Fordham University
Department of Economics
Discussion Paper Series

On the Existence and Stability of Inefficient Boundary Equilibria in the Groves Ledyard Mechanism

Scott E. Page
University of Michigan, Ann Arbor,
Departments of Political Science and Economics, and Centers for the Study of Complex Systems and Policy Studies
Troy Tassier
Fordham University, Department of Economics

Discussion Paper No: 2008-20
October 2008

Department of Economics
Fordham University
441 E Fordham Rd, Dealy Hall
Bronx, NY 10458
(718) 817-4048

On the Existence and Stability of Inefficient Boundary Equilibria in the Groves Ledyard Mechanism

Scott E. Page*
Troy Tassier†

June 15, 2007

Abstract

In this paper, we characterize all interior and boundary equilibria of the Groves-Ledyard mechanism for a large class of economies and demonstrate their stability or lack thereof. We prove that the mechanism implements large numbers of inefficient and stable boundary equilibria, one for each of the efficient, asymmetric, interior equilibria found by Bergstrom, Simon, and Titus (BST). We show that the symmetric equilibrium is stable, and that its stability rests on two unstable dynamics that combine to create stability. The boundary equilibria, but not the asymmetric interior equilibria, are also stable. We further show that both sets of asymmetric equilibria can fail to exist in the presence of a large punishment parameter or a constraint on messages.

1 Introduction

Designing mechanisms that overcome the free rider problem and result in the efficient allocation of public goods has long been seen as the canonical problem in mechanism design. In a seminal paper, Groves and Ledyard offered a solution to the free rider problem (Groves and Ledyard 1976). In their mechanism, creatively constructed incentives induce people to truthfully reveal their preferences, no resources are disposed of, and the outcomes are

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

†Please send correspondence to: Troy Tassier, Department of Economics, E-528 Dealy, Fordham University, Bronx, NY 10458. telephone: (718) 817-4793. e-mail: tassier@fordham.edu

Pareto efficient. In the language of mechanism design, they succeeded in constructing an incentive compatible, balanced, efficient mechanism. They were right to call it a solution to the free rider problem.

However, in this paper, we prove that for many values of their punishment parameter, the mechanism also implements large numbers of inefficient boundary equilibria. This partly mitigates the original claim that the Groves Ledyard mechanism is a solution to the free rider problem. The Groves Ledyard efficiency proof remains correct but it only applies to those equilibria in the interior of the message space. It does not consider the boundary. The boundary equilibria that we discuss in this paper went unnoticed because the incentive structure created what Jackson later characterized as unbounded strategies (Jackson 1992). Hidden deep within the Groves Ledyard mechanism resides a variant of the "name the largest integer game" that leads to an arms race in the space of messages but not to an interior equilibrium.

These boundary equilibria and their interior counterparts fail to exist for large values of the punishment parameter and if there are constraints on the set of possible messages. When punishment is severe, the agents have a strong incentive to conform to the messages of other agents and thus eliminate all of the asymmetric equilibria. And, by constraining messages, we can force agents to remain in the basin of attraction for the symmetric equilibrium. We return to these point in the discussion at the end of this paper.

The direct contribution of this paper is obvious. It moves us closer to a complete understanding of one of the most important mechanisms in the literature. We characterize a new set of inefficient equilibria of the mechanism and prove that these new equilibria added to the set of previously known equilibria comprise all equilibria of the mechanism. Further, we show the new inefficient equilibria are stable as is the symmetric equilibrium of Groves and Ledyard. The other asymmetric equilibria, the BST equilibria, are all unstable. These are all important results.

This paper also contributes to broader methodological questions about mechanism design. Increasingly, papers rely on direct mechanisms. The focus is on proving the best possible outcome for any mechanism. The existence of an incentive compatable, truth telling equilibrium does not rule out the possibility of non truth telling equilibria. If these other equilibria exist, and if they are inefficient, we must ask whether they are likely outcomes of the mechanism.

Most current papers do not yield results as strong as Groves and Ledyard's result that *all* equilibria are efficient. Most papers that use direct mechanisms show what the most effienct mechanism can implement, not that all equilibria of the direct mechanism are efficient. These mechanisms may therefore implement many more equilibria. This is not a new point, the space between what is the best possible outcome and what are the likely outcomes can be quite large, but it is a point worth reiterating in light of our results.

Though we are showing that the mechanism has an abundance of inefficient equilibria, we are not disproving Groves and Ledyard's main result. Their mathematics are correct and their result still holds, though it is misleading. Had Groves and Ledyard bounded the set of possible messages, they too would have found these equilibria. Ignoring the boundary does not help. It may get rid of the equilibria but it does not get rid of the incentive problems they create. Best responses would lead some agents to send higher messages and other agents to send lower and lower messages and no equilibrium would be attained. This is the problem identified by Jackson (1992)

If we do bound the space of messages, there remains the question of whether these inefficient boundary equilibria would ever be realized. The answer would appear to be yes. In the boundary equilibria, agents that play the high message receive higher utility than they receive at the symmetric equilibrium. The high message agents do better than truth telling if they can lead the dynamics of the system toward the boundary equilibrium. Thus, if some agents randomly start out over contributing, they may create incentives for themselves to continue to do so.

Before describing our complete results, we provide some background. The Groves Ledyard mechanism's strength rests on an ingenious penalty function that induces agents to contribute to the public good by punishing them if they deviate from the average contribution levels of the other agents in the economy. At the same time, it rewards agents if they deviate from this average by less than others do, all while maintaining balance. All *interior* equilibria of this mechanism provide efficient levels of the public good. And, for quasi linear preferences, the mechanism has a unique equilibrium, but for more general preferences that allow the benefits from the private goods to depend on the amount of the public good provided, the mechanism implements large numbers of equilibria. We owe this latter result to Bergstrom, Simon, and Titus (1983) (hereafter BST). BST prove that an asymmetric equilibrium exists for each subset of agents containing more than one-half of the agents. In these BST equilibria, which are also efficient, some agents over-contribute and some agents under-contribute relative to their true preferences.[1] With N agents and a single public good, the mechanism implements approximately $\frac{N!}{2}$ equilibria so that even modest numbers of agents result in billions of equilibria.

Two properties of these BST equilibria imply that even more equilibria must exist. The similarity of structure across the BST equilibria indicates that these equilibria should either all be stable or all be unstable. If these equilibria plus the single symmetric equilibria constitute all of the equilibria for the mechanism, then degree theory would be violated. Further, for some values of N, the total number of equilibria found jointly by Groves and Ledyard and by Bergstrom, Simon and Titus is even. This also cannot be. There must be missing equilibria, and in fact there should be at least one for each of the BST equilibria. These missing equilibria are the inefficient boundary equilibria that we describe in this paper. Each inefficient boundary equilibrium can be paired with a unique BST equilibrium. Adding in the symmetric equilibrium this results in an odd number of equilibria. We prove in this paper that these are all of the pure strategy equilibria of the mechanism.

The existence of all of these equilibria raises an important question that was not salient at the time the mechanism was developed: namely the stability of the various equilibria. Since each boundary equilibrium can be paired with a unique BST equilibrium, we would expect one of that pair to be stable and the other to be unstable. That proves to be true. If the inefficient boundary equilibria are unstable, then they matter little and we might say that the Groves Ledyard solution holds except for unstable equilibria. Unfortu-

[1] As an example, let there be three agents A, B, and C. There are three asymmetric BST equilibria: one where A and B over contribute and C under contributes; one where A and C over contribute and B under contributes; and one where B and C over contribute and A under contributes.

nately, as we show in the penultimate section of this paper, the opposite is true. The many efficient asymmetric equilbria are unstable and the inefficient equilibria are stable. By the logic of degree theory, the symmetric equilibrium is also stable.

Our analysis of the stability of the symmetric equilibrium reveals two interesting findings: First, two unstable dynamics combine to form a stable dynamic. These instabilities have intuitive interpretations. One captures the private incentives to shirk. The other captures the mechanism induced incentives to mimic others. The second finding relates to the stability properties being independent of the punishment parameter. When the punishment parameter is large the symmetric equilibrium is unique and therefore stable. As the punishment parameter falls, new asymmetric equilibria are induced, but the symmetric equilibrium remains stable. This is rare. More typically, when asymmetric equilibria arise, they are stable and the symmetric equilibrium becomes unstable (see for example, Brock and Durlauf 2001).

When we say that an equilibrium is stable or unstable, we mean in the traditional sense that relies on linear stability analysis. Stability can only be defined relative to a specific dynamic. In economics, the foundations for those dynamics are often individual level learning rules like best response functions (Van Huyck,Cook, Battalio 1994). Linear stability analysis assumes a mild form of linear best responses on the part of the players. An equilibrium that is stable according to linear stability analysis need not be attained by a given learning rule or by human or artifical agents. This possibility is evident in the literature on the attainability of equilibrium in the Groves Ledyard mechanism, which focuses entirely on the case of quasi linear preferences, a case in which the equilibrium is unique and stable. In an early paper, Muench and Walker (1983) show that best response dynamics need not converge. Experimental results (Chen and Tang 1998) are more nuanced. Convergence to the equilibrium often proves difficult and depends on the size of the punishment parameter. Experiments with artificial agents show that the overshooting anticipated by Muench and Walker can be controlled (Arifovic and Ledyard 2004) provided that the agents moderate their responses.

Given that these results relate to the domain of quasi linear preferences it may appear that these papers have limited applicability to the analysis we undertake here due to the more general preferences that we consider. However note that these papers speak to the difficulty for agents to coordinate on the truth-telling symmetric equilibrium even when there is a unique equilibrium. In our setting where there are many equilibria (and approximately one-half of the equilibria are inefficient) the coordination problem is even more difficult. This directs us to strongly consider results that focus on the attainment of symmetric equilibria without considering the possibility of asymmetric and boundary phenomena. The present paper may be seen as an example in this spirit. BST first hinted at these difficulties for the Groves Ledyard mechanism when they found the first set of asymmetric equilibria of the mechanism and the present paper may be seen as a direct extension of their work.

We also want to make clear that the experimental literature shows that linear stability analysis only tells part of the story. Mathematical stability and attaining or learning to play equilibria are different concepts. To that end, we have a companion paper in which we constructed an agent based model of the Groves Ledyard mechanism (Page and Tassier 2004).[2] Relying on a specific utility function of the more general BST form, we found that artificial agents never located the BST equilibria and only found the symmetric equilibrium in special cases where we tuned the learning rates and initial conditions with great care. These findings echo Arifovic and Ledyard (2004).

The remainder of this paper is organized in five sections. In Section 2 we describe the Groves Ledyard mechanism and solve for the symmetric equilibrium and the asymmetric BST equilibria in closed form. To our knowledge, they have never before been derived. Even though Bergstrom, Simon, and Titus proved that these additional equilibria exist, they were not able to solve for them directly. We then prove that the BST equilibria fail to exist if the punishment parameter in the Groves Ledyard mechanism is above a threshold. In Section

[2]That paper, which was written at the same time as this one, includes solutions for the the boundary and BST equilibrium for a specific functional form so that the results of the agent based model can be compared against the mathematics. It contains none of the general proofs that appear in this paper. It is an agent based implementation of the results of this paper. It shows that the boundary equilibria are likely to be attained under standard learning dynamics. We might add that we owe much of the credit for finding the boundary equilibria that we characterize in this paper to the artificial agents used in that model. They continually located those equilibria calling our attention to the boundary.

3 we prove that the Grove Ledyard and the BST equilibria are the only *interior* equilibria for the mechanism and offer two proofs of efficiency for these equilibria. In Section 4, we formally describe the boundary equilibria and show that they are stable. This characterizes all of the equilibria of the Groves Ledyard mechanism in closed form. In Section 5 we address the stability of the aforementioned interior equilibria using linear stability analysis. We prove that the BST equilibria are all unstable and that the Groves Ledyard and the boundary equilibria are all stable. In the final section, we discuss how a budget constraint can also wipe out the BST equilibria and the boundary equilibria as do a large punishment parameter and a restriction to quasi linear preferences as discussed earlier in the paper.

2 The Groves Ledyard Mechanism

We consider an economy with a single public good and a single private good. We index the agents by $i \in \{1, ...N\}$. Each agent has an initial wealth w_i and consumes a private good x_i and a public good y. In the Groves-Ledyard Mechanism, each agent announces a message, m_i, which is $1/N$th the amount of the public good that the agent would like to have produced. Agents pay a tax $C_i(m)$ which is a function of the vector of messages. The Groves-Ledyard mechanism sets

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

Hereafter to simplify notation, we suppress y and use m to denote the sum of the m_i's.

$$C_i(m) = \alpha_i m + \frac{\gamma}{2}\left[\frac{N-1}{N}(m_i - \bar{m}^i)^2 - \sigma_i\right]$$

The punishment parameter, γ, plays an important role in the analysis. Increasing γ creates an incentive for conformity. For large enough γ, all agents send the same message in equilibrium.

Utility is a function of both goods $U_i(x_i, m)$. An agent's budget constraint requires that $x_i(m) = w_i - C_i(m)$. An equilibrium satisfies

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

To minimize notation, we restrict our attention to the case of agents with identical wealth levels and preferences and assume that $\alpha_i = 1/N$ throughout. Even with identical agents the mechanism generates multiple equilibria, therefore we see no reason to make the analysis more complicated than necessary. We take care to mention when this restriction matters substantively and when it just reduces notation. Our results do not appear to depend on the assumption of identical preferences and in some cases we show how the proofs can be extended to the case of non-identical agents.

Throughout the paper, we rely on a decomposition of $C_i(m)$ into two parts: the *contribution*, $m\alpha_i = m/N$, and the *punishment*, $T_i(\vec{m}) = \frac{\gamma}{2}\left[\frac{N-1}{N}(m_i - \bar{m}^i)^2 - \sigma_i\right]$ where $\bar{m}^i$ equals the average message sent by the $N-1$ agents other than i and σ_i equals $\frac{1}{N-2}$ times the sum of the $(N-1)$ terms of the form $(m_j - \bar{m}^i)^2$. Notice that σ_i does not depend on m_i, so it is out of the agent's control. Thus, for incentive purposes, the contribution m/N and the difference from the average are what is relevant.

2.1 Quasi-Linear Preferences

We begin with the case of quasi linear preferences (Chen and Tang 1998). Here, the Groves Ledyard Mechanism has a unique equilibrium. We can write the utility function of the ith agent as $U_i(x_i, m) = x_i + B(m)$ where $B(m)$ is a concave increasing function of m, the sum of the contributions to the public good. If we add the Groves-Ledyard tax into the budget constraint, we obtain the following equality.

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

Substituting this back into the utility function gives:

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

The first order condition on m_i is as follows:

$$-\frac{1}{N} - \gamma\frac{N-1}{N}(m_i - \bar{m}^i) + B'(m) = 0$$

This system of N equations has a unique equilibrium. Suppose that the agents choose different m_i's. Some of the messages will be above the average message of the others and some will be below the average of the other messages. Thus, $(m_i - \bar{m}^i)$ will be positive for some i and negative for other i. But this cannot hold in equilibrium since $(m_i - \bar{m}^i) = B'(m) - \frac{1}{N}$ for all i. Since the m_i's must all be the same it follows that $m_i = \frac{m}{N}$ for all i. Thus each equation reduces to $\frac{1}{N} = B'(m)$ which has a unique solution at m^* because of the concavity of B. Thus, the equilibrium is unique.

2.2 More General Preferences

We now consider the more general preferences considered by BST in which the Groves Ledyard (hereafter GL) mechanism has many equilibria.[3] These utility functions allow for the amount of the public good to influence the value of the private good. As an example, the value of beach front property (a private good) depends upon the cleanliness of the ocean and the air quality (public goods). This interaction is captured by the function $A(m)$, which is assumed to be concave, continuously differentiable, strictly positive, and strictly increasing. We can thus write the utility of the ith agent as

$$U_i(x_i, \vec{m}) = A(m)x_i + B_i(m)$$

The budget constraint requires that

[3] BST rely on a change of basis to prove their result. Though this simplifies the mathematics, it makes maintaining an intuitive feel for what is happening all but impossible.

$$x_i = w_i - m/N - T_i(\vec{m})$$

If we substitute the first of these two equations into the second we obtain

$$U_i(x_i, y) = A(m)w_i - A(m)m/N - A(m)T_i(\vec{m}) + B_i(m)$$

The first order condition with respect to m_i can be written as follows:

$$A'(m)w_i - A'(m)m/N - A'(m)T_i(\vec{m}) - A(m)/N - A(m)\gamma\frac{N-1}{N}(m_i^* - \bar{m*}^i) + B_i'(m) = 0$$

Assuming strictly concave preferences (this puts restrictions on $A(m)$), then there is a unique allocation which maximizes the sum of the agents' utility functions. Call this (x^*, y^*) and let $A^* = A(m^*)$ and $B^* = B(m^*)$.

2.3 Solving for the BST Equilibria

We now explicitly solve for the equilibria found by BST. In the BST equilibria and in the symmetric equilibrium, the amount of the public good provided is efficient. In the BST equilibria, a minority of the agents sends the same low message and a majority sends the same high message. Let's consider the example where we have k people deviating below by ϵ and $(N-k)$ deviating above by δ. Given that the BST equilibrium is efficient, the sum of the deviations from the symmetric equilibrium add to zero; Thus $\delta = \frac{k\epsilon}{(N-k)}$.

Solving for the equilibria will be accomplished in three steps: First, we state the first order condition for the agents who deviate above the average and the first order condition for the agents who deviate below the average. Second, we solve for the values contained in the Groves Ledyard tax mechanism in terms of ϵ and δ. Third, we substitute these values back into the FOC and solve for ϵ. This procedure gives us the equilibrium deviations from the symmetric equilibrium.

Recall the first order condition for an equilibrium:

$$A'(m)w_i - A'(m)m/N - A'(m)T_i(\vec{m}) - A(m)/N - \gamma\frac{N-1}{N}(m_i^* - \bar{m}_i^*) + B_i'(m) = 0$$

Since the total contribution to the public good is the same in the symmetric equilibrium and the BST equilibria, $A(m)$ does not change. The agents who belong to the minority subset of size k who send the low message have a larger T_i because they differ from the mean by more than the other agents. We denote the punishment paid by the low message agents with $T^l(m)$ and the punishment paid by the high message agents with $T^u(m)$. Assuming all agents have the same wealth, we can write the first order conditions for the two types of agents as

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

for the low message agents and

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

for the high message agents. We can then solve for the elements that make up $T^l(m)$ and $T^u(m)$ where

$T^l(m) = \frac{\gamma}{2}\left[\frac{N-1}{N}(m^l - \bar{m}^{*l})^2 - \sigma^l\right]$

and

$T^u(m) = \frac{\gamma}{2}\left[\frac{N-1}{N}(m^u - \bar{m}^{*u})^2 - \sigma^u\right]$

In the appendix we prove the following claim

Claim 1 *In an efficient allocation of the public good, where* $\delta = \frac{k\epsilon}{(N-k)}$, $T^l(m)$ *and* $T^u(m)$ *are given by the following equations.*

$$T^l(m) = \frac{\gamma}{2}\epsilon^2\frac{N(N-2k)}{(N-2)(N-k)}$$

$$T^u(m) = -\frac{\gamma}{2}\epsilon^2 \frac{N(N-2k)k}{(N-2)(N-k)^2}$$

Notice that the sum of the taxes paid by the k agents sending the lower messages and the $(N-k)$ agents sending the higher messages equals zero, which is an artifact of the mechanism being balanced. Using these values we can then solve for the value of ϵ that characterizes the BST equilibrium.

Claim 2 *The BST equilibria are characterized by k agents sending the message $m^* - \epsilon^{BST}(k)$ and $(N-k)$ agents sending the message $m^* + \frac{k\epsilon^{BST}(k)}{(N-k)}$, where $k < \frac{N}{2}$ and where*

$$\epsilon^{BST}(k) = \frac{2(N-2)(N-k)A(m)}{A'(m)N(N-2k)}$$

pf: see appendix.

This characterizes the BST asymmetric equilibria.[4] Note that the punishment parameter γ does not determine ϵ, ie, the location of these equilibria. ϵ is only a function of N, k, and $A(m)$. Yet, γ does effect whether or not these equilibria exist as we discuss next.

2.4 Structure of the BST Equilibria

We now take a closer look at some characteristics of the BST equilibria. We first discuss the role of the punishment parameter γ. Although it does not impact the location of the BST equilibria we will show that the value of γ determines whether or not these equilibria exist. To show the effect of γ we decompose the utility function into two parts: the public good portion of utility (PGU) which does not depend on γ and the punishment portion of utility (PUNU) which does depend on γ. These can be written as:

$$PGU = A(m^*)(w_i - \frac{m^*}{N}) + B(m^*)$$

and

[4] Notice though that for k larger than $\frac{N}{2}$ we getting the same equilibria that we obtain for k less than $\frac{N}{2}$.

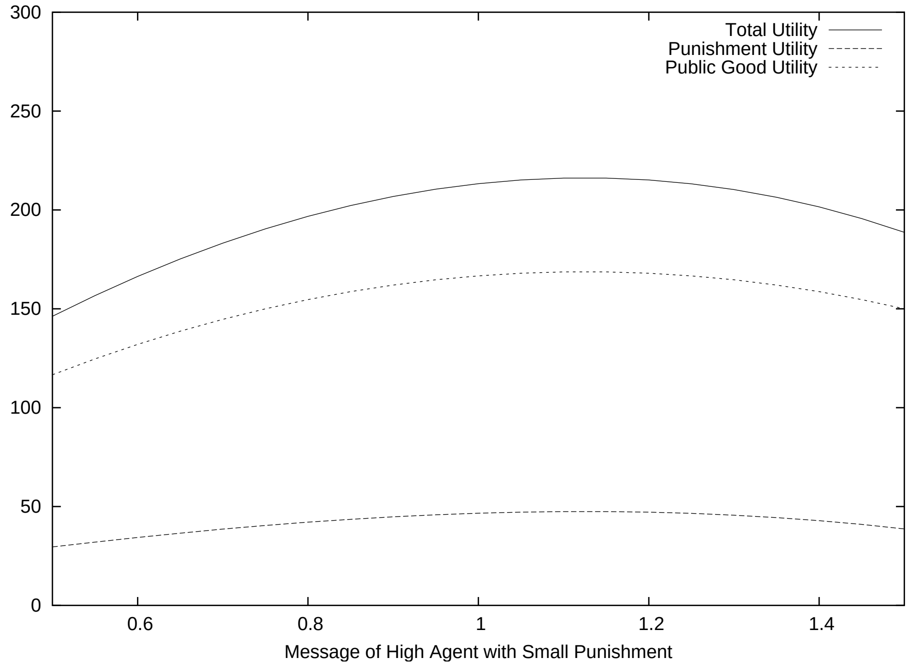

Figure 1: Utility of the high message agents when γ is small.

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

Figures 1 and 2 show this decomposition of utility for the high and low agents for an example utility function with 3 agents at a BST equilibrium (2 agents are playing the high message and one is playing the low message.) Note first that both the public good portion of utility and the punishment portion of utility are at a maximum at the same message for the high agents. Thus total utility is also a maximum at this message (the BST message.) The decomposition of utility for the low agent is much different. For her the public good portion is at a maximum but the punishment portion is at a minimum at the message that gives maximum total utility (the BST message.)

We can see this formally through the derivatives of the punishment portion of utility for the low agent. The first derivative is:

$$-A'(m^*)T^l - A(m^*)\gamma\frac{N-1}{N}(m^l - \bar{m}^l)$$

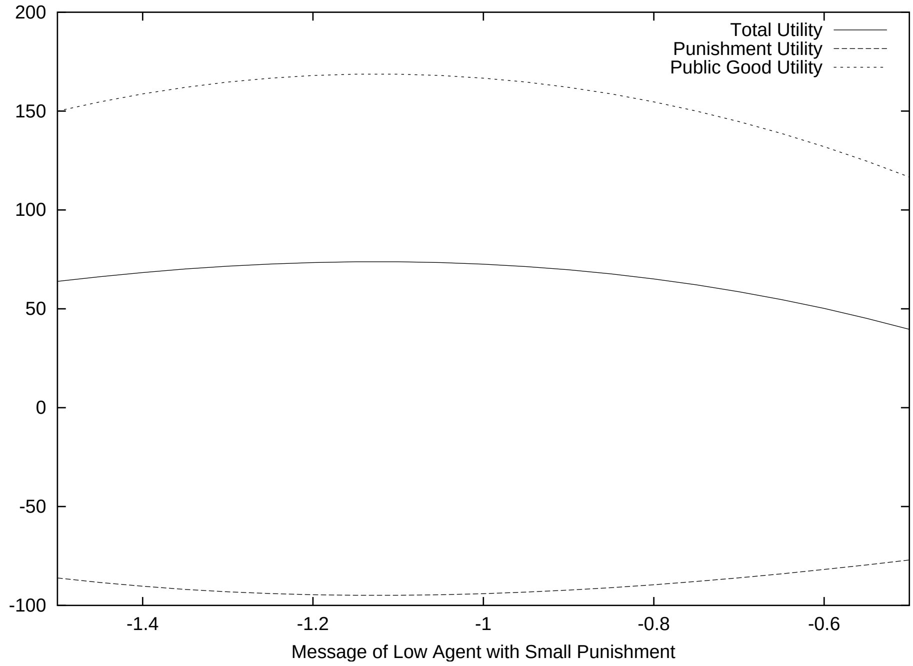

Figure 2: Utility of the low message agent when γ is small.

And the second derivative is:

$$-A''(m^*)T^l - 2A'(m^*)\gamma\frac{N-1}{N}(m^l - \bar{m}^l) - A(m^*)\gamma\frac{N-1}{N}$$

Recall that $m^l - \bar{m}^l = -\frac{\epsilon N}{N-1}$ and that $\epsilon = \frac{2(N-2)(N-k)A(m)}{N(N-2k)A'(m)}$. Substituting these into the 2nd derivative and rearranging yields:

$$-A''(m^*)T^l + \frac{\gamma A(m^*)}{N}[4\frac{(N-2)(N-k)}{(N-2k)} - (N-1)]$$

Note that the public good portion of utility does not depend on γ. Thus as γ increases the PGU remains the same. But, for the low agent, the punishment portion of utility is decreasing linearly in γ. Thus as γ increases the total utility at the BST equilibria is decreasing for the low agents in the punishment portion of utility. Thus for sufficiently large γ, utility is *minimized* at the BST equilibria contribution levels, so these are no longer equilibria. This agrees with our initial intuition. If γ is too large, the pressure to conform is sufficiently great to wipe out the asymmetric equilibria. This can be seen graphically. Using

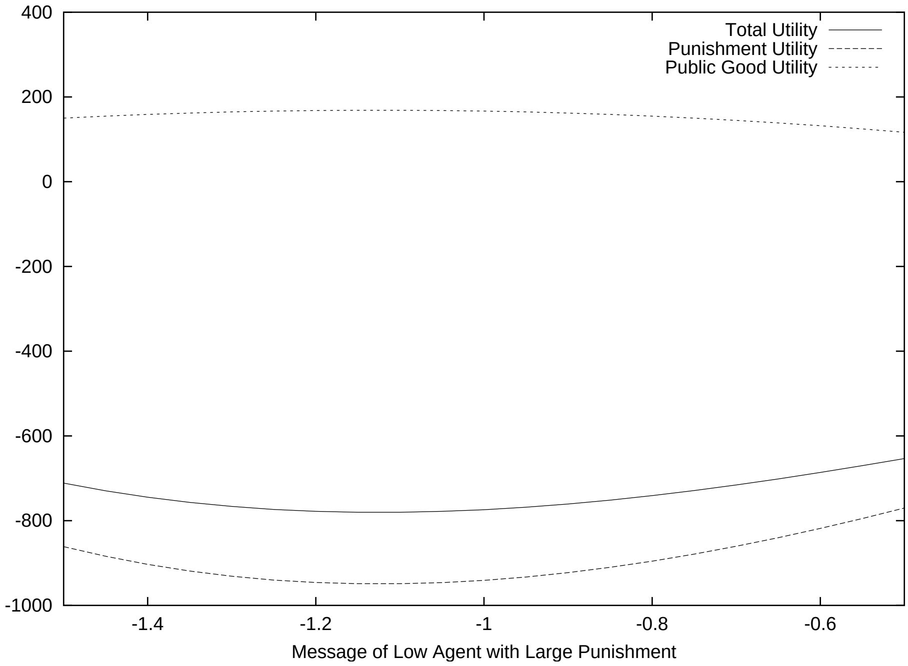

Figure 3: Utility of the low message agent when γ is high.

the same utility function as above we show the utility decomposition for the low agent with a large level of γ in Figure 3.

We can solve for the exact value of γ at which the BST equilibria cease to exist. Let PGU'' be the second derivative of the public good portion of utility. Note that this value is negative at the BST equilibria since the public good portion of utility is at a maximum. Then, using the second derivative of the punishment portion of utility above, the BST equilibria disappear if

$$\gamma > \frac{A''(m^*)T^l - PGU''}{\frac{A(m^*)}{N}[4\frac{(N-2)(N-k)}{(N-2k)} - (N-1)]}$$

This is a complicated expression but an important one as it places a lower bound on γ for the Groves Ledyard mechanism to yield a unique equilibrium. Given how complicated it is, comparative statics are messy. Yet, we can say that as the utility for the public good portion becomes more concave (PGU'' becomes more negative), this lower bound increases. This makes intuitive sense. The more curviture, the more likely the asymmetric equilibria are to exist.

3 Characterization of Interior Equilibria

We next provide an alternative proof that any interior equilibrium must be efficient. We also show that in any interior equilibrium all agents sending a high message (respectively a low message) must send the same message. Together, these claims imply that the symmetric equilibrium together with the asymmetric equilibria found by BST comprise all of the interior equilibria of the mechanism.

To prove these and later claims, we decompose the derivative of the utility function into two parts: the public good part and the punishment part. The public good part of marginal utility (PGMU) equals

$$PGMU(\vec{m}) = A'(m)(w - \frac{m}{N}) - \frac{A(m)}{N} + B'(m)$$

The punishment part of marginal utility (PUNMU) equals

$$PUNMU(\vec{m}) = -A'(m)T_i(\vec{m}) - A(m)\gamma\frac{N-1}{N}(m_i - \bar{m}^i)$$

We first state a claim that any interior equilibrium of the Groves Ledyard Mechanism is efficient, even when agents are not identical. Groves and Ledyard prove a similar claim in their paper. Here we allow the wealth levels to change as well as preferences for the public good which we denote by B_i.

Claim 3 *All interior equilibria of the Groves Ledyard mechanism provide for an efficient level of the public good.*

pf. The amount of the public good is efficient if the sum of the PGMU's for all of the agents equals zero. At an interior equilibrium the sum of each agent's PGMU and PUNMU is zero; therefore, it suffices to show that the sum of the PUNMU's equals zero at any equilibrium. We can write the sum of the PUNMU's as follows

$$\sum_{i=1}^{N}(-A'(m)T_i(\vec{m}) - \gamma\frac{N-1}{N}(m_i - \bar{m}^i))$$

The sum of the $(m_i - \bar{m}^i)$ terms is trivially zero. Each m_i is added once and each is subtracted $(N-1)$ as part of an average of $(N-1)$ messages. It therefore, suffices to show that the sum of the $T_i(\vec{m})$'s equal zero. Groves and Ledyard show this in their original paper. It is a condition of the mechanism being balanced.

The next claim states that at any asymmetric interior equilibrium all agents sending a message higher than average send the same message.

Claim 4 *Let $(m_1, m_2, \ldots m_N)$ be an interior equilibrium of the Groves Ledyard Mechanism with identical agents. If $m_i \geq \bar{m}^i$ and $m_j \geq \bar{m}^j$, then $m_i = m_j$.*

pf. see appendix

This claim appears to depend heavily on the identical agent assumption, but in fact it does not. Without identical agents it would not be the case that all of the agents sending the higher messages choose the exact same messages. However, it would be the case that any asymmetric equilibrium messages higher than the GL equilibrium messages would generate the same PUNMU as the messages under the Groves-Ledyard equilibrium since the amount of the public good provided would be the same in each case. Therefore, all agents sending messages above their GL equilibrium message would be obtaining the same marginal punishment for doing so.

Similarly at any asymmetric interior equilibrium with identical agents all agents sending a lower than average message send the same message.

Claim 5 *Let $(m_1, m_2, \ldots m_N)$ be an interior equilibrium of the Groves Ledyard Mechanism. If $m_i \leq \bar{m}^i$ and $m_j \leq \bar{m}^j$, then $m_i = m_j$.*

pf This proof follows directly from the previous claim.

Using these insights it is possible to construct a simpler direct proof for the efficiency of all interior equilibria in the case of identical agents.

Corollary 1 *All interior equilibria of the Groves Ledyard mechanism provide for an efficient level of the public good when agents are identical.*

pf. Suppose that we have an equilibrium where the amount of the public good is not equal to m^*. It follows that the marginal utility of the public good part of the first order condition cannot equal zero, $PGMU(m^*) \neq 0$. From the previous claims we know that for any interior equilibrium, there must be a set of agents sending a low message and a set agents sending a high message. The punishment portion of the first order condition for each set of agents plus the public good portion must equal zero. From our characterization of the asymmetric equilibria we know that for the agents who send the low messages the punishment portion of the first order condition is given by

$$-A'(m)\frac{\gamma}{2}\epsilon^2\frac{N(N-2k)}{(N-2)(N-k)} - \frac{A(m)\gamma\epsilon(N-1)N}{N(N-1)}$$

and for the agents who send the high message the punishment portion of the first order condition is given by

$$A'(m)\frac{\gamma}{2}\epsilon^2\frac{N(2k-N)k}{(N-2)(N-k)^2} + \frac{A(m)\gamma\epsilon kN(N-1)}{N(N-1)(N-k)}$$

These two values have opposite sign. The first is less than 0 since $k < N/2$ and $A'(m)$ is positive. For the second value the left hand term is negative since $2k - N < 0$ but the right hand term is positive. Thus we need to show that:

$$\frac{A(m)\gamma\epsilon kN(N-1)}{N(N-1)(N-k)} > A'(m)\frac{\gamma}{2}\epsilon^2\frac{N(2k-N)k}{(N-2)(N-k)^2}$$

Recall that $\epsilon = \frac{2(N-2)(N-k)A(m)}{A'(m)N(N-2k)}$. If we cancel terms and substitute for ϵ we find this equation reduces to: $\frac{(2k-N)}{(N-2k)} < 1$ which is true since $k < N/2$. Since they must be equal, this can only occur if they equal zero, which implies that $PGMU(m^*) = 0$. This completes the proof.

Taken together these three claims imply that the only interior equilibria of the Groves Ledyard Mechanism are the symmetric equilibrium and the BST equilibria. Claim 1 restates the Groves Ledyard result that all interior equilibria of the mechanism are efficient. Claims 2 and 3 show that any asymmetric equilibria of the mechanism must have the form described by BST. Thus the BST equilibrium comprise the only set of asymmetric interior equilibria. Further all of these equilibria are efficient. We next consider the possibility of there being other non-interior equilibria if we bound the space of messages.

4 Boundary Phenomena and Equilibria

We now characterize the boundary equilibria of the Groves Ledyard Mechanism. We do this by placing a lower bound on the set of possible messages. Without such a lower bound on messages, there are no other equilibria because the incentive structure creates an unbounded mechanism (Jackson 1992). To see this, suppose that as in the BST equilibria, there are two groups of agents, some sending an identical high message and some sending an identical low message with more of the agents sending the high message. If the high message is above the BST equilibrium value and the amount of the public good is efficient, then the agents sending the high message have an incentive to increase their messages because in doing so they recieve a subsidy from the punishment payments. This creates an incentive for the agents sending the lower message to make their messages even smaller, which in turn causes the high message agents to raise their messages further. No equilibrium will ever be attained as the agents race toward positive and negative infinity.

If we bound the space of messages, then we get equilibria at the boundary as it stops the infinite sequence. Here, we consider the case where messages are restricted to the interval $[-D, \infty)$. This prevents infinite negative amounts of the public good. With this construction it is possible that agents sending higher messages may still want to send infinite positive messages, but they will choose not to do so. For small γ such equilibria always exist provided D is large enough. However, for larger γ or for small D they need not exist.

We can no longer assume that the amount of the public good provided is efficient, but from our previous results, we can assume that k agents, where $k < \frac{N}{2}$, send the message $-D$, and the other $N-k$ agents send messages m^+ that are positive. Let m denote the amount of the public good provided so that m satisfies $-kD + (N-k)m^+ = m$ or $m^+ = \frac{m+kD}{(N-k)}$. Following the method used to solve for the BST equilibria above, we first calculate the difference between the low message $-D$ and the average of the others, $\bar{m}^D$. A straightforward calculations gives that $\bar{m}^D = \frac{m+D}{(N-1)}$. It follows that

$$(-D - \bar{m}^D) = -\frac{m + ND}{(N-1)}$$

For the agents who send higher than average messages the corresponding values are

$$\bar{m}^+ = \frac{(N-k-1)m - kD}{(N-k)(N-1)}$$

$$(m^+ - \bar{m}^+) = \frac{km + NkD}{(N-k)(N-1)}$$

We next calculate the σ^D term within T^D. For the agents at the lower boundary

$$\sigma^D = \frac{1}{N-2}\left[(k-1)\frac{(m+ND)^2}{(N-1)^2} + (N-k)\frac{(ky+NkD)^2}{(N-k)^2(N-1)^2}\right]$$

Which reduces to

$$\sigma^D = \frac{1}{N-2}\frac{(m+ND)^2(N(k-1)+k)}{(N-1)^2(N-k)}$$

We next calculate the σ^+ term within T^+. For the agents choosing the high message

$$\sigma^+ = \frac{1}{N-2}\left[k\frac{(m+ND)^2}{(N-1)^2} + (N-k-1)\frac{(ky+NkD)^2}{(N-k)^2(N-1)^2}\right]$$

Which reduces to

$$\sigma^+ = \frac{1}{N-2}\frac{(m+ND)^2k(N(N-k)-k)}{(N-1)^2(N-k)^2}$$

We can then calculate the punishment for both types of agents. For the agents who send the message at the lower boundary it equals

$$T^D(m) = \frac{\gamma}{2}\frac{(m+ND)^2[(N-1)^2(N-2k)+(2k+1)N]}{N(N-2)(N-k)(N-1)^2}$$

Notice that this is a positive number. For the agents who send the high messages, we obtain

$$T^+(m) = \frac{\gamma}{2}\frac{(m+ND)^2[-N^3+2k(N^2-N+1)]}{N(N-2)(N-k)^2(N-1)^2}$$

Notice that this amount is negative. Therefore, the punishment serves as a subsidy from those sending the low message to those sending the high message. Given that the total amount of the public good is not necessarily efficient, we have that

$$A'(m)(w-m/N) - \frac{A(m)}{N} + B'(m) = X$$

Therefore, the first order conditions for the agents who send the message at the lower boundary is

$$FOC^D = -\frac{A'(m)}{N} - A'(m)T^D(\vec{m}) - A(m)\gamma\frac{N-1}{N}(m^D - \bar{m}^{*D}) + X = 0$$

This can be written as

$$FOC^D = -\frac{A'(m)}{N} - A'(y)T^D(\vec{m}) + A(m)\gamma\frac{(m+ND)}{N} + X = 0$$

For the agents who send the high message we have

$$FOC^+ = -\frac{A'(m)}{N} - A'(m)T^+(\vec{m}) - A(m)\gamma\frac{N-1}{N}(m^+ - \bar{m}^{*+}) = 0$$

Which equals

$$FOC^+ = -\frac{A'(m)}{N} - A'(m)T^+(\vec{m}) - A(m)\gamma\frac{km+NkD}{N(N-k)} = 0$$

In order for a boundary equilibria to exist, three conditions must hold: First, MU for the low message agents must be negative. This implies that the low message agents want to decrease their message more but the boundary prevents them from doing so. Second, for low message agents the payoff at the boundary must be higher than at any interior point. The boundary is a best response for the low message agents. Third, the high message agents must not want to send infinite messages. The message of the high agents must stop at some value. The second and third conditions are trivially satisfied since for low γ the utility function is concave and marginal utility is decreasing in m. From above, the first condition requires

$$-\frac{A'(m)}{N} - A'(m)T^D(m) + A(m)\gamma\frac{(m + ND)}{N} < 0$$

For sufficiently small γ the second and third terms are smaller in absolute terms than the first which is negative. Thus for small γ the low agents may have negative marginal utility at the boundary indicating that they want to further decrease their message. Combining this with the analysis of the BST equilibria above, we see that for sufficiently large γ only the symmetric equilibrium exists. Thus, one can always increase the punishment parameter of the mechanism to enforce an equal sharing of the cost of public good provision.

These boundary equilibria can be shown to be stable. This can be seen by looking at the first order conditions for the low message agents. They are negative. This means that they would like to further decrease their message below the boundary level and if pushed from this boundary they run right back to it. Once there, they create incentives for the high message agents to coordinate on their equilibrium value as well. We omit the formal proof as it is similar to the proofs for the interior equilibria which we cover next.

5 Stability of the Interior Equilibria

We analyze the stability of the interior equilibria using linear stability analysis. With respect to the caveats that we mentioned in the introduction, linear stability analysis remains a powerful tool. Linear stability analysis assumes that if an agent's marginal utility is positive in m_i at a point, then the agent increases its message and if the marginal utility is decreasing that the agent will decrease its message and that the magnitude of the change is proportional to the marginal utility.

$$\dot{m}_i = \frac{\partial U_i}{\partial m_i}$$

In our analysis we again find it useful to decompose the utility into the public good component and the punishment component of marginal utility.

$$PGMU(\vec{m}) = A'(m)(w - \frac{1}{N}) - \frac{A(m)}{N} + B'(m)$$

$$PUNMU(\vec{m}) = -A'(m)T_i(\vec{m}) - A(m)\gamma\frac{N-1}{N}(m_i - \bar{m}^i)$$

We can write the equation of motion as the sum of these two terms.

$$\dot{m}_i = PGMU_i(\vec{m}) + PUNMU_i(\vec{m})$$

As we shall show, peculiarities of the Groves-Ledyard mechanism make this decomposition extremely useful in understanding the dynamics of the system. We first show that the symmetric equilibrium is stable and then show that the BST equilibria are unstable given these equations of motion.

5.1 Stability of the Symmetric Equilibrium

The decomposition of the equation of motion into a public good portion and a punishment portion highlights a remarkable property of the symmetric equilibrium. The punishment contribution to utility creates an unstable dynamic as does the public good contribution to utility, but added together, they create a stable dynamic. The dynamics created by the public good portion of utility allow for drift in the vector of messages as long as the amount of the public good provided remains efficient. The punishment portion of utility creates an incentive for the agents to send the same message, whether or not it is efficient. Therefore, the punishment portion allows drifts from efficiency. The force toward efficiency created by the public good part overpowers the drift away allowed by the punishment part. And the force toward symmetry by the punishment part overpowers the drift away from symmetry allowed by the public good part. Thus, at the symmetric equilibrium, unstable incentives plus unstable incentives yield stable incentives.

To simplify the presentation of our results in this section we show matrices as though there are only three agents but we derive our results and use notation from the N player case, so that our analysis is general. If we compute the Jacobian for the public good contribution to utility we get the following form

$$\frac{\partial PGU_i}{\partial m_j} = \begin{vmatrix} -\theta & -\theta & -\theta \\ -\theta & -\theta & -\theta \\ -\theta & -\theta & -\theta \end{vmatrix}$$

where $-\theta$ equals $A''(m)(w - \frac{1}{N}) - \frac{A'(m)}{N} + B''(m)$. The reason that every entry in the Jacobian has the same value is that the marginal effect of an increase in any agent's message is the same for all players since the costs are split evenly. This value is negative at an efficient equilibrium. As we mentioned, considered in isolation, this is not a stable system. In the case of N agents, $N-1$ of the eigenvalues are 0 and the other has value $-N\theta$. The fact that the non negative eigenvalues are not strictly positive means that the system has other equilibria in the neighborhood of this point. In fact, any set of messages that sum to the efficient amount of the public good is a steady state with respect to the public good portion of utility. If one agent increases its message by ϵ and another decreases its message by ϵ then this new set of messages is a steady state for this portion of the equations of motion.

Next we consider the Jacobian associated with the punishment portion of utility. Recall that the punishment portion of marginal utility at the symmetric equilibrium are zero. It can be shown that the Jacobian is given by

$$\frac{\partial PUNU_i}{\partial m_j} = \begin{vmatrix} -A(m)\gamma\frac{N-1}{N} & A(m)\gamma\frac{1}{N} & A(m)\gamma\frac{1}{N} \\ A(m)\gamma\frac{1}{N} & -A(m)\gamma\frac{N-1}{N} & A(m)\gamma\frac{1}{N} \\ A(m)\gamma\frac{1}{N} & A(m)\gamma\frac{1}{N} & -A(m)\gamma\frac{N-1}{N} \end{vmatrix}$$

This matrix has the form

$$\begin{vmatrix} -(N-1)\omega & \omega & \omega \\ \omega & -(N-1)\omega & \omega \\ \omega & \omega & -(N-1)\omega \end{vmatrix}$$

where, $\omega = A(m)\gamma\frac{1}{N}$.

This matrix has $(N-1)$ eigenvalues equal to $-N\omega$ and one eigenvalue equal to 0. Therefore, this dynamic like the previous dynamic also is not stable. If all of the agents increase or decrease their messages by a common amount, the new messages are a steady state. Because this deviation has to be coordinated among all N agents, this dynamic has only one eigenvalue equal to zero. In the previous dynamic, almost any deviation will lead to a new equilibrium, that is why almost all of the eigenvalues are zero. When we add these two matrices together, we get a matrix that defines a stable dynamic. As mentioned above, the first dynamic forces efficiency; the second forces symmetry. The combined Jacobian at the symmetric equilibrium looks as follows:

$$\frac{\partial U_i}{\partial m_j} = \begin{vmatrix} -(N-1)\omega-\theta & \omega-\theta & \omega-\theta \\ \omega-\theta & -(N-1)\omega-\theta & \omega-\theta \\ \omega-\theta & \omega-\theta & -(N-1)\omega-\theta \end{vmatrix}$$

This matrix has $(N-1)$ eigenvalues of $-N\theta$ and one eigenvalue of $-N\omega$. Therefore, it is stable. The assumption of identical agents again plays only a minor role here. If we give each agent a unique $B_i(m)$, then each row of the public good matrix gets multiplied by a unique constant, $\beta_i > 0$. $N-1$ of the eigenvalues are still zero and the other eigenvalue equals

$-N\sum_{i=1}^{N}\beta_i\theta$ for the public good portion of utility. This nonzero eigenvalue replaces $-N\theta$ as an eigenvalue of the combined Jacobian and the other eigenvalues remain unchanged.

5.2 The Instability of the BST equilibria

We now perform the same analysis for the BST equilibria. Here, the calculations are more involved because for each subset of size $k < N/2$ that deviates from the symmetric equilibrium we get a distinct set of dynamics. Since this creates a potentially infinite set of systems, we consider the case where k equals one in full and show that it is unstable. We then sketch a proof for why the same logic applies for any k.

As in the symmetric case, we decompose the equations of motion into two parts: the public good part and the punishment part, and then combine them to form the Jacobian for the system of equations. The Jacobian for the public goods portion considered alone is the same as for the symmetric equilibrium and takes the form

$$\frac{\partial PGU_i}{\partial m_j} = \begin{vmatrix} -\theta & -\theta & -\theta \\ -\theta & -\theta & -\theta \\ -\theta & -\theta & -\theta \end{vmatrix}$$

This is not a stable system. In the case of N agents, $N-1$ of the eigenvalues are 0 and the other eigenvalue is $-N\theta$. Calculating the Jacobian for the punishment portion of the dynamics at the asymmetric equilibrium is cumbersome. Recall that $T^l(\vec{m})$ is the punishment paid by the agents who send the lower message. In this case, that is just one agent. The equation of motion for the message sent by that agent is given in the third row. With some effort it can be shown that the Jacobian takes the following form:

$$\frac{\partial PUNU_i}{\partial m_j} = \begin{vmatrix} \alpha-(N-1)\beta & \alpha+\beta & \alpha-\beta \\ \alpha+\beta & \alpha-(N-1)\beta & \alpha-\beta \\ -(N-1)\alpha+\beta & -(N-1)\alpha+\beta & -(N-1)\alpha+(N-1)\beta \end{vmatrix}$$

where $\alpha = A''(m)T^{(}\vec{m})/(N-1)$ and $\beta = \gamma A(m)$ It can be shown that his matrix has one eigenvalue equal to 0, one equal to β and $(N-2)$ equal to $-N\beta$. So the punishment portion of utility also is unstable. When we combine the Jacobian for the public goods contribution to the dynamics and the punishment portion of the dynamics we get a matrix of the following form

$$\frac{\partial U_i}{\partial m_j} = \begin{vmatrix} \alpha-(N-1)\beta-\theta & \alpha+\beta-\theta & \alpha-\beta-\theta \\ \alpha+\beta-\theta & \alpha-(N-1)\beta-\theta & \alpha-\beta-\theta \\ -(N-1)\alpha+\beta-\theta & -(N-1)\alpha+\beta-\theta & -(N-1)\alpha+(N-1)\beta-\theta \end{vmatrix}$$

This matrix has one eigenvalue equal to $-N\theta$, $(N-2)$ eigenvalues equal to $-N\beta$, and one eigenvalue equal to β. Therefore, the system is not stable. As before, symmetry of agents does not play a large a role in the dynamics. If each agent has a unique utility for the public good, a unique B_i, then as in the symmetric case, only the $-N\theta$ eigenvalue is affected and the change is only in magnitude not in sign.

This proves that the BST equilibria in which one agent sends a low message and the rest send high messages are unstable, but it does not prove the general case. Moreover, the calculation does not always provide any intuition behind why a system is stable or unstable. Though, in the symmetric case, we found that by decomposing the dynamical system into two parts, we could uncover the causes of stability.

The logic driving the instability of the BST equilibria relies on the decomposition of marginal utility into a public good contribution (PGMU) and the punishment contribution (PUNMU) to marginal utility. At $m-\epsilon$, PUNMU equals

$$-A'(m^*)T^l(m) - A(m^*)\gamma\frac{N-1}{N}(m^l - \bar{m}^l)$$

This can be rewritten as

$$\gamma[\epsilon A(m^*) - \epsilon^2 A'(m^*)\frac{N(N-2k)}{2(N-2)(N-k)}]$$

Similarly, PUNMU at $m + \frac{k\epsilon}{N-k}$ equals

$$-A'(m^*)T^u - A(m^*)\gamma\frac{N-1}{N}(m^u - \bar{m}^u)$$

This can be rewritten as

$$\gamma[-\epsilon A(m^*)\frac{k}{(N-k)} + \epsilon^2 A'(m^*)\frac{N(N-2k)k}{2(N-2)(N-k)^2}]$$

Suppose for a moment that $\epsilon = 0$, (we are at the symmetric equilibrium, m^*.) For $\epsilon = 0$ PUNMU at $m - \epsilon$ and PUNMU at $m + \frac{k\epsilon}{N-k}$ become the same equation. By assumption $A(m^*) > 0$ and $A'(m^*) < 0$; therefore, the punishment portion of the FONC for the agents sending the low message is positive for small values of ϵ. As ϵ increases the punishment portion of FONC eventually decreases to zero and thereafter is negative. The value for which this expression equals zero is the ϵ which gives the BST equilibria. Recall that this is denoted ϵ^{BST}. Similarly, the punishment portion of the FONC for the agents sending the higher message equals 0 at $\epsilon = 0$. For positive ϵ the punishment portion of the FONC is negative. For increases in ϵ the punishment portion of the FONC remains negative until ϵ^{BST} and thereafter it is strictly positive.

This observation confirms our intuition as to why the symmetric equilibrium is stable. If we perturb the stable equilibrium, then the agents sending the low message have a positive PUNMU, therefore, they announce higher messages and the agents sending the higher message have a negative PUNMU and therefore, send lower message, modulo the effect of the amount of the public good.

Near the BST equilibria if the agents are not quite far enough apart (if the low agents are sending too high a message and the high agents are sending too low a message) then the agents sending the lower message will have a positive PUNMU and will increase their message. The agents sending the high message will have a negative PUNMU and decrease their message, moving them away from the BST equilibrium and toward the symmetric equilibrium.

On the other hand, if we are near the BST equilibria but the agents are too far apart (the low message agents are sending too low a message and the high message agents are sending too high a message) then the low message agents will decrease their message further. The high message agents will have positive PUNMU and will increase their messages further. The messages of the low agents move toward $-\infty$ and the messages of the high agents move toward $+\infty$. With this intuition at hand we are ready to prove the instability of the BST equilibria.

To formally show the local instability of the BST equilibria, we must show that given a BST equilibrium m^{eq} and for any $\rho > 0$, that the neighborhood of radius ρ around m^{eq} contains a point, $\hat{m}$, such that beginning from $\hat{m}$, the dynamical system will not converge to m^{eq}. Choose an arbitrary BST equilibrium where k agents send the message $m^* - \epsilon^{BST}(k)$ and $N-k$ agents send the message $m^* + \delta^{BST}(k)$ such that the the allocation of the public good is efficient. Call this $m^{BST}(k)$. Given a $\rho > 0$ choose $\epsilon < \epsilon^{BST}(k)$ such that if k agents send the message $m^* - \epsilon$ and $N-k$ agents send the message $m^* - \frac{(N-k)\epsilon}{k}$, the messages lie in the neighborhood of radius ρ around $m^{BST}(k)$. The agents sending the lower message are sending a message greater than the equilibrium low message and the agents sending the higher message are sending a message lower than the equilibrium high message.

It suffices to show that the agents sending the lower message will increase their message and the agents sending the higher message will decrease their message. This will move the set of messages further from the BST equilibrium. These conditions are sufficient using the following logic: If the messages at each instance of time continue to provide for an efficient amount of the public good then the same argument applies: *at each moment in time, the messages will be even further from the BST equilibrium.* Alternatively, suppose the dynamics could lead to either over or under provision of the public good at some time t_1. Without loss of generality assume over provision of the public good. In this case the public good portion of the FONC will become negative. This will cause the agents sending the higher message to reduce their messages by even more. The agents sending the lower messages will have less incentive to increase their messages and may even have an incentive to decrease their messages if the over provision becomes too severe. Therefore, at some time $t_2 > t_1$ one of two things must happen. Either the low and high messages will converge (in which case the system will converge to the symmetric equilibrium) or the amount of the public good will again be efficient. If the latter occurs, since the agents sending the higher messages have been decreasing their message, the new messages are further from the BST equilibrium at time t_2 then they were at time t_1.

Thus, we need only show that the derivative of the utility function for the agents sending the lower message with respect to their message is positive and that the opposite is true of the derivative of the utility function for the agents sending the higher message. Since the amount of the pubic good is efficient, the value PGMU equals zero. It suffices then to show that PUNMU is positive at $m^* - \epsilon^{BST}(k)$ and negative at $m^* + \frac{k\epsilon^{BST}(k)}{N-k}$ and that therefore, the BST equilibria are locally unstable. We prove this as Claim 6 in the appendix.

6 Conclusion

In this paper, we have fully characterized all interior equilibria for the Groves Ledyard mechanism and located new equilibria that are inefficient and located on the boundary. We also have shown that the equilibria found by Bergstrom, Simon, and Titus (BST) as well as these boundary equilibria fail to exist for large values of the punishment parameter. We have further shown that the BST equilibria are not stable but that the symmetric Groves Ledyard equilibrium is stable. Remarkably, it achieves stability by combining two unstable dynamics. The boundary equilibria, which unlike the other equilibria are not efficient, are also stable. Thus, with boundaries on the message space, the Groves Ledyard mechanism may not generate efficient solutions to the free rider problem, and without boundaries it creates an unbounded mechanism (Jackson 1992) that may not converge to an equilibrium.

The fact that increasing the punishment parameter can give the Groves Ledyard mechanism a unique, stable equilibrium resurrects the mechanism from the implicit critique of Bergstrom, Simon, and Titus. Their result implied that the mechanism created a massive coordination or learning problem in which agents would have to select from among one of possibly billions of equilibria. In that the agents that send the high message get higher utility in those equilibria, this could create a complicated system. However, the fact that a large punishment parameter eradicates this problem does not mean it is an ideal solution. Increasing γ implies that all agents give the same amount. Thus, the mechanism reduces to a tax. This tax is not imposed by the rule of law but by the law of incentives.

The explicit solutions for the BST equilibria show that the agents sending the lower messages send messages that are large and negative for many parameter settings. If a constraint prevents agents from sending negative messages, the agents at the boundary will increase their messages and head back to the symmetric equilibrium. Therefore, constraints on the message space may mitigate the need to make the punishment parameter too large (Page and Tassier 2004.) It need not be the case that the BST equilibria no longer exist, only that the constraints prevent agents from finding these equilibria and the boundary equilibria. Since the BST equilibria are unstable, the real concern is preventing the boundary equilibria from arising. Our analysis suggests that careful selection of a punishment parameter and a message constraint can force the system to the equilibrium. It is not without irony that we can say that despite all of the BST equilibria and all of the inefficient boundary equilibria, that with a budget constraint we can return to what Groves and Ledyard implicitly claimed to have found: *a solution* to the free rider problem, not many, but a single one.

Appendix

Claim 1. *In an efficient allocation of the public good, where* $\delta = \frac{k\epsilon}{(N-k)}$, $T^l(m)$ *and* $T^u(m)$ *are given by the following equations.*

$$T^l(m) = \frac{\gamma}{2}\epsilon^2 \frac{N(N-2k)}{(N-2)(N-k)}$$

$$T^u(m) = -\frac{\gamma}{2}\epsilon^2 \frac{N(N-2k)k}{(N-2)(N-k)^2}$$

pf: We first calculate the difference between the low message m^l and the average of the others, $\bar{m}^l$. In the special case where $\delta = \frac{k\epsilon}{(N-k)}$,

$$(m^l - \bar{m}^l) = -\frac{\epsilon N}{(N-1)}$$

For the agents who send higher than average messages the corresponding values are

$$(m^u - \bar{m}^u) = \frac{\epsilon N k}{(N-1)(N-k)}$$

We next calculate the σ^l term within T^l. For the agents with low messages

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

which can be expanded as

$$\sigma^l \;=\; \frac{1}{N-2}[(k-1)+(N-k)]\frac{(\epsilon+\delta)^2(N-j)(k-1)}{(N-1)^2}$$

which reduces to

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

For the agents with high messages a similar calculation gives

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

We now calculate the punishment for both types of agents as a function of ϵ and δ. For the agents who send the low message the punishment is:

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

Expanding we get

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

which reduces to

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

which further reduces to

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

which is positive since $(N-2k)>0$. In the efficient case, this amount equals

$$T^l(m) = \frac{\gamma}{2}\epsilon^2 \frac{N(N-2k)}{(N-2)(N-k)}$$

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

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

Expanding we get

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

which reduces to

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

which reduces to

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

In the efficient case, this amount equals

$$T^u(m) = -\frac{\gamma}{2}\epsilon^2\frac{N(N-2k)k}{(N-2)(N-k)^2}$$

Claim 2 *The BST equilibria are characterized by* k *agents sending the message* $m^* - \epsilon^{BST}(k)$ *and* $(N-k)$ *agents sending the message* $m^* + \frac{k\epsilon^{BST}(k)}{(N-k)}$*, where* $k < \frac{N}{2}$ *and where*

$$\epsilon^{BST}(k) \ = \ \frac{2(N-2)(N-k)A(m)}{A'(m)N(N-2k)}$$

pf: In an efficient allocation of the public good, where $\delta = \frac{k\epsilon}{(N-k)}$, $T^l(m)$ and $T^u(m)$ are given by the following equations.

$$T^l(m) = \frac{\gamma}{2}\epsilon^2\frac{N(N-2k)}{(N-2)(N-k)}$$

$$T^u(m) = -\frac{\gamma}{2}\epsilon^2\frac{N(N-2k)k}{(N-2)(N-k)^2}$$

By the first order conditions, m^l, m^u, and m^* must satisfy the following two equations

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

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

Recall that in an efficient equilibrium, we have that

$$A'(m^*)w - A'(m^*)/N - \frac{A(m^*)}{N} + B'(m^*) = 0$$

Therefore, at any asymmetric efficient equilibrium it must be that

$$-A'(m^*)T^l - A(m^*)\gamma\frac{N-1}{N}(m^l - \bar{m}^l) = 0$$

For the agents who send the low messages substituting for T^l yields:

$$-A'(m)\frac{\gamma}{2}\epsilon^2\frac{N(N-2k)}{(N-2)(N-k)} = \frac{A(m)\gamma\epsilon(N-1)N}{N(N-1)}$$

which can be simplified as

$$\epsilon^{BST}(k) = \frac{2(N-2)(N-k)A(m)}{A'(m)N(N-2k)}$$

This is the amount that the low message agents deviate below the message sent in the symmetric equilibrium. Similarly, the first order condition for the high message agents reduces to the following when we substitute T^u

$$A'(m)\frac{\gamma}{2}\epsilon^2\frac{N(2k-N)k}{(N-2)(N-k)^2} = \frac{A(m)\gamma\epsilon kN(N-1)}{N(N-1)(N-k)}$$

Which can be reduced further to

$$\epsilon^{BST}(k) \; = \; \frac{2(N-2)(N-k)A(m)}{A'(m)N(N-2k)}$$

Claim 4 *Let $(m_1, m_2, \dots m_N)$ be an interior equilibrium of the Groves Ledyard Mechanism with identical agents. If $m_i \geq \bar{m}^i$ and $m_j \geq \bar{m}^j$, then $m_i = m_j$.*

pf: Set the public good part of marginal utility (PGMU) equal to X, which may be positive, negative, or zero.

$$A'(m)(w - \frac{m}{N}) - \frac{A(m)}{N} + B'(m) = X$$

It follows that the punishment part of marginal utility (PUNMU) equals $-X$.

$$-A'(m)T_i(\vec{m}) - A(m)\gamma\frac{N-1}{N}(m_i - \bar{m}^i) = -X$$

The PGMU's are identical given that the agents have identical wealth and preferences. The PUNMU's differ, we can write them as

$$-A'(m)T_i(\vec{m}) = -X + A(m)\gamma\frac{N-1}{N}(m_i - \bar{m}^i)$$

for agent i, and as

$$-A'(m)T_j(\vec{m}) = -X + A(m)\gamma\frac{N-1}{N}(m_j - \bar{m}^j)$$

for agent j.

Suppose that $m_i > m_j$. It follows then that $\bar{m}^j > \bar{m}^i$ and therefore that $(m_i - \bar{m}^i) > (m_j - \bar{m}^j)$. By the equations above it follows that $T_i(\vec{m}) < T_j(\vec{m})$. However, if we solve for $T_i(\vec{m})$ and $T_j(\vec{m})$ directly we see that $T_i(\vec{m}) > T_j(\vec{m})$, a contradiction.

$$T_i = \frac{\gamma}{2}\left[\frac{N-1}{N}(m_i - \bar{m}^i)^2 - \frac{1}{N-2}\sum_{\ell\neq i}(m_\ell - \bar{m}^i)^2\right]$$

$$T_j = \frac{\gamma}{2}\left[\frac{N-1}{N}(m_j - \bar{m}^j)^2 - \frac{1}{N-2}\sum_{\ell\neq j}(m_\ell - \bar{m}^j)^2\right]$$

Recall that $(m_i - \bar{m}^i) > (m_j - \bar{m}^j)$. It therefore suffices to show that

$$\sum_{\ell\neq i}(m_\ell - \bar{m}^i)^2 < \sum_{\ell\neq j}(m_\ell - \bar{m}^j)^2$$

We will rewrite the left hand side of the inequality in order to show it is less than the right.

$$\sum_{l\neq i}(m_l - \bar{m}^i)^2 = \sum_{l\neq i} m_l^2 - 2\sum_{l\neq i} m_l\bar{m}^i + (N-1)\bar{m}^{i2}$$

$$= \sum_{l\neq j} m_l^2 - 2\sum_{l\neq j} m_l\bar{m}^i + (N-1)\bar{m}^{i2} + m_j^2 - 2m_j\bar{m}^i - m_i^2 + 2m_i\bar{m}^i$$

Let $\Delta = \frac{m_i - m_j}{N-1}$. The previous equation reduces to:

$$= \sum_{l\neq j} m_l^2 - 2\sum_{l\neq j} m_l\bar{m}^i + 2\sum_{l\neq j} m_l\Delta - \sum_{l\neq j} m_l\Delta + (N-1)\bar{m}^{i2} + m_j^2 - 2m_j\bar{m}^i - m_i^2 + 2m_i\bar{m}^i$$

$$= \sum_{l\neq j} m_l^2 - 2\sum_{l\neq j} m_l(\bar{m}^i - \Delta) - 2\sum_{l\neq j} m_l\Delta + (N-1)\bar{m}^i + m_j^2 - 2m_j\bar{m}^i - m_i^2 + 2m_i\bar{m}^i$$

$$= \sum_{l\neq j}(m_l - \bar{m}^j)^2 - 2\Delta\sum_{l\neq j} m_l + (N-1)(\bar{m}^{i2} - \bar{m}^{j2}) + (m_j^2 - m_i^2) + 2\bar{m}^i(m_i - m_j)$$

Given that $\sum_{l \neq j}(m_l - \bar{m}^j)^2$ appears in the expression, it suffices to show that the remaining terms are negative. We can rewrite these terms as:

$$-2\Delta(N-1)(\bar{m}^j - \bar{m}^i) - (N-1)(\bar{m}^{j2} - \bar{m}^{i2}) - (m_i^2 - m_j^2)$$

The first of these three terms is clearly negative. We can rewrite the last two terms as:

$$-(N-1)[\frac{(m-m_j)^2}{(N-1)^2} - \frac{(m-m_i)^2}{(N-1)^2}] - (m_i^2 - m_j^2)$$

$$= -2m\Delta - (m_j^2 + m_i^2)(1 - \frac{1}{N-1}) < 0$$

Claim 6 *PUNMU is positive at* $m^* - \epsilon^{BST}(k)$ *and negative at* $m^* + \frac{k\epsilon^{BST}(k)}{N-k}$.

pf: Recall that

$$\epsilon^{BST}(k) = \frac{2(N-2)(N-k)A(m^*)}{A'(m^*)N(N-2k)}$$

Consider the following deviation $\epsilon = (1-\phi)\epsilon^{BST}(k)$, where $\phi > 0$ is arbitrarily small. PUNMU at $m - \epsilon$ equals

$$\gamma\left[\epsilon A(m^*) - \epsilon^2 A'(m^*)\frac{N(N-2k)}{2(N-2)(N-k)}\right]$$

substituting in the value for ϵ yields

$$\gamma\left[\frac{(1-\phi)2(N-2)(N-k)A(m^*)^2}{A'(m^*)N(N-2k)} - \frac{2(1-\phi)^2(N-2)(N-k)A(m^*)}{A'(m^*)N(N-2k)}\right]$$

which reduces to

$$\gamma\left[((1-\phi)-(1-\phi)^2)\frac{2(N-2)(N-k)A(m^*)^2}{A'(m^*)N(N-2k)}\right]$$

which is strictly positive. Therefore, the agents at $m-\epsilon$ increase their message. To show that the agents at $m + \frac{k\epsilon}{N-k}$ decrease their message, we make the following similar calculation. Recall from above that PUNMU at $m + \frac{k\epsilon}{N-k}$ equals

$$\gamma \left[-\epsilon A(m^*) \frac{k}{(N-k)} + \epsilon^2 A'(m^*) \frac{N(N-2k)k}{2(N-2)(N-k)^2} \right]$$

It follows that PUNMU at ϵ equals

$$\gamma \left[-\frac{2(1-\phi)k(N-2)A(m^*)^2}{A'(m^*)N(N-2k)} + \frac{2(1-\phi)^2 k(N-2)A(m^*)^2}{A'(m^*)N(N-2k)} \right]$$

which can be simplified as

$$\gamma \left[(-(1-\phi) + (1-\phi)^2) \frac{2(k(N-2)A(m^*)^2}{A'(m^*)N(N-2k)} \right]$$

which is negative.

References

Arifovic, Jasmina, and John Ledyard (2004) “Scaled Up Learning in Public Good Games.” *Journal of Public Economic Theory* 6 pp 205-238.

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.

Brock, W. and Durlauf, S. (2001) “Discrete Choice with Social Interactions” *Review of Economics Studies* 68, 2 235-260.

Groves, T. and J. Ledyard (1976) “Optimal Allocation of Public Goods: A Solution to the Free Rider Problem” *Econometrica* 45 (4), 783-809.

Jackson, Matthew O. Jackson (1992) “Implementation in Undominated Strategies: A Look at Bounded Mechanisms” *The Review of Economic Studies*, Vol. 59, No. 4.pp. 757-775.

Muench, Thomas, and Mark Walker (1983) ”Are Groves-Ledyard Equilibria Attainable” *Review of Economics Studies* pp 393-396.

Page, S. and T. Tassier (2004) “Equilibrium Selection and Stability in the Groves Ledyard Mechanism” *Journal of Public Economic Theory*, 6 (2), pp. 311-335..

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.