implementation

William H. Sandholm*
Department of Economics
University of Wisconsin
1180 Observatory Drive
Madison, WI 53706
whs@ssc.wisc.edu
http://www.ssc.wisc.edu/~whs

June 23, 2004

Abstract

We model externality abatement as an implementation problem. A social planner would like to ensure efficient behavior among a group of agents whose actions are sources of externalities. However, the planner has limited information about the agents' preferences, and he is unable to distinguish individual agents except through their action choices. We prove that if a concavity condition on aggregate payoffs is satisfied, the planner can guarantee that efficient behavior is globally stable under a wide range of behavior adjustment processes by administering a variable pricing scheme. Through a series of applications, we show that the concavity condition is naturally satisfied in settings involving negative externalities. We conclude by contrasting the performance of the pricing mechanism with that of a mechanism based on direct revelation and announcement dependent forcing contracts.

JEL Classification Numbers: C61, C72, C73, D62, D82, R41, R48

---

* I thank Jeff Ely, Philippe Jehiel, Larry Samuelson, Jörgen Weibull, three anonymous referees, and a Managing Editor, as well as seminar participants at Harvard, the Royal Institute of Technology (Stockholm), the Stockholm School of Economics, Wisconsin, the 2001 Valencia Game Theory Conference, the 2002 Econometric Society Winter Meeting, the 2002 Case Western Reserve University Learning Conference, and the 2002 Society for Economic Dynamics Meeting for helpful comments and discussions. Financial support from the National Science Foundation (SES-0092145) is gratefully acknowledged.

1. Introduction

When the choices of economic agents directly impose costs upon others, equilibrium behavior is inefficient. Commuters drive too often, on the wrong routes, and at the wrong times. Internet users clog network pipelines during peak hours, and overuse applications with heavy bandwidth requirements. Firms whose production processes create pollution produce too much.

A benevolent social planner might attempt to restore efficiency by altering the agents' incentives. Unfortunately, without precise information about the agents' preferences, the planner cannot know exactly what efficient behavior would entail. He can surely reduce the level of externalities by taxing the activities that generate them. But without clear knowledge of preferences, he cannot know whether the taxes inhibit the activities too much or too little.

Were the planner fully informed about preferences, he would be able to choose prices that render efficient behavior an equilibrium. But this still leaves open the question of whether the agents would play this equilibrium. This issue is especially relevant when the number of agents is large, since the knowledge assumptions that are traditionally used to justify equilibrium play are particularly difficult to accept in these cases. At worst, standard marginal externality pricing may allow multiple equilibria, so that even if equilibrium is reached, socially optimal behavior is not assured.

In this paper, we model externality abatement as an implementation problem. Our model incorporates techniques from evolutionary game theory, assuming that agents dynamically adjust their choices in response to the incentives they currently face. We offer a simple condition on aggregate payoffs under which a planner with limited information about preferences can nevertheless ensure that efficient behavior is globally stable. Finally, we present a series of applications that show that our sufficient condition holds in a variety of models involving negative externalities.

In our model, the members of a continuous population of agents each choose an action from a finite set. Each agent's utility function is the sum of two components. The first component captures the externalities that the agents impose upon one another. It is a function of the agent's own choice and the population's aggregate behavior, and is common across agents. The remaining, idiosyncratic component only depends on the agent's own choice of action, but varies from agent to agent.

The planner knows the common component of payoffs, but he has no information about idiosyncratic payoffs.

At each moment in time, the agents' aggregate behavior is described by a population state, which specifies the number of agents of each type playing each action. We model the agents' behavior adjustment process using a differential equation defined on the set of population states. Rather than fixing a particular functional form for this equation, we assume only that it is a member of a wide class of admissible dynamics. The main requirement defining this class is a mild payoff monotonicity condition: namely, that the average, taken over all types, of the covariances between action growth rates and payoffs is positive.

The planner would like to ensure that the agents behave efficiently. He faces two constraints. As is usual in implementation problems, the planner has limited information about the agents' types. In fact, we assume that the planner has no information about types at all. We also assume that the agents are *anonymous*, in that the planner can only distinguish them through the actions they choose. This assumption can be viewed as a hidden action constraint under which the planner is unable to observe precisely which agents choose which actions, but which leaves him able to tax these choices directly. We call the simplest mechanisms that satisfy both constraints *price schemes*. Price schemes specify prices for undertaking each action as functions of current aggregate behavior. They are easy and inexpensive to implement even when the number of agents is large.

We describe the efficiency of a population state in terms of the total utility that the agents obtain at that state. Like each individual's utility function, the total utility function can be decomposed into two terms. One, the total common payoffs, captures the aggregate effects of externalities on the agents. The other, the total idiosyncratic payoffs, sums the benefits whose exact values are the agents' private information.

The main result of the paper can be described as follows. Suppose that the total common payoff function is concave. Then one can construct a price scheme with the following property: for any realization of types, the set of efficient population states is a global attractor under any adjustment process that is admissible given the payoffs induced by the price scheme. Therefore, regardless of the their preferences, their initial behavior, or their precise method of strategy revision, a population of agents subjected to this price scheme will learn to behave efficiently.

Of course, the usefulness of this result hinges on whether the concavity condition can be expected to hold. To address this question, we consider a number of applications of our implementation result: to models of highway congestion with single and multiple departure times, to a model of computer network congestion, and to a model of pollution. In all cases, we show that the required concavity condition holds, and so that global implementation of efficient behavior is possible.

Our analysis relies upon the notion of a *potential game* (Monderer and Shapley (1996), Sandholm (2001)). Potential games admit potential functions, which serve as Lyapunov functions for all admissible dynamics, and so ensure convergence to equilibrium behavior. Typically, the original externality model does not define a potential game, and so evolutionary dynamics in this model need not converge to equilibrium. However, by altering the agents' payoffs using a price scheme, the planner can create a potential game from one that is not. It is the planner's intervention that ensures that evolution leads to equilibrium play.

Our analysis shows that regardless of the realization of types, the optimal price scheme guarantees that the agents play a potential game. Moreover, the potential function of the game created by the price scheme is the total utility function induced by the (known) common payoff function and the (unknown) realization of types. Hence, no matter which types are realized, evolution under any admissible dynamics increases total utility. The assumption that total common payoffs are concave implies that total utility is concave as well, enabling us to conclude that the evolutionary process converges to its global maximizer. Thus, the price scheme always ensures that the agents learn to behave efficiently.

Our optimal price scheme can be viewed as a form of marginal externality pricing. In the classical approach to externality pricing, dating back to Pigou (1920), one charges agents for the externalities that they would create at the efficient state in order to render this state an equilibrium. In contrast, under our price schemes, prices are specified as a function of current aggregate behavior. This variability allows our price scheme to exhibit two important properties. First, variable pricing enables a planner to ensure that efficient behavior is an equilibrium without knowing what efficient behavior will turn out to be. Second, variable pricing lets the planner guarantee that this efficient equilibrium is unique and globally stable, even when such guarantees are not available in the original game without the price scheme, or in the game obtained by imposing a standard Pigouvian pricing scheme.

In an earlier paper (Sandholm (2002)), we studied an implementation problem set on a highway network during a single peak usage period. In that problem, agents chose among different routes from their hometowns to the towns of their workplaces or elected to stay home; externalities took the form of congestion delays on the streets in the network. We showed that the planner could ensure efficient behavior by employing an appropriate price scheme.

This single departure time congestion model possesses two features that make the corresponding implementation problem relatively simple. First, the single departure time model defines a concave potential game, even before the price scheme is imposed. Global convergence to equilibrium is therefore guaranteed even without pricing, so the only role of the price scheme is to ensure that equilibria always constitute efficient play. Second, each agent's type in the single departure time model represents his net benefit from commuting, and so is one-dimensional. In Sandholm (2002), we took advantage of this feature by using a "reduced form" model of the evolution of behavior, an approach that cannot be extended to settings with multidimensional types.[1]

In the present paper, we eliminate both of these simplifying assumptions. This is important because typical models of negative externalities do not define potential games and do exhibit multidimensional types. For example, suppose we make our congestion model more realistic by allowing commuters to choose their times of departure.[2] Allowing idiosyncratic preferences over when to commute immediately requires vector-valued types in order to describe preferences for commuting at different times of day. Moreover, as we explain in detail below, asymmetries in the externalities that early and late drivers impose upon one another imply that the multiple departure time game is not a potential game. The model of evolutionary implementation developed in this paper addresses both of these complications, allowing us to ensure global convergence to efficient play in settings with uncertain convergence properties and with multidimensional types.

The price schemes we study in this paper are quite different from standard mechanisms based on direct revelation. To facilitate comparisons, we show how when the type distribution is continuous, the implementation problem studied in this paper can be solved using mechanisms that combine direct revelation with announcement dependent forcing contracts. The solution concept we apply to these mechanisms utilizes backward induction and dominance. If the hidden action component of the planner's problem is assumed away, so that the forcing contracts

1 For a more precise description of the reduced form dynamics, see Section 6.

2 Departure time choice is a focus of Vickrey's (1963, 1969) work on congestion pricing; see Arnott, de Palma, and Lindsey (1990, 1993) for more recent results. These authors all emphasize that the full benefits of congestion pricing can only be obtained by modifying commuters' time of use decisions.

are unnecessary, the mechanism that remains can be viewed as an infinite player extension of the Vickrey-Clarke-Groves mechanism. Interestingly, equilibrium transfers under the VCG mechanism and the optimal price scheme are the same, although the game forms that lead to these transfers are quite different.

Each of the solutions we offer to the implementation problem has its advantages. The standard mechanism does not require the concavity condition that is needed to ensure global stability of efficient behavior under the price scheme. Moreover, the effectiveness of our price scheme also depends on the fact that agents' types enter their utility functions in an additively separable way; the standard mechanism does not require this restriction.

On the other hand, the standard mechanism is labor-intensive for the planner, since it is based on direct command and control: the planner must collect reports, compute an optimal assignment of agents to actions, and punish agents who do not obey. Each of these tasks may be demanding when the population is large. The optimal price scheme is only effective in settings with negative externalities, but utilizes indirect control: prices are specified for each action, and the agents themselves decide which actions to choose. Because of this decentralization, the price scheme is easy to administer, and is adept at handling changes in the preferences or composition of the population. Given the different features of the two types of mechanism, the choice between the two should depend on the intended application.

Section 2 describes our model of externalities and defines admissible evolutionary dynamics. Section 3 introduces our definition of evolutionary implementation and presents our main result. The proof of this result is contained in Section 4, and a number of extensions are offered in Section 5. Section 6 applies the implementation theorem to four models of negative externalities. Section 7 solves the implementation problem using a standard mechanism and contrasts this mechanism with the optimal price scheme. Concluding remarks are offered in Section 8.

2. The Model

2.1 Population Games

We consider games with a unit mass of agents, each of whom chooses actions from the same set $S = \{1, \dots, n\}$. Each agent's payoff is the sum of two components: a *common payoff*, which depends on the agent's action choice and the overall distribution of actions in the population, and an *idiosyncratic payoff*, which varies from agent to agent and only depends on the agent's own action choice.[3]

We first define the common component of payoffs. Let $X = \{x \in \mathbf{R}^S_+ \colon \sum_i x_i = 1\}$ denote the set of distributions of agents over actions in S. A *common payoff game* is defined by a C^2 payoff function $F\colon X \to \mathbf{R}^S$, where $F_i(x)$ is the common payoff to action i when the action distribution is x.[4]

In what follows, it will be helpful to be able to speak of the marginal impact that an agent choosing strategy j has on the common payoffs of other agents. This is most naturally done using partial derivatives of the form $\frac{\partial F_i}{\partial x_j}$. However, if common payoffs are only defined on the simplex X, these partial derivatives are undefined. We therefore find it convenient to define payoffs for populations whose total mass may be greater or smaller than one. In particular, we assume without loss of generality that F is defined on the set $\overline{X} = \{x \in \mathbf{R}^S_+ \colon \sum_i x_i \in I\}$, where I is an open interval containing 1.

To capture the idiosyncratic aspects of agents' payoffs, we assign each agent a type in the finite set Θ. $M = \{\mu \in \mathbf{R}^\Theta_+ \colon \sum_\theta \mu_\theta = 1\}$ denotes the set of *type distributions*. The set of *population states* under type distribution μ is given by $Z_\mu = \{z \in \mathbf{R}^{\Theta \times S}_+ \colon \sum_i z_{\theta,i} = \mu_\theta \text{ for all } \theta \in \Theta\}$. The scalar $z_{\theta,i} \in \mathbf{R}$ represents the number of agents who are of type θ and who play strategy i, while the vector $z_\theta \in \mathbf{R}^S$ describes the behavior of the agents of type θ. Similar notational conventions are used throughout the paper.

In this multitype setting, a game is defined by a pair (U, μ), where $\mu \in M$ is a type distribution and $U\colon Z_\mu \to \mathbf{R}^{\Theta \times S}$ is a C^2 payoff function. $U_{\theta,i}(z)$ is the payoff to action i

---

[3] The term "common payoff" is not intended to suggest that the value of this payoff is the same for all agents, only for all agents who choose the same action.

[4] If we define a common payoff game by randomly matching agents to play an $n \times n$ normal form game with payoff bimatrix (A, A'), we obtain a game with linear payoff function $F(x) = Ax$. Games with non-linear payoff functions cannot be derived from random matching games. However, these games can be viewed as the limits of normal form games with large numbers of anonymous players – see Sandholm (2001).

for agents of type θ when the population state is z. A population state $z^* \in X_\mu$ is a *Nash equilibrium* if all agents choose a best response to the play of their opponents:

$$\text{(NE)} \qquad i \in \operatorname*{argmax}_{j \in S} U_{\theta,j}(z^*) \text{ whenever } z^*_{\theta,i} > 0$$

We are particularly interested in games constructed by adding idiosyncratic payoffs to a common payoff game. To accomplish this, we suppose that the type set Θ is a subset of $\mathbf{R}^S$, and interpret θ_i as the idiosyncratic benefit that a agent of type θ obtains by playing action i. We also let $x(z) = \sum_\theta z_\theta \in X$ denote the action distribution that obtains at the population state $z \in Z_\mu$. Then, given any common payoff game F and any type distribution $\mu \in M$, we define the *separable game* (F, μ) to be the multitype game with payoff function

$$U_{\theta,i}(z) = F_i(x(z)) + \theta_i.$$

In this game, the common payoff $F_i(x(z))$ depends on the agent's action choice i and on the current action distribution $x(z)$, but not on the agent's type. This component captures the externalities that the agents' impose upon one another. The idiosyncratic payoff θ_i depends on an agent's action choice and on his type, but does not depend on the population state.

For an example of a separable game, consider the following model of highway congestion with multiple departure times.[5] Each action i represents either the combination of a route to work a and a departure time τ, or the option of staying home. In the former case, the common payoff $F_i(x)$ captures the delay along route a at time τ; this delay depends upon the other agents' driving choices. The type component θ_i captures an agent's idiosyncratic benefit from commuting at time τ, and can also be used to describe an idiosyncratic preference for driving on route a.

2.2 Evolutionary Dynamics

While most applications of game theory begin with the assumption of equilibrium play, this assumption seems especially strong in the large population settings that concern us most. We therefore begin with more primitive assumptions about how agents adjust their behavior in response to the incentives

5 This application and others are studied in detail in Section 6.

they face. We will ultimately prove that if the planner employs our pricing mechanism, the dynamics we specify must converge to equilibrium play. Thus, equilibrium play is obtained as a conclusion rather than as an assumption of our approach.

Since the population's behavior is described by a population state in Z_μ, our evolutionary dynamics are defined on this space. These dynamics are defined by a differential equation

(D) $$\dot{z} = g(z),$$

where g: $Z_\mu \to \mathbf{R}^{\Theta \times S}$ is a function from population states to directions of motion through Z_μ. The component $g_{\theta,i}(z) \in \mathbf{R}$ describes the growth of the use of strategy i by agents of type θ, while the vector $g_\theta(z) \in \mathbf{R}^S$ captures all type θ growth rates at once.

Rather than specify a particular functional form for the dynamics (D), we instead require that these dynamics be drawn from a broad class. Our notion of implementation will require that the behavior that the planner prefers be globally stable under all dynamics from this class.

To define this class of dynamics, we introduce one additional definition: for any pair of vectors $x, y \in \mathbf{R}^S$, we define

$$Cov(x, y) = \tfrac{1}{n}\sum_i \left(x_i - \tfrac{1}{n}\sum_j x_j\right)\left(y_i - \tfrac{1}{n}\sum_j y_j\right)$$

to be the covariance between these two vectors when equal weights are placed on each component. With this definition in hand, we say that the dynamics (D) are *admissible* for the multitype game (U, μ) if the following three conditions hold:

(LC) g is Lipschitz continuous.
(FI) Z_μ is forward invariant under (D).
(PC) $\sum_\theta Cov(g_\theta(z), U_\theta(z)) > 0$ whenever $g(z) \neq 0$.
(NC) If $g(z) = 0$, then z is a Nash equilibrium of (U, μ).

Condition (LC) ensures that there is a unique solution to (D) from each initial condition in Z_μ, while condition (FI) ensures that these solutions do not leave Z_μ. We call condition (PC) *positive correlation*. It requires that whenever the dynamics are not at rest, the average, taken over all types, of the covariance between each type's growth rates and payoffs is positive. This is a very weak monotonicity condition, as it restricts behavior in all subpopulations at once using a single scalar inequality. The last condition, *noncomplacency*, (NC), specifies that whenever the population is at rest, its behavior is in equilibrium. If equilibrium has not been reached, there are agents who would benefit from switching strategies; noncomplacency requires that some of these agents eventually do so.[6]

It is worth noting the relationship between the rest points of admissible dynamics and the Nash equilibria of the underlying game.

Proposition 2.1: *Suppose the dynamics* (D) *are admissible for the game* (U, μ). *Then* $z \in Z_\mu$ *is a rest point of* (D) *if and only if it is a Nash equilibrium of* (U, μ).

This result, established in Sandholm (2001), is proved in the Appendix.

3. Evolutionary Implementation

3.1 The Planner's Problem

We now introduce a social planner who would like to ensure that the population behaves efficiently. The planner faces two constraints. The first is due to hidden information: while the planner knows the common payoff function F, he has no information about agents' idiosyncratic payoffs, described by the unknown measure μ. Information constraints of this kind form the core of most problems in mechanism design and implementation theory.

Our planner also faces another restriction, a form of hidden action constraint that we call *anonymity*. Anonymity requires the planner to employ a mechanism that only conditions an agent's transfers on that agent's action and on the

---

6 An example of an admissible dynamics is the *Brown-von Neumann-Nash dynamics*: see Brown and von Neumann (1950), Weibull (1996), Berger and Hofbauer (2000), or Sandholm (2001). For a general class of dynamics that satisfy these conditions, see Sandholm (2003a).

The best response dynamic (Gliboa and Matsui (1991)) fail condition (LC), which introduces technical complications (in particular, nonuniqueness of solution trajectories). Our analysis can be extended to accommodate this dynamic. Finally, it is well known that the replicator dynamic fails condition (NC) on the boundary of the state space: for example, a state at which all agents choose the same action is a rest point of this dynamic, even if the action is strictly dominated. One can incorporate the replicator dynamic into the analysis by only considering interior initial conditions – see Sandholm (2002).

population's aggregate behavior. As we explained earlier, this condition reflects a need to choose mechanisms that are easy to administer, even when the number of agents is large.

We measure the efficiency of a population state z in terms of the *total utility* $\overline{U}(z)$ obtained by the agents at that state:

$$\overline{U}(z) = \sum_{\theta}\sum_{i} z_{\theta,i}\, U_{\theta,i}(z)$$

The function $\overline{U}$ is defined on the set $Z = \bigcup_{\mu\in M} Z_{\mu}$, which contains all population states that can arise under type distributions in M.

The total utility function for the separable game (F, μ) can be split into two components:

$$\begin{aligned}\overline{U}(z) &= \sum_{\theta}\sum_{i} z_{\theta,i}\,(F_i(x(z)) + \theta_i) \\ &= \sum_{i} x_i(z)\,F_i(x(z)) + \sum_{\theta}\sum_{i} z_{\theta,i}\,\theta_i \\ &= \overline{F}(x(z)) + \overline{I}(z).\end{aligned}$$

We call these two components of $\overline{U}$ the *total common payoffs*, $\overline{F}(x) = \sum_i x_i F_i(x)$, and the *total idiosyncratic payoffs*, $\overline{I}(z) = \sum_{\theta}\sum_{i} z_{\theta,i}\,\theta_i$.

We describe the planner's problem in terms of a *social choice correspondence* (*SCC*), $\phi\colon M \Rightarrow Z$, which maps type distributions μ to sets of population states $\phi(\mu) \subseteq Z_{\mu}$. The *efficient social choice correspondence*, $\phi^*\colon M \Rightarrow Z$, is defined by

$$\phi^*(\mu) = \underset{z\in Z_{\mu}}{\operatorname{arg\,max}}\ \overline{U}(z).$$

For every type distribution μ, $\phi^*(\mu)$ specifies the set of population states that maximize total utility among those that are feasible under μ. It is worth emphasizing that efficiency of a state z may require that for some type $\theta \in \Theta$ and distinct strategies $i, j \in S$, the components $z_{\theta,i}$ and $z_{\theta,j}$ both be strictly positive. In other words, efficiency sometimes demands that different agents of the same type choose different actions.[7]

[7] For example, suppose that agents must choose between two routes, and that efficiency requires that both routes be utilized. Then if all agents are of the same type, the efficient state has the property noted above. Indeed, this property holds generically so long as the set of types is finite. When we study standard mechanisms in Section 7, this fact will lead us to consider a model with a continuum of types; we refer the reader to that section for further discussion of this issue.

3.2 Price Schemes and Global Implementation

The simplest mechanisms that respect the planner's hidden information and anonymity constraints are called *price schemes*. A price scheme is a map $P: \Delta \to \mathbf{R}^n$, where $P_i(x)$ represents the price paid by agents choosing action i when the action distribution is x. Price schemes satisfy the planner's information constraint because payments do not condition on any information about the agents' types, distributional or otherwise. Price schemes respect anonymity because the transfer paid by an agent only depends on his action and the population's aggregate behavior.

The original game faced by the agents is defined by a common payoff function F and a type distribution μ. Introducing a price scheme does not alter the type distribution, but it shifts the common payoff function from F to $F - P$. More explicitly, the price scheme changes the common payoff to strategy i under distribution x from $F_i(x)$ to $F_i(x) - P_i(x)$.

With the definition of a price scheme in hand, we can introduce our notion of implementation. We say that the price scheme P *globally implements* the social choice correspondence ϕ if for each type distribution $\mu \in M$, the set $\phi(\mu)$ is globally stable under any dynamics (D) that are admissible for the game $(F - P, \mu)$.

Proposition 2.1 shows that the Nash equilibria of $(F - P, \mu)$ are precisely the rest points of the game's admissible dynamics. Consequently, global implementation implies unique implementation, since if the set $\phi(\mu)$ is globally stable, it must equal the set of Nash equilibria of $(F - P, \mu)$. However, global implementation demands considerably more than unique implementation: it requires that the agents *learn* to play these equilibria, regardless of their initial behavior and their exact method of strategy adjustment.

3.3 The Optimal Price Scheme

The price scheme we use to establish our implementation results is denoted P^*, and is defined by

$$P_i^*(x) = -\sum_j x_j \frac{\partial F_j}{\partial x_i}(x).$$

Under this scheme, the price an agent pays for choosing action i is equal to marginal impact that the agent currently has on other agents' payoffs by choosing this action.

It is critical that the price scheme always is defined in terms of the *current* levels of common payoffs. Under standard Pigouvian pricing, the planner fixes prices equal to the marginal externalities created at the efficient state; by doing so, he renders this state an equilibrium. In the implementation problem considered here, the planner does not know the type distribution μ, and so is unable to identify the efficient state. One function of the variability in prices is to ensure that the efficient state is an equilibrium, regardless of what this state turns out to be.

The variability in prices also serves another crucial role. To see this, suppose that the planner had enough information to set the standard Pigouvian prices. While doing so would render the efficient state an equilibrium, there is generally no reason to expect that the agents would learn to play this equilibrium if play began at an arbitrary disequilibrium state. A successful price scheme must ensure that the efficient state is not simply an equilibrium; it must be an equilibrium that the agents can easily learn to play. This property depends critically on how prices are defined at out-of-equilibrium states.

Our main result provides a simple sufficient condition for global implementation. Its proof is presented in Section 4.

Theorem 3.1 (Global Implementation):

Suppose that average common payoffs $\bar{F}$ are concave. Then the efficient social choice correspondence ϕ^* can be globally implemented using the price scheme P^*.

Theorem 3.1 shows that if the average common payoff function is concave, then the price scheme P^* ensures that the agents will learn to behave efficiently. As our applications in Section 6 illustrate, the concavity of $\bar{F}$ is a property that arises in a number of models of negative externalities. Thus, in such settings, price schemes can be used to globally implement efficient behavior.

4. Analysis

4.1 Potential Games

Our key tool for proving the implementation theorem is the notion of a *potential game*, introduced in a finite agent setting by Monderer and Shapley (1996), and extended to the continuum of agents setting by Sandholm (2001). In this subsection, we review results from Sandholm (2001) that we need to prove Theorem 3.1.

We say that the common payoff game F is a *potential game* if there is a function f such that

$$\frac{\partial f}{\partial x_i}(x) = F_i(x) \text{ for all } x \in X.$$

We call the function f the *potential function* for the game F. Similarly, we say that the multitype game (U, μ) is a potential game with potential function u if

$$\frac{\partial u}{\partial z_{\theta,i}}(z) = U_{\theta,i}(z) \text{ for all } z \in Z_\mu.$$

Potential games possess two attractive properties. First, the Nash equilibria of a potential game can be characterized in terms of its potential function in a simple way. Second, the potential function serves as a Lyapunov function for all admissible dynamics.

To establish the first claim, consider maximizing the potential function u on the set of population states Z_μ:

$$\max\ u(z) \qquad \text{subject to } \begin{array}{ll} \sum_i z_{\theta,i} = \mu_\theta & \text{for all } \theta \in \Theta \\ z_{\theta,i} \geq 0 & \text{for all } i \in S,\ \theta \in \Theta. \end{array}$$

To solve this program, we first specify its Lagrangian:

$$L(z, \rho, \lambda) = u(z) + \sum_\theta \rho_\theta \left(\mu_\theta - \sum_i z_{\theta,i}\right) + \sum_\theta \sum_i \lambda_{\theta,i}\, z_{\theta,i}\,.$$

Here, $\rho \in \mathbf{R}^\Theta$ and $\lambda \in \mathbf{R}^{\Theta \times S}$ are Lagrange multipliers. We then obtain the following Kuhn-Tucker conditions:

(KT1) $\quad \dfrac{\partial u}{\partial z_{\theta,i}}(z) = \rho_\theta - \lambda_{\theta,i};$

(KT2) $\quad z_{\theta,i}\ \lambda_{\theta,i} = 0;$

(KT3) $\quad \lambda_{\theta,i} \geq 0.$

Because the set Z_μ satisfies constraint qualification, satisfaction of the Kuhn-Tucker conditions is necessary (but not sufficient) for maximization of potential. Interestingly, these conditions also characterize the Nash equilibria of (U, μ).

Proposition 4.1: (Characterization of equilibrium)

(*i*) *The population state $z \in Z_\mu$ is a Nash equilibrium of (U, μ) if and only if there exist multipliers $\rho \in \mathbf{R}^\Theta$ and $\lambda \in \mathbf{R}^{\Theta \times S}$ such that the triple (z, ρ, λ) satisfies* (KT1), (KT2), *and* (KT3) *for all $\theta \in \Theta$ and $i \in S$.*

(*ii*) *If its potential function u is concave on Z_μ, then the set of Nash equilibria of the game (U, μ) is the convex set of states that maximize u on Z_μ.*

Proof: We begin with the proof of (*i*). If z is a Nash equilibrium of (U, μ), then since $U_{\theta,i}(z) = \frac{\partial u}{\partial z_{\theta,i}}(z)$, the Kuhn-Tucker conditions are satisfied by z, $\rho_\theta = \max_i U_{\theta,i}(z)$, and $\lambda_{\theta,i} = \rho_\theta - U_{\theta,i}(z)$. Conversely, let $z \in Z_\mu$, and suppose that the triple (z, ρ, λ) satisfies the Kuhn-Tucker conditions. If $z_{\theta,i} > 0$ for some $\theta \in \Theta$ and $i \in S$, then (KT1) and (KT2) imply that $U_{\theta,i}(z) = \frac{\partial u}{\partial z_{\theta,i}}(z) = \rho_\theta$, while (KT1) and (KT3) imply that $U_{\theta,j}(z) = \rho_\theta - \lambda_{\theta,j} \le \rho_\theta$ for all $j \in S$. Thus, $i \in \operatorname{argmax}_j U_{\theta,j}(z)$, and so z is a Nash equilibrium of (U, μ). Given claim (*i*), claim (*ii*) follows from the sufficiency of the Kuhn-Tucker conditions in concave programs. ■

The most important property of the potential function is that it serves as a Lyapunov function for all admissible dynamics. We say that the function $\Lambda: Z_\mu \to \mathbf{R}$ is a *strict Lyapunov function* for the dynamics (D) if $\frac{d}{dt}\Lambda(z_t) \ge 0$ along any solution trajectory $\{z_t\}$, with a strict inequality whenever z_t is not a rest point of (D). Put differently, a Lyapunov function defines a landscape on the space of population states that the dynamics always ascend.

Proposition 4.2: (Characterization of dynamics)

If (U, μ) is a potential game, then its potential function u is a strict Lyapunov function for any dynamics that are admissible under (U, μ). Hence, every solution trajectory of such dynamics converges to a connected set of Nash equilibria of (U, μ).

Proof: In the Appendix (Lemma A.5), we show if g satisfies condition (FI), then

$$\sum_\theta Cov(g_\theta(z), U_\theta(z)) = \tfrac{1}{n} \sum_\theta \sum_i g_{\theta,i}(z) U_{\theta,i}(z).$$

We can therefore compute that

$$\tfrac{d}{dt}u(z_t) = \sum_\theta \sum_i \tfrac{\partial u}{\partial z_{\theta,i}}(z_t)\tfrac{d}{dt}(z_t)_{\theta,i} = \sum_\theta \sum_i U_{\theta,i}(z_t)\, g_{\theta,i}(z_t) = n\sum_\theta Cov(g_\theta(z_t), U_\theta(z_t)).$$

Condition (PC) then implies that $\tfrac{d}{dt}u(z_t) \geq 0$, with equality only when $g(z_t) = 0$. The second claim then follows from Proposition 2.1 and standard results on Lyapunov functions (e.g., Theorem 7.6 of Hofbauer and Sigmund (1988)). ■

4.2 The Proof of the Implementation Theorem

The proof of the implementation theorem requires two additional lemmas. To derive the first lemma, observe that if the planner imposes the price scheme P^*, the agents play a game with common payoff function

$$F_i(x) - P_i^*(x) = F_i(x) + \sum_j x_j \frac{\partial F_j}{\partial x_i}(x).$$

The right hand side of this expression is simply the partial derivative of total common payoffs $\bar{F}$ with respect to x_i:

$$F_i(x) - P_i^*(x) = F_i(x) + \sum_j x_j \frac{\partial F_j}{\partial x_i}(x) = \frac{\partial}{\partial x_i}\bar{F}(x).$$

This relationship can be expressed equivalently as

$$F(x) - P^*(x) = \nabla\bar{F}(x),$$

and can be interpreted as follows.

Lemma 4.3: *The common payoff game $F - P^*$ is a potential game with potential function $\bar{F}$.*

The second lemma shows that if the common payoff game F is a potential game, then so is any game (F, μ) obtained by introducing idiosyncratic payoffs to F. Moreover, the potential function for (F, μ) is obtained by adding the total idiosyncratic payoffs $\bar{I}(z)$ to the potential function of the original game F.

Lemma 4.4: *Suppose that the common payoff game $\hat{F}$ is a potential game with potential function $\hat{f}$. Then for any type distribution μ, the separable game $(\hat{F}, \mu)$ is a potential game with potential function*

$$\hat{f}(x(z)) + \sum_\theta \sum_i z_{\theta,i}\theta_i = \hat{f}(x(z)) + \bar{I}(z).$$

Proof: Follows from both definitions of potential functions and the fact that

$$\tfrac{\partial}{\partial z_{\theta,i}}\left(\hat{f}(x(z)) + \bar{I}(z)\right) = \tfrac{\partial}{\partial x_i}\hat{f}(x) + \theta_i = F_i(x(z)) + \theta_i. \ \blacksquare$$

By combining these two lemmas, we obtain the following result.

Proposition 4.5: *Fix any type distribution $\mu \in M$. Then the game $(F - P^*, \mu) = (\nabla\bar{F}, \mu)$ is a potential game whose potential function is the total utility function $\overline{U}(z) = \bar{F}(x(z)) + \bar{I}(z)$.*

The proof of Theorem 3.1 now follows in a straightforward manner. Fix a type distribution $\mu \in M$, and suppose that the common payoff function F generates a total common payoff function $\bar{F}$ that is concave. Proposition 4.5 shows that the game $(F - P^*, \mu) = (\nabla\bar{F}, \mu)$ is a potential game with potential function $\overline{U}$. Hence, Proposition 4.2 implies that any dynamics that are admissible under $(F - P^*, \mu)$ converge to the set of Nash equilibria of $(F - P^*, \mu)$. Since $\bar{F}$ is concave by assumption, and since $\bar{I}$ is linear by definition, the potential function $\overline{U}(z) = \bar{F}(x(z)) + \bar{I}(z)$ is concave as well. Thus, Proposition 4.1 (*ii*) tells us that all Nash equilibria of $(F - P^*, \mu)$ maximize $\overline{U}$ on the set Z_μ. We therefore conclude that the efficient set $\phi^*(\mu) = \arg\max_{z \in Z_\mu} \overline{U}(z)$ is globally stable under all dynamics that are admissible for the game $(F - P^*, \mu)$, proving the theorem.

We noted earlier that a successful price scheme must serve two roles, ensuring both that efficient play is always an equilibrium, and that this equilibrium is always essentially unique and globally stable. We now examine each of these roles in turn.

Consider the marginal impact on total utility of a change in the mass of agents who are of type θ and who choose strategy i. Writing x for $x(z)$, we see that

$$(\dagger) \qquad \frac{\partial}{\partial z_{\theta,i}}\overline{U}(z) = \left(F_i(x) + \theta_i\right) + \sum_j x_j \frac{\partial F_j}{\partial x_i}(x).$$

The first term, $F_i(x) + \theta_i$, is simply the agent's own utility; he responds to this without intervention by the planner. The second term, $\sum_j x_j \frac{\partial F_j}{\partial x_i}(x) = -P_i^*(x)$ is the marginal effect that the choice of action i has on the other agents' payoffs. The price scheme P^* forces the agents to internalize this effect, so that payoffs under the price scheme always equal marginal total utilities. Moreover, the first order conditions for maximizing the total utility $\overline{U}$ on the space Z_μ imply that at the efficient state z^*, marginal total utilities satisfy

$$\frac{\partial}{\partial z_{\theta,i}}\overline{U}(z^*) = \max_j \frac{\partial}{\partial z_{\theta,j}}\overline{U}(z^*) \text{ whenever } z_{\theta,i}{}^* > 0.$$

Since these partial derivatives are identical to the payoffs under the price scheme, the population state that maximizes total utility must be an equilibrium under the scheme.

At the same time, if we let $U_{\theta,i}^{P^*}(z) = F_i(x(z)) - P_i^*(z) + \theta_i$ denote the payoffs in the game $(F - P^*, \mu)$, then equation (†) is equivalent to the integrability condition

$$\frac{\partial}{\partial z_{\theta,i}}\overline{U}(z) = U_{\theta,i}^{P^*}(z),$$

which shows that $(F - P^*, \mu)$ is a potential game. Thus, even if dynamics for the original game (F, μ) are badly behaved, dynamics for the new game $(F - P^*, \mu)$ are gradient-like, and hence globally convergent.

The concavity condition on $\overline{F}$ is essential to our results. Without this condition, the game $(F - P^*, \mu)$ could admit multiple stable equilibria, rendering global implementation impossible.[8] Fortunately, our applications in Section 6 will show that this assumption is justified in a variety of settings involving negative externalities.

---

8 However, while concavity of $\overline{F}$ is sufficient for our implementation results, it is not necessary. Concavity is only used to ensure that for all realizations of the type distribution μ, the game $(F - P^*, \mu)$ has a unique component of equilibria. If this condition is assumed directly, the implementation theorem continues to hold.

5. Extensions

To prepare for these applications, we offer a number of useful extensions of Theorem 3.1.

5.1 Multiple Populations

In some applications, different subpopulations of the agents face different sets of choices. For example, in modeling highway congestion, it is natural to divide the population into groups corresponding to origin/destination pairs; the set of actions (i.e., routes) from which an agent chooses depends on which pair of towns he must travel between. At the expense of some additional notation, we can extend our model to allow for multiple populations. Let $R = \{1, \ldots, \bar{r}\}$ denote the set of populations. Population $r \in R$ is of size m^r, and its members choose from strategy set $S^r = \{1, \ldots, n^r\}$. The set of strategy distributions is denoted $X = \{x \in \mathbf{R}_+^{n^1+\ldots+n^{\bar{r}}} : \sum_{i\in S^r} x_i = m^r \text{ for all } r \in R\}$. The common payoff function for strategy $i \in S^r$ is denoted F_i, and idiosyncratic payoff vectors for population r are elements of $\mathbf{R}^{n^r}$. An immediate extension of our previous analysis shows that when average common payoffs are concave, the efficient social choice correspondence can be globally implemented by the price scheme P^*, where

$$P_i^*(x) = -\sum_{r'\in R} \sum_{j\in S^{r'}} x_j \frac{\partial F_j(x)}{\partial x_i} \quad \text{for all } i \in S^r \text{ and } r \in R.$$

Thus, the price an agent pays for choosing strategy i again equals the marginal impact that the agent has on other agents' payoffs; in computing this marginal impact, one aggregates over agents from all populations.

5.2 Participation Constraints

In many applications, agents can avoid being subjected to the price scheme by choosing an outside option. For example, a commuter who would like to avoid paying congestion tolls can always opt to stay at home.

A typical feature of outside options is that choosing them does not create externalities for other agents. But it follows from the definition of P^* that if action j does not generate externalities (i.e., if $\frac{\partial F_i}{\partial x_j} \equiv 0$ for all $i \in S$), then it is optimal not to price it ($P_j^* \equiv 0$). Since outside options are not priced under P^*, our implementation results implicitly include participation constraints.[9]

5.3 Other Social Choice Correspondences

While we have supposed so far that the planner's goal is to maximize the agents' total utility, the planner may have other interests as well. For example, if the planner is regulating a polluting industry, he may care not only about industry profits, but also about the effects of pollution on the environment.

We can incorporate such concerns into the model by defining a function $\overline{U}^O$: $\overline{X} \to \mathbf{R}$, where $\overline{U}^O(x)$ represents the impact of the agents' aggregate behavior on the welfare of outside parties. We can then specify a social choice correspondence that incorporates these external effects:

$$\phi^O(\mu) = \operatorname*{argmax}_{z \in Z_\mu} \overline{U}(z) + \overline{U}^O(x(z))$$

Then an easy extension of our previous analysis establishes the following result.

Theorem 5.1: *Suppose that the function $\overline{F} + \overline{U}^O$ is concave. Then the price scheme*

$$P_i^O(x) = -\left(\sum_j x_j \frac{\partial F_j}{\partial x_i}(x) + \frac{\partial \overline{U}^O}{\partial x_i}(x) \right).$$

globally implements the social choice correspondence ϕ^O.

5.4 Discrete Time

While modeling pricing and behavior adjustment in continuous time is analytically convenient, in many applications it is an idealization. For example, in the highway congestion model, it is more natural to think of time as passing discretely, with each time period capturing the driving decisions made on a single workday. In Appendix A.1, we provide a discrete time formulation of our externality pricing model and prove an extension of Theorem 3.1 for this setting.

---

9 We should also note that because we use a continuum of agents framework, budget balance can always be achieved trivially by distributing the planner's revenues equally among all agents. Since each agent is infinitesimal, doing so does not alter the agents' incentives.

6. Applications

6.1 Highway Congestion

Consider a group of agents who live in a collection of towns connected by a network of streets. Each agent can drive from his hometown to the town of his workplace, or can opt to stay home. An agent who chooses to commute selects a route (i.e., a subset of the streets) leading from home to work. A commuter's total travel time is the sum of the delays on each street, which are functions of the number of drivers on that street.

We model the highway network using a collection $\{R, \{m^r\}_{r\in R}, \{A^r\}_{r\in R}, \{\Phi_a\}_{a\in S}, \{c_\phi\}_{\phi\in\Phi}\}$.[10] R is a set of one or more populations r with masses m^r, one for each home/work location pair. The finite set $\Phi = \bigcup_{r\in R}\bigcup_{a\in S^r}\Phi_a$ contains the streets in the network. Each route $a \in A^r$ corresponds to a set of streets $\Phi_a \subseteq \Phi$ that connects the home/work pair r.

Let $\rho(\phi) = \{a \in \bigcup_{r\in R} A^r : \phi \in \Phi_a\}$ denote the set of routes containing street ϕ. The *utilization* of street $\phi \in \Phi$ is the total mass of the agents who drive on that street:

$$u_\phi(x) = \sum_{a\in\rho(\phi)} x_a .$$

Here, x_a represents the mass of agents choosing route $a \in A^r$.

The cost functions c_ϕ: $\mathbf{R}_+ \to \mathbf{R}$ report the delay on each street ϕ as a function of the number of drivers using that street. We suppose that each function c_ϕ is positive, increasing, and convex. The delay on any complete route a is simply the total delay on the streets Φ_a along this route.

6.1.1 Highway Congestion with a Single Departure Time

We first consider a model in which all commuters drive to work at the same time, a version of which was analyzed in Sandholm (2002). In this case, the strategy set S^r for agents in population r consists of the routes in A^r and the outside option 0, which represents staying home. The common payoff to choosing route a is determined by the delay on that route:

[10] This model of highway congestion builds on models introduced by Beckmann, McGuire, and Winsten (1956) and Rosenthal (1973). Jehiel (1993) shows how Pigouvian pricing schemes can be used to prove that certain highway network games are potential games.

$$F_a(x) \;=\; -\sum_{\phi\in\Phi_a} c_\phi(u_\phi(x)).$$

The common payoff to the outside option is always zero: $F_0(x) \equiv 0$.

Each agent's type $\theta = (\theta_0, \theta_1) \in \mathbf{R}^2$ reflects the benefits he obtains by staying home and by going to work, respectively.[11] Thus, the utility function of a population r agent of type θ is given by

$$U_{\theta,a}(z) = F_a(x(z)) + \theta_1 = -\sum_{\phi\in\Phi_a} c_\phi(u_\phi(x(z))) + \theta_1 \quad \text{for } a \in A^r;$$
$$U_{\theta,0}(z) = F_0(x(z)) + \theta_0 = \theta_0.$$

The planner would like to ensure that the agents' behavior maximizes their total utility, which equals the total benefits obtained by commuters (and by those who stay home) minus the total costs of delay. To do so, he must address two types of inefficiencies: commuting by agents whose benefits from doing so are small, and the overuse of easily congested streets. The planner knows the cost functions c_ϕ, but has no information about the agents' types.

To apply Theorem 3.1, we must check that the total common payoff function $\bar{F}$ is concave.

Proposition 6.1: *In the single departure time congestion model, the function $\bar{F}$ is concave.*

Proof: The total common payoff function is given by

$$\begin{aligned}
\bar{F}(x) \;&=\; \sum_a x_a F_a(x) \\
&= -\sum_a x_a \sum_{\phi\in\Phi_a} c_\phi(u_\phi(x)) \\
&= -\sum_{\phi\in\Phi} \sum_{a\in\rho(\phi)} x_a\, c_\phi(u_\phi(x)) \\
&= -\sum_\phi u_\phi(x) c_\phi(u_\phi(x)).
\end{aligned}$$

[11] We are therefore implicitly assuming that the agents do not have idiosyncratic preferences for the individual routes. This assumption is not needed to apply Theorem 3.1, but was essential to our earlier analysis of this problem in Sandholm (2002) – see below.

Since $u_\phi(x)$ is linear in x, to prove that $\bar{F}$ is concave it is enough to show that $a_\phi(u) = u\,c_\phi(u)$ is convex for all ϕ. But since c_ϕ is increasing and convex, $a''_\phi(u) = u\,c''_\phi(u) + 2c'_\phi(u)$ is positive, so a_ϕ is indeed convex. ■

Theorem 6.2: *In the single departure time congestion model, the efficient SCC ϕ^* can be globally implemented using the price scheme P^*, given by*

$$P^*_a(x) = \sum_{\phi\in\Phi_a} u_\phi(x)\,c'_\phi(u_\phi(x));$$
$$P^*_0(x) \equiv 0.$$

Notice that the optimal price scheme P^* is separable in ϕ: the price of each route can be decomposed into prices $p_\phi(u) = u\,c'_\phi(u)$ for each street ϕ along the route; the price of each street only depends on the level of congestion on that street.

6.1.2 Highway Congestion with Multiple Departure Times

There are two features of the single departure time model that make the implementation problem in this setting relatively simple. First, types in this example are essentially one-dimensional: each agent's decisions are only depend on θ through the difference $\theta_1 - \theta_0$, which represents his net benefit of commuting. In Sandholm (2002), we took advantage of this feature by basing our analysis on "reduced form" evolutionary dynamics, which were defined directly on the space of action distributions $X \subset \mathbf{R}^S$ rather than on the set of population states $Z_\mu \subset \mathbf{R}^{\Theta\times S}$.[12] This reduced form approach cannot be applied in settings with multidimensional types.

Second, even before the price scheme P^* is imposed, the game (F, μ) is a potential game with concave potential function

$$-\sum_{\phi\in\Phi}\int_0^{u_\phi(x(z))} c_\phi(y)\,dy + \bar{I}(z).$$

[12] In particular, we assumed that at each moment in time, the agents who chose to drive were precisely those whose net valuations $\theta_1 - \theta_0$ were highest. This property must hold in equilibrium, but it is a strong assumption to make when specifying out of equilibrium dynamics.

It follows that the equilibria of (F, μ) are essentially unique and globally stable even before the price scheme is imposed. Therefore, the only role of the price scheme is to ensure that equilibria always constitute efficient play.

While abstracting away from the choice of departure time yields a model with convenient features, doing so obscures an issue that is critical in practice: the full benefits of congestion pricing can only be obtained if it is used to alter commuters' time of use decisions. The importance of this issue is emphasized in Vickrey's (1963, 1969) work on congestion pricing, as well as in more recent papers by Arnott, de Palma, and Lindsey (1990, 1993). Furthermore, most congestion pricing schemes currently in operation set prices that vary with the time of day, with the explicit purpose of smoothing peak usage over a longer time span.[13]

Once one introduces time of use decisions, the two convenient features from the single departure time model vanish. If commuters have idiosyncratic preferences not only about staying home versus commuting, but also about when the commute occurs, it is clear that types will be genuinely multidimensional. More interestingly, time of use choice also implies that the commuting game is not a potential game. To see why, recall that the common payoff function F defines a potential game if there exists a potential function f satisfying $\nabla f(x) \equiv F(x)$. Such a function exists precisely when

$$\frac{\partial F_i}{\partial x_j}(x) \equiv \frac{\partial F_j}{\partial x_i}(x)$$

for all strategies i and j.[14] In Sandholm (2001), we call the latter condition *externality symmetry*: it requires that the marginal impact of new strategy j users on current strategy i users is the same as the marginal impact of new strategy i users on current strategy j users. The single departure time model satisfies this condition: the impact that route a drivers have on route b drivers occurs through marginal increases in delays on the streets $\Phi_a \cap \Phi_b$ that the two routes share; the impact of the b drivers on the a drivers is exactly the same. But when there are multiple departure times, externality symmetry fails. Because of the queuing that occurs under congested conditions, drivers who use the network during the early period can increase the

13 For a survey of congestion pricing systems currently in use, see Small and Gomez-Ibañez (1998).

14 This is also a necessary and sufficient condition for the multitype game (F, μ) to be a potential game.

delays experienced by the drivers who use the network during the late period.[15] But while early drivers impose externalities on late drivers, late drivers do not impose externalities on early drivers. Therefore, in the absence of prices, convergence to equilibrium is not guaranteed.

We now introduce a simple modification of our simple departure time model that incorporates the time of use decision, and show that an appropriate price scheme can be used to globally implement efficient behavior in this more complex setting. In this new model, agents not only choose which route to take to work, but also when to drive. The agents' active strategies are pairs (a, τ), where a corresponds to a route $\Phi_a \subseteq \Phi$, and $\tau \in \{1, 2\}$ represents the choice of an early (1) or late (2) departure time; as before, the strategy 0 represents the decision to stay home. Utilization functions for the two periods are defined as follows:

$$u_{\phi,1}(x) = \sum_{a\in\rho(\phi)} x_{a,1};$$
$$u_{\phi,2}(x) = \sum_{a\in\rho(\phi)} x_{a,2}.$$

Each agent's type is a vector $\theta = (\theta_0, \theta_1, \theta_2)$, representing his idiosyncratic payoffs to staying home, commuting early, or commuting late. Once again, the planner has no information about the agents' types.

To capture the dependence of period 2 delays on period 1 behavior, we define the *effective utilization* of street ϕ in period 2 to equal not simply $u_{\phi,2}$, but $u_{\phi,2} + s_\phi(u_{\phi,1})$. We call the functions s_ϕ: $\mathbf{R}_+ \to \mathbf{R}$ *spillover functions*, and assume that they are increasing, convex, and satisfy $s_\phi(0) = 0$. In addition, we assume that $s_\phi'(u) \leq 1$ for all u: an increase in the utilization of a street by du units in period 1 cannot increase the effective utilization of this street by more than du units in period 2. With this additional notation in hand, we define the utilities obtained by the early and late drivers as

$$U_{\theta,(a,1)}(z) = F_{a,1}(x(z)) + \theta_1 = -\sum_{\phi\in\Phi_a} c_\phi(u_{\phi,1}(x(z))) + \theta_1;$$
$$U_{\theta,(a,2)}(z) = F_{a,2}(x(z)) + \theta_2 = -\sum_{\phi\in\Phi_a} c_\phi(u_{\phi,2}(x(z)) + s_\phi(u_{\phi,1}(x(z)))) + \theta_2.$$

[15] Arnott, de Palma, and Lindsey (1993) emphasize the importance of capturing this interdependence between periods when modeling departure time choice.

Despite the complications that time of use choice creates, global implementation of efficient behavior is still possible. To apply Theorem 3.1, we must establish that the total common payoff function is concave.

Proposition 6.3: *In the two departure time congestion model, the function* $\bar{F}$ *is concave.*

Proof: In the Appendix.

Theorem 6.4: *In the two departure time congestion model, the efficient SCC* ϕ^* *can be globally implemented using the price scheme* P^**, given by*

$$P_{a,1}^*(x) = \sum_{\phi \in \Phi_a} \Big(u_{\phi,1}(x) c_\phi'(u_{\phi,1}(x)) + u_{\phi,2}(x) s_\phi'(u_{\phi,1}(x)) c_\phi'(u_{\phi,2}(x) + s_\phi(u_{\phi,1}(x))) \Big);$$
$$P_{a,2}^*(x) = \sum_{\phi \in \Phi_a} u_{\phi,2}(x) c_\phi'(u_{\phi,2}(x) + s_\phi(u_{\phi,1}(x)));$$
$$P_0^*(x) = 0.$$

In the two departure time model, the price scheme P^* guarantees not only that the efficient state is always an equilibrium, but also that it is the unique equilibrium and is globally stable under all admissible dynamics. Once again, the price scheme is separable in ϕ, and so can be administered on a street-by street basis.

6.2 Congestion in a Computer Network

Computer networks are also subject to congestion externalities, but these networks differ from highway networks in important ways. At a first approximation, drivers using a highway network are homogenous in terms of their contributions to and preferences concerning delays. In computer networks, the consumption of network resources and the consequences of delays vary substantially with the network applications in question. For example, a user who is sending an e-mail message uses little of the network's capacity and can tolerate a fair amount of delay, while a user who participates in a videoconference consumes substantial network resources and is quite intolerant of delay.[16]

---

[16] For further discussion of this issue, see Cocchi *et. al*. (1993). Other models of pricing in computer networks include MacKie-Mason and Varian (1995), Shenker *et. al*. (1996), and Crémer and Hariton (1999).

We capture these features of computer networks using a simple model of a single segment of network pipeline. Users of the network choose from a range of activities $\{1, \ldots, A\}$. Each activity a is characterized by a triple $(l_a, w_a, c_a(\cdot))$. The integer $l_a > 0$ represents the length of time that the activity takes to complete. The scalar $w_a > 0$ equals the bandwidth that activity a requires while in progress. Finally, the function $c_a: \mathbf{R}_+ \to \mathbf{R}$ represents the per period cost of delay for activity a as a function of current network utilization. We assume that the functions c_a are positive, increasing, and convex.

An active strategy for an agent consists of an activity $a \in \{1, \ldots, A\}$ and a start time $\tau \in \{0, 1, \ldots, \pi - 1\}$, where π is the number of periods into which each day is divided. Users may choose either an active strategy (a, τ) or the outside option 0. Each agent's type is a vector $\theta \in \mathbf{R}^{A\pi+1}$. The scalar $\theta_{a,\tau}$ represents the benefit an agent obtains from performing activity a at start time τ; θ_0 is the value of his outside option. As always, the agents' types are unknown to the planner.

To define utility functions, let $\rho(\tau) = \{(b, \psi): \tau \in \{\psi, (\psi + l_b - 1)\}\}$ be the set of strategies that use the network during period τ.[17] Total network utilization at time τ is then defined by

$$u_\tau(x) = \sum_{(b,\psi)\in\rho(\tau)} w_b \, x_{b,\psi} \,.$$

Note that the effect of agents who choose activity b on total utilization is scaled by the bandwidth requirement w_b. Utilities are defined by

$$U_{\theta,(a,\tau)}(z) = F_{a,\tau}(x(z)) + \theta_{a,\tau} = -\sum_{\psi=\tau}^{\tau+l_a-1} c_a(u_\psi(x(z))) + \theta_{a,\tau}.$$

In general, the computer network congestion model does not define a potential game. Consider any pair of strategies (a, τ) and (b, ψ) that are active during overlapping sets of periods. In any period ζ during which both strategies are active, the marginal effect of the activity b users on the payoffs of the activity a users is $-w_b\, c_a'(u_\zeta)$; the effect in the reverse direction is $-w_a\, c_b'(u_\zeta)$. For externality symmetry to hold, these expressions must always be equal, which is only true if for each

[17] For convenience, we assume that agents do not choose pairs (b, ψ) such that $\psi + l_b - 1 > \pi - 1$, which use the network during the late periods of one day and the early periods of the next. This assumption is not essential to our analysis.

utilization level u, the marginal delay cost/bandwidth ratio $c_a'(u)/w_a$ is independent of the activity a under consideration. If this strong condition does not hold, then in the absence of prices, convergence to equilibrium play is not assured.

Nevertheless, the planner can still globally implement efficient behavior.

Proposition 6.5: *In the computer network model, the function* $\bar{F}$ *is concave.*

Proof: In the Appendix.

Theorem 6.6: *In the computer network model, the efficient SCC* ϕ^* *can be globally implemented using the price scheme* P^*, *given by*

$$P^*_{a,\tau}(x) = w_a \sum_{\zeta=\tau}^{\tau+l_a-1} \left(\sum_{(b,\psi)\in\rho(\zeta)} x_{b,\psi}\, c_b'(u_\psi(x)) \right);$$

$$P_0^*(x) = 0.$$

Once again, the optimal price scheme P^* takes a very simple form. During each period ζ, network usage is priced at $\sum_{(b,\psi)\in\rho(\zeta)} x_{b,\psi}\, c_b'(u_\psi(x))$ per unit of bandwidth. Hence, the price at time ζ only depends on the use of the network at time ζ. As always, the outside option can be chosen free of charge.

6.3 A Model of Pollution

Our final example shows how a planner can devise price schemes that achieve goals other than maximizing the payoffs of active agents, a possibility discussed in Section 5.3. This example concerns the regulation of an industry whose production generates pollution, and so imposes externalities on outside parties.

Consider a population of firms that posses two production technologies, A and B. These technologies differ in their productivity and in the amounts of pollution they generate. Each firm chooses output levels $a, b \in \{0, 1, \ldots, M\}$ to produce using each technology. Total production by the industry using technologies A and B is given by $a_T(x) = \sum_{a,b} a x_{a,b}$ and $b_T(x) = \sum_{a,b} b x_{a,b}$, where $x_{a,b}$ represents the proportion of firms choosing production plan (a, b). Total production overall equals $q_T(x) = a_T(x) + b_T(x)$. The function p: $\mathbf{R}_+ \to \mathbf{R}$ is the inverse demand curve for the good; it specifies the price $p(q_T)$ of the good as a decreasing function of overall production. The environmental costs of pollution are described by a convex function e: $\mathbf{R}^2_+ \to \mathbf{R}$ of the total production levels a_T and b_T. Each firm's cost function is described by a vector $\theta \in \mathbf{R}_-^{(M+1)^2}$, where $-\theta_{a,b}$ represents the cost of production plan (a, b). The firms' costs functions are not known to the planner.

The utility obtained by a firm of type θ is the difference between its revenues and its costs:

$$U_{a,b}(z)(\theta) = F_{a,b}(x(z)) + \theta_{a,b} = (a + b)\, p(q_T(x(z))) + \theta_{a,b}.$$

Total utility equals industry profits:

$$\overline{U}(z) = \overline{F}(x(z)) + \overline{I}(z) = q_T(x(z))\, p(q_T(x(z))) + \sum_\theta \sum_{a,b} z_{\theta,(a,b)}\, \theta_{a,b}$$

While the planner could aim to maximize total profits, we suppose instead that he prefers to maximize social surplus. To do so, we define

$$\overline{U}^O(x) = \int_0^{q_T(x)} (p(y) - p(q_T(x)))\, dy - e(a_T(x), b_T(x))$$

to be the total effect on outside parties when the distribution of production plans is x. The first term of $\overline{U}^O$ is the consumer surplus obtained when total production is $q_T(x)$. The second term reflects the environmental costs of pollution. The planner's social choice correspondence accounts for industry profits, consumer surplus, and environmental costs:

$$\phi^O(\mu) = \operatorname*{arg\,max}_{z \in Z_\mu} \overline{U}(z) + \overline{U}^O(x(z)).$$

To apply Theorem 5.1 to this model, we must verify that $\overline{F} + \overline{U}^O$ is concave.

Proposition 6.7: *In the pollution model, the function* $\overline{F} + \overline{U}^O$ *is concave.*

Proof: Observe that

$$\begin{aligned} \overline{F}(x) + \overline{U}^O(x) &= q_T(x)\, p(q_T(x)) + \int_0^{q_T(x)} (p(y) - p(q_T(x)))\, dy - e(a_T(x), b_T(x)) \\ &= \int_0^{q_T(x)} p(y)\, dy - e(a_T(x), b_T(x)). \end{aligned}$$

Since a_T, b_T, and q_T are linear, p is decreasing, and e is convex, the result follows. ■

Theorem 6.8: *In the pollution model, the SCC ϕ^O can be globally implemented using the price scheme P^O, defined by*

$$P^O_{a,b}(x) = a\left(\tfrac{\partial}{\partial a_T} e(a_T(x), b_T(x))\right) + b\left(\tfrac{\partial}{\partial b_T} e(a_T(x), b_T(x))\right).$$

To ensure that social surplus is maximized, the planner simply taxes each firm $\frac{\partial}{\partial a_T} e(a_T, b_T)$ for each unit it produces using technology A, and $\frac{\partial}{\partial b_T} e(a_T, b_T)$ for each unit it produces using technology B.[18]

7. Comparison with Standard Mechanisms

The implementation problem studied in this paper involves both hidden information and hidden actions (anonymity). While so far we have considered a mechanism based on variable prices, it is also possible to solve this implementation problem using mechanisms built from more standard components, namely type announcements and announcement dependent forcing contracts. In this section, we present a standard mechanism that solves the implementation problem using a solution concept based on backward induction and dominance. We then compare the performance of this mechanism with that of our optimal price scheme.

The foregoing sections utilized a framework in which the set of types is finite. However, there is no conceptual difficulty extending our earlier results to settings with a continuum of types. This can be accomplished formally using tools developed in Ely and Sandholm (2003) and Sandholm (2004), though at the cost of introducing more sophisticated mathematical techniques.

In this section, we wish to construct an anonymous standard mechanism that solves our implementation problem. If we retain the finite type set framework, we face an immediate difficulty. While we would like to use announcement dependent forcing contracts to ensure obedience, anonymity demands that these contracts yield identical transfers for agents whose announcements and behavior are the same. Consequently, whenever efficiency requires agents of the same type to choose different actions (which, as we argued in Section 3.1, it typically will),

[18] Observe that if there are no pollution externalities ($e(a_T, b_T) \equiv 0$), then the optimal tax on production is always zero ($P^* \equiv 0$). This fact has two noteworthy implications. First, without the planner's intervention, the quantity setting game is a concave potential game, and so admits an essentially unique and globally stable equilibrium. Second, in the absence of externalities, this equilibrium maximizes social welfare.

anonymous forcing contracts are incapable of rendering obedience a dominant strategy.[19]

One way to contend with this issue is to introduce notions of ε-equilibria and ε-efficiency and to employ various approximation results. We prefer instead to utilize a model with a continuum of types. As long as the distribution of types is smooth, efficient states will assign agents of the same type to the same action, precluding the difficulty described above. Moreover, since our evolutionary implementation results can also be established with a continuum of types, direct comparisons of these results with the results for standard mechanisms developed below are valid.

7.1 A Standard Mechanism

To define mechanisms based on direct revelation, we must take as our primitive a set of agents rather than a set of population states. We therefore let $(A, \mathcal{A}, m)$ be a probability space, where A is the set of agents and m is a probability measure. Each agent α has a type $\tau(\alpha) \in \Theta = \mathbf{R}^S$, so the full type profile is denoted $\tau \in \Theta^A$. Each type profile induces a type distribution μ^τ defined by $\mu^\tau(B) = m(\alpha\colon\ \tau(\alpha) \in B)$. For the reasons discussed above, we assume that each possible type realization τ induces a distribution μ^τ that admits a density on $\Theta = \mathbf{R}^S$. We denote action profiles by $s \in S^A$. Each action profile induces an action distribution $x^s \in X$, where $x_i^s = m(\alpha\colon\ s(\alpha) = i)$. Throughout our analysis, we will consider equalities and conditions defined for all agents satisfied as long as they hold on a full measure subset of A.

Apart from the move to continuous type distributions, the implementation problem is unchanged. The planner knows the common payoff function F, but has no information about the type profile τ. He would like to construct a mechanism that implements the efficient social choice correspondence $\phi^*\colon \Theta^A \Rightarrow S^A$, defined by

[19] To see why, suppose as in footnote 7 that all agents are of the same type, that they must choose between two routes, and that efficiency requires that both routes be utilized (say, in equal proportions). One can certainly find route specific transfers that make obedience of an efficient assignment of agents to routes a Nash equilibrium. Indeed, this can be accomplished using the equilibrium prices arising under the optimal price scheme P^*. However, coordinated disobedience is also an equilibrium. Suppose that agents are labeled with numbers in the interval [0, 1] (see Section 7.1 below). Then even if the planner tells agents in the interval [0, 1/2] to take route a and chooses route specific transfers that make obedience a Nash equilibrium, each disobedient allocation in which the agents in the interval $[\alpha, \alpha + 1/2] \subset [0, 1]$ take route a is a Nash equilibrium as well. Moreover, obedience cannot be even a weakly dominant strategy: if all of an agent's opponents take route a, then the agent himself strictly prefers route b, but if the opponents all take b, the agent strictly prefers a.

$$\phi^*(\tau) = \arg\max_{s \in S^A} \left(\bar{F}(x^s) + \sum_i \int_{\{\alpha : s(\alpha) = i\}} \tau_i(\alpha) dm(\alpha) \right).$$

To be feasible, the planner's mechanism cannot condition on any information about realized types, and must respect the agents' anonymity.

Mechanisms from the class we will consider are executed as follows:

Stage 1: Each agent α announces a type $\hat{\tau}(\alpha) \in \Theta$. After receiving these announcements, the planner assigns actions $\hat{s}(\alpha) \in S$ for each agent α.
Stage 2: Upon receiving his assignment, agent α chooses an action $s(\alpha) \in S$. After observing the agents' announcements and assignments, the planner collects transfers $t(\alpha) \in \mathbf{R}$ from each agent α.

Formally, such a mechanism can be described by a pair $(\hat{S}, T)$, where $\hat{S}$: $\Theta^A \to S^A$ and T: $\Theta^A \times S^A \to \mathbf{R}^A$. The mechanism specifies an assignment profile $\hat{s} = \hat{S}(\hat{\tau}) \in S^A$ for each announcement profile $\hat{\tau} \in \Theta^A$, and a transfer profile $t = T(\hat{\tau}, s) \in \mathbf{R}^A$ for each assignment profile/action profile pair $(\hat{\tau}, s) \in \Theta^A \times S^A$. Mechanisms of this kind clearly respect the planner's information constraint, as they only condition on type announcements, not on actual types. To ensure anonymity, we impose two additional requirements. We call the mechanism $(\hat{S}, T)$ *anonymous* if for any two profile pairs $(\hat{\tau}, s)$ and $(\hat{\tau}', s')$, the resulting play sequences $(\hat{\tau}, \hat{s} = \hat{S}(\hat{\tau}), s, t = T(\hat{\tau}, s))$ and $(\hat{\tau}', \hat{s}' = \hat{S}(\hat{\tau}'), s', t' = T(\hat{\tau}', s'))$ have the following properties:

(A1) If $\mu^{\hat{\tau}} = \mu^{\hat{\tau}'}$ and $\hat{\tau}(\alpha) = \hat{\tau}'(\alpha')$, then $\hat{s}(\alpha) = \hat{s}'(\alpha')$.
(A2) If $\mu^{\hat{\tau}} = \mu^{\hat{\tau}'}$, $\hat{\tau}(\alpha) = \hat{\tau}'(\alpha')$, and $s(\alpha) = s'(\alpha')$, then $t(\alpha) = t'(\alpha')$.

Condition (A1) captures two restrictions. When $\hat{\tau}$ and $\hat{\tau}'$ are identical, (A1) requires that agents who send the same announcement receive the same action recommendation. More generally, (A1) asks that recommendations be invariant with respect to permutations of the announcement profile from $\hat{\tau}$ to $\hat{\tau}'$. Similarly, condition (A2) requires that agents who send the same announcement and choose the same action are told to pay the same transfer, and also requires invariance of transfers with respect to synchronized permutations of the announcement and action profiles. If these conditions hold, the mechanism does not condition in any way on the agents' names.

Theorem 7.1 establishes the existence of an anonymous mechanism that implements the efficient social choice correspondence. The proof of this result, which includes the explicit construction of the mechanism, is provided in the Appendix.

Theorem 7.1: *There is an anonymous mechanism* $(\hat{S}, T)$ *that implements the efficient social choice correspondence* ϕ^* *under the following solution concept: elimination of strictly dominated strategies at all stage 2 decision nodes, followed by selection of a profile of weakly dominant strategies in stage 1. In this equilibrium, agents tell the truth and obey their assignments. If the agents announce type profile* $\hat{\tau}$*, the planner specifies the assignment profile* $\hat{s} = \hat{S}(\hat{\tau})$*, and the agents choose the action profile* s*, transfers are given by*

$$t(\alpha) = T(\hat{\tau}, s)(\alpha) = \begin{cases} P^*_{\hat{s}(\alpha)}(x^{\hat{s}}) & \text{if } s(\alpha) = \hat{s}(\alpha), \\ L & \text{otherwise,} \end{cases}$$

where P^* *is the optimal price scheme and* L *is a large real number.*

The mechanism $(\hat{S}, T)$ described in the theorem is constructed as follows. As long as the agents send a profile of announcements $\hat{\tau}$ that is feasible (in that it admits a density), the mechanism specifies an assignment $\hat{s} = \hat{S}(\hat{\tau})$ of agents to actions that is efficient conditional on reports being truthful. Honesty and obedience are ensured using transfer payments.

To guarantee obedience, the mechanism incorporates announcement dependent forcing contracts: any agent who disobeys his assignment is required to pay a stiff fine. Consequently, during the second stage of the mechanism each agent finds it dominant to choose his assigned action.[20]

To ensure honesty, the mechanism charges agents who are assigned to and select action i a price of $P_i^*(x^{\hat{s}})$, where $x^{\hat{s}}$ is the action distribution under $\hat{s}$, and P^* is the optimal price scheme defined in Section 3. To understand the role played by the

[20] In standard moral hazard models, a principal receives a possibly imperfect signal of an agent's action. The term "forcing contract" is used when the signals generated by actions the principal finds desirable and the signals generated by actions the principal finds undesirable have disjoint supports. A forcing contract levies a fine when signals from the latter set occur, coercing the agent into choosing an action from the principal's preferred set. Under our announcement dependent forcing contracts, transfers depend on perfectly observed action choices as well as on type reports solicited during the mechanism's previous stage. By fining agents who disobey their report contingent assignments, the planner is able to compel agents to obey.

optimal price scheme, let $\mu^{\hat{\tau}}$ denote the distribution of type announcements. By the logic of Section 4, state $x^{\hat{s}}$, which is by definition an efficient assignment of agents to actions in the multitype game $(F,\ \mu^{\hat{\tau}})$ is also an equilibrium of the multitype game $(F - P^*,\ \mu^{\hat{\tau}})$.[21] Hence, if an agent of type θ is assigned to action i, it must be that

(*) $$F_i(x^{\hat{s}}) - P_i^*(x^{\hat{s}}) + \theta_i \geq F_j(x^{\hat{s}}) - P_j^*(x^{\hat{s}}) + \theta_j \text{ for all } j \in S.$$

Thus, as long as all agents are obedient, each agent has an incentive to tell the truth in stage 1: doing so ensures that he is assigned to an action i that satisfies equation (*); lying may cause him to be assigned to another action, which can only yield him a lower payoff. In other words, once disobedience is ruled out in stage 2, honesty becomes weakly dominant in stage 1.

We can simplify our analysis by eliminating certain of our requirements for our mechanism. If we drop the anonymity requirement and instead allow the planner to condition transfers on agents' names, then the planner can ensure obedience even when agents of the same type are assigned to different actions. Hence, efficient behavior can be implemented without the restriction to continuous type distributions. If we take the further step of directly assuming obedience, then forcing contracts become unnecessary; in the pure revelation mechanism that remains, truth telling is a weakly dominant strategy. This truncated mechanism can be viewed as an infinite player version of the VCG mechanism (Vickrey (1961), Clarke (1971), Groves (1973)).

The VCG mechanism is commonly described as a mechanism that forces agents to pay for the externalities they impose on others. The fact that the transfer payments are the marginal externality prices seems to validate this description. However, it is important to note that under the VCG mechanism, the externality that an agent must pay for is that which his *announcement* creates for other agents, by way of its effect on the action to which he is assigned. In contrast, under a price scheme, each agent directly chooses an *action*, and the externality he pays for is that due to the direct impact of his action choice on the others' payoffs.[22]

---

[21] For details on the definition and analysis of multitype games with a continuum of types, see the Appendix.

[22] Since each agent in our model is of negligible size, an agent who changes his announcement does not affect the distribution of action assignments. The only externality an agent creates by changing his announcement occurs through its impact on the action to which he himself is assigned, and so his transfer is therefore just the externality that he creates by taking this action. In contrast, in a finite player framework one agent's announcement *can* alter how the planner distributes the other agents over the actions. Consequently, if the standard VCG mechanism is applied to a finite player version of our

7.2 Comparing Standard Mechanisms and Price Schemes

We have offered two rather different mechanisms for solving the planner's problem. We now assess the relative strengths of each.

7.2.1 *Payoff restrictions*. The biggest advantage of the standard mechanism is that it works without any restriction on the common payoff function F. In contrast, the success of the price scheme depends on the concavity of total common payoffs $\bar{F}$, and so restricts application of the mechanism to settings with negative externalities.

7.2.2 *Solution concepts*. The solution concept for the standard mechanism combines backward induction and dominance arguments, while the solution concept for the price schemes is global stability under admissible dynamics. Both of these concepts are quite weak, and so the mechanisms seem comparable in terms of the demands placed on the agents.[23]

7.2.3 *Ease of use and flexibility*. The greatest advantage of the price scheme is the ease with which a planner can administer it. Standard mechanisms are based upon direct command and control. To administer an standard mechanism, the planner must collect reports from each agent, compute an efficient assignment based on these reports, and then monitor behavior to ensure that assignments are obeyed. Each of these tasks may be quite demanding, especially when the number of agents is large. In contrast, price schemes rely on an indirect form of control. The planner sets prices for each action as a function of current aggregate behavior, and allows the agents to decide for themselves how to behave. This decentralization allows the planner to administer price schemes with relative ease.

Decentralization also creates flexibility in the face of variations in preferences. For example, suppose that a number of new commuters move into the planner's region, or that the types of the old commuters change. While the planner would need to rerun the standard mechanism to elicit the agents' new preferences, the optimal price scheme contends with these changes automatically. Hence, price schemes may be preferable to revelation mechanisms when preferences are expected to vary over time.

---

model, agent α's transfer will generally *not* equal the externality he creates by performing the action to which he is assigned—see Sandholm (2003b).

[23] Under the standard mechanism, agents' equilibrium action choices are efficient, while under the price scheme the efficient state is only reached after the adjustment process is complete. For this reason, if revelation and optimization are not too time consuming, time preferences might lead the planner to prefer the standard mechanism.

8. Conclusion

We considered evolutionary implementation in a general model of negative externalities. Our main result showed that a planner with limited information about agents' preferences can ensure the global stability of efficient behavior by administering a variable price scheme.

We studied the implementation problem using a continuum of agents model. In potential games with a continuum of agents, concavity of the potential function implies the existence of a unique component of equilibria that is the global attractor of any admissible dynamics. When we analyzed the optimal price scheme, this fact helped us establish the global stability of the set of efficient states. Moreover, as we saw in Section 6, the continuum model is convenient in applications, where the concavity condition can be verified in intricate models of externalities by applying standard calculus-based techniques. Still, versions of all of our results can be established for models with large, finite populations by applying suitable approximation results.[24]

Our analysis relies on the assumption that the total common payoff function is concave, which ensures uniqueness of equilibrium under the optimal price scheme. While this condition is natural in models of negative externalities, it does not hold more generally. In particular, when externalities are positive, the concavity condition fails, and the game created by the price scheme typically exhibits multiple stable equilibria.[25]

In Sandholm (2003b), we study externality pricing using a finite agent, discrete time model without imposing the restriction that externalities be negative. Unlike in the present paper, behavior adjustment in Sandholm (2003b) is stochastic, with choice probabilities determined by the logit choice rule. We show that if a planner executes an appropriate price scheme, the efficient strategy profiles are the only ones

[24] In particular, the continuum model can be viewed as the limit of finite population models, the equilibria of which converge to the unique component of equilibria of the continuum model when the population size grows large. For the connections between finite and infinite player potential games, see Sandholm (2001). Also, note that in the continuum model, no single agent's action choices can alter the payoffs experienced by the others. While this property slightly simplifies the construction of the optimal price scheme, it is not essential – see Sandholm (2003b).

[25] When there are increasing returns to choosing different actions, coordination on any action can constitute a local welfare optimum, which implies that $\bar{F}$ cannot be concave. Indeed, if the common payoff to each action is increasing and convex in its level of use, and is independent of the use of other actions, then $\bar{F}$ is actually convex. In these cases, $(F - P^*, \mu)$ generally possesses multiple equilibria.

that are stochastically stable: in other words, only socially optimal behavior is observed in a non-negligible fraction of periods after a long enough history of play. Thus, the current paper and Sandholm (2003b) demonstrate the effectiveness of price schemes in two different classes of environments and on two different time scales. The results presented here show that when externalities are negative, price schemes can ensure efficient play within a moderate time span; those in Sandholm (2003b) show that if the planner is very patient, the restriction to negative externalities becomes unnecessary.

Appendix

A.1 Pricing and Evolution in Discrete Time

In this section, we show how our results can be extended to a discrete time setting. We do so using the simplest possible discrete time specification; other specifications are also possible. We suppose that time passes in discrete periods $t \in \mathbf{N}_0$. At the beginning of each period t, each agent chooses an action. At the end of the period, the planner observes the realized distribution of actions, and assigns prices to actions based on this information. Hence, if x_t is the realized action distribution in period t, agents obtain base payoffs of $F(x_t)$ and are charged prices of $P^*(x_t)$ during that period. We model the evolution of behavior using the difference equation

(DD) $\qquad z_{t+1} = z_t + s\, g(z_t).$

As before, g is a vector field that is admissible with respect to the game $(F - P^*,\ \mu)$. The step size parameter s represents the length of a single period; it determines the rate at which the population's behavior adjusts.

Intuitively, one expects that if the step size is sufficiently small, the evolution of behavior under the difference equation (DD) should approximate evolution under the differential equation (D). By formalizing this idea, we can establish the following analogue of Theorem 3.1.

Theorem A.1: *Suppose that the common payoff function $\bar{F}$ is concave, and fix any type distribution μ. Let g be a vector field that is admissible with respect to $(F - P^*, \mu)$. Then for each $\varepsilon > 0$, there exists an $\bar{s} > 0$ such that for any step size $s \in (0, \bar{s}]$,*

(i) Z_μ *is forward invariant under* (DD);

(ii) *The set* $\{z \in Z_\mu\colon \max_{\hat{z}} \overline{U}(\hat{z}) - \overline{U}(z) < \varepsilon\}$ *of* ε*-efficient states is globally stable under* (DD).

Theorem A.1 shows that our implementation results continue to hold in a discrete time framework so long as there is sufficient inertia in the behavior adjustment process.[26] The proof of this result is provided below. It follows immediately from Proposition 4.5 and the following result.

Theorem A.2: *Let* (U, μ) *be a potential game with concave,* C^2 *potential function* u*. Suppose that the vector field* g *is admissible with respect to* (U, μ)*. Then for each* $\varepsilon > 0$*, there exists an* $\bar{s} > 0$ *such that for any step size* $s \in (0, \bar{s}]$,

(i) Z_μ *is forward invariant under* (DD);

(ii) *The set* $Z_\mu^\varepsilon \equiv \{z \in Z_\mu\colon \max_{\hat{z}} u(\hat{z}) - u(z) < \varepsilon\}$ *is globally stable under* (DD).

The proof of Theorem A.2 requires two lemmas.

Lemma A.3: *Suppose that* $s \leq \frac{1}{\sqrt{2}K}$, where K *is the Lipschitz coefficient for* g*. Then* (DD) *is forward invariant on* Z_μ.

Proof: Since the directions of motion specified by g are always tangent to the space Z_μ, the only way that solutions to (DD) can leave Z_μ is if the transition from the current state z to the new state $\hat{z}(s) = z + sg(z)$ causes some component of $\hat{z}(s)$ to become negative. If $g(z) = 0$, then $\hat{z}(s) = z \in Z_\mu$, so this cannot occur. We therefore suppose that $g(z) \neq 0$. Define $\bar{s}(z) = \max\{s\colon \hat{z}(s) \in Z_\mu\}$. It is enough to show that $\bar{s}(z) \geq \frac{1}{\sqrt{2}K}$.

In two places below, we will use the following observation, which follows from the forward invariance condition (FI): if $y \in Z_\mu$ is a state with $y_{\theta,i} = 0$, then $g_{\theta,i}(y) \geq 0$.

Let (θ, i) be a type/strategy pair such that $\hat{z}_{\theta,i}(s) < 0$ when $s > \bar{s}(z)$. Then $g_{\theta,i}(z) < 0$, so the observation implies that $z_{\theta,i} > 0$. Define the state $\tilde{z} \in Z_\mu$ as follows:

---

26 In fact, even the assumption of inertia does not seem essential. By supposing that the step size s is small, we are able to show that under the price scheme P^*, the evolutionary process converges *monotonically* to the set of ε-efficient states. The assumption of inertia rules out large behavior adjustments that overshoot the set of welfare improving states. However, as long as overshooting does not occur too often and becomes rare when payoffs are nearly equalized, the population will tend to congregate at the ε-efficient states.

$$\tilde{z}_{\rho,j} = \begin{cases} z_{\rho,j} & \text{if } \rho \neq \theta; \\ z_{\theta,j} + \frac{1}{n-1} z_{\theta,i} & \text{if } \rho = \theta \text{ and } j \neq i; \\ 0 & \text{if } \rho = \theta \text{ and } j = i. \end{cases}$$

The state $\tilde{z}$ is the closest state to z on the (θ, i) boundary of Z_μ. Notice that

$$\| z - \tilde{z} \| = \sqrt{1 + \tfrac{1}{n-1}}\, z_{\theta,i} \leq \sqrt{2}\, z_{\theta,i}.$$

Thus, appealing again to the observation, we see that

$$| g_{\theta,i}(z) | \leq | g_{\theta,i}(z) - g_{\theta,i}(\tilde{z}) | \leq \| g(z) - g(\tilde{z}) \| \leq K \| z - \tilde{z} \| \leq \sqrt{2}\, K z_{\theta,i}.$$

Hence, if $s \leq \frac{1}{\sqrt{2}K}$, then $\hat{z}_{\theta,i}(s) = z_{\theta,i} + s\, g_{\theta,i}(z) \geq 0$. We thus conclude that $\bar{s}(z) \geq \frac{1}{\sqrt{2}K}$. ❑

Lemma A.4: *For all $\varepsilon > 0$, there exists a $c > 0$ such that if $s > 0$ is sufficiently small,*

(*i*) $u(z + sg(z)) - u(z) \geq -\frac{\varepsilon}{2}$ *for all $z \in Z_\mu$;*

(*ii*) $u(z + sg(z)) - u(z) \geq cs$ *for all $z \in Z_\mu - Z_\mu^{\varepsilon/2}$.*

Proof: Fix $z \in Z_\mu$. Taylor's theorem implies that for every $s \in (0, \frac{1}{\sqrt{2}K})$, there is an $\hat{s} \in [0, s]$ such that

$$u(z + sg(z)) - u(z) = s\left(\nabla u(z) \cdot g(z)\right) + s^2\left(\tfrac{1}{2} g(z)^T \nabla^2 u(z + \hat{s} g(z))\, g(z)\right).$$

Now the definition of potential and Lemma A.5 (in Section A.2 below) show that

$$(1) \qquad \begin{aligned} \nabla u(z) \cdot g(z) &= \sum_\theta \sum_i U_{\theta,i}(z)\, g_{\theta,i}(z) \\ &= n \sum_\theta Cov(g_\theta(z), U_\theta(z)) \end{aligned}$$

Thus, condition (PC) implies that

(2) $\nabla u(z) \cdot g(z) \geq 0$, with equality only when $g(z) = 0$.

Since u is C^2 on the compact set Z_μ, and since g is Lipschitz continuous on Z_μ, there is a constant $k_2 > -\infty$ such that

(3) $\qquad \frac{1}{2}g(z)^T\nabla^2 u(y)\, g(z) \geq k_2$ for all $y, z \in Z_\mu$.

Combining equations (1), (2), and (3), we see that $u(z + sg(z)) - u(z) \geq k_2 s^2$ for all $z \in Z_\mu$. Thus, if $s \leq \sqrt{\left|\frac{\varepsilon}{2k_2}\right|}$, statement (*i*) holds.

We now consider statement (*ii*). Condition (NC) tells us that $g(z) = 0$ only if z is a Nash equilibrium of (U, μ), while Proposition 4.1 (*ii*) implies that all Nash equilibria of (U, μ) maximize u on Z_μ, and so in particular lie in $Z_\mu^{\varepsilon/2}$. Since $Z_\mu^{\varepsilon/2}$ is the inverse image of an open set, it is open, and so $Z_\mu - Z_\mu^{\varepsilon/2}$ is compact. These observations, statement (2), and the continuity of ∇u and g imply that there is a $k_1 > 0$ such that

(4) $\qquad \nabla u(z)\cdot g(z) \geq k_1$ for all $z \in Z_\mu - Z_\mu^{\varepsilon/2}$.

Combining statements (1), (3), and (4), we conclude that if $s \leq \min\left\{\sqrt{\left|\frac{k_1}{2k_2}\right|}, \frac{1}{\sqrt{2K}}\right\}$, then

$$u(z + sg(z)) - u(z) \geq k_1\, s + k_2 s^2 \geq \tfrac{k_1}{2} s \text{ for all } z \in Z_\mu - Z_\mu^{\varepsilon/2}. \;\square$$

We now prove Theorem A.2. Suppose that $s > 0$ is small enough to satisfy the preconditions of Lemmas A.3 and A.4. Statement (*i*) of the theorem then follows immediately from Lemma A.3.

We therefore consider statement (*ii*) of the theorem. Since u is continuous, it is bounded on Z_μ. Hence, Lemma A.4 (*ii*) implies that any solution to (DD) must enter Z_μ^ε in a finite number of periods. Moreover, Lemma A.4 (*i*) shows that if $z_t \in Z_\mu^{\varepsilon/2}$, then $z_{t+1} \in Z_\mu^\varepsilon$, while Lemma A.4 (*ii*) shows that if $z_t \in Z_\mu^\varepsilon - Z_\mu^{\varepsilon/2}$, then $z_{t+1} \in Z_\mu^\varepsilon$. In conclusion, all solutions to (DD) enter the set Z_μ^ε within finite number of periods and never depart. This establishes statement (*ii*), completing the proof of the theorem. ■

A.2 Proofs Omitted from the Text

The Proof of Proposition 2.1

Every rest point of (D) is a Nash equilibrium of (F, μ) by condition (NC). To prove the converse, we introduce a simple lemma.

Lemma A.5: *If g satisfies condition* (FI), *then*

$$\sum_{\theta} Cov(g_\theta(z), U_\theta(z)) = \tfrac{1}{n}\sum_{\theta}\sum_{i} g_{\theta,i}(z) U_{\theta,i}(z).$$

Proof: Condition (FI) implies that the mass of agents of each type θ remains constant over time. Hence, for each $\theta \in \Theta$, we have that $\sum_i \dot{z}_{\theta,i} = \sum_i g_{\theta,i}(z) = 0$. We therefore conclude that

$$\begin{aligned}\sum_{\theta} Cov(g_\theta(z), U_\theta(z)) &= \sum_{\theta}\left(\tfrac{1}{n}\sum_{i}\big(g_{\theta,i}(z) - 0\big)\left(U_{\theta,i}(z) - \tfrac{1}{n}\sum_{j} U_{\theta,j}(z)\right)\right) \\ &= \tfrac{1}{n}\sum_{\theta}\sum_{i} g_{\theta,i}(z) U_{\theta,i}(z). \ \square\end{aligned}$$

Now, observe that condition (FI) and the definition of Nash equilibrium imply that $\{i \in S\colon g_{\theta,i}(z) < 0\} \subseteq \{i \in S\colon z_{\theta,i} > 0\} \subseteq \operatorname{argmax}_i U_{\theta,i}(z)$ for all $\theta \in \Theta$. But condition (FI) also implies that $\sum_i g_{\theta,i}(z) = 0$, so the inclusion yields $\sum_i g_{\theta,i}(z) U_{\theta,i}(z) \le 0$. Summing over θ, and applying Lemma A.5, we see that $0 \ge \sum_\theta \sum_i g_{\theta,i}(z) U_{\theta,i}(z) = n \sum_\theta Cov(g_\theta(z), U_\theta(z))$. We therefore conclude from condition (PC) that $g(z) = 0$. ■

The Proof of Proposition 6.3

In the two departure time congestion game, total common payoffs are given by

$$\begin{aligned}\bar{F}(x) &= \sum_{a,\tau} x_{a,\tau} F_{a,\tau}(x) \\ &= -\sum_{a}\left(x_{a,1}\sum_{\phi\in\Phi_a} c_\phi(u_{\phi,1}(x)) + x_{a,2}\sum_{\phi\in\Phi_a} c_\phi(u_{\phi,2}(x) + s_\phi(u_{\phi,1}(x)))\right) \\ &= -\sum_{\phi\in\Phi}\Big(u_{\phi,1}(x)\, c_\phi(u_{\phi,1}(x)) + u_{\phi,2}(x)\, c_\phi(u_{\phi,2}(x) + s_\phi(u_{\phi,1}(x)))\Big).\end{aligned}$$

Since $u_{\phi,1}$ and $u_{\phi,2}$ are linear in x, to prove that $\bar{F}$ is concave it is enough to show that

$$b_\phi(u, v) = u\, c_\phi(u) + v\, c_\phi(v + s_\phi(u))$$

is convex for all ϕ.

The proof of Proposition 6.1 shows that the first term of b_ϕ is convex. Let $\beta_\phi(u, v) = v\, c_\phi(v + s_\phi(u))$ denote the second term. Its Hessian matrix is given by

$$\nabla^2 \beta_\phi(u,v) = \begin{bmatrix} v\left((s'_\phi)^2 c''_\phi + s''_\phi c'_\phi\right) & s'_\phi\left(vc''_\phi + c'_\phi\right) \\ s'_\phi\left(vc''_\phi + c'_\phi\right) & vc''_\phi + c'_\phi \end{bmatrix}.$$

The eigenvalues of this matrix are

$$\tfrac{1}{2}\left(2c'_\phi + vc''_\phi(1+s'_\phi)^2 + vc'_\phi s''_\phi\right)$$
$$\pm\ \tfrac{1}{2}\sqrt{4(c'_\phi)^2(s'_\phi)^2 - s''_\phi(2vc'_\phi + v^2c''_\phi) - \left(2c'_\phi + vc''_\phi(1+s'_\phi)^2 + vc'_\phi s''_\phi\right)^2}\ .$$

Both eigenvalues are positive if the first term is larger than the second term. Multiplying both sides of this inequality by 2, squaring, and rearranging, we find that the desired inequality will hold if

$$2\left(2c'_\phi + vc''_\phi(1+s'_\phi)^2 + vc'_\phi s''_\phi\right)^2 + s''_\phi(2vc'_\phi + v^2c''_\phi) \geq 4(c'_\phi)^2(s'_\phi)^2$$

Since v, c'_ϕ, c''_ϕ, s'_ϕ, and s''_ϕ are all positive, when we expand the squared sum, every term in the resulting expression is positive. If we do this, and then throw away all terms on the left hand side except the initial one, we see that the previous equality will be true if

$$8(c'_\phi)^2 \geq 4(c'_\phi)^2(s'_\phi)^2.$$

But $s'_\phi \leq 1$ by assumption, so this inequality holds. Therefore, the Hessian $\nabla^2\beta_\phi(u,v)$ is positive semidefinite, the functions $\beta_\phi(u,\ v)$ and $b_\phi(u,\ v)$ are convex, and the function $\overline{F}$ is concave, completing the proof. ■

The Proof of Proposition 6.5

The total common payoffs in the computer network game are given by

$$\begin{aligned}\overline{F}(x) &= -\sum_{a,\tau} x_{a,\tau}\left(\sum_{\psi=\tau}^{\tau+l_a-1} c_a(u_\psi(x))\right) \\ &= -\sum_{\psi}\ \sum_{(a,\tau)\in\rho(\psi)} x_{a,\tau}\, c_a(u_\psi(x)).\end{aligned}$$

It is therefore enough to show that for each fixed time ψ, the function

$$G_\psi(x) = \sum_{(a,\tau)\in\rho(\psi)} x_{a,\tau}\, c_a(u_\psi(x)) = \sum_{(a,\tau)\in\rho(\psi)} x_{a,\tau}\, c_a\left(\sum_{(b,\xi)\in\rho(\psi)} w_b\, x_{b,\xi}\right)$$

is convex in x.

Since the function G_ψ does not depend on the number of agents x_0 who choose the outside option, we ignore this component of x, thereby viewing x as a point in $\mathbf{R}^{A\pi}$. We define the vector $v \in \mathbf{R}^{A\pi}$ and the function C: $\mathbf{R} \to \mathbf{R}^{A\pi}$ as follows:

$$v_{a,\tau} = \begin{cases} w_a & \text{if } (a,\tau) \in \rho(\psi); \\ 0 & \text{otherwise;} \end{cases}$$

$$C_{a,\tau}(u) = \begin{cases} c_a(u) & \text{if } (a,\tau) \in \rho(\psi); \\ 0 & \text{otherwise.} \end{cases}$$

Using this notation, we can express the gradient and the Hessian of G_ψ as

$$\nabla G_\psi(x) = C(u_\psi(x)) + \left(\sum_{(a,\tau)\in\rho(\psi)} x_{a,\tau}\, c'_a(u_\psi(x)) \right) v \quad \text{and}$$
$$\nabla^2 G_\psi(x) = C'(u_\psi(x))\, v^T + v\, C'(u_\psi(x))^T + \left(\sum_{(a,\tau)\in\rho(\psi)} x_{a,\tau}\, c''_a(u_\psi(x)) \right) v\, v^T,$$

where the superscript T denotes transposition.

To establish the proposition, it is enough to show that this Hessian matrix is positive semidefinite. Since the bandwidth vector w is strictly positive, $v \geq 0$, and since each cost function c_a is increasing, $C'(u_\psi) \geq 0$. Moreover, if $C'(u_\psi)$ is not the zero vector, then it and the vector v must have a common positive component, and so $C'(u_\psi(x)) \cdot v > 0$.

The following observation will help us complete the proof.

Observation: *Let $y, z \in \mathbf{R}^m$. If $y \cdot z \neq 0$, then the matrix $y z^T$ is diagonalizable. The eigenvalue 0 has multiplicity $m - 1$ and eigenspace $z_\perp$, while $y \cdot z$ is the remaining eigenvalue, corresponding to the eigenvector y.*

We consider each term of $\nabla^2 G_\psi(x)$ separately. If $C'(u_\psi)$ is the zero vector, then $C'(u_\psi)\, v^T$ is the zero matrix. If it is not, then the observation implies that the matrix $C'(u_\psi)\, v^T$ is diagonalizable, and that its only nonzero eigenvalue is $C'(u_\psi) \cdot v > 0$. Either way, $C'(u_\psi)\, v^T$ is positive semidefinite. The second term of $\nabla^2 G_\psi(x)$ is positive semidefinite as well, since it is just the transpose of the first. Finally, since each cost function c_a is convex, the summation that begins the last term of $\nabla^2 G_\psi(x)$ is positive, while the observation implies that the matrix $v\,v^T$ is positive semidefinite. Hence, $\nabla^2 G_\psi(x)$ is positive semidefinite, so G_ψ is convex and $\bar{F}$ is concave. This completes the proof of the proposition. ■

The Proof of Proposition 7.1

To begin, we extend our definition of separable games from Section 2 to allow continuous type distributions μ; further analysis of these games is presented in Appendix A.3. It will be convenient to describe behavior by specifying the distribution of actions within each subpopulation, where a subpopulation consists of all agents of a single type. We call the resulting map σ in the set $\Sigma = \{\sigma\colon \Theta \to X\}$ a *Bayesian strategy*; the vector $\sigma(\theta) \in X$ is the distribution of actions chosen by agents of type θ. We consider Bayesian strategies equivalent if they agree on a full μ measure set. The action distribution induced by σ is denoted $E_\mu \sigma = \int_\Theta \sigma(\theta)\, d\mu(\theta) \in X$, and payoffs for the separable game (F, μ) are defined by $U_{\theta,i}(\sigma) = F_i(E_\mu\sigma) + \theta_i$. A Bayesian strategy σ is a *Nash equilibrium* of (F, μ) if for μ almost all $\theta \in \Theta$, $\sigma_i(\theta) > 0$ implies that $i \in \text{argmax}_{j\in S}\, F_j(E_\mu\sigma) + \theta_j$.

Since the measure μ admits a density, Nash equilibria are almost surely strict. In particular, if we fix σ and define the disjoint sets $\Theta_i(\sigma) = \{\theta \in \Theta\colon\ F_j(E_\mu\sigma) + \theta_j > F_j(E_\mu\sigma) + \theta_j$ for all $j \neq i\}$, then (i) agents with types in $\Theta_i(\sigma)$ have a strict best response of i under σ, and hence $\sigma_i(\theta) = 1$ for all $\theta \in \Theta_i(\sigma)$, and (ii) $\mu(\cup_i \Theta_i(\sigma)) = 1$, so condition (i) captures the behavior of almost all types under σ.[27]

Now suppose that the common payoff game F is a potential game with potential function f. Given the results from the discrete case, one might expect to obtain a potential function for (F, μ) by summing the potential function and an expression capturing total idiosyncratic payoffs. To carry out this idea, we let $\iota\colon \Theta \to \Theta$ denote the identity function on Θ, and consider the following infinite dimensional optimization problem:

$$\text{(P)} \qquad \max_{\sigma}\ f(E_\mu\sigma) + \sum_i E_\mu \sigma_i \iota_i \qquad \text{subject to} \qquad \sigma_i \geq 0 \quad \text{for all } i \in S; \qquad \sum_i \sigma_i = 1.$$

In Appendix A.3 (Corollary A.8), we show that all solutions to (P) are Nash equilibria of (F, μ), and hence are almost surely strict.

---

[27] This is equality true because the set of θ such that $F(E_\mu\sigma) + \theta$ has two or more equal components has Lebesgue measure zero, and hence μ measure zero.

Fix a type profile $\tau \in \Theta^A$ with induced distribution μ^τ, and let $s \in \phi^*(\tau)$ be an efficient action profile under τ. Let $\sigma^{s,\tau} \in \Sigma$ be the Bayesian strategy induced by s and τ.[28] Since s is efficient given τ, a marshalling of definitions shows that $\sigma^{s,\tau}$ solves

$$\max_{\sigma} \; \bar{F}(E_{\mu^\tau}\sigma) + \sum_i E_{\mu^\tau}\sigma_i \iota_i \quad \text{subject to} \quad \sigma_i \geq 0 \quad \text{for all } i \in S; \qquad \sum_i \sigma_i = 1.$$

Consequently, $\sigma^{s,\tau}$ is a Nash equilibrium of $(\nabla\bar{F}, \mu^\tau) = (F - P^*, \mu^\tau)$, which implies that $\sigma^{s,\tau}$ is almost surely strict. Since μ almost every type has a strict preference, so does m almost every agent: if we define the disjoint sets $A_i = \{\alpha \in A\colon \; \tau(\alpha) \in \Theta_i(\sigma^{s,\tau})\}$, then $m(\cup_i A_i) = 1$, and $s(\alpha) = i$ for m almost all $\alpha \in A_i$.[29]

Formally, our mechanism implements a selection $\phi^{**}\colon \Theta^A \to S^A$ from the correspondence $\phi^*\colon \Theta^A \Rightarrow S^A$. To ensure that the mechanism satisfies the anonymity conditions (A1) and (A2), we require the selection to be invariant with respect to permutations of the agents' names: that is, $\sigma^{\phi^{**}(\tau),\tau} = \sigma^{\phi^{**}(\tau'),\tau'}$ whenever $\mu^\tau = \mu^{\tau'}$.

We now define the mechanism $(\hat{S}, T)$. Actually, the transfer function T is defined in the statement of the theorem, so we need only define the assignment function $\hat{S}$. We do so as follows: given any announcement profile $\hat{\tau}$ such that $\mu^{\hat{\tau}}$ admits a density, $\hat{S}$ specifies the assignment $\hat{s} = \phi^{**}(\hat{\tau})$. Otherwise, $\hat{S}$ specifies the constant assignment $\hat{s}(\alpha) = 1$ for all $\alpha \in A$.

We must verify that the mechanism has the properties stated in the theorem. It is clear from the definition of the transfer function that obedience is dominant at all Stage 2 decision nodes. It is also clear that if agents are truthful and obey their assignments, their behavior is efficient.

We next show that the mechanism is anonymous. Let $(\hat{\tau}, \hat{s} = \hat{S}(\hat{\tau}), s, t = T(\hat{\tau}, s))$ and $(\hat{\tau}', \hat{s}' = \hat{S}(\hat{\tau}'), s', t' = T(\hat{\tau}', s'))$ be two play sequences with the same announcement distribution $\mu^{\hat{\tau}} = \mu^{\hat{\tau}'}$. It is clear that conditions (A1) and (A2) hold when $\mu^{\hat{\tau}}$ does not admit a density. When $\mu^{\hat{\tau}}$ does admit a density, $\hat{s} = \phi^{**}(\hat{\tau})$ and $\hat{s}' = \phi^{**}(\hat{\tau}')$, so our restriction on ϕ^{**} requires that $\sigma^{\hat{s},\hat{\tau}} = \sigma^{\hat{s}',\hat{\tau}'}$. The argument four paragraphs above shows that $\sigma^{\hat{s},\hat{\tau}}$ is a Nash equilibrium of $(\nabla\bar{F}, \mu^{\hat{\tau}})$ and specifies a

---

28 Let G be the sigma algebra on A generated by τ, and define the measure m_i on G by $m_i(B) = m(B \cap \{s = i\})$. Then we can define $\sigma^{s,\tau}\colon \Theta \to X$ formally by letting $\sigma_i^{s,\tau}(\tau(\cdot))\colon A \to \mathbf{R}$ equal the Radon-Nikodym derivative dm_i/dm, since $\sigma_i^{s,\tau}$ satisfies $m_i(B) = E_m 1_B \sigma_i^{s,\tau}(\tau(\cdot))$ for all $B \in$ G.

29 To verify the second claim, note that since $\sigma_i^{s,\tau}(\tau(\cdot)) = 1$ on $A_i \in$ G, it follows that $m(A_i) = E_m 1_{A_i} = E_m 1_{A_i} \sigma_i^{s,\tau}(\tau(\cdot)) = m_i(A_i) = m(A_i \cap \{s = i\})$.

unique action for $\mu^{\hat{\tau}}$ almost every type announcement ($\sigma_i^{\hat{s},\hat{\tau}}(\theta) = 1$ whenever $\theta = \Theta_i(\sigma^{\hat{s},\hat{\tau}})$). It therefore uniquely determines the action assignments under $\hat{s}$ and $\hat{s}'$ for m almost every agent ($\hat{s}(\alpha) = i$ on $\{\alpha\colon \hat{\tau}(\alpha) \in \Theta_i(\sigma^{\hat{s},\hat{\tau}})\}$, and $\hat{s}'(\alpha) = i$ on $\{\alpha'\colon \hat{\tau}'(\alpha') \in \Theta_i(\sigma^{\hat{s},\hat{\tau}})\}$). Thus, for $m \times m$ almost all pairs (α, α'), $\hat{\tau}(\alpha) = \hat{\tau}'(\alpha')$ ($\in \Theta_i(\sigma^{\hat{s},\hat{\tau}})$) implies that $\hat{s}(\alpha) = \hat{s}'(\alpha')$ ($= i$). This is condition (A1), and this argument along with the definition of the transfers imply condition (A2).

To complete the proof, we need only show that given obedience in Stage 2, truth telling is weakly dominant in Stage 1. Suppose that agents besides α make announcements according to the profile $\hat{\tau}$. If $\mu^{\hat{\tau}}$ does not admit a density, then all agents are assigned to action 1 regardless of agent α's announcement, so all announcements are optimal for α. If $\mu^{\hat{\tau}}$ admits a density, then $\sigma = \sigma^{\phi^{**}(\hat{\tau}),\hat{\tau}}$ is a Nash equilibrium of $(F - P^*, \mu^{\hat{\tau}})$ with action distribution $x = x^{\phi^{**}(\hat{\tau})} = E_{\mu^{\hat{\tau}}}\sigma$. In particular, for m almost every agent α, we know that if agent α is of type $\tau(\alpha)$ and $\sigma_i(\tau(\alpha)) = 1$, then

$$F_i(x) - P_i^*(x) + \tau_i(\alpha) \geq F_j(x) - P_j^*(x) + \tau_j(\alpha) \text{ for all } j \in S.$$

If agent α announces his true type $\tau(\alpha)$, then he is assigned to action i, which the previous inequality reveals to be his optimal action given that others are obedient. (In contrast, if he announces some other type θ', he is assigned to an action j that is optimal for agents of type θ', but that may not be optimal for him.) Hence, if assignments are obeyed in Stage 2, it is optimal for α to truthfully reveal his type in Stage 1 regardless of his opponents' announcements. This completes the proof of the theorem. ■

A.3 Separable Games with Continuous Type Distributions

This section establishes the missing claim from the proof of Theorem 7.1: that a Bayesian strategy that solves program (P) is a Nash equilibrium of (F, μ). We begin our analysis by reviewing some concepts from functional analysis. Let B and B' be two Banach spaces, and let $l(B, B')$ denote the set of bounded linear operators from B to B'. If $\Lambda \in l(B, B')$, we let $\langle \Lambda, \hat{b} \rangle$ denote the value of Λ at $\hat{b} \in B$. Given a function $\Phi\colon B \to B'$, we call $D\Phi(b)$ the (*Fréchet*) *derivative* of Φ at b if

$$\left\| \left(\Phi(b + \hat{b}) - \Phi(b) \right) - \langle D\Phi(b), \hat{b} \rangle \right\|_{B'} \in o\left(\left\| \hat{b} \right\|_B \right) \text{ when } \left\| \hat{b} \right\|_B \text{ is sufficiently small.}$$

Let μ be an absolutely continuous probability distribution on $\Theta = \mathbf{R}^n$, where $n = \#S$. If ϕ is a function from $\mathbf{R}^n$ to $\mathbf{R}^k$, we define its L^2 norm by

$$\|\phi\| = \sqrt{\sum_{i=1}^{k} E_\mu (f_i)^2}\,,$$

and we let $L_k^2 = \{\phi\colon \mathbf{R}^n \to \mathbf{R}^k : \|\phi\| < \infty\}$. The *Riesz Representation Theorem* says that if $\Lambda \in l(L_k^2, \mathbf{R})$, then there exists a $\phi \in L_k^2$ such that $\langle \Lambda, \hat{\phi}\rangle = \sum_{i=1}^{k} E_\mu \phi_i \hat{\phi}_i$ for all $\hat{\phi} \in L_k^2$.

Let F be a common payoff game with potential function f. We study the separable game (F, μ), whose payoff functions $U_{\theta,i}(\sigma) = F_i(E_\mu\sigma) + \theta_i$ are defined for each Bayesian strategy $\sigma\colon \Theta \to X$, where $\Theta = \mathbf{R}^n$ and $X \subset \mathbf{R}^n$ is the simplex. (For further discussion of the definition of this game, see the proof of Theorem 7.1.)

Let ι denote the identity function on $\mathbf{R}^n$, and consider the program

$$\text{(P)} \qquad \max_{\sigma} f(E_\mu\sigma) + \sum_{i=1}^{n} E_\mu \sigma_i \iota_i \quad \text{subject to} \quad \sigma_i \geq 0 \quad \text{for all } i \in S; \qquad \sum_{i=1}^{n} \sigma_i = 1.$$

If we define $\phi\colon L_n^2 \to \mathbf{R}$ by $\phi(\sigma) = f(E_\mu\sigma) + \sum_{i=1}^{n} E_\mu \sigma_i \iota_i$, define $\psi\colon L_n^2 \to L_{n+1}^2$ by $\psi(\sigma) = (\sigma_1, \ldots, \sigma_n, \sum_{i=1}^{n} \sigma_i)$, and let $C = \{\pi \in L_{n+1}^2\colon\ \pi_i \geq 0$ for $i = 1, \ldots, n$ and $\pi_{n+1} = 0\}$, we can rewrite program (P) as

$$\max_{\sigma} \phi(\sigma) \ \text{ subject to } \psi(\sigma) \in C.$$

The Lagrangian for this program can be expressed as

$$\text{(L)} \qquad D\phi(\sigma) + \Lambda \circ D\psi(\sigma) = 0,$$

where the linear operator $\Lambda \in l(L_{n+1}^2, \mathbf{R})$ captures the shadow values associated with the constraints. It is easily verified that

$$\langle D\phi(\sigma), \hat{\sigma}\rangle = \sum_{i=1}^{n} E_\mu \left(\frac{\partial f}{\partial x_i}(E_\mu \sigma)\hat{\sigma}_i + \iota_i \hat{\sigma}_i \right) \text{ and that}$$
$$\langle D\psi(\sigma), \hat{\sigma}\rangle = (\hat{\sigma}_1, \ldots, \hat{\sigma}_n, \sum_{i=1}^{n} \hat{\sigma}_i) = \psi(\hat{\sigma})$$

for all $\hat{\sigma} \in L_n^2$. If we let $(\lambda_1, \ldots, \lambda_n, -\rho) \in L_{n+1}^2$ be the Riesz representation of Λ, then

$$\langle \Lambda \circ D\psi(\sigma), \hat{\sigma}\rangle = \sum_{i=1}^{n} E_\mu (\lambda_i - \rho)\hat{\sigma}_i.$$

Therefore, if we define $\check{1}$: $\mathbf{R}^n \to \mathbf{R}^n$ by $\check{1}(\theta) = (1, \dots, 1)$, we can express the Riesz representation of the Lagrangian (L), along with the appropriate complementary slackness conditions, as

(KT1′) $\dfrac{\partial f}{\partial x_i}(E_\mu \sigma)\check{1}_i + \iota_i = \rho - \lambda_i$;

(KT2′) $\lambda_i \sigma_i = 0$;

(KT3′) $\lambda_i \geq 0$.

for $i = 1, \dots, n$.

We can now state the auxilliary results we need to prove Theorem 7.1.

Proposition A.6: *Bayesian strategy $\sigma \in L_n^2$ is a Nash equilibrium of (F, μ) if and only if (σ, λ, ρ) satisfies* (KT1′), (KT2′), *and* (KT3′) *for some $(\lambda, \rho) \in L_{n+1}^2$.*

Proposition A.7: *If $\sigma \in L_n^2$ solves* (P), *then it satisfies* (KT1′), (KT2′), *and* (KT3′).

Corollary A.8: *If $\sigma \in L_n^2$ solves* (P), *then it is a Nash equilibrium of (F, μ).*

Proofs: The proof of Proposition A.6 is similar to that of Proposition 4.1, and Corollary A.8 follows immediately from the two propositions. We therefore need only prove Proposition A.7.

Let $\Sigma = \{\sigma \in L_n^2\colon\ \psi(\sigma) \in C\}$ be the set of Bayesian strategies. We say that $\hat{\sigma} \in L_n^2$ *points into Σ at σ*, denoted $\hat{\sigma} \in P(\Sigma, \sigma)$ if there exist a sequence $\{\sigma_k\} \subset \Sigma$ and a sequence $\{a_k\} \subset \mathbf{R}_+$ such that $a_k(\sigma_k - \sigma) \to \hat{\sigma}$. If we let $Z_i(\sigma) = \{\theta \in \mathbf{R}^n\colon\ \sigma_i(\theta) = 0\}$, then it is easily verified that

$$P(\Sigma, \sigma) = \{\hat{\sigma} \in L_n^2\colon\ \hat{\sigma}_k \geq 0 \text{ on } Z_i(\sigma) \text{ and } \textstyle\sum_{i=1}^n \hat{\sigma}_i = 0\},$$

and that this set is closed and convex.

Now let

$$P^-(\Sigma, \sigma) = \{\Lambda \in l(L_n^2, \mathbf{R})\colon\ \langle \Lambda, \hat{\sigma} \rangle \leq 0 \text{ for all } \hat{\sigma} \in P(\Sigma, \sigma)\}.$$

Theorem 1 of Guignard (1969) tells us that if σ solves program (P), then

$$D\phi(\sigma) \in P^-(\Sigma, \sigma).$$

Intuitively, this condition says that if σ solves (P), then the Riesz representation of $D\phi(\sigma)$, $(\ldots, \frac{\partial f}{\partial x_i}(E_\mu \sigma)\check{1}_i + \iota_i , \ldots)$, must form a weakly obtuse angle with every $\hat{\sigma}$ that points into Σ at σ.

If we represent $\Lambda \in l(L_n^2, \mathbf{R})$ by $\pi \in L_n^2$, we can compute that

$$\begin{aligned} P^-(\Sigma,\sigma) &= \{\Lambda\colon \textstyle\sum_{i=1}^n E_\mu \pi_i \hat{\sigma}_i \leq 0 \text{ for all } \hat{\sigma} \in P(\Sigma, \sigma)\} \\ &= \{\Lambda\colon \pi_i = \max_j \pi_j \text{ on } \mathbf{R}^n - Z_i(\sigma) \text{ for all } i\} \\ &= \{\Lambda\colon \pi = (\rho - \lambda_i, \ldots, \rho - \lambda_n), \text{ with } \lambda_i \sigma_i = 0 \text{ and } \lambda_i \geq 0 \text{ for all } i\}. \end{aligned}$$

Thus, $D\phi(\sigma) \in P^-(\Sigma,\sigma)$ if and only if (KT1′), (KT2′), and (KT3′) hold for some $(\lambda, \rho) \in L_{n+1}^2$. We therefore conclude that if σ solves (P), it satisfies the Kuhn-Tucker conditions. ■

References

Arnott, R., A. de Palma, and R. Lindsey (1990). "Economics of a Bottleneck," *Journal of Urban Economics* **27**, 111-130.

Arnott, R., A. de Palma, and R. Lindsey (1993). "A Structural Model of Peak Period Congestion: A Traffic Bottleneck with Elastic Demand," *American Economic Review* **83**, 161-179.

Beckmann, M., C. B. McGuire, and C. B. Winsten (1956). *Studies in the Economics of Transportation*. New Haven: Yale University Press.

Berger, U., and J. Hofbauer (2000). "The Nash Dynamics," unpublished manuscript, Universität Wien.

Brown, G. W. and J. von Neumann (1950). "Solutions of Games by Differential Equations," In H. W. Kuhn and A. W. Tucker, Eds., *Contributions to the Theory of Games I*. Annals of Mathematics Studies 24. Princeton: Princeton University Press.

Clarke, E. (1971). "Multipart Pricing of Public Goods," *Public Choice* **2**, 19-33.

Cocchi, R., S. Shenker, D. Estrin, and L. Zhang (1993). "Pricing in Computer Networks: Motivation, Formulation, and Example," *IEEE/ACM Transactions on Netowrking* **1**, 614-627.

Crémer, J., and C. Hariton (1999). "The Pricing of Critical Applications in the Internet," unpublished manuscript, Université de Toulouse 1.

Ely, J. C., and W. H. Sandholm (2001). "Evolution in Bayesian Games I: Theory," unpublished manuscript, Northwestern University and University of Wisconsin.

Groves, T. (1973). "Incentives in Teams," *Econometrica* **41**, 617-631.

Guignard, M. (1969). "Generalized Kuhn-Tucker Conditions for Mathematical Programming Problems in a Banach Space," *SIAM Journal on Control* **7**, 232-241.

Hofbauer, J. and K. Sigmund (1988). *Theory of Evolution and Dynamical Systems*. Cambridge: Cambridge University Press.

Jehiel, P. (1993). "Equilibrium on a Traffic Corridor with Several Congested Modes," *Transportation Science* **27**, 16-24.

MacKie-Mason, J. K., and H. Varian (1995). "Pricing the Internet." In Brian Kahin and James Keller, Eds., *Public Access to the Internet*. Cambridge: MIT Press.

Monderer, D., and L. S. Shapley (1996). "Potential Games," *Games and Economic Behavior* **14**, 124-143.

Pigou, A. C. (1920). *The Economics of Welfare*. London: Macmillan.

Rosenthal, R. W. (1973). "A Class of Games Possessing Pure Strategy Nash Equilibria," *International Journal of Game Theory* **2**, 65-67.

Sandholm, W. H. (2001). "Potential Games with Continuous Player Sets," *Journal of Economic Theory* **97**, 81-108.

Sandholm, W. H. (2002). "Evolutionary Implementation and Congestion Pricing," *Review of Economic Studies* **69**, 667-689.

Sandholm, W. H. (2003a). "Excess Payoff Dynamics, Potential Dynamics, and Stable Games," unpublished manuscript, University of Wisconsin.

Sandholm, W. H. (2003b). "Pigouvian Pricing and Stochastic Evolutionary Implementation," unpublished manuscript, University of Wisconsin.

Sandholm, W. H. (2004). "Evolution in Bayesian Games II: Applications," unpublished manuscript, University of Wisconsin.

Shenker, S., D. Clark, D. Estrin, and S. Herzog (1996). "Pricing in Computer Networks: Reshaping the Research Agenda," *Telecommunications Policy* **20**, 183-201.

Small, K., and Gomez-Ibañez, J. A. (1998). "Road Pricing for Congestion Management: The Transition from Theory to Policy," In K. J. Button and E. T. Verhoef, Eds., *Road Pricing, Traffic Congestion, and the Environment*: Chetlenham, UK: Edward Elgar.

Vickrey, W. (1961). "Counterspeculation, Auctions, and Competitive Sealed Tenders," *Journal of Finance* **16**, 8-37.

Vickrey, W. (1963). "Pricing in Urban and Suburban Transport," *American Economic Review Papers and Procedings* **52**, 452-465.

Vickrey, W. (1969). "Congestion Theory and Transport Investment," *American Economic Review Papers and Procedings* **59**, 251-260.

Weibull, J. W. (1996). In "The Work of John Nash in Game Theory," *Journal of Economic Theory* **69**, 153-185.