

DESIGNING STABLE MECHANISMS FOR ECONOMIC ENVIRONMENTS[†]

PAUL J. HEALY* AND LAURENT MATHEVET**

ABSTRACT. Motivated by past experimental results, we study the design of mechanisms that implement Lindahl or Walrasian allocations and whose Nash equilibria are dynamically stable for a wide class of adaptive dynamics. Previous research has focused on designing supermodular mechanisms; we show that supermodularity can generate vacuous stability predictions and focus instead on designing *contractive* mechanisms. We provide necessary and sufficient conditions for a mechanism to Nash implement Lindahl or Walrasian allocations, show that these conditions are inconsistent with the contraction property when individual strategy spaces are one-dimensional, and then show how to use additional dimensions to achieve dynamic stability.

Keywords: Mechanism design; implementation; stability; learning.

JEL Classification: C62; C72; C73; D02; D03; D51.

I INTRODUCTION

Mechanism design theory began as the evaluation of resource allocation processes in general equilibrium settings (Hurwicz, 1959). Following the impossibility results for dominant strategy implementation (Hurwicz, 1972; Gibbard, 1973; Satterthwaite, 1975), Groves and Ledyard (1977) showed by construction that optimal allocations can be achieved in economic environments with public goods when Nash equilibrium behavior is assumed. Subsequent mechanisms implementing Walrasian or Lindahl allocations in Nash equilibrium were provided by Hurwicz (1979b) and Walker (1981), among others. The problem of implementing optimal allocations in economic environments was solved, and the implementation literature moved on to study which objectives could be implemented in more general environments and other equilibrium concepts (Maskin, 1999; Moore and Repullo, 1988, e.g.).

[†]The authors thank Yaron Azrieli, Yan Chen, Vince Crawford, Federico Echenique, Guillaume Frechette, Matt Jackson, John Ledyard, Stephen Morris, Charlie Plott, William Sandholm, Steve Spear, Matt Van Essen, and Mark Walker for their helpful comments and conversations.

*Dept. of Economics, The Ohio State University, 1945 North High street, Columbus, OH 43210, U.S.A.; healy.52@osu.edu.

**Dept. of Economics, University of Texas, 1 University Station, C3100, Austin, TX 78712, U.S.A.; lmath@mail.utexas.edu.

Despite the impressive accomplishments of this literature, there remains a sizeable gap between the abstract mechanisms used to prove sufficiency results and the types of mechanisms that practitioners would be willing to apply in real-world settings. Even the relatively simple Groves-Ledyard, Hurwicz, and Walker mechanisms have not found significant field applications. Intuition alone is unlikely to close the gap between theory and applicability; instead, laboratory and field tests can provide key behavioral insights into the failure of these mechanisms, which can then be overcome by modifying the design process appropriately.

Early laboratory tests have already provided one persistent insight for mechanisms in economic environments: when the mechanism is played repeatedly, subjects' behavior is well approximated by myopic learning dynamics, rather than static equilibrium concepts (Chen and Plott, 1996; Chen and Tang, 1998; Chen and Gazzale, 2004; Healy, 2006). This literature supports the claim that play will converge to equilibrium in mechanisms that induce supermodular games, and behavior will fail to converge in mechanisms that do not (Chen, 2005, 2008).¹ In this paper we take the next step, incorporating this behavioral observation into the theory; we illustrate how one can design mechanisms that not only (fully) Nash implement Pareto optimal allocations in economies with or without public goods, but are also globally stable for a wide class of learning dynamics. We do this in four steps:

- (1) We argue that supermodularity is not an appropriate notion of stability for mechanisms in many admissible environments because its stability predictions are often vacuous. Instead, we require that mechanisms induce contractive games, which are games whose best response functions are contraction mappings. We show that contractive games are globally stable under a wide class of adaptive learning dynamics (Theorem 1).²
- (2) We identify strong necessary and sufficient conditions on the functional form of a mechanism that Nash implements Lindahl or Walrasian allocations when the type space is rich (Theorems 2–5).
- (3) We prove that if a mechanism asks each agent to transmit a one-dimensional real-valued message, then it cannot satisfy both the contractive property and the necessary condition for implementing Lindahl or Walrasian allocations; dynamic stability requires a larger strategy space (Theorem 6).

¹Supermodular games are games in which players' best responses are increasing functions of others' actions and strategy spaces are compact; see Topkis (1979), Topkis (1998), or Milgrom and Roberts (1990).

²The contraction-mapping approach was first suggested by Van Essen (2009b), who designs a mechanism that is both contractive and supermodular.

- (4) We show how to design mechanisms that are contractive and implement Lindahl or Walrasian allocations using two-dimensional strategy spaces for each agent. Examples of such mechanisms are provided (Theorems 7 and 8).

Our necessary and sufficient conditions for Nash implementing Lindahl or Walrasian allocations provide a new understanding about the types of mechanisms that can be used in general equilibrium settings. With one-dimensional message spaces, the necessary condition is quite strong: agents' announcements must represent individual purchases of the non-numeraire good at prices determined by others' messages. Thus, in the public goods setting, the choice of announcement is *equivalent* to the choice of the public good level taking prices as given. In this way, the mechanism must parallel the consumer's optimization problem given in the very definitions of Walrasian and Lindahl equilibrium. Sufficiency is obtained by assuming in addition that every possible Lindahl or Walrasian allocation can be reached by some announcement.³

The motivation for studying dynamic stability in mechanism design is manifold. First and foremost, this paper continues the dialogue between theory and data that hopefully will converge on acceptable mechanisms for real-world application. To aid in the continuation of this dialogue, we provide a *recipe* for designing stable mechanisms. New ingredients can be added as new behavioral regularities are discovered. In contrast, a single example of a stable mechanism is less desirable because these new behavioral regularities may render that particular mechanism ineffective.

Dynamic stability of equilibrium also has appeal independent from the existing experimental results. If equilibrium is arrived at through iterated applications of best-response in players' internal logic, or through iterations of pre-play communication, then stable equilibria are the most likely to arise. They are also the most robust to perturbations in opponents' logic or pre-play communication. Thus, we also view stability as a device to make static Nash implementation more robust.

It is important to study stability under the widest possible class of dynamics since experimental evidence suggests that the process of learning can vary dramatically from one environment to another. Existing work on economic environments by Vega-Redondo (1989); de Trenquallye (1989); and Kim (1993, 1996), for example, focus on particular learning dynamics that may or may not be descriptive in various settings.

³These necessity and sufficiency results were suggested by Brock (1980) (see also Groves and Ledyard, 1987), though not proved generally. A proof of the necessity result can be gleaned from the differential geometry arguments of Reichelstein and Reiter (1988). Our contribution is in making these conditions precise for both one-dimensional and higher-dimensional strategy spaces, providing transparent proofs for each, and showing the intuition behind the results.

Requiring full implementation—where every equilibrium maps to a desirable outcome, and vice-versa—is also important in a dynamic context, because it guarantees that agents do not settle on equilibria whose outcomes are not desirable. By dealing with full implementation, our paper presents an advantage over Mathevet (2007), who builds supermodular mechanisms in Bayesian environments but focuses on weak implementation and minimizing the size of the equilibrium set.

We focus attention on Nash implementation in economic environments, where the problem of stability is relatively long-standing. Muench and Walker (1983) study the Groves-Ledyard mechanism as the number of agents grows large. If the punishment parameter remains small then the mechanism becomes highly unstable under standard best-response dynamics. If the punishment parameter grows sufficiently fast then payoffs become arbitrarily ‘flat’. In either case, attainability and sustainability of equilibrium becomes a concern. If preferences are not quasilinear, then the Groves-Ledyard mechanism may have many undesirable equilibria even with small numbers of agents (Bergstrom et al., 1983); however, these equilibria may not be a concern since they are unstable and disappear when the punishment parameter is sufficiently large (Page and Tassier, 2004). Chen and Tang (1998) show that the mechanism also becomes supermodular in quasilinear environments with a large punishment parameter, though the critical requirement of a compact strategy space for supermodular games is neglected. As we argue below, supermodularity may not guarantee any meaningful stability properties, especially with an unbounded strategy space.

Regardless of its stability properties, a major drawback of the Groves-Ledyard mechanism is that it is not individually rational; agents’ final utility may be worse than that of their initial endowments. Hurwicz (1979a) proves that, under a mild continuity requirement, if one wants to implement Pareto optimal and individually rational outcomes in economic settings, then one must implement the Walrasian or Lindahl equilibrium allocations. From this view, the mechanisms of Hurwicz (1979b) and Walker (1981) that Nash-implement Lindahl allocations are preferable. We refer to such mechanisms as *Nash-Lindahl* mechanisms.

Unfortunately, the Nash-Lindahl mechanisms of Hurwicz and Walker are known to have poor stability properties. Experimental results (Chen and Tang, 1998; Healy, 2006) confirm that this severely hinders performance. Kim (1987) (following Jordan, 1986) shows that for a certain class of preferences, *all* Nash-Lindahl mechanisms must be unstable for at least one preference profile in the class. As mentioned above, Vega-Redondo

(1989); de Trenqualye (1989) and Kim (1993, 1996) all design Nash-Lindahl mechanisms that are stable for particular dynamics under various restrictions on preferences.

The first carefully-controlled laboratory experiments of Nash-Lindahl mechanisms were performed by Chen and Plott (1996).⁴ The subsequent experimental research (Chen and Tang, 1998; Chen and Gazzale, 2004; Healy, 2006) suggests that supermodularity is a sufficient condition for subjects to converge to Nash equilibrium.⁵ Supermodularity (following Milgrom and Roberts, 1990) in this context requires a compact strategy space, and implies monotone best-responses. Based on the learning result from Milgrom and Roberts (1990), Chen (2002) provides a family of “supermodular” Nash-Lindahl mechanisms, though their strategy spaces are not compact. Unfortunately, the compact strategy space assumption is important for the Milgrom-Roberts stability result.

Even with a compact strategy space the Milgrom-Roberts stability result may be of little use. It is easy to imagine a supermodular game with three equilibria: an unstable interior equilibrium and two stable corner equilibria at the smallest and largest points in the strategy space. The equilibria to which play will converge are the corner equilibria. Full implementation would require that these corner allocations are Lindahl allocations for every environment in which the corner equilibria exist. This is a very stringent condition; all existing Nash-Lindahl mechanisms avoid corner equilibria by using unbounded strategy spaces. But in that case, “supermodularity” does not guarantee stability, even if the equilibrium is unique. We illustrate these points in Section III and focus instead on designing mechanisms whose best response functions are contraction mappings. We show in Theorem 1 that this guarantees stability for a wide family of adaptive dynamics even when the strategy space is unbounded.

There has been comparatively little work on implementing Walrasian allocations, presumably because the competitive mechanism generally performs well, both in the field and in the laboratory. But there are cases where an alternative mechanism may be desirable. If the number of agents is small then the price-taking assumption becomes tenuous; a mechanism with a game-theoretic foundation is more likely to succeed. Dynamic stability then guarantees that adaptively-adjusting agents can still arrive at the

⁴Scherr and Babb (1975); Smith (1979); Harstad and Marrese (1983); and Tideman (1983) ran earlier experiments and many authors tested inefficient public goods processes such as the voluntary contributions mechanism (see Ledyard, 1995) but Chen and Plott (1996) were the first to test directly a theoretically optimal mechanism without modifications in a controlled laboratory setting.

⁵Unpublished experimental work by John Ledyard and Jasmina Arifovic show that supermodularity is not necessary; Healy (2006) suggests that a strictly weaker dominant diagonal condition may be sufficient.

Walrasian allocations. Furthermore, adaptive learning models in the competitive mechanism focus on tâtonnement-like adjustment processes where stability is, in general, not guaranteed (Scarf, 1960; Hirota, 1985); in fact this instability has been observed in the laboratory (Anderson et al., 2004; Crockett et al., 2009). We focus instead on designing game-theoretic mechanisms that are stable under a family of learning dynamics that is known to be reasonably descriptive. To our knowledge, the only other paper to focus on the design of dynamically stable (approximate) Walrasian mechanisms is Walker (1984).

Various methods for generating stability directly through the solution concept have also been studied. For example, dominant strategy equilibria are certainly dynamically stable for nearly any reasonable learning process. Unfortunately, standard impossibility results severely limit its applicability and the set of available mechanisms (Green and Laffont, 1977; Roberts, 1979). Furthermore, if the dominant strategy is not strict, then myopically-adapting agents may converge to undesirable Nash equilibria, as was observed in tests of the Vickrey-Clarke-Groves mechanism (Healy, 2006).

Abreu and Matsushima (1992) provide a mechanism that has a (unique) dominance-solvable equilibrium whose outcome is a lottery placing an arbitrarily large weight on the desired allocation. From a theoretical standpoint, the result is very strong; equilibrium in iteratively undominated strategies implies convergence of a wide class of learning dynamics to the equilibrium point. Yet their mechanism is of limited practical use. As the mechanism becomes more precise (placing more weight on the desired allocation) the dimensionality of the message space becomes infinite. Furthermore, laboratory tests of the mechanism (Sefton and Yavas, 1996, inspired by Glazer and Rosenthal, 1992) find that subjects do not move toward the equilibrium over 14 periods of play. This suggests that the speed at which iterated dominance is respected by learning is slow, or non-existent. These results are in line with the findings of McKelvey and Palfrey (1992); Stahl and Wilson (1995); Nagel (1995); and others, showing that subjects do not learn to respect iterated dominance arguments over relatively short time frames. These results apparently limit the applicability of mechanisms that rely on iterated-deletion solution concepts.⁶

Sandholm (2002) studies stable Nash implementation of efficient resource utilization in congestion games. These games admit a strictly concave potential function (Monderer and Shapley, 1996) which then guarantees a unique Nash equilibrium of a simple

⁶Bergemann and Morris (2009) consider rationalizable implementation, which is equivalent to iterated deletion of strictly dominated strategies when the strategy space is finite. They show that virtual implementation in iteratively undominated strategies requires a social choice function to select agents' favorite social choice outcome when preferences are identical.

pricing mechanism that is globally stable for a large family of adaptive dynamics. Sandholm (2005, 2007) extends this result to more general externality-abatement problems. A common feature of these environments is that agents must choose one action from a fixed set of available actions and externalities occur only through the number of people choosing each action. In settings with continuous levels of public and private goods and rich type spaces, it is typically not possible to use potential function techniques to guarantee stability.

Cabrales (1999) shows that, in the canonical mechanism of Maskin (1999), adaptive Markovian dynamics (placing positive probability on better responses to opponents' last-period strategies) will converge to and remain at the Nash equilibrium. Cabrales and Serrano (2009) extend this result, proving that a quasimonotonicity condition is necessary for implementation in the steady states of these dynamics; when a no-worst-alternative condition is also satisfied implementation can be achieved using a variation on the canonical mechanism. In experiments of the standard canonical mechanism with three agents and only three states of the world, Cabrales et al. (2003) find that the mechanism successfully implements the given social choice function in 68% of trials. Adding a fine to make the equilibrium strict raised the success rate to 80%. Whether or not these positive results would extend to larger environments remains an open question. Cabrales (1999) also shows that the Abreu-Matsushima mechanism is vulnerable to 'drift' when agents use adaptive Markovian dynamics, since the equilibrium features non-equilibrium best responses for each agent. Sjöström (1994) provides a mechanism that exactly implements most social choice functions assuming two rounds of iterated deletion of dominated strategies (one weak and one strict), but Cabrales and Ponti (2000) prove that monotone adaptive dynamics (in the sense of Samuelson and Zhang, 1992) converge to other Nash equilibria that do not result in the social choice outcome.

A small strand of the literature has focused on implementation with bounded rationality in games without repetition. Eliaz (2002) studies necessary and sufficient conditions for Nash implementation when some agents tremble in their strategy choice. Tumennasan (2008) finds that the quasimonotonicity and No Worst Alternative conditions (Cabrales and Serrano, 2009) are necessary and almost sufficient for implementation in logit quantal response equilibrium (McKelvey and Palfrey, 1995) as the noise in players' strategies is reduced to zero.

The structure of the paper is as follows: We introduce a two-good general equilibrium model and the basic definitions of implementation in Section II. In Section III

we demonstrate the weakness of supermodularity as a stability concept, formally define the contraction property, and demonstrate its stability properties. We then provide necessary and sufficient conditions for a mechanism to implement Lindahl or Walrasian allocations in Section IV; we first study the case of mechanisms with one-dimensional strategy spaces for each agent, and then generalize the results to higher-dimensional mechanisms. The procedure for designing contractive Lindahl and Walrasian mechanisms is given in Section V. We conclude with a brief discussion in Section VI.

II THE MODEL

Economic Environments

Consider a two-good general-equilibrium economy in which agents $i \in \{1, \dots, n\} = \mathcal{I}$ have endowments $\omega_i = (\omega_i^x, \omega_i^y) \in \mathbb{R}^2$, make net trades $z_i = (x_i, y_i) \in \mathbb{R}^2 - \{\omega_i\}$, and have preferences over net trades representable by a utility function $u_i(x_i, y_i | \theta_i)$ where $\theta_i \in \Theta_i$ identifies i 's type drawn from i 's type space Θ_i . We assume that for each i and θ_i u_i is increasing in x_i (the numeraire good) for all y_i and differentiable in both goods; in our discussion of stability, we restrict attention to the special case of quasilinear preferences where $u_i(x_i, y_i | \theta_i) = v_i(y_i | \theta_i) + x_i$. Let $\omega = (\omega_1, \dots, \omega_n)$, $z = (z_1, \dots, z_n)$ and $\theta = (\theta_1, \dots, \theta_n) \in \Theta = \times_i \Theta_i$ and let $p \in \mathbb{R}$ represent the price of the non-numeraire good, normalizing the numeraire price to one. A net trade vector $z \in \mathbb{R}^{2n}$ is *balanced* if $\sum_i z_i = 0$.

Unlike most general equilibrium models, ours does not restrict the feasible consumption set to the positive orthant. Since no mechanism can Nash implement Walrasian or Lindahl equilibria when boundary equilibria are permitted (see Hurwicz, 1979a or Jackson, 2001), it is necessary to rule out such equilibria either by allowing unbounded consumption bundles, or by restricting preferences so that boundary equilibria never obtain. The latter approach is more common (see Groves and Ledyard, 1977, for example), but may be incompatible with our notion of dynamic stability. We discuss this in Section V.

As specified, the model describes an exchange economy with purely private goods. But we can easily reinterpret the model to allow the second good to be a purely public good by making four changes: (1) every feasible net trade must be such that $y_i = y_j$ for all agents i and j , (2) $\omega_i^y = \omega_j^y$ for all i and j , (3) there is a single firm, capable of producing y units of the public good from $c(y)$ units of the numeraire, that aims to maximize the profit function $py - c(y)$, and (4) an allocation is now said to be balanced if $c(y) + \sum_i x_i = 0$. In this paper, we assume a constant marginal cost of production $\kappa > 0$ so that $c(y) = \kappa y$.

A Walrasian equilibrium of a private goods economy at type vector θ is a net trade vector z^* and a price p^* such that z^* is balanced and maximizes each $u_i(\cdot, \cdot | \theta_i)$ subject to the budget constraint $z_i p^* \leq 0$. Here z^* is referred to as a Walrasian equilibrium allocation.

A Lindahl equilibrium of a public goods economy is a net trade vector z^* (the Lindahl equilibrium allocation) and a vector of individual prices $p^* = (p_1^*, \dots, p_n^*)$ such that z^* is balanced, maximizes each $u_i(\cdot, \cdot | \theta_i)$ subject to the budget constraint $z_i p_i^* \leq 0$, and maximizes the firm's profit of $(\sum_i p_i^*)y - c(y)$.

Note that Lindahl equilibria are of the same dimensionality as Walrasian equilibria; the latter consists of $2n$ quantities and only one price while the former has $n + 1$ quantities but needs n prices.

Mechanisms & Implementation

A *social choice correspondence* $f : \Theta \rightarrow \mathbb{R}^{2n}$ maps type profiles into sets of net trades. For example, f might identify all Pareto optimal net trades for each θ (the *Pareto correspondence*), all Walrasian equilibrium allocations (the *Walrasian correspondence*), or, in a public goods setting, all Lindahl equilibrium allocations (the *Lindahl correspondence*).

A *mechanism* $\Gamma = (\mathcal{M}, h)$ consists of a *message space* $\mathcal{M} = \times_i \mathcal{M}_i$ and an *outcome function* $h : \mathcal{M} \rightarrow \mathbb{R}^{2n}$ mapping each message profile $m = (m_1, \dots, m_n)$ into a net trade vector z . A mechanism Γ is also called a *game form*; when combined with a particular type profile θ , the mechanism induces a well-specified game with strategy spaces \mathcal{M}_i for each i and utilities over strategy profiles given by

$$U_i(m | \theta_i) := u_i(h(m) | \theta_i).$$

We let

$$\beta_i(m | \theta_i) = \{m_i \in \mathcal{M}_i : U_i(m_i, m_{-i} | \theta_i) \geq U_i(m'_i, m_{-i} | \theta_i) \forall m'_i \in \mathcal{M}_i\}$$

represent i 's best-response correspondence and define $\beta = (\beta_1, \dots, \beta_n)$. The *Nash correspondence* $v : \Theta \rightarrow \mathcal{M}$ identifies the set of pure-strategy Nash equilibrium message profiles m^* of Γ at each θ ; formally, $v(\theta) = \{m \in \mathcal{M} : m \in \beta(m | \theta)\}$. A mechanism (\mathcal{M}, h) is said to *(Nash) implement* a social choice correspondence f if, for all $\theta \in \Theta$,

$$(1) \quad h(v(\theta)) = f(\theta).$$

We sometimes refer to (1) as *full implementation*; if $h(v(\theta)) \subseteq f(\theta)$ we say that Γ *weakly Nash implements* f and if $h(v(\theta)) \cap f(\theta) \neq \emptyset$ then Γ *partially Nash implements* f .

In the case of economic environments with two goods, the outcome function h can equivalently be written as a pair of functions of the form $x_i(m)$ and $y_i(m)$ for each $i \in \mathcal{I}$. In this paper, we consider mechanisms for which $\mathcal{M}_i \subseteq \mathbb{R}^{K_i}$ for some $K_i \in \{0, 1, \dots\}$ for each i . When \mathcal{M}_i has $J_i < K_i$ dimensions that enter into the y_i function, and $K_i - J_i$ dimensions that do not, then we may, for notation's sake, partition agent i 's strategy space into $\mathcal{M}_i = \mathcal{R}_i \times \mathcal{S}_i$ with $\mathcal{R}_i \subseteq \mathbb{R}^{J_i}$ and $\mathcal{S}_i \subseteq \mathbb{R}^{K_i - J_i}$. Letting $\mathcal{R} = \times_i \mathcal{R}_i$ and $\mathcal{S} = \times_i \mathcal{S}_i$ we have that $y_i : \mathcal{R} \times \mathcal{S}_{-i} \rightarrow \mathbb{R}$ and $x_i : \mathcal{R} \times \mathcal{S} \rightarrow \mathbb{R}$. In a public goods setting, it must be that $y : \mathcal{R} \rightarrow \mathbb{R}$ since no s_j may enter into y .

Given any mechanism with functions $y_i(m)$, it is without loss of generality that we can express i 's net trade of the numéraire as

$$(2) \quad x_i(m) = -q_i(m_{-i})y_i(m) - g_i(m)$$

so that the per-unit 'price' term q_i does not depend on m_i . Thus, any mechanism can be equivalently described by a list of functions of the form $q_i(m_{-i})$, $g_i(m)$, and $y_i(m)$ for each i . This formulation makes explicit the 'price' and 'penalty' components of $x_i(m)$.

III NOTIONS OF STABILITY

Past experimental results suggest that behavior in repeated mechanisms in economic environments is best described through learning dynamics rather than through repeated-game equilibria.⁷ Formally, a learning dynamic is a function $m : \{1, 2, \dots\} \rightarrow \mathcal{M}$ specifying a strategy profile $m(t)$ for each point in 'time' $t \in \{1, 2, \dots\}$.⁸ For example, a Cournot best-response dynamic would be a function satisfying $m(t) \in \beta(m(t-1)|\theta)$ for all $t > 1$.

If a single learning dynamic were known to describe human behavior in these dynamic settings, then a desirable mechanism would be one such that for all θ , the sequence of outcomes $h(m(t))$ converges quickly to $f(\theta)$. Since the debate over which dynamic best describes behavior is far from resolved, a more practical approach is to design mechanisms that induce games such that $h(m(t))$ converges to $f(\theta)$ for a large family of dynamics. The existing literature takes this approach by requiring a mechanism to induce *supermodular* games, which are known to have certain stability properties. We examine the shortcomings of this approach in the next subsection, and instead focus on designing mechanisms that induce *contractive* games, described in the subsequent subsection.

⁷We say 'suggest' rather than 'prove' here because it is possible—though in our view extremely unlikely—that the dynamic patterns observed in experiments are supported as repeated-game equilibria with complicated and unobserved off-path punishment strategies.

⁸This definition could be generalized to allow for continuous or finite time intervals.

Supermodular Games

Consider the game induced by some mechanism $\Gamma = (\mathcal{M}, h)$ at type profile θ . If each U_i is twice differentiable everywhere then, following Milgrom and Roberts (1990), this game is said to be *supermodular* if

- (1) $\partial^2 U_i / \partial m_{ik} \partial m_{il} \geq 0$ for all $i \in \mathcal{I}$ and $k \neq l \in \{1, \dots, K_i\}$,
- (2) $\partial^2 U_i / \partial m_{ik} \partial m_{jl} \geq 0$ for all $i \neq j \in \mathcal{I}$, $k \in \{1, \dots, K_i\}$, and $l \in \{1, \dots, K_j\}$, and
- (3) \mathcal{M}_i is a compact interval in \mathbb{R}^{K_i} for all i .

Properties (1) and (2) guarantee that β_i is increasing in others' strategies. If conditions (1) and (2) are satisfied but (3) is not then, we say the game is *open-supermodular*. Milgrom and Roberts (1990) prove that for every supermodular game, there is a smallest and largest Nash equilibrium, denoted here by \underline{m}^* and \overline{m}^* , and if a given learning dynamic is 'adaptive'—roughly, if it selects undominated strategies against a not-too-distant history of past play—then that dynamic will converge to the interval $[\underline{m}^*, \overline{m}^*]$. If the game has a unique equilibrium ($\underline{m}^* = \overline{m}^*$), then the equilibrium point is globally stable under all adaptive learning dynamics.

Unfortunately, the usefulness of this stability result is sometimes quite limited—a fact not well appreciated in the existing literature. Since the strategy space is required to be compact, then \underline{m}^* and \overline{m}^* may well be corner equilibria. In this case, the stability result is vacuous, placing no restrictions on the path of adaptive learning dynamics.

To illustrate, consider a simple two-player game with $\mathcal{M}_i = [-100, 100]$ for each i , parameter $\alpha \in \mathbb{R}$, and

$$U_i(m_i, m_j) = -(m_i - \alpha m_j)^2$$

for $i \in \{1, 2\}$. The unique best-response of agent i is $\beta_i(m) = \alpha m_j$. If $\alpha \geq 0$, then this game is clearly supermodular. If $\alpha \in (0, 1)$, then the game also has a unique Nash equilibrium at $m_0^* = (0, 0)$ and, by Milgrom and Roberts (1990), all adaptive dynamics converge to m^* .

If $\alpha > 1$, then the game is still supermodular but now it has three Nash equilibria: $\underline{m}^* = (-100, -100)$, $m_0^* = (0, 0)$, and $\overline{m}^* = (100, 100)$. In this case, the stability theorem of Milgrom and Roberts (1990) is vacuous. Furthermore, the interior equilibrium is unstable under most adaptive dynamics; a simple Cournot best-response process initiated away from m_0^* will converge monotonically to either \underline{m}^* or \overline{m}^* . If these corner equilibria are eliminated by setting $\mathcal{M}_i = \mathbb{R}$ for each i (making it an open-supermodular game), then m_0^* continues to be unstable and most adaptive dynamics will diverge. These games with $\alpha > 1$, despite their supermodularity, have a fundamental instability generated by

the *slopes* of their best response functions. Clearly, it is the absolute magnitude of the slopes—and not their sign—that determine dynamic stability.

Existing work on supermodular mechanism design fails to appreciate this issue.⁹ Chen (2002) designs an open-supermodular mechanism that Nash implements the Lindahl correspondence with $\mathcal{M}_i = \mathbb{R}^2$ for each i . Given the discussion above, however, it is unclear whether (open-)supermodularity will guarantee the convergence of adaptive learning dynamics to equilibrium strategy profiles; the mechanism may induce steeply-sloped best response functions that lead to divergent dynamics.

The following example clearly demonstrates that designing a supermodular (or open-supermodular) mechanism does not guarantee dynamic stability.

Example 1. Let $\mathcal{M}_i = \mathbb{R}^1$ for each i and suppose n is even. Take any quasilinear public goods environment of the form $v_i(y|\theta_i) + x_i$ where there exists some $B > 1$ such that $v_i'' \in (-B, -1/B)$ for every i and θ_i . Consider the mechanism with free parameter $\gamma > 0$ given by

$$q_i(m) = \begin{cases} \frac{\kappa}{n} - \gamma \sum_{j \neq \{i, i + \frac{n}{2}\}} m_j & \text{if } i \leq n/2 \\ \frac{\kappa}{n} + \gamma \sum_{j \neq \{i, i - \frac{n}{2}\}} m_j & \text{if } i > n/2 \end{cases},$$

$$g_i(m) \equiv 0,$$

and

$$y(m) = \sum_{i=1}^{n/2} m_i - \sum_{i=n/2+1}^n m_i$$

and for each i .

One can show that this mechanism implements the Lindahl correspondence using a proof very similar to that of Walker (1981). Calculating the slopes of the individual best-response functions, however, gives

$$\frac{\partial \beta_i}{\partial m_j} = \frac{\gamma}{-v_i''} - 1 \geq \frac{\gamma}{B} - 1$$

for each $j \notin \{i, i + n/2\}$ and $\partial \beta_i / \partial m_{i+n/2} = 0$. If $\gamma > B$, then β_i is non-decreasing in m_j for all θ_i and so the game is open-supermodular. If $\gamma > 2B$, however, then the slopes of the best response functions are larger than one, and the Cournot dynamics are unstable. Figure I shows a simulation of the best-response dynamics for the case of $n = 4$, $v_i'' \equiv 1$ for each i , $\kappa = 4$, and $\gamma = 2$. The dynamic diverges exponentially and is unstable.

⁹See Chen and Tang (1998); Chen (2002); and Healy (2006), for example, or Chen (2008) for a survey. Mathevet (2007) is an exception; he considers Bayesian implementation and studies a direct mechanism ($\mathcal{M}_i = \Theta_i$) with compact type spaces.

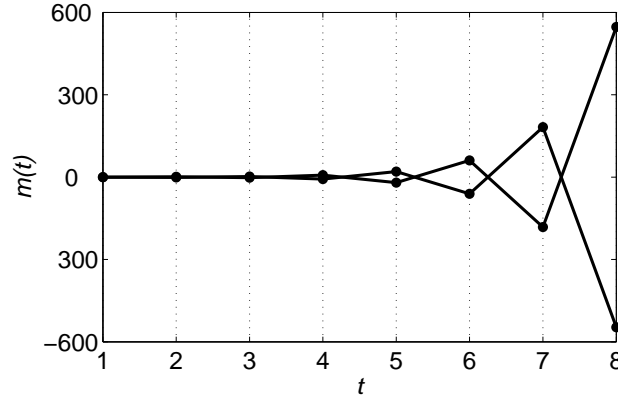


FIGURE I. Simulated best-response dynamics for an unstable supermodular mechanism.

Contractive Games

Given the difficulties with supermodularity we follow a different tack suggested first by Van Essen (2009b). In the simple examples above, what distinguishes stability from instability is whether or not $\alpha \in (-1, 1)$. This coincides with whether or not the best-response functions form a contraction mapping in \mathcal{M} . If a game's best response function is a contraction mapping, then its equilibrium will be unique, and we show below that a wide range of learning dynamics will converge to this unique equilibrium from any starting point. Therefore, the contraction property will serve as our notion of stability when designing stable mechanisms for economic environments.

Definition 1. If (\mathcal{M}, d) is a complete metric space with metric d , then a (single-valued) function $\beta : \mathcal{M} \rightarrow \mathcal{M}$ is a d -contraction mapping if there is some constant $\xi \in (0, 1)$ such that for all $m, m' \in \mathcal{M}$,

$$d(\beta(m), \beta(m')) \leq \xi d(m, m').$$

When the metric d is understood, we simply refer to β as a *contraction mapping*. When β describes the best-response function of a particular game, we say that the game is *contractive*. The following useful lemma provides a simple sufficient condition for a continuously differentiable function β to be a contraction mapping.

Lemma 1. If $\mathcal{M} \subseteq \mathbb{R}^K$ for some $K \in \{1, 2, \dots\}$, then a continuously differentiable function $\beta : \mathcal{M} \rightarrow \mathcal{M}$ is a contraction mapping if $\sup_{m \in \mathcal{M}} \|D\beta(m)\| < 1$, where $D\beta(m) = [\partial\beta_k/\partial m_l]_{k,l}$ is the $K \times K$ matrix of derivatives of β and $\|\cdot\|$ is any matrix norm.

The proof of this lemma follows easily from Conlisk, 1973. Using the absolute row-sum norm, for example, one can show that β is a contraction mapping, if $\sum_l |\partial \beta_k(m) / \partial m_l| < 1$ at every m for each dimension k . In games, this sufficient condition is somewhat simplified, because agent i 's best response function does not respond to changes in his own messages; therefore, the row sum condition given above only needs to sum over those dimensions $l \neq k$.

Note that supermodularity and contractiveness are logically unrelated; a game may satisfy one, both, or neither of these properties. In the linear two-player example above, supermodularity corresponds to $\alpha \geq 0$, while contractiveness corresponds to $\alpha \in (-1, 1)$.

Since mechanisms induce different games for different type profiles $\theta \in \Theta$, we must extend our definition of a contractive game when describing mechanisms:

Definition 2. Let (\mathcal{M}, d) be a complete metric space. A mechanism $\Gamma = (\mathcal{M}, h)$ with outcome function h is *d-contractive on Θ* , if for every $\theta \in \Theta$, the induced game with preferences $U_i(m|\theta_i)$ has a single-valued best-response function $\beta(\cdot, \theta) : \mathcal{M} \rightarrow \mathcal{M}$ that is a *d-contraction mapping*.

Again, we drop the reference to d when the metric is understood. Contractiveness is a strong property to require of a mechanism; by the Banach fixed point theorem, it guarantees the existence of a *unique* Nash equilibrium of Γ at θ . This equilibrium is globally stable under the Cournot best-response dynamic. This means that, if Γ also Nash implements some social choice function f , then the outcome $f(\theta)$ will in fact be realized in the limit, when agents' play is described by Cournot best-response.

Clearly, the processes that best describe dynamic human behavior are more complex and subtle than the simple Cournot best-response dynamic, so guaranteeing stability for a larger family of dynamics is desirable. In this vein, we provide a contraction-mapping analogue of the Milgrom-Roberts stability result for supermodular games: There is a family of *adaptive best response dynamics (ABR dynamics)* such that every dynamic in this family is globally stable in any game with a contractive best response function.

To describe ABR dynamics formally, let $H(t', t) = \{m(s) : t' \leq s < t\}$ denote the history of play from time t' up to (but not including) t and let m^* denote the unique Nash equilibrium of the game under consideration. Fix a metric d . For any $r \geq 0$ let $B(r|m^*) = \{m \in \mathcal{M} : d(m, m^*) \leq r\}$ be the closed ball with center m^* and radius r . Given any bounded set $\mathcal{M}' \subset \mathcal{M}$ define

$$B(\mathcal{M}') = \bigcap \{B(r|m^*) : \mathcal{M}' \subseteq B(r|m^*)\}$$

to be the smallest closed ball centered at m^* that includes \mathcal{M}' .

Definition 3. A learning dynamic $\{m(t)\}$ is an *adaptive best-response dynamic (ABR dynamic)* if $(\forall t')(\exists \hat{t} > t')(\forall t \geq \hat{t}), m(t) \in B(\beta(B(H(t', t))))$.

To understand this definition, consider first two points in time t' and t . Take the point $m' \in H(t', t)$ that is farthest from the equilibrium m^* , and consider *all* points in \mathcal{M} that are closer to the equilibrium than m' . Calculate the best response to each of those points, and among those calculated best responses, let m'' be the farthest from the equilibrium. The requirement that $m(t) \in B(\beta(B(H(t', t))))$ simply states that the date- t strategy cannot be farther from m^* than m'' . Thus, players observe history $H(t', t)$, form a ‘belief’ that the next profile will be in $B(H(t', t))$, and choose any profile that is either a best response to this belief, or is at least no farther from equilibrium than any best response to this belief.

The quantifiers then say that for any date t' , there is some later date \hat{t} , after which the dynamic ignores the history of play prior to t' . Thus, the effect of early strategies must eventually vanish.

The ABR family of dynamics encompasses most variants of Cournot dynamics and many processes with bounded memory, including the k -period best response dynamic suggested by Healy (2006) to be a reasonable approximation of dynamic behavior in repeated economic mechanisms.

Theorem 1. If a game is contractive then all adaptive best-response dynamics converge to the unique Nash equilibrium.

Formal proofs appear in the appendix.

If actual play of repeated mechanisms is well described by an ABR dynamic (as Healy, 2006 suggests) then Theorem 1 implies that one should design mechanisms that not only Nash implement desirable social choice functions, but are also contractive on Θ .

IV CHARACTERIZATIONS OF IMPLEMENTING MECHANISMS

Before adding stability to the list of mechanism design constraints, we first identify necessary and sufficient conditions on the functional form of a mechanism, if it is to Nash implement Walrasian or Lindahl allocations. The necessary condition is particularly useful in understanding how much freedom is available in the construction of new mechanisms and, therefore, how one might consider new constraints such as dynamic stability. These characterizations also have implications about the complexity that is necessary for a mechanism to Nash implement Walrasian or Lindahl allocations, complementing earlier results by Reichelstein and Reiter (1988).

One-Dimensional Mechanisms

In this section, we restrict attention to *one-dimensional* mechanisms where $\mathcal{M}_i = \mathbb{R}^1$ for each $i \in \mathcal{I}$. Our proofs rely heavily on differentiability arguments, so we also require that mechanisms have twice-continuously differentiable outcome functions.

Assumption 1 (Differentiability). For each agent i , the message space \mathcal{M}_i equals \mathbb{R}^1 and, for each message vector $m \in \mathcal{M}$, the functions x_i and y_i are twice continuously differentiable in m_i at m .

Consider now an agent i who believes the other agents will submit a message vector m_{-i} . Given a twice continuously differentiable mechanism with outcome functions $(y_i, q_i, g_i)_i$, agent i is able to trace out a continuously differentiable manifold of (x_i, y_i) pairs in \mathbb{R}^2 that he can unilaterally achieve by varying m_i , holding fixed m_{-i} . Identifying an agent's best-response message m_i^* in response to m_{-i} is then equivalent to identifying the point on this manifold that maximizes i 's utility.

Our next assumption explicitly rules out cases where agent i 's manifold becomes arbitrarily flat in (x_i, y_i) -space by requiring that $\partial y_i / \partial m_i$ is uniformly bounded away from zero. To the authors' knowledge, this does not rule out any existing mechanisms; most use linear functions such as $y_i(m) = \sum_j m_j$.

Assumption 2 (Responsive y_i). For each i there exists some $\varepsilon_i > 0$ such that for all $m \in \mathcal{M}$, $|\partial y_i(m) / \partial m_i| \geq \varepsilon$.

Under Assumption 2, each agent i 's manifold can now be thought of as the graph of a single-valued mapping from y_i into x_i . We denote this mapping by

$$\chi_i(y_i | m_{-i}) := x_i(y_i^{-1}(y_i | m_{-i}), m_{-i}),$$

where $y_i^{-1}(y_i | m_{-i})$ identifies the unique m_i such that $y_i(m_i, m_{-i}) = y_i$. This construction highlights the fact that each agent i can choose *any* level of $y_i \in \mathbb{R}$ because the function $y_i(\cdot, m_{-i})$ is bijective for all m_{-i} .

We show an example of $\chi_i(y_i | m_{-i})$ in Figure II. At the point m^* , the outcome $(x_i(m^*), y_i(m^*))$ is realized by agent i . As i differentially changes his message m_i , he differentially changes his allocation (x_i, y_i) along the graph of χ_i . The downward slope of this graph at m^* —which we label $P_i(m^*)$ —therefore serves as the *effective price* of y_i charged by the mechanism at m^* . Formally,

$$(3) \quad P_i(m) = -\frac{\partial x_i(m) / \partial m_i}{\partial y_i(m) / \partial m_i}.$$

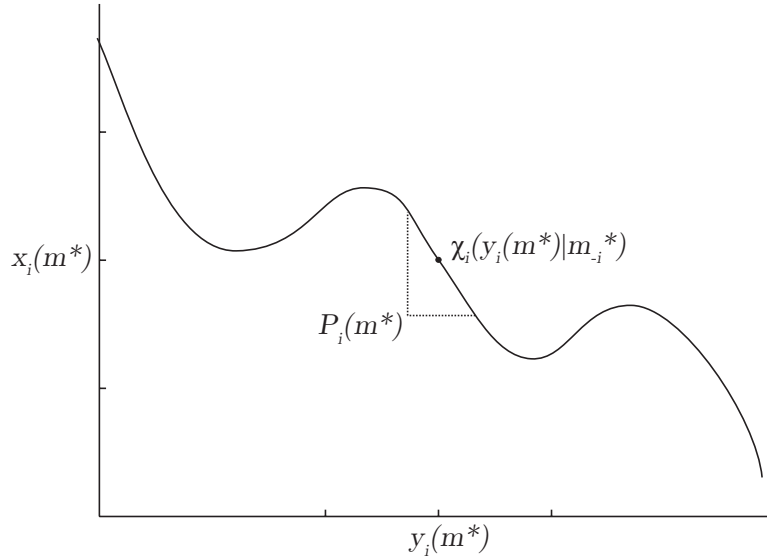


FIGURE II. The mapping $\chi_i(y_i|m_{-i})$ and the effective price $P_i(m)$ at m^* .

The effective price in a mechanism serves the same role *locally* as prices in a Walrasian or Lindahl equilibrium.

If m^* is a Nash equilibrium then the standard first-order conditions imply that

$$(4) \quad \frac{\partial u_i(x_i, y_i | \theta_i) / \partial y_i}{\partial u_i(x_i, y_i | \theta_i) / \partial x_i} = P_i(m^*),$$

so that the marginal rate of substitution between y_i and x_i equals the effective price of the mechanism at m^* .¹⁰

If this mechanism Nash implements a Walrasian or Lindahl equilibrium, then the marginal rate of substitution must also equal the Walrasian or Lindahl price. Thus, the effective prices at the equilibrium message profile m^* must match the Walrasian or Lindahl price for each environment θ . This leads to the following observation:

Observation (The Triple Tangency Property). If a mechanism Nash-implements Walrasian or Lindahl allocations then at any Nash equilibrium point m^* each agent's indifference curve in (x_i, y_i) -space must be tangent to both the mechanism's outcome manifold $\chi_i(\cdot | m_{-i}^*)$ and the Walrasian or Lindahl equilibrium price hyperplane.

The Triple Tangency Property is illustrated in panel (A) of Figure III; for type θ_i the point z_i is both a Nash equilibrium outcome and a Walrasian allocation at equilibrium

¹⁰Recall that \mathcal{M} is open so there are no boundary Nash equilibria.

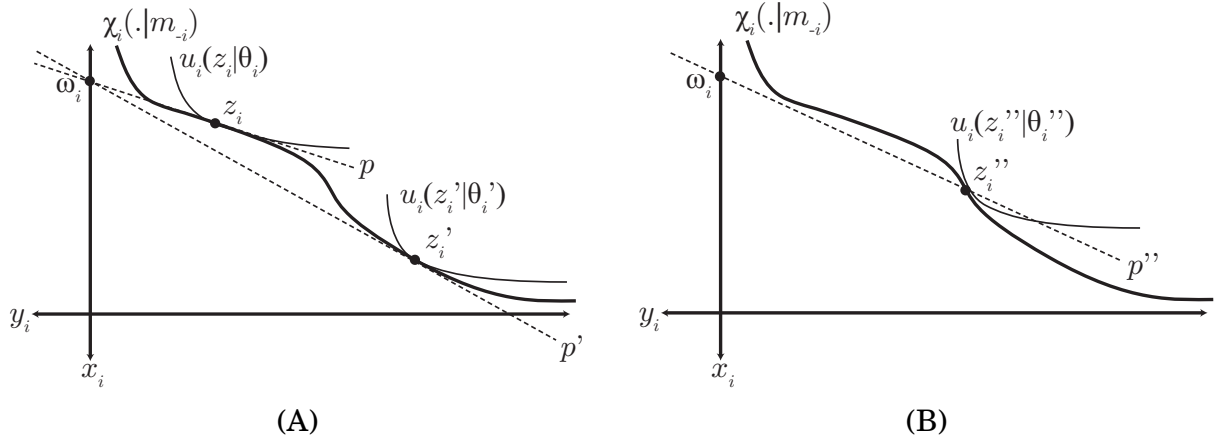


FIGURE III. (A) The Triple Tangency Property, and (B) a ‘bad’ Nash equilibrium.

price p . Similarly, z_i' is a Nash equilibrium point and a Walrasian allocation (at price p') for type θ_i' .

Now consider panel (B) if figure III. If the type space is sufficiently ‘rich’—meaning that every outcome z is a Nash equilibrium outcome for some environment—then there will exist some $\theta'' \in \Theta$ such that the point z_i'' is also a Nash equilibrium outcome. This must be a ‘bad’ Nash equilibrium, however, because i ’s indifference curve is not tangent to the price hyperplane passing through ω_i and so z_i'' cannot possibly be Walrasian or Lindahl allocation under θ'' . Because of this bad Nash equilibrium the mechanism represented by χ_i does not fully implement the Walrasian or Lindahl correspondence.

If the type space is rich, then *every* point z_i along χ_i can be made into a Nash equilibrium outcome by selecting an appropriate type profile. Thus, a mechanism that Nash implements the Walrasian or Lindahl correspondence must not allow for points such as z_i'' (in Figure III) where χ_i is not tangent to the hyperplane passing through z_i'' and ω_i . But this can only be accomplished if χ_i is linear and passes through ω_i . This observation generates our key necessary condition.

Assumption 3 (Rich Type Space). $\nu(\Theta) = \mathcal{M}$.¹¹

Theorem 2 (Necessity). Under Assumptions 1, 2, and 3, if a mechanism $\Gamma = (\mathcal{M}_i, q_i, g_i, y_i)_i$ weakly Nash implements the Walrasian or Lindahl correspondence with $\mathcal{M}_i = \mathbb{R}^1$ for each i then for every $i \in \mathcal{I}$ and every $m \in \mathcal{M}$,

$$(5) \quad x_i(m) \equiv -q_i(m_{-i})y_i(m)$$

¹¹Assumption 3 may appear unsatisfactory because it places restrictions on the equilibrium set rather than on the primitives of the model; in an online appendix we provide two linked assumptions on the primitives—one for the type space and one for the mechanism—that together imply Assumption 3.

and

$$(6) \quad \sum_i q_i(m_{-i}) = \kappa,$$

so that q_i does not depend on m_i and $g_i(m) \equiv 0$.

Theorem 2 is stated for weak implementation, but obviously applies to full implementation as well. This theorem gives a strong but intuitive result: If a mechanism is to Nash implement the Walrasian or Lindahl correspondence, then each agent's message-choosing problem in the mechanism (taking others' messages as fixed) must be identical to the quantity-choosing problem in an exchange economy when prices are taken as given. In the case of a mechanism, the quantity y_i is chosen indirectly through the choice of m_i , and the 'price' is determined endogenously as a function of m_{-i} . In exchange economies, agents choose y_i directly and face exogenously-given prices.

Conversely, if an agent has the ability to change both his chosen quantity *and* his per-unit price then such a mechanism cannot fully implement the Walrasian or Lindahl allocations.

For the case of public goods economies, compare Theorem 2 with the mechanisms of Groves and Ledyard (1977), Walker (1981), and Tian (1990). All three are one-dimensional mechanisms in which q_i depends only on m_{-i} , but the Groves-Ledyard mechanism has a non-trivial penalty function g_i while the latter two do not. Consequently, Walker's and Tian's mechanisms Nash implement the Lindahl correspondence while the Groves-Ledyard mechanism does not (its outcomes are Pareto optimal but not Lindahl).

For private goods economies, Theorem 2 is in fact excessively strong. If no agent is allowed to affect their own per-unit price, if all agents must have the same price at every equilibrium message, and if every message is an equilibrium message for some type profile, then the only admissible price function q_i is a constant function that depends on *no* agents' reports. But clearly such a mechanism cannot fully implement the Walrasian correspondence on a rich type space, so we arrive at a contradiction. This proves the following corollary:

Corollary 1. Under Assumptions 1–3 there does not exist a one-dimensional mechanism that Nash implements the Walrasian correspondence.

Corollary 1 was first proven by Reichelstein and Reiter (1988) using substantially different mathematical techniques. They provide a tight lower bound on the dimensionality of a mechanism that Nash implements the Walrasian correspondence. This lower

bound is larger than the number of agents, ruling out the one-dimensional case and proving the corollary.

Finally, we show that for public goods environments the necessary condition of Theorem 2 becomes sufficient for weak implementation of the Lindahl correspondence. Full implementation is achieved if for each Lindahl equilibrium there is some message $m' \in \mathcal{M}$ that maps to it.

Assumption 4. For every $(x_i^*, y_i^*, p_i^*)_i \in \mathbb{R}^{2n+1}$ that is a Walrasian or Lindahl allocation for some $\theta \in \Theta$ there is a message $m' \in \mathcal{M}$ such that

$$(x_i(m'), y_i(m'), q_i(m'))_i = (x_i^*, y_i^*, p_i^*)_i.$$

Note that Assumption 4 is slightly stronger than requiring $f(\Theta) \subseteq h(\mathcal{M})$ because it also requires that every possible Lindahl price be achievable by the q_i functions.

Theorem 3 (Sufficiency). In a public goods environment, if a mechanism Γ satisfies Assumptions 1 and 2 and equations (5) and (6) of Theorem 2 then Γ weakly Nash implements the Lindahl correspondence. If, in addition, Γ satisfies Assumption 4 then Γ fully Nash implements the Lindahl correspondence.

Thus, under Assumptions 1–4, equations (5) and (6) are necessary and sufficient for full implementation of the Lindahl correspondence.

The proof of sufficiency for weak implementation is intuitive: Since $y(m_i, m_{-i})$ is bijective in m_i , choosing m_i is equivalent to choosing y with $q_i(m_{-i})$ fixed. Equations (5) and (6) guarantee that the mechanism's outcome manifolds (the graphs of χ_i) are linear and pass through the endowment. If χ_i also passes through the Lindahl allocation then χ_i is uniquely determined and must exactly correspond to the Lindahl budget hyperplane defined by the Lindahl price p_i^* . Choosing m_i along χ_i is thus identical to choosing y along the budget hyperplane and so the budget-constrained optimal choice of y must be equal to the mechanism-constrained optimal choice of $y(\cdot, m'_{-i})$.

Assumption 4 guarantees that any Lindahl equilibrium allocation can be reached by some message $m' \in \mathcal{M}$ with $q_i(m'_{-i})$ equalling the Lindahl price for each i . Because the Lindahl equilibrium allocation is budget-constrained optimal, y is bijective, and the linear mechanism mimics this budget constraint, the m'_i mapping to $y(m'_i, m'_{-i})$ must be a best response for agent i . Thus, every Lindahl allocation is a Nash equilibrium and full implementation is achieved.

To demonstrate the gap between weak implementation and full implementation, consider the equal-tax voluntary contribution mechanism, where $\mathcal{M}_i = \mathbb{R}^1$ for each i , $y(m) =$

$\sum_i m_i$ and $x_i(m) = -\kappa y(m)/n$ (see Groves and Ledyard, 1980 or Healy, 2006). The hypotheses of the first part of Theorem 3 and satisfied, so this mechanism weakly Nash implements the Lindahl correspondence. But Assumption 4 fails at any θ that has a Lindahl equilibrium $(x_i^*, y_i^*, p_i^*)_i$ with $p_i^* \neq p_j^*$ for some i, j (which is true generically) because $q_i(m_{-i}) = q_j(m_{-j})$ for every $m \in \mathcal{M}$. In these environments the mechanism has no Nash equilibrium and full implementation fails.

Higher-Dimensional Mechanisms

When the strategy space of each agent has multiple dimensions, recall that we can further break apart a mechanism by distinguishing those strategy space dimensions that affect the quantity of the non-numeraire good from those that do not. Specifically, we let $\mathcal{M}_i = \mathcal{R}_i \times \mathcal{S}_i$, where, for each i , $\mathcal{R}_i \subseteq \mathbb{R}^{J_i}$ represents those dimensions that affect $y_i(r, s_{-i})$ and $\mathcal{S}_i \subseteq \mathbb{R}^{K_i - J_i}$ be those dimensions that do not.

With the partitioning of the strategy spaces into \mathcal{R}_i and \mathcal{S}_i we can modify equation 2 slightly and write any mechanism's numeraire outcome function as

$$(7) \quad x_i(r, s) = -q_i(r, s)y_i(r, s_{-i}) - g_i(r, s).$$

(Again, in a public goods environment y depends only on r .) Unlike equation 2, this formulation allows the 'price' term q_i to depend on agent i 's message. We will show, however, that along the equilibrium set of announcements agent i cannot affect q_i , even though he may be able to affect q_i at points off the equilibrium set.

We now reformulate our previous assumptions for the case of multiple dimensions.

Assumption 1' (Differentiability). For each agent i and each message vector $m \in \mathcal{M}$ the functions x_i and y_i are twice continuously differentiable in every dimension of m_i .

Assumption 2' (Responsive y_i). For each i there exists some $\varepsilon_i > 0$ such that for all $(r, s) \in \mathcal{M}$ and all dimensions $k \in \{1, \dots, J_i\}$, $|\partial y_i(r, s_{-i})/\partial r_{ik}| \geq \varepsilon$.

As for Assumption 3 with extra dimensions, it becomes overly restrictive to assume that $v(\theta) = \mathcal{M}$ because the dimensionality of $v(\Theta)$ may be strictly less than that of \mathcal{M} . Specifically, for any Γ we can rule out two types of messages as messages that can never be Nash equilibria of Γ for any $\theta \in \Theta$.

The first type of message $m = (r, s)$ that cannot be an equilibrium are those for which s is not a best response given r . Because s only enters into the determination of the numeraire, and utility is strictly increasing in the numeraire, the mechanism designer can identify those values of s (if any) that are best responses to each r . Formally, it must

be that

$$s_i \in \sigma_i(r, s_{-i}) := \arg \max_{s'_i \in \mathcal{S}_i} x_i(r, s'_i, s_{-i})$$

for each i , or

$$s \in \sigma(r) := \{s^* \in \mathcal{S} : (\forall i \in \mathcal{I}) s_i^* \in \sigma_i(r, s_{-i}^*)\}.$$

Thus, $\sigma(r)$ represents the pure-strategy equilibria of a ‘transfer-maximizing game’ in which agents pick s_i to maximize x_i given r . If $s \notin \sigma(r)$ then the pair (r, s) cannot be a Nash equilibrium of the mechanism for any θ .

The second type of messages that cannot be equilibria are those for which some agent has different effective prices along different dimensions of his strategy space. Formally, for each agent i and dimension $k \in \{1, \dots, J_i\}$ define the effective price along dimension k at message (r, s) by

$$P_{ik}(r, s) := -\frac{\partial x_i(r, s)/\partial r_{ik}}{\partial y_i(r, s_{-i})/\partial r_{ik}}$$

and note that by the same first-order argument as in the one-dimensional case, it must be that $P_{ik}(r, s)$ equals i ’s marginal rate of substitution between y_i and x_i at any equilibrium point (r, s) . Therefore, if a point (r', s') is such that $P_{ik}(r', s') \neq P_{il}(r', s')$ then (r', s') cannot be a Nash equilibrium.

Given these two restrictions, we now define

$$(8) \quad \mathcal{M}^* := \{m \in \mathcal{M} : (\forall i \in \mathcal{I})(\forall k, l \in \{1, \dots, J_i\}) P_{ik}(m) = P_{il}(m)\} \\ \cap \{m = (r, s) \in \mathcal{M} : s \in \sigma(r)\}$$

to be the set of ‘candidate equilibrium’ points in \mathcal{M} . The natural extension of our rich type space assumption (Assumption 3) is that every candidate equilibrium is an equilibrium for some environment.

Assumption 3’. $v(\Theta) = \mathcal{M}^*$.¹²

Finally, we define *regular* candidate equilibrium points for which our theorem will apply. Let \mathcal{R}^* be the projection of \mathcal{M}^* onto \mathcal{R} .

Definition 4. A candidate equilibrium $(r^*, s^*) \in \mathcal{M}^*$ is *regular* if $\sigma(r)$ is locally threaded by some differentiable function $\zeta = \times_i \zeta_i$ for each i ; formally (r^*, s^*) is regular if for each i there is some open set $\mathcal{R}_i^0 \subset \mathcal{R}_i^*$ containing r_i^* and a differentiable function $\zeta_i : \mathcal{R} \times \mathcal{S}_{-i} \rightarrow \mathcal{S}_i$ such that $\zeta_i(r^*, s_{-i}^*) = s_i^*$ and $(r'_i, r_{-i}^*, \zeta_i(r'_i, s_{-i}^*), s_{-i}^*) \in \mathcal{M}^*$ for all $r'_i \in \mathcal{R}_i^0$.

¹²As with Assumption 3, we provide in an online appendix two linked assumptions—one on the type space and one on the mechanism—that together imply Assumption 3’.

A candidate equilibrium (r^*, s^*) is regular if differential deviations in r_i can always be accompanied by a differential change in s_i such that the joint deviation does not lead to a strategy profile outside of \mathcal{M}^* . We refer to ζ_i as i 's *adjustment function*. An example of a non-regular equilibrium would be one for which a differential change in r_i leads to a point at which $\sigma(r)$ is empty or has a jump discontinuity.

We now prove a higher-dimensional analog to Theorem 2.

Theorem 4 (Necessity). Suppose a mechanism $\Gamma = (\mathcal{M}_i, x_i, y_i)_{i \in \mathcal{I}}$ Nash implements the Lindahl or Walrasian correspondences and satisfies Assumptions 1', 2', and 3'. Writing the mechanism as

$$x_i(r, s) = -q_i(r, s)y_i(r, s_{-i}) - g_i(r, s),$$

it must be the case that for every regular point $(r^*, s^*) \in \mathcal{M}^*$ with adjustment functions $(\zeta_i)_i$,

$$\frac{dq_i(r^*, \zeta_i(r^*, s_{-i}^*), s_{-i}^*)}{dr_{ik}} = 0 \quad \forall i \in \mathcal{I}, k \in \{1, \dots, J_i\}$$

and

$$g_i(r^*, s^*) = 0.$$

Thus, higher dimensional mechanisms may allow agents to affect their own prices and face non-trivial penalty functions, but the penalty function must equal zero on the equilibrium set, and each agent's price must not change as the agent unilaterally changes r_i and adjusts s_i appropriately. At off-equilibrium or non-regular equilibrium points, however, we derive no restrictions on the shape of the mechanism. It is this freedom that allows us to introduce global stability properties into a mechanism. Intuitively, one should be able to take a mechanism satisfying the restrictions of Theorem 4 and alter the mechanism on $\mathcal{M} \setminus \mathcal{M}^*$ so that any adaptive dynamic process that wanders off of \mathcal{M}^* will eventually return back to the appropriate point in \mathcal{M}^* , restoring the equilibrium.

Finally, we identify sufficient conditions for Nash implementation of the Walrasian or Lindahl correspondence that will be useful in constructing stable mechanisms.

Assumption 4'. For every $(x_i^*, y_i^*, p_i^*)_i \in \mathbb{R}^{2n+1}$ that is a Walrasian or Lindahl allocation for some $\theta \in \Theta$ there is a message $m' \in \mathcal{M}^*$ such that

$$(x_i(m'), y_i(m'), q_i(m'))_i = (x_i^*, y_i^*, p_i^*)_i.$$

Theorem 5 (Sufficiency). If Γ is a mechanism satisfying Assumptions 1' and 2' and for every i ,

$$(1) \quad q_i(r, s) \equiv q_i(r_{-i}, s_{-i})$$

- (2) $\sum_i q_i(r_{-i}, s_{-i}) = \kappa$ for all $(r, s) \in \mathcal{M}^*$,
- (3) $g_i(r, s) \geq 0$ for every $(r, s) \in \mathcal{M}$, and
- (4) $g_i(r, s) = 0$ if $s \in \sigma(r)$,

then Γ weakly implements the Walrasian or Lindahl correspondence. If, in addition, Γ satisfies Assumption 4' then Γ fully Nash implements the Walrasian or Lindahl correspondence.

Note that Theorem 3 is a direct corollary of Theorem 5 for the case where $\mathcal{M}_i = \mathbb{R}^1$ for each i ; we prove only the latter in the appendix.

V CONTRACTIVE MECHANISMS

From this point forward, we restrict attention to concave, quasilinear preferences of the form $u_i(x_i, y_i | \theta_i) = v_i(y_i | \theta_i) + x_i$ with $v_i'' < 0$. In order to establish stability results under our assumptions, however, we must place bounds on the concavity of agents' preferences. By way of contradiction, suppose instead that v_i'' is arbitrarily close to zero. Nash implementing a Walrasian or Lindahl allocation requires the first order condition $v_i'(y_i(m^*) | \theta_i) = q_i(m^*)$ to be satisfied at the equilibrium point m^* . Consider any differential change in m_j for some $j \neq i$ that alters q_i . Agent i 's best response to this change must alter v_i' by an equal amount to restore the first-order condition. When v_i'' is very small, however, this requires a very large shift in $y(m)$, which can only be accomplished by a large change in m_i .¹³ Since the response in m_i is larger than the original shift in m_j , the mechanism cannot be contractive.

Proposition 1. Suppose preferences are quasilinear and Assumptions 1–3 (or 1'–3') are satisfied. If for every $\varepsilon < 0$ there is some i , θ_i , and y such that $v_i''(y | \theta_i) \in (\varepsilon, 0)$ then no mechanism that implements Walrasian or Lindahl allocations is contractive on Θ .

Given Proposition 1, we must explicitly assume that the concavity of v_i is bounded away from zero if we are to construct contractive mechanisms.¹⁴ Recall that we use unbounded feasible consumption sets in our model of the economy. Proposition 1 shows that this is necessary when studying stability; the alternative assumption that indifference curves do not cross the consumption set boundaries (Groves and Ledyard, 1977) would require v_i'' to approach zero at some y for every i and θ .

¹³Clearly there is an offsetting effect if $\partial y_i / \partial m_i$ or $\partial y_i / \partial m_j$ is large near m^* . But a rich type space requires that this offsetting effect be present at every point m . No one y function can compensate for every possible 'flat spot' in v_i' .

¹⁴These bounds are inconsistent with Assumption 6A in the appendix but, depending on the mechanism, may or may not be consistent with Assumptions 3 or 3'.

Assumption 5. For all types $\theta \in \Theta$, all agents i have quasilinear preferences of the form $v_i(y_i|\theta_i) + x_i$ where $v'_i > 0$ and there is some $B > 0$ such that $v''_i \in (-B, -1/B)$.

Contractive One-Dimensional Mechanisms

We now show that there cannot exist a mechanism with one-dimensional strategy spaces ($\mathcal{M}_i = \mathbb{R}^1$ for each i) that Nash implements the Lindahl or Walrasian correspondence under our maintained assumptions.

Theorem 6. Under Assumptions 1–5, there does not exist a mechanism with $\mathcal{M}_i = \mathbb{R}^1$ for each i that is both contractive and Nash implements the Lindahl or Walrasian correspondence.

Note that proof for the Walrasian correspondence is trivial since there does not exist *any* one-dimensional mechanism that implements the Walrasian allocations.

Inspection of the proof reveals that Theorem 6 holds true even if v''_i can take any value in $(-\infty, 0)$; the bounds on v''_i from Assumption 5 are needed in the sequel to generate higher-dimensional mechanisms that are contractive.

Contractive Mechanisms for the Lindahl Correspondence

This above impossibility result for single-dimensional mechanisms forces us to consider higher-dimensional mechanisms when seeking contractive implementation of the Lindahl correspondence. We proceed by taking any two-dimensional mechanism (where $m_i = (r_i, s_i) \in \mathbb{R}^2$ for each i) of the form $y(r) = \sum_i r_i$ with $q_i(r_{-i}, s_{-i})$ and $g_i(r, s)$ unspecified. The first-order condition that $\partial U_i / \partial r_i = \partial U_i / \partial s_i = 0$ for any $m^* \in \mathcal{M}^*$ allows us to use the Implicit Function Theorem to derive closed-form expressions for the slope of the various best-response functions. If $\rho_i(r_{-i}, s_{-i})$ is i 's best response value of r_i and $\sigma(r_{-i}, s_{-i})$ is i 's best response value of s_i then, for example, the slope of ρ_i with respect to r_j is given by

$$(9) \quad \frac{\partial \rho_i}{\partial r_j} = \left(\frac{\partial^2 U_i}{\partial r_i \partial s_i} \frac{\partial^2 U_i}{\partial s_i \partial r_j} - \frac{\partial^2 U_i}{\partial s_i^2} \frac{\partial^2 U_i}{\partial r_i \partial r_j} \right) / \left(\frac{\partial^2 U_i}{\partial r_i^2} \frac{\partial^2 U_i}{\partial s_i^2} - \left(\frac{\partial^2 U_i}{\partial r_i \partial s_i} \right)^2 \right).$$

Using Lemma 1 (and the row-sum matrix norm), the mechanism is contractive if

$$(10) \quad \sum_{j \neq i} \left(\left| \frac{\partial \rho_i}{\partial r_j} \right| + \left| \frac{\partial \rho_i}{\partial s_j} \right| \right) < 1$$

and

$$(11) \quad \sum_{j \neq i} \left(\left| \frac{\partial \sigma_i}{\partial r_j} \right| + \left| \frac{\partial \zeta_i}{\partial s_j} \right| \right) < 1.$$

Thus, designing a contractive mechanism amounts to satisfying the sufficiency conditions of Theorem 5 and then verifying that the particular functional forms of y_i , q_i , and g_i are such that (10) and (11) are satisfied for every $\theta \in \Theta$.

We now present one example of a mechanism satisfying these conditions for Lindahl implementation. Let $\mathcal{M}_i = \mathcal{R}_i \times \mathcal{S}_i$ for each i with $\mathcal{R}_i = \mathcal{S}_i = \mathbb{R}^1$, choose $\delta > 0$, and set

$$(12) \quad y(r) = \sum_i r_i,$$

$$(13) \quad q_i(m_{-i}, s_{-i}) = \frac{\kappa}{n} + \frac{1}{\delta} (r_{i-1} - r_{i+1}) + \delta \frac{n-1}{n^2} \left(s_{i-1} - \frac{1}{n} r_{i+1} \right),$$

and

$$(14) \quad g_i(r, s) = \frac{1}{2} \left(s_i - \frac{1}{n} r_{i+1} \right)^2 + \frac{\delta}{2} \left(s_{i-1} - \frac{1}{n} r_i \right)^2.$$

Theorem 7. The mechanism defined by equations (12)–(14) fully Nash implements the Lindahl correspondence. Under Assumption 5, it is contractive on Θ when $\delta \geq Bn^2$.

That this mechanism fully implements the Lindahl correspondence follows from Theorem 5 because $g_i \geq 0$ with $g_i = 0$ when $s \in \sigma(r)$, q_i depends only on m_{-i} , $\sum_i q_i = \kappa$ if $s \in \sigma(r)$, y is bijective in r_i , and for every Lindahl equilibrium there is some message $m' \in \mathcal{M}$ that maps to the Lindahl equilibrium allocation and prices. The contractive property of the mechanism is proven by checking that equations (10) and (11) are satisfied when $\delta \geq Bn^2$; this is done in the appendix.

One downside of this mechanism is that it fails to balance the budget at certain off-equilibrium message profiles. This occurs both because the penalty terms (g_i) may all be strictly positive and the price terms (q_i) may sum to something other than the marginal cost. Both of these sources of imbalance can be corrected, however, by adding appropriate terms to the penalty functions. For the case of $n \geq 4$, this can be done by taking the g_i function from equation (14) and modifying it to equal

$$(15) \quad \hat{g}_i(r, s) = g_i(r, s) + \Delta_i(r, s)$$

where

$$\Delta_i(r, s) = -\frac{1+\delta}{2} \left(s_{i+1} - \frac{1}{n} r_{i+2} \right)^2 - \frac{\delta}{n} \left(s_{i+1} - \frac{1}{n} r_{i+2} \right) \left(\sum_{j \neq i} r_j \right) - \frac{\delta}{n} \left(s_{i+2} - \frac{1}{n} r_{i+3} \right) r_{i+1}.$$

Corollary 2. If $n \geq 4$ and $\delta \geq Bn^2$, then the mechanism defined by equations (12)–(15) fully Nash implements the Lindahl correspondence, is contractive on Θ , and is budget balanced for all $m \in \mathcal{M}$.

Chen (2002) provides a parameterized mechanism such that for large parameter values, the mechanism induces an open-supermodular game (see Section III) with boundedly-concave quasilinear preferences. Van Essen (2009a) shows that Chen’s mechanism is also contractive for some parameters, and Van Essen (2009b) provides a mechanism that is contractive for large parameter values and induces a ‘supermodular’ game with an open strategy space. Van Essen’s mechanism is similar to Chen’s, and they both have the advantage of simplicity. However, none of them is balanced out of equilibrium, and our method for correcting imbalance by adding terms to the penalty function will not apply.¹⁵ To our knowledge, the mechanism in (12)–(15) is the first contractive mechanism that is also budget balanced out of equilibrium. Although our example mechanism is somewhat complex—especially with the budget-balancing adjustment—our impossibility result for one-dimensional mechanisms (Theorem 6) suggests that complexity cannot be substantially improved.

Which contractive mechanism should be used in a given field application may well depend on which issues the designer deems most relevant, and further laboratory experiments may help to understand the relevant trade-offs. Our approach above also shows how other contractive mechanisms can be constructed to satisfy whatever additional requirements one might wish to impose.

Contractive Mechanisms for the Walrasian Correspondence

The process of designing a contractive Walrasian mechanism is nearly identical to the process of designing a contractive Lindahl mechanism, though the exact functional forms obviously must differ. Again we employ the Implicit Function Theorem approach and search for y_i , q_i , and g_i functions satisfying the sufficiency conditions of Theorem 5 and the contraction conditions (10) and (11) for each $\theta \in \Theta$.

¹⁵This is because the added terms in i ’s penalty function would necessarily depend on i ’s message, thereby affecting the best responses

The example mechanism we provide is two-dimensional with $\mathcal{M}_i = \mathbb{R}_i \times \mathcal{S}_i = \mathbb{R}^2$ for each i . The outcome functions are defined by

$$(16) \quad y_i(r, s_{-i}) = \left(r_i - \frac{1}{n-1} \sum_{j \neq i} r_j \right) - \left(s_{i+1} - \frac{\delta-1}{\delta} r_i \right),$$

$$(17) \quad q_i(r_{-i}, s_{-i}) = \frac{1}{\sqrt{\delta}} \left(s_{i+1} + \frac{\delta-1}{\delta} \frac{1}{n-1} r_{i+1} \right),$$

and

$$(18) \quad g_i(r, s) = \left(s_i - \frac{\delta-1}{\delta} \frac{1}{n-1} \sum_{j \neq i} r_j \right)^2$$

for each i .

Theorem 8. The mechanism defined by equations (16)-(18) fully Nash implements the Walrasian correspondence. For $\delta > B^2$, it is contractive on Θ .

Since $\sum_i y_i = 0$, the mechanism is balanced in the non-numeraire good both in and out of equilibrium. Out of equilibrium, however, the mechanism may not be balanced in the numeraire. Our ‘trick’ of balancing the Lindahl mechanism does not apply here; it is not known whether there exists a contractive Walrasian mechanism that is also budget balanced out of equilibrium.

VI DISCUSSION

From a theoretical perspective these newly-constructed mechanisms have nearly all the features one might ask; they implement Pareto optimal and individually rational allocations for a wide range of economic environments, they are dynamically stable for a large family of adaptive learning dynamics, and the individual message spaces are of minimal dimension necessary for dynamic stability. The mechanism for public goods environments also balances the budget both in and out of equilibrium.

Extending our results successfully to an environment with multiple, simultaneous public goods decisions seems unlikely without strong assumptions about the strength of complementarities between the various public goods. Even with additively separable preferences over the multiple public goods the problem is made difficult by the fact that several public goods levels are simultaneously changing; requiring contractiveness in each good separately is no longer sufficient for stability, and requiring contractiveness for all goods simultaneously becomes a strong condition as the number of goods is increased. Thus, we conjecture that stable mechanisms may exist for small numbers

of simultaneous public goods problems when complementarities are restricted, but the limits of existence may not extend far beyond the one-good case studied here.

The theorems in this paper make heavy use of the rich type space assumption. Sufficiently weakening this assumption opens the door for mechanisms to have q_i depend on r_i or g_i to be non-zero, which in turn will make dynamic stability an easier requirement to satisfy. For example, with only two possible type profiles (each with a unique Lindahl equilibrium) the Triple Tangency Property only needs to be satisfied at two points; away from those two points the mechanism can be ‘bent’ arbitrarily to satisfy the desired stability properties. As the type space grows this flexibility clearly diminishes.

The obvious next step for future research is to return to the lab with these newly-constructed mechanisms to understand what additional requirements they should be asked to satisfy. Perhaps bounds on mechanism complexity or the limits on the magnitude of off-equilibrium punishments will be identified as the next important factor for the theory to incorporate. Eventually these mechanisms can be field-tested on a small scale and perhaps the theory will be refined further as a result. Regardless of the exact path the process takes, it is only through a continuing dialogue between theory and experiments that practical mechanisms can be constructed for real-world application.

APPENDIX

Proof of Theorem 1

The proof follows by induction. Pick a starting time t_0 . By definition of an ABR dynamic, for each point in time t_n there exists some later point in time $t_{n+1} > t_n$ such that for all $t \geq t_{n+1}$, $m(t) \in B(\beta(B(H(t_n, t))))$. For each $n \in \{0, 1, 2, \dots\}$ let $\mathcal{M}_n = H(t_n, t_{n+1})$ be the history of play from t_n to t_{n+1} .

For any metric d on \mathcal{M} , any set $\mathcal{M}' \subseteq \mathcal{M}$, and any point $m' \in \mathcal{M}$ the d -Hausdorff distance between \mathcal{M}' and the singleton set $\{m'\}$ is given by

$$h_d(\mathcal{M}', m') = \sup_{m \in \mathcal{M}'} d(m, m').$$

Therefore, for any set $\mathcal{M}' \subseteq \mathcal{M}$, $h_d(\mathcal{M}', m^*) = h_d(B(\mathcal{M}'), m^*)$. Thus,

$$\begin{aligned} \xi h_d(\mathcal{M}_0, m^*) &= \xi h_d(B(\mathcal{M}_0), m^*) \\ &\geq h_d(\beta(B(\mathcal{M}_0)), m^*) \\ &= h_d(B(\beta(B(\mathcal{M}_0))), m^*) \\ &\geq h_d(\mathcal{M}_1, m^*), \end{aligned}$$

where the first inequality comes from the contraction property of β and the last inequality follows from the fact that $\mathcal{M}_1 \subset B(\beta(B(\mathcal{M}_0)))$. Taking any n and $n + 1$, we can use a similar argument to show that $\xi h_d(\mathcal{M}_n, m^*) \geq h_d(\mathcal{M}_{n+1}, m^*)$. Therefore, for all n ,

$$\xi^n h_d(\mathcal{M}_0, m^*) \geq h_d(\mathcal{M}_n, m^*),$$

which implies that the sequence \mathcal{M}_n converges to $\{m^*\}$, and so any ABR dynamics converges to m^* .

Q.E.D.

Proof of Theorems 2 and 4

Theorem 2 follows from Theorem 4 so we only prove the latter here.

For any $\theta \in \Theta$ let $p_i(\theta)$ be agent i 's price for good y_i at the Walrasian or Lindahl equilibrium for environment θ . For any $m \in \nu(\Theta)$ let $\phi(m) \in \Theta$ identify an environment θ for which m is an equilibrium. Thus, $p_i(\phi(m))$ is the Walrasian or Lindahl price that must be charged to agent i in the environment $\phi(m)$. Pick any regular equilibrium point $m^* = (r^*, s^*)$ in \mathcal{M}^* and, for notational simplicity, let $y_i^* = y_i(r^*)$ and $x_i^* = x_i(m^*)$. The proof then follows from three important observations that must be true at m^* for each $i \in \mathcal{I}$:

- (1) Because m^* is a Nash equilibrium for some $\theta \in \Theta$ the following first-order condition is satisfied for each $k \in \{1, \dots, J_i\}$:

$$(19) \quad \frac{\partial u_i(x_i^*, y_i^* | \theta_i)}{\partial y_i} \frac{\partial y_i(r_i^*)}{\partial r_{ik}} = \frac{\partial u_i(x_i^*, y_i^* | \theta_i)}{\partial x_i} \left[-\frac{\partial x_i(r^*, s^*)}{\partial r_{ik}} \right].$$

- (2) If m^* maps to a Walrasian or Lindahl equilibrium for some $\theta \in \Theta$ then it must be that the transfers collected by the mechanism equals the transfers of the numéraire required by the Walrasian or Lindahl equilibrium:

$$(20) \quad x_i(r^*, s^*) = -p_i(\phi(r^*, s^*))y_i(r^*).$$

- (3) If m^* maps to a Walrasian or Lindahl equilibrium for some $\theta \in \Theta$ then the Walrasian or Lindahl price must equal the marginal rate of substitution of y_i in terms of x_i :

$$(21) \quad \frac{\partial u_i(x_i^*, y_i^* | \theta_i) / \partial y_i}{\partial u_i(x_i^*, y_i^* | \theta_i) / \partial x_i} = p_i(\phi(r^*, s^*)).$$

Dividing both sides of (19) by $\partial u_i/\partial x_i$, inserting equation (21), and rearranging gives

$$(22) \quad \frac{\partial x_i(r^*, s^*)}{\partial r_{ik}} = -p_i(\phi(r^*, s^*)) \frac{\partial y_i(r^*)}{\partial r_{ik}}.$$

for each i and k .

Since (r^*, s^*) is a regular equilibrium point, differential changes in r_i accompanied by the requisite change in s_i lead to other points in \mathcal{M}^* at which the above equations hold. Now take the total derivative of (20) with respect to r_i (allowing for the adjustment in s_i); since s_i maximizes x_i the envelope theorem guarantees that $dx_i/dr_{ik} = \partial x_i/\partial r_{ik}$ and so the total derivative is

$$(23) \quad \frac{\partial x_i(r^*, s^*)}{\partial r_{ik}} = -p_i(\phi(r^*, s^*)) \frac{\partial y_i(r^*)}{\partial r_{ik}} - \frac{dp_i(\phi(r^*, s^*))}{dr_{ik}} y_i(r^*).$$

Comparing equations (22) and (23), it must be that either $y_i(r^*) = 0$ or $dp_i(\phi(r^*, s^*))/dr_{ik} = 0$ for all k .

If $y_i(r^*) \neq 0$ but $dp_i(\phi(r^*, s^*))/dr_{ik} = 0$ for all k then, by (20),

$$g_i(r^*, s^*) = [p_i(\phi(r^*, s^*)) - q_i(r^*, s^*)] y_i(r^*)$$

and so $g_i(r^*, s^*)$ can be expressed as $h_i(r^*, s^*) y_i(r^*)$ for some function h_i such that $dh_i/dr_{ik} = 0$ for all k . But then $x_i(r^*, s^*)$ can be re-written as:

$$x_i(r^*, s^*) = -[q_i(r^*, s^*) + h_i(r^*, s^*)] y_i(r^*).$$

Label the bracketed term as $\tilde{q}_i(r^*, s^*)$ and we have that

$$x_i(r^*, s^*) = -\tilde{q}_i(r^*, s^*) y_i(r^*)$$

with $d\tilde{q}_i/dr_{ik} = 0$, giving the result.

If $y_i(r^*) = 0$ then by equation (20) we have $g_i(r^*, s^*) = 0$. It remains to show that $dq_i(r^*, s^*)/dr_{ik} = 0$. Since dy_i/dr_{ik} is bounded away from zero any perturbation of r_{ik} leads to $y_i \neq 0$; by regularity of (r^*, s^*) , small perturbations lead to other regular equilibria with $y_i \neq 0$ at which $dq_i/dr_{ik} = 0$. Since q_i is continuously differentiable it must be that $dq_i(r^*, s^*)/dr_{ik} = 0$ as well.

Q.E.D.

Proof of Theorems 3 and 5

Consider the case of Lindahl equilibrium. Under the maintained assumptions an allocation $(x_1^*, \dots, x_n^*, y^*)$ is Lindahl equilibrium allocation at θ if there exists some $(p_i^*)_i$ such that

- (1) for each i , $(x_i^*, y^*) \in \arg \max_{x_i, y} u_i(x_i, y | \theta_i)$ subject to $x_i = -p_i^* y$, and
(2) $\sum_i p_i^* = \kappa$.

For the first part of the theorem, fix a Nash equilibrium $m^* = (r^*, s^*)$ of Γ at θ and let $p_i^* = q_i(m_{-i}^*)$ for each i . Then Condition 2 is satisfied by hypothesis. Condition 1 can be rewritten as

$$(24) \quad y^* \in \arg \max_y u_i(-q_i(m_{-i}^*)y, y | \theta_i).$$

Since y is bijective in r_i for each m_{-i} , this is equivalent to

$$(r_i^*, s_i^*) \in \arg \max_{(r_i, s_i)} u_i(-q_i(r_{-i}^*, s_{-i}^*)y(r_i, r_{-i}^*, s_{-i}^*), y(r_i, r_{-i}^*, s_{-i}^*) | \theta_i).$$

Because $g_i \geq 0$ and $g_i = 0$ at any equilibrium point and u_i is increasing in the first argument, Condition 1 is also equivalent to

$$(r_i^*, s_i^*) \in \arg \max_{(r_i, s_i)} u_i(-q_i(r_{-i}^*, s_{-i}^*)y(r_i, r_{-i}^*, s_{-i}^*) - g_i(r_i, r_{-i}^*, s_i, s_{-i}^*), y(r_i, r_{-i}^*, s_{-i}^*) | \theta_i).$$

But this is clearly satisfied since (r_i^*, s_i^*) is a best response to (r_{-i}^*, s_{-i}^*) . Thus,

$$(x_1(m^*), \dots, x_n(m^*), y(m^*))$$

is a Lindahl equilibrium allocation at θ with prices $(p_i(m_{-i}^*))_i$.

For the second part of the theorem, fix a Lindahl allocation (x^*, y^*) with prices $(p_i^*)_i$ such that message $m' = (r', s')$ maps to (x^*, y^*) and $q_i(m'_{-i}) = p_i^*$ for each i ; Assumption 4 guarantees that at least one such m' exists.

Condition 1 for Lindahl equilibria is equivalent to equation (24); since $s' \in \sigma(r')$ this is equivalent to

$$(r'_i, s'_i) \in \arg \max_{(r_i, s_i)} u_i(-q_i(r'_{-i}, s'_{-i})y(r_i, r'_{-i}, s'_{-i}) - g_i(r_i, r'_{-i}, s_i, s'_{-i}), y(r_i, r'_{-i}, s'_{-i}) | \theta_i).$$

But this implies that (r'_i, s'_i) is a best response for each i , so (r', s') is a Nash equilibrium of Γ at θ .

The proof for Walrasian equilibria is identical, setting $\kappa = 0$.

Q.E.D.

Proof of Theorem 6

We know that there cannot exist *any* one-dimensional mechanism that Nash implements the Walrasian correspondence, contractive or not. For the public goods setting, suppose by way of contradiction that the mechanism $(y, (q_i, g_i)_{i=1}^n)$ Nash implements the Lindahl correspondence and is contractive. By Theorem 4 we know that $g_i \equiv 0$ and so for any

quasilinear environment

$$u_i(x_i, y|\theta_i) = v_i(y|\theta_i) + x_i$$

with $v_i'' < 0$ we have

$$U_i(r_i, r_{-i}) = v_i(y(r)|\theta_i) - q_i(r_{-i})y(r).$$

Agent i 's best-response is given by $\rho_i(r_{-i})$ and satisfies the first-order condition

$$v_i'(y(\rho_i, r_{-i})|\theta_i) = q_i(r_{-i})$$

for all r_{-i} . Take any r^* and θ for which r^* is a Nash equilibrium at θ . By the Implicit Function Theorem the slope of ρ_i at r^* with respect to each r_j is

$$\frac{\partial \rho_i}{\partial r_j} = \frac{\partial q_i / \partial r_j - v_i''(y|\theta_i) \partial y / \partial r_j}{v_i''(y|\theta_i) \partial y / \partial r_i}.$$

For the mechanism to be contractive it is necessary (though not sufficient) that for all i and $j \neq i$

$$(25) \quad \left| \frac{\partial y / \partial r_j}{\partial y / \partial r_i} - \frac{\partial q_i / \partial r_j}{v_i''(y|\theta_i) \partial y / \partial r_i} \right| < 1.$$

Now select the agent j^* such that $|\partial y(r^*) / \partial r_{j^*}| \geq |\partial y(r^*) / \partial r_i|$ for all i . In order to satisfy equation 25 it must be that $\partial y / \partial r_{j^*}$ and $\partial q_i / \partial r_{j^*}$ have the opposite sign for all i and that $\partial q_i / \partial r_{j^*} \neq 0$ for all i . Therefore, each $\partial q_i / \partial r_{j^*}$ has the same sign for all $i \neq j^*$ (and $\partial q_{j^*} / \partial r_{j^*} = 0$) so that $\sum_i \partial q_i / \partial r_{j^*} \neq 0$. But since all r are Nash equilibria for some θ and the mechanism implements Lindahl allocations it must be that $\sum_i q_i(r_{-i}) = \kappa$ for all r and, therefore, that $\sum_i \partial q_i / \partial r_{j^*} = 0$; this is a contradiction.

Q.E.D.

Proof of Theorem 7

Step 1: We first show that the mechanism is contractive.

To show how to construct contractive mechanisms more generally, we begin the proof using arbitrary y , q_i , and g_i functions; we insert the specific functions given in equations when necessary.

In a two-dimensional mechanism agent i 's utility function at θ defined over strategy profiles (r, s) is given by

$$U_i(r, s|\theta_i) = v_i(y(r)|\theta_i) - q_i(r_{-i}, s_{-i})y(r) - g_i(r, s).$$

Henceforth we drop the dependence on θ for notational brevity. For each i and (r_{-i}, s_{-i}) define $\rho_i(r_{-i}, s_i) \in \mathcal{R}_i$ and $\sigma_i(r_{-i}, s_{-i}) \in \mathcal{S}_i$ to be i 's best-response messages. Using the Implicit Function Theorem, we solve for the slope of each ρ_i and σ_i by differentiating the identities

$$\frac{\partial U_i(\rho_i, \sigma_i, r_{-i}, s_{-i})}{\partial r_i} \equiv \frac{\partial U_i(\rho_i, \sigma_i, r_{-i}, s_{-i})}{\partial s_i} \equiv 0$$

with respect to each r_j and s_j . The resulting system of equations (for each i and j) is given by

$$\begin{bmatrix} \frac{\partial^2 U_i}{\partial r_i^2} & \frac{\partial^2 U_i}{\partial r_i \partial s_i} \\ \frac{\partial^2 U_i}{\partial s_i \partial r_i} & \frac{\partial^2 U_i}{\partial s_i^2} \end{bmatrix} \begin{bmatrix} \frac{\partial \rho_i}{\partial r_j} & \frac{\partial \rho_i}{\partial s_j} \\ \frac{\partial \sigma_i}{\partial r_j} & \frac{\partial \sigma_i}{\partial s_j} \end{bmatrix} = - \begin{bmatrix} \frac{\partial^2 U_i}{\partial r_i r_j} & \frac{\partial^2 U_i}{\partial r_i s_j} \\ \frac{\partial^2 U_i}{\partial s_i r_j} & \frac{\partial^2 U_i}{\partial s_i s_j} \end{bmatrix},$$

which has a unique solution if the left-most matrix is invertible. In that case the inverse is given by

$$\left(\frac{\partial^2 U_i}{\partial r_i^2} \frac{\partial^2 U_i}{\partial s_i^2} - \left(\frac{\partial^2 U_i}{\partial r_i \partial s_i} \right)^2 \right)^{-1} \begin{bmatrix} \frac{\partial^2 U_i}{\partial s_i^2} & -\frac{\partial^2 U_i}{\partial r_i \partial s_i} \\ -\frac{\partial^2 U_i}{\partial r_i \partial s_i} & \frac{\partial^2 U_i}{\partial r_i^2} \end{bmatrix},$$

and so invertibility requires that for each i and j ,

$$\left(\frac{\partial^2 U_i}{\partial r_i \partial s_i} \right)^2 \neq \frac{\partial^2 U_i}{\partial r_i^2} \frac{\partial^2 U_i}{\partial s_i^2}.$$

Letting $y(r)$ be linear (so that $y(r) = \sum_i \alpha_i r_i$ for some $(\alpha_1, \dots, \alpha_n) > 0$), the relevant second derivatives for solving this system are

$$\begin{aligned} \frac{\partial^2 U_i}{\partial r_i \partial r_j} &= v_i''(y(r)) \alpha_i \alpha_j - \alpha_i \frac{\partial q_i}{\partial r_j} - \frac{\partial^2 g_i}{\partial r_i \partial r_j}, \\ \frac{\partial^2 U_i}{\partial r_i \partial s_i} &= -\frac{\partial^2 g_i}{\partial r_i \partial s_i}, \\ \frac{\partial^2 U_i}{\partial r_i \partial s_j} &= -\alpha_i \frac{\partial q_i}{\partial s_j} - \frac{\partial^2 g_i}{\partial r_i \partial s_j}, \\ \frac{\partial^2 U_i}{\partial r_i^2} &= v_i''(y(r)) \alpha_i^2 - \frac{\partial^2 g_i}{\partial r_i^2}, \\ \frac{\partial^2 U_i}{\partial s_i \partial r_j} &= -\frac{\partial^2 g_i}{\partial s_i \partial r_j}, \\ \frac{\partial^2 U_i}{\partial s_i^2} &= -\frac{\partial^2 g_i}{\partial s_i^2}, \end{aligned}$$

and

$$\frac{\partial^2 U_i}{\partial s_i \partial s_j} = -\frac{\partial^2 g_i}{\partial s_i \partial s_j}.$$

Note that if $\partial^2 g_i / \partial r_i^2 \geq 0$ and $\partial^2 g_i / \partial s_i^2 > 0$ then the utility function is strictly concave, so the best-response is always unique.

With these partial derivatives, the invertibility conditions become

$$-v_i''(y(r)) \neq \frac{\left(\frac{\partial^2 g_i}{\partial r_i \partial s_i}\right)^2 - \frac{\partial^2 g_i}{\partial r_i^2} \frac{\partial^2 g_i}{\partial s_i^2}}{\alpha_i^2 \frac{\partial^2 g_i}{\partial s_i^2}}$$

and the slopes of the best-response functions are given by

$$\begin{bmatrix} \frac{\partial \rho_i}{\partial r_j} & \frac{\partial \rho_i}{\partial s_j} \\ \frac{\partial \sigma_i}{\partial r_j} & \frac{\partial \sigma_i}{\partial s_j} \end{bmatrix} = \left(\frac{\partial^2 U_i}{\partial r_i^2} \frac{\partial^2 U_i}{\partial s_i^2} - \left(\frac{\partial^2 U_i}{\partial r_i \partial s_i} \right)^2 \right)^{-1} \begin{bmatrix} \frac{\partial^2 U_i}{\partial r_i \partial s_i} \frac{\partial^2 U_i}{\partial s_i \partial r_j} - \frac{\partial^2 U_i}{\partial s_i^2} \frac{\partial^2 U_i}{\partial r_i \partial r_j} & \frac{\partial^2 U_i}{\partial r_i \partial s_i} \frac{\partial^2 U_i}{\partial s_i \partial s_j} - \frac{\partial^2 U_i}{\partial s_i^2} \frac{\partial^2 U_i}{\partial r_i \partial s_j} \\ \frac{\partial^2 U_i}{\partial r_i \partial s_i} \frac{\partial^2 U_i}{\partial r_i \partial r_j} - \frac{\partial^2 U_i}{\partial r_i^2} \frac{\partial^2 U_i}{\partial s_i \partial r_j} & \frac{\partial^2 U_i}{\partial r_i \partial s_i} \frac{\partial^2 U_i}{\partial r_i \partial s_j} - \frac{\partial^2 U_i}{\partial r_i^2} \frac{\partial^2 U_i}{\partial s_i \partial s_j} \end{bmatrix}.$$

We now insert our particular y , q_i , and g_i functions. Specifically, let

$$y(r) = \sum_i r_i,$$

$$q_i(r_{-i}, s_{-i}) = \frac{\kappa}{n} + \frac{1}{\delta} (r_{i-1} - r_{i+1}) + \delta \frac{n-1}{n^2} \left(s_{i-1} - \frac{1}{n} r_{i+1} \right),$$

and

$$g_i(r, s) = \frac{1}{2} \left(s_i - \frac{1}{n} r_{i+1} \right)^2 + \frac{\delta}{2} \left(s_{i-1} - \frac{1}{n} r_i \right)^2,$$

which gives a best response function of $\sigma_i(r_{-i}, s_{-i}) = r_{i+1}/n$. Clearly, at any best response point (and, hence, any equilibrium point), $g_i(r, s) = 0$. Also, we can calculate the slopes of σ_i directly: $\partial \sigma_i / \partial r_{i+1} \equiv 1/n$, $\partial \sigma_i / \partial r_j \equiv 0$ for $j \neq i+1$, and $\partial \sigma_i / \partial s_j \equiv 0$ for all j .

The non-zero second partial derivatives of g_i are

$$\begin{aligned} \frac{\partial^2 g_i}{\partial r_i^2} &= \delta \frac{1}{n^2}, \\ \frac{\partial^2 g_i}{\partial r_i \partial s_{i-1}} &= -\delta \frac{1}{n}, \\ \frac{\partial^2 g_i}{\partial s_i^2} &= 1, \end{aligned}$$

and

$$\frac{\partial^2 g_i}{\partial s_i \partial r_{i+1}} = -\frac{1}{n}$$

(all other second-partial derivatives of g_i are identically zero). The non-zero derivatives of q_i are

$$\begin{aligned}\frac{\partial q_i}{\partial r_{i-1}} &= \frac{1}{\delta}, \\ \frac{\partial q_i}{\partial r_{i+1}} &= -\frac{1}{\delta} - \delta \frac{n-1}{n^3},\end{aligned}$$

and

$$\frac{\partial q_i}{\partial s_{i-1}} = \delta \frac{n-1}{n^2}.$$

To calculate the slopes of ρ_i we need to appeal to the general Implicit Function Theorem argument above. The invertibility condition reduces to

$$v_i''(y(r)) \neq \frac{\delta}{n^2},$$

which holds for positive δ since $v_i'' < 0$.

Plugging the derivatives of q_i and g_i into the slope formulas above gives

$$\begin{aligned}\frac{\partial \rho_i}{\partial r_{i-1}} &= \frac{v_i''(y(r)) - \frac{1}{\delta}}{-v_i''(y(r)) + \delta \frac{1}{n^2}}, \\ \frac{\partial \rho_i}{\partial r_{i+1}} &= \frac{v_i''(y(r)) + \frac{1}{\delta} + \delta \frac{n-1}{n^3}}{-v_i''(y(r)) + \delta \frac{1}{n^2}},\end{aligned}$$

and

$$\frac{\partial \rho_i}{\partial r_j} = \frac{v_i''(y(r))}{-v_i''(y(r)) + \delta \frac{1}{n^2}}$$

for all $j \notin \{i-1, i+1\}$; and

$$\frac{\partial \rho_i}{\partial s_{i-1}} = \frac{\delta \frac{1}{n^2}}{-v_i''(y(r)) + \delta \frac{1}{n^2}}$$

and $\partial \rho_i / \partial s_j = 0$ for all $j \neq i-1$. As a check, one can use the Implicit Function Theorem method to verify that $\partial \sigma_i / \partial r_{i+1} = 1/n$, $\partial \sigma_i / \partial s_{i+1} = 0$, and $\partial \sigma_i / \partial r_j = \partial \sigma_i / \partial s_j = 0$ for all $j \neq i+1$, as was derived directly above.

Using Lemma 1 and the *column-sum* matrix norm, a sufficient condition for this mechanism to be contractive is that

$$\sum_{j \neq i} \left(\left| \frac{\partial \rho_j}{\partial r_i} \right| + \left| \frac{\partial \sigma_j}{\partial r_i} \right| \right) < 1$$

and

$$\sum_{j \neq i} \left(\left| \frac{\partial \rho_j}{\partial s_i} \right| + \left| \frac{\partial \sigma_j}{\partial s_i} \right| \right) < 1.$$

As δ grows large we have that $\partial\rho_i/\partial r_{i+1}$ converges to $(n-1)/n$, $\partial\rho_i/\partial r_j$ converges to zero for all $j \neq i+1$, $\partial\rho_i/\partial s_{i-1}$ converges to one, and $\partial\rho_i/\partial s_j$ converges to zero for all $j \neq i-1$. Therefore,

$$\begin{aligned} \lim_{\delta \rightarrow \infty} \sum_{j \neq i} \left(\left| \frac{\partial \rho_j}{\partial r_i} \right| + \left| \frac{\partial \sigma_j}{\partial r_i} \right| \right) &= \lim_{\delta \rightarrow \infty} \left(\left| \frac{\partial \rho_{i-1}}{\partial r_i} \right| + \left| \frac{\partial \sigma_{i-1}}{\partial r_i} \right| \right) \\ &= \left| \frac{n-1}{n} \right| + \left| \frac{1}{n} \right| \\ &= 1 \end{aligned}$$

and

$$\begin{aligned} \lim_{\delta \rightarrow \infty} \sum_{j \neq i} \left(\left| \frac{\partial \rho_j}{\partial s_i} \right| + \left| \frac{\partial \sigma_j}{\partial s_i} \right| \right) &= \lim_{\delta \rightarrow \infty} \left| \frac{\partial \rho_{i+1}}{\partial s_i} \right| \\ &= 1. \end{aligned}$$

If we can verify that both of these sums approach one from below then for large but finite values of δ the mechanism will be contractive.

For the first condition we know that $|\partial\sigma_{i-1}/\partial r_i| = 1/n$ for all δ , so it suffices to check that $|\partial\rho_{i-1}/\partial r_i|$ converges to $|(n-1)/n|$ from below. This is true if

$$\left| \frac{v''_{i-1}(y(r)) + \frac{1}{\delta} + \delta \frac{n-1}{n^3}}{-v''_{i-1}(y(r)) + \delta \frac{1}{n^2}} \right| < \left| \frac{n-1}{n} \right|$$

for large but finite δ . By rearranging this reduces to

$$\left| v''_{i-1}(y(r)) + \frac{1}{\delta} + \delta \frac{n-1}{n^3} \right| < -\frac{n-1}{n} v''_{i-1}(y(r)) + \delta \frac{n-1}{n^3}.$$

If δ is large (specifically, if either

$$\delta > \frac{\left(B + \sqrt{B^2 - 4 \frac{n-1}{n^3}} \right) n^3}{2(n-1)}$$

or $B^2 < 4(n-1)/n^3$) then the term inside the absolute value becomes positive and the expression reduces to

$$\frac{1}{\delta} < -\frac{2n-1}{n} v''_{i-1}(y(r)).$$

Since v''_{i-1} is negative and bounded away from zero this inequality is satisfied for large enough δ (specifically, for $\delta > n/(2B(n-1))$), but this is implied by the previous lower bound on δ whenever $B \geq 4(n-1)/n^3$.

For the second condition we check that $\partial \rho_{i+1} / \partial s_i$ converges to one from below but does not equal one for finite δ ; but this is true for all $\delta > 0$ since

$$\frac{\partial \rho_{i+1}}{\partial s_i} = \frac{\delta \frac{1}{n^2}}{-v''_{i+1}(y(r)) + \delta \frac{1}{n^2}}$$

and v''_{i+1} is negative and bounded away from zero.

Thus, for large δ the best response mapping is a contraction mapping. A bit of algebra confirms that $\delta \geq Bn^2$ is sufficient to satisfy the contraction requirements above.

Step 2: We now prove that the mechanism Nash implements the Lindahl correspondence.

To see that a unique Lindahl equilibrium exists for all θ , note that the three necessary and sufficient conditions for any Lindahl equilibrium are:

- (1) Given p_i^* and y^* , it must be that $x_i^* = -p_i^* y^*$ for all i ;
- (2) This implies that that $\partial v_i(y^*) / \partial y = p_i^*$ for each i ; and
- (3) Linearity of the firm's profit function then implies that $\sum_i \partial v_i(y^*) / \partial y = \sum_i p_i^* = \kappa$.

Using these conditions we can derive the unique equilibrium in three steps:

- (1) Since $v''_i \in (-B, -1/B)$ for each i there is one unique y^* satisfying the third necessary condition;
- (2) Given the unique y^* , there is one unique p_i^* for each i satisfying the second condition; and
- (3) Given y^* and p_i^* there is one unique x_i^* for each i satisfying the first condition.

Since the mechanism is contractive it also has a unique Nash equilibrium (r^*, s^*) for every θ . Now take the equilibrium message (r^*, s^*) and let $p_i^* = q_i(r_{-i}^*, s_{-i}^*)$. Then $x_i(r^*, s^*) = -p_i^* y(r^*)$ for each i , satisfying the first condition. Furthermore, the first-order condition for maximization in r_i at an equilibrium point imply that

$$v'_i(y(r^*)) \frac{\partial y_i(r^*)}{\partial r_i} = p_i^* \frac{\partial y_i(r^*)}{\partial r_i} + \frac{\partial g_i(r^*, \sigma(r^*))}{\partial r_i},$$

but since $\partial y_i(r) / \partial r_i \neq 0$ and $\partial g_i / \partial r_i = 0$ at the equilibrium point implies that $v'_i(y(r^*)) = p_i^*$, satisfying the second condition. Finally, it is easy to check that $\sum_i q_i(r_{-i}^*, s_{-i}^*) = \kappa$ at the equilibrium point since $s_i^* = r_{i+1}^* / n$ and so the third condition is satisfied. Thus, the unique equilibrium point is equal to the unique Lindahl allocation.

Q.E.D.

Proof of Theorem 8

Step 1: We first show that the mechanism is contractive.

Agent i 's utility over strategies in this mechanism is given by:

$$U_i(r, s|\theta) = v_i(y_i(r, s_{-i})|\theta) - \frac{1}{\sqrt{\delta}} \left(s_{i+1} + \frac{\delta-1}{\delta} \frac{1}{n-1} r_{i+1} \right) y_i(r, s_{-i}) - \left(s_i - \frac{\delta-1}{\delta} \frac{1}{n-1} \sum_{j \neq i} r_j \right)^2.$$

To find i 's best-response function, we compute the first-order conditions:

$$\frac{\partial U_i(r, s|\theta)}{\partial r_i} \equiv v'_i(y_i(\cdot)) \frac{\partial y_i}{\partial r_i} - \frac{1}{\sqrt{\delta}} \left(s_{i+1} + \frac{\delta-1}{\delta} \frac{1}{n-1} r_{i+1} \right) \frac{\partial y_i}{\partial r_i} = 0$$

$$\frac{\partial U_i(r, s|\theta)}{\partial s_i} \equiv s_i - \frac{\delta-1}{\delta} \frac{1}{n-1} \sum_{j \neq i} r_j = 0$$

So,

$$\rho_i(r_{-i}, s_{-i}) = \frac{\delta}{2\delta-1} \left(\frac{1}{n-1} \sum_{j \neq i} r_j + s_{i+1} + v_i'^{-1} \left(\frac{1}{\sqrt{\delta}} \left(s_{i+1} + \frac{\delta-1}{\delta} \frac{1}{n-1} r_{i+1} \right) \middle| \theta \right) \right)$$

$$\sigma_i(r_{-i}, s_{-i}) = \frac{\delta-1}{\delta} \frac{1}{n-1} \sum_{j \neq i} r_j$$

The sufficient conditions for the mechanism to be contractive are then

$$(26) \quad \sum_{j \neq i} \left| \frac{\partial \rho_i(r_{-i}, s_{-i})}{\partial r_j} \right| + \sum_{j \neq i} \left| \frac{\partial \rho_i(r_{-i}, s_{-i})}{\partial s_j} \right| = \frac{\delta}{2\delta-1} \left(\sum_{j \neq i, i+1} \left| \frac{1}{n-1} \right| + \left| \frac{1}{n-1} + \frac{\frac{1}{\sqrt{\delta}} \frac{\delta-1}{\delta} \frac{1}{n-1}}{v_i''(\cdot|\theta)} \right| + \left| 1 + \frac{\frac{1}{\sqrt{\delta}}}{v_i''(\cdot|\theta)} \right| \right) < 1$$

and

$$(27) \quad \sum_{j \neq i} \left| \frac{\partial \sigma_i(r_{-i}, s_{-i})}{\partial r_j} \right| + \sum_{j \neq i} \left| \frac{\partial \sigma_i(r_{-i}, s_{-i})}{\partial s_j} \right| = (n-1) \left| \frac{\delta-1}{\delta} \frac{1}{n-1} \right| < 1.$$

The second condition is satisfied since $\delta > 0$.

For the first condition, recall that $v_i''(\cdot|\theta) \in (-B, -1/B)$, so if $\delta > B^2$ then

$$\frac{1}{n-1} + \frac{\frac{1}{\sqrt{\delta}} \frac{\delta-1}{\delta} \frac{1}{n-1}}{v_i''(\cdot|\theta)} > 0, \text{ and } 1 + \frac{\frac{1}{\sqrt{\delta}}}{v_i''(\cdot|\theta)} > 0.$$

Therefore, the left-hand side of (26) is equal to

$$\frac{\delta}{2\delta-1} \left(\frac{n-2}{n-1} + \frac{1}{n-1} + \frac{\frac{1}{\sqrt{\delta}} \frac{\delta-1}{\delta} \frac{1}{n-1}}{v_i''(\cdot|\theta)} + 1 + \frac{\frac{1}{\sqrt{\delta}}}{v_i''(\cdot|\theta)} \right),$$

or

$$\frac{\delta}{2\delta-1} \left(2 + \frac{\frac{1}{\sqrt{\delta}} \frac{\delta-1}{\delta} \frac{1}{n-1}}{v_i''(\cdot|\theta)} + \frac{\frac{1}{\sqrt{\delta}}}{v_i''(\cdot|\theta)} \right).$$

This quantity is less than

$$\frac{\delta}{2\delta-1} \left(2 - \frac{1}{\sqrt{\delta}} \left(1 + \frac{\delta-1}{\delta} \frac{1}{n-1} \right) \right),$$

which is less than one if $\delta > (B(n-1)/n)^2$.

A sufficient requirement for both conditions is therefore that $\delta > B^2$.

Step 2: We now prove that this mechanism Nash implements the Walrasian correspondence.

Take a Nash equilibrium (r^*, s^*) for $\theta \in \Theta$. Then, from the previous step, we know $s_i^* = \frac{\delta-1}{\delta} \frac{1}{n-1} \sum_{j \neq i} r_j^*$. As a result,

$$q_i(r_{-i}^*, s_{-i}^*) = \frac{1}{\sqrt{\delta}} \left(\frac{\delta-1}{\delta} \frac{1}{n-1} \sum_{j \neq i+1} r_j^* + \frac{\delta-1}{\delta} \frac{1}{n-1} r_{i+1}^* \right) = \frac{1}{\sqrt{\delta}} \frac{\delta-1}{\delta} \frac{1}{n-1} \sum_{i=1}^n r_i^* = Q^*$$

for all $i \in \mathcal{I}$, and

$$(28) \quad x_i(r^*, s^*) = -Q^* y_i(r^*, s^*).$$

By definition of Nash equilibrium, $u_i(r^*, s^*) \geq u_i(r_i, r_{-i}^*, s^*)$ for all r_i . That is,

$$(29) \quad v_i(y_i(r^*, s^*)|\theta) - Q^* y_i(r^*, s^*) \geq v_i(y_i(r_i, r_{-i}^*, s^*)|\theta) - Q^* y_i(r_i, r_{-i}^*, s^*)$$

for all r_i . Since y_i is a surjection from \mathbb{R} onto \mathbb{R} (in r_i), (29) implies

$$(30) \quad v_i(y_i(r^*, s^*)|\theta) - Q^* y_i(r^*, s^*) \geq v_i(y_i|\theta) - Q^* y_i$$

for all y_i . Finally, we verify that allocation $[y_i(r^*, s^*), x_i(r^*, s^*)]_i$ is balanced. By definition,

$$\begin{aligned} \sum_i y_i(r^*, s^*) &= \sum_i r_i^* - \frac{1}{n-1} \sum_i \sum_{j \neq i} r_j^* - \sum_i s_{i+1}^* + \frac{\delta-1}{\delta} \frac{1}{n-1} (n-1) \sum_i r_i^* \\ &= -\sum_i s_{i+1}^* + \frac{\delta-1}{\delta} \frac{1}{n-1} (n-1) \sum_i r_i^* \\ &= -\sum_i \frac{\delta-1}{\delta} \frac{1}{n-1} \sum_{j \neq i+1} r_j^* + \frac{\delta-1}{\delta} \frac{1}{n-1} (n-1) \sum_i r_i^* \\ &= 0. \end{aligned}$$

Since $x_i^*(r^*, s^*) = -Q^* y_i(r^*, s^*)$, $\sum_i x_i^*(r^*, s^*) = 0$. It follows from balancedness, (28), and (30) that allocation $[y_i(r^*, s^*), x_i(r^*, s^*)]_i$ is a Walrasian allocation. To complete the proof,

start with the Walrasian allocation $[Y_i^*, X_i^*]_i$ of some environment θ . From step 1, the mechanism is contractive, hence a Nash equilibrium exists. From the previous argument, this Nash equilibrium must correspond to a Walrasian allocation, which is necessarily $[Y_i^*, X_i^*]_i$.

Q.E.D.

REFERENCES

- Abreu, D., Matsushima, H., 1992. Virtual implementation in iteratively undominated strategies I: Complete information. *Econometrica* 60, 993–1008.
- Anderson, C. M., Plott, C. R., Shimomura, K.-I., Granat, S., 2004. Global instability in experimental general equilibrium: The scarf example. *Journal of Economic Theory* 115, 209–249.
- Bergemann, D., Morris, S., 2009. Rationalizable implementation, princeton working paper.
- Bergstrom, T., Simon, C. P., Titus, C. J., 1983. Counting groves-ledyard equilibria via degree theory. *Journal of Mathematical Economics* 12, 167–184.
- Brock, W. A., 1980. The design of mechanisms for efficient allocation of public goods. In: Lawrence R. Klein, M. N. a. S. C. T. (Ed.), *Quantitative Economics and Development: Essays in Honor of Ta-Chung Liu*. Economic Theory, Econometrics, and Mathematical Economics. Academic Press, New York, pp. 45–80.
- Cabrales, A., 1999. Adaptive dynamics and the implementation problem with complete information. *Journal of Economic Theory* 86, 159–184.
- Cabrales, A., Charness, G., Corchón, L. C., 2003. An experiment on nash implementation. *Journal of Economic Behavior and Organization* 51, 161–193.
- Cabrales, A., Ponti, G., 2000. Implementation, elimination of weakly dominated strategies and evolutionary dynamics. *Journal of Economic Dynamics and Control* 3, 247–282.
- Cabrales, A., Serrano, R., 2009. Implementation in adaptive better-response dynamics, iMDEA working paper.
- Chen, Y., 2002. A family of supermodular Nash mechanisms implementing Lindahl allocations. *Economic Theory* 19, 773–790.
- Chen, Y., 2005. Dynamic stability of Nash-efficient public goods mechanisms: Reconciling theory and experiments. In: Zwick, R., Rapoport, A. (Eds.), *Experimental Business Research*. Vol. II. Kluwer Academic Publishers, Norwell, MA.
- Chen, Y., 2008. Incentive compatible mechanisms for pure public goods: A survey of experimental research. In: Plott, C. R., Smith, V. (Eds.), *The Handbook of Experimental Economics Results*. Vol. 1. North Holland, Amsterdam.
- Chen, Y., Gazzale, R., 2004. When does learning in games generate convergence to Nash equilibria? The role of supermodularity in an experimental setting. *American Economic Review* 95 (5), 1505–1535.
- Chen, Y., Plott, C. R., 1996. The Groves-Ledyard mechanism: An experimental study of institutional design. *Journal of Public Economics* 59, 335–364.

- Chen, Y., Tang, F.-F., 1998. Learning and incentive-compatible mechanisms for public goods provision: An experimental study. *Journal of Political Economy* 106, 633–662.
- Conlisk, J., 1973. Quick stability checks and matrix norms. *Economica* 40, 402–409.
- Crockett, S., Oprea, R., Plott, C. R., 2009. Extreme walrasian dynamics: The gale example in the lab, university of Santa Cruz Working Paper.
- de Trenquallye, P., 1989. Stable implementation of Lindahl allocations. *Economics Letters* 29, 291–294.
- Eliasz, K., 2002. Fault-tolerant implementation. *Review of Economic Studies* 69, 589–610.
- Gibbard, A., 1973. Manipulation of voting schemes: A general result. *Econometrica* 41, 587–602.
- Glazer, J., Rosenthal, R. W., 1992. A note on abreu-matsushima mechanisms. *Econometrica* 60 (6), 1435–1438.
- Green, J., Laffont, J.-J., 1977. Characterization of satisfactory mechanisms for the revelation of the preferences for public goods. *Econometrica* 45, 783–809.
- Groves, T., Ledyard, J. O., 1977. Optimal allocation of public goods: A solution to the ‘free-rider’ problem. *Econometrica* 45, 783–809.
- Groves, T., Ledyard, J. O., 1980. The existence of efficient and incentive compatible equilibria with public goods. *Econometrica* 48 (6), 1487–1506.
- Groves, T., Ledyard, J. O., 1987. Incentive compatibility since 1972. In: Groves, T., Radner, R., Reiter, S. (Eds.), *Information, Incentives, and Economic Mechanisms*. University of Minnesota Press, Minneapolis.
- Harstad, R. M., Marrese, M., 1983. Behavioral explanations of efficient public good allocations. *Journal of Public Economics* 19 (3), 367–383.
- Healy, P. J., 2006. Learning dynamics for mechanism design: An experimental comparison of public goods mechanisms. *Journal of Economic Theory* 129 (1), 114–149.
- Hirota, M., 1985. Global stability in a class of markets with three commodities and three consumers. *Journal of Economic Theory* 36, 186–192.
- Hurwicz, L., 1959. Optimality and informational efficiency in resource allocation processes. In: Arrow, K. J., Karlin, S., Suppes, P. (Eds.), *Mathematical Methods in the Social Sciences: Proceedings of the First Stanford Symposium*. Vol. IV of *Stanford Mathematical Studies in the Social Sciences*. Stanford University Press, Stanford, CA.
- Hurwicz, L., 1972. On informationally decentralized systems. In: McGuire, C. B., Radner, R. (Eds.), *Decision and Organization: A Volume in Honor of Jacob Marschak*. North-Holland, Amsterdam.
- Hurwicz, L., 1979a. On allocations attainable through Nash equilibria. *Journal of Economic Theory* 21, 140–165.
- Hurwicz, L., 1979b. Outcome functions yielding Walrasian and Lindahl allocations at Nash equilibrium points. *Review of Economic Studies* 46, 217–225.
- Jackson, M. O., 2001. A crash course in implementation theory. *Social Choice and Welfare* 18, 655–708.
- Jordan, J. S., 1986. Instability in the implementation of walrasian allocations. *Journal of Economic Theory* 39, 301–328.

- Kim, T., 1987. Stability problems in the implementation of Lindahl allocations. Ph.D. thesis, University of Minnesota.
- Kim, T., 1993. A stable Nash mechanism implementing Lindahl allocations for quasi-linear environments. *Journal of Mathematical Economics* 22 (4), 359–371.
- Kim, T., 1996. A stable Nash mechanism for quasi-additive public-good environments. *The Japanese Economic Review* 47, 144–156.
- Ledyard, J. O., 1995. Public goods: A survey of experimntal research. In: Kagel, J., Roth, A. (Eds.), *Handbook of Experimental Economics*. Princeton University Press, Princeton, NJ.
- Maskin, E., 1999. Nash equilibrium and welfare optimality. *Review of Economic Studies* 66, 23–38.
- Mathevet, L. A., 2007. Supermodular bayesian implementation: Learning and incentive design, mimeo.
- McKelvey, R. D., Palfrey, T. R., 1992. An experimental study of the centipede game. *Econometrica* 60 (4), 803–836.
- McKelvey, R. D., Palfrey, T. R., 1995. Quantal response equilibria for normal form games. *Games and Economic Behavior* 10 (1), 6–38.
- Milgrom, P., Roberts, J., 1990. Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica* 58, 1255–1277.
- Monderer, D., Shapley, L. S., 1996. Potential games. *Games and Economic Behavior* 14, 124–143.
- Moore, J., Repullo, R., 1988. Subgame perfect implementation. *Econometrica* 46, 1191–1220.
- Muench, T., Walker, M., 1983. Are Groves-Ledyard equilibria attainable? *Review of Economic Studies* 50, 393–396.
- Nagel, R., 1995. Unraveling in guessing games: An experimental study. *American Economic Review* 85 (5), 1313–1326.
- Page, S. E., Tassier, T., 2004. Equilibrium selection and stability for the groves ledyard mechanism. *Journal of Public Economic Theory* 6 (2), 311–335.
- Reichelstein, S., Reiter, S., 1988. Game forms with minimal message spaces. *Econometrica* 56 (3), 661–692.
- Roberts, K., 1979. The characterization of implementable social choice rules. In: Laffont, J.-J. (Ed.), *Aggregation and Revelation of Preferences*. North-Holland, Amsterdam.
- Samuelson, L., Zhang, J., 1992. Evolutionary stability in asymmetric games. *Journal of Economic Theory* 57, 363–391.
- Sandholm, W., 2002. Evolutionary implementation and congestion pricing. *Review of Economic Studies* 69, 667–689.
- Sandholm, W., 2005. Negative externalities and evolutionary implementation. *Review of Economic Studies* 72, 885–915.
- Sandholm, W., 2007. Pigouvian pricing and stochastic evolutionary stability. *Journal of Economic Theory* 132, 367–382.
- Satterthwaite, M., 1975. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory* 10, 187–217.

- Scarf, H., 1960. Some examples of global instability of the competitive equilibrium. *International Economic Review* 1, 157–172.
- Scherr, B. A., Babb, E. M., 1975. Pricing public goods: An experiment with two proposed pricing systems. *Public Choice* 21, 35–53.
- Sefton, M., Yavas, A., 1996. Abreu-matsushima mechanisms: Experimental evidence. *Games and Economic Behavior* 16, 280–302.
- Sjöström, T., 1994. Implementation in undominated nash equilibria without integer games. *Games and Economic Behavior* 6, 502–511.
- Smith, V., 1979. Incentive compatible experimental processes for the provision of public goods. In: Smith, V. (Ed.), *Research in Experimental Economics*. JAI Press, Greenwich, CT.
- Stahl, D. O., Wilson, P. W., 1995. On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior* 10, 218–254.
- Tian, G., 1990. Completely feasible and continuous implementation of the lindahl correspondence with a message space of minimal dimension. *Journal of Economic Theory* 51, 443–452.
- Tideman, T. N., 1983. An experiment in the demand-revealing process. *Public Choice* 41 (3), 378–401.
- Topkis, D. M., 1979. Equilibrium points in nonzero-sum n -person submodular games. *SIAM Journal on Control and Optimization* 17, 773–787.
- Topkis, D. M., 1998. *Supermodularity and Complementarity*. Princeton University Press, Princeton, NJ.
- Tumennasan, N., 2008. To err is human: Implementation in quantal response equilibria, brown University Working Paper.
- Van Essen, M., 2009a. A note on the stability of chen's lindahl mechanism, university of Arizona Working Paper.
- Van Essen, M., 2009b. A simple supermodular mechanism that implements lindahl allocations, university of Arizona Working Paper.
- Vega-Redondo, F., 1989. Implementation of Lindahl equilibrium: An integration of static and dynamic approaches. *Mathematical Social Science* 18, 211–228.
- Walker, M., 1981. A simple incentive compatible scheme for attaining Lindahl allocations. *Econometrica* 49, 65–71.
- Walker, M., 1984. A simple auctioneerless mechanism with walrasian properties. *Journal of Economic Theory* 32, 111–127.

[The Following Section Would Appear as an Online Appendix]

The Rich Type Space Assumptions

In this subsection, we break Assumption 3' into two separate (but linked) assumptions that together imply Assumption 3'; in the case of one-dimensional mechanisms these assumptions imply Assumption 3.

Assumption 6. There exists some $\rho \in \{2, 4, 6, \dots\}$ such that

(A) all ρ -th order preferences are admissible:

$$\Theta_\rho := \{\theta \in \Theta : (\forall i) (\exists (\alpha_i, \beta_i) \in (\mathbb{R}_{++} \times \mathbb{R})) \text{ s.t. } u_i(x_i, y_i | \theta_i) = (-\alpha_i y_i^\rho + \beta_i y_i) + x_i\} \subseteq \Theta,$$

and

(B) for all $r \in \mathcal{R}$, all $s \in \sigma(r)$, and all $i \in \mathcal{I}$ there exists some finite $\gamma_i(r) > 0$ such that for all $r'_i \in \mathcal{R}_i$ and $s'_i \in \sigma_i(r'_i, r_{-i}, s_{-i})$,

$$|x_i(r'_i, r_{-i}, s'_i, s_{-i}) - x_i(r, s)| \leq \gamma_i(r) \max\left\{|y_i(r'_i, r_{-i}) - y_i(r)|^\rho, |y_i(r'_i, r_{-i}) - y_i(r)|^{1/\rho}\right\},$$

Assumption 6A simply requires that all polynomial (quasilinear) preferences of order ρ be permitted in the type space. To interpret Assumption 6B, let $\rho = 2$ and consider changes in m_i that lead to large changes in y_i . In this case, the squared term in the maximand applies, and so the assumption places quadratic upper and lower bounds on the change in x_i . For changes in m_i , that lead to small changes in y_i the upper and lower bounds are square-root bounds. In either case, the requirement is strictly weaker than requiring that χ_i be Hölder continuous of degree ρ or that χ_i be Lipschitz continuous. The bounds on χ_i imposed by this assumption are demonstrated in figure IV. Note that as ρ increases Assumption 6B becomes strictly weaker though Assumption 6A requires more 'exotic' preferences in the economy.

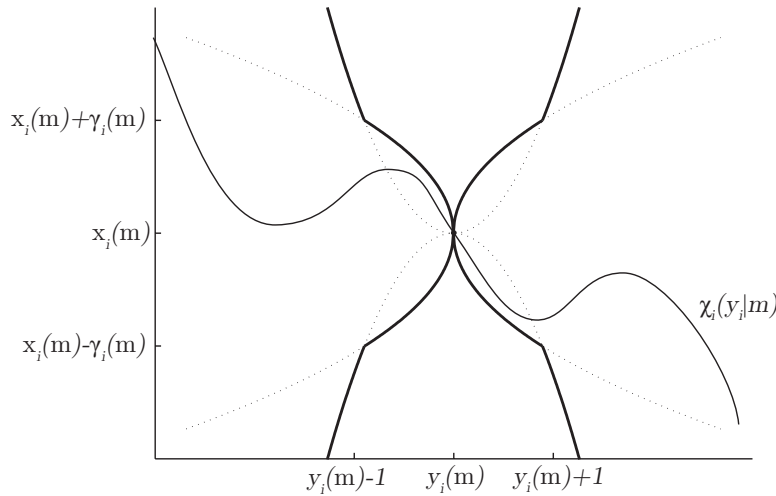


FIGURE IV. The bounds on $\chi_i(y_i | m_{-i})$ imposed by Assumption 6B for $\rho = 2$.

Given these modified assumptions, we can now prove that Assumption 3' (or Assumption 3) holds.

Proposition 2. Take any mechanism $\Gamma = (\mathcal{M}_i, x_i, y_i)_{i \in \mathcal{I}}$ and ρ satisfying Assumptions 1', 2' and 6B and any type space Θ satisfying Assumption 6A. If $\rho \leq 2$ then Assumption 3' is satisfied: $\nu(\Theta) = \mathcal{M}^*$. If $\rho > 2$ then $\{(r, s) \in \mathcal{M}^* : y_i(r) \neq 0 \forall i\} \subseteq \nu(\Theta)$.

Proof of Proposition 2:

Define \mathcal{M}^{**} by

$$\mathcal{M}^{**} = \{(r, s) \in \mathcal{M}^* : y_i(r)^{\rho-2} \neq 0 \forall i\}.$$

Note that if $\rho \in \{1, 2\}$ then $\mathcal{M}^* = \mathcal{M}^{**}$ (using the convention that $0^0 = 1$). We know that $\nu(\Theta) \subseteq \mathcal{M}^*$; Proposition 2 can then be proven by showing that $\mathcal{M}^{**} \subseteq \nu(\Theta)$. This is done by constructing a mapping $\phi : \mathcal{M}^{**} \rightarrow \Theta_\rho$ such that $m \in \nu(\phi(m))$ for all $m \in \mathcal{M}^{**}$. Thus,

$$\mathcal{M}^{**} \subseteq \nu(\phi(\mathcal{M}^{**})) = \nu(\Theta_\rho) \subseteq \nu(\Theta),$$

giving the result.

Specifically, consider the mapping $\phi : \mathcal{M}^{**} \rightarrow \Theta_\rho$ such that $\phi_i(m^*) = (\alpha_i(m^*), \beta_i(m^*)) \in \mathbb{R}_+ \times \mathbb{R}$ for each $m^* \in \mathcal{M}^{**}$ and

$$u_i(x_i, y_i | \phi_i(m^*)) = v_i(y_i | \phi_i(m^*)) + x_i,$$

where

$$v_i(y_i | \phi_i(m^*)) = -\frac{\alpha_i(m^*)}{\rho} y_i^\rho + \beta_i(m^*) y_i$$

and, for a given value of $\alpha_i(m^*)$ (to be determined later in the proof), $\beta_i(m^*)$ is given by

$$(31) \quad \beta_i(r^*, s^*) := \alpha_i(r^*, s^*) y_i^{\rho-1}(r) + P_{ik}(r^*, s^*)$$

(recall that P_{ik} is the effective price function defined in equation (3) and does not depend on k since $m^* \in \mathcal{M}^{**}$).

We now fix an arbitrary $m^* = (r^*, s^*) \in \mathcal{M}^{**}$ and show that m_i^* is a best-response to m_{-i}^* for each i in environment $\phi(m^*) = (\alpha_i(m^*), \beta_i(m^*))_{i \in \mathcal{I}}$. This is done in two steps; first we verify that m_i^* is a local optimum in response to m_{-i}^* for each i and then we show m_i^* can be made a global optimum by increasing $\alpha_i(m^*)$ sufficiently, allowing $\beta_i(m^*)$ to adjust appropriately as $\alpha_i(m^*)$ changes.

Given $\phi_i(m^*)$, i 's objective is to choose (r_i, s_i) to maximize

$$(32) \quad -\frac{\alpha_i(m^*)}{\rho} y_i(r_i, r_{-i}^*)^\rho + \beta_i(m^*) y_i(r_i, r_{-i}^*) + x_i(r_i, s_i, r_{-i}^*, s_{-i}^*).$$

For local optimality, the first-order conditions for each s_{ik} are already satisfied at m^* by the construction of \mathcal{M}^{**} (see equation 8). As for r_{ik} , agent i 's first-order condition for utility maximization at (r^*, s^*) with respect to each r_{ik} is

$$\left[-\alpha_i(r^*, s^*) y_i^{\rho-1}(r) + \beta_i(r^*, s^*) \right] \frac{\partial y_i(r)}{\partial r_{ik}} + \frac{\partial x_i(r, s)}{\partial r_{ik}} = 0.$$

But the construction of β_i (equation 31) guarantees that this is satisfied at $(r, s) = (r^*, s^*)$ for any $\alpha_i(r^*, s^*)$, so the first-order conditions are satisfied for all $m^* \in \mathcal{M}^{**}$.

To describe the second-order conditions for local optimality, we show that the matrix of second-partial derivatives of i 's objective function will be negative definite for sufficiently large $\alpha_i(m^*)$. Shortening notation, let \mathbf{X}_r and \mathbf{X}_s be the column vectors of partial derivatives of x_i with respect to r_i and s_i , respectively, and let \mathbf{X}_{rr} , \mathbf{X}_{rs} , and \mathbf{X}_{ss} represent the matrices of cross-partial derivatives of x_i . Similarly define \mathbf{Y}_r and \mathbf{Y}_{rr} as the partial and cross-partial derivatives of y_i , respectively. Using this notation, the matrix of second partial derivatives of the objective function (32) (after inserting the definition of $\beta_i(m^*)$ from equation 31) is given by the $K_i \times K_i$ matrix

$$\mathbf{H}_i = \left[\begin{array}{c|c} -\alpha_i(m^*)(\rho-1)y_i(r^*)^{\rho-2}(\mathbf{Y}_r \cdot \mathbf{Y}_r^T) + P_{ik}(m^*)\mathbf{Y}_{rr} + \mathbf{X}_{rr} & \mathbf{X}_{rs} \\ \hline \mathbf{X}_{rs}^T & \mathbf{X}_{ss} \end{array} \right],$$

where again $P_{ik}(m^*)$ does not depend on k since $m^* \in \mathcal{M}^{**}$. Now take any direction $(\mathbf{d}_r, \mathbf{d}_s) \neq \mathbf{0}$ of deviation from m_i^* . Since $m^* \in \mathcal{M}^{**}$ implies $s^* \in \sigma(r^*)$, we know that any deviation with $\mathbf{d}_r = \mathbf{0}$ will not yield strictly higher utility, hence $(\mathbf{0}, \mathbf{d}_s)^T \cdot \mathbf{H}_i \cdot (\mathbf{0}, \mathbf{d}_s) \leq 0$. For any direction $(\mathbf{d}_r, \mathbf{d}_s)$ with $\mathbf{d}_r \neq \mathbf{0}$ we have

$$\begin{aligned} (\mathbf{d}_r, \mathbf{d}_s)^T \cdot \mathbf{H}_i \cdot (\mathbf{d}_r, \mathbf{d}_s) &= -\alpha_i(m^*)(\rho-1)y_i(r^*)^{\rho-2} \mathbf{d}_r^T (\mathbf{Y}_r \cdot \mathbf{Y}_r^T) \mathbf{d}_r + K_i(m^*) \\ &= -\alpha_i(m^*)(\rho-1)y_i(r^*)^{\rho-2} (\mathbf{d}_r^T \mathbf{Y}_r)^2 + K_i(m^*) \end{aligned}$$

where

$$K_i(m^*) = \mathbf{d}_r^T [P_{ik}(m^*)\mathbf{Y}_{rr} + \mathbf{X}_{rr}] \mathbf{d}_r + 2\mathbf{d}_r^T \mathbf{X}_{rs} \mathbf{d}_s + \mathbf{d}_s^T \mathbf{X}_{ss} \mathbf{d}_s.$$

Since x_i and y_i are continuously differentiable and $\partial y_i / \partial r_i$ is bounded away from zero, $K_i(m^*)$ is finite for all m^* . Because $y_i(r^*)^{\rho-2} \neq 0$, α_i can be chosen to be any function satisfying

$$\alpha_i(m^*) > K_i(m^*) ((\rho-1)y_i(r^*)^{\rho-2})^{-1} (\mathbf{d}_r^T \mathbf{Y}_r)^{-2}$$

for all $m^* \in \mathcal{M}^{**}$, so that $(\mathbf{d}_r, \mathbf{d}_s)^T \cdot \mathbf{H}_i \cdot (\mathbf{d}_r, \mathbf{d}_s) < 0$. Thus, m_i^* is a local best-response to m_{-i}^* for large enough $\alpha_i(m^*)$.

We now construct $\phi_i(m^*)$ by increasing $\alpha_i(m^*)$ until m_i^* is a *global* best-response to m_{-i}^* . Since m_i^* is a local best-response, there is some neighborhood $\mathcal{N}_i(m^*)$ of m_i^* on which m_i^* maximizes i 's utility given $\alpha_i(m^*)$. Although increasing α_i may change the neighborhood around m^* on which m_i^* is a local best-response, the neighborhood can only increase in size as α_i is increased. Thus, we ignore this dependence of $\mathcal{N}_i(m^*)$ on α_i and show that any $m'_i \notin \mathcal{N}_i(m^*)$ yields a lower payoff than m_i^* when α_i is sufficiently large.

To proceed, pick any m'_i and m''_i such that $m_i^* \in (m'_i, m''_i) \subset \mathcal{N}_i(m^*)$ and, to shorten notation, let $y_i^* = y_i(r^*)$, $x_i^* = x_i(m^*)$, $y'_i = y_i(r'_i, r_{-i}^*)$, $x'_i = x_i(m'_i, m_{-i}^*)$, $y''_i = y_i(r''_i, r_{-i}^*)$, and $x''_i = x_i(m''_i, m_{-i}^*)$.

To show that $u_i(x_i^*, y_i^*) - u_i(x'_i, y'_i) \geq 0$ for some α'_i , we expand this expression to get

$$\alpha'_i \left[\left(\frac{\rho-1}{\rho} y_i^{*\rho} + \frac{1}{\rho} y_i'^{\rho} \right) - (y_i^{*\rho})^{\frac{\rho-1}{\rho}} (y_i'^{\rho})^{\frac{1}{\rho}} \right] + P_{ik}(m^*) (y_i^* - y'_i) \geq (x'_i - x_i^*),$$

which, by Assumption B, is true if

$$(33) \quad \alpha'_i \left[\left(\frac{\rho-1}{\rho} y_i^{*\rho} + \frac{1}{\rho} y_i'^{\rho} \right) - (y_i^{*\rho})^{\frac{\rho-1}{\rho}} (y_i'^{\rho})^{\frac{1}{\rho}} \right] + P_{ik}(m^*) (y_i^* - y_i') \geq \gamma_i(m^*) \rho \max \left\{ |y_i^* - y_i'|^\rho, |y_i^* - y_i'|^{\frac{1}{\rho}} \right\}$$

(the extra ρ before the maximizing operator is needed for a later step). But the term in square brackets is the difference between the weighted arithmetic mean and the weighted geometric mean of the two points $y_i^{*\rho}$ and $y_i'^{\rho}$; by the AM-GM inequality this difference is positive. Thus, there is some finite α'_i at which inequality (33) is true. Similarly, there is some finite α''_i at which the expression $u_i(x_i^*, y_i^*) - u_i(x_i'', y_i'') \geq 0$ is true. Let $\alpha_i(m^*) = \max\{\alpha'_i, \alpha''_i\}$ and now fix $\phi_i(m^*) = (\alpha_i(m^*), \beta_i(m^*))$.

Suppose that $y_i' < y_i''$ (the proof for the case where $y_i'' < y_i'$ is symmetric) and pick any $y_i \geq y_i''$. Suppose that

$$(34) \quad \alpha_i(m^*) \left[\left(\frac{\rho-1}{\rho} y_i^{*\rho} + \frac{1}{\rho} y_i^\rho \right) - (y_i^{*\rho})^{\frac{\rho-1}{\rho}} (y_i^\rho)^{\frac{1}{\rho}} \right] + P_{ik}(m^*) (y_i^* - y_i) - \gamma_i(m^*) \rho \max \left\{ |y_i^* - y_i|^\rho, |y_i^* - y_i|^{\frac{1}{\rho}} \right\} \geq 0,$$

which is true for $y_i = y_i''$ (see inequality (33)). Then the derivative of the left-hand side of this inequality is positive, implying that the inequality is true for all $y_i \geq y_i''$; to see this, take the derivative of the left-hand side and multiply by $(y_i - y_i^*) > 0$ to get either

$$(35) \quad \alpha_i(m^*) \left[y_i^{*\rho} - y_i^{*\rho-1} y_i + y_i^\rho - y_i^* y_i^{\rho-1} \right] + P_{ik}(m^*) (y_i^* - y_i) - \gamma_i(m^*) \rho (y_i - y_i^*)^\rho$$

or

$$(36) \quad \alpha_i(m^*) \left[y_i^{*\rho} - y_i^{*\rho-1} y_i + y_i^\rho - y_i^* y_i^{\rho-1} \right] + P_{ik}(m^*) (y_i^* - y_i) - \gamma_i(m^*) \frac{1}{\rho} (y_i - y_i^*)^{1/\rho}.$$

In either case, the expression is greater than the left-hand side of (34) because

$$\left[y_i^{*\rho} - y_i^{*\rho-1} y_i + y_i^\rho - y_i^* y_i^{\rho-1} \right] \geq \left[\left(\frac{\rho-1}{\rho} y_i^{*\rho} + \frac{1}{\rho} y_i^\rho \right) - (y_i^{*\rho})^{\frac{\rho-1}{\rho}} (y_i^\rho)^{\frac{1}{\rho}} \right]$$

reduces to

$$\left(\frac{\rho-1}{\rho} y_i^{*\rho} + \frac{1}{\rho} y_i^\rho \right) \geq (y_i^{*\rho})^{\frac{\rho-1}{\rho}} (y_i^\rho)^{\frac{1}{\rho}},$$

which is just the AM-GM inequality again. Thus, both (35) and (36) are positive. By continuity, (34) is positive for all $y_i \geq y_i''$ and so deviations resulting in $y_i \geq y_i''$ are not profitable. A symmetric argument shows that deviations to $y_i \leq y_i'$ are also not profitable. Since we already know that deviations resulting in $y_i \in (y_i', y_i'')$ are unprofitable, the proof is complete.

Q.E.D.