Discussion Paper No. 396

Technical Note to "Strategy-proof Allocation Mechanisms"

by

Mark A. Satterthwaite
Graduate School of Management
Northwestern University

and

Hugo Sonnenschein
Department of Economics
Princeton University

August, 1979

Technical Addendum to
"Strategy-Proof Allocation Mechanisms"

by

Mark A. Satterthwaite
Northwestern University

Hugo Sonnenschein
Princeton University

In our paper [1] we presented a proof of Lemma 3 that is only valid for the case where each agent's constraint set $B_j(u)$ is a $\ell-1$ dimensional manifold. Our results, however, are in terms of the general case where $B_j(u)$ may have any dimension m_j, $0 \leq m_j \leq \ell-1$. Therefore this addendum presents a general proof of Lemma 3 that, because of its length, was not included in our main paper.

> Lemma 3. If σ is strategy-proof and satisfies BA, then, for all $i \neq j$, $\tilde{\tilde{S}}^{ij} \supset S^{ij}$ where $\tilde{\tilde{S}}^{ij}$ denotes the closure of $\tilde{S}^{ij}$.

Proof. Suppose the Lemma is not true. A $\bar{u} \in \mathcal{R}$ therefore exists such that:

a. $i\ A(\bar{u})\ j$ and not $i\ \tilde{A}(\bar{u})\ j$ and

b. a neighborhood $N(\bar{u}) = N(\bar{u}_1) \times N(\bar{u}_2) \times \ldots \times N(\bar{u}_n) \subset \mathcal{R}$ exists for which $u \in N(\bar{u})$ implies $i\ A(u)\ j$ and not $i\ \tilde{A}(u)\ j$.

Regularity and BA imply that we may select a neighborhood $\hat{N} = N(\sigma_j(\bar{u})) \subset X$ so that (a) corresponding to each x in $B_j(\bar{u}) \cap \hat{N}$ is an admissible utility function $u_j^x \in N(\bar{u}_j)$ that has its maximum on $B_j(\bar{u})$ at x and (b) for all

$u \in N(\bar{u})$, the manifold $B_j(u) \cap \hat{N}$ is smooth, continuously differentiable in u, and m-dimensional where $0 \leq m \leq \ell-1$. Note that because σ is strategy-proof $\sigma_j(\bar{u}\backslash u_j^x) = x$.

First we consider the harder case where $1 \leq m \leq \ell-1$. The second case where m=0 is deferred until the end. Pick an arbitrary $v_i \in C^2(X)$. For all $u \in N(\bar{u})$, not $i\ \tilde{A}(u)\ j$ by hypothesis. This means that $D_{(v_1)}u_j\sigma_j(u) = 0$ for all $u \in N(\bar{u})$. Define $\hat{B}_j(\lambda) = \{x \in \hat{N} | \exists u_j' \in N(\bar{u}_j) \text{ s.t. } x = \sigma_j(\bar{u}\backslash u_j', \bar{u}_i + \lambda v_i)\}$ where the scalar λ is contained in some neighborhood Λ of zero and $(\bar{u}\backslash u_j', \bar{u}_i + \lambda v_i) = (\bar{u}_1, \ldots, \bar{u}_{j-1}, u_j', \bar{u}_{j+1}, \ldots, \bar{u}_{i-1}, \bar{u}_i + \lambda v_i, \bar{u}_{i+1}, \ldots, \bar{u}_n)$. Because $\hat{B}_j(\lambda)$ is m-dimensional within $\hat{N}$, it is representable as the solution of $\ell-m \equiv L$ continuously differentiable functions: $\hat{B}_j(\lambda) = \{x \in \hat{N} | f_k(x,\lambda) = 0\ \forall_k \in M\}$ where $M = \{1,\ldots,L\}$. We show as a first step that not $i\ \tilde{A}(u)\ j$ for all $u \in N(\bar{u})$ implies, for all $k \in M$ and all $x \in \hat{B}_j(0)$,

$$\text{(1)} \qquad \frac{\partial f_k(x,\lambda)}{\partial \lambda} = 0$$

when evaluated at $\lambda = 0$. In the second step we show that (1) implies $D_{(v_1)}\sigma_j(\bar{u}) = 0$, which contradicts the hypothesis that $i\ A(\bar{u})\ j$ and therefore completes the proof for this case where $1 \leq m \leq \ell-1$.

That not $i\ \tilde{A}(u)\ j$ for all $u \in N(\bar{u})$ implies (1) is seen as follows. Pick an $x \in \hat{B}_j(0)$ and let $u_j^x \in N(\bar{u}_j)$ be maximized on $\hat{B}_j(0)$ at x. Define $\sigma_j(\lambda) = \sigma_j(\bar{u}_1, \ldots, \bar{u}_{i-1}, \bar{u}_i + \lambda v_i, \bar{u}_{i+1}, \ldots, \bar{u}_{j-1}, u_j^x, \bar{u}_{j+1}, \ldots, \bar{u}_n)$. Since σ is strategy-proof $\sigma_j(\lambda)$ solves, for all $\lambda \in \Lambda$, the maximization problem $\max_{y \in \hat{N}} u_j^x(y)$ subject to $f_k(y,\lambda) = 0, \forall k \in M$. Regularity implies that the usual first order conditions hold; thus

$$\text{(2)} \qquad \nabla u_j^x[\hat{\sigma}_j(\lambda)] = \sum_{k \in K} \delta_k(\lambda) \nabla f_k[\hat{\sigma}_j(\lambda), \lambda]$$

$$f_k(\hat{\sigma}_j(\lambda), \lambda) = 0 \qquad \forall k \in M$$

where ∇u_j^x and ∇f_k are the gradients with respect to x of u_j^x and f_k respectively and the $\delta_k(\lambda)$ are the Lagrangian multipliers as functions of λ.[1]

By hypothesis, not $i \tilde{A}(\bar{u}\backslash u_j^x) j$ because $(\bar{u}\backslash u_j^x) \in N(\bar{u})$. Suppose not $i \tilde{A}(u\backslash u_j^x) j$ does not imply $\partial f_k(x,0)/\partial\lambda = 0$ i.e. a nomempty $M_1 \subset M$ exists such that $\partial f_k(x,0)/\partial\lambda \neq 0$ if and only if $k \in M_1$. Feasibility of the solution $\hat{\sigma}_j(\lambda)$ implies that, for all $k \in M$ and $\lambda \in \Lambda$, $f_k(\hat{\sigma}_j(\lambda),\lambda) = 0$. Differentiation with respect to λ gives, for all $k \in M$,

$$(3) \qquad \nabla f_k[\hat{\sigma}_j(0),0] \cdot \frac{d\hat{\sigma}_j(0)}{d\lambda} + \frac{\partial f_k[\hat{\sigma}_j(0),0]}{\partial\lambda} = 0.$$

This implies that, for all $k \in M_1$,

$$(4) \qquad \nabla f_k[\hat{\sigma}_j(0),0] \cdot \frac{d\hat{\sigma}_j(0)}{d\lambda} = \begin{cases} \text{not } 0 & \text{if } k \in M_1 \\ 0 & \text{if } k \notin M_1 \end{cases} .$$

The assumption not $i \tilde{A}(\bar{u}\backslash u_j^x) j$ implies that

$$(5) \qquad \frac{du_j^x[\hat{\sigma}_j(0)]}{d\lambda} = \nabla u_j^x[\hat{\sigma}_j(0)]. \frac{d\hat{\sigma}_j(0)}{d\lambda} = 0$$

Substituting (2) into (5) in conjunction with (4) results in

$$(6) \qquad \sum_{k \in M_1} \delta_k(0) \nabla f_k[\hat{\sigma}_j(0),0] \cdot \frac{d\hat{\sigma}_j(0)}{d\lambda} = 0 .$$

where $\nabla f_k \cdot d\hat{\sigma}_j/d\lambda \neq 0$ for each $k \in M_1$. Pick scalar weights α_k such that (a),

---

[1] We follow the convention that $\nabla u_j^x[\hat{\sigma}_j(\lambda)]$ represents the gradient of u_j^x evaluated at $\hat{\sigma}_j(\lambda)$. Similarly $\nabla f_k[\hat{\sigma}_j(\lambda),\lambda]$ represents the gradient of f with respect to x evaluated at $(\hat{\sigma}_j(\lambda),\lambda)$.

$$(7) \qquad \sum_{k \in M_1} \alpha_k \nabla f_k [\hat{\sigma}_j(0), 0] \cdot \frac{d\hat{\sigma}_j(0)}{d\lambda} \neq 0$$

and (b), for each $k \notin M_1$, $\alpha_k = 0$. Define $v_j(x) \equiv \sum_{k \in M} \alpha_k f_k(x,0)$.

Broad applicability implies that, for a small enough scalar $\gamma > 0$, $\hat{u}_j^x = u_j^x + \gamma v_j$ is an element of $N(\bar{u}_j)$. That $\hat{\sigma}_j(0)$ maximizes $\hat{u}_j^x$ on $\hat{B}_j(0)$ is a consequence of regularity and the fact that the first order conditions continue to hold at $\hat{\sigma}_j(0)$:

$$(8) \qquad \begin{aligned} \nabla \hat{u}_j^x[\hat{\sigma}_j(0)] &= \nabla u_j^x[\hat{\sigma}_j(0)] + \gamma \nabla v_j[\hat{\sigma}_j(0)] \\ &= \sum_{k \in M} [\delta_k(0) + \gamma \alpha_k] \cdot \nabla f_k[\hat{\sigma}_j(0), 0] \end{aligned}$$

where the second line is obtained by substituting from (2) and (7) and rearranging. But

$$\nabla \hat{u}_j^x[\hat{\sigma}_j(0)] \cdot \frac{d\hat{\sigma}_j(0)}{d\lambda} = \nabla u_j^x[\hat{\sigma}^j(0)] \cdot \frac{d\hat{\sigma}_j(0)}{d\lambda} + \gamma \nabla v_j[\hat{\sigma}_j(0)] \cdot \frac{d\hat{\sigma}_j(0)}{d\lambda}$$

$$(9) \qquad = \gamma \nabla v_j[\hat{\sigma}_j(0)] \cdot \frac{d\hat{\sigma}_j(0)}{d\lambda}$$

$$\neq 0 .$$

where the inequality follows from (7) and the definition of $v_j(x)$. This implies that $D_{(v_1)} \hat{u}_j^x \sigma_j(\bar{u} \backslash \hat{u}_j^x) \neq 0$ because $\hat{\sigma}_j(\lambda)$ is a feasible (if not optimal) path for agent $\hat{u}_j^x$ to follow when λ varies. Thus $i \tilde{A}(\bar{u} \backslash \hat{u}_j^x)\ j$ and if $\partial f_k(x,0)/\partial \lambda \neq 0$ for some $x \in \hat{B}_j(0)$, a $\hat{u}_j^x$ exists such that (a) $(\bar{u} \backslash \hat{u}_j^x) \in N(\bar{u})$ and (b) $i \tilde{A}(\bar{u} \backslash \hat{u}_j^x)$, which contradicts the hypothesis that not $i \tilde{A}(u)\ j$ for all $u \in N(\bar{u})$. Therefore $\partial f_k(x,0)/\partial \lambda = 0$ for all $x \in \hat{B}_j(0)$. A useful implication of this follows directly from (3): for all $k \in M$,

$$\nabla f_k[\hat{\sigma}_j(0),0]\cdot\frac{d\hat{\sigma}_j(0)}{d\lambda} = 0\ . \tag{10}$$

This completes the first step of this case's proof.

Step two of the proof is to show that if not i $\tilde{A}(\bar{u})$ j and if $\partial f_k(x,0)/\partial\lambda = 0$ for all $x \in \hat{B}_j(0)$ and $k \in M$, then, for all $v_i \in C^2(X)$, $D_{(v_i)}\ \sigma_j(\bar{u}) = 0$, which is to say that not i $A(\bar{u})$ j. Let $\hat{\sigma}_j(\lambda) = \sigma_j(\bar{u}\backslash\bar{u}_i + \lambda v_i)$, i.e. in the notation of the first step $\bar{u}_j = u_j^x$. Suppose the result is not true, i.e. a v_i does exist such that $D_{(v_i)}\ \sigma_j(\bar{u}) = d\hat{\sigma}_j(0)/d\lambda \neq 0$.

Rotate and translate the coordinate system of X so that (a) the origin is $x_0 \equiv \hat{\sigma}_j(0) \equiv \sigma_j(\bar{u})$, (b) the x_1 axis is in the direction $d\hat{\sigma}_j(0)/d\lambda$, and (c) the x_2 axis is in the direction $\nabla\bar{u}_j[\hat{\sigma}_j(0)]$. Requirements (b) and (c) are consistent because not i $\tilde{A}(\bar{u})$ j implies orthogonality of $\nabla\bar{u}_j[\hat{\sigma}_j(0)]$ and $d\hat{\sigma}_j(0)/d\lambda$. The directions of the remaining $\ell-2$ axes may be set arbitrarily, provided orthogonality is preserved.

The first order conditions (2) may be differentiated with respect to λ to obtain:

$$
\begin{bmatrix}
U_{11} & & U_{12} & \cdots & U_{1\ell} & \frac{\partial f_1(x_0,0)}{\partial x_1} & \cdots & \frac{\partial f_L(x_0,0)}{\partial x_1} \\
U_{21} & & U_{22} & \cdots & U_{2\ell} & \frac{\partial f_1(x_0,0)}{\partial x_2} & \cdots & \frac{\partial f_L(x_0,0)}{\partial x_2} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
U_{\ell 1} & & U_{\ell 2} & \cdots & U_{\ell\ell} & \frac{\partial f_1(x_0,0)}{\partial x_\ell} & \cdots & \frac{\partial f_L(x_0,0)}{\partial x_\ell} \\
\frac{\partial f_1(x_0,0)}{\partial x_1} & & \frac{\partial f_1(x_0,0)}{\partial x_2} & \cdots & \frac{\partial f_1(x_0,0)}{\partial x_\ell} & 0 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{\partial f_L(x_0,0)}{\partial x_1} & & \frac{\partial f_L(x_0,0)}{\partial x_2} & \cdots & \frac{\partial f_L(x_0,0)}{\partial x_\ell} & 0 & \cdots & 0
\end{bmatrix}
$$

(11)

$$
\cdot
\begin{bmatrix}
\frac{d\hat{\sigma}_{j1}(0)}{d\lambda} \\
\frac{d\hat{\sigma}_{j2}(0)}{d\lambda} \\
\cdots \\
\frac{d\hat{\sigma}_{j\ell}(0)}{d\lambda} \\
-\frac{d\delta_1(0)}{d\lambda} \\
\cdots \\
-\frac{d\delta_L(0)}{d\lambda}
\end{bmatrix}
=
\begin{bmatrix}
\sum_{k\in M} \delta_k(0)\frac{\partial f_k(x_0,0)}{\partial x_1 \partial\lambda} \\
\sum_{k\in M} \delta_k(0)\frac{\partial f_k(x_0,0)}{\partial x_2} \\
\cdots \\
\sum_{k\in M} \delta_k(0)\frac{\partial f_k(x_0,0)}{\partial x_\ell \partial\lambda} \\
-\frac{\partial f_1(x_0,0)}{\partial\lambda} \\
\cdots \\
-\frac{\partial f_L(x_0,0)}{\partial\lambda}
\end{bmatrix}
$$

where

$$U_{IJ} = \frac{\partial^2 \bar{u}_j(x_0)}{\partial x_I \partial x_J} - \sum_{k \in M} \delta_k(0) \frac{\partial^2 f_k(x_0,0)}{\partial x_I \partial x_J} \quad .$$

The choice of coordinate system immediately entails three conclusions. First, the orientation of the x_1 axis means that

$$(12) \qquad \frac{\partial \hat{\sigma}_{jI}(0)}{\partial \lambda} = \begin{cases} \text{not } 0 & \text{if } I = 1 \\ 0 & \text{if } I \in \{2,\ldots,\ell\} \end{cases} \quad .$$

Second, equation (10) together with (12) implies that $\partial f_k(x_0,0)/\partial x_1 = 0$ for all $k \in M$. Third, the proof's first step established that, for all $x \in \hat{B}_j(0)$ and all $k \in M$, $\partial f_k(x,0)/\partial\lambda = 0$; therefore $\partial^2 f_k(x_0,0)/\partial x_1 \partial\lambda = 0$ for all $k \in M$ because (a) movement in the direction of $d\hat{\sigma}(0)/d\lambda$ is, by definition of the coordinate system, movement out along the x_1 axis and (b) the orthoganality of $d\hat{\sigma}(0)/d\lambda$ and each $\nabla f_k(x_0,0)$ implied by (10) means that the movement out along the x_1 axis is movement within $\hat{B}_j(0)$. In addition, the second order conditions for $\hat{\sigma}_j(0)$ to be a regular maximum require that $U_{11} \neq 0$. These restrictions on the values that terms within (11) may take reduce the first of the $\ell + L$ equations of (11) to

$$(13) \qquad U_{11} \frac{d\hat{\sigma}_{j1}(0)}{d\lambda} = 0.$$

Since $U_{11} \neq 0$, necessarily $d\hat{\sigma}_{j1}(0)/d\lambda = 0$, which contradicts the assumption $i\ A(\bar{u})\ j$. This completes the proof for the case $1 \leq m \leq \ell-1$.

The proof for m=0 is simpler. Dimensionality of zero for $B_j(u)$ for all $u \in N(\overline{u})$ implies that $B_j(u)$ is just a point in X, i.e. j's allocation is imposed on him by the other agents. The hypothesis that i $A(\overline{u})$ j means that i can move the point $B_j(u)$ in X, i.e. a $v_i \in C^2(X)$ exists such that

$$(14) \qquad \frac{d\,\hat{\sigma}_j(0)}{d\lambda} = \frac{d\hat{B}_j(0)}{d\lambda} \neq 0$$

where $\hat{\sigma}_j(\lambda) = \sigma_j(\overline{u}\backslash\overline{u}_i + \lambda v_i)$ and $\hat{B}_j(\lambda) = B_j(\overline{u}\backslash\overline{u}_i + \lambda v_i)$. The hypothesis that not i $\tilde{A}(\overline{u})$ j implies that

$$(15) \qquad \nabla\overline{u}_j[\sigma_j(\overline{u})] \cdot \frac{d\hat{B}_j(0)}{d\lambda} = 0.$$

Pick a $v_j \in C^2(X)$ such that

$$(16) \qquad \nabla v_j[\sigma_j(\overline{u})] \cdot \frac{d\hat{B}_j(0)}{d\lambda} \neq 0.$$

BA guarantees the existence of a scalar $\gamma > 0$ small enough such that $\hat{u}_j \equiv \overline{u}_j + \partial v_j \in N(\overline{u}_j)$. Thus

$$(17) \qquad \nabla\hat{u}_j[\sigma_j(\overline{u})] \cdot \frac{d\hat{B}_j(0)}{d\lambda} = \{\nabla\overline{u}_j[\sigma_j(\overline{u})] + \nabla v_j[\sigma_j(\overline{u})]\} \cdot \frac{d\hat{B}_j(0)}{d\lambda} \neq 0 .$$

which is to say that i $\tilde{A}(\overline{u}\backslash\hat{u}_j)$ j. Moreover, since $\hat{u}_j \in N(\overline{u}_j)$, $(\overline{u}\backslash\hat{u}_j) \in N(\overline{u})$. Therefore i $\tilde{A}(\overline{u}\backslash\hat{u}_j)$ j contradicts the hypothesis that not i $\tilde{A}(u)$ j for all $u \in N(\overline{u})$. This completes the proof for the second case where m=0 and thus completes the proof of the entire Lemma.

---

[1] It is permissible to write $d\hat{B}_j(0)/d\lambda$ because $\hat{B}_j$ is a point.

Reference

[1] Satterthwaite, M. and H. Sonnenschein, "Strategy-Proof Allocation Mechanisms," Discussion Paper No. 395, Center for Mathematical Studies in Economics and Management Science, Northwestern University, August 1979.