Equilibrium Selection and Stability for the Groves Ledyard Mechanism

Scott E. Page*

Troy Tassier†

September 4, 2003

Abstract

In their seminal paper Groves and Ledyard (1976) construct a balanced incentive compatible mechanism that solves the free-rider problem. In subsequent research, Bergstrom, Simon, and Titus (1983) prove that there exist numerous asymmetric equilibria in addition to the symmetric equilibrium. In the present paper we explicitly solve for the additional equilibria and use computational experiments to examine the structure and stability of the set of equilibria of the Groves Ledyard Mechanism. We find that all of the equilibria found by Bergstrom, Simon, and Titus are unstable and that for a high level of the punishment parameter these equilibria do not exist. Further we find that there exists an additional boundary equilibrium for each of the equilibria found by Bergstrom, Simon, and Titus. The boundary equilibria are all stable.

*Departments of Political Science and Economics and Centers for the Study of Complex Systems and Policy Studies University of Michigan, Ann Arbor, MI 48106.

†Department of Economics, Fordham University, Bronx, NY 10458.

1 Introduction

In their seminal paper describing a solution to the free rider problem Groves and Ledyard (1976) construct a balanced, incentive compatible mechanism for an economy with both public and private goods. The paper provides a proof that any equilibrium for their mechanism produces an efficient amount of the public good. This was a major contribution to the field of mechanism design in that it showed how carefully constructed payoff functions could overcome incentive issues, including the free rider problem.

The goal of Groves and Ledyard was to construct a mechanism that could overcome the free rider problem. They did not claim that their mechanism generated a unique or stable equilibrium. However, it is relatively straightforward to show that their mechanism does generate a unique equilibrium when agent preferences have a quasi-linear form. Thus, in a restricted class of functions, their mechanism does solve the free rider problem in the stronger sense of generating a unique equilibrium. Subsequent research by Bergstrom, Simon, and Titus (1988) (hereafter BST) showed that the Groves Ledyard Mechanism has an enormous number of equilibria for a more general class of preferences that allow for the value of the private good to depend on the amount of the public good provided. Specifically, BST show that the Groves Ledyard Mechanism has an additional equilibrium for every possible subset of agents consisting of fewer than half of the agents. So, for N larger than a few dozen there are literally billions and billions of these equilibria.

The existence of billions of equilibria is one thing, but the tendency for a system of dynamically adjusting agents to locate those equilibria is quite another. For example, it might well be that many of the equilibria are unstable, so they might never be attained. Or, even if stable, these equilibria might have relatively small basins of attraction compared to the symmetric Groves Ledyard equilibrium

for a given learning dynamic. In either case, the original Groves Ledyard solution might be the best predictor for the system. Further, since these other equilibria are efficient then the differences between these billions of equilibria and the Groves Ledyard equilibrium are purely allocative. In each of these equilibria, some of the agents may do better than they would in the Groves Ledyard Equilibrium, and some may do worse. The question is how much better.

In this paper, we analyze the Groves Ledyard Mechanism mathematically and computationally to test whether these billions of equilibria matter. We construct a model with agents who use a variant of the best response dynamic called Q-learning and explore the Groves Ledyard mechanism using a specific functional form consistent with the more general preference structure assumed by BST. We find, much to our great surprise, that the system never settles onto any of the BST equilibria and that for a sufficiently large punishment parameter of the mechanism, the BST equilibria disappear. Moreover, if started at a BST equilibrium and perturbed, the system does not return to the equilibrium. In short, we find that all of the BST equilibria are unstable if they exist at all.

This observation leads to a conclusion that the Groves Ledyard Mechanism must have even more equilibria than BST claim. And, in fact, for our specific functional form, we find that there exists one boundary equilibrium for each of the BST equilibria. Thus, there are twice as many additional equilibria as originally thought. Moreover, these new boundary equilibria are stable, have reasonably sized basins of attraction given our assumptions about dynamics, and, if we bound the space of possible messages, they are not efficient.

The remainder of this paper is organized as follows. We first present the Groves Ledyard mechanism and then consider its application for a specific functional form. Using that functional form, we solve for the BST equilibria. In the original

BST paper, the equilibria were found using a change of basis and were not explicitly calculated. We find, somewhat counter intuitively, that the location of these equilibria does not depend in any way on the punishment parameter of the Groves Ledyard mechanism. We then construct a computational model of the Groves Ledyard mechanism. We show that the agents in the model either find the symmetric Groves Ledyard Equilibrium or head off to the boundary. We then test the stability of the BST equilibria and find that for low values of the punishment parameter these equilibria are unstable in our model. We further find that for high values of the punishment parameter the BST equilibria do not exist. We then discuss why these equilibria fail to exist and why they are unstable when they do. Finally we discuss how one can remove all equilibria but the symmetric Groves Ledyard equilibrium. One can accomplish this by setting the punishment parameter to a sufficiently high level or by setting a moderate tax and imposing a budget constraint. This second approach is important because the first can be interpreted as saying that if you force agent to all choose the same level of the public good, they will do it.

2 The Groves Ledyard Mechanism

We consider a restriction of the Groves Ledyard mechanism to an economy with a single good of each type. We index the agents by $i \in \{1, \dots, N\}$. Each agent has an initial wealth w_i and consumes a private good x_i and a public good y . Utility is a function of both goods $[U_i(x_i, y)]$. Agent i pays a percentage of the public good denoted by α_i , which we assume to be $\frac{1}{N}$. In the Groves Ledyard Mechanism, each agent sends a message m_i which is the amount of the public good that the agent would like to have produced. The vector of these messages can be denoted by $\vec{m}$. The Groves Ledyard mechanism sets

$$y(\vec{m}) = \sum_{i=1}^N m_i$$

Agents pay a tax $C_i(\vec{m})$ which is the following function of the vector of messages

$$C_i(\vec{m}) = \frac{y(\vec{m})}{N} + \frac{\gamma}{2} \left[\frac{N-1}{N} (m_i - \bar{m}^i)^2 - \sum_{j \neq i} \frac{1}{N-2} (m_j - \bar{m}^i)^2 \right]$$

where $\bar{m}^i$ equals the average message sent by the $N - 1$ agents other than i . Notice that the summation term does not depend on m_i , so it is out of the agent's control. Thus for incentive purposes all that matter are the tax share $\alpha_i y(\vec{m})$ and the difference between the message of agent i and the average message sent by agents other than i . The punishment parameter γ describes pressures for conformity. By increasing γ , the mechanism can induce everyone to send the same message. For low γ there will be less pressure to converge on a common message. Thus, γ will play an important role in the dynamics of the system. For the case of quasi-linear preferences (which have a unique equilibrium) Chen and Tang(1998) have studied the impact of varying γ in experimental settings and found that increasing γ leads to agents being more likely to converge to the equilibrium.

An agent's budget constraint requires that $x_i(\vec{m}) = w_i - C_i(\vec{m})$. An *equilibrium* of this mechanism satisfies

$$U_i(x_i(\vec{m}), y(\vec{m})) \geq U(x(\hat{m}_i, m_{-i}), y(\hat{m}_i, m_{-i})) \quad \forall \hat{m}_i \text{ and } \forall i$$

Because the messages are the agent's contributions to the public good, we hereafter replace the m_i 's with y_i and define $\bar{y}_i$ to be the average contribution to the public good of the agents other than i .

2.1 Quasi-Linear Preferences

We first show that with quasi-linear preferences, the Groves Ledyard Mechanism has a unique equilibrium. Though the unique quasi-linear case is not the focus of this paper, it is useful as a point of comparison for the functional form that we consider. With quasi-linear preferences, the utility function for the agents can be written as

$$U_i(x_i, y) = x_i + B(y)$$

where $B(y)$ is a concave increasing function of y , the sum of the contributions to the public good. The budget constraint requires that

$$x_i = w_i - \frac{y}{N} - \frac{\gamma}{2} \left[\frac{N-1}{N} (y_i - \bar{y}_i)^2 - \sum_{j \neq i} \frac{1}{N-2} (y_j - \bar{y}_i)^2 \right]$$

If we substitute the budget constraint into the utility function, we obtain

$$U_i(x_i, y) = w_i - \frac{y}{N} - \frac{\gamma}{2} \left[\frac{N-1}{N} (y_i - \bar{y}_i)^2 - \sum_{j \neq i} \frac{1}{N-2} (y_j - \bar{y}_i)^2 \right] + B(y)$$

The first order condition on y_i is as follows:

$$-\frac{1}{N} - \gamma \frac{N-1}{N} (y_i - \bar{y}_i) + B'(y) = 0$$

That this system of N equations has a unique equilibrium is easy to see. First, note that if $y_i = \frac{y^*}{N}$ for all i , then this equation reduces to $\frac{1}{N} = B'(y^*)$ which has a unique solution because of the concavity of $B(\cdot)$. Next, suppose that the agents send different messages, then some of the $(y_i - \bar{y}_i)$ terms will be positive and some will be negative. But this cannot be since $\gamma \frac{(N-1)}{N} (y_i - \bar{y}_i) = B'(y) - \frac{1}{N}$ for all i . Thus in the

case of quasi-linear preferences the equilibrium is unique and symmetric.

2.2 The BST Equilibria

We now turn to a specific functional form that belongs to the more general class of preferences considered by BST and show that it generates multiple equilibria. BST rely on a change of basis to prove their result. Though that makes the math easier, it makes it harder for the reader to maintain an intuition for the equilibria they find. BST choose the following general class of utility functions

$$U_i(x_i, y) = A(y)x_i + B_i(y)$$

With this functional form, as more of the public good is provided, the private good is more valuable. This construction makes sense if we think of public goods like roads and private goods like automobiles. The value of an automobile increases as more roads are built. It also makes sense if we think of the public good as clean air and the private good as land.

Here, we consider a specific functional form from their general class. We choose $A(y) = y$ and $B_i(y) = \psi y - \phi y^2$. We also assume that all agents have the same wealth, $w_i = w$ for all i . This simplifies the algebra without compromising the analysis. We can then write the utility function for agent i as:

$$U_i(x_i, y) = yw - yC_i + \psi y - \phi y^2$$

The first order condition on y_i now becomes

$$w - C_i - \frac{y}{N} - y\gamma \frac{N-1}{N}(y^i - \bar{y}^i) + \psi - 2\phi y = 0$$

We first solve for the symmetric equilibrium where $y_i = \frac{y}{N}$ for all i . The first order condition reduces to

$$w - \frac{y}{N} - \frac{y}{N} + \psi - 2\phi y = 0$$

which has a unique solution which we denote y^* . We call this the symmetric Groves Ledyard equilibrium. We next construct the BST equilibria from this equilibrium as follows. We first choose a subset of agents of size k , where k is less than $\frac{N}{2}$. We assume that those agents send the low message $y^l < y^*/N$. We define ϵ to be equal to $\frac{y^*}{N} - y^l$. We then assume that the other $N - k$ players send a high message $y^u = \frac{y^*}{N} + \delta$, where $(N - k)\delta = k\epsilon$. This last assumption implies that

$$ky^l + (N - k)y^u = y^*$$

so that the amount of the public good produced equals y^* . Equivalently, we have that $\delta = \frac{k\epsilon}{(N-k)}$. It suffices to show that for some $\epsilon > 0$ the conditions for an equilibrium are satisfied. To simplify notation define the *penalty* to be that portion of the Groves Ledyard tax not equal to one N th of the public good. For the agents sending the low message define this as

$$T^l = \frac{\gamma}{2} \left[\frac{N-1}{N} (y^l - \bar{y}^l)^2 - \sum_{j \neq i} \frac{1}{N-2} (y_j - \bar{y}^l)^2 \right]$$

where $\bar{y}^l$ equals the average contribution of the agents excluding one of the low contributing agents. Define T^u and $\bar{y}^u$ similarly. It is straightforward to show that the agents who contribute less pay a larger penalty. There are fewer of them so they differ from the mean of the others by more than those who contribute more. The first order conditions for the two types of agents can be written as follows:

$$w - T^l - \frac{y}{N} - y\gamma \frac{N-1}{N}(y^l - \bar{y}^l) + B'(y) = 0$$

$$w - T^u - \frac{y}{N} - y\gamma \frac{N-1}{N}(y^u - \bar{y}^u) + B'(y) = 0$$

Using the respective taxes for each type of agent and the first order conditions above we can calculate ϵ explicitly. In the appendix we show that $\epsilon = \frac{2(N-2)(N-k)y^*}{N(N-2k)}$.

Notice though that for k larger than $\frac{N}{2}$ we obtain the same equilibria that we get for k less than $\frac{N}{2}$. Therefore, there is an equilibrium for each subset of k agents where k is strictly greater than $\frac{N}{2}$. These are precisely the BST equilibria, but here we have them in closed form for this particular functional form. Notice that these equilibria are independent of γ which determines the extent of the penalty.

2.3 Examples

To provide some intuition for how these equilibria depend upon the various parameters, we consider some examples. First, suppose that $N = 6$ and $k = 1$. Solving for ϵ gives $\epsilon = \frac{2(N-2)(N-k)y^*}{N(N-2k)} = \frac{5y^*}{3}$ so each agent contributing less would send a message that contributed a negative amount to the public good. For $k = 2$, the value of ϵ equals $\frac{8y^*}{3}$ so as the number of agents in the group sending the low messages increases, the amount by which they are less than the symmetric equilibrium increases. In other words, more equal sized groups making high and low contributions demand greater distortions in the asymmetric equilibria; the messages sent by the agents are farther from the symmetric equilibrium. Also note that the BST equilibria are farther from the symmetric equilibrium as N increases. For $N = 4$ and $k = 1$, ϵ equals $\frac{3y^*}{2}$. Thus

as N increases there are at least two effects: First, there are more equilibria. Second, the asymmetric equilibria have larger allocative distortions.

3 Convergence/ Dynamics: Computational Results

In this section we begin to explore the dynamic properties of the equilibria discussed above. Specifically we address the ability of agents to coordinate on and converge to an equilibrium, the basins of attraction of the equilibria, and the stability of the equilibria.

We investigate two classes of utility functions: First we examine the quasi-linear case with preferences identical to Chen and Tang (1998). The experiments of Chen and Tang have 5 agents each having preferences of the form: $a_i Y - b_i Y^2 + c_i - C_i(\vec{m})$. Second, we examine the non-linear BST preferences described above. Specifically, we use the following utility function: $U_i(x_i, y) = A(y)x_i + B(y) = yx_i + \psi y - \phi y^2$. These simulations with the BST preferences have 3, 4, or 5 agents with parameters $w_i = w = 1$, $\psi = 2$, and $\phi = 1$. We bound the minimum message that can be sent to be the sum of the wealth of the other agents in the economy, $-N * w$. We vary the punishment parameter, γ , across simulations in order to view its effect.

3.1 Convergence

As an initial condition of each experiment assume that in period one each agent plays a random message. Agents then play a response to these random initial messages. The initial random messages are chosen from a uniform distribution over $[-2w, 2w]$.

We investigate two dynamic processes: best response and a partial adjustment process that is a simple version of Q-learning (Watkins 1989.) With best response agents choose the message in period t , m_i^t , such that utility is maximized given the

vector of messages in period $t - 1$, $\vec{m}^{t-1}$. Specifically let the best response of agent i in period t be: $R_i^t(\vec{m}^{t-1})$ such that: $U_i(m_i, \vec{m}^{t-1}) > U_i(\hat{m}_i, \vec{m}^{t-1}) \forall \hat{m}_i$.

With the partial adjustment dynamic agents send a message that is a weighted combination of last periods message and this periods best response as a function of a parameter q . Specifically the agents send the message $m_i^t = (1 - q)R_i^t + qm_i^{t-1}$ with $q \in [0, 1)$. Thus if $q = 0$ the partial adjustment dynamic is equivalent to the best response dynamic.

In computational models with discrete time it is sometimes difficult for agents to converge to an equilibria. With best response agents sometimes systematically over-shoot and under-shoot the equilibrium and thus do not converge.¹ Using a partial adjustment dynamic smoothes the best response of agents across time and allows the best response of agents in discrete time to approach the best response of agents in continuous time. Thus with partial adjustment it is easier for agents to find an equilibrium in some cases. This appears to be the case here.

When we use the parameters and quasi-linear utility function of Chen and Tang (1998) we found that our simulations converge to the unique symmetric equilibrium in all trials with sufficiently high q . When q is too low agents enter into message cycles that do not converge to an equilibrium. Interestingly the level of q needed to ensure convergence depends on the level of γ . For γ greater than 12 the agents converge for any level of q . For $\gamma = 1$, q needs to be greater than approximately 0.55 to ensure convergence.²

We now investigate the Groves Ledyard Mechanism when agents have BST preferences. When the agents play a pure best response they do not converge to

¹For a specific discussion of the partial adjustment dynamic employed here see Van Huyck, Cook, and Battalio (1994). For a discussion of overadjustment in a specific public goods experimental setting see Smith (1979).

²The code used to implement all experiments in this paper is available from the authors upon request.

Figure 1: Cycles result when the agents play a pure best response to the messages of other agents.

either the symmetric or the asymmetric equilibria. Agents oscillate between sending a message that is too high or too low, relative to an equilibrium. The history of messages for one of the agents is shown in Figure 1.

However, if we use agents who partially adjust they do converge to the symmetric equilibrium in a large fraction of the runs. Table 1 shows the percent of 500 runs that converge to the symmetric equilibrium with $q = 0.9$ for 3, 4, and 5 agents and levels of γ between 0.25 and 1.00.³

In the runs that do not converge to the symmetric equilibrium the agents either cycle or converge to a boundary equilibrium that is different than the equilibria described by BST.⁴ In the boundary equilibrium one of the agents sends the minimum

³For the three levels of γ less than 1 a substantially lower value of q allows convergence. However at $\gamma = 1$ the runs often do not converge and instead cycle for low values of q . Thus we chose a higher value of q in order to keep the results consistent across the parameter values.

⁴Note that only 8 of the 500 runs did not converge for $N = 5$, $\gamma = .75$; 7 did not converge for $N = 5$, $\gamma = 1.0$; and 5 did not converge for $N = 4$, $\gamma = 1.0$. Thus almost all runs converge to either the symmetric equilibrium or one of the boundary equilibria.

Table 1: Percent of runs that converge to the symmetric equilibrium.

N	$\gamma = .25$	$\gamma = .50$	$\gamma = .75$	$\gamma = 1.0$
3	84	84	81	71
4	90	89	59	4
5	93	77	0	0

message and the remaining agents send the same positive message. For example in the case with three agents and $\gamma = .25$ one agent sends the minimum message of -3 while the other two agents send the message 2.16. Note that if the message space was not bounded from below the agent sending the minimum message would continue decreasing her message. Consequently the agents sending the positive message would increase their message, prompting a lower message from the first agent, and so on. In this way the agents would perpetually move toward positive and negative infinity if the lower bound was not in place. The agents never converge to any of the asymmetric BST equilibria. Thus, if the agents converge, they either converge to the symmetric equilibrium or they go to the boundary in all of the trials of our simulation.

Surprisingly, as γ increases the agents converge to the symmetric equilibrium less often. The reason for this result becomes clear after we gain a greater understanding of the structure of the asymmetric equilibria discussed in the next two subsections.

3.2 Stability of the Asymmetric Equilibria

The fact that none of the runs converge to any of the BST equilibria suggests two possibilities: First the BST equilibria could be unstable. By unstable we mean that if the messages of the agents are perturbed slightly from the messages of a given equilibrium the agents do not return to that equilibrium. Second, the equilibria could be stable but have small basins of attraction relative to the symmetric and boundary equilibria.

We can test the stability of the equilibria by perturbing some of the agents slightly away from their equilibrium messages and seeing if the systems returns to the equilibrium. We do this in the following way for the case with three agents: We leave two of the agents at their equilibrium messages and move the other agent away from their equilibrium message by a small amount. In our experiments we find that the symmetric equilibrium and the boundary equilibria are stable but the BST equilibria are not.

We begin by testing the stability of the asymmetric equilibria. Using our results from the previous section, we find $\epsilon = 3/2$ for the case where $N = 3$ and $k = 1$. Thus there is an asymmetric equilibrium when one agent sends the message $3/8 - 3/2 = -9/8$ and two agents send the message $3/8 + 3/4 = 9/8$. Since there can be one subset of agents greater than $N/2$ for $N = 3$ this is the only configuration for the BST equilibria in this example. We then verify that the BST equilibria were in fact equilibria by starting the agents at the equilibrium and verifying that no agent chose to deviate. We find this to be true for low γ . But an interesting thing happens for sufficiently high γ ; the low agent chooses to deviate by sending a lower than equilibrium message and the system then converges to a boundary equilibrium. This implies that the BST equilibrium is not actually an equilibrium for high γ . This is a point we return to in a moment.

For low γ the BST equilibria are unstable. In our experiments, for any single agent deviating an arbitrary amount from the asymmetric equilibria, the agents converge to either the symmetric equilibrium (if the message of an agent is perturbed toward the symmetric equilibrium) or a boundary equilibrium (if the message of an agent is perturbed away from the symmetric equilibrium). Thus as one would expect the basins of attraction for the symmetric and the boundary equilibria shift at the unstable BST equilibrium.

To better understand why the BST equilibria cease to exist for high γ consider the following decomposition of the agent utility function. Agents care about the amount of the public good provided and about their respective taxes. We can write the public good portion of utility as:

$$PGU = y(w_i - \frac{y}{N}) + B(y)$$

And we can write the punishment portion of the utility as:

$$PUNU = y(-\frac{\gamma}{2}[\frac{N-1}{N}(m_i - \bar{y}_i)^2 - \sum_{j \neq i} \frac{1}{N-2}(m_j - \bar{y}_i)^2])$$

To better understand the structure of the Groves Ledyard Mechanism in general and the BST equilibria in particular we turn to a graphical representation. Figure 2 and Figure 3 show the utility for the high and low agents at a BST equilibrium decomposed into the punishment and public good portion of utility when there are three agents and $\gamma = 0.25$. For the high agents (Figure 2) both the punishment portion and the public good portion of utility are at a maximum at the equilibrium message of $9/8$. Thus total utility also is at a maximum. For the low agent (Figure 3) the public good portion of utility is at a maximum (its second derivative is negative) but the punishment portion of utility is at a minimum (its second derivative is positive) at the equilibrium message of $-9/8$. Note that the punishment utility is increasing from the BST equilibrium to the boundary.

Now recall that the location of the BST equilibria is independent of γ as we showed earlier. Additionally, note that the public good portion of utility for the agents is independent of γ but the punishment portion of utility decreases linearly in γ . So, as we increase γ the total utility of the low agent decreases. To see this compare Figure 4 with $\gamma = 0.50$ to Figure 3 with $\gamma = 0.25$. If we increase γ to

Figure 2: Utility of the high message agents when $\gamma = 0.25$.

Figure 3: Utility of the low message agent when $\gamma = 0.25$.

Figure 4: Utility of the low message agent when $\gamma = 0.5$.

a sufficiently high level the total utility becomes a local minimum, (ie, its second derivative becomes positive.) Thus for sufficiently large γ the BST equilibria do not exist. If γ is sufficiently large the low agent prefers the boundary message to the message of the BST equilibrium.

We display the non-existence of the BST equilibria in Figure 5. Here we increase γ to 1.25. As can be seen in the figure the boundary message of -2 yields higher utility than the BST low message of -1.125. The message of -1.125 becomes a local minimum at this level of γ . (This can be better seen in Figure 6 with an even higher level of the punishment parameter, $\gamma = 5$.)

A similar logic explains why the agents converge to the boundary equilibria more frequently for high γ . For high γ , whenever there are sufficiently many agents sending a high message an agent sending a low message wants to move to the lower boundary instead of sending an interior message. This can be seen by comparing figures 5 and 3. Thus the pressure toward the boundary induced by high γ makes it

Figure 5: Utility of the low message agent when $\gamma = 1.25$.

Figure 6: Utility of the low message agent when $\gamma = 5.0$.

Figure 7: Basin of attraction of the symmetric equilibrium for $\gamma = 0.25$ and $N = 3$. The initial message of the third agent is at the symmetric equilibrium, 0.375.

less likely that the agents converge to the symmetric equilibrium.

3.3 Basins of Attraction

To demonstrate when the agents converge to the symmetric equilibrium we present a portion of its basin of attraction in Figure 7. The figure shows the initial message of agent 1 on the horizontal axis and of agent 2 on the vertical axis. The third agent sends an initial message at the symmetric equilibrium, 0.375. As one can see, as long as agents 1 and 2 send sufficiently similar initial messages the agents converge to the symmetric equilibrium when $\gamma = 0.25$. But if one of the agents sends an extreme positive or negative message relative to the other agents (the area in the top left and bottom right of the figure) the agents converge to the boundary equilibrium.

When $\gamma = 1.00$ the symmetric basin of attraction shrinks as can be seen in Figure 8. Note that two areas that converge to the symmetric equilibrium with

Figure 8: Basin of attraction of the symmetric equilibrium for $\gamma = 1.00$ and $N = 3$. The initial message of the third agent is at the symmetric equilibrium, 0.375.

$\gamma = 0.25$ now converge to the boundary equilibrium when $\gamma = 1.00$. These areas are where agents 1 and 2 send similar high initial messages and when one of the agents sends a low message. To see how the dynamics differ γ changes consider the paths of the agent messages for the initial messages of (2, 375, -5) in Figures 9 and 10. When $\gamma = 0.25$ the system converges to the symmetric equilibrium but when $\gamma = 1.00$ the systems converges to the boundary equilibrium.

To understand why this happens first note that the sum of the messages in the initial period is larger than the efficient amount of $9/8$; the provision of the public good is larger than the efficient level with these initial messages. Thus all agents have an incentive to lower their messages in order to bring the public good portion of utility closer to the efficient level. And since the agents are sending different messages the agents have an incentive to move closer to the average message in order to lower their taxes and increase their punishment portion of utility. Thus in regard to the

Figure 9: Agent messages for the first 20 periods with $\gamma = 0.25$ for the initial configuration of $(2, .375, -.5)$. Here the agents converge to the symmetric equilibrium after approximately 50 periods.

Figure 10: Agent messages for the first 20 periods with $\gamma = 1.00$ for the initial configuration of $(2, .375, -.5)$. Here agents converge to the boundary equilibrium.

punishment portion of utility agent 1 would like to decrease her message (since she is above the average of the other agents) and agents 2 and 3 would like to increase their message (since they are below the average of the other agents). Agent 1 clearly wants to lower her message since this action increases both her public good and punishment utility. But agents 2 and 3 may want to increase or decrease their messages depending on whether the public good portion or the punishment portion of utility dominates.

When $\gamma = 0.25$ the effect of the punishment portion of utility is relatively minor and the public good portion of utility is relatively more important. Thus agents 2 and 3 also want to decrease their messages in the first period as shown in Figure 9. But when γ is higher the punishment portion of utility is more important and agent 2 increases her utility slightly in the first period because of an increase in γ . But notice the action of agent 3; agent 3 actually decreases her message more when $\gamma = 1.00$ than when $\gamma = 0.25$. To understand the action of agent 3 remember that for sufficiently high γ a boundary message may be optimal for an agent sending a below average message. This is the case here; the best response of agent 3 is to send the boundary message. (Remember that agent 3 does not move all the way to the boundary even though it is a best response because her actions are dampened by the Q-learning dynamic.) This momentum to the boundary for agent 3 and the convergence of the messages of agents 1 and 2 bring the agents near the boundary equilibrium to which the agents eventually converge.

3.4 Inducing the Symmetric Equilibrium

Recall that the provision of the public good is the efficient level at the asymmetric BST equilibria. Note that this does not imply that the aggregate utility of the agents owing to distributional concerns. The low message agents receive lower utility than the high message agents. At the boundary equilibria there are the same asymmetries

in utility but in addition the provision of the public good is not efficient. Therefore, aggregate utility may be substantially lower than in the symmetric case. Consider our example with 3 agents and $\gamma = 0.25$. At the boundary equilibria the agents sending the high messages receive utility of 2.32 compared to the utility received in the symmetric equilibrium of 1.69. But the agent sending the low message receives lower utility of 0.40. At the boundary equilibria the provision of the public good is greater than the efficient level (1.200 compared to 1.125) and aggregate utility is lower than in the symmetric equilibrium (5.04 compared to 5.07). As mentioned above this is because the agent sending the message at the lower boundary would like to send an even lower message but the boundary prevents her from doing so.

Because the symmetric equilibrium yields higher aggregate utility it is preferable to the population as a whole. But notice that there is a possibility of coalition formation by a subset of the agents. Any subset of agents greater than $N/2$ in number can decide to send the high message of the boundary equilibrium. Once they do so the best response for the remaining agents is to send the lower boundary message. And the coalition of high message agents receives higher utility than they would at the symmetric equilibrium and the coalition of low message agents receive lower utility and pays significantly higher taxes. This argument shows that a move toward the boundary benefits some of the agents. Therefore we cannot appeal to the notion of Pareto dominance to select the symmetric equilibrium. And moreover, the symmetric equilibrium would not lie in the core. However further analysis of the payoff function of the agents reveals that we can guarantee convergence to the symmetric equilibrium by adjusting the punishment parameter γ .

There are two methods to induce symmetry: by ramping up the punishment parameter and by using a budget constraint together with a moderate punishment parameter. As we previously showed it is possible for the taxes of an agent to be

Figure 11: Utility and budget constraint of the low message agent when $\gamma = 0.25$.

greater than her wealth. If we constrain the agents such that they may only send messages such that $w_i \geq T_i(m_i)$ we limit the space of messages that are possible as a function of agent wealth and the punishment parameter γ . Agent wealth less taxes for the low agent at the BST equilibrium is shown in Figure 11 (with $\gamma = 0.25$) and Figure 12 (with $\gamma = 0.50$).

If we restrict agents to sending messages such that $w_i > T_i$ then as γ increases agents are further restricted in the magnitude of negative messages that they can send. Thus one can increase γ to a level that is sufficient to make the symmetric equilibrium the unique equilibrium. Further, this can be done with a γ such that the asymmetric equilibria would still exist were the budget constraint not in place. For instance for $N = 3$ and $\gamma = 0.25$ the message of the BST equilibria is allowed but for $\gamma = 0.50$ the low message of the BST equilibria violates the budget constraint. In the case with $\gamma = 0.50$ the only allowable messages for the low agent are between approximately -0.35 and 0.65 when the other agents send the equilibrium message

Figure 12: Utility and budget constraint of the low message agent when $\gamma = 0.50$.

$9/8$. Thus for sufficiently high γ the BST equilibria violates the budget constraint of $w_i \geq T_i$.

We show how the budget constraint affects equilibrium selection in Table 2. Note that all runs with γ at 0.75 or greater converge to the symmetric equilibrium. In addition almost all of the runs with $\gamma = 0.25$ and $\gamma = 0.50$ also converge to the symmetric equilibrium. Because of the imposition of the budget constraint the space of the allowable messages shrinks. And most of the message restrictions affect the

Table 2: Percent of runs that converge to the symmetric equilibrium when the budget constraint is imposed.

N	$\gamma = .25$	$\gamma = .50$	$\gamma = .75$	$\gamma = 1.0$
3	99	100	100	100
4	100	99	100	100
5	97	100	100	100
10	100	100	100	100
20	100	100	100	100

Figure 13: A hypothetical fitness landscape for the Groves Ledyard Mechanism. The two peaks represent the boundary equilibria (left) and the symmetric equilibrium (right). When γ is low the budget constraint allows for the existence of the boundary equilibria. But when γ is high the budget constraint shifts to the right and makes the boundary non-optimal. Thus the unique equilibrium for sufficiently high γ is the symmetric Groves Ledyard equilibrium.

basin of attraction for the boundary equilibria not the symmetric equilibrium. Thus by increasing γ we can restrict the set of equilibria to include only the symmetric equilibrium.

This is shown most clearly in Figure 13. In this figure we show a hypothetical fitness landscape for the Groves Ledyard Mechanism. The left peak represents one of the boundary equilibria for a low message agent and the right peak represents the symmetric equilibrium. For the low message agents the only allowable messages lie to the right of the budget constraint. When γ is sufficiently low, the budget constraint allows the low messages required to sustain the boundary equilibria. But if γ is sufficiently large the budget constraint binds at a message that makes the boundary non-optimal. Thus the low agents prefer to send a message within the

Table 3: Preference Parameters

Agent	ψ	$\delta =$
1	1.95	0.95
2	2.00	1.00
3	2.05	1.05

basin of attraction of the symmetric equilibrium. This restriction on the messages of the low agents makes it optimal for the high message agents to move into the basin of attraction of the symmetric equilibrium as well. Thus, by setting γ at a sufficiently high level, we can restrict the set of possible messages to rule out the BST and boundary equilibria.

3.5 Heterogeneous Agents

Up until now the agents considered have been homogeneous with respect to preferences. We now briefly consider heterogeneous agents and see whether using the same learning dynamic they exhibit behavior that leads to qualitatively different equilibrium selection results. We consider the case $N = 3$ and set the preference parameters of agent utility as described in Table 3.

Note that agent 2 has the same preferences as in the previous experiments but agents 1 and 3 have slightly different parameters. This change in preferences changes the location of the efficient equilibrium of the agents with respect to γ . When $\gamma = 0.25$ the efficient equilibrium is the vector of messages $(.59, .42, .11)$. And for levels of γ equal to .50, .75, and 1.00 the vectors of equilibrium messages are: $(.48, .39, .26)$, $(.45, .38, .30)$, and $(.43, .38, .32)$. As γ increases the equilibrium messages of the agents converge closer to each other but the total amount of the public good provided remains constant at approximately 1.13.

Table 4 shows that introducing heterogeneity does not change equilibrium

Table 4: Percent of runs that converge to the efficient equilibrium with heterogeneous agents.

N=3	$\gamma = .25$	$\gamma = .50$	$\gamma = .75$	$\gamma = 1.0$
no budget constraint	82	84	84	70
with budget constraint	88	100	100	100

selection that much. When the budget constraint is not enforced the agents converge to the efficient equilibrium in 70-80% of the runs and when the budget constraint is enforced the agents allow converge to the efficient equilibrium in 100% of the runs for $\gamma = .50$ or greater. Note that 12% of the runs do not converge to the efficient equilibrium for $\gamma = .25$. In all of the runs that do not converge to the efficient equilibrium the agents fail to converge to any equilibrium; the agents cycle without ever settling down. This suggests that it may be more difficult to converge for heterogeneous agents. We did not observe agents having the same difficulties of convergence for the homogeneous case above.

Next we further increase the heterogeneity of the agents. Again agent 2 has the same preferences, but agent one has $\psi = 1.9$ and $\delta = .9$ and agent 3 has $\psi = 2.1$ and $\delta = 1.1$. When we do this the agents fail to converge in any of the runs. Note however that this does not change the stability of the equilibria. These results speak to the ability of agents to coordinate. Small amounts of agent heterogeneity do not affect the ability of agents to converge to an equilibrium nor does it affect the likelihood of selecting a particular equilibrium. But for high degrees of agent heterogeneity it may be harder for agents to coordinate their messages and settle on an equilibrium even for a small number of agents.

4 Conclusion

In this paper we examine the stability of the equilibria generated by the Groves Ledyard Mechanism for public good provision using mathematical and computational techniques. Our analysis highlights five primary results: First, we show that the many equilibria proposed by Bergstrom, Simon, and Titus also are efficient. Second, of these equilibria none are stable given our learning dynamic. Third, for large values of the punishment parameter, the solutions proposed by BST cease to be equilibria. They are local minima in the space of possible messages. Fourth, we find that when the BST equilibria do exist there are also boundary equilibria – one for each of the BST equilibria. Finally, we find that when we include a budget constraint, we increase the likelihood that a system of learning agents locates at the symmetric equilibrium.

Given the emphasis of this special issue, it is perhaps equally or more important to discuss the methodological implications of this paper. We set out to explore the characteristics of a model with agents who learn in a system that has many equilibria. At the outset we expected to learn how changing the parameters of the Groves Ledyard Mechanism alters the basins of attraction for the various equilibria. Instead, we find an additional set of boundary equilibria that had previously gone unnoticed. Moreover, we find that these equilibria were both stable and inefficient, while the BST equilibria were the opposite, unstable and efficient, if they even existed. Thus, the Groves-Ledyard result that all equilibria of their mechanism are efficient only applies to the interior equilibria.

This paper makes two methodological contributions: First, it provides additional support for the idea of generative social science put forth by Epstein (2003) and others. A generative social science emphasizes the attainment of as opposed to the existence of equilibria. The BST equilibria exist, but because they are unstable, they cannot be attained. Other equilibria, those at the boundary, were not known to exist,

but their existence became clear once they were attained. Second, the paper can be interpreted as a challenge to those skeptics who argue that "simulations don't prove anything." While that may be partially true, our computational investigations show that something we thought was true, the existence of the BST equilibria, was not. In retrospect, we see that Bergstrom, Simon, and Titus failed to note that a second derivative could be positive for some parameter values. Thus the BST equilibria exist for a smaller set of utility function parameter combinations than previously thought. By constructing a computational model, we were able to check many parts of the BST paper and the Groves Ledyard Mechanism simultaneously. Errors in either paper would have stood a reasonable chance of coming to light in our computational implementation. Therefore, at a minimum, this paper makes clear the benefits of using computational models as a check for logical gaps in mathematical models and for pointing us toward new results.

References

- Bergstrom, T. C., C. P. Simon, and C. J. Titus (1983) "Counting Groves-Ledyard Equilibria via Degree Theory" *Journal of Mathematical Economics* 12, 167-184.
- Chen, Y. and F-F Tang (1998) "Learning and Incentive Compatible Mechanisms for Public Goods Provision: An Experimental Study" *Journal of Political Economy* 106 (3), 633-662.
- Groves, T. and J. Ledyard (1977) "Optimal Allocation of Public Goods: A Solution to the Free Rider Problem" *Econometrica* 45 (4), 783-809.
- Smith, V. (1979) "Incentive Compatible Experimental Processes for the Provision of Public Goods" *Research in Experimental Economics* vol 1, Greenwich CT, JAI Press.
- Van Huyck, J. B., J. P. Cook, R. C. Battalio (1994) "Selection Dynamics, Asymptotic Stability, and Adaptive Behavior" *Journal of Political Economy* 102 (5), 975-1005.
- Watkins, C. (1989) "Learning from Delayed Rewards", thesis University of Cambridge, England.

Appendix

To solve for ϵ we first calculate the the taxes of each type of agent, T^l and T^u . First calculate the difference between the low contribution y^l and the average of the others, $\bar{y}^l$. This equals $-\frac{N-k}{(N-1)}(\delta + \epsilon)$. In the special case where $\delta = \frac{k\epsilon}{(N-k)}$, this becomes $-\frac{\epsilon N}{(N-1)}$. For the agents who contribute more than average the corresponding values are $-\frac{k}{(N-1)}(\delta + \epsilon)$ and $-\frac{\epsilon Nk}{(N-1)(N-k)}$. We next calculate the second term in the tax $\sum_{j \neq i} (y_j - \bar{y}^i)^2$ for each group. For the agents that contribute the lower amount this equals

$$(\delta + \epsilon)^2 \left[\frac{(k-1)(N-k)}{(N-1)(N-2)} \right]$$

For the agents who contribute more this equals

$$(\delta + \epsilon)^2 \left[\frac{(k)(N-k-1)}{(N-1)(N-2)} \right]$$

We can then calculate the total tax for both types of agents. For the agents who contribute less the penalty equals.

$$(\delta + \epsilon)^2 \left[\frac{(N-k)^2}{N(N-1)} - \frac{(k-1)(N-k)}{(N-1)(N-2)} \right]$$

Expanding we get

$$(\delta + \epsilon)^2 \left[\frac{(N-2)(N-k)^2 - (k-1)(N-k)N}{N(N-1)(N-2)} \right]$$

which reduces to

$$(\delta + \epsilon)^2 (N-k) \left[\frac{(N-1)(N-2k)}{N(N-1)(N-2)} \right]$$

which further reduces to

$$(\delta + \epsilon)^2 \left[\frac{(N-k)(N-2k)}{N(N-2)} \right]$$

which is positive since $(N-2k) > 0$. And, given our restriction on δ this can be further reduced to

$$\epsilon^2 \frac{N(N-2k)}{(N-2)(N-k)}$$

A similar calculation for the taxes paid by those who send the higher messages gives

that it equals

$$(\delta + \epsilon)^2 \left[\frac{k^2}{N(N-1)} - \frac{k(N-k-1)}{(N-1)(N-2)} \right]$$

Expanding we get

$$(\delta + \epsilon)^2 \left[\frac{(N-2)k^2 - Nk(N-k-1)}{N(N-1)(N-2)} \right]$$

which reduces to

$$(\delta + \epsilon)^2 k \left[\frac{(N-1)(2k-N)}{N(N-1)(N-2)} \right]$$

which further reduces to

$$(\delta + \epsilon)^2 \left[\frac{k(2k-N)}{N(N-2)} \right]$$

And, given our restriction on δ this can be reduced to

$$-\epsilon^2 \frac{N(N-2k)k}{(N-2)(N-k)^2}$$

For the first order conditions to hold, $y^* = ky^l + (N-k)y^u$ satisfies the following two equations

$$w - \frac{1}{N} - T^l - \frac{y^*}{N} - y^* \gamma \frac{N-1}{N} (y^l - \bar{y}^l) + B'(y) = 0$$

and

$$w - \frac{1}{N} - T^u - \frac{y^*}{N} - y^* \gamma \frac{N-1}{N} (y^u - \bar{y}^u) + B'(y) = 0$$

However because y^* is also the symmetric equilibrium we have that

$$w - \frac{1}{N} - \frac{y^*}{N} + B'(y) = 0$$

Therefore, in any asymmetric equilibrium it must be that

$$-T^l - y^* \gamma \frac{N-1}{N} (y^l - \bar{y}^l) = 0$$

and

$$-T^u - y^* \gamma \frac{N-1}{N} (y^u - \bar{y}^u) = 0$$

Substituting in the expanded expression for T^l the first equation becomes

$$-\frac{\gamma}{2}\epsilon^2 \frac{N(N-2k)}{(N-2)(N-k)} - \frac{y^*}{N} = \frac{y^*\gamma\epsilon(N-1)N}{N(N-1)}$$

which can be rewritten

$$\epsilon \frac{N(N-2k)}{2(N-2)(N-k)} = y^*$$

which can be reduced to the following expression for ϵ

$$\epsilon = \frac{2(N-2)(N-k)y^*}{N(N-2k)}$$

The second equation that must be satisfied for the asymmetric equilibrium is for the agents who contribute more to the public good. Substituting in the expanded expression for T^u , we get

$$\frac{\gamma}{2}\epsilon^2 \frac{N(2k-N)k}{(N-2)(N-k)^2} = \frac{y^*\gamma\epsilon k N(N-1)}{N(N-1)(N-k)}$$

Which can be reduced to

$$\epsilon = \frac{2(N-2)(N-k)y^*}{N(N-2k)}$$

This is the same value for ϵ as for the agents who contribute less. Therefore, these are all asymmetric equilibria.