Institutions, Incentives and Behavior: Essays in Public Economics and Mechanism Design

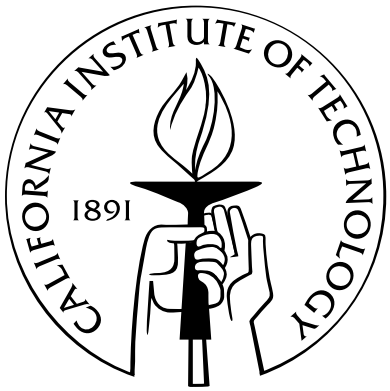
Thesis by

Paul J. Healy

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2005

(Submitted May 17, 2005)

For Meredith

Acknowledgements

I wish to thank John Ledyard for his support, funding, and encouragement. I am also indebted to Colin Camerer, Preston McAfee, and Federico Echenique for their helpful comments and frequent guidance. I am particularly grateful to John Ledyard and Charlie Plott (and the EEPS Lab at Caltech) for funding the two sets of experiments presented. Ken Binmore's energy and encouragement have had a very positive impact on my work, and for that, I thank him. Many people have contributed to my research through innumerable stimulating conversations, including (but not limited to) Kim Border, Matt Jackson, Tom Palfrey, David Grether, Simon Wilkie, Tim Cason, Ivana Komunjer, Chris Chambers, and Ernst Fehr. This work has also benefitted from research support provided by the ARCS Foundation. Finally, I am grateful for the research assistance provided by Isa Hafalir, Joel Grus, and Basit Kahn, each of whom assisted in running the experiments presented in Chapter 4.

Abstract

The economic outcomes realized by a society are a function of the institutions put in place, the incentives they create, and the behavior of agents in the face of those incentives. Selecting the appropriate institutions for a given economy is particularly important in the domain of public economics, where individual incentives are often inconsistent with efficiency. Three major concerns in institutional design are addressed. First, do agents select the equilibrium strategies at which efficient allocations obtain? Second, does the repeated game nature of a long-lived institution impact behavior? Third, what degree of coercion is necessary for a planner to guarantee that the allocation selected by a mechanism can be enforced? Answering these questions helps to understand which institutions are most appropriate in various environments. In Chapter 2, five public goods mechanisms are experimentally tested in a repeated game environment. Behavior is well approximated by a model in which agents best respond to an average of recently observed data. This model provides various sufficient conditions a mechanism must satisfy for play to converge to an efficient equilibrium. In Chapter 3, it is assumed that the designer of a one-shot mechanism must allow agents a 'no trade' option in which they are free to contribute nothing but enjoy the public good produced by others' contributions. It is shown that a large set of

economies exist in which there is some agent at every allocation who prefers this option. Even in economies where this is not true, it becomes true as the economy is replicated, making it impossible to implement any allocation except the endowment in large economies.

In the final chapter, a model of group reputations is developed to explain why moral hazard problems are significant in some laboratory experiments and less significant in others. If firms believe that either all workers are selfish or all workers are reciprocal, then selfish workers may have an incentive to develop a 'group reputation' of being reciprocal for a fixed number of periods in order to extract higher wages. As predicted, only in those experiments in which this incentive is sufficiently large is the moral hazard problem mitigated.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The economic outcomes realized by a society are a function of the institutions put in place, the incentives they create, and the behavior of agents in the face of those incentives. In situations where a social planner or government agency has both a notion of desirable outcomes and an ability to put in place certain institutions, it is imperative that the planner understand the interactions between these three elements. Economic theory has provided a solid foundation for this understanding, but it is necessarily constrained by a need for tractability. As a consequence, economic models make various specific assumptions about both the ability of the planner to select institutions and the response of agents to the incentives those institutions create. In order to move this theoretical research toward the domain of application, it is necessary to understand the robustness and the realism of these assumptions.

In economic environments where goods can be made excludable and where the agent that consumes a good is the only one who receives benefit, it is well-known that under very general assumptions, establishing and maintaining property rights is sufficient to guarantee that the selfish private exchange of goods can lead to efficient

allocations for the entire society.[1] In such situations, the benevolent social planner need not consider complex mechanisms for the achievement of socially efficient outcomes; enforcing the rights of individuals to own and trade property is sufficient. This theoretical result is particularly appealing because its assumptions about individual behavior and information are minimal and it does not require that the planner actively engage in the reallocation of property. Although the theory still grapples with exactly how prices form and how economies adjust from their initial state to an efficient allocation, it is clear that under the simplest of behavioral assumptions, agents will still arrive at these optimal outcomes.

Unfortunately, such desirable results do not obtain when the consumption and production choices of one agent have an impact on the welfare of another. The provision of a pure public good is a particularly stark example where all agents necessarily consume a single good, so the production of one agent necessarily impacts the benefits of others. Property rights alone are then insufficient for the realization of efficient outcomes because each agent has an incentive to let the others fund a given level of the public good. Samuelson [90] formally demonstrates how individual incentives lead agents to select allocations different from the socially efficient allocation so that generically, an efficient allocation cannot be an equilibrium state of the economy. In such situations, more complex institutions are needed to realize the desired outcomes. Samuelson conjectures that *no* decentralized institution can be effective when agents have the ability to misrepresent their preferences. Fueled by debates about

[1] It will be assumed throughout that the social planner's goal is efficiency, as formalized by the notion of Pareto optimality. A similar methodology could be used to study institutions, incentives, and behavior under alternative desiderada and many of the results herein would still apply.

the viability of a socialist system, various authors have worked to identify and study decentralized institutions and to determine whether or not Samuelson's conjecture is universally true.

Early research focused on models of economic *adjustment* that are guaranteed to move the economy to a welfare-maximizing state, without concern for individuals' incentives to reveal their preferences truthfully.[2] An adjustment process is said to be Pareto satisfactory if it selects a unique Pareto optimal outcome and if every Pareto optimal outcome can be attained from some initial state. Thus, the classic welfare theorems for private goods economies prove that the perfectly competitive mechanism is Pareto satisfactory. At this point, authors were primarily concerned with the amount of information that a Pareto satisfactory mechanism must acquire from its agents in order to operate under the assumption that agents would always be willing to provide *truthfully* such information when asked of them.

As the mechanism design literature progressed, authors became increasingly concerned with the assumption of truthful revelation.[3] The additional requirement of incentive compatibility was imposed on mechanisms, stating that the vector of informational messages sent by each agent be a Nash equilibrium strategy profile of the 'message-sending game'. This led to the key negative result of the literature: there does not exist an incentive compatible, Pareto satisfactory mechanism where truthful revelation of one's preferences is an equilibrium behavior. Groves & Ledyard [43] demonstrated that this impossibility result could be circumvented by using abstract

[2]See Hurwicz [48] or Hurwicz [50] for a review of this development.

[3]This concern about incentives dates back at least as far as Samuelson [90].

message spaces (rather than asking for agents' preference profiles directly) and assuming that agents will always select Nash equilibrium messages of the one-shot game defined by the mechanism. Indeed, every Nash equilibrium message profile of this proposed mechanism maps to a Pareto optimal outcome of the economy without any *a priori* information about the characteristics of the economy. Maskin [67] then provided a general theorem characterizing the desiderata (including Pareto optimality, for example) that can successfully be implemented in this way. These results show that when the Nash equilibrium assumption is used, Samuelson's concerns about the incentives to misrepresent one's preferences can be successfully avoided.

The first step in taking these theoretical solutions to real world problems is determining the conditions under which agents will behave according to the Nash equilibrium prediction. If a mechanism is proposed whose Nash equilibrium outcomes are guaranteed to be Pareto optimal, it is of use in real world settings only if Nash equilibrium is an accurate predictor of behavior. Previous experimental research on behavior in games clearly indicates that the Nash equilibrium concept is highly predictive in some situations and terribly inaccurate in others. It is generally true that in market-like trading interactions, the selfish utility maximizing model performs quite well, while in situations where one's strategies directly and obviously affect others' payoffs, behavior often deviates from the selfish prediction in a way consistent with models of fairness, inequality aversion, or reputation building. It is therefore difficult to extrapolate from these observations a prediction about behavior in mechanisms for the provision of a public good, and it certainly appears plausible that the existence of

a public good in the economy will lead players to deviate from the Nash prediction.

A second necessary extension of the theoretical work is to consider situations where the mechanism is repeated periodically. Institutions are often long-lived and agents' behavior may not be identical through time. Players may learn more about the preferences of others. They may attempt to build reputations. They might use past information to predict what others will do in an upcoming iteration. They may expect that others will use current announcements to shape their play in later periods and will thus be motivated to deviate in the present to positively affect the future. Repetition of a one-shot game opens the door to a much larger set of equilibrium predictions and off-equilibrium motivations, raising questions about the validity of the original solution in such a setting. It may be that repetition will 'undo' certain mechanisms by generating unexpected inefficiencies through dynamic behavior, while improving the outcomes of some inefficient mechanisms through the selection of repeated game equilibria that Pareto dominate the one-shot prediction.

These two concerns are summarized by two empirical questions: Will agents play the Nash equilibrium prediction, and will repetition of the mechanism alter behavior? Chapters 2 and 4 experimentally consider these two questions in the domains of public goods provision and contracting under moral hazard, respectively. The first of these chapters identifies a particular 'best response' learning model of behavior that is fairly consistent with actual play of a repeated mechanism. This model not only provides insight about which particular mechanisms will converge to their equilibrium points, but also provides sufficient conditions for a mechanism to guarantee that equilibrium

outcomes will eventually obtain. The latter chapter examines a situation in which repeated game behavior may improve the outcomes of a mechanism through rational reputation-building in the presence of stereotyping behavior. Specifically, employees may select levels of effort that are socially optimal for long periods of time because they believe their employer will erroneously categorize them (and, through stereotyping, their co-workers) as 'reciprocal' players who irrationally respond to higher wages with increased effort. The employer, believing that its workers may be reciprocal, will attempt to motivate these workers with higher pay and in fact, will receive higher effort in response for all but the final period, even if all of the workers are in fact selfish-minded individuals. Thus, in an institution where selfish behavior is generally thought to cause highly inefficient outcomes, the repeated nature of the interaction can lead to significant welfare gains for all parties involved.

Finally, the standard model of mechanism design assumes that the social planner needs only to identify desirable allocations in order for such outcomes to be realized. This paradigm implicitly assumes that the planner is equipped with some ability to enact the chosen allocations through a credible and undesirable action, should some agents not comply with the allocation. This action may take the form of explicit penalties for deviating, such as fines or imprisonment, or even the seizure of the agents' endowments and the forceful reallocation of assets. It may also take the form of an alternative allocation that any deviating agent would prefer less than the one suggested. In the example of providing a public good, it may be credible for the planner to provide no public good if even one agent deviates from the chosen

allocation. Here, any allocation preferred by every agent to their initial endowment can be credibly enforced by this possibility. On the other extreme, if the planner has no credible outside option, agents can freely ignore the suggested allocation and voluntarily provide any level of the public good they prefer – an outcome that is generically different from optimality.

Chapter 3 examines an intermediate case where the central planner cannot credibly commit to cancel all production of the public good if a single agent deviates, but does have the ability to decline the contribution of an individual if that contribution is not consistent with the chosen allocation. In other words, if the central planner requests a certain amount of money from an agent to be put into production of the public good, and that agent responds by sending a different amount of money, the planner can credibly 'tear up the check' and produce only the level of public good that can be achieved with all others' contributions. In such a situation, the planner is unable to guarantee that there exists an allocation that is mutually agreeable by all agents, and therefore, must expect that there will be situations in which one or more agents will have an incentive to not comply with the proposed allocation. Even worse, it is shown that as an economy becomes large, it is guaranteed that at least one individual will always prefer noncompliance. Thus, the ability to identify desirable allocations through a mechanism is hardly sufficient for those allocations to obtain when the planner has this level of enforcement available.

Chapter 2

Learning Dynamics for Mechanism Design

As mentioned in Chapter 1, many mechanisms have been identified whose equilibria generate efficient allocations in economies with pure public goods.[1] In general, mechanisms that require stronger equilibrium concepts are more restricted in their ability to select desirable outcomes. Theoretical results are unclear about how these trade-offs should be resolved in practice. For example, consider an environment where agents have little to no information about each others' preferences and the level of a certain public good is to be re-evaluated at regular intervals. If a social planner were asked to choose a particular mechanism in this setting, which would she prefer? Are dominant strategy equilibria necessary in this environment? Will mechanisms with stable Nash equilibria converge quickly to an efficient outcome, even though preferences are private information?

In the current chapter, five public goods mechanisms with various equilibrium

[1] Chapter 2 is reprinted with minor modifications from a *Journal of Economic Theory* forthcoming article by Paul J. Healy entitled 'Learning dynamics for mechanism design: An experimental comparison of public goods mechanisms', Copyright 2005, with permission from Elsevier.

properties are experimentally tested in a repeated game setting.[2] Specifically, the Voluntary Contribution, Proportional Tax, Groves-Ledyard, Walker, and Vickrey-Clarke-Groves mechanisms are all compared in an identical laboratory environment.[3] The goal of this research is to compare behavior across mechanisms and identify a simple learning dynamic that approximates actual behavior and correctly predicts when actions will converge to the efficient equilibria. Armed with this information, the social planner will then be able to select a mechanism whose desirable equilibrium properties should be realized in practice.

Previous experimental studies have concluded that learning dynamics play an important role in the repeated play of a mechanism. Two general observations suggest that behavior may be consistent with a learning model based on some form of best response play. First, convergence is observed only in game forms known to be supermodular, where best response play predicts convergence. Second, tests of dominant strategy mechanisms suggest that agents tend to play weakly dominated strategies that are best responses to previously observed strategy choices.[4] Motivated by these observations, the current chapter develops a simple model of best response play and finds that its predictions well approximate observed subject behavior.

The key six results of this chapter are as follows:

1. Subject behavior is well approximated by a model in which agents best respond

[2]This is, to the author's knowledge, the largest set of public goods mechanisms to be tested side-by-side to date.

[3]The cVCG mechanism refers to the Vickrey-Clarke-Groves mechanism in cases where the level of the public good is selected from a continuum. In contrast, the Pivot mechanism refers to the VCG mechanism when the public project choice is binary. The details of all five mechanisms appear in Section 2.4.

[4]An overview of previous results is provided in Section 2.1, as well as in a survey by Chen [19].

to the average strategy choice over the last few periods. This model is shown to be significantly more accurate than the stage game equilibrium prediction.

2. Half of all decisions in the cVCG mechanism are at the demand-revealing dominant strategy point, while the remainder cluster around weakly dominated best response strategies that are payoff-equivalent to the dominant strategy prediction.

3. Behavior converges close to equilibrium in the Groves-Ledyard and Voluntary Contribution mechanisms.

4. Behavior does not systematically converge in the Proportional Tax and Walker mechanisms.

5. Because of the stability results, the cVCG mechanism is found to be the most efficient. The instability in the Walker mechanism often leads to payoffs below that of the initial endowment.

6. Finally, most strategy profiles observed to be stable or asymptotically stable are approximately equilibrium strategy profiles.

Note that the model presented in result 1 successfully predicts results 2 through 6. This indicates that the model is a reasonable and tractable tool for predicting subject behavior and convergence properties of public goods mechanisms.

A brief overview of the previous experimental literature is given in the next section. The learning model and its testable implications are then introduced in Section

2.2. Details of the experimental design are outlined in Section 2.3 and a complete description of each mechanism in use is given in Section 2.4. Results and data analysis appear in Section 2.5, and Section 2.6 offers concluding remarks.

2.1 Previous Experiments

This section briefly summarizes previous experimental results on public goods mechanisms. One theme spanning these results is that behavior is, at least qualitatively, consistent with a model of best response play. This observation partially motivates the construction of the particular class of best response models in the following section.

The earliest studies of public goods provision have focused on the Voluntary Contribution mechanism. A wide variety of specifications and treatment variables have been examined, and this line of research continues to generate interesting results about preferences and behavior. A comprehensive summary of this literature is provided by Ledyard [61], who concludes that "in the initial stages of finitely repeated trials, subjects generally provide contributions halfway between the Pareto-efficient level and the free riding level," and that "contributions decline with repetition." For example, in an early paper by Isaac, McCue & Plott [51], payoffs drop from 50% of the maximum in the first period to 9% by the fifth period. Strategies quickly converge toward the free-riding dominant strategy through repetition.

In the decades since the theoretical development of public goods mechanisms designed to solve the 'free-rider' problem, experimental tests have focused primarily on

Nash mechanisms. Several studies, mostly due to Yan Chen, explore properties of the Groves-Ledyard mechanism. For example, Chen & Plott [21] study the effect of the punishment parameter and find that strategies converge rapidly to equilibrium for large parameter values – an observation consistent with known convergence results for best response dynamics in supermodular games. The authors conclude that a best response model that uses information from all previous periods is more accurate than one in which agents best respond to only the previous period.[5] Chen & Tang [22] compare the Groves-Ledyard mechanism to the Walker mechanism and find that the Groves-Ledyard mechanism is significantly more efficient, apparently due to dynamically unstable behavior in the Walker mechanism. This instability is both observed in subject behavior and predicted by best response dynamic models.

The first well-controlled laboratory test of the dominant strategy Pivot mechanism (the VCG mechanism with a binary public choice) is run by Attiyeh, Franciosi & Isaac [4]. Subjects are given positive or negative values for a proposed project and must submit a message indicating their demand. Although revealing one's demand is a (weak) dominant strategy, only ten percent of observations are consistent with this prediction, with thirteen of twenty subjects *never* revealing their true value.

Kawagoe & Mori [55] extend the Attiyeh *et al.* result by comparing the above treatment to one in which subjects are given a payoff table. The effect of having players choose from the table is significant, as demand revelation increases to 47%. Since the equilibrium is a weak dominant strategy in the sense that all agents have

[5]This model, due to Carlson [13] and Auster [5], is a best response model (defined in Section 2.2) in which predictions equal the simple average (*not* the empirical frequency) of *all* previous periods.

other strategies that are best responses to the equilibrium, the authors argue that subjects have difficulty discovering the undominated property of truth-telling because non-revelation strategies may also be best responses.

Most recently, Cason *et al.* [14] compare the Pivot and cVCG mechanisms with two players and show that deviations from truthful revelation, when they occur, tend to result in weakly dominated Nash equilibria. Specifically, assume the first subject in the Pivot mechanism announces her value truthfully. If the good is produced when the second player announces truthfully, it is also produced when he overstates his value. Thus, there exists a wide range of false announcements that can be equilibrium strategies for player two, even though they are weakly dominated by truth-telling. In the cVCG mechanism with only one preference parameter, on the other hand, agents have a strict dominant strategy to reveal truthfully. In the experiment, only half of the observed subject pairs play the dominant strategy equilibrium in the Pivot mechanism, while 81 percent reveal truthfully in the cVCG mechanism. Behaviorally, this explanation is consistent with an evolutive model where agents select payoff maximizing strategies rather than an eductive model where agents solve for the equilibria of the game.

The results of these previous studies indicate that dynamically stable Nash equilibria and strict dominant strategies are good predictors of behavior, but unstable equilibria generate unstable behavior and weakly dominated best responses may draw players away from dominant strategy equilibria. These observations are consistent with a history-dependent best response model. The goal of the current chapter is to

refine this conjecture and identify a tractable model that, if not a perfect description of behavior, can at least predict the convergence properties of a mechanism in the repeated environment.

2.2 Setup and Environment

The general environment in use is as follows:

A set of agents is given by $\mathcal{I} = \{1, \ldots, n\}$. Each has preferences for consumption of a private good $\mathbf{x} = (x_1, \ldots, x_n)$ and a single public good y that can be represented by the differentiable function $u_i(y, x_i; \boldsymbol{\theta}_i)$, where $\boldsymbol{\theta}_i \in \Theta_i$ indicates the 'type' of agent $i \in \mathcal{I}$. Specifically, $\boldsymbol{\theta}_i$ is a vector of utility parameters held by agent i. Each Θ_i is assumed to be convex and $\Theta = \times_{i=1}^{n} \Theta_i$. Unless stated otherwise, preferences are assumed throughout to be quasilinear, so that $u_i(x_i, y; \boldsymbol{\theta}_i) = v_i(y; \boldsymbol{\theta}_i) + x_i$ where $v_i(y; \boldsymbol{\theta})$ is strictly concave in y.

A public goods allocation is an $(n+1)$-tuple of the form $(y, x_1, \ldots, x_n)$. No public good exists initially, although a linear technology can be used to build $y \geq 0$ units of the public good at a cost of κy units of the private good. Given an initial endowment of the private good ω_i, consumption of the private good is given by $x_i = \omega_i - \tau_i$ for each i, where τ_i represents a transfer payment paid by agent i. Therefore, the public goods allocation is equivalently expressed as $(y, \tau_1, \ldots \tau_n)$. A vector of transfer payments is feasible if $\sum_i \tau_i \geq \kappa y$ and budget balanced if the constraint is met with equality.[6]

[6]Individual budget constraints are not imposed in the following analysis, so that τ_i may be larger than ω_i.

A mechanism is represented as a game form, indexed by g, in which agents choose a message $m_{g,i}$ from a strategy space $\mathcal{M}_{g,i}$ that is assumed here to be convex. The vector of all messages is denoted $\mathbf{m}_g \in \mathcal{M}_g = \times_{i=1}^{n} \mathcal{M}_{g,i}$. When there is no confusion, the g subscript will be dropped. For a given agent i and a vector of messages $\mathbf{m}_{-i} = (m_1, \ldots, m_{i-1}, m_{i+1}, \ldots, m_n)$, a best response message for agent i is that which maximizes i's utility under the assumption that the other agents send messages $\mathbf{m}_{-i}$. The set of best responses to $\mathbf{m}_{-i}$ in mechanism g is denoted $\mathcal{B}_{g,i}(\mathbf{m}_{-i}; \boldsymbol{\theta}_i)$. Define $\mathcal{B}_g(\mathbf{m}, \boldsymbol{\theta}) = \times_{i=1}^{n} \mathcal{B}_{g,i}(\mathbf{m}_{-i}; \boldsymbol{\theta}_i)$ to be the set of message profiles that are best responses to the profile $\mathbf{m}$. Any fixed point of the best response profile mapping is a Nash equilibrium strategy profile of the game. Formally, any equilibrium strategy profile, denoted $\mathbf{m}^*(\boldsymbol{\theta}) = (m_1^*(\boldsymbol{\theta}), \ldots, m_n^*(\boldsymbol{\theta}))$, satisfies

$$\mathbf{m}^*(\boldsymbol{\theta}) \in \mathcal{B}(\mathbf{m}^*(\boldsymbol{\theta}), \boldsymbol{\theta})$$

Note that this solution concept requires each player's equilibrium strategy to be a function of the other players' types if m_i^* varies with $\boldsymbol{\theta}_j$ for $j \neq i$. If the equilibrium message does *not* depend on the types of other agents, the Nash equilibrium is in dominant strategies. The set of Nash equilibria for a given type profile and game is given by $\mathcal{E}_g(\boldsymbol{\theta})$.

The vector of received messages $\mathbf{m}$ in mechanism g maps to a unique outcome of the form $(y_g(\mathbf{m}), \boldsymbol{\tau}_g(\mathbf{m}))$, where $y_g : \mathcal{M}_g \to \mathbb{R}_+$ determines the level of the public good chosen and $\boldsymbol{\tau}_g : \mathcal{M}_g \to \mathbb{R}^n$ determines the vector of transfer payments of the

private good to be paid by each agent.[7] The strategy space (inputs) and outcome function (outputs) completely characterize the mechanism. All mechanisms considered here are feasible and some are budget balanced.

The objective of the mechanism designer is to implement a social choice correspondence $\mathcal{F}: \Theta \twoheadrightarrow \mathbb{R}_+ \times \mathbb{R}^n$ with certain desirable properties. Let $\left(y^{\mathcal{F}}(\boldsymbol{\theta}), \boldsymbol{\tau}^{\mathcal{F}}(\boldsymbol{\theta})\right) \in \mathcal{F}(\boldsymbol{\theta})$ represent a particular public goods allocation satisfying the properties for preference parameters $\boldsymbol{\theta}$. In the public goods environment, the appropriate social choice correspondence for a utilitarian planner is the mapping $\mathcal{P}(\boldsymbol{\theta})$ that picks the set of Pareto optimal allocations $\left(y^{\mathcal{P}}(\boldsymbol{\theta}), \boldsymbol{\tau}^{\mathcal{P}}(\boldsymbol{\theta})\right)$ satisfying

$$y^{\mathcal{P}}(\boldsymbol{\theta}) \in \arg\max_{y \in \mathbb{R}_+} \left[\sum_{i=1}^{n} v_i(y; \boldsymbol{\theta}_i) - c(y) \right]$$

and such that $\boldsymbol{\tau}^{\mathcal{P}}(\boldsymbol{\theta})$ is budget balanced. A mechanism g implements $\mathcal{F}$ if, for every $\boldsymbol{\theta} \in \Theta$, the outcome function selects allocations in $\mathcal{F}(\boldsymbol{\theta})$ at every equilibrium message $\mathbf{m}^*(\boldsymbol{\theta})$. If g implements $\mathcal{P}(\boldsymbol{\theta})$, then it is said to be efficient. If $y_g(\mathbf{m}^*(\boldsymbol{\theta})) = y^{\mathcal{P}}(\boldsymbol{\theta})$ for some $\left(y^{\mathcal{P}}(\boldsymbol{\theta}), \boldsymbol{\tau}^{\mathcal{P}}(\boldsymbol{\theta})\right) \in \mathcal{P}(\boldsymbol{\theta})$ and $\boldsymbol{\tau}_g(\mathbf{m}^*(\boldsymbol{\theta}))$ is feasible but not budget balanced, then the mechanism is only outcome efficient. The surplus transfer payments in this case are assumed to be wasted and yield no value to any agent in the economy.

2.2.1 A Best Response Model of Behavior

A history-based best response learning model assumes that each agent i forms predictions about the strategies others will use in period t based on the observed strategies

[7] Set-valued mechanisms may be defined, but implementation in this context is assumed to require the selection of a unique outcome.

of the other players in periods 1 through $t-1$, denoted $\mathbf{m}_{-i}^1$ through $\mathbf{m}_{-i}^{t-1}$. Agent i's prediction about the strategy to be played by agent j in period t is represented by the function $\psi_j^i\left(m_j^1,\ldots,m_j^{t-1}\right)\in\mathcal{M}_j$, which maps each possible history of agent j into a unique pure strategy for every $j\neq i$.[8] Agent i is then assumed to select a best response to his predictions. Letting $\boldsymbol{\psi}:\mathcal{M}^{t-1}\to\mathcal{M}$ represent the vector of predictions generated from the history of play up to period t, the strategy profile occurring in period t will be an element of $\mathcal{B}\left(\boldsymbol{\psi}\left(\mathbf{m}^1,\ldots,\mathbf{m}^{t-1}\right),\boldsymbol{\theta}\right)$. If ψ_j^i is undefined for some $j\neq i$, then let the best response model predict any strategy that is a best response to some m_j. In short, best response learning models assume that players are utility maximizers, but that their predictions are myopic and may be inaccurate.

When $\mathcal{M}$ is a convex set in $\mathbb{R}^n$, a *k-period average best response dynamic* assumes that

$$\psi_j^i\left(\left\{m_j^s\right\}_{s=1}^{t-1}\right)=\bar{m}_j^{t,k}=\frac{1}{k}\sum_{s=t-k}^{t-1}m_j^s \tag{2.1}$$

for all $i,j\in\mathcal{I}$ when $t>k$, and $\psi_j^i=\emptyset$ when $t\leq k$. Let $\bar{\mathbf{m}}^{t,k}=\left(\bar{m}_j^{t,k}\right)_{j=1}^n$. In this model, agents best respond to the prediction that the average message of the previous k periods will be played in the current period. Note that $\psi_j^i\in\mathcal{M}_j$ by the convexity of $\mathcal{M}_j$.

Behaviorally, the k-period average model implies that agents best respond to an estimate of the current *trend* in the messages of other agents. Here, the estimate of trend is given by a simple moving average filter. Other filters may be used to

[8]Most dynamic models (such as fictitious play) are based on mixed strategy predictions over a finite strategy space. The best response models suggested here generate pure strategy predictions over a continuous strategy space.

determine the current trend in $\mathbf{m}_{-i}$, such as exponential smoothing or time-weighted moving averages. Although these various trend models will produce slightly different results, the implication of such models is that agents form unique, pure-strategy predictions about the decisions of others using previous observations that are not highly sensitive to period-by-period fluctuations in the history of play.

One simple fact immediate from the definition of the k-period dynamic is that strictly dominated strategies will not be observed. This provides the first testable proposition.

Proposition 2.1 *In the k-period best response dynamic, no strictly dominated strategy is observed in any period $t > k$.*

Given that this dynamic is suggested as a model capable of predicting convergence in public goods mechanisms, it is of interest to study the limiting behavior of this process. The following propositions and corollaries establish the relationship between the k-period average best response model and Nash equilibrium. Note that several of these theoretical results will be verified empirically in Section 2.5.

Proposition 2.2 *If a strategy is observed in $k+1$ consecutive periods of the k-period average best response dynamic, then it is a Nash equilibrium.*

Proposition 2.2 immediately implies the following important corollary:

Corollary 2.3 *All rest points of the k-period best response dynamic are Nash equilibria.*

The following proposition shows that convergence implies that the limit point is a Nash equilibrium.

Proposition 2.4 *Given some* $\boldsymbol{\theta} \in \Theta$*, let* $\{\mathbf{m}^t\}_{t=1}^{\infty}$ *be a sequence of strategy profiles consistent with the* k*-period average best response dynamic that converges to a profile* $\mathbf{q} \in \mathcal{M}$*. If the best response correspondence* $\mathcal{B}(\cdot, \boldsymbol{\theta})$ *is upper hemi-continuous at* $\mathbf{q}$ *and non-empty on* $\mathcal{M}$*, then* $\mathbf{q}$ *is a Nash equilibrium strategy profile at* $\boldsymbol{\theta}$*.*

This follows from the fact that $\left\{\bar{\mathbf{m}}^{t,k}\right\}_{t=1}^{\infty}$ must converge to $\mathbf{q}$, so $\left\{\mathcal{B}\left(\bar{\mathbf{m}}^{t,k}, \boldsymbol{\theta}\right)\right\}_{t=1}^{\infty}$ converges to a set containing $\mathbf{q}$.

Corollary 2.5 *Given some* $\boldsymbol{\theta} \in \Theta$*, let* $\{\mathbf{m}^t\}_{t=1}^{\infty}$ *be consistent with the* k*-period average best response dynamic that converges to a profile* $\mathbf{q} \in \mathcal{M}$*. If* y_g*,* $\boldsymbol{\tau}_g$*, and each* $u_i(\cdot; \boldsymbol{\theta}_i)$ *are continuous and single-valued, then* $\mathbf{q}$ *is a Nash equilibrium strategy profile at* $\boldsymbol{\theta}$*.*

This corollary is a simple application of the Theorem of the Maximum, which guarantees that the best response correspondence is upper hemi-continuous and non-empty under the given conditions. The notion of asymptotic stability requires that the dynamic path from *all* initial points in some neighborhood of $\mathbf{q}$ converge to $\mathbf{q}$. By Proposition 2.4, this is clearly sufficient for $\mathbf{q}$ to be a Nash equilibrium.

Corollary 2.6 *If, for some* $\boldsymbol{\theta} \in \Theta$*,* $\mathbf{q} \in \mathcal{M}$ *is asymptotically stable according to the* k*-period average best response dynamic and* $\mathcal{B}$ *is upper hemi-continuous and non-empty, then* $\mathbf{q}$ *is a Nash equilibrium strategy profile at* $\boldsymbol{\theta}$*.*

One might conjecture that the dynamic is more stable (in a global sense) under larger values of k. However, simple games can be constructed in which cycles can occur under a particular value of k, but globally stable obtains for the $k-1$ and $k+1$ dynamics.[9] It is natural to then ask what properties of a game are sufficient for global stability to obtain for *all* values of k. In the class of supermodular games (Topkis [96],) the monotonicity of the best response correspondence guarantees global stability of the 1-period best response dynamic. The following proposition demonstrates that this result extends to the k-period dynamic.

Proposition 2.7 *In a supermodular game, if, for some* $\boldsymbol{\theta} \in \Theta$, $\{\mathbf{m}^t\}_1^\infty$ *is consistent with a k-period average best response dynamic, then* $\liminf \mathbf{m}^t \geq \underline{\mathcal{E}}(\boldsymbol{\theta})$ *and* $\limsup \mathbf{m}^t \leq \overline{\mathcal{E}}(\boldsymbol{\theta})$*, where* $\underline{\mathcal{E}}(\boldsymbol{\theta})$ *and* $\overline{\mathcal{E}}(\boldsymbol{\theta})$ *are the smallest and largest pure strategy Nash equilibrium profiles at* $\boldsymbol{\theta}$*.*

Corollary 2.8 *If a supermodular game has a unique pure strategy Nash equilibrium, then the k-period average best response dynamic is globally asymptotically stable.*

The proof of Proposition 2.7 appears in the chapter appendix (Section 2.7.)[10] This result is of particular significance because it is consistent with the claim of Chen [18] and Chen & Gazzale [20] that supermodularity is sufficient for convergence in a variety of environments tested in the laboratory.

[9] Simply pick a 2-player game with $\mathcal{M}_i = [0,1]$ and where the best response function for each player equals 0 on some very small neighborhood around $(k-1)/k$ and equals 1 everywhere else. Start the dynamic with each player playing 0 for the first k periods. Cycles then emerge that jump in and out of the unique equilibrium *ad infinitum*. See also Bear [7].

[10] As an alternative method of proof, it can be established that any sequence $\{\mathbf{m}^t\}_1^\infty$ consistent with the k-period dynamic must satisfy the *adaptive dynamics* conditions of Milgrom & Roberts [71]. Proposition 2.7 then follows from Theorem 8 of that paper.

There are also two sets of 'dominant diagonal' conditions under which the 1-period dynamic is known to be globally stable. In both cases, this stability result can be shown to extend to all k-period dynamic models. The first is due to Gabay & Moulin [38], and is summarized in the following proposition:[11]

Proposition 2.9 *Assume that* $\mathcal{M} = [0, +\infty)^n$*, and for some* $\boldsymbol{\theta} \in \Theta$*, each* u_i *is twice continuously differentiable and, for every* $\mathbf{m} \in \mathcal{M}$*,* u_i *satisfies*

Pseudo-concavity: $\frac{\partial u_i}{\partial m_i}(\mathbf{m}, \boldsymbol{\theta}_i) \cdot (m_i - m_i') \geq 0 \Rightarrow u_i(\mathbf{m}, \boldsymbol{\theta}_i) \geq u_i((m_i', \mathbf{m}_{-i}), \boldsymbol{\theta}_i)$,

Coercivity: $\lim_{m_i \to +\infty} \left| \frac{\partial u_i}{\partial m_i}(\mathbf{m}, \boldsymbol{\theta}_i) \right| = +\infty$*, and*

Strict Diagonal Dominance: $\left| \frac{\partial^2 u_i}{\partial m_i^2}(\mathbf{m}, \boldsymbol{\theta}_i) \right| > \sum_{j \neq i} \left| \frac{\partial^2 u_i}{\partial m_i \partial m_j}(\mathbf{m}, \boldsymbol{\theta}_i) \right|$.

There exists a unique Nash equilibrium $\mathbf{m}^*(\boldsymbol{\theta})$ *of the mechanism and every sequence* $\{\mathbf{m}^t\}_1^\infty$ *consistent with a* k*-period best response dynamic converges to* $\mathbf{m}^*(\boldsymbol{\theta})$.

The second, more direct condition for stability of the 1-period model requires that the best response correspondence be a single-valued, linear function of the form $\mathcal{B}(\mathbf{m}, \boldsymbol{\theta}) = \mathbf{A}(\boldsymbol{\theta})\ \mathbf{m} + \mathbf{h}(\boldsymbol{\theta})$, where $\mathbf{h}(\boldsymbol{\theta}) \in \mathbb{R}^n$ and $\mathbf{A}(\boldsymbol{\theta}) = [a_{ij}(\boldsymbol{\theta})]_{i,j=1}^n$ is a real matrix for which there exists a strictly positive n-vector $\mathbf{d}$ such that $d_i > \sum_{j=1}^n d_j |a_{ij}(\boldsymbol{\theta})|$ for $i = 1, \ldots, n$. When $\mathbf{A}(\boldsymbol{\theta})$ is non-negative, this positive dominant diagonal condition is both necessary and sufficient for global stability (see Murata [75, Chapter 3] for details.) Using the distributed lag methods of Bear [6] and [7], this same condition is easily shown to be necessary and sufficient for global convergence of any best response

[11] See the chapter appendix for a proof. A similar result was previously established by Rosen [86] using the *diagonal strict concavity* condition on utilities.

dynamic whose predictions are weighted averages of past observations, including the k-period dynamics.

Note that the global stability properties of the k-period dynamic are similar to properties of other well-known learning models. Consider, for example, the fictitious play dynamic, which is best suited for games with small, finite strategy spaces. As with the k-period model, fictitious play has stable Nash equilibrium points (Brown [11]) and is globally stable in supermodular games (Milgrom & Roberts [72, Theorem 8],) but is also capable of off-equilibrium limit cycles (Shapley [93].)

2.3 Experimental Design

The five public goods mechanisms under consideration were tested in a laboratory environment using human subjects. All experiments were run at the California Institute of Technology during the 2002-03 academic year using undergraduates recruited via E-mail. Most subjects had participated in economics experiments, though none had experience with the particular game forms in the current study. Four sessions were run with each mechanism for a total of twenty sessions.[12] Each session consisted of five subjects interacting through computer terminals. Subjects only participated in one session in which they played a single mechanism fifty times against the same four cohorts. Each iteration of the mechanism is referred to as a period. Multiple sessions were run simultaneously so that more than five subjects would be in the lab

[12]Four additional sessions with the cVCG mechanism were run, but had to be discarded due to a software failure. These data are very similar to the reported sessions and feature a slightly higher frequency of demand revelation.

at the same time. Each subject knew she was grouped with four others, but could not discern which individuals were in her group.[13] Instructions were given to the subjects at the beginning of the experiment and read aloud by the experimenter. Participants were then given time to ask any clarifying questions.[14]

Before the experiment, subjects were given their preference parameters and initial endowments privately on a slip of paper. An incomplete information environment was used because the interest of this study is to identify mechanisms whose efficiency properties are robust to the assumptions of complete information. If the learning process of agents converges rapidly to a mechanism's efficient equilibrium, then a social planner need not worry about the informational assumptions of the equilibrium concept.

After receiving their private information, subjects logged into the game from their computer terminal using an Internet browser program. The software interface includes two useful tools for the subjects to use at any time. First, a history window is available that displays the results of all past periods. Subjects can see all previous outcomes, including the message they sent, taxes paid, public good levels, and profits. The entire vector of messages submitted by the other agents in previous periods is not shown; only the relevant variables used in calculating the tax, value and payoff functions are provided. Subjects can open this window at any time and are also shown the same information at the conclusion of each period. The second tool, called the 'What-If

[13] In one cVCG session, only one group of subjects was in the laboratory at the same time. The subjects were well separated to prevent communication or other out-of-experiment effects.

[14] Instructions and data are available at http://kakutani.caltech.edu/pj. Subjects were not deceived in any way during this experiment.

Scenario Analyzer,' allows subjects to enter hypothetical messages into a calculator-like program to view what levels of $y(\mathbf{m})$, $\tau_i(\mathbf{m})$ and profit would result. Each subject is shown only her own hypothetical tax, value, and profits so that subjects cannot deduce the preference parameters of other subjects. Instead of a practice period, subjects were given five minutes to experiment with the "What-If Scenario Analyzer" and ask questions.

The benefit of the 'What-If Scenario Analyzer' is that it enables subjects to perform searches over the strategy space before selecting their strategies. In this sense, it is similar to giving subjects a complete payoff table, although the current tool provides more feedback than payoffs alone. The interest of this study is to understand the learning dynamics involved as subjects resolve the uncertainty about the strategies of others. With inexperienced subjects, this process may be confounded with subjects figuring out how strategies map to outcomes. By experimenting with the 'What-If' tool before the experiment, the subjects become well informed about the game they are playing, but may be unsure what strategies others will be using. This reduces the potential confound in the observed dynamics.

Once the experiment begins, subjects enter their message (twice for confirmation) into the computer each period. The feedback at the end of the period is identical to the history information described above. Total earnings are kept at the bottom of the screen at all times, along with the current period number and the total number of periods in the game. Subjects' earnings were tallied in 'francs' and converted to dollars at the end of the experiment. Conversion rates from francs to actual dollar

Player:	1	2	3	4	5
a_i	1	8	2	6	4
b_i	34	116	40	68	44
ω_i	200	140	260	250	290

Table 2.1: Preference parameters $\boldsymbol{\theta}_i = (a_i, b_i)$ and ω_i used in all sessions.

earnings varied by mechanism between 650 to 800 francs per dollar so that typical subjects would earn forty to fifty cents for each of their fifty decisions, plus a \$5 show-up fee. Sessions typically lasted from 90 minutes to two hours.

Each subject in a session was assigned a unique player type, differing only by their utility parameters and endowments. The same five player types were used in every session. Quasilinear preferences were induced with concave quadratic values for the public good given by $v_i(y; \boldsymbol{\theta}_i) = -a_i y^2 + b_i y$ and an initial endowment of the private good, ω_i. The vector $\boldsymbol{\theta}_i = (a_i, b_i)$ and the endowment ω_i are positive for all $i \in \mathcal{I}$. The quasilinear, quadratic structure of the preferences is common knowledge across all subjects, although the vectors of individual coefficients $\boldsymbol{\theta}_i = (a_i, b_i)$ and endowment ω_i are private information. The chosen player type profile $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_5)$ and endowments are identical across all periods, sessions and mechanisms. These values are given in Table 2.1.

The marginal cost of the public good is chosen to be constant at $\kappa = 100$ in every session. As will be shown in the next section, these parameter values have been chosen to provide distinct predictions between various mechanisms. Given the quasilinear preferences, the Pareto optimal level of the public good is uniquely solved by $y^{\mathcal{P}}(\boldsymbol{\theta}) = 4.8095$. From an experimental design standpoint, a non-focal value

Mechanism	Outcome Functions	Strategy Space
Voluntary Contribution	$y(\mathbf{m}) = \sum_i m_i$ $\tau_i(\mathbf{m}) = \kappa m_i$	$\mathcal{M}_i = [0, 6]$
Proportional Tax	$y(\mathbf{m}) = \sum_i m_i$ $\tau_i(\mathbf{m}) = \kappa y(\mathbf{m})/n$	$\mathcal{M}_i = [0, 6]$
Groves-Ledyard	$y(\mathbf{m}) = \sum_i m_i$ $\tau_i(\mathbf{m}) = \frac{\kappa y(\mathbf{m})}{n} + \frac{\gamma}{2}\left(\frac{n-1}{n}(m_i - \mu_i)^2 - \sigma_i^2\right)$ $\mu_i = \frac{1}{n-1}\sum_{j\neq i} m_j \quad \sigma_i^2 = \frac{1}{n-2}\sum_{j\neq i}(m_j - \mu_i)^2$	$\mathcal{M}_i = [-4, 6]$
Walker	$y(\mathbf{m}) = \sum_i m_i$ $\tau_i(\mathbf{m}) = \left(\frac{\kappa}{n} + m_{(i-1) \bmod n} - m_{(i+1) \bmod n}\right) y(\mathbf{m})$	$\mathcal{M}_i = [-10, 15]$
cVCG	$y(\hat{\boldsymbol{\theta}}) = \arg\max_{y\geq 0}\left(\sum_i v_i(y; \hat{\boldsymbol{\theta}}_i) - \kappa y\right)$ $\tau_i(\hat{\boldsymbol{\theta}}) = \frac{\kappa y(\hat{\boldsymbol{\theta}})}{n} - \sum_{j\neq i}\left(v_j(y(\hat{\boldsymbol{\theta}}); \hat{\boldsymbol{\theta}}_j) - \frac{\kappa y(\hat{\boldsymbol{\theta}})}{n}\right)$ $+ \max_{z\geq 0}\sum_{j\neq i}\left(v_j(z; \hat{\boldsymbol{\theta}}_j) - \frac{\kappa z}{n}\right)$	$\mathcal{M}_i = \Theta_i$ $= \mathbb{R}^2$

Table 2.2: The five mechanisms tested.

for the Pareto optimum is preferred so that public good levels observed at or near Pareto optimal levels cannot alternatively be explained by subjects choosing integer strategies, for example.

2.4 The Mechanisms

The following section describes each of the five mechanisms in detail. The outcome functions and strategy spaces are presented in Table 2.2. A reader familiar with the details of public goods mechanisms may skip the discussion of the Voluntary Contribution, Proportional Tax, Groves-Ledyard, and Walker mechanisms, although the cVCG mechanism in use here has interesting properties that are critical in understanding the results presented below.

2.4.1 Voluntary Contribution Mechanism

In this simple mechanism, each player i announces m_i, the number of units of the public good to be added to the total. The sum of the contributions represents the realized level of the public good, and the tax paid by agent i is the cost of her contribution to the public good. In this mechanism, each agent has a 'Robinson Crusoe' ideal point, denoted $\tilde{y}_i$, representing the amount of public good she would contribute in the absence of contributions by others. The best response function for each agent is then given by

$$\mathcal{B}_i\left(\mathbf{m}_{-i};\boldsymbol{\theta}_i\right) = \tilde{y}_i - \sum_{j\neq i} m_j. \tag{2.2}$$

In the quadratic environment, $\tilde{y}_i = \left(b_i - \kappa\right)/2a_i$. Using the parameters of the experiment, the vector of Robinson Crusoe ideal points is

$$\tilde{\mathbf{y}} = \left(-33,\ 1,\ -15,\ -2\frac{2}{3},\ -7\right).$$

Since player 2 is the only agent for which $\tilde{y}_i > 0$, he is the only player who does not have a dominant strategy of contributing zero. The unique Nash equilibrium of this game is therefore $\mathbf{m}^*\left(\boldsymbol{\theta}\right) = (0, 1, 0, 0, 0)$, which results in a suboptimally low level of the public good. Under the k-period dynamic, this equilibrium must obtain by period $2k + 1$. Note the if the message space were unbounded, then no equilibrium would exist and the k-period model would diverge.

2.4.2 Proportional Tax Mechanism

The Proportional Tax mechanism is an alternative to the Voluntary Contribution mechanism in which each agent must pay an equal share of the total cost. Under this scheme, agents' Robinson Crusoe ideal points are given by $\tilde{y}_i = (b_i - \kappa/n)/2a_i$, which is necessarily larger than under the Voluntary Contribution mechanism. Specifically,

$$\tilde{\mathbf{y}} = (7,\ 6,\ 5,\ 4,\ 3)\,.$$

This mechanism has the same message space and best response function as the Voluntary Contribution mechanism (eq. 2.2), but no agents have a dominant strategy in this mechanism since $\tilde{\mathbf{y}} \gg \mathbf{0}$. The unique pure strategy Nash equilibrium is the corner strategy profile $\mathbf{m}^*(\boldsymbol{\theta}) = (6, 0, 0, 0, 0)$, which results in a suboptimally *large* level of the public good.[15] Note that although no agent has a dominant strategy, players 3, 4, and 5 are only willing to contribute when the total contributions of all others is well below their ideal points.

The Proportional Tax mechanism is also of interest because it provides the foundation of both the Groves-Ledyard and the Walker mechanisms. One can show that if the Proportional Tax mechanism had an interior Nash equilibrium, it would select an optimal level of the public good. The problem is that, generically, interior equilibria do not exist. The Groves-Ledyard and Walker mechanisms are essentially variants of the Proportional Tax mechanism with an additional 'penalty' term in the transfer

[15] If $\mathcal{M} = \mathbb{R}$, then no equilibrium would exist. If the upper bound of the message space were chosen to be greater than 7, then the equilibrium would be $(7, 0, 0, 0, 0)$.

function chosen to guarantee the existence of an interior equilibrium point.

2.4.3 Groves-Ledyard Mechanism

The mechanism of Groves & Ledyard [43] was the first constructed whose Nash equilibria yield fully efficient outcomes. The mechanism requires all agents pay an equal share of the cost plus a penalty term based on deviations from the average of the others' contributions and on the variance of those contributions. Each agent's unique best response message in the quadratic, quasilinear environment is given by

$$\mathcal{B}_i\left(\mathbf{m}_{-i};\boldsymbol{\theta}_i\right)=\frac{b_i-\kappa/n}{2a_i+\gamma\frac{n-1}{n}}+\left(\frac{\gamma/n-2a_i}{2a_i+\gamma\frac{n-1}{n}}\right)\sum_{j\neq i}m_j.$$

In the experimental environment, this mechanism has a unique pure strategy of

$$\mathbf{m}^*\left(\boldsymbol{\theta}\right)=\left(1.0057,\,1.1524,\,0.9695,\,0.8648,\,0.8171\right),$$

which results in $y\left(\mathbf{m}^*\left(\boldsymbol{\theta}\right)\right)=4.8095=y^{\mathcal{P}}\left(\boldsymbol{\theta}\right)$.[16] The message space $\mathcal{M}_i=[-4,6]$, which is identical to that used by Chen & Plott [21] and Chen & Tang [22], is sufficiently wide so that the equilibrium is not near a corner of the strategy space.

One nice property of the Groves-Ledyard mechanism is that it can possess the global stability properties of Propositions 2.7 and 2.9. In the quadratic environment, the equilibrium is supermodular if $\gamma>2n\max_i\left(a_i\right)$ and satisfies the positive dominant diagonal condition if $\gamma>\left[\left(n-2\right)/\left(n-1\right)\right]n\max_i\left(a_i\right)$. Using the experiment

[16] Note that the equilibrium strategy profile has less variance for larger values of γ. Here, γ is chosen to be large for stability reasons.

parameters, these conditions are $\gamma > 80$ and $\gamma > 30$, respectively. Since $\gamma = 100$ is used in the experiment, both sufficient conditions for stability are satisfied. Therefore, all k-period best response dynamics are globally stable in this setting.

2.4.4 Walker Mechanism

One important theoretical drawback of the Groves-Ledyard mechanism is that it may select some efficient allocations that do not Pareto dominate the initial endowment. Partially in response to this issue, Walker [100] developed a 'paired difference' mechanism that implements Lindahl allocations in Nash equilibrium. Since Lindahl allocations are guaranteed to Pareto dominate the initial endowment and tax each agent based on their marginal willingness to pay, Walker's mechanism appears to provide the most desirable solution to the free-rider problem.

In the quadratic environment, the unique best response function is given by

$$\mathcal{B}_i\left(\mathbf{m}_{-i};\boldsymbol{\theta}_i\right) = \frac{b_i - \left(\kappa/n + m_{(i-i) \bmod n} - m_{(i+1) \bmod n}\right)}{2a_i} - \sum_{j \neq i} m_j.$$

Solving for the equilibrium with the given parameters,

$$\mathbf{m}^*\left(\boldsymbol{\theta}\right) = \left(12.276,\, -1.438,\, -6.771,\, -2.200,\, 2.943\right),$$

which gives the Lindahl allocation of $y = 4.8095$ and

$$\tau = (117.26,\, 187.8,\, 99.855,\, 49.469,\, 26.567)$$

To accommodate the disperse equilibrium messages, the message space is expanded to $\mathcal{M}_i = [-10, 15]$ for each $i \in \mathcal{I}$.

Although this mechanism implements Lindahl allocations in Nash equilibrium, its equilibria are known to have instability problems. In the quadratic environment, the 1-period best response dynamic can be represented by the system of difference equations $\mathbf{m}^t = \mathbf{A}\mathbf{m}^{t-1} + \mathbf{h}$, where the row sums of $\mathbf{A}$ all equal $-(n-1)$. This matrix is irreducible and non-positive, so by Harriff *et al.* [45, Corollary 1], its dominant eigenvalue must then equal $-(n-1)$, which is greater than 1 in absolute value. Thus, Cournot best response is unstable under any parameter choice $\boldsymbol{\theta}$.[17] Chen & Tang [22] also argue that the cost of small deviations from equilibrium in the Walker mechanism are lower than in the Groves-Ledyard mechanism, making the former less robust to experimentation.

2.4.5 Continuous VCG (cVCG) Mechanism

The cVCG mechanism represents a particular selection from the class of dominant strategy incentive compatible mechanisms developed independently by Vickrey [99], Clarke [24], and Groves [42]. In these direct mechanisms (where $\mathcal{M}_i = \Theta_i$), truth-telling weakly dominates all other strategies. However, given any $\mathbf{m}_{-i}$, there exist messages $\mathbf{m}_i \neq \boldsymbol{\theta}_i$ such that $\mathbf{m}_i \in \mathcal{B}_i(\mathbf{m}_{-i}; \boldsymbol{\theta}_i)$. As will be demonstrated, with two preference parameters, all points on a particular line in the strategy space that inter-

[17] If $\mathbf{A}$ were non-negative, instability of the one-period model would be sufficient (and necessary) for instability in the k-period model. This result does not extend to the case of non-positive matricies. However, for the experiment parameters, the k-period model is unstable for all values of k in the range under consideration, and it is conjectured that *no* value of k will guarantee stability in this environment. See Harriff *et al.* [45, p. 359] for the relevant theorems and counter-examples.

sects the 'truth-telling' strategy are best responses to $\mathbf{m}_{-i}$. Consequently, a best response learning model predicts that agents select messages from the best response line that are not necessarily the truth-telling equilibrium. Given that the dominant strategy equilibrium is a zero-dimensional set, the best response set is one-dimensional, and the strategy space is two-dimensional, it is easy to distinguish between equilibrium, best response, and random (or, unexplained) strategy choices.

In the cVCG mechanism, $\mathcal{M}_i = \Theta_i$, which equals $\mathbb{R}^2_+$ in the experiment, and any message $\mathbf{m}_i$ can be equivalently expressed as an announced parameter value $\hat{\boldsymbol{\theta}}_i = (\hat{a}_i, \hat{b}_i)$. Agents are free to misrepresent preferences by announcing $\hat{\boldsymbol{\theta}}_i \neq \boldsymbol{\theta}_i$. The outcome function takes the vector of announced parameter values $\hat{\boldsymbol{\theta}}$ and solves for the Pareto optimal level of the public good on the assumption that $\hat{\boldsymbol{\theta}}$ is the true vector of preference parameters.

Each agent's tax is comprised of three parts: an equal share of the cost of production, a reward equal to the net utility of all other agents assuming their preference announcements are truth-telling, and a penalty equal to the maximum possible net utility of all others under their given preference announcement. The third term necessarily dominates the second, so the sum of transfers is always weakly greater than the cost of the project.[18]

This mechanism is constructed so that each agent i prefers an announcement $\hat{\boldsymbol{\theta}}_i$ that yields a Pareto optimal level of the public good under the assumption that $\hat{\boldsymbol{\theta}}_{-i} = \boldsymbol{\theta}_{-i}$. To see this directly, note that the first-order condition for utility maximization

[18]This mechanism is known to be budget-balanced in the quadratic environment when all agents have the same slope parameter. This is not true in the current environment.

with quadratic preferences is that

$$\frac{dy}{d\hat{\boldsymbol{\theta}}_i}\left(\left[b_i+\sum_{j\neq i}\hat{b}_j-\kappa\right]-y(\hat{\boldsymbol{\theta}})\left[2\left(a_i+\sum_{j\neq i}\hat{a}_j\right)\right]\right)=\mathbf{0}.$$

Since $dy/d\hat{\boldsymbol{\theta}}_i \neq \mathbf{0}$ for all $\hat{\boldsymbol{\theta}}_i \in \Theta_i$ and all $i \in \mathcal{I}$, utility maximization is achieved by setting the term in parentheses to zero through manipulation of $y(\hat{\boldsymbol{\theta}})$. The necessary and sufficient condition for maximization is therefore

$$y(\hat{\boldsymbol{\theta}}_i,\hat{\boldsymbol{\theta}}_{-i})=\frac{b_i+\sum_{j\neq i}\hat{b}_j-\kappa}{2\left(a_i+\sum_{j\neq i}\hat{a}_j\right)}=y(\boldsymbol{\theta}_i,\hat{\boldsymbol{\theta}}_{-i}). \tag{2.3}$$

Thus, any announcement by player i that results in the same level of the public good as would have obtained under truth-telling is necessarily a best response.

Since $\hat{\boldsymbol{\theta}}_i=\boldsymbol{\theta}_i$ satisfies (2.3) for all $\hat{\boldsymbol{\theta}}_{-i}$, truth-telling is a dominant strategy. Given a particular value of $\hat{\boldsymbol{\theta}}_{-i}$, however, there exists a range of $\hat{\boldsymbol{\theta}}_i \neq \boldsymbol{\theta}_i$ that satisfy condition (2.3). This set of values is given by the best response correspondence

$$\mathcal{B}_i\left(\hat{\boldsymbol{\theta}}_{-i};\boldsymbol{\theta}_i\right)=\left\{(\hat{a}_i,\hat{b}_i)\in\Theta_i:\left(\hat{b}_i-b_i\right)=2\left(\hat{a}_i-a_i\right)y(\boldsymbol{\theta}_i,\hat{\boldsymbol{\theta}}_{-i})\right\}. \tag{2.4}$$

Clearly, the manifold of best responses to $\hat{\boldsymbol{\theta}}_{-i}$ is a line through $\not\leq_i$ that must contain $\boldsymbol{\theta}_i$. The slope of this line depends on $\hat{\boldsymbol{\theta}}_{-i}$, so the best response line rotates about $\boldsymbol{\theta}_i$ as $\hat{\boldsymbol{\theta}}_{-i}$ varies.[19] If an agent holds a prediction that places non-zero probability on multiple values of $\hat{\boldsymbol{\theta}}_{-i}$, then the dominant strategy point becomes the unique best

[19]Note that $\hat{\boldsymbol{\theta}}_i$ affects agent i's utility only through the value of $y(\hat{\boldsymbol{\theta}}_i,\hat{\boldsymbol{\theta}}_{-i})$, so indifference curves in agent i's strategy space correspond to level curves of the $y(\cdot,\hat{\boldsymbol{\theta}}_{-i})$ function. The set $\mathcal{B}_i(\hat{\boldsymbol{\theta}}_{-i};\boldsymbol{\theta}_i)$ is therefore the level set of i's most preferred quantity of the public good, given $\hat{\boldsymbol{\theta}}_{-i}$.

response.

It is important to reiterate the fact that the set of Nash equilibria of this mechanism extends beyond the dominant strategy equilibrium. As a simple example, if $n-1$ agents submit $\hat{\boldsymbol{\theta}}_i = \boldsymbol{\theta}_i$ while the n^{th} agent announces $\hat{\boldsymbol{\theta}}_n \in \mathcal{B}_n\left(\boldsymbol{\theta}_{-n}; \boldsymbol{\theta}_n\right) \setminus \{\boldsymbol{\theta}_n\}$, then a Nash equilibrium with a weakly dominated strategy has obtained. Since $y(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_{-n}) = y\left(\boldsymbol{\theta}\right) = y^{\mathcal{P}}\left(\boldsymbol{\theta}\right)$, this equilibrium is also outcome efficient.

Although this mechanism is known to be inefficient due to its lack of budget balance, the size of the predicted inefficiency varies with the parameter choices. The results obtained in the laboratory may be sensitive to the choice of preference parameters. In the current experiment, equilibrium efficiency is over 99%. The discussion in Section 2.5.8 will highlight the significance of this (or any) fixed parameter choice in analyzing the results.

2.5 Results

2.5.1 Calibrating the Parameter k

Using the observed data, best response model predictions for each period $t > k$ are generated for $k \in \{1, \ldots, 10\}$ and compared to the observed message. To focus further analysis of the best response models, the value of k that minimizes the mean absolute deviation between the best response prediction and the data is selected from

$k \in \{1, \ldots, 10\}$. Define k^* to be the parameter that minimizes

$$\sum_{g,i=1}^{5} \sum_{s=1}^{4} \frac{1}{51 - t_{\min}} \sum_{t=t_{\min}}^{50} \left(\inf_{\hat{m}_i \in \mathcal{B}_{g,i}\left(\bar{\mathbf{m}}^{t,k}_{g,s,-i}; \boldsymbol{\theta}_i\right)} \left\| m^t_{g,s,i} - \hat{m}_i \right\| \right),$$

where g represents each of the five mechanisms under consideration, s indexes the 4 identical sessions of each mechanism, and $\|\cdot\|$ is the standard Euclidean norm.[20] Since the first k periods of each model are used to seed the initial beliefs, they must be excluded from analysis. Consequently, $t_{\min}$ must be strictly larger than k. In four of the five mechanisms, $\mathcal{B}_{g,i}(\bar{\mathbf{m}}^{t,k}_{g,s,-i}; \boldsymbol{\theta}_i)$ is unique and $\mathcal{M}_{g,i} \subseteq \mathbb{R}^1$, so the term in parentheses reduces to a simple absolute difference. In the cVCG mechanism, this term represents the orthogonal distance from the observed message to the appropriate best response line.

Table 2.3 reports the average deviation for various values of k and $t_{\min}$. Note that for every value of k considered, the average score decreases in $t_{\min}$, indicating that the models are less accurate in early periods than in later periods. Therefore, comparisons between models should only be made for fixed values of $t_{\min}$.

Given that messages are serially dependant and the nature of this dependence

[20] A 'scoring rule' such as the quadratic scoring rule characterized by Selten [92] would be more appropriate if the behavioral models generated probabilistic predictions of play. With deterministic behavioral models and a continuum strategy space, the scoring rule simply counts the number of observations that *exactly* match the prediction. In the Walker mechanism, for example, 47,025 individual predictions are generated across the 10 models and only one of them is exactly correct. Since the strategy space is endowed with a distance metric, a notion of error based on that metric is used rather than a measure of error in the space of (degenerate) probability distributions. This is similar in spirit to the notion of error in econometric models, where mean squared deviation is most often used because it is tractable and can be interpreted as an estimate of variance. These considerations do not apply here, and the mean absolute deviation measure is more robust to outliers.

	First time period used in calculating average minimum deviation ($t_{\min}$)									
k	2	3	4	5	6	7	8	9	10	11
1	**1.407**	1.394	1.284	1.151	1.104	1.088	1.072	1.054	1.054	1.049
2	-	**1.240**	1.135	0.991	0.967	0.949	0.932	0.922	0.913	0.910
3	-	-	**1.097**	0.963	0.940	0.925	0.904	0.888	0.883	0.875
4	-	-	-	**0.952**	0.932	0.915	0.898	0.877	0.866	0.861
5	-	-	-	-	**0.924**	0.911	**0.895**	**0.876**	**0.860**	**0.853**
6	-	-	-	-	-	**0.911**	0.897	0.881	0.868	0.854
7	-	-	-	-	-	-	0.899	0.884	0.873	0.863
8	-	-	-	-	-	-	-	0.884	0.874	0.864
9	-	-	-	-	-	-	-	-	0.879	0.870
10	-	-	-	-	-	-	-	-	-	0.875

Table 2.3: Calculated average quadratic score for various k-period best response models. Boldfaced entries represent, for each value of $t_{\min}$, the smallest average quadratic score among the 10 models tested. Note that the measure cannot be calculated for $k \geq t_{\min}$ since k periods are used to "seed" the model.

is unknown, no appropriate notion of significance is applicable to this analysis.[21] The objective of this subsection is to make further analyses tractable by selecting a single value k^* to represent the class of k-period best response models. Therefore, statistical significance of the difference in quality of fit between best response models is unimportant in this context; choosing the minimum-deviation model is sufficient.

Result 2.1 *Among the k-period best response models with $k \in \{1, \ldots, 10\}$, the 5-period model is estimated to be the most accurate.*

Support. The result follows immediately from inspection of the average deviation measures in Table 2.3. The measures are strictly decreasing in k for all $t_{\min} \leq 5$ (for

[21] Serial dependence is clear from inspection of correlograms. Several models of serial dependence were estimated, including various time trend regressions, GARCH models, and a variety of stochastic differential equations. None of these procedures fit the data well or generated an uncorrelated error structure.

which the $k = 5$ model cannot be calculated). For every $t_{\min} \geq 6$, $k = 5$ minimizes the average score, with one minor exception.[22] ■

Other types of best response models were also considered. For example, a more general k-period model that includes a discount factor so that more recent observations receive greater weighting slightly outperforms the undiscounted $k = 5$ model, but the increase in accuracy is marginal, considering the added parameter. Empirical analysis of the best response models will henceforth be limited to the undiscounted 5-period model. [23]

2.5.2 Best Response in non-VCG Mechanisms

Due to the substantial difference between the structure of the first four mechanisms and that of the cVCG mechanism, results pertaining to the latter will be considered separately.

Given that each of these public goods mechanisms was developed under the assumption that agents play Nash equilibrium strategies, the static Nash equilibrium serves as a key benchmark against which the best response models may be tested, even though the experimental environment is one of incomplete information. This is particularly true for the Groves-Ledyard and Walker mechanisms, where the Nash equilibrium is the only point at which efficient outcomes obtain. If a dynamic best response model is found to provide significant improvement in predictive power over

[22] Table 2.3 was also generated using a squared deviation metric. In this case, $k = 8$ yields slightly smaller error measures than $k = 5$ when $t_{\min} = 8$ or 10, but $k = 5$ is more often the minimizer.

[23] Complete analysis was performed on all models in $k \in \{1, \ldots, 10\}$, and results for $k \geq 2$ are similar to the case of $k = 5$. As is apparent from Table 2.3, the $k = 1$ model is notably less accurate than the others because the smoothing achieved by the $k \geq 2$ models provides a better fit.

Nash equilibrium, then mechanism design theory is improved by insisting that mechanisms converge quickly under this dynamic.

2.5.3 Comparison of Best Response and Equilibrium Models

The goal of this section is to determine whether the error of the best response model is significantly smaller than the error of the Nash equilibrium prediction. Standard parametric tests are inappropriate for this data because a one-sample runs test for randomness indicates that neither time series of errors is randomly drawn from a zero-median distribution, and tests for correlation indicate that the errors are serially dependent.[24] This non-stationarity implies that statistics aggregated across time may be easily misinterpreted.[25] For example, the average prediction error across all periods does not estimate the expected error in any one period; analysis of the average must be considered specific to the length of the experiment.[26] For these reasons, empirical analysis is performed on each player type in each period individually, with data aggregated only across the four sessions of each mechanism. The results of these period-by-period tests cannot be aggregated across time.

The prediction error of each model averaged across the four sessions is presented in

[24]The runs test indicates that the best response model errors for 16 of the 20 total player types are not evenly scattered about zero at a significance level of 5%. Each of the 4 player types with model errors apparently randomly drawn from a zero-median distribution were from different mechanisms, indicating that the assumption of mean-zero random errors for all player types in any one mechanism is likely invalid. The errors of the equilibrium model were not evenly scattered about zero for 19 of 20 player types at the 5% significance level. Tests of first-order correlation indicate that the errors are serially correlated for all 20 player types in both models.

[25]The dependence also implies that neither model fully captures the true dynamics of subject behavior in repeated games.

[26]This is a point occassionally forgotten in past analyses of time series data in experiments, leading to results that likely depend on the somewhat arbitrary choice of experiment length.

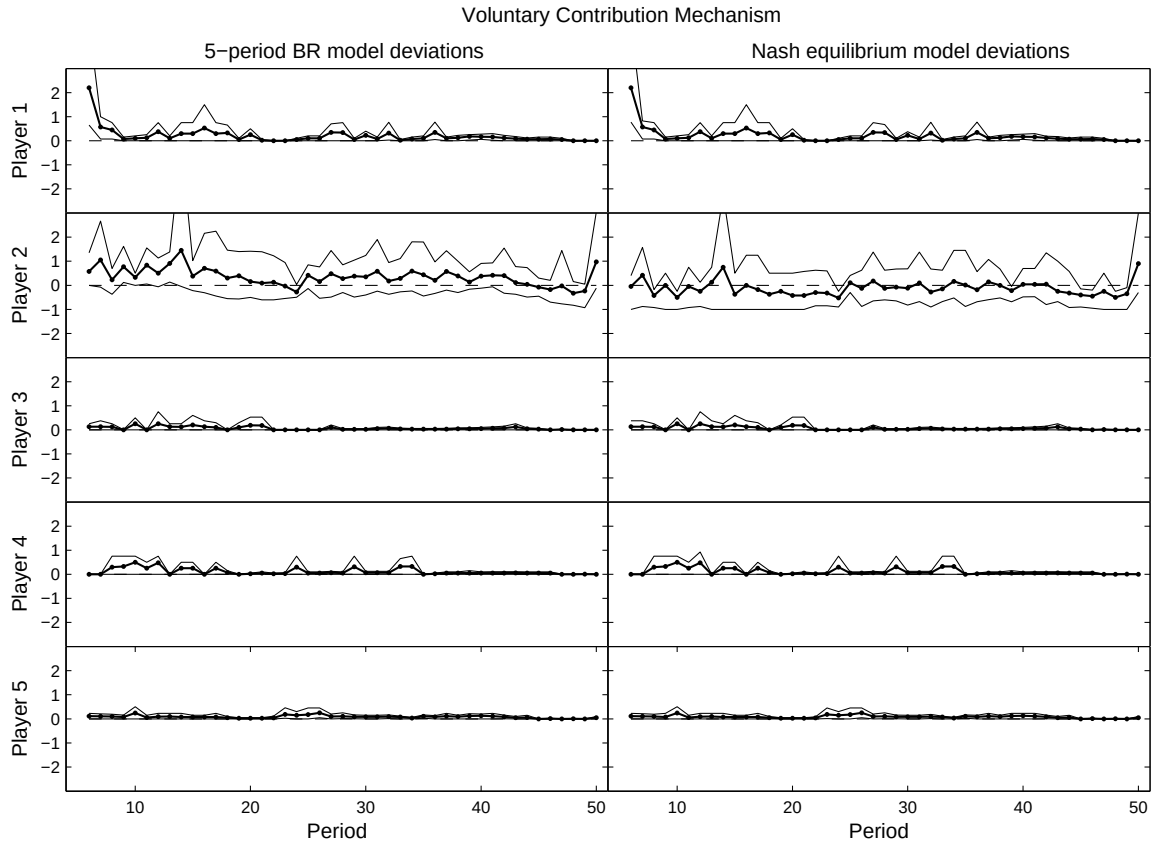


Figure 2.1: Average model error for the 5-period best response model and Nash equilibrium models in the Voluntary Contribution mechanism.

Figures 2.1 through 2.4. Around each data point is a 95% confidence interval generated by the bias-corrected and accelerated bootstrapping method.[27] These graphs begin to illustrate the superiority of the best response model over the Nash equilibrium model. While the two predictions are often very similar, there are certain player types for whom the equilibrium model systematically under- or over-predicts the observed strategies. The best response model appears both more accurate and more precise than the equilibrium model whenever differences between the two are observed.

The statistical analysis is aimed at testing the null hypothesis

$$H_0 : \mathbb{E}\left[\left|m_i^t - m_i^*\left(\boldsymbol{\theta}\right)\right|\right] \leq \mathbb{E}\left[\left|m_i^t - \mathcal{B}_i\left(\bar{\mathbf{m}}_{-i}^{t,k}; \boldsymbol{\theta}_i\right)\right|\right] \tag{2.5}$$

for each player type i and period $t > k$, where the expectation is taken across the four sessions of each mechanism. A non-parametric permutation test for a difference in means between the two model errors is performed in each period for each player type in each mechanism. Each test was based on a simulated distribution of 2,000 draws, more than enough to minimize the variation in estimated p-values due to random sampling.

The power of the permutation test depends on this difference between the predictions of the two models. If the two models have very similar predictions, the outcome of the test will not yield strong posteriors about the truth of the alternative hypothesis. In the following analysis, tests will only be run when there is enough power to

[27] Two thousand draws are used in each period for each player type in each mechanism, which is more than enough to eliminate any bootstrap sampling error. See Efron & Tibshirani [28] for details on the bootstrapping method and related statistical tests.

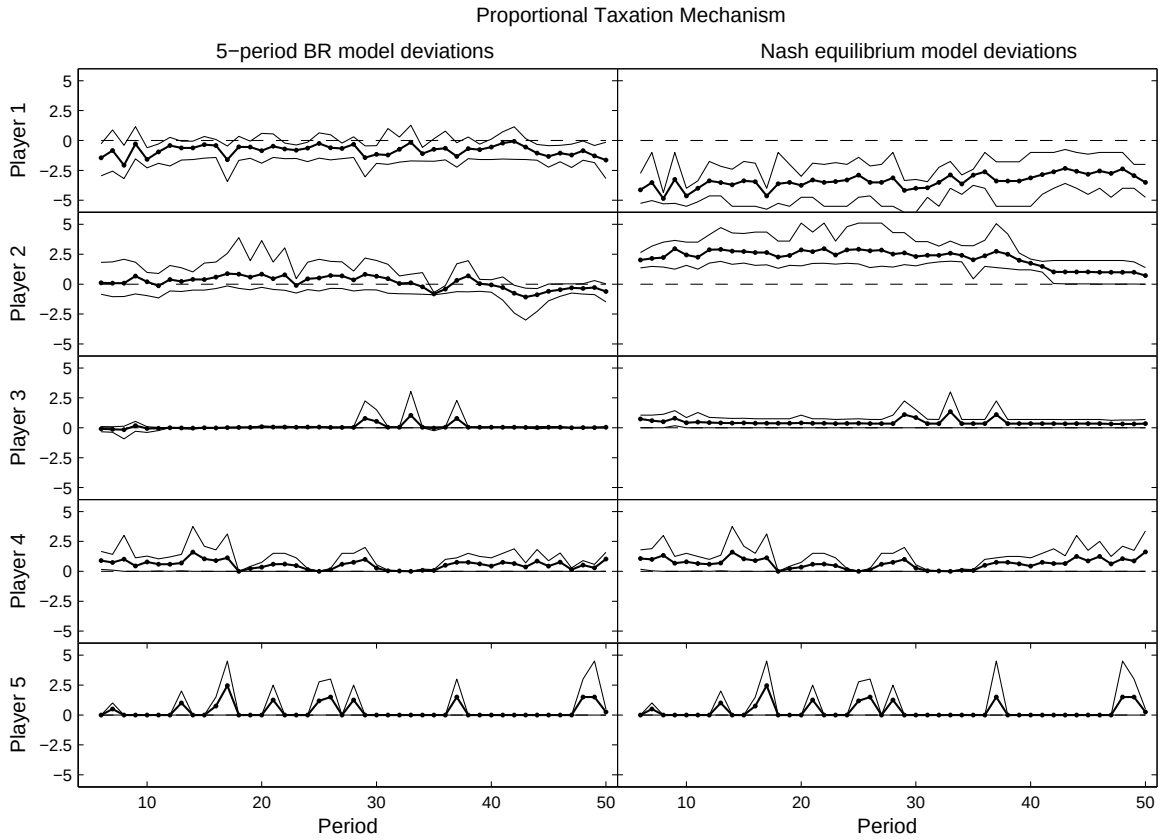


Figure 2.2: Average model error for the 5-period best response model and Nash equilibrium models in the Proportional Taxation mechanism.

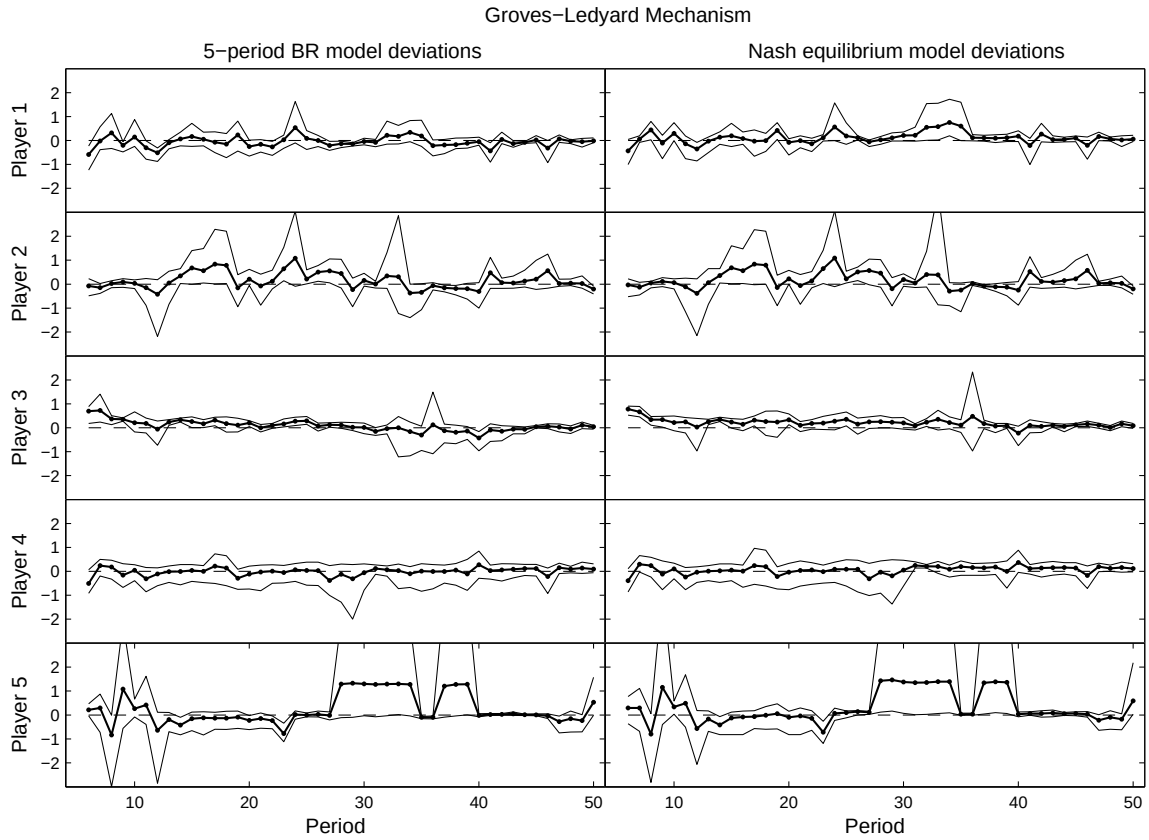


Figure 2.3: Average model error for the 5-period best response model and Nash equilibrium models in the Groves-Ledyard mechanism.

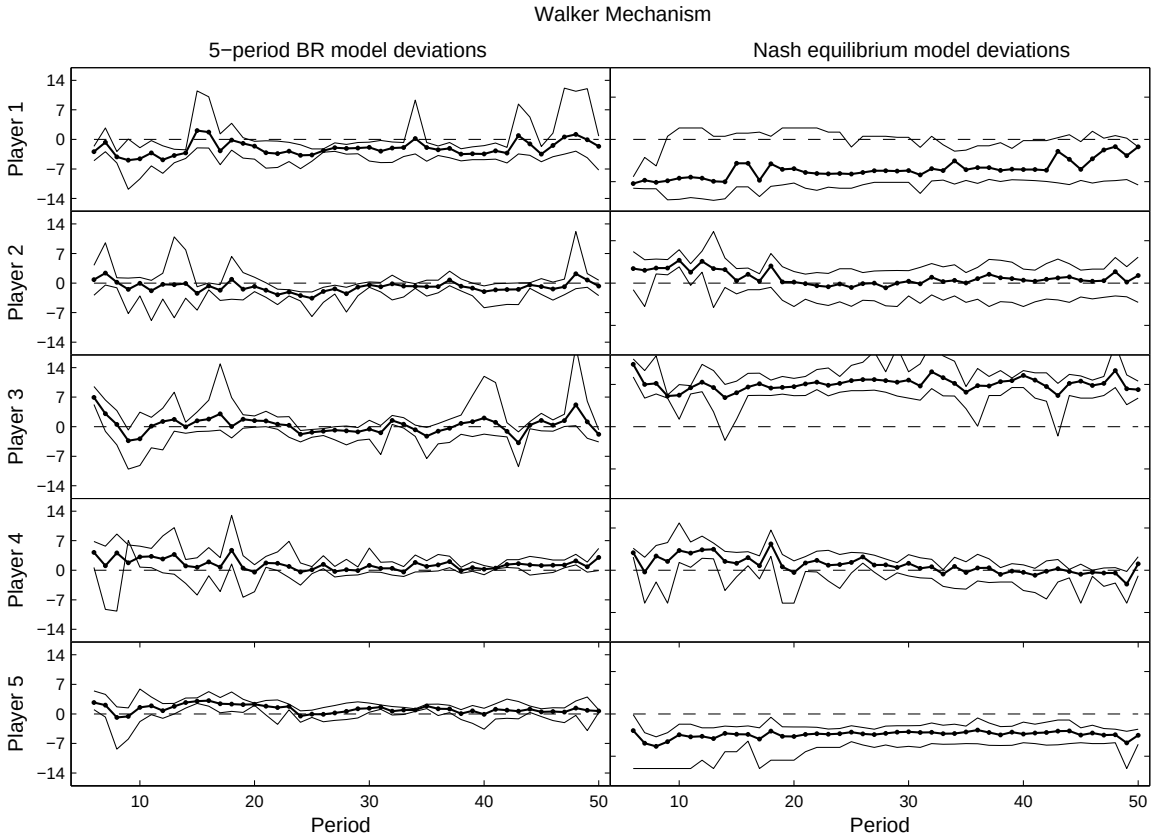


Figure 2.4: Average model error for the 5-period best response model and Nash equilibrium models in the Walker mechanism.

conclude that $\mathbb{P}[H_A \text{ True} \mid \text{Reject } H_0] \geq 90\%$. Assuming diffuse priors on the truth of H_A and a standard significance level of 5%, this is equivalent to requiring a test power of at least 45%. Without this level of power, it is likely that a rejection of H_0 is due to random sampling rather than an actual difference in model errors. If one additionally required that $\mathbb{P}[H_A \text{ True} \mid \text{Do Not Reject } H_0] \leq 10\%$, a power of 89.4% would be needed.

In order to identify what difference between model predictions is necessary to guarantee a test power of 45%, a simulation of the permutation test is performed for

various differences between model predictions. Specifically, four independent, normally distributed random messages $(w_1, \ldots, w_4)$ are generated with mean μ_a and variance σ_w^2. These represent four observations of a particular player type in a particular period. This is repeated 100 times and the permutation test is performed in each repetition on the hypotheses $\tilde{H}_0 : \mathbb{E}[|w - \mu_b|] \leq \mathbb{E}[|w - \mu_a|]$, where μ_a and μ_b represent the predictions of two different models, the first of which is correct in the sense that it predicts the true mean of the data. An estimate of the power of the permutation test is given by the percentage of simulated tests that correctly reject $\tilde{H}_0$. The simulation is repeated for various values of $(\mu_a - \mu_b)$ and σ_w^2, and the estimated power of the test is plotted as a function of $(\mu_a - \mu_b)/\sigma_w$ in Figure 2.5.[28] From this graph, it is clear that the distance between the two predictions should be at least 1.75 standard deviations of the data in order to keep the probability of incorrect rejections of H_0 to under 10%.

Figures 2.6 through 2.9 display the p-value of the permutation test for each player type in each period, along with the estimated power of each test (from Figure 2.5), and for those tests with power greater than 45%, whether or not the test rejects the null hypothesis at the 5% and 10% significance levels.

Result 2.2 *The 5-period best response model is overall a more accurate model than the Nash equilibrium model for the non-VCG mechanisms.*

Support. In the Voluntary Contribution mechanism (Figure 2.6), players 1, 3, 4, and 5 have a strict dominant strategy, so the power of the test is zero for these

[28] It should be noted that if the mean of the data were $\mu_w \neq \mu_a$ and $\mu_w > \mu_a > \mu_b$, then the test would have more power. If $\mu_a > \mu_w \geq (\mu_a + \mu_b)/2$, the test would have less power.

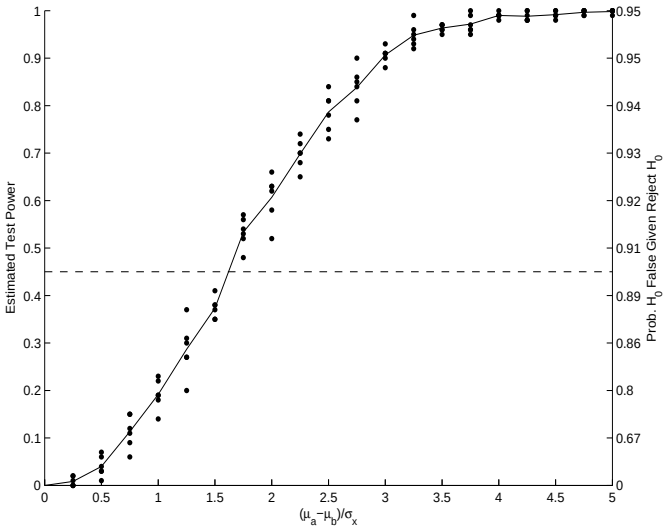


Figure 2.5: Simulated power of the permutation test given in Equation for various differences in model predictions as a ratio of the data's standard deviation.

player types. Although the p-values for player 2 never indicate a rejection of the null hypothesis, the power of the test is always below 40%, rendering its conclusions ambiguous. In the Proportional Tax mechanism (Figure 2.7,) players 3, 4, and 5 have little incentive to contribute, given the contributions of others so that the free-riding equilibrium strategy is most often a best response. For players 1 and 2, best response is occasionally far from equilibrium, providing enough power for the permutation tests to be conclusive. For player 1, all 16 tests with sufficient power reject the null hypothesis at the 10% level, and 15 of 16 reject at the 5% level. Player 2's results are similar, although the data revert toward equilibrium in the final periods (see Figure 2.2 as well.) The rapid convergence of the Groves-Ledyard mechanism (Figure 2.8) to equilibrium, which is accompanied by the convergence of best response predictions to equilibrium, reduces the power of the test in most periods. The ten tests with

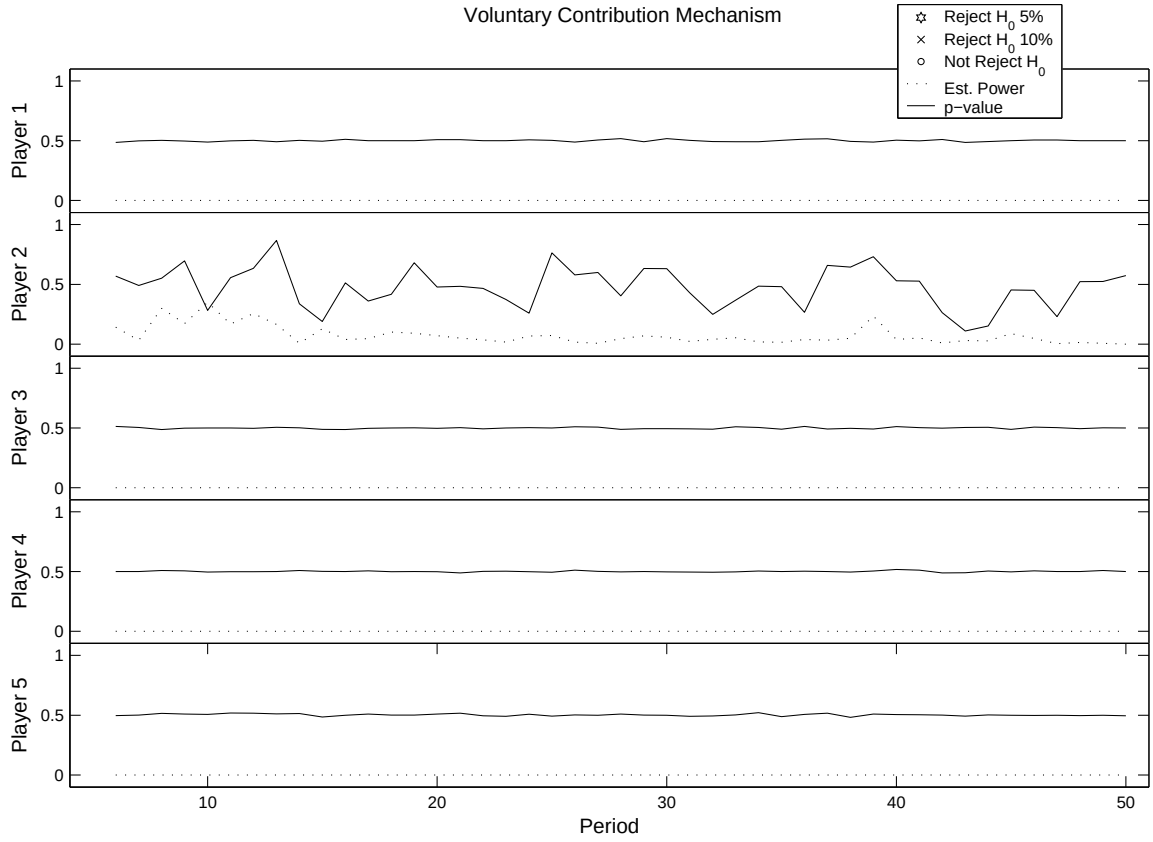


Figure 2.6: p-values and estimated power for the permutation tests in the Voluntary Contribution mechanism. Stars, Xs, and Os represent test results for those tests with a power of at least 0.45. Star represents rejection of H_0 at the 5% level, X represents rejection at the 10% level, and O represents no rejection of H_0. Note that this mechanism has no test with power $\geq 45\%$.

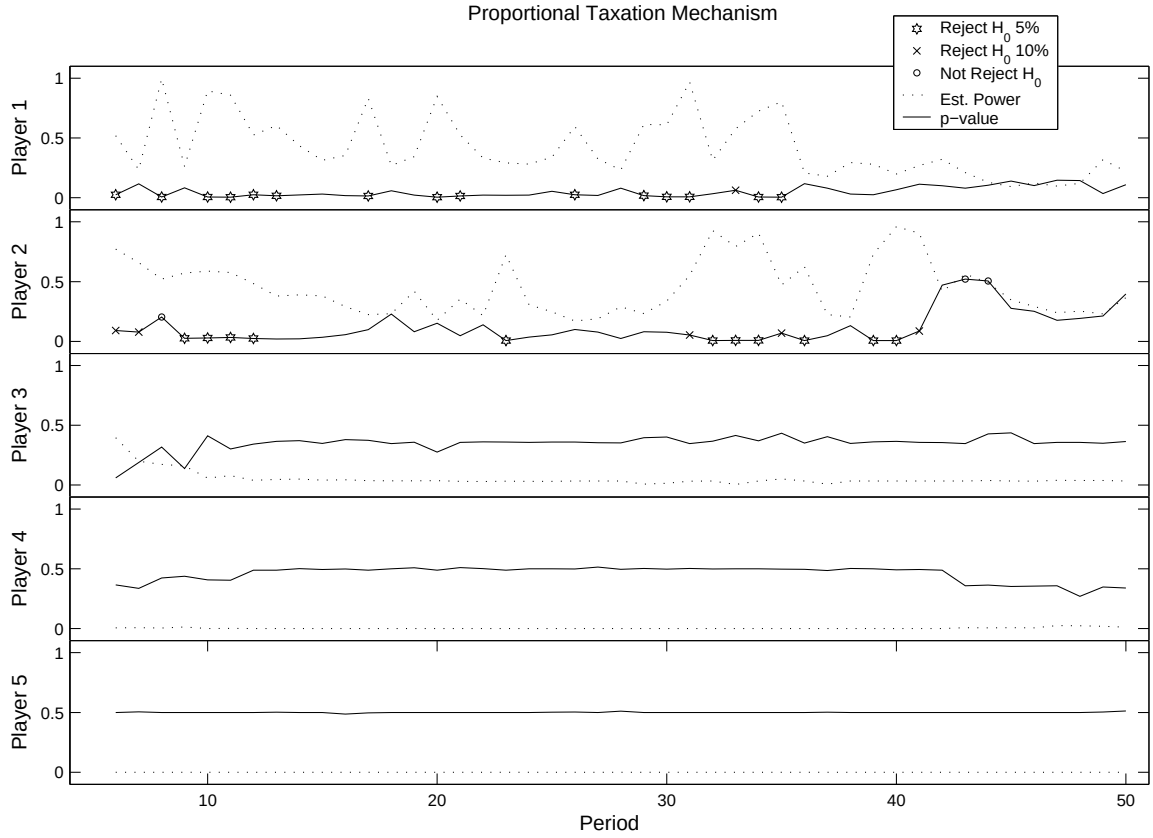


Figure 2.7: p-values and estimated power for the permutation tests in the Proportional Tax mechanism. Stars, Xs, and Os represent test results for those tests with a power of at least 0.45. Star represents rejection of H_0 at the 5% level, X represents rejection at the 10% level, and O represents no rejection of H_0.

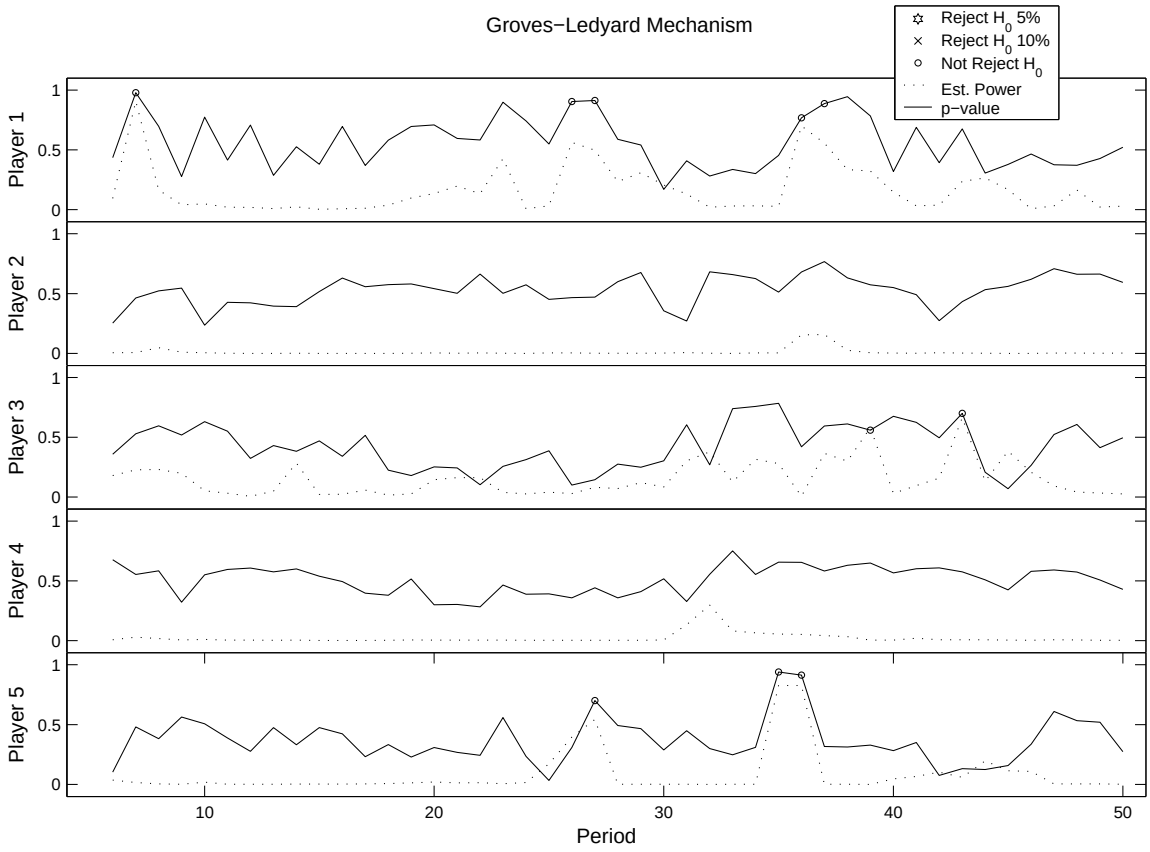


Figure 2.8: p-values and estimated power for the permutation tests in the Groves-Ledyard mechanism. Stars, Xs, and Os represent test results for those tests with a power of at least 0.45. Star represents rejection of H_0 at the 5% level, X represents rejection at the 10% level, and O represents no rejection of H_0.

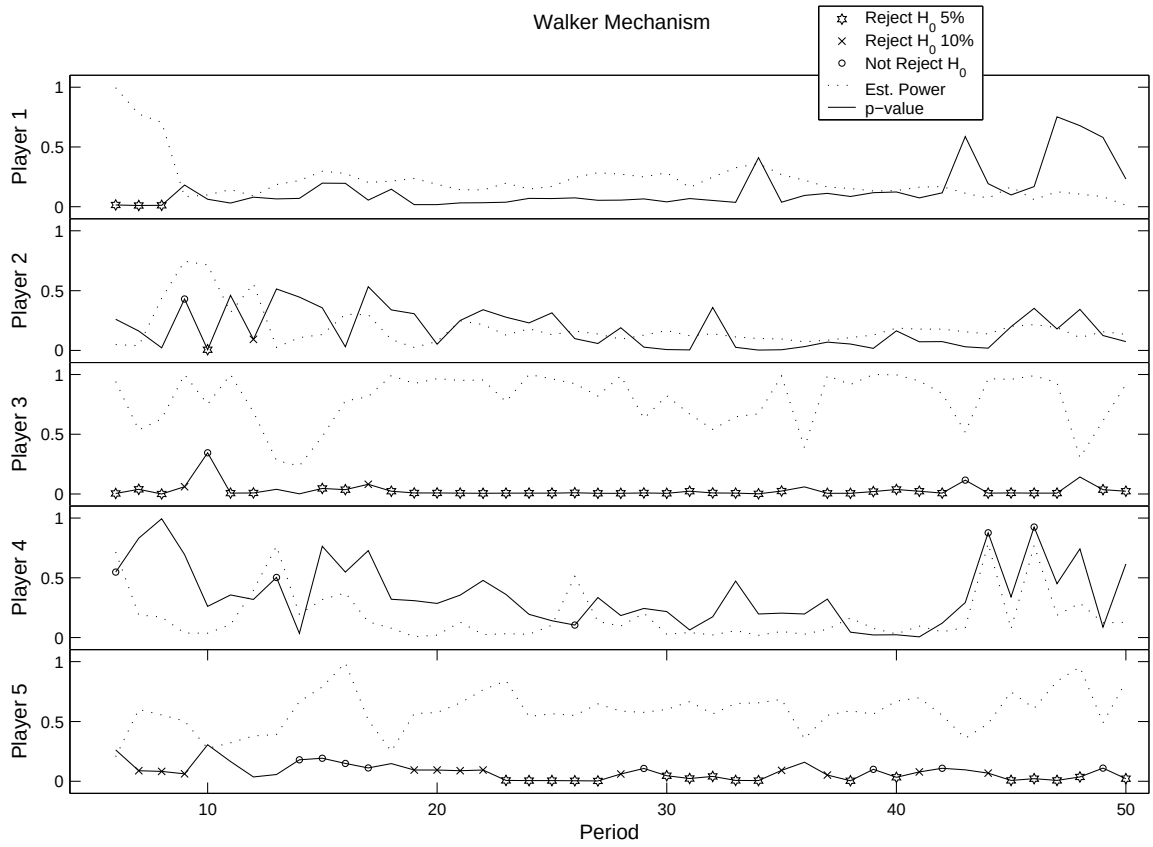


Figure 2.9: p-values and estimated power for the permutation tests in the Walker mechanism. Stars, Xs, and Os represent test results for those tests with a power of at least 0.45. Star represents rejection of H_0 at the 5% level, X represents rejection at the 10% level, and O represents no rejection of H_0.

sufficient power (out of 175 possible) do favor the Nash equilibrium model, though the test power being well below 89.4% in all ten tests prevents conclusive rejection of H_A. The Walker mechanism (Figure 2.9) provides the most testing power due to the lack of convergence of the data. Of the 41 tests due to player 3, only two fail to reject the null hypothesis at the 10% level, while 37 tests reject H_0 at the 5% level. Of the 37 tests due to player 5, eight fail to reject H_0 at the 10% level, while 12 tests reject H_0 at the 10% level and 17 tests reject at the 5% level. The p-values of all 37 tests are below 0.20. These rejections are scattered evenly throughout the session, indicating no particular pattern over time. Of the other three player types, players 1 and 2 have low p-values on average, with five rejections of H_0 and only one failure to reject. Player 4 shows mixed support overall, and in the few tests with sufficient power, shows fairly strong support for equilibrium behavior. This can also be seen in Figure 2.4. ■

The significance of this result lies in its implications for implementation in a repeated interaction setting, where the assumption that agents play the stage game equilibrium is less accurate than a simple best response behavioral assumption. Mechanisms constructed under the assumption of equilibrium behavior may fail to implement the desired outcome due to instability in the behavioral process. Although the best response model does not provide a complete description of human behavior, a mechanism designer who assumes this simple dynamic will be able to more 'accurately implement the desired outcomes than a designer who assumes static equilibrium behavior.

2.5.4 Best Response in the cVCG Mechanism

In this direct mechanism, agents announce both $\hat{a}_i$ and $\hat{b}_i$. The decisions of subjects can be grouped into three categories: 'full' revelation ($\hat{a}_i = a_i$ *and* $\hat{b}_i = b_i$, which is the dominant strategy), 'partial' revelation ($\hat{a}_i = a_i$ *or* $\hat{b}_i = b_i$), and no revelation. The equilibrium model predicts that all messages will be of the first type. The best response model predicts that all three types of messages are possible, but messages that aren't fully revealing must lie along the best response surface. The following subsections look at (a) what percentage of messages are fully revealing, and (b) whether non-revealing messages are scattered randomly or are centered around the line of best responses.

2.5.5 Frequency of Revelation

Previous experimental tests of dominant strategy mechanisms with a weak dominant strategy indicate that around half of all subjects play their dominant strategy. Table 2.4 indicates that this result holds true in the current study. Rates of both full and partial revelation in the cVCG mechanism are reported. Note a message that varies from truth-telling by any amount is encoded as non-revealing.

Result 2.3 *Truthful revelation in the cVCG mechanism is observed in the majority of decisions. This frequency increases in the final periods.*

Support. Refer to Table 2.4. On average, 54% of all observed messages are full revelation strategies, with the frequency increasing to 59% in the final 10 periods. Average partial revelation rises from 64% to 72% over the last 10 periods. Half of

Session	Periods	Player 1		Player 2		Player 3		Player 4		Player 5		Average	
		Full	Partial	Full	Partial	Full	Partial	Full	Partial	Full	Partial	Full	Partial
Session 1	All 50	0.82	0.94	0.76	0.78	0.00	0.04	0.88	0.90	0.62	0.76	0.62	0.68
	Last 10	1.00	1.00	1.00	1.00	0.00	0.00	1.00	1.00	1.00	1.00	0.80	0.80
Session 2	All 50	0.92	1.00	0.28	0.30	1.00	1.00	0.00	0.04	0.88	0.94	0.62	0.66
	Last 10	1.00	1.00	0.60	0.60	1.00	1.00	0.00	0.00	0.70	0.70	0.66	0.66
Session 3	All 50	0.40	0.80	0.04	0.06	1.00	1.00	0.48	0.48	0.70	0.72	0.52	0.61
	Last 10	0.00	1.00	0.00	0.00	1.00	1.00	0.50	0.50	0.90	0.90	0.48	0.68
Session 4	All 50	0.00	0.68	0.94	0.98	0.92	1.00	0.04	0.08	0.04	0.34	0.39	0.62
	Last 10	0.00	1.00	1.00	1.00	1.00	1.00	0.00	0.10	0.00	0.60	0.40	0.74
Average	All 50	0.54	0.86	0.51	0.53	0.73	0.76	0.35	0.38	0.56	0.69	0.54	0.64
	Last 10	0.50	1.00	0.65	0.65	0.75	0.75	0.38	0.40	0.65	0.80	0.59	0.72

Table 2.4: Frequency of revelation of both parameters ('Full' revelation) and of only one parameter ('Partial' revelation) by each subject in the cVCG mechanism for all 50 periods and for the last 10 periods.

the twenty subjects fully reveal in at least nine of the last ten periods, and twelve at least partially reveal. Three subjects never choose full revelation, and an additional three subjects reveal fully only twice. Every subject reveals partially at least twice over the course of the experiment. ■

Analysis of individual data also reveals that in 98% of the cases where a subject only partially reveals, it is the $\hat{b}_i$ term that is misrepresented. This is likely due to the fact that altering the linear term has a more transparent effect on payoffs than the quadratic term.

2.5.6 Misrevelation & Weakly Dominated Best Responses

Recall from Section 2.4.5 that given $\hat{\boldsymbol{\theta}}_{-i}$, agent i has a line of best responses through $\boldsymbol{\theta}_i$ that are payoff equivalent to truth-telling. The slope of this line depends on $\hat{\boldsymbol{\theta}}_{-i}$, so the best response line is sensitive to a player's prediction about the strategies of the others. Since the k-period best response model provides a point prediction of $\hat{\boldsymbol{\theta}}_{-i}$, it is easily testable in this framework. In particular, the model cannot be rejected if misrevelation messages are centered around the particular best response line suggested by the k-period average prediction.

A convenient method for analyzing the data is to convert each two-dimensional message $(\hat{a}_i^t, \hat{b}_i^t)$ into polar coordinates $(\hat{\phi}_i^t, \hat{r}_i^t)$ with the origin at the truthful revelation point (a_i, b_i). Here, $\hat{\phi}_i^t$ represents the angle from (a_i, b_i) to $(\hat{a}_i^t, \hat{b}_i^t)$ and $\hat{r}_i^t$ represents the distance between these points. Fully revealing observations are not included in this analysis since they are consistent with both best response and equilibrium play.

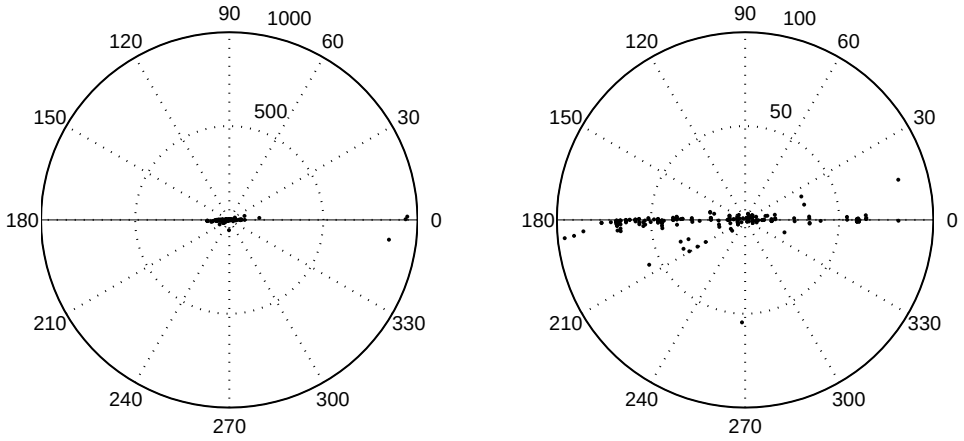


Figure 2.10: Polar-coordinate representation of the cVCG data with the origin at the truth-telling point and the horizontal axis corresponding to the 5-period best response surface. Two different scalings of the same graph are presented.

For each period $t > k$ and player i, the k-period best response model identifies a particular angle $\phi_i^B\left(\bar{\boldsymbol{\theta}}_{-i}^{t,k};\boldsymbol{\theta}_i\right)$ such that any announcement $\hat{\boldsymbol{\theta}}_i^t$ with $\hat{\phi}_i^t = \phi_i^B$ is a best response to the average of the previous k values of $\hat{\boldsymbol{\theta}}_{-i}$. Figure 2.10 graphs $(\hat{\phi}_i^t - \phi_i^B, \hat{r}_i^t)$ in polar coordinates, where ϕ_i^B is the angle generated from the 5-period model. In this figure, the origin ($\hat{r}_i^t = 0$) represents a full revelation announcement and the horizontal axis ($\hat{\phi}_i^t - \phi_i^B$) represents a message consistent with the 5-period best response model. If subjects play the dominant strategy equilibrium, the data should scatter evenly around the origin. If subjects follow the 5-period best response model, the data should scatter around the horizontal axis.

Unfortunately, the removal of all full-revelation observations reduces the average sample size to less than two observations per period – not enough to perform a statistical test. Qualitatively, the evidence strongly supports the 5-period best response model. Half of all non-equilibrium observations are within 1.3 degrees of the best

response line and 81% are within 10 degrees of the prediction. This analysis includes all partial revelation observations for which $\hat{\phi}_i^t$ is necessarily a multiple of $\pi/2$. After removing these observations, just over half of the remaining data are within 0.83 degrees of the best response prediction, and 79% are within 10 degrees.

Figure 2.11 shows the time-series representations of $\hat{r}_i^t$ and $\hat{\phi}_i^t - \phi_i^B\left(\bar{\boldsymbol{\theta}}_{-i}^{t,k}; \boldsymbol{\theta}_i\right)$ for each player type in the cVCG mechanism. The 95% confidence intervals are again formed by the bias-corrected and accelerated bootstrapping method with 2,000 draws. The average distance from truth-telling is frequently large, highly variable, and does not converge toward zero for three of the five player types. The graphs of $\hat{\phi}_i^t - \phi_i^B(\bar{\boldsymbol{\theta}}_{-i}^t; \boldsymbol{\theta}_i)$ across time show that the off-equilibrium data are centered at or near the best response prediction, with more stability in later periods. Again, small sample sizes prevent clean statistical analyses.

The tendency for the angular deviation to be slightly positive by about six degrees (visible in both figures) arises from the partial revelation observations. Around 87% of the best response lines are between 83° and 85°, while 20% of all off-equilibrium observations are partial revelation strategies located at 90°.

2.5.7 Testing Theoretical Predictions of the Model

In Section 2.2, various theoretical properties of the k-period average best response model are derived. Each of these may be tested empirically to confirm that the important implications of this behavioral assumption are observed in the laboratory.

In the cVCG mechanism, the best response line for each player is characterized by

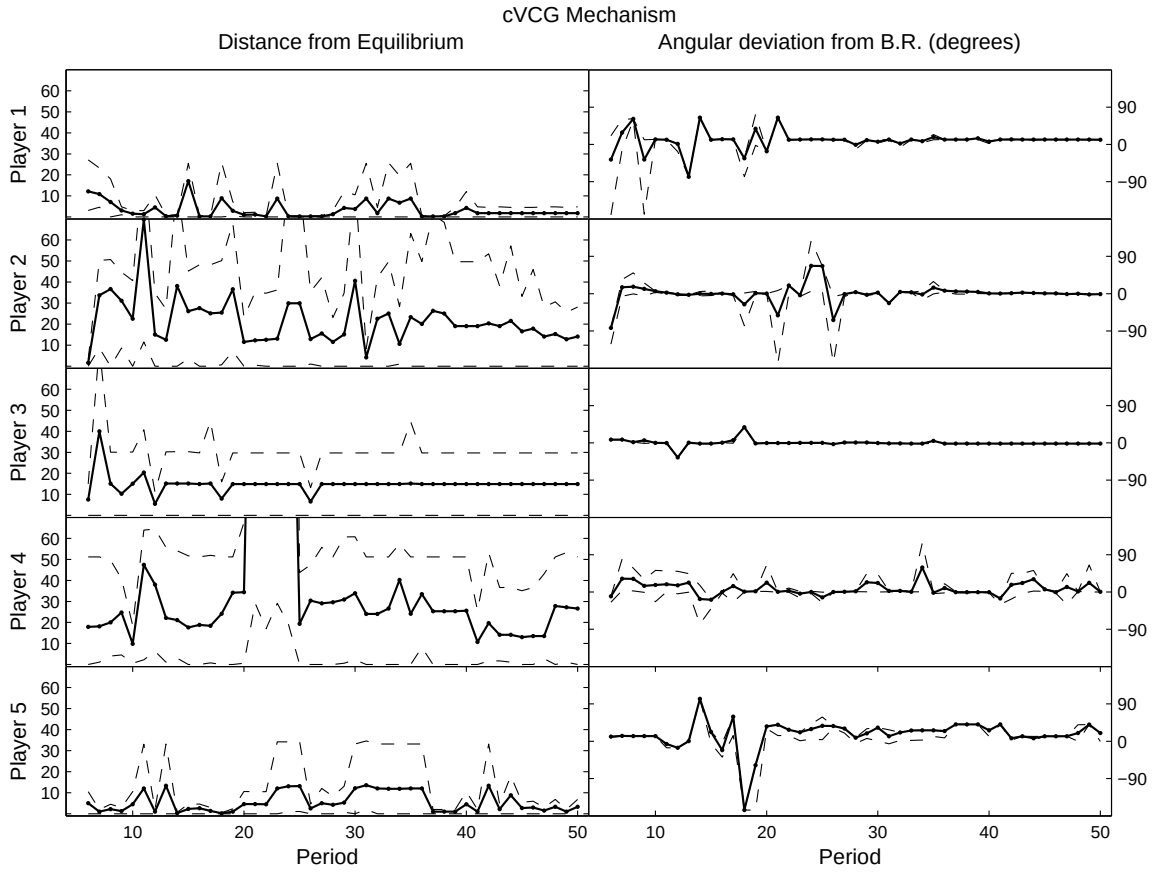


Figure 2.11: Time series of average distance from equilibrium and angular deviation from the best response line for each player type in the cVCG mechanism with 95% confidence intervals.

equation (2.4). Given $\hat{\boldsymbol{\theta}}_{-i}$, each of these is a single equation of two variables, so the set of Nash equilibria of the game is the set of solutions to a system of n equations with $2n$ variables. Clearly, there exists a large number of such solutions. An observed strategy profile $\hat{\boldsymbol{\theta}}$ is an ε-equilibrium if, for each $i \in \mathcal{I}$, the loss in utility between announcing $\boldsymbol{\theta}_i$ and $\hat{\boldsymbol{\theta}}_i$ is less than ε. This does not necessarily indicate that $\hat{\boldsymbol{\theta}}_i$ is close to $\boldsymbol{\theta}_i$ since $\hat{\boldsymbol{\theta}}_i$ may be a neighborhood of $\mathcal{B}_i\left(\hat{\boldsymbol{\theta}}_{-i}; \boldsymbol{\theta}_i\right)$ that is far from $\boldsymbol{\theta}_i$.

Result 2.4 *Weakly dominated ε-Nash equilibria are observed, while the dominant strategy equilibrium is not.*

Support. Setting $\varepsilon = 1$, 30.5% of observed strategy profiles in the cVCG mechanism are weakly dominated ε-Nash equilibria.[29] At $\varepsilon = 5$, 67% of the profiles are ε-equilibria. Across the last 12 periods, ε-equilibria are observed 93.8% of the time for $\varepsilon = 5$. In the first session, subjects play a particular ε-equilibrium (for $\varepsilon = 1/2$) in each of the final 19 periods. In *none* of the 200 repetitions of the cVCG mechanism is the truth-telling dominant strategy equilibrium observed. ■

Beyond providing further support for a best response model of behavior, this result has greater implications: it suggests that elimination of weakly dominated strategies leads to the elimination of certain Nash equilibria that are observed in the laboratory. This equilibrium selection technique is consequently inappropriate as a realistic equilibrium selection algorithm. The following result indicates that elimination of *strictly* dominated strategies *is* consistent with observed behavior:

[29] Agents typically earn over 300 francs per period, so $\varepsilon = 1$ represents a deviation from optimality of less than 0.33%.

Result 2.5 *(Proposition 2.1) Messages quickly converge to, and do not significantly deviate from, strictly dominant strategies in any period* $t > 5$.

Support. Players 1, 3, 4, and 5 in the Voluntary Contribution mechanism are the only players with strictly dominant strategies. From Figure 2.1, it is clear that these players quickly converge to their equilibrium strategies. The bootstrapped confidence intervals must lie in the strategy space. Of the 180 confidence interval lower bounds after period five, only 15 are different from zero, and only one is greater than 0.1. In the last nine periods, the *upper bound* of the intervals are all no greater than 0.25, and no greater than 0.1 in the last four periods. ■

The following results indicate that convergence to, or repetition of, a message profile are often indicative of a Nash equilibrium, as predicted by the best response model of behavior.

Result 2.6 *(Proposition 2.2) If a strategy profile is observed in* 6 *consecutive periods, then it is most likely a Nash equilibrium.*

Support. In the non-cVCG mechanisms, there are 754 messages m_i^t such that $m_i^t = m_i^{t-1} = \cdots = m_i^{t-5}$. Of those, 74.8% are Nash equilibrium messages. 80.1% of such messages are within 1 unit of Nash equilibrium. In the cVCG mechanism, 45% of the 375 such messages are ε-equilibria with $\varepsilon = 1$. Setting $\varepsilon = 5$ increases the frequency to 82.1%. ■

Result 2.7 *(Proposition 2.4) If a sequence of strategy profiles converges to a point* q*, then* q *is most likely a Nash equilibrium strategy profile.*

Support. Of the 20 groups across the 5 mechanisms, only one played the same strategy profile in all of the last 10 periods, indicating convergence to a particular strategy profile; the first session of the cVCG mechanism converged to an ε-equilibrium (for $\varepsilon \geq 1/2$) in all of the final 19 periods. One group in the Proportional Tax mechanism played a particular non-equilibrium strategy in 15 of the final 25 periods, while another group played the Nash equilibrium profile in 7 of the final 10 periods. ■

Finally, Propositions 2.7 and 2.9 are confirmed empirically by the convergence of the data to equilibrium in the Groves-Ledyard mechanism, which is supermodular *and* satisfies the dominant diagonal condition for the given parameters.

Overall, the above results indicate that the dynamic properties of observed behavior are generally in line with the theoretical properties of the k-period best response dynamic. The k-period dynamic is apparently a reasonably accurate yet tractable model for predicting repeated game behavior and convergence in these settings.

2.5.8 Efficiency & Public Good Levels

The ability to compare data across a fairly large number of mechanisms leads to the natural question of which mechanisms generate the most efficient outcomes. In fact, this study provides a unique opportunity to do so since no other experiment to date has tested as many processes side-by-side. It should be understood, however, that any experimental result may be very sensitive to changes in parameters. For example, it may be the case that if the efficiency of the dominant strategy equilibrium of the cVCG mechanism were lower, then subjects would play it less often, possibly reducing

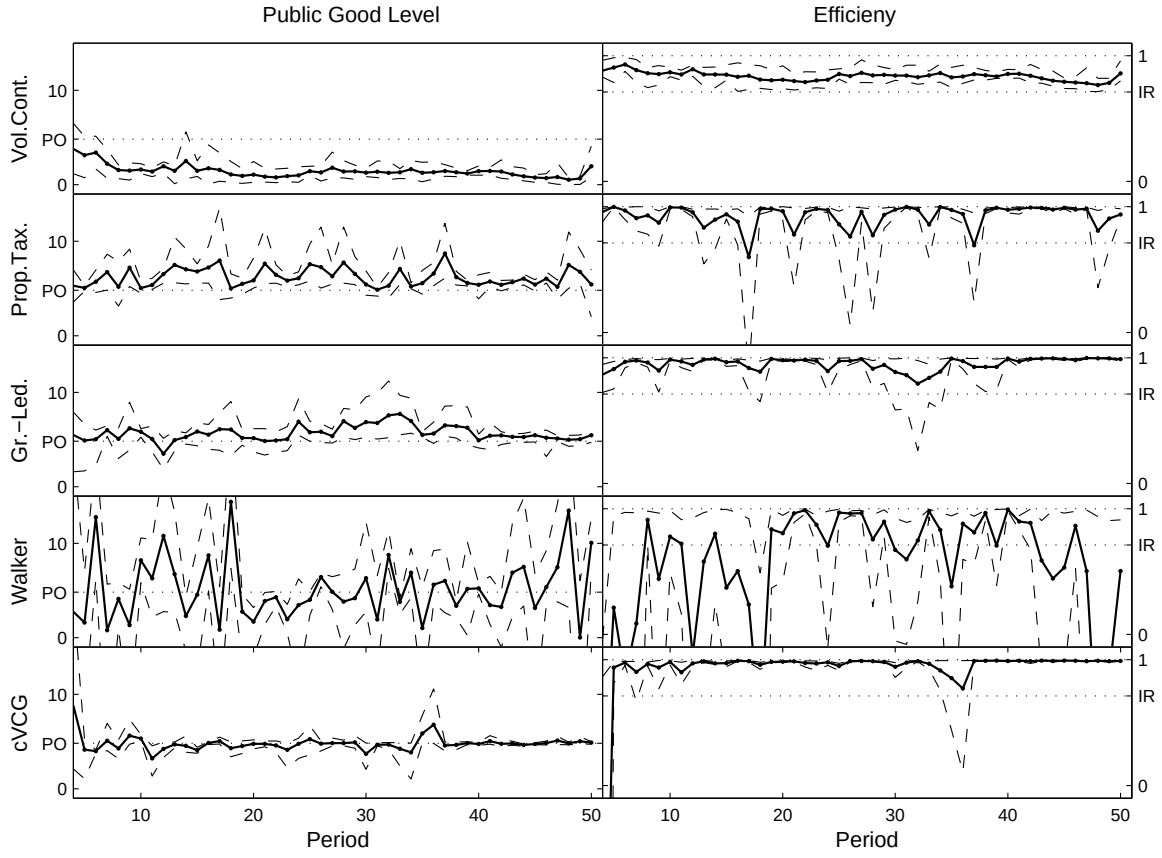


Figure 2.12: The average level of public good and realized efficiency for each mechanism in each period along with 95% confidence intervals. PO represents the Pareto Optimal level of the public good (4.8095) and IR represents the efficiency of the initial endowments (71.18%.)

realized efficiencies even lower. With this note of caution, the average public good level and realized efficiency are presented in Figure 2.12 along with 95% bootstrap confidence intervals.

Result 2.8 *For the given parameters, the average public good* levels *are closest to the Pareto optimal level in the cVCG mechanism, followed by the Groves-Ledyard mechanism. Overall efficiency in these two mechanism is also higher than the oth-*

ers, with the Walker mechanism often resulting in efficiency below that of the initial endowment.

Support. The average public good level is not significantly different from the Pareto optimum ($y^{\mathcal{P}} = 4.8095$) in 43 of the 50 periods for the cVCG mechanism, and 35 of 50 periods in both the Groves-Ledyard and Walker mechanisms. However, the realized public good level in the Walker mechanism is highly variable. The average public good level in the cVCG mechanism is not significantly different from the Pareto optimum in 22 of the final 25 periods, whereas the same is true in only 15 of the final 25 periods for the Groves-Ledyard mechanism. As predicted, the Voluntary Contribution mechanism significantly under-provides, and the Proportional Tax mechanism over-provides the public good. The cVCG mechanism is the most efficient, followed by the Groves-Ledyard mechanism. The average efficiency of the Walker mechanism is not significantly greater than the efficiency of the initial endowments (71.18%) in 38 of the 50 periods, and is significantly lower in four periods. This trend does not disappear across time. ■

Recall from Section 2.4.5 that there exists a large set of Nash equilibria of this mechanism that are necessarily outcome efficient, so the high level of efficiency realized by the cVCG mechanism is due to best response behavior and not preference revelation. Also, the fact that the Walker mechanism often realizes efficiencies at or below the initial endowment is particularly surprising, given that it is the only outcome-efficient mechanism tested whose equilibrium guarantees outcomes that Pareto dominate the endowment.

2.5.9 Open Questions

The fundamental difficulty of testing the efficiency of mechanisms in the laboratory lies in their sensitivity to parameter choice. Past research has focused on changing punishment parameters of the mechanisms, but it is unknown how behavior differs when preferences are varied within a mechanism. In particular, the role of the incentive-efficiency trade-off in guiding behavior is not known.

The current set of experiments makes use of the 'What-If Scenario Analyzer' tool that enables subjects to calculate hypothetical payoffs. This tool is provided as an alternative to the payoff tables often provided in experiments. The effect of this tool on dynamic behavior is not well understood. The version of the software in use for this study does not store data about what hypothetical scenarios subjects considered; only actual decisions are tracked. The possibility of studying hypothetical explorations is an exciting extension to this research that may provide additional understanding about the learning dynamics in use.

On the theoretical front, much work remains with respect to dynamics in public goods mechanisms. The literature is far behind that of market dynamics for private goods, where stability has been extensively analyzed for several decades. Kim [56] shows that mechanisms for Nash implementing Lindahl allocations must be locally unstable for some environment. However, restricting attention to quasilinear environments, a dynamically stable game form is introduced that implements the Lindahl correspondence. Similarly, de Tranqualye [27] and Vega-Redondo [98] introduce mechanisms that converge to the Lindahl allocation under Cournot best response

behavior. Some mechanisms such as that studied by Smith [94] use convergence of tatonnemont-like dynamics as a stopping rule in each period of the mechanism. However, there have been only limited attempts to seriously consider dynamic issues in the theoretical mechanism design literature.

2.6 Conclusion

Motivated by the observation that many of the results of previous experimental studies are consistent with a simple best response dynamic model, this chapter experimentally compares five different public goods mechanisms in order to test this conjecture. In particular, dynamically stable and unstable Nash mechanisms are compared along with a weak dominant strategy mechanism whose best response properties provide an opportunity to distinguish between best response and equilibrium behavior. This latter mechanism, though tested in simpler forms in previous experiments, has never been tested in the laboratory with more than one preference parameter.

The results of these experiments support the best response behavioral conjecture, particularly as an alternative to the static equilibrium hypothesis. Strategies converge to Nash equilibria that are asymptotic attractors in a best response dynamic and diverge from equilibria that are not. In the weak dominant strategy mechanism, behavior tends to track a rotating best response line through the strategy space, implying that subjects who do not understand the undominated properties of truthful revelation instead seek a best response strategy, resulting in convergence to weakly dominated Nash equilibria. This result implies that elimination of weakly dominated

strategies is an inappropriate tool for game theoretic analysis.

Although outcomes and efficiency are sensitive to parameters, the continuous VCG mechanism performs well in both categories, as does the Groves-Ledyard mechanism. The Walker mechanism, due to its instability, generates efficiencies often below that of the initial endowment.

The implications for mechanism design are straightforward. Most theorists have ignored dynamic stability in designing mechanisms. The significant contribution of this chapter is that it bridges the behavioral hypotheses that have existed separately in dominant strategy and Nash equilibrium mechanism experiments. The finding that a 5-period average best response dynamic is a reasonably accurate behavioral model in all of these settings implies not only that dynamic behavior should be considered in theoretical research, but it also provides some guidance as to which behavioral models are appropriate. In particular, Nash implementation mechanisms should satisfy dynamic stability and dominant strategy mechanisms should satisfy the strict dominance property if either is to be considered for real-world use in a repeated interaction setting.

2.7 Appendix

Proof of Proposition 2.7. Let $\underline{\mathbf{m}} = \inf \mathcal{M}$ and $\overline{\mathbf{m}} = \sup \mathcal{M}$. Define $\underline{\mathcal{B}}(\mathbf{m}, \boldsymbol{\theta}) = \inf \mathcal{B}(\mathbf{m}, \boldsymbol{\theta})$ and $\overline{\mathcal{B}}(\mathbf{m}, \boldsymbol{\theta}) = \sup \mathcal{B}(\mathbf{m}, \boldsymbol{\theta})$ as the infimal and supremal best responses to a given message profile $\mathbf{m}$. Since the game is supermodular, $\underline{\mathbf{m}}$ and $\overline{\mathbf{m}}$ are finite elements of $\mathcal{M}$, and $\underline{\mathcal{B}}$ and $\overline{\mathcal{B}}$ are elements of $\mathcal{B}$ for all $\mathbf{m}$ and $\boldsymbol{\theta}$. Furthermore, $\underline{\mathcal{B}}$

and $\overline{\mathcal{B}}$ are non-decreasing functions of $\mathbf{m}$ and there exists a smallest Nash equilibrium $\underline{\mathcal{E}}(\boldsymbol{\theta})$ and a largest Nash equilibrium $\overline{\mathcal{E}}(\boldsymbol{\theta})$.

Consider the sequence $\{\underline{\mathbf{m}}^t\}_1^\infty$ where $\underline{\mathbf{m}}^t = \underline{\mathbf{m}}$ for $t = 1, \ldots, k$ and $\underline{\mathbf{m}}^t = \underline{\mathcal{B}}\left(\underline{\bar{\mathbf{m}}}^{t,k}, \boldsymbol{\theta}\right)$ for $t > k$. In order to establish by induction that the sequence is monotone increasing, assume that $\underline{\mathbf{m}}^t \geq \underline{\mathbf{m}}^s$ for some $t > k$ and all $s < t$. This is certainly true for $t = k+1$. Since $\underline{\mathbf{m}}^t \geq \underline{\mathbf{m}}^{t-k}$, then $\underline{\bar{\mathbf{m}}}^{t+1} \geq \underline{\bar{\mathbf{m}}}^t$. By monotonicity of $\underline{\mathcal{B}}$, it must be that $\underline{\mathbf{m}}^{t+1} \geq \underline{\mathbf{m}}^t$. This implies that $\underline{\mathbf{m}}^{t+1} \geq \underline{\mathbf{m}}^s$ for all $s < t$. By induction, it is established that $\{\underline{\mathbf{m}}^t\}_1^\infty$ is a monotone increasing sequence. Since $\overline{\mathbf{m}}$ is finite, $\{\underline{\mathbf{m}}^t\}_1^\infty$ must converge to some point $\underline{\mathbf{m}}^* \in \mathcal{M}$. By Proposition 2.4, $\underline{\mathbf{m}}^*$ is a Nash equilibrium profile.

Assume that for some $t > k$, $\underline{\mathbf{m}}^s \leq \underline{\mathcal{E}}(\boldsymbol{\theta})$ for all $s \leq t$, which is true for $t = k + 1$. Then $\underline{\bar{\mathbf{m}}}^t \leq \underline{\mathcal{E}}(\boldsymbol{\theta})$ and, by monotonicity of $\underline{\mathcal{B}}$, $\underline{\mathbf{m}}^t \leq \underline{\mathcal{B}}(\underline{\mathcal{E}}(\boldsymbol{\theta}), \boldsymbol{\theta}) \leq \underline{\mathcal{E}}(\boldsymbol{\theta})$. This implies that $\underline{\mathbf{m}}^s \leq \underline{\mathcal{E}}(\boldsymbol{\theta})$ for all $s \leq t + 1$. Therefore, the sequence $\{\underline{\mathbf{m}}^t\}_1^\infty$ is bounded above by $\underline{\mathcal{E}}(\boldsymbol{\theta})$.

Since $\{\underline{\mathbf{m}}^t\}_1^\infty$ converges to some equilibrium point, it must be that $\lim \underline{\mathbf{m}}^t = \underline{\mathcal{E}}(\boldsymbol{\theta})$. Similar induction arguments establish that the sequence $\{\overline{\mathbf{m}}^t\}_1^\infty$ of k-period average best responses starting from $\overline{\mathbf{m}}$ must converge to $\overline{\mathcal{E}}(\boldsymbol{\theta})$.

Now consider any arbitrary sequence $\{\mathbf{m}^t\}_1^\infty$. If $\underline{\mathbf{m}}^s \leq \mathbf{m}^s \leq \overline{\mathbf{m}}^s$ for all s less than some $t > k$, then by monotonicity of $\underline{\mathcal{B}}$ and $\overline{\mathcal{B}}$, it must be that $\underline{\mathbf{m}}^t \leq \mathbf{m}^t \leq \overline{\mathbf{m}}^t$. Since this hypothesis is true for $t = k + 1$, induction implies that $\underline{\mathbf{m}}^t \leq \mathbf{m}^t \leq \overline{\mathbf{m}}^t$ for all t. These bounds establish the result in the limit. ■

Proof of Proposition 2.9. Recall that $\mathcal{M}_i = [0, \infty)$ and, for each $\boldsymbol{\theta}$ in some $\Theta_0 \subset \Theta$,

$$\lim_{m_i \to +\infty} \left| \frac{\partial u_i}{\partial m_i} (\mathbf{m}) \right| = +\infty$$

and $[\partial^2 u_i / \partial m_i \partial m_j]_{i,j=1}^n$ satisfies diagonal dominance on a set $\Theta_0 \subseteq \Theta$.

Fix $\boldsymbol{\theta} \in \Theta_0$. Gabay & Moulin [38, Theorem 4.1] show that there must exist an unique Nash equilibrium $\mathbf{m}^* (\boldsymbol{\theta})$ and that diagonal dominance implies $\mathcal{B} (\mathbf{m}, \boldsymbol{\theta})$ is single-valued and strictly non-expansive in the sup-norm, so that for all $\mathbf{m}, \mathbf{m}' \in \mathcal{M}$,

$$\left\| \mathcal{B} (\mathbf{m}, \boldsymbol{\theta}) - \mathcal{B} (\mathbf{m}', \boldsymbol{\theta}) \right\|_\infty < \left\| \mathbf{m} - \mathbf{m}' \right\|_\infty , \tag{2.6}$$

where $\|\mathbf{m}\|_\infty = \sup_i |m_i|$. If $\{\mathbf{m}^t\}_1^\infty$ is consistent with the k-period dynamic, then by (2.6),

$$\begin{aligned} \left\| \mathbf{m}^t - \mathbf{m}^* (\boldsymbol{\theta}) \right\|_\infty &< \left\| \bar{\mathbf{m}}^t - \mathbf{m}^* (\boldsymbol{\theta}) \right\|_\infty \\ &= \left\| \frac{1}{k} \sum_{s=1}^{k} \left(\mathbf{m}^{t-s} - \mathbf{m}^* (\boldsymbol{\theta}) \right) \right\|_\infty \\ &\leq \frac{1}{k} \sum_{s=1}^{k} \left\| \mathbf{m}^{t-s} - \mathbf{m}^* (\boldsymbol{\theta}) \right\|_\infty \\ &\leq \sup_{1 \leq s \leq k} \left\| \mathbf{m}^{t-s} - \mathbf{m}^* (\boldsymbol{\theta}) \right\|_\infty \end{aligned}$$

for every $t > k$. If

$$\lim_{t \to \infty} \sup_{1 \leq s \leq k} \left\| \mathbf{m}^{t-s} - \mathbf{m}^* (\boldsymbol{\theta}) \right\|_\infty = 0, \tag{2.7}$$

then convergence of $\{\mathbf{m}^t\}_1^\infty$ to $\mathbf{m}^* (\boldsymbol{\theta})$ is established.

Take any $\mathbf{q} \in \mathcal{M}^k \subseteq \mathbb{R}^{nk}$. For any such $\mathbf{q}$, there exists a unique sequence $\{\mathbf{m}^t\}_1^\infty$ consistent with the k-period dynamic such that $\left(\mathbf{m}^1, \ldots, \mathbf{m}^k\right) = \mathbf{q}$. Define $G\left(\mathbf{q}, \boldsymbol{\theta}\right) = \left(\mathbf{m}^{k+1}, \ldots, \mathbf{m}^{2k}\right)$ to be the next k terms of the k-period dynamic, all of which can be uniquely determined given $\mathbf{q}$ and $\boldsymbol{\theta}$. Iterated application of G generates a sequence $\{\mathbf{q}^r\}_1^\infty$ where $\mathbf{q}^1 = \mathbf{q}$ and $\mathbf{q}^{r+1} = G\left(\mathbf{q}^r, \boldsymbol{\theta}\right)$. Define $\mathbf{q}^*\left(\boldsymbol{\theta}\right) = \left(\mathbf{m}^*\left(\boldsymbol{\theta}\right), \ldots, \mathbf{m}^*\left(\boldsymbol{\theta}\right)\right)$ and note that $\mathbf{q}^*\left(\boldsymbol{\theta}\right)$ is a fixed point of $G\left(\cdot, \boldsymbol{\theta}\right)$. Condition (2.7) can now be rewritten as $\lim_{r\to\infty} \left\|\mathbf{q}^r - \mathbf{q}^*\left(\boldsymbol{\theta}\right)\right\|_\infty = 0$.

The following demonstrates that G is strictly non-expansive. Pick any points $\mathbf{q} = \left(\mathbf{m}^1, \ldots, \mathbf{m}^k\right)$ and $\hat{\mathbf{q}} = \left(\hat{\mathbf{m}}^1, \ldots, \hat{\mathbf{m}}^k\right)$. If $G\left(\mathbf{q}, \boldsymbol{\theta}\right) = \left(\mathbf{m}^{k+1}, \ldots, \mathbf{m}^{2k}\right)$ and $G\left(\hat{\mathbf{q}}, \boldsymbol{\theta}\right) = \left(\hat{\mathbf{m}}^{k+1}, \ldots, \hat{\mathbf{m}}^{2k}\right)$, then by (2.6),

$$\begin{aligned}
\left\|\mathbf{m}^{k+1} - \hat{\mathbf{m}}^{k+1}\right\|_\infty &< \left\|\frac{1}{k}\sum_{s=1}^{k}\left(\mathbf{m}^s - \hat{\mathbf{m}}^s\right)\right\|_\infty \\
&\leq \frac{1}{k}\sum_{s=1}^{\infty}\left\|\mathbf{m}^s - \hat{\mathbf{m}}^s\right\|_\infty \\
&\leq \sup_{1\leq s\leq k}\left\|\mathbf{m}^s - \hat{\mathbf{m}}^s\right\|_\infty .
\end{aligned}$$

Similarly,

$$\left\|\mathbf{m}^{k+2} - \hat{\mathbf{m}}^{k+2}\right\|_\infty < \sup_{2\leq s\leq k+1}\left\|\mathbf{m}^s - \hat{\mathbf{m}}^s\right\|_\infty .$$

By replacing $\left\|\mathbf{m}^{k+1} - \hat{\mathbf{m}}^{k+1}\right\|_\infty$ in the argument of the supremum with

$$\sup_{1\leq s\leq k}\left\|\mathbf{m}^s - \hat{\mathbf{m}}^s\right\|_\infty ,$$

the inequality becomes

$$\left\|\mathbf{m}^{k+2}-\hat{\mathbf{m}}^{k+2}\right\|_{\infty}<\max\left\{\sup_{2\leq s\leq k}\left\|\mathbf{m}^{s}-\hat{\mathbf{m}}^{s}\right\|_{\infty},\sup_{1\leq s\leq k}\left\|\mathbf{m}^{s}-\hat{\mathbf{m}}^{s}\right\|_{\infty}\right\}$$

$$=\sup_{1\leq s\leq k}\left\|\mathbf{m}^{s}-\hat{\mathbf{m}}^{s}\right\|_{\infty}.$$

Applying this reasoning to $\left\|\mathbf{m}^{k+t}-\hat{\mathbf{m}}^{k+t}\right\|_{\infty}$ for all $t=2,\ldots,k$ gives

$$\left\|\mathbf{m}^{k+t}-\hat{\mathbf{m}}^{k+t}\right\|_{\infty}<\sup_{1\leq s\leq k}\left\|\mathbf{m}^{s}-\hat{\mathbf{m}}^{s}\right\|_{\infty}\ \forall t=1,\ldots,k,$$

or

$$\sup_{1\leq s\leq k}\left\|\mathbf{m}^{k+s}-\hat{\mathbf{m}}^{k+s}\right\|_{\infty}<\sup_{1\leq s\leq k}\left\|\mathbf{m}^{s}-\hat{\mathbf{m}}^{s}\right\|_{\infty}.$$

This is equivalent to

$$\left\|G\left(\mathbf{q},\boldsymbol{\theta}\right)-G\left(\hat{\mathbf{q}},\boldsymbol{\theta}\right)\right\|_{\infty}<\left\|\mathbf{q}-\hat{\mathbf{q}}\right\|_{\infty},$$

so $G\left(\cdot,\boldsymbol{\theta}\right)$ is strictly non-expansive for each $\boldsymbol{\theta}\in\Theta_{0}$ and for all $\mathbf{q},\hat{\mathbf{q}}\in\mathcal{M}^{k}$. By an application of Edelstein's Theorem (see Ortega & Rheinbolt [78, p. 404],) the sequence $\{\mathbf{q}^{r}\}_{1}^{\infty}$ must converge to $\mathbf{q}^{*}\left(\boldsymbol{\theta}\right)$.[30] This implies that (2.7) holds, completing the proof. ■

[30]Edelstein's Theorem requires compactness. In this case, the set

$$\left\{\mathbf{m}\in M:\left\|\mathbf{m}-\mathbf{m}^{*}\left(\boldsymbol{\theta}\right)\right\|_{\infty}\leq\left\|\mathbf{m}^{1}-\mathbf{m}^{*}\left(\boldsymbol{\theta}\right)\right\|_{\infty}\right\}$$

is compact for each $\{\mathbf{m}^{t}\}_{1}^{\infty}$ and, by (2.6), all subsequent elements of the sequence lie in this compact set. By restricting attention to this set, Edelstein's Theorem applies.

Chapter 3

Equilibrium Participation in Public Goods Allocations

The previous chapter addressed certain concerns in mechanism design related to the assumption of equilibrium behavior and the robustness of predictions in repeated interaction scenarios. The current chapter takes a different tack, focusing instead on the enforceability of the outcomes selected by a particular mechanism. The key issues related to enforceability are the credible options of the social planner and the available options of the players. In economies with private goods, Hurwicz [48] assumes that the mechanism designer must allow the agents a 'no-trade' option, which leads naturally to the individual rationality constraint that agents must prefer the chosen allocation to their initial endowment. With public goods, exercising a no-trade option may allow an agent to consume some level of the public good produced by those who participate. Thus, Green & Laffont [41, p. 121] argue that individual rationality is instead founded on the ethical belief that each agent has a natural right to her endowment and the welfare its consumption would generate.

The current chapter reconsiders the mechanism design problem with public goods

when the mechanism designer must allow a no-trade option. The resulting constraint – called *equilibrium participation* – requires the mechanism to select an outcome such that every agent prefers to contribute their requested transfer payment rather than withhold it. If an agent withholds her transfer payment, then the level of the public good is reduced to that which can be feasibly produced with the remaining transfers.

In order to induce all agents to choose participation over non-participation, a mechanism can satisfy equilibrium participation by making those agents with the strongest free-riding incentive responsible for the largest share of the production inputs. This is demonstrated in example 3.5 of Section 3.2.4. However, if several agents have strong free-riding incentives, they cannot all be made responsible for the lion's share of production. This problem is exacerbated in larger economies. This is the intuition behind the two main results of this chapter: (1) there are many finite economies in which only the endowment satisfies equilibrium participation, and (2) as *any* classical public goods economy is replicated, the set of outcomes satisfying equilibrium participation converges to the endowment.

The negative results of this chapter imply that coercion is absolutely necessary for mechanisms to successfully implement desirable outcomes. If an agent opts out of the mechanism outcome, some punishment system must be in place so that the dissenting agent cannot free ride on the production of others. This can be obtained explicitly through fines and sanctions, or implicitly by threatening to produce nothing if any agent defects. If explicit coercion is unavailable and implicit threats incredible, then mechanism design cannot avoid the standard free-rider problem.

The next section reviews the relevant literature. The notation and key definition of the chapter are provided in Section 3.2. General properties of the set of allocations satisfying equilibrium participation are explored in Section 3.3, followed by an analysis of the constraint in classical, quasi-concave economies with convex technology in Section 3.4. The main result on convergence to the endowment in large economies is proven in Section 3.5. Concluding comments and open questions are discussed in Section 3.6.

3.1 Relation to Previous Literature

Several authors have tried with limited success to define a notion of the core that is appropriate in a public goods economy. Such definitions must make assumptions about the behavior of non-dissenting coalitions when some coalition blocks an allocation. In the original definition by Foley [36], only the dissenting coalition may produce the public good; non-dissenters withdraw their contributions to production. This maximizes the threat to dissenters and many allocations remain in the core.[1] [82] assumes that non-dissenting agents select levels of production that are 'rational' for themselves (under various meanings) and finds that the subsequent definition of the core may be empty.

Champsaur, Roberts & Rosenthal [15] define the φ-core as the allocations that remain unblocked when blocking coalitions are given the power to tax the remaining agents an amount up to φ, which depends on the *proposed* blocking allocation. If φ

[1] [74] shows that Foley's core does not converge to the set of Lindahl equilibria in large economies.

were a function of the *original* allocation, then this notion of blocking (for single-agent coalitions) could encompass the definition of equilibrium participation. Though the results for both definitions are similarly negative, they are logically independent.

Saijo [89] analyzes the mechanism design problem when the utility of autarkic production is used as a welfare lower bound instead of the utility of the endowment. His notion of *autarkic individual rationality* requires each agent's final utility level to be weakly greater than that which the agent could achieve in isolation with his endowment and access to the production technology. Whereas Ledyard & Roberts [62] demonstrate that the standard notion of individual rationality is incompatible with incentive compatibility among the class of Pareto optimal mechanisms, Saijo [89] shows that autarkic individual rationality is incompatible with incentive compatibility for *all* mechanisms, optimal or not.

Other authors have proposed various models of the outside options of agents in a mechanism design setting. The most general of these is Jackson & Palfrey [53], where an unspecified function maps from any given outcome to another (possibly identical) outcome. The necessary and sufficient conditions of Maskin [67] are then extended in a simple way to accommodate this 'reversion function.' This approach unifies several existing attempts to model renegotiation and participation in the outcomes of mechanisms in private goods settings, such as Ma, Moore & Turnbull [64], Maskin & Moore [68], and Jackson & Palfrey [52]. It also encompasses pubic goods models with an exogenous status quo outcome or mechanism, as in Perez-Nievas [81].

The issue of enforceability has been addressed in the literature on Bayesian mech-

anism design (where agents have a non-degenerate common knowledge prior belief over the set of possible preference profiles) through the means of an external 'budget breaker' who receives a large transfer from the agents when undesirable performance is observed. This concept was introduced by Holmstrom [47] as a way for managers to incentivize teams of agents. Eswaran & Kotwal [31] argue that such schemes create a strong incentive for the budget breaker to bribe a single agent to deviate. For example, consider a situation where a central planner uses Walker's [100] mechanism to determine the level of some public good. Assume that if some agent submits a transfer smaller than that required by the mechanism outcome, then the planner can credibly commit to giving all received transfers to some disinterested third party, rather than putting those funds into production or refunding them to the agents. If this third party receives some benefit from these transfers, then she has an incentive to bribe one agent in the economy to withhold his transfer. If agents in the economy expect that this budget breaker will offer such a bribe, then the mechanism outcome cannot be supported as an equilibrium because the agents will rationally expect that some agent will be bribed. In the current chapter, it is assumed that the use of such budget breakers is not admissable, either because no dissinterested agent can be found or because the incentive to bribe is sufficiently large so as to make this an ineffective enforcement device.[2]

Finally, it is worth noting that concepts such as dominant strategy incentive compatibility and ex-post equilibrium do not encompass the definition of equilibrium

[2]Alternatively, it could be assumed that the planner has a strong preference for efficiency, so that the use of a budget breaker is simply not credible off the equilibrium path.

participation. Although these concepts do require that the mechanism outcome be preferred by each individual to all other outcomes in the range of the mechanism, there is no guarantee that the allocation obtaining after an agent opts out is in the mechanism's range. Indeed, most 'standard' public goods mechanisms (such as the Groves-Ledyard, Walker, and cVCG mechanisms presented in Section 2.4) do not include the opt-out points in their range. Therefore, the fact that an allocation is selected as part of an equilibrium decision does not preclude the possibility that agents will later prefer to free-ride on the contributions of others.

3.2 Notation & Definitions

This chapter uses the following notational conventions:[3]

$\mathbb{R}$	The real line: $(-\infty, \infty)$
$\mathbb{R}_+$	The non-negative real line: $[0, \infty)$
$\mathbb{R}^n$, $\mathbb{R}^n_+$	The n-fold Cartesian products of $\mathbb{R}$ and $\mathbb{R}_+$, respectively

3.2.1 Environments

Consider the following environment with one private good and one public good:

[3] If $\mathbf{x}$ and $\mathbf{x}'$ are in $\mathbb{R}^n$, then $\mathbf{x} \geq \mathbf{x}' \Leftrightarrow x_i \geq x'_i$ for all i, $\mathbf{x} > \mathbf{x}' \Leftrightarrow \mathbf{x} \geq \mathbf{x}'$ and $x_i > x'_i$ for some i, and $\mathbf{x} \gg \mathbf{x}' \Leftrightarrow x_i > x'_i$ for all i.

$I \geq 2$	The number of individuals
$\mathcal{I} = \{1, \ldots, I\}$	The set of individuals, indexed by i
$\mathbf{x} \in \mathbb{R}_+^I$	An allocation of the private good; $\mathbf{x} = (x_1, \ldots, x_I)$
$y \in \mathbb{R}_+$	A level of the public good
$\mathbf{z} = (\mathbf{x}; y) \in \mathbb{R}_+^{I+1}$	An allocation
$\mathcal{Z} \subset \mathbb{R}_+^{I+1}$	The set of all possible allocations
$\omega \in \mathcal{Z}$	The initial endowment: $\omega_i > 0$ for $i \in \mathcal{I}$, $\omega_{I+1} = 0$
$\mathbf{t} = \omega - \mathbf{x}$	The transfers paid by the agents. $T = \sum_i t_i$, $T_{-i} = \sum_{j \neq i} t_j$
$\succeq_i$	The complete, transitive preference relation of i on $\mathcal{Z} \times \mathcal{Z}$
$\succ_i$	The strict preference relation of i
$u_i : \mathcal{Z} \to \mathbb{R}$	Utility representation of $\succeq_i$
$\mathcal{Y} \subseteq \mathbb{R}^2$	The set of production possibilities: $\mathcal{Y} \cap \mathbb{R}_+^2 = \{(0,0)\}$
	$\varphi \in \mathcal{Y}$ & $\varphi' \leq \varphi \Rightarrow \varphi' \in \mathcal{Y}$ (comprehensive), $\mathcal{Y}$ closed
$F : \mathbb{R}_+ \to \mathbb{R}_+$	The production function: $F(T) = \sup\{y : (-T, y) \in \mathcal{Y}\}$
$c : F(\mathbb{R}_+) \to \mathbb{R}_+$	The cost function: $c(y) = \inf\{T \geq 0 : (-T, y) \in \mathcal{Y}\}$
$\mathbf{e} = \left(\{\succeq_i\}_{i \in \mathcal{I}}, \mathcal{Y}, \omega\right)$	An economy with I agents
$\mathcal{E}_I$	The set of all economies with I agents

Given an economy $\mathbf{e}$, let $\mathcal{Z}(\mathbf{e}) \subseteq \mathcal{Z}$ be the set of *feasible* allocations of the form $\mathbf{z} = \omega + (-\mathbf{t}; y)$, where

$$y \geq 0$$

and $\mathbf{t} \in \mathbb{R}^I$ satisfies

$$\mathbf{t} \leqq \omega,$$

$$T \geq 0,$$

and

$$(-T; y) \in \mathcal{Y}.$$

A feasible allocation $(\mathbf{x}; y)$ is *balanced* if $y = F(T)$.

The following assumptions are used at various points in the chapter:

A1 (Monotonicity) If $(x_i', y') \geq (x_i, y)$, then $(\mathbf{x}'; y') \succeq_i (\mathbf{x}; y)$.

A2 (Convexity) If $\mathbf{z}' \succeq_i \mathbf{z}$, then $\alpha \mathbf{z}' + (1 - \alpha)\mathbf{z} \succeq_i \mathbf{z}$ for all $\alpha \in (0, 1)$.

A3 (Continuity) For every $\mathbf{z} \in \mathcal{Z}(\mathbf{e})$, $\{z' \in \mathcal{Z}(\mathbf{e}) : \mathbf{z}' \succeq_i \mathbf{z}\}$ and $\{\mathbf{z}' \in \mathcal{Z}(\mathbf{e}) : \mathbf{z}' \preceq_i \mathbf{z}\}$ are closed.

A4 (Increasing marginal cost) $\mathcal{Y}$ is convex.

A5 (Differentiable utility) Preferences $\succeq_i$ can be represented by a differentiable utility function u_i.

A6 (Differentiable cost) The function F is differentiable.

Denote the set of 'classical' economies satisfying A1 through A4 by $\mathcal{E}_I^C$. Let $\mathcal{E}_I^D$ denote the set of differentiable economies satisfying A1 through A6. Note that under A4 and A6, $c'(y) = 1/F'(T)$.

3.2.2 Mechanisms

The following defines a mechanism and its possible outcomes:

$\mathcal{S}_i$	The set of strategies of i: $\mathcal{S} = \prod_{\mathcal{I}} \mathcal{S}_i$
$\tau : \mathcal{S} \to \mathbb{R}^I$	Transfer function
$\eta : \mathcal{S} \to \mathbb{R}_+$	Outcome function
$\Gamma = (\mathcal{S}, \eta, \tau)$	A mechanism
$\mu_\Gamma(\mathbf{e})$	Equilibrium correspondence mapping Γ and $\mathbf{e}$ into subsets of $\mathcal{S}$
$\mathcal{O}^\mu_\Gamma(\mathbf{e})$	$= \{(\mathbf{x}; y) \in \mathcal{Z} : [\exists \mathbf{s} \in \mu_\Gamma(\mathbf{e})]\ \mathbf{x} = \omega - \tau(\mathbf{s})\ \&\ y = \eta(\mathbf{s})\}$
$\bar{\mathcal{O}}^\mu_\Gamma(\mathbf{e})$	$= \{(\mathbf{x}; y) \in \mathcal{O}^\mu_\Gamma(\mathbf{e}) : y = F(\sum_i (\omega_i - x_i))\}$

The sets $\mathcal{O}^\mu_\Gamma(\mathbf{e})$ and $\bar{\mathcal{O}}^\mu_\Gamma(\mathbf{e})$ represent the set of outcomes and balanced outcomes, respectively, of an economy $\mathbf{e}$.

Definition 3.1 *Γ is decisive under μ if, for all $\mathbf{e} \in \mathcal{E}_I$, $\mathcal{O}^\mu_\Gamma(\mathbf{e}) \neq \emptyset$.*

Definition 3.2 *Γ is feasible under μ if it is decisive under μ and, for all $\mathbf{e} \in \mathcal{E}_I$, $\mathcal{O}^\mu_\Gamma(\mathbf{e}) \subseteq \mathcal{Z}(\mathbf{e})$.*

Definition 3.3 *Γ is balanced under μ if it is feasible under μ and, for all $\mathbf{e} \in \mathcal{E}_I$, $\mathcal{O}^\mu_\Gamma(\mathbf{e}) = \bar{\mathcal{O}}^\mu_\Gamma(\mathbf{e})$.*

The set of *Pareto optimal* allocations for e is given by

$$\mathcal{PO}(\mathbf{e}) = \{\mathbf{z} \in \mathcal{Z}(\mathbf{e}) : [\nexists \mathbf{z}' \in \mathcal{Z}(\mathbf{e})]\, \mathbf{z}' \succ \mathbf{z}\}.$$

Definition 3.4 *Γ is efficient under μ if it is decisive under μ and, for all $\mathbf{e} \in \mathcal{E}_I$, $\mathcal{O}^\mu_\Gamma(\mathbf{e}) \subseteq \mathcal{PO}(\mathbf{e})$.*

If preferences are strictly monotonic, efficient mechanisms must be balanced.

3.2.3 Implementation

In general, if $\mathcal{G}$ is a *social choice correspondence (SCC)* mapping each economy $\mathbf{e}$ to a subset of the feasible allocations $\mathcal{Z}(\mathbf{e})$, then Γ *implements* $\mathcal{G}$ *under* μ if $\mathcal{O}_{\Gamma}^{\mu}(\mathbf{e}) \subseteq \mathcal{G}(\mathbf{e})$ for every $\mathbf{e}$ and Γ *fully implements* $\mathcal{G}$ *under* μ if $\mathcal{O}_{\Gamma}^{\mu}(\mathbf{e}) = \mathcal{G}(\mathbf{e})$ for every $\mathbf{e}$. For example, if $\mathcal{IR}_i(\mathbf{e}) = \{(\mathbf{x}; y) \in \mathcal{Z}(\mathbf{e}) : (\mathbf{x}; y) \succeq_i (\omega; 0)\}$, then $\mathcal{IR}(\mathbf{e}) = \bigcap_{\mathcal{I}} \mathcal{IR}_i(\mathbf{e})$ is the SCC that selects all points in the economy that are weakly preferred to the endowment by all individuals. If Γ implements $\mathcal{IR}(\mathbf{e})$ under μ, then all agents are made weakly better off by participating in Γ and playing a strategy in $\mu_{\Gamma}(\mathbf{e})$.

Hurwicz [48] and Ledyard & Roberts [62] have shown that no mechanism implements $\mathcal{PO}(\mathbf{e}) \cap \mathcal{IR}(\mathbf{e})$ in dominant strategies for private or public goods economies, respectively. Hurwicz [49] shows that if a mechanism implements $\mathcal{PO}(\mathbf{e}) \cap \mathcal{IR}(\mathbf{e})$ in Nash equilibrium, then $\mathcal{O}_{\Gamma}^{\mu}(\mathbf{e})$ is the set of Walrasian (or Lindahl) allocations.

3.2.4 The Participation Decision

Consider a situation in which agents in economy e participate in a mechanism Γ that is balanced and efficient under μ and receive the outcome $(\omega - \tau; \eta) \in \mathcal{O}_{\Gamma}^{\mu}(\mathbf{e})$. If each agent i has the freedom to either contribute τ_i or exercise a 'no-trade' option by withholding τ_i, then the mechanism outcome induces an I-player, two-strategy game. Assume that the final public goods level is the maximum feasible, given the contributions received. If all agents prefer to contribute τ_i over exercising their no-

trade option, then full participation is a Nash equilibrium of the induced participation game and the allocation $(\omega - \tau; \eta)$ will be fully realized.

Clearly, there may exist a conflict between the goal of the social planner and the opt-out incentives of the agents. This is clearly seen by the following example:

Example 3.5 *Let* $\mathcal{I} = \{1, 2\}$. *Define*

$$u_1(x_1, y) = x_1 + 21y - 2y^2$$

and

$$u_2(x_2, y) = x_2 + 77y - 9y^2.$$

Fix $\omega_i = 50$ *for each* i *and let* $F(T) = T/10$.

In this example, $\mathcal{PO}(\mathbf{e}) = \{(\mathbf{x}; y) : y = 4 \ \& \ t_1 + t_2 = 40\}$. At the optima, the marginal rate of substitution is 5 for both agents, so the consumers' Lindahl prices are equal. Suppose an efficient mechanism under μ selects the Lindahl solution $\tau = (20, 20)$ and $\eta = 4$. The induced participation game is given in panel (a) of Figure 3.1. Clearly, agent 1 has an incentive to withhold her requested transfer, resulting in a suboptimal outcome of $y = 2$ in equilibrium.

Now consider another efficient mechanism under μ that selects $\eta = 4$ and $\tau = (30, 10)$. In the induced participation game, shown in panel (b) of Figure 3.1, it is an equilibrium for both agents to participate. Agent 1 no longer has an incentive to opt out because her contribution is responsible for a larger share of the production.

Although this redistribution of production 'responsibility' is an effective trick to

$t_1 \setminus t_2$	20	0
20	$82, 194$	$64, 168$
0	$84, 148$	$50, 50$

(a)

$t_1 \setminus t_2$	10	0
30	$72, 204$	$65, 200$
0	$69, 108$	$50, 50$

(b)

Figure 3.1: The induced participation game for Example 3.5 from (a) the equal-price Lindahl allocation, and (b) an unequal-price optimal allocation.

offset free-riding incentives, feasibility constraints limit how many agents can have their tax burden sufficiently increased. Furthermore, some agents may prefer to always defect, regardless how much of the burden they must bear. These difficulties are key to the negative results of the chapter.

Consider the more general case of two players and a constant marginal cost. If an allocation z is proposed such that $t_i > 0$ for each i and $F(T) > 0$, then the allocation that obtains when agent 1 opts out is given by

$$\mathbf{z}^{(-1)} = \left((\omega_1, x_2) \, ; y^{(-1)} \right),$$

where

$$y^{(-1)} = F(t_2).$$

The opt-out point $\mathbf{z}^{(-2)}$ is similarly defined. Panel (a) of Figure 3.2 provides a graphical example of these points in the Kolm triangle diagram (Kolm [58]; see Thomson [95] for a detailed exposition.) For the proposal z to satisfy equilibrium participation, both agents must prefer z to their 'opt-out' points $\mathbf{z}^{(-i)}$, as in the figure.

In the case where $t_1 < 0$ while $t_2 > 0$, then $y^{(-2)} = 0$ since negative quantities

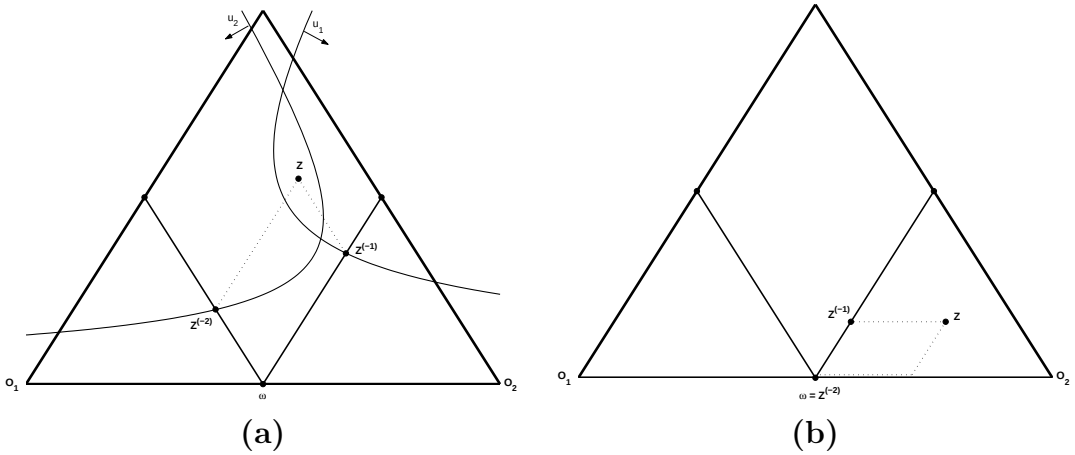


Figure 3.2: (a) The point $z \succeq_i z^{(-i)}$ for each $i \in 1, 2$, so it satisfies equilibrium participation. (b) The points $z^{(-i)}$ when t_1 is negative.

of the public good are not admissible. However, $y^{(-1)} = y$ since agent 1 is not asked to contribute any private good. In this case, it is assumed that the negative transfer rejected by agent 1 is either redistributed among the other agents or destroyed, rather than affecting the level of the public good.[4] Under A1, agent 1 will always prefer participation when $t_1 < 0$ and agent 2 will prefer participation only if $(\mathbf{x}; y) \in \mathcal{IR}_2(\mathbf{e})$. The case of a negative transfer is demonstrated graphically in panel (b) of Figure 3.2.

Generalizing the concepts of the two-player example provides the key definition of this chapter.

Definition 3.6 *For any $I = 1, 2, \ldots$ and any economy $\mathbf{e} \in \mathcal{E}_I$, a feasible allocation $(\mathbf{x}; y) \in \mathcal{Z}(\mathbf{e})$ such that $\mathbf{x} = \omega - t$ satisfies* equilibrium participation for agent i *(EP_i) if and only if*

$$(\mathbf{x}; y) \succeq_i \left(\mathbf{x}^{(-i)}; y^{(-i)}\right),$$

[4]Whether the transfer is redistributed or destroyed will not affect the i's participation decision since $\succeq_i$ depends only on x_i and y.

where

$$x_i^{(-i)} = \omega_i,$$

$$y^{(-i)} = \begin{cases} F(T_{-i}) & \text{if} \quad t_i \geq 0,\ T_{-i} \geq 0,\ \text{and}\ y \geq F(T_{-i}) \\ 0 & \text{if} \quad T_{-i} < 0 \\ y & \text{otherwise} \end{cases}, \tag{3.1}$$

and

$$\left(\mathbf{x}^{(-i)}; y^{(-i)}\right) \in \mathcal{Z}(\mathbf{e}).$$

The allocation $(\mathbf{x}; y) \in \mathcal{Z}(\mathbf{e})$ *satisfies* equilibrium participation (EP) *if and only if it satisfies* EP_i *for all* $i \in \mathcal{I}$.

There are four possible cases in this definition. When $t_i \geq 0$, $T_{-i} \geq 0$, and $y \geq F(T_{-i})$, removing agent i's transfer necessarily reduces production, but not to zero. If $T_{-i} < 0$, then $t_i > 0$ and removing i's transfer results in $y^{(-i)} = 0$. If $t_i < 0$ or $y < F(T_{-i})$, then y can be produced in the absence of i's transfer, so $y^{(-i)} = y$.

For any economy $\mathbf{e} \in \mathcal{E}_I$, let

$$\mathcal{EP}_i(\mathbf{e}) = \{\mathbf{z} \in \mathcal{Z}(\mathbf{e}) : \mathbf{z} \text{ satisfies } \mathrm{EP}_i\},$$

and define

$$\mathcal{EP}(\mathbf{e}) = \bigcap_{i \in \mathcal{I}} \mathcal{EP}_i(\mathbf{e}).$$

Referring back to the example of Figure 3.2, $\mathbf{z} \in \mathcal{EP}(\mathbf{e})$ in panel (a), but in panel

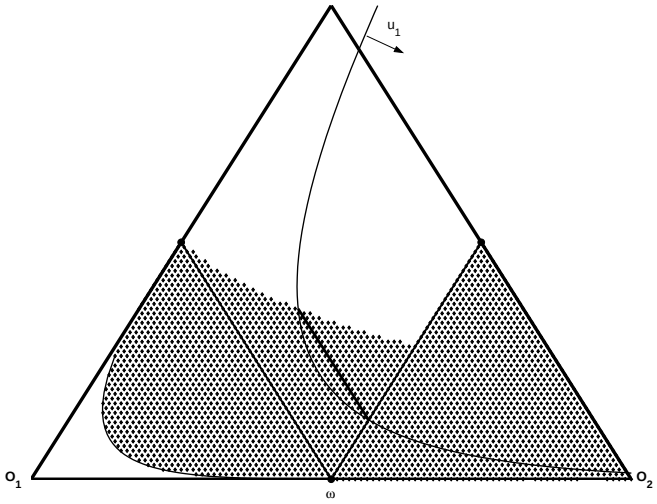


Figure 3.3: The set of balanced allocations satisfying equilibrium participation for agent 1.

(b), $\mathbf{z} \notin \mathcal{EP}_1(\mathbf{e})$, so $\mathbf{z} \notin \mathcal{EP}(\mathbf{e})$.

3.3 Properties of $\mathcal{EP}(\mathbf{e})$

The shaded region of Figure 3.3 demonstrates a typical equilibrium participation set for agent 1 in a two-agent classical economy. Note that $\mathcal{EP}(\mathbf{e})$ is closed and has a continuous boundary, but need not be convex. Clearly, $\mathcal{EP}(\mathbf{e})$ is non-empty for every $\mathbf{e} \in \mathcal{E}_I$ and every I since $(\omega; 0) \in \mathcal{EP}(\mathbf{e})$.

As an alternative to equilibrium participation, consider an environment in which agents can freely choose $t_i \in [0, \omega_i]$, resulting in $y = F(T)$. The set of Nash equilib-

rium allocations is given by

$$\mathcal{NE}(\mathbf{e}) = \left\{ \begin{array}{c} (\mathbf{x}^*; y^*) \in \mathcal{Z}(\mathbf{e}) : \mathbf{x}^* \leq \omega \text{ and} \\ [\forall i \in \mathcal{I}] \quad [\forall t_i' \geq 0] \quad (x_i^*, y^*) \succeq_i \left(\omega_i - t_i', F\left(T_{-i}^* + t_i'\right)\right) \end{array} \right\}.$$

The notion of equilibrium participation is now shown to be more stringent than the standard notion of individual rationality, but less restrictive than the Nash equilibrium requirement.

Proposition 3.7 *Under monotone increasing preferences (A3), all allocations satisfying equilibrium participation also satisfy individual rationality ($\mathcal{EP}(\mathbf{e}) \subseteq \mathcal{IR}(\mathbf{e})$.)*

Proof. Consider a point $(\mathbf{x}; y)$ such that $(x_i, y) \succeq_i \left(\omega_i, y^{(-i)}\right)$ for all $i \in \mathcal{I}$. Note that $y^{(-i)} \geq 0$ for each i, so A3 implies that $\left(\omega_i, y^{(-i)}\right) \succeq_i (\omega_i, 0)$. By transitivity, $(x_i, y) \succeq_i (\omega_i, 0)$ for every i, proving the result. ■

Proposition 3.8 *All Nash equilibria of the voluntary contributions game satisfy equilibrium participation ($\mathcal{NE}(\mathbf{e}) \subseteq \mathcal{EP}(\mathbf{e})$.)*

Proof. From any Nash equilibrium point, the 'opt-out' allocation for agent i in the participation game is simply $\left(\omega_i, F\left(T_{-i}^*\right)\right)$. Since the definition requires that $(x_i^*, y^*) \succeq_i \left(\omega_i, F\left(T_{-i}^*\right)\right)$ for all i by considering $t_i' = 0$, the point $(\mathbf{x}^*; y^*)$ must satisfy equilibrium participation. ■

In mechanism design with public goods, the most common goal is to implement $\mathcal{PO}(\mathbf{e})$. There exist several mechanisms whose Nash equilibria are guaranteed to be Pareto optimal when utility is transferable. However, if the outcomes of these

mechanism fail to satisfy equilibrium participation, then their desirable properties are of little use in environments where agents cannot be coerced to submit their transfers. The following class of examples shows the potential difficulty of finding points in $\mathcal{PO}(\mathbf{e}) \cap \mathcal{EP}(\mathbf{e})$.

Example 3.9 *Let $I \geq 2$. Define $u_i(x_i, y) = v_i(y) + x_i$, where each $v_i(y)$ is continuous and differentiable. Assume $F(T) = T/\kappa$, and let $v_i'(y) < \kappa$ for all $i \in \mathcal{I}$ and $y \geq 0$. Assume that there is a unique $y^o > 0$ such that $\sum_i v_i'(y) > \kappa$ for $y < y^o$ and $\sum_i v_i'(y) < \kappa$ for $y > y^o$. Finally, assume that $\sum_{j \neq i} \omega_j < \kappa y^o$ for each $i \in \mathcal{I}$.*[5]

In this example, no agent is willing to unilaterally fund any amount of the public good at any level and therefore refuses to contribute in any participation game. To see this, pick any allocation $(\mathbf{x}; y) \neq (\omega; 0)$, so $\mathbf{t} \neq \mathbf{0}$. If all agents participate in this allocation, then each agent i receives

$$u_i(x_i, y) = v_i(y) + \omega_i - t_i.$$

If i withholds her transfer, she receives

$$u_i\left(x_i^{(-i)}, y^{(-i)}\right) = v_i\left(y^{(-i)}\right) + \omega_i.$$

There must be some agent i with $t_i > 0$. If $y = 0$ or $T_{-i} \leq 0$, then $y^{(-i)} = 0$ and EP_i is

[5] *One such example is $\kappa = 1$ and*

$$v_i(y) = \begin{cases} \frac{3}{2I} y & \text{if} \quad y \leq 1 \\ \frac{1}{2I} y + \frac{1}{I} & \text{if} \quad y \geq 1 \end{cases}$$

for each i. Here, $y^o = 1$. The point of non-differentiability in v_i is of no consequence.

not satisfied. If $y > 0$ and $T_{-i} > 0$, but $y \leq F(T_{-i})$ then $y^{(-i)} = y$ and EP_i again fails. Therefore, consider the case where $y > 0$, $T_{-i} > 0$, and $y > F(T_{-i})$, so $y^{(-i)} = F(T_{-i})$. By withholding, agent i saves $t_i = \kappa\left(y - y^{(-i)}\right)$ in transfer payments. Her loss in value due to the reduction in public goods production is $v_i(y) - v_i\left(y^{(-i)}\right) = \int_{y^{(-i)}}^{y} v_i'(s)\,ds$, which is less than $\kappa\left(y - y^{(-i)}\right)$ since $v_i'(y) < \kappa$ for all y. Therefore, she will prefer to withhold her transfer regardless of t_i and the allocation will not satisfy equilibrium participation for agent i. In this economy, no allocation can satisfy EP_i for every i, so $\mathcal{EP}(\mathbf{e})$ is simply the endowment. This class of examples proves the following proposition:

Proposition 3.10 *For every $I \geq 2$, there exists economies $\mathbf{e}$ in $\mathcal{E}_I^C$ such that no allocation except the endowment satisfies equilibrium participation ($\mathcal{EP}(\mathbf{e}) = \{\omega\}$).*

The following shows that the notion of voluntary participation implicit in the definition of EP may preclude any optimal allocation from obtaining.

Proposition 3.11 *For every $I \geq 2$, there exists economies $\mathbf{e}$ in $\mathcal{E}_I^C$ in which no allocation $z \in \mathcal{Z}(\mathbf{e})$ can be selected such that the equilibrium of the resulting participation game is Pareto optimal.*

The proof of this result is simple. Any Pareto optimal allocation in the above class of examples must choose $y^o > 0$, from which any agent will defect. Furthermore, optimal allocations cannot obtain *after* an agent defects; if any one agent is consuming $x_i = \omega_i$, then $\sum_{j \neq i} \omega_j < \kappa y^o$ guarantees that y^o cannot be feasibly produced by the remaining agents.

Note that example 3.9 does not represent a knife-edge case. A wide range of economies fits its assumptions and a number of similar examples can be constructed. The key factor is marginal utilities must be smaller than marginal costs at all levels of y.

Since Proposition 3.11 indicates that EP is inconsistent with Pareto optimality, it is natural to ask whether there can exist *any* non-trivial mechanisms that satisfy this constraint.[6] In other words, is there a mechanism and a μ that implements $\mathcal{EP}(\mathbf{e})$ in μ? The results of Gibbard [40], Satterthwaite [91], K. Roberts [85] and Zhou [102] indicate that dominant strategy implementation of $\mathcal{EP}(\mathbf{e})$ is futile, even in classical economies. More positive results may be obtained when μ is weakened to the Nash equilibrium concept; it is simple to show that $\mathcal{EP}(\mathbf{e})$ satisfies Maskin's definition of monotonicity (see Maskin [67], giving the following result[7]):

Proposition 3.12 *The set of allocations satisfying equilibrium participation ($\mathcal{EP}(\mathbf{e})$) can be non-trivially implemented in Nash equilibrium when $I \geq 3$.*

The proof of this proposition for full implementation relies on Maskin's mechanism which is not a particularly 'natural' game form. Proposition 3.8 shows that $\mathcal{EP}(\mathbf{e})$ can be implemented by the voluntary contribution mechanism since $\mathcal{NE}(\mathbf{e}) \subseteq \mathcal{EP}(\mathbf{e})$. However, this mechanism does not fully implement $\mathcal{EP}(\mathbf{e})$. Note that in economies like those of Example 3.5, $\mathcal{EP}(\mathbf{e}) = \{\omega\}$, making implementation of $\mathcal{EP}(\mathbf{e})$ trivial.

[6] A non-trivial mechanism is defined as one that selects something other than the initial endowment in at least one environment.

[7] The other sufficient condition, 'no-veto power,' is trivially satisfied in economic environments such as this one.

3.4 Quasi-Concave Economies

3.4.1 Necessary and Sufficient Conditions

The additional structure gained by adding assumptions A1 through A6 allows for the derivation of separate necessary and sufficient conditions for an allocation to satisfy equilibrium participation. Although these conditions are not tight, they require only 'local' information about the gradients of utilities and derivative of the production function.

Proposition 3.13 *For any economy in $\mathcal{E}_I^D$, if equilibrium participation is satisfied at a point $(\mathbf{x}; y) = (\omega + \mathbf{t}; y)$, then*

$$\frac{\partial u_i\left(\omega_i; F\left(T_{-i}\right)\right)/\partial y}{\partial u_i\left(\omega_i; F\left(T_{-i}\right)\right)/\partial x_i} \geq c'\left(y^{(-i)}\right) \tag{3.2}$$

for all $i \in \mathcal{I}$ such that $t_i, T_{-i} \geq 0$ and $y \geq F\left(T_{-i}\right)$.

A similar condition is now shown to be sufficient for a point to satisfy equilibrium participation. Whereas the necessary condition compares the marginal rate of substitution to marginal costs at the drop-out point, the sufficient condition compares these quantities at the proposed allocation:

Proposition 3.14 *For any economy in $\mathcal{E}_I^D$, if a point $(\mathbf{x}; y) = (\omega + \mathbf{t}; y)$ satisfies*

$$\frac{\partial u_i\left(\mathbf{x}; y\right)/\partial y}{\partial u_i\left(\mathbf{x}; y\right)/\partial x_i} \geq c'\left(y\right) \tag{3.3}$$

for all i such that t_i,$T_{-i} \geq 0$ and $y \geq F(T_{-i})$ and

$$u_j(\mathbf{x}; y) \geq u_j(\omega; 0) \tag{3.4}$$

for all j such that $T_{-j} < 0$, then equilibrium participation is satisfied at $(\mathbf{x}; y)$.

Unlike the necessary condition, equation (3.4) implies that information about the utilities of some agents at both the suggested allocation and the endowment is needed. This may be undesirable from the standpoint of mechanism design since additional information is necessary to determine that the condition is met.[8] The following condition shows how equation (3.4) could be replaced by a stronger version of equation (3.3) to give a single condition sufficient for all agents that uses only information about preferences and costs at the selected allocation.

Proposition 3.15 *For any economy in $\mathcal{E}_I^D$, if a point $(\mathbf{x}; y) = (\omega + \mathbf{t}; y)$ satisfies*

$$\frac{\partial u_i(\mathbf{x}; y)/\partial y}{\partial u_i(\mathbf{x}; y)/\partial x_i} \geq \frac{t_i}{F(T)} \tag{3.5}$$

for all i, then equilibrium participation is satisfied at $(\mathbf{x}; y)$.

Figure 3.4 demonstrates the interpretation of these conditions. The quantity $(\partial u_i/\partial y)/(\partial u_i/\partial x_i)$ is the slope of the gradient of u_i, while c' is the slope of the normal to the production possibilities frontier. In the figure, F is reflected around the y-axis and horizontally shifted so that its graph represents the production pos-

[8] Of course, there could exist mechanisms whose outcomes satisfy Equilibrium Participation without satisfying this sufficient condition.

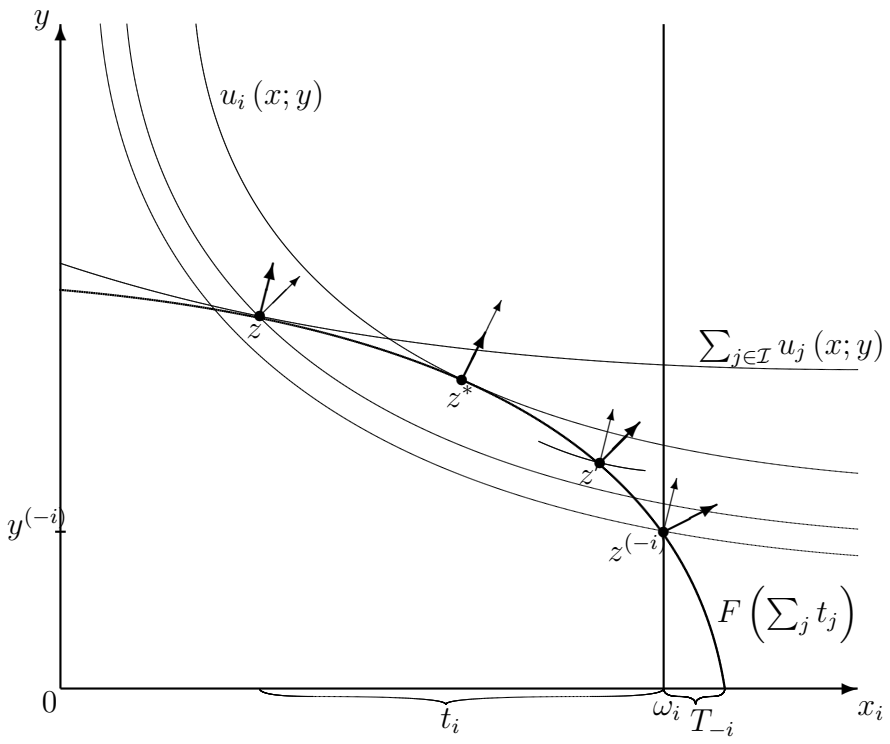


Figure 3.4: An example with quasi-concave utilities and convex production sets. z is Pareto optimal, z^* is i's most-preferred feasible allocation, and $z^{(-i)}$ is i's drop-out point. $z^{(-i)}$ satisfies the sufficient condition for EP. z^* and z' satisfy the sufficient condition.

sibilities set for agent i, given the endowments. If agent i withholds t_i, then the allocation $z^{(-i)}$ results. In this case, i will prefer the Pareto optimal point z to $z^{(-i)}$.

The necessary condition for equilibrium participation is satisfied in the figure since the gradient of utility has a steeper slope than the normal to F at $z^{(-i)}$. The sufficient condition is satisfied at z' since the gradient of utility is steeper than the normal to F at z', but this condition fails at the optimal point, z. In fact, the sufficient condition is satisfied for any point along F between $z^{(-i)}$ and z^*, but nowhere left of z^*. This is

intuitive; z' is closer to z^* (i's most preferred point) than $z^{(-i)}$, so i will not opt out of z'.

The Samuelson [90] condition for an interior optimum requires z to be to the left of z^*, where the sufficient condition fails. Thus, equilibrium participation requires that $z^{(-i)}$ be sufficiently to the right of z^*, causing t_i to be large. As in the opening example, large transfers are needed to incentivize participation, but feasibility may constrain how large the transfer can be or how many agents can have these inflated transfers. Clearly, this constraint will be more restrictive in larger economies, as will be demonstrated in Section 3.5.

3.4.2 Quasi-Linear Preferences

The transferable utility environment is especially important in mechanism design as the absence of wealth effects is useful in guaranteeing the ability to satisfy incentive compatibility constraints through transfer payments. It also allows a more precise quantification of the minimal transfer needed to satisfy equilibrium participation.

Assume agents have utility functions $u_i(x_i, y) = v_i(y) + x_i$, where $v_i' > 0$ and $v'' \leq 0$, and let the production function be strictly increasing and concave, so $c(y)$ is strictly increasing and convex. Let y_i^* be the unique solution to $c'(y) = v_i'(y)$. Equilibrium participation at a public good level of $\hat{y}$ requires that

$$t_i \leq \int_{y^{(-i)}}^{\hat{y}} v_i'(y)\, dy.$$

It must be that if t_i is non-negative, then

$$\int_{y^{(-i)}}^{\hat{y}} c'(y)\,dy \leq t_i,$$

with equality if the allocation is non-wasteful. In order for $\hat{y}$ to satisfy equilibrium participation for agent i when $\hat{y} > y_i^*$, it must be the case that

$$\int_{y^{(-i)}}^{y_i^*} \left(v_i'(y) - c'(y)\right) dy \geq \int_{y_i^*}^{\hat{y}} \left(c'(y) - v_i'(y)\right) dy, \tag{3.6}$$

both of which are non-negative quantities.

For an optimal allocation y^o, equation (3.6) provides an exact requirement on how 'far' $y^{(-i)}$ must be from y_i^* to guarantee equilibrium participation. This is demonstrated in Figure 3.5, in which $y^{(-i)}$ is the largest value satisfying (3.6) for the optimal point y^o. The necessary and sufficient conditions from equations (3.2) and (3.3) are also intuitive in this figure; if $y^{(-i)} > y_i^*$, then the necessary condition fails because marginal costs are everywhere larger than the marginal benefit between $y^{(-i)}$ and y^o, and the sufficient condition is satisfied for any $y \in [y^{(-i)}, y_i^*)$ since marginal costs are less than the marginal benefit at every public good level between y and $y^{(-i)}$.

3.5 Equilibrium Participation in Large Economies

The analysis of finite economies indicates that the large transfers needed to guarantee equilibrium participation for optimal allocations conflict with the feasibility constraints, particularly for larger economies. There is a fundamental difficulty in

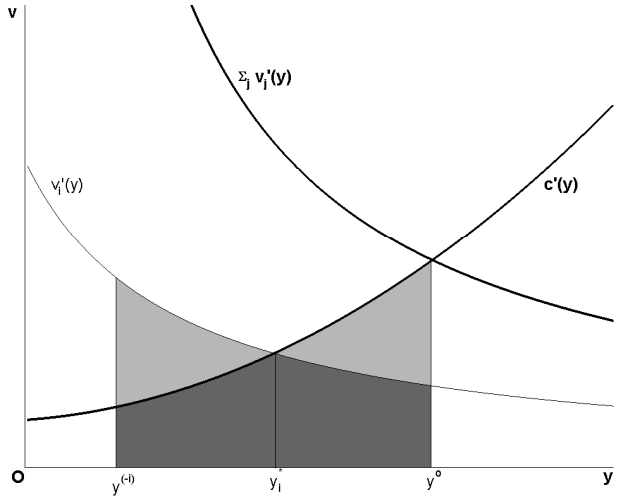


Figure 3.5: The Pareto optimal point y^o exactly satisfies equilibrium participation; if $y^{(-i)}$ were any larger, EP would fail.

the notion of a replica public goods economy. If each replicated agent is given the same endowment, then the total available production input grows without bound. Unless preferences bound the level of production, agents in large economies can find themselves consuming an infinite ratio of public to private goods.

Muench [74], Milleron [73], and Conley [25] discuss the difficulty of replicating public goods economies and offer various possible methods.[9] Milleron [73] provides an intriguing notion of replication; by splitting a fixed endowment among the replicates and adjusting preferences so that agents' concerns for the private good are relative to the size of their endowment, the fundamental difficulties of replication are mitigated. In essence, as the economy is replicated and agents are given a smaller share of the endowment, their preferences adjust proportionally to become more sensitive to the

[9]These authors are examining the convergence of the core of the economy to the Lindahl equilibrium. See Foley [36] for the appropriate definitions.

private goods holding. Thus, a very small shift in the absolute holdings of the private good is more significant to an agent with a small endowment in a big economy than to an agent with a big endowment in a small economy.

Formally, consider a base economy $\mathbf{e} \in \mathcal{E}_I$ with I unique agents such that $\mathbf{e} = \left(\{\succeq_i\}_{i\in\mathcal{I}}, \mathcal{Y}, \omega\right)$. A replica economy $\mathbf{e}^R$ is defined by replicating R times each $i \in \mathcal{I}$. Each replicate of consumer type i, denoted by the pair (i, r) for $r = 1, \ldots, R$, is endowed with ω_i/R units of the private good and a preference relation $\succeq_{i,r}$ such that $(x_{i,r}, y) \succeq_{i,r} \left(x'_{i,r}, y'\right)$, if and only if $(Rx_i, y) \succeq_i (Rx'_i, y')$ because of the scaling of endowments. This assumption on preferences of replicates mimics the approach of Milleron [73] and guarantees that private good consumption trade-offs are significant, even as the magnitude of those trade-offs becomes arbitrarily small. Finally, the production technology of $\mathbf{e}^R$ is assumed to be identical to that of $\mathbf{e}$.

This intuition that equilibrium participation becomes oppressively restrictive as an economy is replicated is confirmed by the following theorem:

Theorem 3.16 *For any economy satisfying A1, A3, and A4 (continuous, monotone preferences and increasing, continuous production technology,) the set of allocations satisfying equilibrium participation converges to the initial endowment as the economy is infinitely replicated.*

The proof of this theorem, available in the chapter appendix (Section 3.7,) demonstrates how the shrinking endowment restricts the amount any agent can be asked to pay in the limit. This, in turn, limits the agent's effect on production. Since agents in large economies care about small changes in their private goods consumption, but not

in the level of the public good, agents eventually prefer to opt-out as their individual effect on production vanishes.

This result is sensitive to the definitions of a replica economy. Consider instead a more standard notion of replication in which $\omega_{i,r} = \omega_i$ for each type i and replicate r, and assume $(x_{i,r}, y) \succeq_{i,r} \left(x'_{i,r}, y'\right)$ if and only if $(x_i, y) \succeq_i (x'_i, y')$. To see that the theorem no longer holds, construct a simple base economy $\mathbf{e} \in \mathcal{E}_I$ with an agent i for whom $(0, F(\omega_i)) \succ_i (\omega_i, 0)$. Here, the allocation $(\mathbf{x}, y)$ where $x_i = 0$, $x_j = \omega_j$ for all $j \neq i$, and $y = F(\omega_i)$ satisfies equilibrium participation. This economy can be replicated arbitrarily often, but the sequence of allocations $\left(\mathbf{x}^R, y^R\right)$ such that $x^R_{i,1} = 0$, $x^R_{j,r} = \omega_j$ for all $(j, r) \neq (i, 1)$, $y^R = F(\omega_i)$ satisfies equilibrium participation for every R, but does not converge to the endowment.[10]

Note that this result holds in economies where the set of Pareto optimal allocations remains far from the endowment as the economy grows, so that notion of approximate efficiency is of no benefit. For large economies, it is necessary that the committee or government has the power of coercion in order to overcome the free-rider problem.

3.6 Conclusion

If a mechanism is to implement a desired social choice correspondence with public goods when agents have available a no-trade alternative, it must select an allocation impervious to agents withdrawing their transfers. The incompatibility between equi-

[10] If the limit economy is represented by a measure space of consumers, however, this example fails because the contributions of a single individual are of measure zero and will not affect production of the public good.

librium participation and Pareto optimality is established through simple quasilinear examples, indicating that optimality is unobtainable under the standard assumptions used in mechanism design. In many economies, only the initial endowment is insusceptible to agents withdrawing. Even in those economies for which non-trivial allocations satisfy equilibrium participation, the set of equilibrium participation allocations eventually shrinks to the endowment as the economy is replicated.

The above analysis leaves open important questions about participation in public goods allocations. Perhaps it is possible to characterize those economies for which optimality is not inconsistent with equilibrium participation. If this class of such economies is reasonable to assume as the set of possible economies, then the negative results may be avoided with small numbers of agents. Similarly, there may exist a wide range of economies for which Pareto optimality may be well approximated under equilibrium participation. If such 'approximately desirable' outcomes could be identified, perhaps there exists a more natural mechanism that can implement these outcomes in Nash equilibrium. Given that the equilibrium participation constraint can be thought of as a restriction on the size of transfers, it is conceivable that a total transfer maximizing solution to this system of restrictions may be identified and used to maximize the total size of the public good in a given economy.

Finally, empirical observation demonstrates that non-trivial quantities of public goods are regularly provided in large economies. Governments and other voluntarily established methods of coercion exist as enforcement devices to guarantee that welfare improving allocations are attained. The next chapter provides a repeated-game

justification for endogenous enforcement, even when interactions are anonymous and individual reputations cannot be tracked. Such a process is a naturally occurring phenomenon within the larger private ownership/competitive mechanism framework, rather than a formally defined allocation mechanism. A larger model of how allocation mechanisms evolve in time has yet to be developed.

3.7 Appendix

Proof of Proposition 3.13. Pick any agent i such that $t_i, T_{-i} \geq 0$ and $y \geq F(T_{-i})$. Equilibrium participation implies that

$$u_i(\omega_i - t_i, F(T_{-i} + t_i)) \geq u_i(\omega_i, F(T_{-i})).$$

By quasi-concavity of u_i,

$$\nabla u_i(\omega_i, F(T_{-i})) \cdot (-t_i, F(T_{-i} + t_i) - F(T_{-i})) \geq 0,$$

or

$$\frac{F(T_{-i} + t_i) - F(T_{-i})}{t_i} \geq \frac{\partial u_i(\omega_i; F(T_{-i}))/\partial x_i}{\partial u_i(\omega_i; F(T_{-i}))/\partial y}.$$

Thus, by concavity of F,

$$\frac{\partial u_i(\omega_i; F(T_{-i}))/\partial x_i}{\partial u_i(\omega_i; F(T_{-i}))/\partial y} \leq F'(T_{-i}).$$

Inverting this inequality gives the necessary condition. ■

Proof of Proposition 3.14. By monotonicity, equilibrium participation is trivially satisfied for all j such that $t_j < 0$ or $y < F(T_{-j})$. Equation (3.4) guarantees equilibrium participation when $T_{-j} < 0$.

Now consider some $i \in \mathcal{I}$ such that $t_i, T_{-i} \geq 0$ and $y \geq F(T_{-i})$, but for whom equilibrium participation fails. For this agent,

$$u_i(\omega_i, F(T_{-i})) > u_i(\omega_i - t_i, F(T_{-i} + t_i)), \tag{3.7}$$

so that

$$\nabla u_i(\mathbf{x}; y) \cdot (t_i, F(T_{-i}) - F(T_{-i} + t_i)) > 0.$$

This is equivalent to

$$\frac{\partial u_i(\mathbf{x}; y)/\partial x_i}{\partial u_i(\mathbf{x}; y)/\partial y} > \frac{F(T_{-i} + t_i) - F(T_{-i})}{t_i}, \tag{3.8}$$

so applying the concavity of F at $T_{-i} + t_i$ and inverting the resulting relationship gives

$$\frac{\partial u_i(\mathbf{x}; y)/\partial y}{\partial u_i(\mathbf{x}; y)/\partial x_i} < \frac{1}{F'(T_{-i} + t_i)}.$$

Equation (3.3) implies that (3.7) cannot hold, so by the contrapositive of this argument, $(\mathbf{x}; y)$ must satisfy EP_i. ■

Proof of Proposition 3.15. For agents with $T_{-i} < 0$, $y^{(-i)} = 0$, but $F(T_{-i}) < 0$. By replacing $F(T_{-i})$ with zero in the proof of Proposition 3.14, the argument is

identical through equation (3.8). At this point, the subsequent relationship with $F'(T)$ cannot be derived from $F(T)/t_i$ when $T_{-i} < 0$, so inverting (3.8) gives the alternative sufficient condition

$$\frac{\partial u_i(\mathbf{x}; y)/\partial y}{\partial u_i(\mathbf{x}; y)/\partial x_i} \geq \frac{1}{F(T)/t_i} \tag{3.9}$$

for all i such that $T_{-i} < 0$. Since this is a stronger condition than (3.3), it is also sufficient for *every* agent. ■

Proof of Theorem 3.16. By way of contradiction, assume that there exists some economy e and some sequence $\left\{\left(\mathbf{x}^R; \hat{y}^R\right)\right\}_{R=1}^{\infty}$ in $\mathcal{EP}\left(e^R\right)$ for each R such that $\left|\hat{y}^R\right|$ fails to converge to zero.

For each (i, r), let $t_{i,r}^R = \omega_{i,r}^R - x_{i,r}^R$. For any $\left(\mathbf{x}^R; \hat{y}^R\right) \in \mathcal{EP}\left(e^R\right)$, if $\hat{y}^R < F\left(\sum_{i,r} t_{i,r}^R\right)$, then by monotonicity, $\left(\mathbf{x}^R; y^R\right) \in \mathcal{EP}\left(e^R\right)$, where $y^R = F\left(\sum_{i,r} t_{i,r}^R\right)$. In other words, if a wasteful allocation $(\mathbf{x}; \hat{y})$ satisfies equilibrium participation, so does the transfer-equivalent non-wasteful allocation $(\mathbf{x}; y)$. Thus, the sequence $\left\{\left(\mathbf{x}^R; y^R\right)\right\}_{R=1}^{\infty}$ satisfies equilibrium participation for each R and $\left\{\left|y^R\right|\right\}_{R=1}^{\infty}$ also fails to converge to zero. This implies that there exists a subsequence $\left\{\left(\mathbf{x}^{R(k)}; y^{R(k)}\right)\right\}_{k=1}^{\infty}$ such that $\left|y^{R(k)}\right| > \varepsilon$ for some $\varepsilon > 0$ all $k \in \mathbb{N} = \{1, 2, \ldots\}$. Letting $c(y)$ represent the minimal cost of producing y (which is the inverse of F,) non-convergence guarantees that $c\left(y^{R(k)}\right) \geq c(\varepsilon) > 0$ for each k since c is an increasing function and $\mathcal{Y} \cap \mathbb{R}_+^2 = \{0\}$.

For any k, if $R(k) > \max_{i \in \mathcal{I}} (\omega_i / c(\varepsilon))$, then no one agent (i, r) can unilaterally

fund $y^{R(k)}$ using $t_{i,r}^{R(k)}$ since

$$\begin{aligned} t_{i,r}^{R(k)} &\leq \max_{i \in \mathcal{I}} \omega_i / R\left(k\right) \\ &< c\left(\varepsilon\right) \\ &\leq c\left(y^{R(k)}\right). \end{aligned}$$

Letting

$$k^* = \max\left\{k \in \mathbb{N} : R\left(k\right) \leq \max_{i \in \mathcal{I}} \left(\omega_i / c\left(\varepsilon\right)\right)\right\},$$

there exists at least one sequence of agents $\{(i_k, r_k)\}_{k=1}^{\infty}$ such that

$$t_{i_k,r_k}^{R(k)} \geq c\left(y^{R(k)}\right) / \left(R\left(k\right) I\right)$$

for all k, and $T_{-(i_k,r_k)} > 0$ for all $k > k^*$. In other words, there exists a sequence of agents such that at each k, the identified agent is paying a transfer which is more than the average transfer of $c\left(y^{R(k)}\right) / \left(R\left(k\right) I\right) > 0$, and the sum of the others' transfers is eventually positive as individual (i_k, r_k)'s budget constraint becomes restrictive. For example, $\{(i_k, r_k)\}_{k=1}^{\infty}$ might identify the agent (i, r) in each k for whom $t_{i,k}^{R(k)}$ is maximal among all agents (this particular sequence may not have a well-defined limit, but any selection of agents paying an above average proportion of the cost is sufficient.)

Since each $\left(\mathbf{x}^{R(k)}; y^{R(k)}\right)$ satisfies equilibrium participation for all (i, r), it must be

the case that

$$\left(\omega_{i,r} - t^{R(k)}_{i_k,r_k}, y^{R(k)}\right) \succeq_{i_k,r_k} \left(\omega_{i,r}, \left(y^{R(k)}\right)^{-(i_k,r_k)}\right),$$

or equivalently,

$$\left(\omega_i - R\left(k\right) t^{R(k)}_{i_k,r_k}, y^{R(k)}\right) \succeq_{i_k} \left(\omega_i, \left(y^{R(k)}\right)^{-(i_k,r_k)}\right).$$

Note that for $k > k^*$,

$$\left(y^{R(k)}\right)^{-(i_k,r_k)} = F\left(\sum_{j,s} t^{R(k)}_{j,s} - t^{R(k)}_{i_k,r_k}\right).$$

By continuity of the production function, $\left(y^{R(k)}\right)^{-(i_k,r_k)}$ becomes arbitrarily close to $y^{R(k)}$ as k grows. However, since $t^{R(k)}_{i_k,r_k} > c\left(\varepsilon\right)/\left(R\left(k\right)I\right)$, then $R\left(k\right) t^{R(k)}_{i_k,r_k}$ is bounded below by $c\left(\varepsilon\right)/I > 0$ at all k. By monotonicity of preferences, it must be the case that

$$\left(\omega_i - \frac{c\left(\varepsilon\right)}{I}, y^{R(k)}\right) \succeq_i \left(\omega_i - R\left(k\right) t^{R(k)}_{i_k,r_k}, y^{R(k)}\right) \succeq_i \left(\omega_i, \left(y^{R(k)}\right)^{-(i_k,r_k)}\right).$$

By continuity of preferences, convergence of $\left(y^{R(k)}\right)^{-(i_k,r_k)}$ to $y^{R(k)}$ implies that for large enough k,

$$\left(\omega_i - \frac{c\left(\varepsilon\right)}{I}, y^{R(k)}\right) \succeq_{i_k} \left(\omega_i, y^{R(k)}\right).$$

However, this violates monotonicity. Since there cannot be an infinite subsequence of allocations with $\left|y^{R(k)}\right| > \varepsilon$ for any $\varepsilon > 0$, it must be the case that $y^R \to 0$ as

$R \to \infty$. Feasibility then requires that $\left\|\mathbf{x}^R - \omega^R\right\|_\infty \to 0$, completing the proof. ■

Chapter 4

Group Reputations and Stereotypes as Contract Enforcement Devices

Incomplete contracts are frequently observed despite the well-known incentive distortions they create. For example, if a product's quality is not verifiable, sellers have a clear incentive to deliver lower-quality goods. Rational buyers recognize this incentive and adjust their demand accordingly. The resulting transaction is often Pareto dominated by the 'cooperative' outcome of high quality goods sold at higher prices. Despite these difficulties, cooperative market interactions continue to take place in the absence of complete, verifiable contracts; buyers often trust sellers to deliver a high quality product and sellers often respond in kind.

Economic theory has struggled to explain the success of the marketplace in the face of enforcement difficulties. An appealing argument is that repeated interactions act as an enforcement device when the cost of damaging a valuable long-term relationship outweighs the immediate benefit of poor performance, as in the models of Klein & Leffler [57] or MacLeod & Malcomson [65]. By adding a small probability of

agents being unconditionally cooperative, Kreps *et al.* [59] show how false reputation-building by selfish agents can lead to full cooperation in early periods of the finitely repeated prisoners' dilemma. These models implicitly require that trading partners' identities be known in order for reputation-building to occur.

Experimental studies show, however, that cooperation can emerge even when interactions are anonymous. In tests of moral hazard in the labor market by Ernst Fehr and many others (see Fehr *et al.* [34] and [35], Charness [16], Fehr & Falk [32], Charness *et al.* [17], Gächter & Falk [39], and Hannan *et al.* [44], among others), wages and effort levels are observed to be substantially higher than the stage game equilibrium prediction even though transactions are anonymous. Furthermore, these studies show little evidence of reversion to the equilibrium in the final periods. Therefore, the authors conclude that fairness norms (such as a natural preference for 'gift exchange') solve the moral hazard problem, not reputation-building.

Not all experimental studies confirm these results. Lynch et al. [63], Engelmann & Ortmann [29], and Rigdon [84] find behavior consistent with, or converging toward, the stage game equilibrium. Even some studies purporting the existence of fairness preferences include some sessions with strong end-game effects, as in Fehr *et al.* [35] and Riedl & Tyran [83]. These apparently contradictory results leave open the question of what forces are at work to offset the shirking incentive.

The current chapter provides three novel results. First, in a direct replication of Fehr *et al.* [34], high wages and efforts observed in early periods collapse dramatically to the stage game equilibrium in the last period. Second, a group reputation

(or 'stereotyping') model is developed that explains this behavior. Third, a new experimental environment is tested in which individuals have insufficient incentives to maintain the group's reputation. As predicted by the model, actual subjects play the stage game equilibrium in every period. Furthermore, the results of many previous experimental papers are consistent with the model's predictions.

The stereotyping model works as follows:

Assume, *à la* Kreps *et al.* [59], that some percentage of workers are unconditional cooperators whose efforts are always positively correlated with their wage. If firms believe that worker types are perfectly correlated – even if that belief is empirically unsupported – then a single defection by one worker leads to the belief that all workers are 'selfish,' destroying the reputation of the entire group and causing low wages in all subsequent periods.[1] Under the payoff structure used by Fehr *et al.* [34], selfish workers have an incentive to maintain a group reputation until the very last period (unless firms are very certain *a priori* that the workers are selfish.) By changing the payoff structure, one can eliminate the existence of such group reputation equilibria. Indeed, this is the phenomenon observed in a new set of experiments.

The assumption of stereotyping behavior, though irrational, is a well documented phenomenon in the social psychology literature. People use stereotypes to economize on cognitive resources in the processes of evaluating and recalling information about other individuals (McGarty et al. [70, p. 3–5]). Effectively, stereotyping is an inexpensive internal reputation management system. By assigning attributes to groups rather than individuals, decision makers can easily estimate the attributes of individ-

[1]This is similar in spirit to the contagion mechanism of Kandori [54].

ual group members. The danger of this cognitive shortcut is, of course, potentially harmful inaccurate beliefs about individuals.

The following section formally introduces the gift exchange market under various payoff structures and explores the equilibrium predictions of the stage game. Sections 4.2 and 4.3 describe experimental sessions in which high wages and effort are observed only under certain conditions, though all sessions show strong signs of repeated game considerations. Section 4.4 simplifies the structure of the gift exchange market for game theoretic analysis and Subsection 4.4.2 formally introduces the stereotype-reputation model, along with evidence that such a model has solid foundations in the field of social psychology. Subsection 4.4.3 compares the predictions of the stereotype-reputation model to past gift exchange market experiments, and Section 4.5 concludes the chapter.

4.1 The Gift-Exchange Market

A single play of a gift exchange market (GEM) has the following structure:

A finite set $\mathcal{J}$ of firms and a finite set $\mathcal{I}$ of workers participate in a two-stage posted-price labor market. Let $J = |\mathcal{J}|$ and $I = |\mathcal{I}|$, with $I > J$. In the first stage, each firm $j \in \mathcal{J}$ posts at most one wage offer w_j in the market. The set of allowable offers is given by $\mathcal{W} \cup \{\phi\}$, where $\mathcal{W} \subseteq \mathbb{R}$ is a compact set of non-negative wage offers and ϕ represents no wage offer. Workers $i \in \mathcal{I}$ may accept any outstanding offers w_j in the market at any time. After w_j has been accepted by some worker, firm j may not post any further wage offers, so each firm may hire at most one worker. After

a fixed amount of time (or after all firms have had a wage offer accepted,) the first stage ends and unmatched agents earn zero profit.

In the second stage, each worker that accepted a wage offer selects an effort level e_i from a linearly ordered set $\mathcal{E}$ such that $\inf \mathcal{E} \in \mathcal{E}$. Let $\chi_i(w) \in \{0,1\}$ denote whether or not worker i accepts the given wage offer w and let $e_i(w) \in \mathcal{E}$ denote the effort level chosen. Note that by making workers the second mover, effort choices are quite naturally dependent on wages.

The monetary payoffs realized by each firm j and worker i are given by π and u, respectively, each mapping strategy pairs from $(\mathcal{W} \cup \{\phi\}) \times (\{0,1\} \times \mathcal{E})$ into $\mathbb{R}$ such that unmatched agents receive zero profit. The functions π and u are identical across agents and are common knowledge. Assume u is monotone decreasing in e and increasing in w, and π is monotone increasing in e and decreasing in w.

4.1.1 Stage Game Equilibrium

Define $e_{\min} = \inf \mathcal{E}$ and $w^* = \inf \{w \in \mathcal{W} : u(w,(1,e_{\min})) \geq 0\}$. Here, $e_{\min}$ is the minimal effort choice and w^* is the reservation wage at $e_{\min}$. Without loss of generality, assume that $e_i(w) = e_{\min}$ whenever $\chi_i(w) = 0$ since effort choices are irrelevant when no wage is accepted. In the second stage of the one-shot game, consider the strategy for each worker i given by

$$(\chi_i^*(w), e_i^*(w)) = \begin{cases} (1, e_{\min}) & \text{if} \quad w \geq w^* \\ (0, e_{\min}) & \text{if} \quad w < w^* \end{cases},$$

where $e_{\min}$ is chosen in response to any acceptable wage offer. This is the workers' unique dominant strategy since u is decreasing in e and $u(w,(1,e)) < u(w,(0,e_{\min})) = 0$ for each $e \in \mathcal{E}$ when $w < w^*$. Note that while the choice of whether to accept a given wage offer is dependent upon w, the choice of $e_{\min}$ is not.

In the current set of environments, $\pi(w^*,(1,e_{\min})) > 0$, so the firms have an incentive to participate. The unique subgame perfect equilibrium outcome is for each firm j to offer $w_j = w^*$ and for every worker i to accept w^* with an effort level of $e_{\min}$.[2] Since $I > J$, involuntary unemployment will still result.[3]

4.1.2 Three Specifications

Three specifications of the GEM are tested experimentally. Each varies in the functional form of agent payoffs and in the information feedback conditions, although the game forms are identical across specifications. In all three treatments, $I = 9$, $J = 6$, $\mathcal{E} = \{1,2,\ldots,10\}$ and $\mathcal{W} = \{5,10,15,\ldots\}$. Utilities are chosen such that $w^* = 30$, so the stage game equilibrium always predicts a wage of 30 and an effort choice of $e_{\min} = 1$.

[2]There exists a Pareto-dominated 'no-trade' Nash equilibrium to this game in which all J firms choose to make no wage offers ($w_j = \phi$ for all $j \in \mathcal{J}$) and all I workers choose to reject all wage offers ($\chi_i(w) \equiv 0$ for all $i \in \mathcal{I}$.) Unless otherwise indicated, further discussion of the stage game equilibrium will refer only to the subgame perfect equilibrium with full employment.

[3]In equilibrium, workers know that w^* will be the only wage offer in the market and will therefore accept w^* immediately. Allocation of firms to workers is assumed to be random in the situation of multiple simultaneous acceptances, so that the set of unemployed workers will be randomly selected in equilibrium.

4.1.3 Treatment 1: High MRS Ratio, Anonymous IDs (HRA)

The first variant of this game – denoted HRA – is the original gift exchange market studied by FKR. Here, the payoffs for matched firms and workers are given by

$$\pi_1(w,(1,e)) = (126-w)\left(\frac{1}{10}e\right) \tag{4.1}$$

and

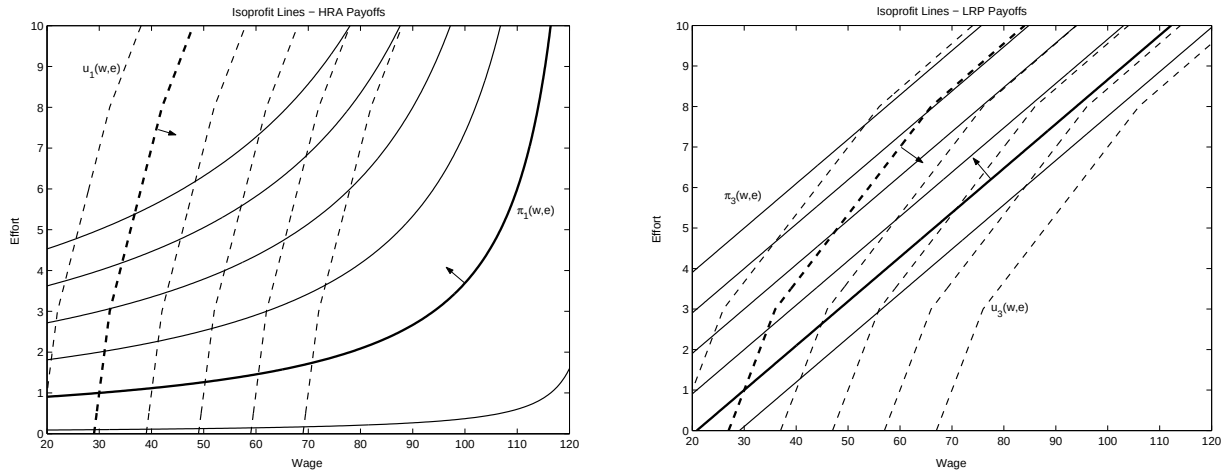
$$u_1(w,(1,e)) = w - 26 - c(e), \tag{4.2}$$

where the cost of effort is given by

$$c(e) = \begin{cases} -1+e & \text{if} \quad e \in \{1,2,3\} \\ -4+2e & \text{if} \quad e \in \{4,5,6,7\} \\ -12+3e & \text{if} \quad e \in \{8,9,10\} \end{cases}.$$

In the experiment, π_1 and u_1 are denoted in francs which are then converted to dollars at a rate of 12 francs per dollar.

At the equilibrium strategy profile $(w^*, e_{\min})$, each firm's marginal rate of substitution is $(\partial\pi_1/\partial e)/(\partial\pi_1/\partial w) = -1$ and each worker's MRS is $-1/96$. The ratio of these values is quite high, indicating that each party in a transaction can have a dramatic effect on the payoffs of the counterparty at little cost to themselves. As strategies move farther from equilibrium, the cost of affecting the other agent's profit increases. This is clear from the graph of the level curves of π_1 and u_1 in the space $\mathcal{E} \times \mathcal{W}$, given in Panel (a) of Figure 4.1.



(a) **(b)**

Figure 4.1: Isoprofit lines for workers and firms in (a) HRA – the gift exchange market used by Fehr, Kirchsteiger & Riedl, and (b) LRP, the new design with quasilinear payoffs.

In HRA, wage offers are displayed for all agents to see, but the identity of the firm offering each wage is known only to the firms. Similarly, the acceptance of wage offers is public information, but the identity of the accepting worker is known only among the workers. Finally, the effort level decision of each matched worker is made after the market is closed and revealed only to the hiring firm. No other firms or workers observe this decision. Thus, it is impossible for firms to develop informative reputations about individual workers in this environment. In sessions using HRA, subjects are paid one U.S. Dollar for every 12 'francs' earned in the experiment.

4.1.4 Treatment 2: High MRS Ratio, Public IDs (HRP)

The second variant – HRP – alters the information structure and payoff conversion rates of HRA. First, all agents observe the player ID number associated with each wage offer and with each worker accepting any given wage offer. Second, effort level decisions are made immediately after a worker accepts a wage offer and this decision is posted (along with the worker's ID number) for all agents to observe. This not only provides information for the formation of individual reputations across periods, but also allows all agents to observe the realization of strategies chosen by each worker, given the accepted wage offer before the market closes. Finally, the conversion rate between experimental currency and actual payoffs is increased to four francs per dollar for the workers and nine francs per dollar for the firms so that consequences of strategy choices have increased saliency.

The payoff functions of HRP are identical to HRA, so that $\pi_2 \equiv \pi_1$ and $u_2 \equiv u_1$. Thus, the high ratio of the marginal rates of substitution is maintained.

4.1.5 Treatment 3: Low MRS Ratio, Public IDs (LRP)

The third variant, LRP, alters HRP to make the payoffs of both agents quasilinear in wages. The cost of effort function is tripled to reduce the disparity between the effect of a change in effort on workers and firms. Finally, a linear rescaling of the 'value of effort' to the firms is used to adjust payoffs for the experimental environment. These changes serve to reduce the equilibrium MRS ratio so that workers cannot dramatically affect the payoff of the firms without significant penalty to themselves,

and vice-versa. The payoff functions for LRP are given by

$$\pi_3(w,(1,e)) = 126\left(\frac{11}{40}+\frac{2.9}{40}e\right) - w \tag{4.3}$$

and

$$u_3(w,(1,e)) = w - 26 - 3c(e). \tag{4.4}$$

The conversion rate of 12 francs per dollar is used for all subjects.

The ratio of marginal rates of substitution at the equilibrium is three in this design, as opposed to 96 in the previous treatments. Keeping the ratio greater than one guarantees that there still exists wage-effort pairs that Pareto dominate the equilibrium so that players still have an incentive to attempt cooperation. The level curves of the payoff functions are given in Panel (b) of Figure 4.1.

4.2 Experimental Design

The three designs were tested experimentally at the California Institute of Technology Laboratory for Experimental Economics and Political Science (EEPS) using Caltech undergraduate students recruited via e-mail. Subjects were randomly divided into two groups of six firms and nine workers, with each group separated into a different room. The instructions did not make reference to firms, workers, wages, or effort levels.[4] To avoid a labor market framing, subjects were labelled as buyers and sellers,

[4]For the HRA session, the instructions provided by Fehr et al. [34] were used. For HRP and LRP, the instructions were appropriately modified. Copies of the instructions are available upon request.

and their task was to post prices for a good in a market and choose a 'conversion rate' (rather than an effort level) that affected payoffs. The term 'conversion rate' is used in HRA and HRP to emphasize that sellers, by their choice of e, are choosing the percentage of $(126 - w)$ that their buyer will be paid. In LRP, the effort level choice can no longer be thought of as a conversion rate on firms' profits, so the generic name, 'X' was instead used in the instructions to identify this choice variable.

Telephone conversations between two experimenters was used as the means of transmitting information between rooms during the market stage of each period. All decisions were posted on the blackboard in both rooms.[5] When effort levels were not publicly viewable, the worker wrote his effort decision on an index card that was delivered to the appropriate subject in the other room. The first session (S1) consisted of HRA repeated over twelve periods. The second (S2) ran HRP for twelve periods, while the third and fourth sessions (S3 and S4) ran LRP for twelve periods. The fifth session (S5) was divided into two parts.[6] First, LRP was played for six periods. Immediately following, the same subjects read instructions and participated in HRA for six periods.[7] The treatment-switching design in S5 tests whether or not social norms or reputations developed in LRP affect behavior in HRA which can then be compared to behavior in S1. If behavior is substantially different between HRA and LRP within S5, then differences in the structure of the two markets apparently

[5]In later sessions, the market information was projected on a screen and transmitted by computer. This had no effect on the actual procedures of the experiment.

[6]Sessions were run on the following dates: S1 on 12/03/2002, S2 on 12/05/2002, S3 on 08/05/2004, S4 on 08/04/2004, and S5 on 01/14/2003.

[7]Although subjects were informed that they would participate in two different experiments, they were not given specific information or instructions about the second treatment until the conclusion of the first.

cause differences in behavior. Each session lasted between 90 minutes and two hours. In sessions S1 and S5, subjects earned an average of $35, while earnings in S2 were as high as $130 due to the reduced exchange rate. In S3 and S4, average earnings were around $25 because of the reduced cooperation in that treatment.

4.3 Experimental Results

See Figures 4.2, 4.3 and 4.4 for a complete representation of the data from the five experimental sessions.[8] These results show that effort does appear to be an increasing function of wages. In HRA and HRP, wage-effort pairs are well above equilibrium, but play converges to the stage game equilibrium in the final period. In LRP, players are unable to coordinate on high wage-effort pairs and the stage game equilibrium is observed across all periods. In HRA played after LRP, subjects are able to coordinate on high wage-effort pairs, indicating coordination is easier in the latter environment. These results are inconsistent with a pure fairness hypothesis, but due to the information structure of HRA, must also be inconsistent with individual reputation building. Thus, the results are indicative of a group reputation effect.

In previous experiments, a strong positive correlation between wages and efforts has been observed. This key wage-effort relationship is traditionally taken as evidence

[8]In session S4, two subjects acting as workers had not been matched with many wage offers in the first several periods and consequently, had accumulated very little earnings by the 7^{th} and 8^{th} periods. These subjects, informed that they would not have to pay their losses to the experimenter, began to accept the smallest possible wages and offer the highest possible effort in an attempt to create maximal wealth for the (anonymous) firms. After 4 such actions, one worker was removed from the experiment and the other immediately (and voluntarily) stopped participating. These four data points are removed from analysis, but likely affected beliefs in the market for the remainder of the session.

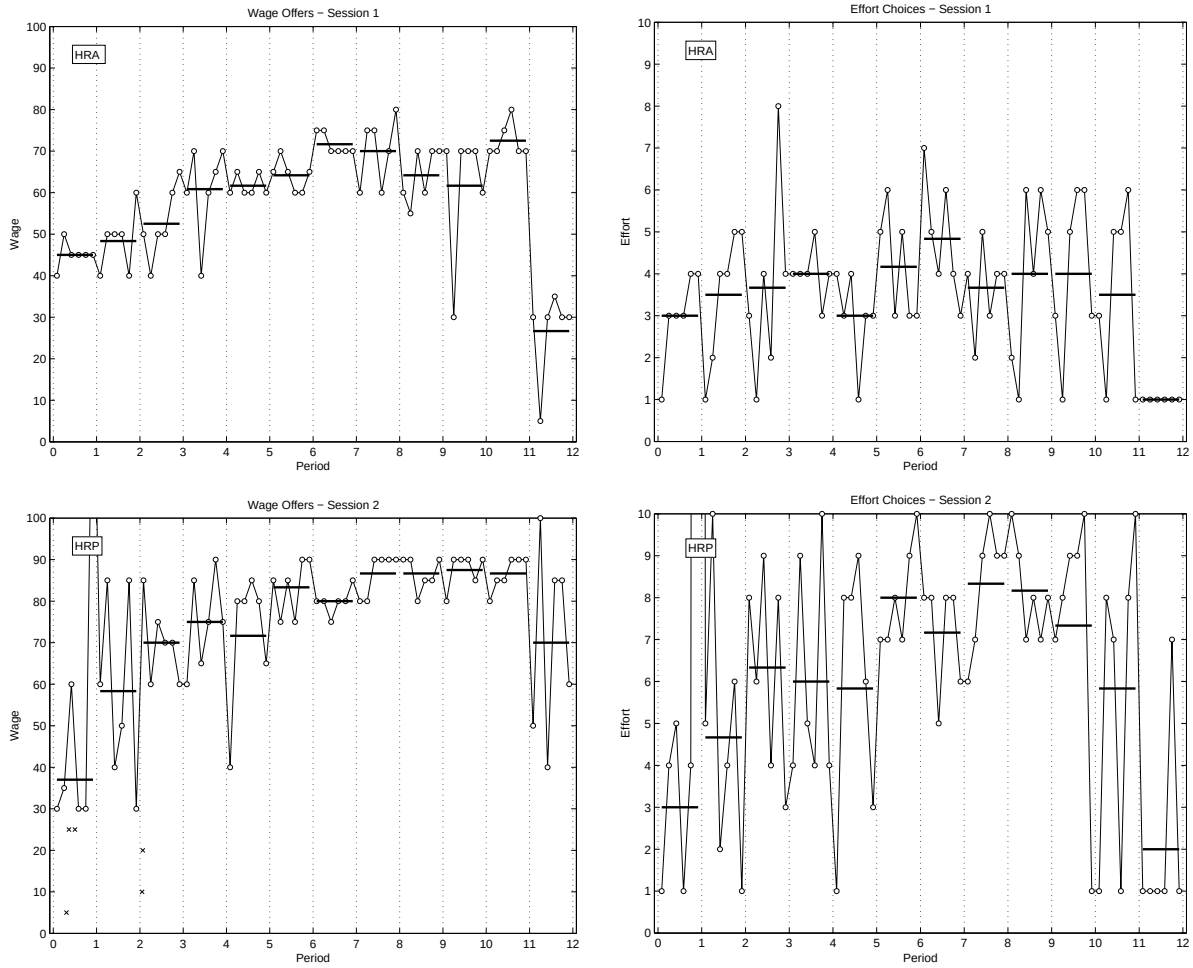


Figure 4.2: Wage and effort levels across time in sessions S1 (a replication of the Fehr, Kirchsteiger and Riedl (1993) experiment,) and S2 (the same design with individual reputations added.) Solid lines represent period averages and x's represent unaccepted bids.

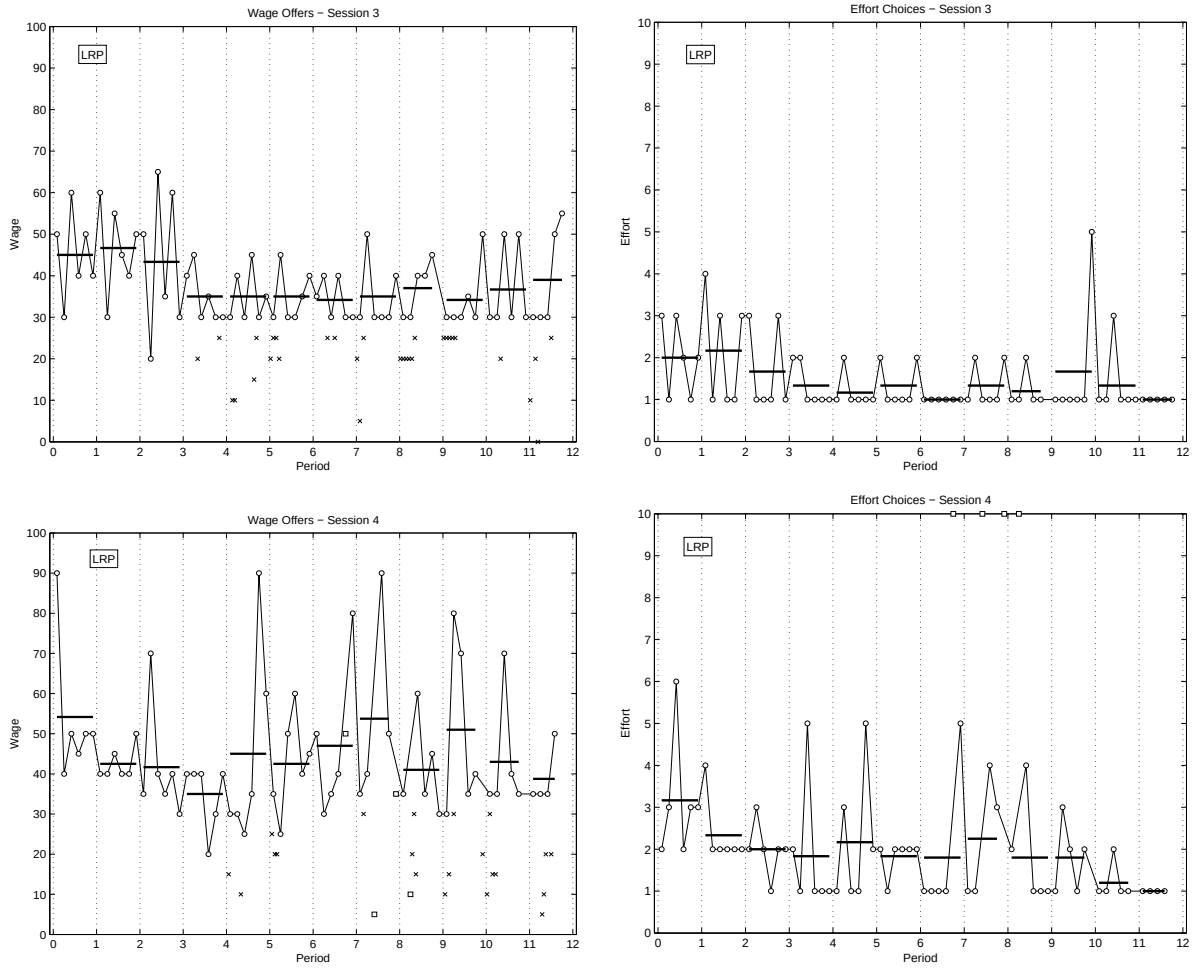


Figure 4.3: Wage and effort levels across time in sessions S3 and S4 (with quasilinear payoffs.) Solid lines represent period averages and x's represent unaccepted bids. Four data points in S4 (represented by squares) are removed from analysis. See footnote 8.

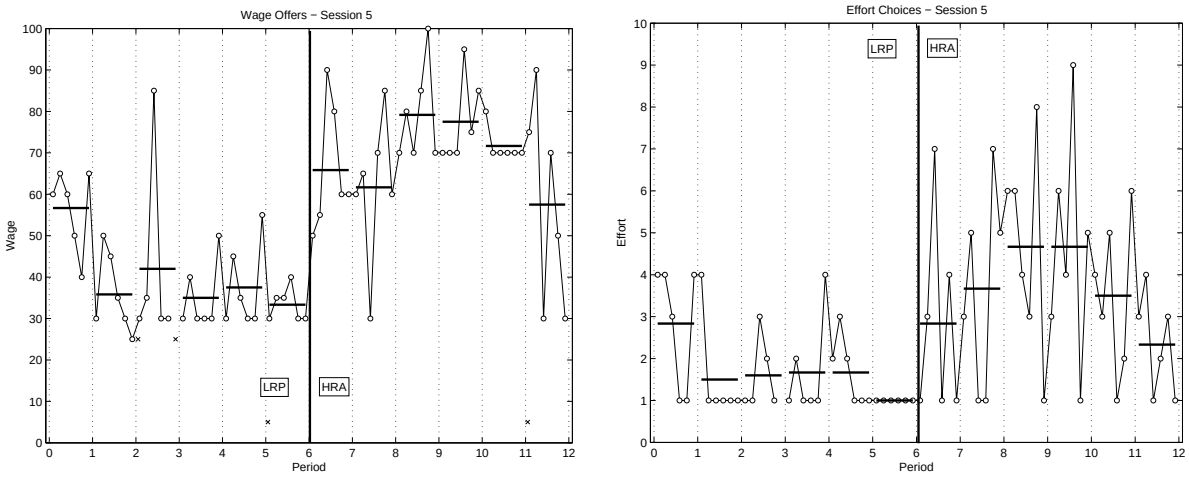


Figure 4.4: Wage and effort levels across time in session S5, switching from LRP to HRA after period 6. Solid lines represent period averages and x's represent unaccepted bids.

of reciprocal-minded subjects, but is also consistent with repeated game explanations in which selfish subjects either mimic reciprocal agents in early periods (a reputation effect) or use punishment strategies to enforce cooperation (a folk theorem effect.) Regardless of the cause, this correlation is also observed in the current set of experiments.

Session (Treatment)	RankCorr(w, e)	p-Value
S1 (HRA)	0.546	7.07×10^{-7}
S2 (HRP)	0.641	1.64×10^{-9}
S3 (LRP)	0.604	2.64×10^{-7}
S4 (LRP)	0.446	0.0001
S3 (LRP)	0.499	0.0023
S3 (HRA)	0.511	0.0015

Table 4.1: Spearman rank correlation coefficients between effort and wages for each treatment

Result 4.1 *In all treatments in all sessions, wages and efforts of matched firms and workers are positively correlated as predicted by both fairness and repeated game explanations.*

Support. Spearman rank correlation coefficients between wages and effort are calculated for each treatment. For each, the coefficient is estimated to be at or greater than 0.446 and significantly positive at the 0.5 percent level. The estimates of the individual coefficients are given in Table 51. ■

Although the wage-effort correlation is inconclusive about the strategies employed, the following set of results provide evidence in support of a group-reputation repeated game explanation.

Result 4.2 *In the replications of the FKR gift exchange market (HRA), strategies converge toward the stage-game equilibrium in the final period, providing evidence of repeated game strategies.*

Support. See Figures 4.2 and 4.4. In session S1, the average wage for periods 4 through 11 is 65.83. In period 11, the final wage offer of 70 was given a minimal effort level. In the final period (period 12), all wages except one and all efforts are at or below the equilibrium values, and the average wage was less than the equilibrium prediction.[9] In the latter half of session S5, average wages decline in the final periods and two wage offers appear at equilibrium levels in the last period, although the average wage offer only declines to 57.5. Similarly, average effort choices decline in

[9]One worker accidentally accepted a sub-equilibrium wage offer in his haste to participate in the market.

the final periods and two last-period effort choices are at the equilibrium prediction, although the average declines to only 2.33. ■

For any model of fairness to accommodate these results, agents' relative preference for fairness must be time-dependent. Furthermore, subject pool effects are significant in the last period since the HRA represents an exact replication of the Fehr *et al.* [34] experiment using different subjects. The percentage of purely fair-minded agents is apparently lower in the population used for the current study, although high wage-effort choices are observed in early periods across subject pools.

Result 4.3 *In the first 11 periods of the 12-period replication of the FKR experiment (HRA), the average wage increases in time, while average effort does not significantly change.*

Support. Each transaction is numbered chronologically from 1 to 72, with six transactions per period. The rank correlations between transaction numbers and wages and between transaction numbers and effort are estimated for the first 66 transactions (11 periods) using Spearman rank correlation coefficients. Rank correlation between wages and transaction number is estimated at 0.6905 with a 2-sided p-value of 1.4×10^{-10}, while correlation between effort and transaction number is estimated at 0.1711 with a p-value of 0.1696. Rank correlations are also estimated using period number as a proxy for time instead of transaction number, yielding very similar results. ■

That wages increase in time is not predicted by either theory under consideration, but it is not necessarily inconsistent with a reputation-building equilibrium. On the

other hand, fairness explanations are time-independent, suggesting that pure fairness concerns are inadequate to explain behavior.

The following results indicate that the ability of players to achieve outcomes that Pareto dominate the stage game equilibrium is not robust to the payoff specifications.

Result 4.4 *In LRP, the minimum effort level is played more often than all other strategies combined, and the effort level regresses to the stage game equilibrium strategy in the final two periods, suggesting that behavior is highly sensitive to the market's payoff structure.*

Support. Of the 169 effort decisions in LRP treatments, 60.4 percent are at $e = 1$. Over 91 percent of all observations are at effort levels 1, 2, or 3. Only 5 effort choices are observed at $e \geq 5$. In the penultimate period, the minimal effort is observed in 12 of the 17 transactions, with an average effort of 1.411. In the final period, *every one* of the 15 sellers selects $e = 1$. ■

Result 4.5 *In LRP, the firm's stage game equilibrium strategy ($w = 30$) is the modal observation and the frequency of this strategy increases with time.*

Support. The subgame perfect strategy of $w = 30$ occurs in 32 percent of the 169 accepted wage offers in treatment LRP – more often than any other strategy is used. Wage offers of $w \leq 40$ constitute 68 percent of all observations and 86.4 percent of accepted wage offers are no greater than 50. ■

If fairness considerations are independent of payoff specifications, then fair-minded workers will always be willing to reciprocate in all gift exchange environments. This is

not observed in the LRP experiment. The following result indicates that the difference in environment is responsible for the differences in behavior.

Result 4.6 *In the treatment-switching session (S5), average wages and effort levels increase after switching from LRP to HRA, indicating that changing the payoff structure back to that of FKR induces the same group of subjects to choose high wages and effort.*

Support. To avoid problems with non-stationarities in the time series, each wage and effort from LRP is compared to the wage and effort from HRA with the same time identifier (or "transaction number"). These differences are analyzed using a Wilcoxon signed rank sum test. For wages, the HRA values are significantly greater than those from LRP, with an estimated z-statistic of 4.693. Similarly, HRA effort choices are significantly greater with a z-statistic of 3.652. Thus, significance in both cases is better than 0.015 percent. ■

This result is obvious from Figure 4.4. It is clear that the change in behavior immediately follows the change of the market in the seventh period. The fact that the same group of subjects generates two very different sets of data in two similar markets played consecutively implies that the difference in behavior is due to differences in the two markets. Therefore, the structure of HRA appears much more conducive to high wage-effort pairs than that of LRP.

Result 4.7 *In a replication of FKR's experiment with full information about past strategy choices (HRP), wages and effort are significantly greater than in the original design (HRA).*

Support. Again, wages and effort were paired between sessions according to their time identifier and a Wilcoxon signed rank sum test was performed in the differences. Wages are significantly higher in HRP, with a z-statistic of 5.925. Effort is also significantly higher in HRP, with a z-statistic of 5.401.[10] ■

Recall that the only differences between HRP and HRA are that the worker ID numbers and effort levels are publicly displayed with each transaction and the payoffs are increased to 4 francs/dollar for the workers and 9 francs/dollar for the firms. Thus, individual reputations and/or increased payoffs lead to more cooperation. Comparing HRP to LRP, however, reveals that cooperative behavior is sensitive to the payoff specifications, and not to the information structure of the game. This is consistent with a reputation equilibrium hypothesis in which firms are more likely to engage in trusting behavior if the relative payoff is greater.

4.4 A Reputation Model With Stereotypes

The results of Section 4.3 indicate that the stage game equilibrium prediction obtains in the LRP environment, but not in HRA or HRP until the final period. However, the remarkable reversion to the stage game prediction in the final period of both HRA and HRP indicates that a simple model of fairness is inadequate to explain the high effort and wages observed in early periods. The final period crash is reminiscent of a repeated game sequential equilibrium with reputations, as in Kreps *et al.* [59] and Kreps & Wilson [60].

[10] That this difference is more significant than that documented in the previous result is also a consequence of the larger sample size.

As Proposition 4.1 will demonstrate, a standard sequential equilibrium cannot explain the observed behavior since worker decisions are not linked to ID numbers. However, the psychology literature on categorical thinking and stereotyping indicates that it may be appropriate to assume that firms erroneously believe worker types are correlated. Proposition 4.2 shows that under such an assumption, the observed behavior in *all three* sessions is consistent with a sequential equilibrium explanation. Furthermore, the results of many past studies are also explainable in this framework.

4.4.1 The Basic Framework

To apply more complex game theoretic arguments to the gift exchange market, a simplified version of the GEM is developed. This 'mini-GEM' consists of a stage game repeated over T periods. In each period t, let $\mathcal{W} = \{\underline{w}, \overline{w}\}$ with $\overline{w} > \underline{w}$, and let $\mathcal{E}(\overline{w}) = \{\underline{e}, \overline{e}\}$, and $\mathcal{E}(\underline{w}) = \{\underline{e}\}$ with $\overline{e} > \underline{e}$, so that allowable effort choices are a function of the wage.[11] A fraction J/I of the workers are chosen with equal probability each period and matched with a wage offer. Let $j(i,t)$ denote the firm matched to worker i (if any) in period t, and $i(j,t)$ denote the worker matched to firm j. Define $w_{j,t}$ and $e_{i,t}$ to be the wage and effort in period t of firm j and worker i, respectively. Unmatched workers have no strategy choice and receive zero payoff.

Workers' types ($\boldsymbol{\theta}_i$) are selected from the set $\Theta = \left\{\underline{\boldsymbol{\theta}}, \overline{\boldsymbol{\theta}}\right\}$, where $\underline{\boldsymbol{\theta}}$ represents a 'selfish' worker and $\overline{\boldsymbol{\theta}}$ represents a 'reciprocal' or 'fair' worker. Worker payoffs for

[11] The restriction of $\mathcal{E}(\underline{w}) = \{\underline{e}\}$ is for simplicity of exposition. As will be apparent, allowing $\mathcal{E}(\underline{w}) = \{\underline{e}, \overline{e}\}$ does not alter the equilibrium analysis since selecting a high effort in response to a low wage can be assumed to perfectly signal that the agent is *not* reciprocal, thus causing reduced current period payoffs *and* reduced continuation payoffs. Furthermore, Clark & Sefton [23] experimentally demonstrate that in a sequential prisoners' dilemma, $\overline{e}$ is played in response to $\underline{w}$ in less than 5% of decisions.

type $\underline{\boldsymbol{\theta}}$ in each stage t are given by the function $u\left(w_{j(i,t),t}, e_{i,t}|\underline{\boldsymbol{\theta}}\right)$ whose values exactly match the monetary payoff specifications of the game given $\chi = 1$. Instead, type $\overline{\boldsymbol{\theta}}$ workers receive payoffs given by

$$u\left(w_{j(i,t),t}, e_{i,t}|\overline{\boldsymbol{\theta}}\right) = \begin{cases} -1 & \text{if} \quad \left(w_{j(i,t),t}, e_{i,t}\right) = (\overline{w}, \underline{e}) \\ 0 & \text{otherwise} \end{cases}$$

Thus, selfish workers receive utility for only their monetary payoffs and reciprocal workers (type $\overline{\boldsymbol{\theta}}$) are penalized whenever they fail to reciprocate. Time discounting is ignored since experimental subjects are paid for all decisions at the end of the session.

Figure 4.5 shows the extensive form of one period of the mini-GEM for one matched pair of players with $\mathcal{W} = \{30, 100\}$ and $\mathcal{E} = \{1, 10\}$, using either the HRA or LRP payoffs. It is clear that $(\underline{w}, \underline{e})$ is the unique Nash equilibrium profile of either stage game when the worker is selfish ($\boldsymbol{\theta}_i = \underline{\boldsymbol{\theta}}$). However, if the firm believes that $\overline{e}$ will obtain with sufficiently high probability, then $\overline{w}$ is the best response. This depends on the firm's subjective probability that the worker is a reciprocating agent and the relative payoffs for cooperation ($\overline{w}$,$\overline{e}$), defection ($\overline{w}$,$\underline{e}$), and equilibrium ($\underline{w}$,$\underline{e}$).[12]

In each period of the repeated game, a subset of J workers is randomly selected to be paired with the J firms, while $I - J$ workers remain 'unemployed' for the period and receive zero payoff. A sequential equilibrium of this game is defined as a pairing of a strategy profile and a system of beliefs such that each agent is playing optimally at every information set, given the strategies of others and the system of

[12]The terms cooperation, defection, and equilibrium are taken from the prisoners' dilemma literature since this game is effectively a sequential prisoners' dilemma.

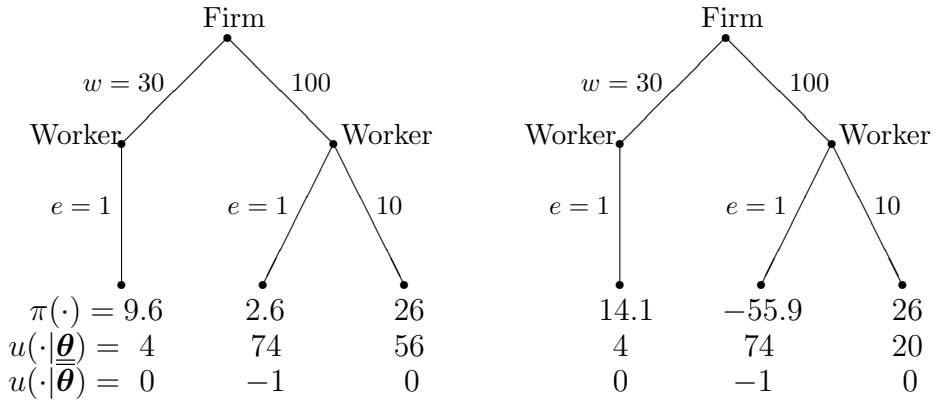


Figure 4.5: **(a)** The HRA mini-game, and **(b)** the LRP mini-game, assuming $\mathcal{W} = \{30, 100\}$ and $\mathcal{E} = \{1, 10\}$.

beliefs and the beliefs at each information set are derived from previous beliefs and action probabilities in accordance with Bayes' Law.

For notational simplicity, let $A = \pi(\underline{w}, \underline{e}) - \pi(\overline{w}, \underline{e})$, $B = \pi(\overline{w}, \overline{e}) - \pi(\overline{w}, \underline{e})$, $C = u(\overline{w}, \underline{e}) - u(\overline{w}, \overline{e})$, and $D = u(\overline{w}, \underline{e}) - u(\underline{w}, \underline{e})$. Here, A/B serves as a measure of how tempted the firms may be to gamble on the worker types by offering a high wage; if A/B is near zero, then offering a high wage is more appealing. Similarly, C/D is a measure of the workers' temptation to defect.

Under the assumption that worker types are independent and $e_{i,t}$ is only observed by $j(i,t)$, only the beliefs of firm $j(i,t)$ are changed by i's choice in period t. Let $p_{j,t}$ denote firm j's probability in period t that the randomly assigned worker will have type $\overline{\boldsymbol{\theta}}$. If $p_{j,t+1}$ is not substantially lower than $p_{j,t}$ when $e_{i(j,t),t} = \underline{e}$, then firm j's strategy in $t+1$ will be the same, regardless of $e_{i(j,t),t}$. If this is the case and $i(j,t)$ is selfish, then $\underline{e}$ will be chosen. Furthermore, tight restrictions on C and D are needed

to guarantee that selfish workers would indeed prefer to maintain a false reputation, giving the following result:.0

Proposition 4.1 *If worker decisions are anonymous and private, worker types are independent, and $C/D > 1/I$, then there does not exist a reputation equilibrium of the T-period repeated mini-GEM in which all workers choose $\overline{e}$ with probability 1 in all periods $t < T$ and all firms choose $\overline{w}$ in every period.*

The proof of this proposition follows from the arguments used in the proof of Proposition 4.2 which appears in the chapter appendix (Section 4.6.) Note that the equilibrium described in this proposition describes observed behavior in HRA. However, letting $\overline{w} = 100$, $\underline{w} = 30$, $\overline{e} = 10$, and $\underline{e} = 1$, the weak condition on C and D is satisfied for both HRA and LRP, so no pure strategy reputation equilibrium can exist in these two mini-games. Thus, for the experimental data to be explained by this model, either there exists a mixed strategy equilibrium that predicts frequent high effort levels, or firms believe that worker types are instead, highly correlated. Mixed strategy reputation equilibria are unlikely to exist, and if they do, are unlikely to generate high effort levels through many periods.[13] Therefore, a model of perceived

[13]To see this, let $q_{i',s}$ be the probability that agent i' chooses $\overline{e}$ in response to $\overline{w}$ in each period s, and let $q_{i',s} = 1$ if i' did not participate in period s. Note that if i' selects $\underline{e}$ in response to $\overline{w}$ in some period $s < T-1$, then the beliefs of the firms in period $T-1$ about agent i', (denoted $p_{i',T-1}$), will be zero. Given some worker $i \in \mathcal{I}$, define the average belief assigned to the other workers going into the final period to be

$$\bar{p}_{-i,T} = \frac{1}{I-1} \sum_{\substack{i' \in \mathcal{I}\setminus\{i\}: \\ p_{i',T-1}>0}} \frac{p_1}{p_1 + (1-p_1)\prod_{s=1}^{T-1} q_{i',s}}.$$

In period T, worker i will choose $\overline{e}$ in response to $\overline{w}$ only if

$$\frac{A}{B} - \frac{1}{I} < \frac{I-1}{I}\bar{p}_{-i,T} < \frac{A}{B}.$$

type correlation is examined.

4.4.2 The Model With Stereotypes

Although there exists experimental support for reputation equilibria in repeated games (for example, see Camerer & Weigelt [12], Neral & Ochs [76], and Andreoni & Miller [3]), the added assumption that firms view workers' types as highly correlated is apparently irrational. For example, Fehr *et al.* [35] and Fehr & Falk [32] find support for heterogenous player types in gift exchange markets, so rational, well-informed subjects should not expect homogeneity or strong correlation. On the other hand, McEvily *et al.* [69] show that subjects make inferences about the trustworthiness of opponents based on whether or not *different* opponents were trustworthy in early periods. Furthermore, when the group of opponents is chosen according to some unrelated criterion, the effect becomes even more pronounced. Thus, decision makers use past behavior to make inferences about the future behavior of others, especially when there is any reason to think those individuals are part of a group; subjects perceive correlation between their anonymous competitors.

The existing social psychology literature also supports the claim that people in social situations infer more correlation than is warranted – a phenomenon known as 'illusory correlation.' By design, the gift exchange market separates firms and workers into groups before the experiment begins, creating an initial identification of group

However, $\bar{p}_{-i,T}$ is a random variable determined not only by the results of earlier mixed strategy play by all other agents, but also by which agents are randomly selected to participate in each period. Therefore, it is unlikely to expect that any agent will maintain a false reputation in later periods. Knowing this, agents will have less incentive to develop reputations in early periods, unraveling the equilibrium.

membership among the subjects. A subject acting as a firm may see the group of firms as her 'ingroup' and the group of workers as the 'outgroup.' This partitioning leads naturally to categorical thinking (i.e., stereotyping) on the part of subjects, even if it is common knowledge that the outgroup is heterogeneous. As Pendry & Macrae [80, p. 926] note, "while true that outgroups are commonly perceived to be less heterogeneous in composition than ingroups, outgroup members nonetheless still display appreciable degrees of variability. Acknowledging the variability of social groups, however, is no antidote to stereotypical thinking." Thus, firms who are aware of seller heterogeneity may not act rationally on this information.

Although experimental psychology has established that perceivers are less likely to apply existing stereotypes when the actions of the perceived affect the outcomes of the perceiver (see, for example, Neuberg & Fiske [77] or Erber & Fiske [30]), this result disappears when cognitive resources are depleted by multiple task requirements. For example, Pendry & Macrae [79] find that subjects who are first primed with a stereotype of an unknown group and then learn a long list of attributes describing various group members later recall stereotype-inconsistent attributes from the list at least as frequently as stereotype-consistent attributes. However, when subjects are also required to remember an 8-digit number while learning the list of attributes, they recall stereotype-consistent information significantly more frequently than inconsistent information. Thus, as summarized by Macrae & Bodenhausen [66, p. 105], "judgement becomes more stereotypic under cognitive load."

Since firms in the gift exchange market are continually using cognitive resources

watching the market, devising strategies, and computing payoffs, they are indeed likely to think categorically about the group of workers even though firms' payoffs depend on the behavior of workers. Furthermore, Yzerbyt *et al.* [101] find that when subjects are exposed to information about a group member inconsistent with a formed stereotype, the stereotype shifts more dramatically when the subject is under a high cognitive load. This indicates that stereotypes formed by (busy) firms are likely to change noticeably when confronted with a sudden change in behavior by a single worker.

Finally, a study by Ruscher *et al.* [88] shows that when groups are perceived to be in competition rather than individuals (e.g., when firms see themselves as *collectively* in competition with firms), then subjects tend to pay more attention to stereotype-consistent information regarding individuals in the outgroup and stereotype-inconsistent information for members of their ingroup. Additionally, Rothgerber [87] and Brewer *et al.* [10] find that "competition has the potential to create stereotypes where none or very few existed before," as summarized by Corneille & Yzerbyt [26, p. 118]. This emphasizes that there need not be existing stereotypes of the group of workers for the firms to develop stereotypes when placed in this competitive market situation.

In total, the evidence from past experiments in economics and social psychology provides reasonable support for the assumption that firms hold a collective reputation of the workers rather than tracking individual histories, and this reputation is particularly sensitive to reputation-inconsistent behaviors by individual workers. Thus, a model in which firms believe workers' types are highly correlated, though

perhaps not rational, has an established psychological foundation. Taken to an extreme, the assumption of perfectly correlated types (so that any observation of $\underline{e}$ given $\overline{w}$ causes firms to believe that all agents are selfish) is in fact, sufficient to generate predictions highly consistent with the observed data in all three sessions. The following proposition formalizes this claim:.1

Proposition 4.2 *If the common knowledge prior* $p_1 = \Pr[\boldsymbol{\theta}_i = \overline{\boldsymbol{\theta}}]$ *is at least as large as* A/B*, then there exists a pure strategy reputation equilibrium of the* T*-period repeated mini-GEM with perfectly correlated types if and only if* $C/D \leq J/I$*. In this equilibrium, all firms offer* $\overline{w}$ *in every period, all selfish workers play* $\overline{e}$ *in every period* $t < T$ *and* $\underline{e}$ *in* T*, and all reciprocal workers play* $\overline{e}$ *in every period.*

As an example, let $\overline{w} = 100$, $\underline{w} = 30$, $\overline{e} = 10$, and $\underline{e} = 1$. In the HRA mini-game, $C/D \approx 0.257$ and $J/I = 2/3$. Thus, a pure strategy reputation equilibrium exists in this game. Since $A/B \approx 0.299$, if firms have at least a 3/10 prior probability that the workers are of type $\overline{\boldsymbol{\theta}}$, then cooperation should be observed by the firm in every period and by the worker in every period except the last. In the LRP mini-game, $C/D \approx 0.771$, which is greater than 2/3, so no pure strategy reputation equilibrium will exist in this mini-game. Also note that $A/B \approx 0.855$, so the restriction on beliefs would be much tighter even if J/I were greater than C/D.

In order to apply these theoretical results to the experimental data, the mini-GEM must be embedded back into the full GEM specification. While the appropriate values of $\underline{w}$ and $\underline{e}$ are clearly the stage game equilibrium values w^* and $e_{\min}$, the selection of $\overline{w}$ and $\overline{e}$ is less transparent. If the existence of the reputation equilibrium is

robust to this choice for a particular specification, then the result makes more general predictions about behavior under that specification. By Proposition 4.2, the pure strategy reputation equilibrium exists for some prior p_1 if and only if $A/B < 1$ and $C/D \leq J/I$.

Figure 4.6 displays the value of $1 - A/B$ for each of the possible high wage-effort pairs $(\overline{w}, \overline{e})$ for which $A/B < 1$ and $C/D \leq 2/3$ in HRA and similarly for LRP. The value $1 - A/B$ represents the measure of the interval $[A/B, 1]$ on which prior beliefs support the reputation equilibrium using $(\overline{w}, \overline{e})$ as the 'high' strategies.[14] Larger values of $1 - A/B$ suggest that a wider range of beliefs will generate a reputation equilibrium. Reputation equilibria cannot exist for pairs $(\overline{w}, \overline{e})$ at which the graph reports a value of zero. These graphs demonstrate that the HRA specification has many more wage-effort pairs capable of supporting a pure strategy reputation equilibrium than the LRP specification, and the condition on prior beliefs is much less restrictive in HRA. Therefore, existence of reputation equilibria in the HRA are significantly more robust to perturbations of initial beliefs and the choice of $\overline{w}$ and $\overline{e}$.

In generalizing the analysis of the mini-GEM to the full GEM specification, the concept of a reciprocal worker is less concrete, given the larger strategy space. In the previous literature, a positive correlation between effort and wages is taken as an indication of reciprocity. This can be operationalized by assuming that reciprocal workers play a known pure strategy that is monotone increasing in w. For example, when $\mathcal{E} = \{1, 2, \ldots, 10\}$ and $\mathcal{W} = \{5, 10, 15, \ldots\}$, assume that reciprocal workers play

[14] Note that the set of (w, e) that Pareto dominate $(\underline{w}, \underline{e})$ is given by $\{(w, e) \in W \times E : A/B < 1 \ \& \ C/D < 1\}$. Thus, the sets where $1 - A/B > 0$ (depicted in Figure 4.6) are strict subsets of this Pareto set.

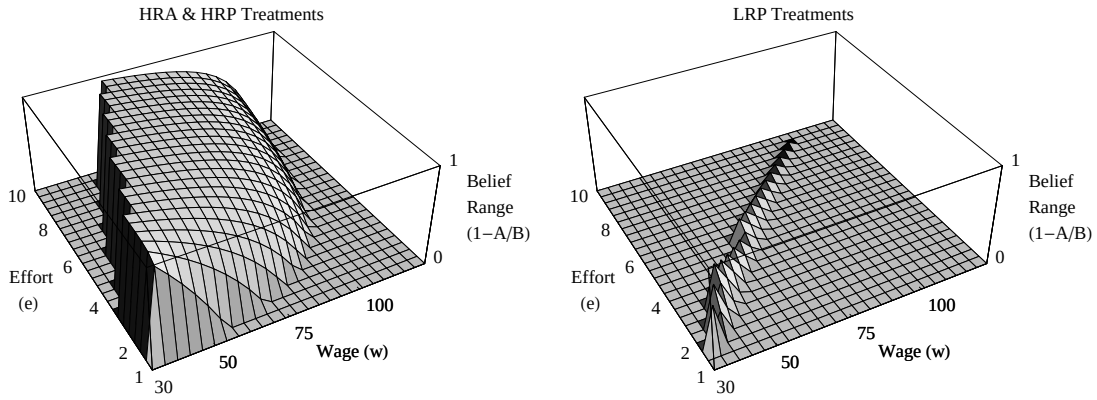


Figure 4.6: Size of the range of beliefs that support a pure-strategy reputation equilibrium in (a) the HRA and HRP treatments, and (b) the LRP treatment for various values of $\overline{w}$ and $\overline{e}$. A zero-sized belief range implies that no pure strategy reputation equilibrium exists.

a strategy weakly increasing in w given by

$$\left(\tilde{\chi}\left(w\right),\tilde{e}\left(w\right)\right)=\begin{cases}(0,1) & \text{if} \quad w<30\\ \left(1,\frac{w-20}{10}\right) & \text{if} \quad w\in\{30,40,\ldots\}\\ \left(1,\frac{w-15}{10}\right) & \text{if} \quad w\in\{35,45,\ldots\}\end{cases}$$

If a worker is known to be reciprocal, firms' profits are maximized at $\hat{w}=75$ in HRA and at $\hat{w}=35$ in LRP. If the firm is uncertain about the worker's type, $\hat{w}$ is weakly increasing in his belief. For example, if the firm believes the worker is reciprocal with probability 0.2, then $\hat{w}$ is 55 in HRA and 30 in LRP. In a pure strategy reputation equilibrium, the firms' beliefs do not update until after the final period, so the value of $\hat{w}$ associated with the prior belief p_1 can be sustained as a choice for $\overline{w}$ in the

repeated game. Thus, $\overline{w} = \hat{w}$ is an appropriate choice for the mini-GEM analysis.[15] For this particular example, the HRA again supports a reputation equilibrium while the LRP does not.

4.4.3 Application to Previous Experiments

Since the original FKR experiment in 1993, a variety of tests of the gift exchange market have been performed by various authors. Most frequently, 'no-loss' profit functions of the form $\pi(w, e) = (v - w)\, e$ are used, where v is a fixed constant. This results in a situation such as that depicted in panel (a) of Figure 4.6 in which many (w, e) pairs are capable of supporting a reputation equilibrium with a wide range of beliefs. The most common result in these experiments is high wages and effort in every period, with little or no indication of reversion to stage game equilibrium (e.g., Fehr et al. [33], Charness [16], Fehr & Falk [32], Charness et al. [17], Gächter and Falk [39], and Hannan et al. [44],) indicating that most or all workers are indeed reciprocal-minded. However, the data provided by Fehr et al. [35] show strong signs of a final-period crash under the 'no-loss' payoffs. In particular, 16 out of 26 workers choose $e_{\min}$ in the final period after high wages and effort are observed in previous periods.[16] Interestingly, wages remain high in one session despite frequent observations of $e_{\min}$ throughout the game, indicating that a model of perfect correlation is in fact, not

[15] If $\tilde{e}(w)$ has a relatively steep positive slope and $\mathcal{W}$ is unbounded above, then $\hat{w}$ may diverge to infinity. In this case, it is necessary to bound $\mathcal{W}$ by some $\overline{w}$.

[16] This fact is deduced from the data in the appendix of the paper.

appropriate for this data, although a weaker degree of correlation may suffice.[17]

Several experiments have done away with the 'no-loss' condition by using linear profits of the form $\pi(w, e) = ve - w$. This does not necessarily imply that reputation equilibria are eliminated. For example, panel (a) of Figure 4.7 shows the pairs and beliefs that support reputation equilibria with the payoffs $\pi(w, e) = 10 - w + 5e$ and $u(w, e) = 10 - e + 5w$ when $\mathcal{W} = \mathcal{E} = [0, 10]$, as in Brandts & Charness [9]. From Figure 4.7 it is clear that the environment supports reputation equilibria with correlation and in fact, the data show that high wages and effort move toward equilibrium on average in the final period.[18]

Riedl & Tyran [83] and Rigdon [84] also use quasilinear payoffs, and the set of wage-effort pairs sustainable as reputation equilibria with correlation is significantly smaller and larger prior beliefs are required, as demonstrated by panels (b) and (c) of Figure 4.7. In Riedl & Tyran, average wages are constant around 45 in all periods, with average efforts around 6 and several sessions featuring crashes in effort in the final period.[19] At the pair $(45, 6)$, firms' initial probability estimate that workers are reciprocal needs to be over 88 percent to support a reputation equilibrium. In Rigdon's experiment, effort decays to equilibrium early in the session, with wages following. Here, workers and firms were either unable to coordinate on a reputation equilibrium or beliefs were insufficient for such an equilibrium to exist.

[17] The fact that one worker was able to submit $e_{\min}$ in every period without destroying the group reputation indicates that it should not be an equilibrium for other selfish workers to submit higher efforts in every period, unless exactly two simultaneous occurrences of $e_{\min}$ are necessary to destroy the group reputation.

[18] Individual data is not presented, so it is unclear whether the group collectively chose slightly lower strategies or if the separation predicted by the group reputation model obtained.

[19] This fact is found in the data provided in the appendix of the paper.

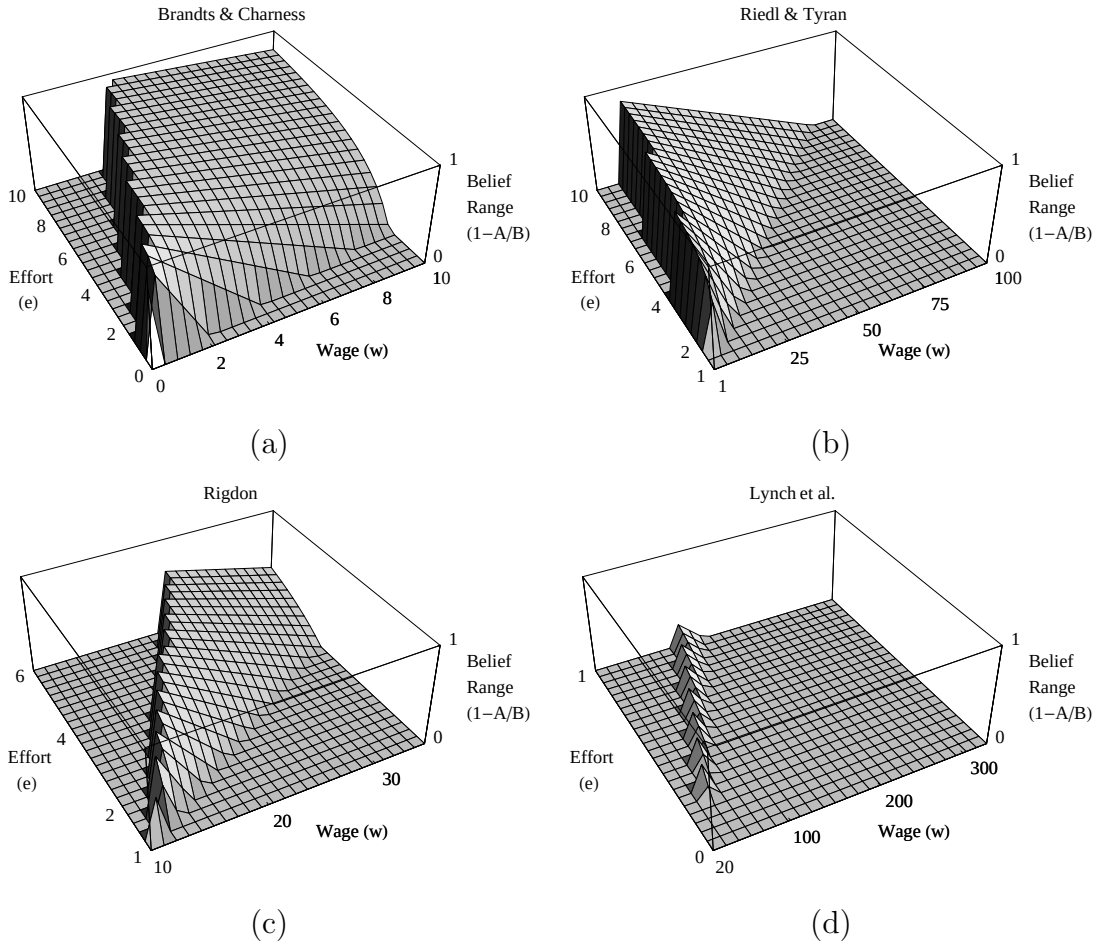


Figure 4.7: Size of the range of beliefs that support a pure-strategy reputation equilibrium in (a) the 'excess supply of labor' treatment of Brandts & Charness (2003), (b) Riedl & Tyran (2003), (c) Rigdon (2002), and (d) Lynch, Miller, Plott & Porter (2001), where 'effort' is a binary choice.

One previous experiment closely matches the conditions of the LRP environment. Lynch *et al.* [63] use a quasilinear environment with only two effort (quality) choices. Graphing the strategy pairs supporting a reputation equilibrium with correlation (panel (d) of Figure 4.7) demonstrates that at the high effort level ($\overline{e} = 1$), a reputation equilibrium can only be sustained for a very small number of wages, and only with very high prior beliefs. As predicted, wages and effort converge early to the stage game equilibrium. Lynch *et al.* conclude from their data that "a seller's demand depends not only upon his/her own 'reputation' for delivering [high quality], but also upon the market 'reputation' (p. 276)." Thus, the authors acknowledge that group reputations play an important role in these settings.

4.5 Conclusion

Although fair-minded, reciprocal behavior has been suggested as a solution to problems of contractual incompleteness, the findings of this chapter and previous experimental studies question the robustness and pervasiveness of reciprocal incentives. Inefficient equilibria are observed in some, but not all, laboratory environments.

The experimental data from this study indicate that repeated game concerns emerge in these settings despite the inability of agents to track individual reputations. A model of reputations with stereotyping recaptures the observed behavior nicely. Conditions are derived under which the problems of contractual incompleteness may be mitigated. The assumption that agents use categorical thinking (stereotyping) to assess the expected performance of others has a solid foundation in the social

psychology literature. The game theoretic construct of player types provides a natural way to introduce this concept in an economic domain.

This model introduces further testable hypotheses that warrant investigation. Empirical work on consumer behavior may confirm the existence of stereotypes. For example, do customers who have had bad experiences at one auto mechanic show a reduced demand for all mechanics? Experimental studies could be used to more directly isolate the stereotype-formation phenomenon. Scoring rules could be used to elicit beliefs from subjects who purchase from a sequence of sellers under moral hazard. Functional MRI studies may provide neurological evidence for stereotype formation and its economic consequences. A variety of tests could be constructed to further examine the validity and limits of the stereotyping assumption.

Although much is to be learned about the role of categorical thinking in economic contexts, it is clear that the introduction of such a model into the environment of incomplete contracting provides more explanatory power than either a simple model of fairness or a model of individual reputation building. Thus, the 'irrational' process of stereotype formation may indeed be a powerful force by which market failure is averted.

4.6 Appendix

Assume $p_1 \geq A/B$ and let p_t be the firms' shared belief that the workers are reciprocal.

Since types are correlated, if all workers play $q_t = \Pr\left[e_{i,t} = \overline{e} | w_{j(i,t),t} = \overline{w}\right]$ in

period t, then

$$p_{t+1} = \frac{p_t}{p_t + (1 - p_t)\, q_t}$$

when $e_{i,t} = \overline{e}$, which is always weakly greater than p_t. If the realization of any worker's strategy is $e_{i,t} = \underline{e}$, then $p_{t+1} = 0$. Thus, firms' beliefs weakly increase until a worker reveals that he is perfectly rational, at which point all firms know that all workers are selfish and play reverts to the fully selfish subgame perfect equilibrium.[20] In each period t, define $\mathcal{M}_t = \left\{ i \in \mathcal{I} : w_{j(i,t),t} = \overline{w} \right\}$ to be the set of workers receiving high wages in period t and note that $|\mathcal{M}_t| = J$ in every period in the proposed equilibrium.

Period T

All selfish workers play $e_{i,T} = \underline{e}$ and all reciprocal workers play $e_{i,T} = \overline{e}$ given $\overline{w}$.

If $p_T \geq A/B$, then each firm is willing to gamble on the workers by offering a high wage since

$$p_T \pi\left(\overline{w}, \overline{e}\right) + \left(1 - p_T\right) \pi\left(\overline{w}, \underline{e}\right) > \pi\left(\underline{w}, \underline{e}\right).$$

If $p_T = 0$, firms offer low wages.

Period $T - 1$

Selfish workers prefer to choose a mixed strategy

$$q_{T-1} = \Pr\left[e_{i,T-1} = \overline{e} | w_{j(i,T-1),T-1} = \overline{w} \right]$$

[20] Note that in the HRA specification, only one firm sees that effort choice of any given worker. However, if firm j observes $\underline{e}$ in period t, then he will offer $\underline{w}$ in period $t+1$, instantly signalling to the other firms that $p_{t+1} = 0$. Thus, as long as firms such that $p_{t+1} = 0$ offer wages first in period $t+1$, then the above result holds. If not, then other firms will update their beliefs to zero by period $t+2$. The former assumption is used here.

such that $p_T \geq A/B$ whenever the realization of all J workers' strategies in $\mathcal{M}_{T-1}$ is $e_{i,T-1} = \overline{e}$. The reader may verify that this occurs when

$$q_{T-1} \leq \left(\frac{p_{T-1}}{1-p_{T-1}} \frac{1-(A/B)}{A/B} \right)^{1/J}.$$

However, if $p_{T-1} \geq A/B$, then the right-hand side of this expression is weakly greater than 1, so the inequality is satisfied. Thus, regardless of the workers' strategies, firms will offer high wages in the final period whenever all workers in $\mathcal{M}_{T-1}$ provide high effort. Each worker $i \in \mathcal{M}_{T-1}$ has an expected payoff over the final two periods given by

$$q_{i,T-1} \left[u(\overline{w}, \overline{e}) + \frac{J}{I} \left(\begin{array}{c} \prod_{i' \in \mathcal{M}_{T-1} \setminus \{i\}} q_{i',T-1} u(\overline{w}, \underline{e}) \\ + \left(1 - \prod_{i' \in \mathcal{M}_{T-1} \setminus \{i\}} q_{i',T-1}\right) u(\underline{w}, \underline{e}) \end{array} \right) \right]$$
$$+ (1 - q_{i,T-1}) \left[u(\overline{w}, \underline{e}) + (J/I)\, u(\underline{w}, \underline{e}) \right].$$

Note that this payoff is increasing in $q_{i,T-1}$ if and only if

$$u(\overline{w}, \overline{e}) + \frac{J}{I} \left(\begin{array}{c} \prod_{i' \in \mathcal{M}_{T-1} \setminus \{i\}} q_{i',T-1} u(\overline{w}, \underline{e}) \\ + \left(1 - \prod_{i' \in \mathcal{M}_{T-1} \setminus \{i\}} q_{i',T-1}\right) u(\underline{w}, \underline{e}) \end{array} \right) - u(\overline{w}, \underline{e}) - \frac{J}{I} u(\underline{w}, \underline{e}) \geq 0,$$

which is true if and only if

$$\left(\prod_{i' \in \mathcal{M}_{T-1} \setminus \{i\}} q_{i',T-1} \right) \frac{J}{I} \geq \frac{C}{D}.$$

Note that if $q_{i',T-1} = 0$ for any $i' \in \mathcal{M}_{T-1} \setminus \{i\}$, then $q_{i,T-1} = 0$ is a best response regardless of C and D since i' is fully revealing the workers' type to the firms. If $C/D > J/I$, then worker i's payoff is necessarily decreasing in $q_{i,T-1}$, so i will choose low effort with certainty. Thus, when $C/D > J/I$, $q_{i,T-1} = 0$ for all $i \in \mathcal{M}_{T-1}$ must be true in any equilibrium. If $C/D \leq J/I$, then there exists an equilibrium in which $q_{i,T-1} = 1$ for all $i \in \mathcal{M}_{T-1}$.

Assume now that $p_{T-1} \geq A/B$, $C/D \leq J/I$, and firms know all types of workers will offer high effort with probability 1. Suppose each firm j offers a high wage with probability $r_{j,T-1}$, giving firm j an expected profit over the last two periods of

$$r_{j,T-1}\pi\left(\overline{w},\overline{e}\right) + \left(1 - r_{j,T-1}\right)\pi\left(\underline{w},\underline{e}\right) + p_{T-1}\pi\left(\overline{w},\overline{e}\right) + \left(1 - p_{T-1}\right)\pi\left(\overline{w},\underline{e}\right).$$

This is strictly increasing in $r_{j,T-1}$ (regardless of the strategies of the other firms since p_T is guaranteed to equal p_{T-1}), so every firm will choose to offer high wages with probability 1.

Period $T - k$

Assume $p_{T-k} \geq A/B$ for $k > 1$. As before, regardless of strategies chosen by the workers, if the realization of all workers' strategies in $\mathcal{M}_{T-k}$ is high effort, then $p_{T-k+1} \geq A/B$, and high wages and effort will be realized in period $T - k + 1$ with probability 1. Thus, the expected payoff over the last $k + 1$ periods to a worker

$i \in \mathcal{M}_{T-k}$ when each $i' \in \mathcal{M}_{T-k}$ plays $q_{i',T-k}$ is

$$q_{i,T-k}\left[u\left(\overline{w},\overline{e}\right)+\frac{J}{I}\left(\begin{array}{c}\prod_{i'\in\mathcal{M}_{T-k}\setminus\{i\}}q_{i',T-k}\left[\left(k-1\right)u\left(\overline{w},\overline{e}\right)+u\left(\overline{w},\underline{e}\right)\right]\\+\left(1-\prod_{i'\in\mathcal{M}_{T-k}\setminus\{i\}}q_{i',T-k}\right)\left[ku\left(\underline{w},\underline{e}\right)\right]\end{array}\right)\right]$$

$$+\left(1-q_{i,T-k}\right)\left[u\left(\overline{w},\underline{e}\right)+\left(J/I\right)ku\left(\underline{w},\underline{e}\right)\right].$$

This is increasing in $q_{i,T-k}$ if and only if

$$u\left(\overline{w},\overline{e}\right)+\frac{J}{I}\left(\begin{array}{c}\prod_{i'\in\mathcal{M}_{T-k}\setminus\{i\}}q_{i',T-k}\left[\left(k-1\right)u\left(\overline{w},\overline{e}\right)+u\left(\overline{w},\underline{e}\right)\right]\\+\left(1-\prod_{i'\in\mathcal{M}_{T-k}\setminus\{i\}}q_{i',T-k}\right)\left[ku\left(\underline{w},\underline{e}\right)\right]\end{array}\right)$$

$$-u\left(\overline{w},\underline{e}\right)-\left(J/I\right)ku\left(\underline{w},\underline{e}\right)>0.$$

As in period $T-1$, this expression is positive if and only if

$$\left(\prod_{i'\in\mathcal{M}_{T-k}\setminus\{i\}}q_{i',T-k}\right)\frac{J}{I}>\frac{C}{D},$$

so when $C/D > J/I$, only $q_{i,T-k} = 0$ for all $i \in \mathcal{M}_{T-k}$ can be an equilibrium strategy. If $C/D \leq J/I$, then there exist equilibria in which $q_{i,T-k} = 1$ for all $i \in \mathcal{M}_{T-k}$.

Assume now that $p_{T-k} \geq A/B$, $C/D \leq J/I$, and firms know all types of workers will offer high effort with probability 1. Suppose each firm j offers a high wage with

probability $r_{j,T-k}$, giving firm j and expected profit over the last $k+1$ periods of

$$r_{j,T-k}\pi\left(\overline{w},\overline{e}\right)+\left(1-r_{j,T-k}\right)\pi\left(\underline{w},\underline{e}\right)$$
$$+\left(k-1\right)\pi\left(\overline{w},\overline{e}\right)+p_{T-k}\pi\left(\overline{w},\overline{e}\right)+\left(\underline{e}-p_{T-k}\right)\pi\left(\overline{w},\underline{e}\right).$$

This is strictly increasing in $r_{j,T-k}$ (regardless of the strategies of the other firms), so every firm will choose to offer high wages with probability 1. Thus, when $p_1 \geq A/B$ and $C/D \leq J/I$, there exists a sequential equilibrium in which high wages and high effort obtain in every period before the last. If $C/D > J/I$, then there exists no such equilibrium.

Appendix A

Experiment Instructions

A.1 Instructions from Chapter 2

Experiment Overview

[All Treatments]

You are about to participate in an experiment in the economics of decision making. If you listen carefully and make good decisions, you could earn a considerable amount of money that will be paid to you in cash at the end of the experiment.

You have been paired with four other individuals in the room. In each period, each of you will be given an amount of cash, denominated in francs. This will be denoted by the letter E, for "Endowment." Your group will participate in a process in which the level of some good Y will be chosen by the collective decisions of your group members. Each of you has a value for Y that depends on how much Y is chosen by your group. This will be called $V(Y)$. You may not have the same value for a given level of Y as anyone else in your group. Furthermore, your decision in the process may result in you being charged some number of francs T. This number

T may be determined by your decision, the decisions of the others in your group, and the level of Y chosen by your group.

At the end of each period, you will be paid for the number of francs you have earned for the period, which is equal to your value for Y plus your initial endowment of francs minus the amount T that you paid. Mathematically, you will be paid based on

$$earnings\ in\ francs = V(Y) + E - T$$

At the end of the experiment, your total earnings in francs will be converted to dollars using a pre-specified exchange rate, which will be told to you by the experimenter.

The rules for the experiment are as follows. No talking or communicating with other participants. If you are using a computer, do not use any software other than that explicitly required by the experiment. Feel free to ask questions by raising your hand or signalling to the experimenter.

The process will now be explained in detail.

The Process

[Voluntary Contribution Treatment]

The process through which your group will choose the level of Y is as follows. Each of you will choose how many units of Y you would personally like to add to the total. For example, let person 1 choose $A1$, person 2 choose $A2$, and so on. At the

end of the period, the total level of Y will be given by

$$Y = A1 + A2 + A3 + A4 + A5$$

In each period, each person can add anywhere from 0 to 6 units of Y. You may add "partial" units, such as 3.45 units, for example.

Each unit of Y costs 100 francs in this experiment, so you will be charged 100 frances for each unit you add to the total. For example, if you are person 1 and you choose to add $A1 = 5$ units, then your payment will be $100 * A1 = 500$ francs. This payment corresponds to "T" from the above instructions. Therefore, $T = 500$ for person 1 in this example period.

$$T = 100 * Addition$$

At the end of the period, the total level of Y will be given by $Y = A1 + A2 + A3 + A4 + A5$ and you will earn $V(Y) + E - T$.

This process is computerized. In each period, you will input your decision into a computer program that will calculate Y and determine your payment and your earnings based on the decisions of your group. The computer will also keep track of your earnings and the results of all previous periods for your reference. The computer interface includes a tool called the "What-If Scenario Analyzer." This is a special calculator adapted to help you figure out how much you would earn in certain hypothetical scenarios. Feel free to use this tool to help you make decisions. If you

have any questions about the computer interface at any time, please raise your hand.

Questions?

[Proportional Tax Treatment]

The process through which your group will choose the level of Y is as follows. Each of you will choose how many units of Y you would personally like to add to the total. For example, let person 1 choose $A1$, person 2 choose $A2$, and so on. At the end of the period, the total level of Y will be given by

$$Y = A1 + A2 + A3 + A4 + A5$$

In each period, each person can add anywhere from 0 to 6 units of Y. You may add "partial" units, such as 3.45 units, for example.

Each unit of Y costs 100 francs in this experiment, and the total cost of Y will be split evenly among the 5 members of your group. For example, if the total level of Y is 10 units, then the total cost is $1,000$ francs. Dividing this cost evenly means each person must pay $1000/5 = 200$ francs. This payment corresponds to "T" from the above instructions. Therefore, $T = 200$ for each person in the group for this example period.

$$T = \frac{100 * Y}{5}$$

At the end of the period, the total level of Y will be given by $Y = A1 + A2 + A3 + A4 + A5$ and you will earn $V(Y) + E - T$.

This process is computerized. In each period, you will input your decision into

a computer program that will calculate Y and determine your payment and your earnings based on the decisions of your group. The computer will also keep track of your earnings and the results of all previous periods for your reference. The computer interface includes a tool called the "What-If Scenario Analyzer." This is a special calculator adapted to help you figure out how much you would earn in certain hypothetical scenarios. Feel free to use this tool to help you make decisions. If you have any questions about the computer interface at any time, please raise your hand.

Questions?

[Groves-Ledyard Treatment]

The process through which your group will choose the level of Y is as follows. Each of you will choose how many units of Y you would personally like to add to the total. For example, let person 1 choose $A1$, person 2 choose $A2$, and so on. At the end of the period, the total level of Y will be given by

$$Y = A1 + A2 + A3 + A4 + A5$$

In each period, each person can add anywhere from -4 to 6 units of Y. You may add "partial" units, such as 3.45 units, for example. A negative addition implies that you want to take units away from the total of Y.

The payment each person must make is dependent upon the average of the other 4 additions and the variance of those other 4 additions. Specifically, let S be the average of the other 4 messages and V be the variance of the other 4 messages. For

person 2, these would be calculated by the formulas

$$S_2 = \frac{A1 + A3 + A4 + A5}{4}$$
$$V_2 = \frac{(A1 - S_2)^2 + (A3 - S_2)^2 + (A4 - S_2)^2 + (A5 - S_2)^2}{3}$$

Roughly speaking, V measures how "spread out" the other 4 messages are.

The payment each person must make is as follows. First, we calculate the total cost of Y per person. Each unit of Y costs 100 francs, so the total cost is $100 * Y$ francs. This is then divided equally among the 5 people, so the cost per person is $(100 * Y)/5$. Second, we calculate the additional payment each person must make based on their S and V. This extra payment is determined by how far away your addition is from the average of everyone else's and on the variance of everyone else's messages. Putting it all together, if a person adds A units to the level of Y, then the payment each person must make is given by

$$T = \left(\frac{100\ Y}{5}\right) + \left(40\ \ (A - S)^2 - 50\ V\right)$$

Note that if V were large and $(A - S)$ were small, then the second half of your payment could be negative. This means that you may end up paying more or less than 1/5 of the total cost, and this all depends on your decision and the decisions of the others in your group.

At the end of the period, the total level of Y will be given by $Y = A1 + A2 + A3 + A4 + A5$ and you will earn $V(Y) + E - T$.

This process is computerized. In each period, you will input your decision into a computer program that will calculate Y and determine your payment and your earnings based on the decisions of your group. The computer will also keep track of your earnings and the results of all previous periods for your reference. The computer interface includes a tool called the "What-If Scenario Analyzer." This is a special calculator adapted to help you figure out how much you would earn in certain hypothetical scenarios. It allows you to enter values of S, V, and A just to see what you would earn if that was the real outcome of the period. Feel free to use this tool to help you make decisions. If you have any questions about the computer interface at any time, please raise your hand.

Questions?

[**Walker Mechanism Treatment**]

The process through which your group will choose the level of Y is as follows. Each of you will choose how many units of Y you would personally like to add to the total. For example, let person 1 choose $A1$, person 2 choose $A2$, and so on. At the end of the period, the total level of Y will be given by

$$Y = A1 + A2 + A3 + A4 + A5$$

In each period, each person can add anywhere from -10 to 15 units of Y. You may add "partial" units, such as 3.45 units, for example. A negative addition implies that you want to take units away from the total of Y.

The payment T each person must make is dependent upon the total level of Y chosen by the group as well as the individual proposals of two other people in the group. This is calculated by the following procedure. First, we calculate the total cost of Y per person. Each unit of Y costs 100 francs, so the total cost is $100 * Y$ francs. This is then divided equally among the 5 people, so the cost per person is $(100 * Y)/5$. Second, each person pays an additional amount based on the difference in proposals of two other players in the following way. Each of you has been given a Player number of the form PLR#. Take for example PLR3. Since PLR4 has the next-highest Player number, we refer to PLR4 as the player "above" PLR3 and we refer to PLR2 as the player "below" PLR3. Similarly, PLR3 is above PLR2 and PLR1 is below PLR2. Since there is no PLR0 or PLR6, we say that PLR5 is below PLR1 and PLR1 is above PLR5. The additional amount that a given person must pay is the proposal of the player below them minus the proposal of the player above them, all multiplied by the total of the proposals Y.

Combined, the total payment T that PLR3 must make, for example, will be given by

$$T = \frac{100Y}{5} + (PLR2\text{'s proposal} - PLR4\text{'s proposal}) * Y$$

The table at the end of these instructions gives the formula for T for all 5 players, to avoid confusion.

Recall that at the end of the period, the total level of Y will be given by $Y = A1 + A2 + A3 + A4 + A5$ and you will earn $V(Y) + E - T$.

This process is computerized. In each period, you will input your decision into

a computer program that will calculate Y and determine your payment and your earnings based on the decisions of your group. The computer will also keep track of your earnings and the results of all previous periods for your reference. The computer interface includes a tool called the "What-If Scenario Analyzer." This is a special calculator adapted to help you figure out how much you would earn in certain hypothetical scenarios. It allows you to enter values of S, V, and A just to see what you would earn if that was the real outcome of the period. Feel free to use this tool to help you make decisions. If you have any questions about the computer interface at any time, please raise your hand.

Questions?

Player #	Payment Calculation
1	$\frac{100Y}{5} + (PLR5\text{'s proposal} - PLR2\text{'s proposal}) * Y$
2	$\frac{100Y}{5} + (PLR1\text{'s proposal} - PLR3\text{'s proposal}) * Y$
3	$\frac{100Y}{5} + (PLR2\text{'s proposal} - PLR4\text{'s proposal}) * Y$
4	$\frac{100Y}{5} + (PLR3\text{'s proposal} - PLR5\text{'s proposal}) * Y$
5	$\frac{100Y}{5} + (PLR4\text{'s proposal} - PLR1\text{'s proposal}) * Y$

[cVCG Treatment]

All five people in your group value the level of Y by the formula $V(Y) = -CY^2 + DY$. However, people in your group may have different values of C and D. Check your private information slip for your values of C and D. In each period, all players in the group tell the central computer two *non-negative* numbers: A and B. The central computer will assume that each player's value for Y is given by $-AY^2 + BY$

and will then choose the level of Y that maximizes the combined value of all five players. Notice that if everyone submits $A = C$ and $B = D$ to the central computer, then the central computer will maximize the true combined value of all five players. On the other hand, if some players submit $A \neq C$ or $B \neq D$, then the computer will choose a different level of Y. Since A and B cannot be negative, if you enter negative values into the central computer, it will treat them as zeros.

In addition to choosing the level of Y, the computer will also choose how much each person must pay for the chosen level of Y. In total, every unit of Y costs 100 francs. Thus, the group must pay a *total* of at least $100 * Y$ francs. Each person will pay an equal share of the total cost plus an additional amount depending on the values of A and B chosen by all the members of the group. The additional amount each person must pay is given by the following process used by the computer:

1. Assume that everyone's $V(Y)$ is given by $-AY^2 + BY$ (the computer does *not* know the true C or D for any player.)

2. Calculate the Y that maximizes the sum of the reported values, net of costs. This is given by $Y^* = \frac{(B1+B2+B3+B4+B5)-100}{2*(A1+A2+A3+A4+A5)}$.

3. For each person, calculate the sum of everyone else's value in the group for the amount Y^*, minus $\frac{4}{5} * 100 * Y^*$, which represents how much of the cost of Y^* the other four must pay. Call this amount $U^{-i}(Y^*)$.

4. For each person, calculate the level of Y that maximizes the sum of *everyone else's value*, net of costs. We'll call this value Z^*. For example, for player 2,

this is given by $Z^* = \frac{(B1+B3+B4+B5)-100}{2*(A1+A3+A4+A5)}$.

5. For each person, calculate the sum of everyone else's value in the group for the amount Z^*, minus $\frac{4}{5} * 100 * Z^*$, which represents how much of the cost of Z^* the other four must pay. Call this amount $U^{-i}(Z^*)$.

6. Each person must pay $U^{-i}(Z^*) - U^{-i}(Y^*)$.

In words, each person must pay the difference between everyone else's value if that person weren't playing the game ($U^{-i}(Z^*)$) and everyone else's value if that person was playing the game ($U^{-i}(Y^*)$). This payment can be summarized by the phrase, "you must pay the change in everyone else's value that occurred because of your presence." Remember that the computer makes the above calculation based on the values of A and B that are reported and *not* on the actual values C and D.

In total, the amount each person must pay is $T = \frac{100*Y}{5} + [U^{-i}(Z^*) - U^{-i}(Y^*)]$. This process is guaranteed to collect at least enough money to cover the cost of Y. However, it is often the case that it collects more money than is needed to cover the cost of Y. If an excess is collected, the extra money is not refunded or used in any way.

At the end of the period, you receive your value for the level of Y chosen (which is $-CY^2 + DY$) plus E minus T.

This process is computerized. In each period, you will input your decision into a computer program that will calculate Y and determine your payment and your earnings based on the decisions of your group. The computer will also keep track of your earnings and the results of all previous periods for your reference. The

computer interface includes a tool called the "What-If Scenario Analyzer". This is a special calculator adapted to help you figure out how much you would earn in certain hypothetical scenarios. Feel free to use this tool to help you make decisions. If you have any questions about the computer interface at any time, please raise your hand.

Questions?

A.2 Instructions from Chapter 4

General Instructions

[All Treatments]

The experiment you will participate in is part of a research project used to analyze the decision behavior in markets. The instructions are simple, and if you read them carefully and make appropriate decisions, you can earn a considerable amount of money. At the end of the whole experiment, all the profits you have made by your decisions will be added up and paid to you in cash. The experiment you will participate in consists of two stages. In the first stage six of you act as buyers, and nine of you as sellers. In the second stage, the sellers will determine the value of the goods for the buyers (for details of the second stage see below). We have distributed two kinds of instructions – information for the buyers, and information for the sellers, respectively. This information is for private use only – you are not allowed to reveal this information to anyone. Furthermore, you will find at the end of these instructions a second sheet (your record sheet) that is used to document your decisions. Insert your buyer or seller number there, as well as your name and the

date.

Specific Instructions for Buyers

[HRA Treatment]

At the market, a good is traded, and each seller sells the same good. A seller can sell this good to any buyer, and a buyer can buy it from any seller. The market is organized in the following way: we open the market for a trading period (a "trading day"), and each trading day lasts three minutes. As a buyer you can offer a price that must be divisible by 5, for example, prices like 15, 60, 80, 275 are allowed, but prices like 48, 67, 124, 83 are not. These offers will be announced to the sellers by us over the telephone. The sellers will not know your identity, that is, your buyer number; they will only know the price offered. If a seller accepts your offer, all buyers are informed about this acceptance. In this case, an agreement is concluded, and the good is bought by you at the offered price. During each trading day you can buy one unit of the good. Therefore, a trading day ends for you when your offer is accepted. Note also that each seller can sell one unit of the good per day at most. If your offer is not accepted, you are free to change your offer, that is, to make a new offer. But the new price you offer must be higher than all the prices that have not been accepted. Each seller may accept an offer or not, but he cannot make a counteroffer.

After three minutes the day ends, and you cannot buy any more of the good. Then the second stage of thee experiment will be conducted. After this, a new trading day is opened. On the whole, there will be twelve trading days. In the second stage of the experiment, the seller who has sold the good to you on this day can fix the value

that the good will have for you. You as a buyer get a certain amount of experimental money (reselling price) from us for each unit you have bought. This reselling price is noted in the upper part of sheet 2. Your profit (measured in experimental money) is the difference between the reselling price and the price at which you have bought the good. If you bought the good for 20 and the reselling price is 30, you make a profit of $30 - 20 = 10$ (measured in experimental money). How much one unit of experimental money is worth to you depends on "your" seller. By the choice of a conversion rate, he decides how much real money you receive from us for one unit of experimental money. Which conversion rates he is allowed to choose are noted on the lower part of sheet 2. If he chooses, for example, the rate 0.5, you will get \$5 for 10 units of experimental money.

Sellers have two kinds of costs: production costs and decision costs. The latter are associated with the decision about the conversion rate. Production costs are noted in the middle of sheet 2, and decision costs on the lower part of sheet 2. As you can see from sheet 2, the higher the conversion rate "your" seller chooses, the greater are his decision costs. The profit of the sellers paid in dollars is given by the formula: profit = (price - production costs - decision costs). Suppose, for example, that you have bought the good for 75. The production costs of the seller are 60, and he chooses a conversion rate of 0.6 (which is associated with decision costs of 5), the profits of "your" seller are given by $75 - 60 - 5 = \$10$. Do you have any questions?

[HRP Treatment]

At the market, a good is traded, and each seller sells the same good. A seller can

sell this good to any buyer, and a buyer can buy it from any seller. The market is organized in the following way: we open the market for a trading period (a "trading day"), and each trading day lasts three minutes. As a buyer you can offer a price that must be divisible by 5, for example, prices like 15, 60, 80, 275 are allowed, but prices like 48, 67, 124, 83 are not. These offers will be announced to the sellers by us over the computer and displayed at the front of the room. The sellers will see your identity, that is, your buyer number, and the price you offered. If a seller accepts your offer, all buyers are informed about this acceptance. In this case, an agreement is concluded, and the good is bought by you at the offered price. During each trading day you can buy one unit of the good. Therefore, a trading day ends for you when your offer is accepted. Note also that each seller can sell one unit of the good per day at most. If your offer is not accepted, you are free to change your offer, that is, to make a new offer. But the new price you offer must be higher than all the prices that have not been accepted. Each seller may accept an offer or not, but he cannot make a counteroffer.

After each transaction, the seller who has sold the good to you on this day can fix the value that the good will have for you. You as a buyer get a certain amount of experimental money (reselling price) from us for each unit you have bought. This reselling price is noted in the upper part of your record sheet. Your profit (measured in experimental money) is the difference between the reselling price and the price at which you have bought the good. If you bought the good for 20 and the reselling price is 30, you make a profit of $30 - 20 = 10$ (measured in experimental money).

How much one unit of experimental money is worth to you depends on "your" seller. By the choice of a conversion rate, he decides how much money you receive from us for one unit of experimental money. Which conversion rates he is allowed to choose are noted on the lower part of the record sheet. If he chooses, for example, the rate 0.5, you will get \$5 for 10 units of experimental money. Note that the conversion rate choice and the ID number of the seller will be visible by all buyers.

Sellers have two kinds of costs: production costs and decision costs. The latter are associated with the decision about the conversion rate. Production costs are noted in the middle of sheet 2, and decision costs on the lower part of sheet 2. As you can see from sheet 2, the higher the conversion rate "your" seller chooses, the greater are his decision costs. The profit of the sellers paid in dollars is given by the formula: profit = (price - production costs - decision costs). Suppose, for example, that you have bought the good for 75. The production costs of the seller are 60, and he chooses a conversion rate of 0.6 (which is associated with decision costs of 5), the profits of "your" seller are given by $75 - 60 - 5 = \$10$.

At the end of the experiment, you (and the sellers) will be paid for your earnings at a rate of 12-to-1, meaning every 12 dollars you earn in the experiment is worth 1 actual dollar that will be paid to you at the end of the experiment.

After three minutes the trading day ends, and you cannot buy any more of the good. After this, a new trading day is opened. On the whole, there will be twelve trading days. Do you have any questions?

[LRP Treatment]

At the market, a good is traded, and each seller sells the same good. A seller can sell this good to any buyer, and a buyer can buy it from any seller. The market is organized in the following way: we open the market for a trading period (a "trading day"), and each trading day lasts three minutes. As a buyer you can offer a price that must be divisible by 5, for example, prices like 15, 60, 80, 275 are allowed, but prices like 48, 67, 124, 83 are not. These offers will be announced to the sellers by us over the computer and projected in their room along with the buyers' ID numbers. If a seller accepts your offer, all buyers are informed about this acceptance and the ID number of the seller who accepted it. At that point, the seller then chooses a number 'x' that affects how valuable the good is to you. Higher values of 'x' make the good more valuable, but cost the seller more money. This choice is then transmitted, along with the seller's ID number, back to this room and the transaction is concluded. The good is bought by you at the offered price and your value is affected by 'x'. You have to note the accepted price and the seller's choice of 'x' on your record sheet.

During each trading day you can buy one unit of the good. Therefore, a trading day ends for you when your offer is accepted. Note also that each seller can sell one unit of the good per day at most. If your offer is not accepted, you are free to change your offer, that is, to make a new offer. But the new price you offer must be higher than all the prices that have not been accepted. Each seller may accept an offer or not, but he cannot make a counteroffer. After three minutes the day ends, and you cannot buy any more of the good. After this, a new trading day is opened. On the whole, there will be twelve trading days.

Your profit is the fixed value of the good (which is shown on your record sheet,) multiplied by the number 'x' that your seller will determine, minus the price you pay to buy the good. Mathematically, your profit is given by the formula

$$\text{buyer profit} = \text{value} * x - \text{price}.$$

The seller's profit is the price they get for the good, minus a fixed production cost, minus an 'additional cost' based on their choice of 'x'. The formula for their profit is

$$\text{seller profit} = \text{price} - \text{production cost} - \text{additional cost}.$$

Your record sheet lists the value of the good to the buyers, the production cost to the sellers, and what the 'additional cost' for the seller is for each choice of 'x'. The higher the choice of 'x', the greater are the 'additional costs.'

If, for example, your value for the good is 400, the seller chooses 'x' to be 0.49, then your value times 'x' equals 196. If the price you paid was 175, then your profit is $196 - 175 = 21$. If the seller's production cost is 100 and his additional cost from choosing $x = 0.49$ is 6, then the seller's profit is $175 - 100 - 6 = 69$. This example appears on your record sheet.

At the end of the experiment, your earnings will be converted to dollars at a rate of _____. Do you have any questions?

Specific Instructions for Sellers

[HRA Treatment]

At the market, a good is traded, and each seller sells the same good. A seller can sell this good to any buyer, and a buyer can buy it from any seller. The market is organized in the following way: we open the market for a trading period (a "trading day"), and each trading day lasts three minutes. Every buyer can offer a price that will be relayed to us by telephone. We list these offers on the blackboard, and you can accept one of these offers. If, e.g., a price of 50 is offered and you as seller number 5 want to accept this offer, you just say: "Number 5 sells for 50." In this case, the transaction is concluded. The good is sold to the buyer who made the offer of 50. The buyer will not know your identity. He will just know that his offer is accepted. You have to note your accepted price on sheet 2.

You can sell one unit of the good on each trading day. Therefore, the trading day ends for you after the acceptance of an offer. Note also that each buyer can buy, at most, one unit of the good per trading day. Each seller may accept an offer or not, but the sellers cannot make counteroffers. After three minutes the trading day ends, and the second stage of the experiment is conducted. After this, a new trading day is opened. In total there will be twelve trading days. At the second stage of the experiment, you can fix the value the good will have for the buyers. Buyers receive a certain amount of experimental money (reselling price) from us for each unit that they have bought. This reselling price is noted in the middle of sheet 2.

The profit of a buyer (measured in experimental money) is the difference between the reselling price and the price at which he has bought the good from you. If "your" buyer has bought the good for 20 and the reselling price is 30, he makes a

profit of $30 - 20 = 10$ (measured in experimental money.) How much one unit of experimental money is worth for "your" buyer depends on you. By the choice of a conversion rate, you decide how much real money "your" buyer gets from us for one unit of experimental money. If you choose, e.g., the rate 0.5, your buyer gets \$5 for 10 units of experimental money. Which conversion rates you are allowed to choose, is noted on the lower part of sheet 2. You have to write down your decision on the upper part of sheet 2. Do not announce your decision publicly.

You, as a seller, have two kinds of costs: production costs and "decision costs." The latter are associated with your decision about the conversion rate. Of course, you incur costs only in the case of a deal. If you do not trade on a certain day, your costs are zero for this day. Production costs are noted on the upper part of sheet 2. Decision costs depend on your choice of the conversion rate. The higher the conversion rate you decide to give "your" buyer, the greater are your decision costs. The costs, which are associated with the conversion rate, are noted in the lower part of sheet 2.

Your profit paid in dollars is given by the formula profit = price - production costs - decision costs. If, for example, you sell your good for 75, while your production costs are 60, and you choose a conversion rate of 0.6 which leads to a decision cost of 5, your profit is given by $75 - 60 - 5 = \$10$. Do you have any questions?

[HRP Treatment]

At the market, a good is traded, and each seller sells the same good. A seller can sell this good to any buyer, and a buyer can buy it from any seller. The market is

organized in the following way: we open the market for a trading period (a "trading day"), and each trading day lasts three minutes. Every buyer can offer a price that will be relayed to us by computer and displayed at the front of the room. We list these offers on the screen, along with buyer ID numbers, and you can accept one of these offers. If, e.g., a price of 50 is offered and you as seller number 5 want to accept this offer, you just say: "5 sells for 50." In this case, the transaction is concluded. The good is sold to the buyer who made the offer of 50. The buyers will see your decision and your ID number. You have to note your accepted price on the record sheet.

You can sell one unit of the good on each trading day. Therefore, the trading day ends for you after the acceptance of an offer. Note also that each buyer can buy, at most, one unit of the good per trading day. Each seller may accept an offer or not, but the sellers cannot make counteroffers.

After each transaction, you can fix the value the good will have for the buyers. Buyers receive a certain amount of experimental money (reselling price) from us for each unit that they have bought. This reselling price is noted in the middle of the record sheet. The profit of a buyer (measured in experimental money) is the difference between the reselling price and the price at which he has bought the good from you. If "your" buyer has bought the good for 20 and the reselling price is 30, he makes a profit of $30 - 20 = 10$ (measured in experimental money.) How much one unit of experimental money is worth for "your" buyer depends on you. By the choice of a conversion rate, you decide how much money "your" buyer gets from us for one unit

of experimental money. If you choose, e.g., the rate 0.5, your buyer gets \$5 for 10 units of experimental money. Which conversion rates you are allowed to choose, is noted on the lower part of the record sheet. You have to write down your decision on the upper part of the record sheet.

You, as a seller, have two kinds of costs: production costs and "decision costs." The latter are associated with your decision about the conversion rate. Of course, you incur costs only in the case of a deal. If you do not trade on a certain day, your costs are zero for this day. Production costs are noted on the upper part of the record sheet. Decision costs depend on your choice of the conversion rate. The higher the conversion rate you decide to give "your" buyer, the greater are your decision costs. The costs, which are associated with the conversion rate, are noted in the lower part of the record sheet.

Your profit is given by the formula profit = price - production costs - decision costs. If, for example, you sell your good for 75, while your production costs are 60, and you choose a conversion rate of 0.6 which leads to a decision cost of 5, your profit is given by $75 - 60 - 5 = \$10$.

At the end of the experiment, you (and the buyers) will be paid for your earnings at a rate of 12-to-1, meaning every 12 dollars you earn in the experiment is worth 1 actual dollar that will be paid to you at the end of the experiment.

After three minutes the trading day ends, and the second stage of the experiment is conducted. After this, a new trading day is opened. In total there will be twelve trading days. Do you have any questions?

[LRP Treatment]

At the market, a good is traded, and each seller sells the same good. A seller can sell this good to any buyer, and a buyer can buy it from any seller. The market is organized in the following way: we open the market for a trading period (a "trading day"), and each trading day lasts three minutes. Every buyer can offer a price (in multiples of 5) that will be relayed to us by computer and projected at the front of the room, along with the ID number of the buyer. We list these offers on the screen, and you can accept one of these offers. If, e.g., a price of 50 is offered and you as seller number 5 want to accept this offer, you just say: "Seller 5 sells for 50." At that point, you then choose a number 'x' that affects how valuable the good is to the buyer. Higher values of 'x' make the good more valuable to the buyer, but cost you more money. This choice is then transmitted, along with the your ID number, back to the buyers and the transaction is concluded. The good is sold by you at the offered price and you pay an additional cost for your choice of 'x'. You have to note your accepted price and your choice of 'x' on your record sheet.

During each trading day you can sell one unit of the good. Therefore, a trading day ends for you when you accept an offer. Note also that each buyer can buy one unit of the good per day at most. If a buyer's offer is not accepted, the buyer is free to change his offer, but the new price must be higher than all the prices that have not been accepted. Each seller may accept an offer or not, but you cannot make a counteroffer. After three minutes the day ends, and you cannot accept any offers. After this, a new trading day is opened. On the whole, there will be twelve trading

days.

The profit of a buyer is the buyer's fixed value of the good (which is shown on your record sheet,) multiplied by the number 'x' that you will determine, minus the price they pay to buy the good. Mathematically, the buyer's profit is given by the formula

$$\text{buyer profit} = \text{value} * x - \text{price}.$$

Your profit is the price you get for the good, minus a fixed production cost, minus an 'additional cost' based on your choice of 'x'. The formula for your profit is

$$\text{seller profit} = \text{price} - \text{production cost} - \text{additional cost}.$$

In the second stage of the experiment, your job is to choose the value 'x', and your 'additional cost' depends on this decision. Your record sheet lists the value of the good to the buyers, your production cost, and what the 'additional cost' is for each choice of 'x'. The higher the choice of 'x', the greater are your 'additional costs.

If, for example, you sell your good for 175, while your production costs are 100, and you choose the value of 'x' as 0.49 which leads to a decision cost of 6, your profit is given by $175 - 100 - 6 = 69$. This example appears on your record sheet.

At the end of the experiment, your earnings will be converted to dollars at a rate of _____. Do you have any questions?

Buyer Number: Name: Date & Time:

Period	Selling Price (1)	Price (2)	Exp. Money Profit (3)=(1)-(2)	Conv. Rate "CR" (4)	Profit (5)=(3)x(4)	Total Profit
Example	405	205	200	0.5	100	100
1	126					
2	126					
3	126					
4	126					
5	126					
6	126					
7	126					
8	126					
9	126					
10	126					
11	126					
12	126					

Production Cost of Sellers: 26

Seller Profit = Price - Production Cost - Decision Cost "DC"

CR	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
DC	0	1	2	4	6	8	10	12	15	18

Figure A.1: Record Sheet: Buyers, HRA and HRP Treatments

Seller Number: Name: Date & Time:

Period	Conversion Rate "CR"	Price (1)	Production Cost (2)	Decision Cost "DC"	Profit (4)=(1)-(2)-(3)	Total Profit
Example	0.6	175	100	5	70	70
1			26			
2			26			
3			26			
4			26			
5			26			
6			26			
7			26			
8			26			
9			26			
10			26			
11			26			
12			26			

Reselling Price of Buyers: 126

Buyer Profit = (Reselling Price - Price) x Conversion Rate "CR"

CR	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
DC	0	1	2	4	6	8	10	12	15	18

Figure A.2: Record Sheet: Sellers, HRA and HRP Treatments

Buyer Number: Name: Date & Time:

Period	Value (1)	"x" (2)	Value * "x" (3)=(1)*(2)	Price (4)	Profit (5)=(3)-(4)	Total Profit
Example	400	0.49	196	175	21	21
1	126					
2	126					
3	126					
4	126					
5	126					
6	126					
7	126					
8	126					
9	126					
10	126					
11	126					
12	126					

Production Cost of Sellers: 26

Seller Profit = Price - Production Cost - Additional Cost "AC"

AC	0	3	6	12	18	24	30	36	45	54
x	0.35	0.42	0.49	0.57	0.64	0.71	0.78	0.86	0.93	1.00

Figure A.3: Record Sheet: Buyers, LRP Treatment

Seller Number: Name: Date & Time:

Period	Price (1)	Prod'n Cost (2)	Additional Cost "AC" (3)	"x" (4)	Profit (4)=(1)-(2)-(3)	Total Profit
Example	175	100	6	0.49	69	69
1		26				
2		26				
3		26				
4		26				
5		26				
6		26				
7		26				
8		26				
9		26				
10		26				
11		26				
12		26				

Value to Buyers: 126
Buyer Profit =Value * x - Price

AC	0	3	6	12	18	24	30	36	45	54
x	0.35	0.42	0.49	0.57	0.64	0.71	0.78	0.86	0.93	1.00

Figure A.4: Record Sheet: Sellers, LRP Treatment

Bibliography

[1] Akerlof, G. Labor contracts as partial gift exchange. *Quart. J. Econ.*, 97 (1982), 543–569.

[2] Akerlof, G. Gift exchange and efficiency-wage theory: Four views. *AEA Papers and Proceedings*, 74 (1984), 79–83.

[3] Andreoni, J., Miller, J. H. Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *Econ. J.*, 103 (1993), 570–585.

[4] Attiyeh, G., Franciosi, R., Isaac, R. M. Experiments with the pivot process for providing public goods. *Public Choice*, 102 (2000), 95–114.

[5] Auster, R. D. The invariably stable cobweb model. *Rev. Econ. Stud.*, 38 (1971), 117–121.

[6] Bear, D. V. T. The matrix multiplier and distributed lags. *Econometrica*, 31 (1963), 514–529.

[7] Bear, D. V. T. Distributed lags and economic theory. *Rev. Econ. Stud.*, 33 (1966), 235–243.

[8] Bernheim, B. D., Whinston, M. D. Incomplete contracts and strategic ambiguity. *Amer. Econ. Rev.*, 88 (1998), 902–932.

[9] Brandts, J., Charness, G. Do labour market conditions affect gift exchange? Some experimental evidence (2004). Forthcoming, Econ. J.

[10] Brewer, M. B., Weber, J. G., Carini, B. Person memory in intergroup contexts: Categorization versus individuation. *J. Personality Soc. Psych.*, 69 (1995), 29–40.

[11] Brown, G. W. Iterative solution of games by fictitious play. In *Activity Analysis of Production and Allocation*, T. C. Koopmans, ed. New York: Wiley (1951).

[12] Camerer, C., Weigelt, K. Experimental tests of a sequential squilibrium reputation model. *Econometrica*, 61 (1988), 1–36.

[13] Carlson, J. A. The stability of an experimental market with supply-response lag. *Southern Econ. J.*, 33 (1967), 305–321.

[14] Cason, T., Saijo, T., Sjostrom, T., Yamato, T. Secure implementation experiments: Do strategy-proof mechanisms really work? (2003). California Institute of Technology Department of Humanities and Social Sciences, Working Paper.

[15] Champsaur, P., Roberts, D. J., Rosenthal, R. W. On cores of economies with public goods. *Int. Econ. Rev.*, 16 (1975), 751–764.

[16] Charness, G. Attribution and reciprocity in a simulated labor market: An experimental investigation (1998). University of California, Santa Barbara Department of Economics Working Paper.

[17] Charness, G., Frechette, G., Kagel, J. How robust is laboratory gift exchange? (2004). Forthcoming, Exper. Econ.

[18] Chen, Y. Dynamic stability of Nash-efficient public goods mechanisms: Reconciling theory and experiments (2004). Forthcoming, *Experimental Business Research, Volume II*, R. Zwick and A. Rapoport, Eds. Kluwer Academic Publishers: Norwell, MA.

[19] Chen, Y. Incentive compatible mechanisms for pure public goods: A survey of experimental research. In *The Handbook of Experimental Economics Results*, C. R. Plott, V. Smith, eds. Amsterdam: Elsevier Press (2004).

[20] Chen, Y., Gazzale, R. When does learning in games generate convergence to Nash equilibria? The role of supermodularity in an experimental setting (2004). Forthcoming, *Amer. Econ. Rev.*

[21] Chen, Y., Plott, C. R. The Groves-Ledyard mechanism: An experimental study of institutional design. *J. Pub. Econ.*, 59 (1996), 335–364.

[22] Chen, Y., Tang, F.-F. Learning and incentive-compatible mechanisms for public goods provision: An experimantal study. *J. Polit. Economy*, 106 (1998), 633–662.

[23] Clark, K., Sefton, M. The sequential prisoner's dilemma: Evidence on reciprocation. *Econ. J.*, 111 (2001), 51–68.

[24] Clarke, E. Multipart pricing of public goods. *Public Choice*, 2 (1971), 19–33.

[25] Conley, J. P. Convergence theorems on the core of a public goods economy: Sufficient conditions. *J. Econ. Theory*, 62 (1994), 161–185.

[26] Corneille, O., Yzerbyt, V. Dependence and the formation of stereotyped beliefs about groups: from interpersonal to intergroup perception. In *Stereotypes as Explanations: The formation of meaningful beliefs about social groups*, C. McGarty, V. Y. Yzerbyt, R. Spears, eds. Cambridge: Cambridge University Press (2002).

[27] de Trenqualye, P. Stable implementation of Lindahl allocations. *Econ. Letters*, 29 (1989), 291–294.

[28] Efron, B., Tibshirani, R. *An Introduction to the Bootstrap*. New York: Chapman & Hall (1993).

[29] Engelmann, D., Ortmann, A. The robustness of laboratory gift exchange: A reconsideration (2002). CERN-EI Working Paper.

[30] Erber, R., Fiske, S. T. Outcome dependency and attention to inconsistent information. *J. Personality Soc. Psych.*, 47 (1984), 709–726.

[31] Eswaran, M., Kotwal, A. The moral hazard of budget-breaking. *RAND J. Econ.*, 15 (1984), 578–581.

[32] Fehr, E., Falk, A. Wage rigidity in a competitive incomplete contract market. *J. Polit. Economy*, 107 (1999), 106–134.

[33] Fehr, E., Kirchler, E., Weichbold, A., Gachter, S. When social norms overpower competition – Gift exchange in experimental labor markets. *J. Lab. Econ.*, 16 (1998), 324–351.

[34] Fehr, E., Kirchsteiger, G., Riedl, A. Does fairness prevent market clearing? An experimental investigation. *Quart. J. Econ.*, 108 (1993), 437–459.

[35] Fehr, E., Kirchsteiger, G., Riedl, A. Gift exchange and reciprocity in competitive experimental markets. *Europ. Econ. Rev.*, 42 (1998), 1–34.

[36] Foley, D. K. Lindahl's solution and the core of an economy with public goods. *Econometrica*, 38 (1970), 66–72.

[37] Fryer, R., Jackson, M. O. Categorical cognition: A psychological model of categories and identification in decision making (2003). California Institute of Technology working paper.

[38] Gabay, D., Moulin, H. On the uniqueness and stability of Nash-equilibria in noncooperative games. In *Applied Stochastic Control in Econometrics and Management Science*, A. Bensoussan, P. Kleindorfer, C. S. Tapiero, eds. Amsterdam: North-Holland (1980).

[39] Gachter, S., Falk, A. Reputation and reciprocity: Consequences for the labour relation. *Scand. J. Econ.*, 104 (2002), 1–27.

[40] Gibbard, A. Manipulation of voting schemes: A general result. *Econometrica*, 41 (1973), 587–602.

[41] Green, J. R., Laffont, J.-J. *Incentives in public decision-making.* Amsterdam: North-Holland (1979).

[42] Groves, T. Incentives in teams. *Econometrica*, 41 (1973), 617–633.

[43] Groves, T., Ledyard, J. O. Optimal allocation of public goods: A solution to the 'free-rider' problem. *Econometrica*, 45 (1977), 783–809.

[44] Hannan, R. L., Kagel, J. H., Moser, D. V. Partial gift exchange in an experimental labor market: Impact of subject ppulation dfferences, poductivity dfferences, and efort rquests on bhavior. *J. Lab. Econ.*, 20 (2002), 923–951.

[45] Harriff, R., Bear, D. V. T., Conlisk, J. Stability and speed of adjustment under retiming of lags. *Econometrica*, 48 (1980).

[46] Healy, P. J. Learning dynamics for mechanism design: An experimental comparison of public goods mechanisms (2005). Forthcoming, J. Econ. Theory.

[47] Holmstrom, B. Moral hazard in teams. *Bell J. Econ.*, 13 (1982), 324–340.

[48] Hurwicz, L. On informationally decentralized systems. In *Decision and Organization: A Volume in Honor of Jacob Marschak*, C. B. McGuire, R. Radner, eds. Amsterdam: North-Holland (1972).

[49] Hurwicz, L. On allocations attainable through Nash equilibria. *J. Econ. Theory*, 21 (1979), 140–165.

[50] Hurwicz, L. Economic design, adjustment processes, mechanisms, and institutions. *Econ. Design*, 1 (1994), 1–14.

[51] Isaac, R. M., McCue, K. F., Plott, C. R. Public goods provision in an experimental environment. *J. Pub. Econ*, 26 (1985), 51–74.

[52] Jackson, M. O., Palfrey, T. R. Efficiency and voluntary implementation in markets with repeated pairwise bargaining. *Econometrica*, 66 (1998), 1353–1388.

[53] Jackson, M. O., Palfrey, T. R. Voluntary implementation. *J. Econ. Theory*, 98 (2001), 1–25.

[54] Kandori, M. Social norms and community enforcement. *Rev. Econ. Stud.*, 59 (1992), 63–80.

[55] Kawagoe, T., Mori, T. Can the pivotal mechanism induce truth-telling? An experimental study. *Public Choice*, 108 (2001), 331–354.

[56] Kim, T. *Stability problems in the implementation of Lindahl allocations*. Ph.D. thesis, University of Minnesota (1987).

[57] Klein, B., Leffler, K. B. The role of market forces in assuring contractual performance. *J. Polit. Economy*, 89 (1981).

[58] Kolm, S.-C. *La Valeur Publique*. Paris: Dunod (1970).

[59] Kreps, D., Milgrom, P., Roberts, J., Wilson, R. Rational cooperation in the finitely repeated prisoners' dilemma. *J. Econ. Theory*, 27 (1982), 245–252.

[60] Kreps, D., Wilson, R. Reputation and imperfect information. *J. Econ. Theory*, 27 (1982), 253–279.

[61] Ledyard, J. O. Public goods: A survey of experimental research. In *Handbook of Experimental Economics*, J. Kagel, A. Roth, eds. Princeton, NJ: Princeton University Press (1995).

[62] Ledyard, J. O., Roberts, J. On the incentive problem with public goods (1975). Northwestern University Center for Mathematical Studies in Economics and Management Science, Discussion Paper No. 116.

[63] Lynch, M., Miller, R., Plott, C. R., Porter, R. Product quality, informational efficiency, and regulations in experimental markets. In *Information, Finance and General Equilibrium: Collected Papers on the Experimental Foundations of Economics and Political Sciences*, C. R. Plott, ed., volume 3. Northampton, MA: Elgar (2001).

[64] Ma, C.-T., Moore, J., Turnbull, S. Stopping agents from cheating. *J. Econ. Theory*, 46 (1988), 355–372.

[65] MacLeod, W. B., Malcomson, J. M. Implicit contracts, incentive compatibility, and involuntary unemployment. *Econometrica*, 57 (1989), 447–480.

[66] Macrae, C. N., Bodenhausen, G. V. Social cognition: Thinking categorically about others. *Annual Rev. Pscyh.*, 51 (2000), 93–120.

[67] Maskin, E. Nash equilibrium and welfare optimality. *Rev. Econ. Stud.*, 66 (1999), 23–38.

[68] Maskin, E., Moore, J. Implementation and renegotiation. *Rev. Econ. Stud.*, 66 (1999), 39–56.

[69] McEvily, B., Weber, R. A., Bicchieri, C., Ho, V. Can groups be trusted? An experimental study of collective trust (2002). Carnegie Mellon University Behavioral Decision Research, Working Paper 308.

[70] McGarty, C., Yzerbyt, V. Y., Spears, R. Social, cultural and cognitive factors in stereotype formation. In *Stereotypes as Explanations: The formation of meaningful beliefs about social groups*, C. McGarty, V. Y. Yzerbyt, R. Spears, eds. Cambridge: Cambridge University Press (2002).

[71] Milgrom, P., Roberts, J. Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica*, 58 (1990), 1255–1277.

[72] Milgrom, P., Roberts, J. Adaptive and sophisticated learning in normal form games. *Games Econ. Behav.*, 3 (1991), 82–100.

[73] Milleron, J.-C. Theory of value with public goods: A survey article. *J. Econ. Theory*, 5 (1972), 419–477.

[74] Muench, T. J. The core and the Lindahl equilibrium of an economy with a public good: An example. *J. Econ. Theory*, 4 (1972), 241–255.

[75] Murata, Y. *Mathematics for stability and optimization of economic systems.* Academic Press (1977).

[76] Neral, J., Ochs, J. The sequential equilibrium theory of reputation building: A further test. *Econometrica*, 60 (1992), 1151–1169.

[77] Neuberg, S. L., Fiske, S. T. Motivational influences on impression formation: Outcome dependency, accuracy-driven attention, and individuating processes. *J. Personality Soc. Psyc.*, 53 (1987), 431–444.

[78] Ortega, J. M., Rheinboldt, W. C. *Iterative solution of nonlinear equations in several variables.* New York: Academic Press (1970).

[79] Pendry, L. F., Macrae, C. N. Stereotypes and mental life: The case of the motivated but thwarted tactician. *J. Exper. Psych.*, 30 (1994), 303–325.

[80] Pendry, L. F., Macrae, C. N. Cognitive load and person memory: The role of perceived group variability. *Europ. J. Soc. Pscyh.*, 29 (1999), 925–942.

[81] Perez-Nievas, M. *Interim efficient allocation mechanisms.* Ph.D. thesis, Universidad Carlos III de Madrid (2002).

[82] Richter, D. K. The core of a public goods economy. *Int. Econ. Rev.*, 15 (1974), 131–142.

[83] Riedl, A., Tyran, J.-R. Tax liability side equivalence in gift-exchange labor markets (2004). Forthcoming, J. Public Econ.

[84] Rigdon, M. Efficiency wages in an experimental labor market. *Proc. Nat. Acad. Sci.*, 99 (2002), 13348–13351.

[85] Roberts, K. The characterization of implementable social choice rules. In *Aggregation and Revelation of Preferences*, J.-J. Laffont, ed. Amsterdam: North-Holland (1979).

[86] Rosen, J. B. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica*, 33 (1965), 520–533.

[87] Rothgerber, H. External intergroup threat as an antecedent to perceptions of in-group and out-group homogeneity. *J. Personality Soc. Psych.*, 73 (1997), 1206–1211.

[88] Ruscher, J. B., Fiske, S. T., Miki, H., van Manen, S. F. Individuating processes in competition: Interpersonal versus intergroup. *Personality Soc. Psych. Bull.*, 17 (1991), 595–605.

[89] Saijo, T. Incentive compatibility and individual rationality in public good economies. *J. Econ. Theory*, 55 (1991), 203–212.

[90] Samuelson, P. A. The pure theory of public expenditure. *Rev. Econ. Statist.*, 36 (1954), 387–389.

[91] Satterthwaite, M. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *J. Econ. Theory*, 10 (1975), 187–217.

[92] Selten, R. Axiomatic characterization of the quadratic scoring rule. *Exper. Econ.*, 1 (1998), 43–61.

[93] Shapley, L. S. Some topics in two-person games. In *Advances in Game Theory*, M. Dresher, L. S. Shapley, A. W. Tucker, eds. Princeton: Princeton University Press (1964).

[94] Smith, V. Incentive compatibile experimental processes for the provision of public goods. In *Research in Experimental Economics*, V. Smith, ed. Greenwich, CT: JAI Press (1979).

[95] Thomson, W. Economies with public goods: An elementary geometric exposition. *J. Public Econ. Theory*, 1 (1999), 139–176.

[96] Topkis, D. M. Equilibrium points in nonzero-sum n-person submodular games. *SIAM J. Control and Optimization*, 17 (1979), 773–787.

[97] Topkis, D. M. *Supermodularity and Complementarity.* Princeton, NJ: Princeton University Press (1998).

[98] Vega-Redondo, F. Implementation of Lindahl equilibrium: An integration of static and dynamic approaches. *Math. Soc. Sci.*, 18 (1989), 211–228.

[99] Vickrey, W. Counterspeculation, auctions and competitive sealed tenders. *J. Finance*, 16 (1961), 8–37.

[100] Walker, M. A simple incentive compatible scheme for attaining Lindahl allocations. *Econometrica*, 49 (1981), 65–71.

[101] Yzerbyt, V. Y., Coull, A., Rocher, S. J. Fencing off the deviant: The role of cognitive resources in the maintenance of stereotypes. *J. Personality Soc. Psych.*, 77 (1999), 449–462.

[102] Zhou, L. Impossibility of strategy-proof mechanisms in economies with public goods. *Rev. Econ. Stud.*, 58 (1991), 107–119.

[103] Zhou, L. The set of Nash equilibria of a supermodular game is a complete lattice. *Games Econ. Behav.*, 7 (1994), 295–300.