**Inequality, Risk Sharing, and**

**the Boundaries of Collective Organizations**

Edward S. Prescott
Federal Reserve Bank of Richmond

and

Robert M. Townsend
University of Chicago
Federal Reserve Bank of Chicago

Revised June 2000

## Abstract

This paper studies the boundaries and interactions of collective economic organizations that share risk and mitigate moral hazard. Each type of organization is identified with a set of information, technology, and contracting possibilities. A mechanism design, or agency, problem is then solved to determine the optimal organizational structure. Information-constrained optimal distributions of organizations are shown to be functions of the underlying primitives, in particular, the distribution of Pareto weights, and hence degree of inequality. More generally, the impact of inequality on organizational form and allocations is shown to depend on hypothesized interactions among technology, information, and collusion. These hypotheses and their implications on organizations and allocations could be distinguished in cross-sectional, time series data.

The authors acknowledge comments from participants at seminars at CREST, DELTA, European Institute of Florence, Richmond Fed, Minneapolis Fed, Midwest Economics Association, LSE, Yale, U. of Venice, U. of Iowa, Arizona St., Penn St., U. of Pennsylvania, and the Northwestern Summer Macro Workshop. Townsend would like to thank the NSF and NICHD for financial support. The views expressed here are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Chicago, the Federal Reserve Bank of Richmond, or the Federal Reserve System.

## 1. Introduction

In this paper, we study collective organizations that share risk and mitigate incentive problems. Partnerships are one such commonly observed type of organization. They exist in many fields, including medicine, law, and agriculture.[1] The large Japanese and Korean industrial conglomerates that provide mutual credit and insurance are another such type of collective organization. Close-knit networks of family members and friends, while often informal, are equally important examples.

The answers to substantive economic questions depend on understanding why these organizations form, why they coexist with other types of organizations, and how they trade with the rest of the world. For example, would limiting the formation of the Japanese and Korean conglomerates improve these countries' economic performance? While presently fashionable to criticize these conglomerates for inefficiency and corruption, it was not too long ago that they were lauded for their economic performance. Surely, a serious study of the design of industrial credit and insurance requires more than an inference based on the correlation between the performance of firms and the macroeconomy.

For less-developed, agricultural economies the questions are different, though equally important. It is well known that agriculture in these countries is a risky enterprise. It is also well known that there is a considerable degree of risk sharing in these economies, even if it is far from perfect. Are organizations such as families, kinship networks, and farming partnerships an important source of risk sharing? Are they an efficient method for production? Their coexistence with more individualistic types of organization suggests that the tradeoffs are not trivial. Indeed, in the ICRISAT villages of India, some individuals are members of farming partnerships while simultaneously farming their own land.

Risk aversion, private information, monitoring, and collusion are the ingredients we use to deliver a theory of collectives. The idea is that in an economy with moral hazard, there will likely be better insurance among the agents with common information or superior monitoring technologies. In our approach, these clusters are joint assignments, teams, or organizations that we put explicitly into the commodity space.

Our emphasis on monitoring and joint assignments places this paper in the theory-of-the-firm research tradition initiated by Alchian and Demsetz (1972). Unlike much of this literature, however, we emphasize the connection between risk sharing, inequality, and organizational formation. We find that equilibrium organizational structure delivers strong implications for variability and co-movements of consumption and effort allocations. These implications are strong enough that they can be tested with sufficiently rich data.

We also find a strong connection between the distribution of wealth and the organizational structure of the economy. There has been very little work on this connection, though a notable exception is Legros and Newman (1996), who used the core concept to study an example economy with risk neutrality. Using the Pareto mapping, we add to their work by studying a class of economies that are different along several dimensions, including risk aversion. As we will see, even within the class of models considered in this paper, the mapping between inequality and organizational structure is highly model dependent.

We proceed first by defining here a general class of regimes or contracting environments. These vary in the combinations of collusion against outsiders, internal commitment among members, the internal and external information and monitoring structure, and the relation of that information structure to production and the operation of various possible technologies. Some of these regimes can be thought of as individualistic while other regimes display a version of the collectivity referred to earlier. The equilibrium regime will be a solution to a mechanism design problem. The predicted regime will depend not only on the usual primitives of preferences and technology, but also on the distribution of Pareto weights, which are related to the distribution of wealth in a competitive equilibrium.

In the first class of models, those of Section 2, a collective organization is defined as a group of agents who have perfect information about one another, pool resources, and can without cost enforce internal insurance and labor assignments. At the same time, there is an outsider who does not have perfect information about the group members' activities and the outsider is presumed to be unable to elicit this information from a revelation game.

The alternative mode of organization for this small economy is more individualistic in the sense that each agent can enter into an individual contract with the outsider to operate one or various technologies. In not forming a group, the agents are *presumed* to no longer have perfect information on each other and to lose the ability to write contracts among members. However, the payoff to one agent from the contract with the outsider may depend on his performance relative to that of other agents.

The comparison between the group and the relative performance regimes is precisely the comparison made by Holmstrom and Milgrom (1990), Itoh (1993), and Ramakrishnan and Thakor

(1991). Unlike these papers, we assume neither transferable utility nor symmetric equilibria. We show that without these restrictions the relative merits of the two types of organization depend on where the economy is on the Pareto frontier, that is, on the distribution of Pareto weights across potential members of the small economy and of the agents relative to the outsider. Groups are more likely to emerge the greater the internal dispersion of weights between small economy workers and, with exceptions, the poorer or less utility weight all members/workers have relative to the outsider. The exception occurs when the utility weights of local members are so high relative to outsiders that there are no binding incentive constraints. We also show that there are striking implications for risk sharing and production. There will be perfect risk sharing within members of a group, and if reallocation of inputs is allowed there will also be efficiency in production across the technologies operated by members of the group. The relation of the group to the outsider will be determined as in a principal-agent relationship, but with a “mongrel” consumer.

In the second class of models, those of Section 3, we define a collective organization as a set of production technologies that are operated jointly by a set of agents, working in equal amounts. This joint production may come at a cost, but common information on labor effort applied to these technologies can be transmitted to outsiders, allowing such "groups" to escape incentive problems that limit access to external insurance. At the same time though, other technologies may be operated individually by these same agents, and the evident moral hazard problem does limit insurance on these. The local economy again solves an optimal assignment problem in which the scope of a collective organization is endogenous. All, some, or even none of the technologies may be operated by a collective organization.

As in Section 2, we find that the poorer the agents are relative to the outsider, the more likely collective organizations are to form. Unlike in Section 2, however, we find that equality of the agents' Pareto weights is conducive to the formation of collective organizations. The effect of unequal weights on collective organizations, however, also depends on the parameters of preferences and technology. We provide an example where collective organizations emerge at unequal weights and another example where they do not. We also show that if jointly operated technologies coexist with individually operated ones, then a given agent has some shocks which are completely insured and others which are not. This implication can be subjected to econometric tests jointly with predictions about the optimal assignment.

Both the first and the second model can also be formulated as a general equilibrium problem with a continuum of small local economies. This formulation allows each small economy to interact, possibly as a collective entity, with the rest of the economy. This also allows us to vary the characteristics of the small economies as in cross-sectional data and hence to be more precise about the likelihood that a particular regime will be seen in the data as a function of observables. Allocations of the large economy can also be decentralized to some degree, but in the interest of brevity we do not work this out in detail here. See the earlier working paper and, especially, Prescott and Townsend (2000).

**2. Collective Organizations as Collusive Cooperatives**

In the first model, there is a set of agents (households or workers) who can work on one or several technologies. Efforts on these technologies determine the probability distribution of outputs. The agents can potentially come together to constitute a cooperative or a collective organization relative to a non-working outsider. Two regimes are considered and compared. In the first regime the agents have full information about each other's efforts, outputs and consumption, and they can collude against the outsider by agreeing to an internal effort and consumption allocation, one which is Pareto optimal for the group, conditioned on the group's agreement with the outsider. We call this setup the group regime and identify it as a collective economic organization.

In the second regime two changes are made. First, collusion between the agents is prohibited *a priori*. The outsider or principal can prevent any reallocation of effort or consumption. He determines and knows all transfers. Second, agents no longer observe each other's efforts. Instead, they only observe what the principal does, namely, the other agents' outputs. We call this setup the relative performance regime and identify it as the non-cooperative type of organization.

This section proceeds by first analyzing the relative performance and group regimes separately. In this setup, there are three people, two agents and an outsider. Next, the models are combined into a single mechanism design problem where the regime is endogenous.

**2.1 Environment of a Small Three-Person Economy**

Assume for simplicity that there are only two agents and two technologies. Each agent *i* has preferences over own consumption, $c_i$, and total own effort, $e_i$. Define utility as $U_i(c_i, T_i - e_i)$ where $T_i$ is agent *i*'s time endowment. Total own effort is the sum of agent *i's* efforts across all technologies $j = 1, 2$. Let $e_{ij}$ denote agent *i's* effort on technology *j*, so $e_i = \sum_j e_{ij}$. It is also useful to define

$e_{i\bullet} = (e_{i1}, e_{i2})$ as the *vector* across the two technologies of agent *i's* efforts and to define $a_j = \sum_i e_{ij}$ as the total effort applied to technology *j* by all agents. This implies that $(a_1, a_2) = e_{1\bullet} + e_{2\bullet}$.

Output on technology $j$, $q_j$, is a function of the total effort put on that technology, $a_j$, and a random shock. The technologies *j* can differ in size or productivity. More generally, the joint probability of outputs $q_1$ and $q_2$, given total technology efforts $a_1$ and $a_2$, is described by the density function $p(q_1, q_2 \mid a_1, a_2)$. For expositional ease, the same function will often be written with conditioning variable $e_1, + e_2$. Finally, correlation in technology returns, as from common aggregate shocks, is already consistent with this notation.

Two different specifications of the labor assignments are considered. In the first specification, agents can each work both technologies, whether done in the group regime by colluding against an outsider, or done in the relative performance regime in assignments recommended by an outsider. The point is that information about who is working which technologies does not distinguish the two regimes. In the second labor specification, agents must work only on their assigned technologies, for simplicity agent one on technology one and agent two on technology two. This second specification is easily incorporated by restricting efforts to satisfy $e_{ij} = 0$ for $i \neq j$. Here again, however, the technology specification does not pin down the regime.

The two agents in this small economy are imagined to be dealing with an outsider. Two concepts of the outsider are useful. The first is that the outsider is a principal who cannot work the technology himself. The principal's utility is a function of technology outputs minus agents' consumption, namely $W(q_1 + q_2 - c_1 - c_2)$. The alternative concept of the outsider is that he is simply the rest of the economy. In this case, *W* is necessarily linear and $q_1 + q_2 - c_1 - c_2$ is the surplus generated by a group. In a larger closed economy with multiple groups the summation of surpluses across groups must be zero. We shall defer consideration of this interpretation of the outsider until Section 4.

We proceed in the following sections by writing down programming problems that determine the entire class of information-constrained optimal allocations. A particular Pareto optimum is determined by the Pareto weights associated ultimately with the wealth or status of the individual agents and of the principal. The choice object in each program is a potentially random device that assigns consumptions $c_i \in C_i$, to both agents; outputs $q_i \in Q_i$, on both technologies; and effort of agent *i* on technology *j*, $e_{ij} \in E_{ij}$ with indices running over agents *i* and technologies *j*. The set of *a priori* feasible points is $S = C_1 \times C_2 \times Q_1 \times Q_2 \times E_{1\bullet} \times E_{2\bullet}$, where $E_{i\bullet} = E_{i1} \times E_{i2}$. In the second technological specification, where agents may only work their own technology, we restrict $E_{ij} = \{0\}, i \neq j$. Note also that sets $C_i$, $Q_i$ and $E_{ij}$ can take on a continuum of values, but often we will restrict ourselves to sets with a finite number of values.

For expositional clarity define $c = (c_1, c_2)$ and $q = (q_1, q_2)$ as the vectors of consumptions and outputs, respectively. As mentioned earlier $e_{i\bullet}$ is the vector of agent *i's* technology-specific efforts $(e_{i1},\ e_{i2})$. Then the choice object can be written as $p(c, q, e_{1\bullet}, e_{2\bullet})$, the probability with which a particular consumption, output, and labor allocation is chosen. As we focus attention on the case of finite sets, this notation is obvious. But if we allowed continua, we would call $p$ a probability measure and define $p(B)$ to be the probability of Borel set $B$ in a Borel-algebra of *S*. We stress as well that solutions often are deterministic in the sense that $c_i$ is a nonstochastic function of output vector $q = (q_1, q_2)$ and deterministic numbers $e_{1\bullet}$ and $e_{2\bullet}$ are recommended by the contract. Each program maximizes a weighted sum of the expected utilities of the two agents subject to constraints: a participation constraint for the outsider; incentive constraints on individual efforts if it is a relative performance model, or collusion constraints if it is a group model; nature constraints which assure that the endogenous probability of outputs is consistent with the underlying technology specification $p(q_1, q_2 \mid e_{1\bullet} + e_{2\bullet})$; and a set of constraints that ensures that the choice object $p$ is a probability measure. The theoretical advantage of the lottery approach is that all possible gains from contracting are exhausted. The practical advantage is that the constrained maximization problems are linear programs, and as long as the dimensions of *S* are not too large then solutions to parameterized economies can be computed. Again, for expositive purposes we do restrict our attention to the finite case, except where noted.[2]

**2.2 Relative Performance Regime**

It is useful to begin the discussion with the relative performance model. In this model one agent cannot observe the other agent's effort nor can the two agents collude against the outsider. Program 2.2.1 below describes the problem under the technological specification that agents can work both technologies.

**Program 2.2.1:**

The problem is to maximize a $l = (l_1, l_2)$ weighted sum of utilities

(2.2.1) $$\sum_{c,q,e_{1\bullet},e_{2\bullet}} p(c, q, e_{1\bullet}, e_{2\bullet})[l_1 U_1(c_1, T_1 - e_1) + l_2 U_2(c_2, T_2 - e_2)]$$

by choice of the $p(c,q,e_{1\bullet},e_{2\bullet})$ subject to a set of constraints. Constraint (2.2.2) below is the participation constraint for the principal:

$$\text{(2.2.2)} \qquad \sum_{c,q,e_{1\bullet},e_{2\bullet}} p(c,q,e_{1\bullet},e_{2\bullet})W(q_1+q_2-c_1-c_2) \geq \overline{W}.$$

By solving the problem for the feasible range of promised utility $\overline{W}$ for the outsider and over the $l$-weights for the two internal members the entire Pareto frontier can be calculated.

Thus, $l$ and $\overline{W}$ can be treated as parameters to be identified in data. Weights $l$, in particular, specify the degree of internal equality and the parameter $\overline{W}$ specifies the poverty of the working agents relative to the outsider.

Constraints (2.2.3) below are the incentive constraints for agent one. They imply that for every effort $e_{1\bullet}$ assigned with positive probability, obedience is weakly preferred to deviations $\hat{e}_{1\bullet}$. As the summand indicates, agent one makes his decision independent of the other agent, agent two, since he does not observe the other agent's effort or recommendation:

$$\text{(2.2.3)} \qquad \begin{aligned} &\sum_{c,q,e_{2\bullet}} p(c,q,e_{1\bullet},e_{2\bullet})U_1(c_1,T_1-e_1) \\ &\geq \sum_{c,q,e_{2\bullet}} p(c,q,e_{1\bullet},e_{2\bullet})\frac{p(q \mid \hat{e}_{1\bullet}+e_{2\bullet})}{p(q \mid e_{1\bullet}+e_{2\bullet})}U_1(c_1,T_1-\hat{e}_1),\ \forall e_{1\bullet},\hat{e}_{1\bullet}. \end{aligned}$$

Similar to (2.2.3) are equations (2.2.4), agent two's incentive constraints:

$$\text{(2.2.4)} \qquad \begin{aligned} &\sum_{c,q,e_{1\bullet}} p(c,q,e_{1\bullet},e_{2\bullet})U_2(c_2,T_2-e_2) \\ &\geq \sum_{c,q,e_{1\bullet}} p(c,q,e_{1\bullet},e_{2\bullet})\frac{p(q \mid e_{1\bullet}+\hat{e}_{2\bullet})}{p(q \mid e_{1\bullet}+e_{2\bullet})}U_2(c_2,T_2-\hat{e}_2),\ \forall e_{2\bullet},\hat{e}_{2\bullet}. \end{aligned}$$

Equation (2.2.5) below guarantees that allocations respect the probability distributions generated by the technology:

$$\text{(2.2.5)} \qquad \forall \bar{e}_{1\bullet},\bar{e}_{2\bullet},\bar{q},\ \sum_{c} p(c,\bar{q},\bar{e}_{1\bullet},\bar{e}_{2\bullet}) = p(\bar{q} \mid \bar{e}_{1\bullet}+\bar{e}_{2\bullet})\sum_{c,q} p(c,q,\bar{e}_{1\bullet},\bar{e}_{2\bullet}).$$

Finally, equation (2.2.6) below ensures that the choice variable be a probability measure:

$$\text{(2.2.6)} \qquad \sum_{c,q,e_{1\bullet}e_{2\bullet}} p(c,q,e_{1\bullet},e_{2\bullet}) = 1.$$

The time line for the realization of a contract deserves to be made explicit. First, actions $e_i$ are recommended to agent *i*, possibly at random, and without knowledge of recommendation $e_i$ to the other agent *i*. Agent $i$ then decides on some actual actions. Mother nature then determines outputs, and final consumptions $c$ are determined, almost certainly moving with random outputs $q$ and possibly with additional randomness as well.

Only one slight modification to Program 2.2.1 is required in order to set up the constrained maximization problem which incorporates the restriction that agents must work their own technology; restrict the set *S* of feasible points to those where $e_{12} = e_{21} = 0$. Everything else, the objective function and the constraints, are unchanged. The program for this second technological specification will be referred to as Program 2.2.2, but because of the similarities with Program 2.2.1 it is not shown explicitly.

The properties of the relative performance model have been extensively analyzed by Holmstrom (1982) and Mookherjee (1984) and are well known in the literature. Consequently, we do not analyze the model in any detail except to note the main observation about these models. The value of relative performance depends on the correlation in the technologies' returns. When returns are correlated, output comparison is very informative about effort deviations.[3] When returns are uncorrelated, however, output comparison reveals no information about an agent's effort. In this case, when $W(\cdot)$ is linear, and the principal absorbs all risk apart from variations for incentive purposes, the optimal contract contains no dependence of one agent's consumption on the other agent's output realization.

### 2.3 Group Regime

We now consider the case where agents can observe costlessly each other's efforts and are in a group that can collude perfectly in its dealings with the outsider. Consider first the specification where labor effort can be shared across the technologies. The idea is to start with Program 2.2.1 and its solution $p$, and then imagine what the two agents would do if they could jointly specify aggregate and individual efforts on the two technologies, could specify internal consumption as a function of outputs across the two technologies, had perfect information about efforts as well as outputs, and had the ability to perfectly and costlessly enforce all internal agreements.[4]

Given the contract $p(c, q, e_{1\bullet}, e_{2\bullet})$ what they would *not* do is end up in a situation where they could improve at least one member's utility without making the other member worse off. That is, they would seek to find some conditional Pareto optimal group allocation. An allocation satisfies this criterion if it solves the group's internal Pareto problem. This problem is to maximize some sum of weighted utilities of the two agents subject to two resource constraints. The first constraint is that individual efforts sum to the aggregate effort that the group members agree internally to do. The second constraint is that internal consumptions add to the sum of consumption implicit in the agreement with the outsider. Remember, the outsider makes transfers to the group members as a function of observed outputs *q*. Thus the two agents decide on total effort, that is, $\tilde{e}_g$; effort of each group member, $\tilde{e}_1$ and $\tilde{e}_2$; and effort over the two technologies, $\tilde{a}_1$ and $\tilde{a}_2$. Further, upon receiving $c_g$ units of consumption, the group divides it between its members by choosing vector $\tilde{c} = (\tilde{c}_1, \tilde{c}_2)$ such that $\tilde{c}_1 + \tilde{c}_2 = c_g$, a restriction we impose subsequently. If we let $m = (m_1, m_2)$ denote the Pareto weights *within* the group, group members choose these efforts and consumptions to maximize that weighted sum of their utilities, subject to the group transfer rule $p$ to the outsider. Using lottery notation $\tilde{p}(\tilde{c}, c_g, \tilde{e}_1, \tilde{e}_2)$, we write the group's problem as

**Program 2.3.1:**

Maximize

(2.3.1) $$\sum_{\tilde{c}, c_g, \tilde{e}_{1\cdot}, \tilde{e}_{2\cdot}} \tilde{p}(\tilde{c}, c_g, \tilde{e}_{1\cdot}, \tilde{e}_{2\cdot})[m_1 U_1(\tilde{c}_1, T_1 - \tilde{e}_1) + m_2 U_2(\tilde{c}_2, T_2 - \tilde{e}_2)]$$

by choice of the $p(\tilde{c}, c_g, \tilde{e}_{1\bullet}, \tilde{e}_{2\bullet}) \geq 0$, subject to constraints defining lotteries

(2.3.2) $$\sum_{\tilde{c}, c_g, \tilde{e}_{1\bullet}, \tilde{e}_{2\bullet}} \tilde{p}(\tilde{c}, c_g, \tilde{e}_{1\bullet}, \tilde{e}_{2\bullet}) = 1,$$

and subject to a technology constraint

(2.3.3) $$\forall \bar{c}_g,\ \bar{\tilde{e}}_{1\bullet}, \bar{\tilde{e}}_{2\bullet},\ \sum_{\tilde{c}} \tilde{p}(\tilde{c}, \bar{c}_g, \bar{\tilde{e}}_{1\bullet}, \bar{\tilde{e}}_{2\bullet}) = p(\bar{c}_g \mid \bar{\tilde{e}}_1, \bar{\tilde{e}}_2) \sum_{\tilde{c}, c_g} \tilde{p}(\tilde{c}, c_g, \bar{\tilde{e}}_{1\bullet}, \bar{\tilde{e}}_{2\bullet}),$$

where $p(\bar{c}_g \mid \bar{\tilde{e}}_1, \bar{\tilde{e}}_2)$ is the probability of receiving aggregate $\tilde{c}_g$ given efforts $\bar{\tilde{e}}_1$, $\bar{\tilde{e}}_2$ and is determined under $p$ (the contract between the outsider and the two agents) by

(2.3.4) $$p(\bar{c}_g \mid \bar{\tilde{e}}_1, \bar{\tilde{e}}_2) = \sum_{(c_1, c_2 \mid c_1 + c_2 = \bar{c}_g)} p(c_1, c_2 \mid q, e_{1\bullet}, e_{2\bullet}) p(q \mid \bar{\tilde{e}}_{1\bullet}, \bar{\tilde{e}}_{2\bullet}),$$

given the allocation $p(c_1, c_2, q, e_{1\bullet}, e_{2\bullet})$, and recommended efforts $e_{1\bullet}, e_{2\bullet}$. As a conditionally Pareto optimal agreement, however, the weights $m$ are decided just after the agreement $p$ is entered into with the principal and before efforts are recommended under the contract. Weights $m$ are thus endogenous, unlike exogenous weights $l$. But as weights $l$ are varied cross sectionally across small economies, so also will endogenous weights $m$ move. See below for specifics. On the other hand, fixing weights $m$, program 2.3.1 is the classical full-information program of the risk-sharing literature.

These assumptions on how groups operate change the information structure in the group model from that of the relative performance model. Here, in particular, the group members have perfect information about their labor efforts $\tilde{e}_{1\bullet}, \tilde{e}_{2\bullet}$. The question naturally arises as to whether the outsider can take advantage of this by using a direct revelation mechanism that implements the full-information optima. If members were not allowed to collude, then such an allocation could be incentive compatible as in Harris and Townsend (1981). But as Itoh (1993) has argued in a similar context, collusion effectively rules this out. The two agents would simply decide on the allocation they would like to implement, after the contract with the principal has been signed.

One can ask whether collusion among the agents could take place at some other point along the time line of the contract. For example, could the two agents collude ex ante at the time the contract is signed by committing with the principal to carry out full-information allocations? As much as all parties to the contract might like this arrangement, the two agents would re-solve given the full-information contract for some conditionally Pareto optimal allocation in the group, much the way a single agent takes actions conditioned on the contract with the principal in the classic principal-agent model. We assume this type of collusion cannot be prevented ex ante.

We do allow commitment in all other aspects, however, including agreement among the two working agents to a particular full-information within-group conditional Pareto optimal allocation, that is, equivalently to what weights $(m_1, m_2)$ to use in Program 2.3.1 above. Again, with this foresight, we can think of these weights $(m_1, m_2)$ as endogenous to or implicitly part of the original contract $p$, and we now <u>*include*</u> $m$ as an argument in contract $p$.

Ex post-collusion after outputs are realized is feasible, but it accomplishes nothing new. Given that the two agents have agreed to a conditional Pareto optimal allocation and end up with specified consumption bundles as a function of realized outputs net of the payoff to the principal, there is no possible gain for both agents together. For most specifications of utility functions, changes that benefit one agent must by construction hurt the other. Again, we suppose the agents have commitment devices that preclude ex post bargaining.

Thus, given the outside contract $p$, what the group will do can be rationally anticipated by all, just as the deviation of effort can be anticipated in the standard principal-agent model. More formally, the group's induced and potentially random choice of efforts and their induced final distribution of consumption can also be put into the initial random assignment function $p(c, q, e_{1\bullet}, e_{2\bullet}, m)$. Then, conditional on an assigned $m$, recommended technology-specific efforts $(a_1, a_2)$ should actually maximize the $m$-weighted sum of expected utilities relative to any other such choice. Similarly, the group's sharing rule for agent-specific consumption and efforts can be anticipated.

Precisely how the group's internal allocation decisions can be embedded into the allocation is easiest to see when preferences are separable. With separable preferences individual consumption and efforts can be substituted out as a function of group aggregates and internal Pareto weights $m$. Consider the case of preferences $U_i(c_i, T_i - e_i) = U_i(c_i) + V_i(T_i - e_i)$. In this case, the solution to the effort and consumption allocation problem will be determined entirely by the weights $m = (m_1, m_2)$, by the aggregate consumption $c_g$ that materializes, and by the aggregate effort $e_g$ that the group decides to implement. Thus we write in the obvious notation $c_i = c_i(c_g, m)$ and $e_i = e_i(e_g, m)$, and these become the obvious choice variables. Further, given weights $m$, any effort $a_1$, $a_2$ on the technologies with $a_1 + a_2 = e_g$ must be maximal and hence satisfy the incentive constraints (2.3.7) that will be listed below.

By the preceding arguments we can substitute out individual consumptions and individual efforts and instead consider the choice variable $p(c_g, q, a_1, a_2, m)$. While this substitution is only valid for separable preferences, this case is sufficiently interesting that we write the program out explicitly.[5] In summary then, we let $l$ be the ex ante vector of Pareto weights for the agents, $\overline{W}$ be the reservation value for the outsider, both to be varied exogenously, and let $m$ be the ex post internal weights as assigned by the contract. The program for the determination of an optimal principal-group contract is then

**Program 2.3.2:**

Maximize the objective function

(2.3.5) $$\sum_{c_g, q, a_1, a_2, m} p(c_g, q, a_1, a_2, m) \sum_i l_i [U_i(c_i(c_g, m)) + V_i(T_i - e_i(e_g, m))]$$

by the choice of $p(c_g, q, a_1, a_2, m) \geq 0$, subject to

(2.3.6) $$\sum_{c_g, q, a_1, a_2, m} p(c_g, q, a_1, a_2, m) W(q_1 + q_2 - c_g) \geq \overline{W},$$

(2.3.7)
$$\sum_{c_g, q} p(c_g, q, a_1, a_2, m) \sum_i m_i [U_i(c_i(c_g, m)) + V_i(T_i - e_i(e_g, m))]$$
$$\geq \sum_{c_g, q} p(c_g, q, a_1, a_2, m) \frac{p(q \mid \hat{a}_1, \hat{a}_2)}{p(q \mid a_1, a_2)} \sum_i m_i [U_i(c_i(c_g, m)) + V_i(T_i - e_i(\hat{e}_g, m))],$$
$$\forall m, a_1, a_2, \hat{a}_1, \hat{a}_2, \text{ where } \hat{e}_g = \hat{a}_1 + \hat{a}_2,$$

(2.3.8) $$\forall \bar{a}_1, \bar{a}_2, \bar{q}, \bar{m}, \ \sum_{c_g} p(c_g, \bar{q}, \bar{a}_1, \bar{a}_2, \bar{m}) = p(\bar{q} \mid \bar{a}_1, \bar{a}_2) \sum_{c_g, q} p(c_g, q, \bar{a}_1, \bar{a}_2, \bar{m}),$$

(2.3.9) $$\sum_{c_g, q, a_1, a_2, m} p(c_g, q, a_1, a_2, m) = 1.$$

The main difference between this group regime, Program 2.3.2 here, and Program 2.2.1, the relative performance regime, is the set of incentive constraints, namely equations (2.3.7) in Program 2.3.2 versus equations (2.2.3) and (2.2.4) in Program 2.2.1.[6] Constraints (2.3.7) here ensure that *given* the group's contract with the outsider and given its internal Pareto weights $m$, there is no alternative action pair conditionally Pareto superior for the group. The ability of the group to redistribute consumption and efforts is already incorporated in the internal distribution rules. The incentive constraints represent the preferences of a fictitious "mongrel" consumer in which the within-group Pareto weight vector $m$ is an essential parameter. The within-group weight is a choice variable, which need not be equal to the Pareto weights $l$. The potential difference between these $m$-values and $l$-values will be discussed later.

Now suppose we are under the second technology specification, that is, each agent works only his own technology. Individual efforts no longer solely depend on aggregate effort and the internal distribution weight $m$. Rather, we assign $e_1$ and $e_2$ directly as in going from Program 2.2.1 to Program 2.2.2 earlier. Some simplifications for consumption still apply, namely, we can write $c_i = c_i(c_g, m), \; i = 1, 2$. This new program is labeled Program 2.3.3 for further reference, but it is not written out. Comparisons between unwritten Programs 2.2.2 and 2.3.3 would appear quite similar to the comparison between written Programs 2.2.1 and 2.3.2.

**2.4 Theoretical and Empirical Implications of the Relative Performance and Group Regimes**

As we discussed earlier, the relative performance model is well understood. If outputs from the production technologies are uncorrelated, a risk neutral outsider should enter into a risk-sharing incentive contract separately with each agent, and there are literally two principal-agent problems. The econometric implications of such classic agency problems are straightforward. Here and below we imagine that Pareto weights can be varied as if in a nontrivial cross section and that the static contracts analyzed here can be replicated over many time periods without explicit inter-temporal dynamics.

The relation between output and consumption will be determined by the weight of agent *i* relative to the principal, as embodied in $\overline{W}$, and by the production technology $p(q_i \mid a_i)$. Over ranges where the principal's $\overline{W}$ is low and the welfare weight $l_i$ of the agent *i* is high, the agent will be asked to work little. If there is then no incentive problem, there is full insurance between the agent and the principal. If for alternative $\overline{W}$'s the agent is assigned higher, non-trivial effort, then the relation between output and consumption will be determined by the likelihood ratio $p(q_i \mid \hat{a}_i) / p(q_i \mid a_i)$, but there is still no relation between consumption the two agents. Indeed, the principal is using consumption to reward or punish each agent separately based on the principal's inference of effort, and the principal does not want to undercut this by tying consumption variations together. Levels of consumptions are tied together through the overall resource constraint (2.2.2) only.

If outputs are positively correlated, there will be a relationship between the consumptions of the two agents. High outputs on both technologies are likely to be associated, causing consumption to be high for both. But with high output on one technology unlikely to be associated with low output on the other, if one agent receives a low output and the other agent receives a high output then the agent with the low output will be punished by receiving a low level of consumption. This is the benefit of using relative performance evaluation, an implication that distinguishes it from full risk sharing. Again, correlated outputs is a force for using the relative performance regime.[7]

In contrast, the salient feature of the group regime is that internally it is exactly a full-commitment, full-information economy. As noted in the derivations of Program 2.3.2 from Program 2.3.1, this feature has implications for the internal allocation of consumption and effort, and also for the internal allocation of effort into production technologies. We briefly review these here.

For simplicity, we now imagine that there is a continuum of possible allocations of consumption and effort within a group. We also continue with the assumption that utility functions are separable. (Exceptions to separability will be noted as appropriate.) Finally, we imagine that no agent hits binding corner constraints, that is, consumption, leisure, and work levels all remain positive. Then, given the optimum associated with $m = (m_1, m_2)$ weight, the allocations within a group must satisfy

(2.4.1) $$m_1 U_1'(c_1) = m_2 U_2'(c_2),$$

(2.4.2) $$m_1 V_1'(T_1 - e_1) = m_2 V_2'(T_2 - e_2),$$

along with adding up constraints $c_1 + c_2 = c_g$ and $e_1 + e_2 = e_g$.

As is apparent in (2.4.1), the $m$ weights will determine relative levels of consumption; roughly, the higher the weight $m_1$ relative to $m_2$, the higher will be consumption of agent one relative to agent two. Of course, consumption of agents one and two will also move around with outputs $q_1$ and $q_2$ but only in so far as aggregate consumption $c_g$ moves. Holding aggregate group consumption $c_g$ constant, variations in $q_1$ and $q_2$ will not influence $c_i$. This full insurance implication is immune to correlations in outputs over the two technologies and with separability in consumption it is even immune to non-reallocable effort.

In particular, this dependence of individual consumption on group consumption implies that agents within the group should pass econometric tests for full insurance as in Altug-Miller (1990), Cochrane (1991), Deaton (1993), Mace (1991), and Townsend (1994), for example, but only with respect to group consumption.[8] These tests regress individual consumption on aggregate consumption, often at the village or national level, to see if the functional relationship implicit in equation (2.4.1) holds. The model here thus provides guidance about the level of aggregation at which the tests should be performed and implies that it is necessary to carefully measure who is a member of a group and who is not (Morduch, 1994).

Though often overlooked, equation (2.4.2) on the allocation of labor/leisure is also subject to econometric tests of precisely the same form as those for consumption. Even with nonseparability in consumption and leisure, we are left with potential econometric tests of labor supply as in Chiappori (1992). More generally, there are joint implications for effort and consumption levels, as high $m$ agents should not be required to work hard though they should eat well.

Assuming that labor is reallocable, similar arguments can be used to devise tests for production efficiency across the two technologies. Technology specific labor effort should not depend on household composition; Benjamin (1992) has conducted such tests, for example. More generally, we can see how these production implications are also tied to consumption/leisure implications. One can test jointly for consumption smoothing and efficiency in production, for example, assuming group membership is correctly specified. But note, however, that if the specification is such that labor is not reallocable over technologies, then we lose these production efficiency implications.

This exhausts the implications for Program 2.3.1 for within group allocation of consumption, efforts, and inputs, but there remains the relationship of the group relative to the outsider. In part, this appears as a classic principal-agent relationship, as discussed earlier, except that the agent is a mongrel consumer, as in incentive constraints (2.3.7), working the two technologies. At this level of generality, one can see that group consumption $c_g$ is unlikely to be constant, that is, the group is not fully insured, but rather $c_g$ should move with $q_1$ and $q_2$, for incentive reasons. Put differently, the insurance transfer is determined by the difference between aggregate consumption $c_g$ and aggregate output, $q_1$ plus $q_2$. That is, if aggregate consumption $c_g$ can be written as a deterministic relationship $c_g(q_1,q_2)$, the transfer is then $q_1+q_2-c_g(q_1,q_2)$. The exact transfer will be determined by the production technology $p(q_1,q_2 \mid a_1,a_2)$, by the weight $\overline{W}$ of the principal, and the weight $l=(l_1,l_2)$ of the agents.[9] This leads to joint restrictions that can be econometrically evaluated. This will be more evident with some examples.

**Numerical Example 2.4.1**

Suppose preferences for agent *i* are separable in consumption and leisure and of the power form $U_i(c_i,T_i-e_i)=c_i^{0.5}-(e_i)^{0.5}$, for $i=1,2$ where we suppress notation for the time endowment. We suppose here as well that each agent *i* can work only his own technology *i*. Consumption for each agent *i* is gridded over the range 0 to 25 at intervals of one. This determines set $C_i$. Actions on each technology can take on a low value of $a_l=2$ or a high value of $a_h=6$. Because agents can work only their own technology, we need only keep track of each agent's effort on his own technology. Accordingly, we treat elements of set $E_i$ as a one-dimensional vector and write the set $E_i = \{2, 6\}$.

Like effort, output on either technology *i* can only take two values. It can take on a low value of $q_l = 2$ or a high value of $q_h = 20$. This determines sets $Q_1$ and $Q_2$. The grid of $m_1$ was set at intervals of 0.01 on [0.0, 1.0] for 101 points in total. (Of course, the Pareto weight on the second agent is $m_2 = 1 - m_1$.)

For this example, output returns are presumed uncorrelated. On each technology *i* the probability distribution of output given effort, $p(q_i \mid a_i)$, is described by

| | $q_l$ | $q_h$ |
|---|---|---|
| $a_l$ | 0.7 | 0.3 |
| $a_h$ | 0.3 | 0.7 |

Note that higher efforts make high output more likely, so we expect *a priori* that high output is to be rewarded. Finally, the principal is assumed to be risk neutral and the level $\overline{W}$ of his participation constraint is set equal to 20 units.

For the power preferences used in this example the consumption sharing rule is

$$c_i = \frac{m_i^{1/(1-a)}}{m_1^{1/(1-a)} + m_2^{1/(1-a)}} c_g \qquad (2.4.3)$$

with $a = 0.5$. It can be seen that consumption $c_i$ explicitly increases in the internal weight $m_i$ for a given level of $c_g$. Sufficient repeated data on the $c_i$ and $c_g$ would allow one to estimate the $m_i$ and parameter $a$ (a regression in log levels does the same).

Hereafter, we make the discrete equivalent substitution of (2.4.3) for $c_i$ and focus on group consumption $c_g$ and how it varies with $q_1$ and $q_2$. Figure 2.4.1 shows (expected) group consumption, conditional on outputs $q_1$ and $q_2$ on technologies one and two, that is, $c_g(q_1, q_2)$. Expected rather than actual group consumption is shown only because actual solutions delivered lotteries over adjacent consumption grid points, lotteries that would be degenerate for a sufficiently fine grid. The Pareto weight $l_1$ on agent one is on the *x*-axis. The weight of agent two is not shown but again with obvious normalization it is just $l_2 = 1 - l_1$.

There are four possible outputs, reflecting all the different combinations of high or low outputs on the two technologies. The legend describes which of the four lines corresponds to which output combination. Not shown in the graph is the optimal labor assignment, but here for the full range of $l_1 \in [0.0, 0.5]$ both agents are assigned the high labor effort.

Starting from the left side of the graph, at $l_1 = 0.0$, agent one receives no weight within the group.[10] Not shown in the graph is individual consumption. Of course, at $l_1 = 0.0$, agent one's consumption is zero for all outputs, while agent two consumes the entire group consumption $c_g$.[11] Low consumption and high effort for agent one have little consequence for group utility because agent one has low weight. Essentially, agent one is a "serf".

Internally, having agreed on the distribution of welfare implicit in $m$, agent one abides by the agreement, which is to work hard and to consume little or nothing. Notice that $c_g(q_1, q_2)$ lines for $(q_h, q_h) = (20, 20)$ and $(q_l, q_h) = (2, 20)$ nearly coincide in Figure 2.4.1 at $l_1 = 0.0$ (any difference is due to numerical approximation). Coincident also are the lines for $(q_l, q_l) = (2, 2)$ and $(q_h, q_l) = (20, 2)$. Thus, group consumption does *not* depend on output from the technology utilized by agent one. The risk neutral principal provides full insurance on technology one because internal monitoring and perfect commitment take care of potential incentive problems for agent one. In contrast, agent two, the so-called "lord" of the group, has a high $l_2$ weight. At and near $l_1 = 0.0$, and $l_2$ near 1.0, the "mongrel consumer's" utility is nearly identical to that of agent two. Since the mongrel consumer cares (mostly) about the effort of agent two, the group must be given incentives to make him work hard. Thus, group consumption and agent two's consumption vary positively with the output of agent two on technology two.

This logic prevails more generally, as $l_1$ increases toward the symmetric weight $l_1 = 0.5$. Over this range group consumption $c_g$ depends primarily on output of agent two from technology two, namely $q_2$, though output $q_1$ of agent one from technology one becomes increasingly important. Accordingly, for $0.0 < l_1 < 0.5$, group consumption $c_g$ is ordered with technology outputs:

$$c_g(q_1 = 2, q_2 = 2) < c_g(q_1 = 20, q_2 = 2) < c_g(q_1 = 2, q_2 = 20) < c_g(q_1 = 20, q_2 = 20).$$

Again, the ordering of $c_g$ with respect to $q_1$ and $q_2$ reflects the relative importance of each agent as described above. At $l_1 = 0.5$ the consumption allocation is nearly symmetric.[12] This movement of $c_g$ with outputs $q_1$ and $q_2$ evident in Figure 2.4.1 also conveys information about weights $l_1$, (and hence weights $m_1$), beyond the information contained in sharing rule (2.4.3).

Another parameter to be identified, if possible, is the outsider utility $\overline{W}$. Roughly, individual consumption increases as $\overline{W}$ decreases, but there are interesting implications for labor effort as well. At low enough levels of the principal's utility $\overline{W}$, and hence high weights for the agents, it is no longer optimal for both agents to work hard. At $l_1 = 0.0$ agent one still works hard and consumes virtually nothing. There is still no incentive issue there. But agent two now also works the low amount, so there is no need to give him an incentive to work hard. The allocation in this extreme case is fully equivalent to the allocation of the full information version of the model! Neither individual nor group consumption is a function of the output vector. As $l_1$ increases, however, the group cares more and more about agent one's welfare and is more tempted to have him shirk. Consequently, incentives need to be introduced to induce agent one's effort. Furthermore, as $l_1 \to 0.5$ agent two is eventually assigned the high effort as well. At this point, group consumption depends on the output of both technologies as in the earlier experiment.

Of course, at yet lower $\overline{W}$ agents receive higher levels of consumption, as the principal is extracting less surplus. At sufficiently low $\overline{W}$ both agents work the low effort in either regime. The two regimes could not then be distinguished by consumption and effort data, even though we recover parameter $\overline{W}$.

We now return to the potential divergence in Program 2.3.3 between $l$ weights in the objective function and $m$ weights in the group incentive constraint. Figure 2.4.2 plots the expected value of $m_1$ as a function of $l_1$ for $\overline{W} = 20$ and for $l_1 \in [0.0, 0.5]$. Starred points in the graph indicate $m_1$ that are realized with positive probability. When $l_1 = 0.10$ in Figure 2.4.2, the optimal contract contains a lottery over $m_1 = 0.0$ and $m_1 = 0.03$. For this parameterization and both realizations of $m$, the within-group distribution of weights is more unequal than in the master program objective function. When $l_1 = 0.15$ the optimal solution is a lottery over $m_1 = 0.24$ and $m_1 = 0.25$. In this case for both realizations of $m_1$ the within-group distribution is more symmetric than is $l_1$. In this sense, then, the ultimate degree of inequality is endogenous, though it is determined in large part by the degree of inequality in $l$, more generally. Figure 2.4.2 allows us to go back and forth between implications as a function $m$ and implications as a function of $l$.

It seems difficult to say in general whether it is optimal for the $m$ -weights to be greater or less than the $l$ -weights. Figure 2.4.2 displaying all the computed solutions make clear there are no strong theorems. We are able to say, however, that the effect is due to the incentive constraint, that is, to the attempt to weaken it. Surprisingly, the consumption grid makes this effect easier to see than it would be otherwise. For example, at $l_1 = 0.10$ group consumption is in a low enough range (0 to 10) and the consumption grid is sufficiently coarse that the consumption sharing rule $c_i(c_g, m)$ is the same whether or not $m_1$ equals 0.00, 0.03, or 0.10, the latter being the value of $l_1$. In this special case, the only effect of setting $m \neq l$ is to alter the relative disutilities of the two agents' efforts in Program 2.3.3's version of the incentive constraints (2.3.7).[13] In particular, setting $m_1 < l_1$ makes it easier to implement the high effort on the first technology because the group cares less about the first agent's effort. Similarly, setting $m_1 < l_1$ makes it harder to implement the high effort on the second technology because the group cares more about the second agent's effort. In this example, the first effect dominates.

In general, the first effect need not dominate. As Figure 2.4.2 illustrates, for $l_1 = 0.15$ the optimal allocation contains a lottery over $m_1 = 0.24$ and $m_1 = 0.25$. Again, aggregate consumption and the coarseness of the consumption grid are such that the consumption sharing rule is the same for $m_1 = 0.15$, $m_1 = 0.24$, and $m_1 = 0.25$ so the only effect of setting $m \neq l$ is the direct effect in the incentive constraint from altering disutility of effort on incentives. Unlike the previous case, the program now prefers to set $m_1 > l_1$, making it easier to implement effort on technology two and harder to implement effort on technology one.

For certain classes of preferences and technologies, however, we know that $m$ and $l$ must be equal. To see this return to Program 2.3.2, in which labor is transferable across technologies. Assume that sets of feasible consumption and efforts are continua, just as we did in the analysis of internal group sharing rules, but retain the CRRA preference specification of the example just given. These preferences aggregate in the sense of Gorman (1954). The relation for consumption was noted earlier in (2.4.3). For effort the relation is similar,

(2.4.4) $$e_i = \frac{m_i^{1/(1-a)}}{m_1^{1/(1-a)} + m_2^{1/(1-a)}} e_g, i = 1,2.$$

Individual allocations of consumption and effort are linear in group aggregates.

If we substitute these expressions back into the weighted utility function of the incentive constraint (2.3.7) in Program 2.3.2, a common constant, $k(m)$, can be pulled out from both sides of the equation, leaving only a utility function expressed in aggregates $c_g$ and $e_g$. Similarly, in the objective function (2.3.5) of Program 2.3.2 a scalar $k(l)$ can be pulled out, so the objective function consists of that scalar multiplied by a function of the aggregates $c_g$ and $e_g$. This constant makes no difference to the maximization problem, that is, to the information-constrained optimal choice of $c_g$ and $e_g$, though it is still related to the internal distribution of consumption and effort. Varying the weights within the group will not affect the group's schedule of payments to the outsider in any way.

Essentially, the weights $m$ and $l$ disappear from the problem of determining aggregates and it is as if the outsider were facing a single agent who has the choice of effort over the two technologies. The consumption and labor allocation to this "single agent" is determined as in the well-understood, classic, principal-agent model, while the consumption and labor allocations to the individuals should not depend on the payment schedule to the outsider. Both implications are testable in cross-sectional data.

The reason that CRRA preferences have this feature is that income expansion paths are linear. Equivalently, the distribution of Pareto weights does not effect the equilibrium marginal rates of substitution between effort and consumption. Consequently, all types have the same preference ordering over aggregate consumptions and aggregate effort. For further discussion of the Gorman class of preferences, of which CRRA is one type, see Gorman (1954) or the exposition in Townsend (1993).

Normally, when agents cannot work the technologies of each other, as in Program 2.3.3, this way of proceeding fails. In example 2.4.1, preferences do not aggregate in the example economy with CRRA preferences because labor is not re-allocable. Other special preference specifications, however, provide exceptions. Exponential utility of the form $U_i(c_i, e_i) = -\exp(-(c_i - e_i)/g_i)$ also Gorman aggregates. Holmstrom and Milgrom (1990) use these preferences along with linearity assumptions on compensation schedules and assumptions on the technology to develop an environment where utility is transferable. In this environment, they then compared the relative performance and group regimes. They found that not only does the distribution of wealth not matter for regime choice but absolute levels of wealth do not matter either. They then showed that correlation in technology outputs is the critical factor for determining the relative merits for the two regimes. At a high enough level of correlation, the relative performance regime dominates. We turn now to this issue and demonstrate how in other formulations wealth effects play a salient role. Also, and related, we do not restrict ourselves to symmetric equilibria, as in Itoh (1993) and Ramakrishan and Thakor (1991).

**Numerical example 2.4.2**

When outputs across two technologies are correlated, the relative performance regime can dominate sometimes. To see this we consider a comparison of Programs 2.2.2 and 2.3.3, the relative performance and group regimes when the technology is such that agents must work their own technologies. The economy is identical to the one described in Example 2.4.1, in that high effort increases the probability of high output, except that here outputs are correlated, as shown in Table 2.4.1 below.[14] The upper left quadrant shows how probabilities change as agent one's effort increases from low to high, $a_l$ to $a_h$, agent two's effort is set at $a_1$, and the output on technology two is $q_1$. The remaining quadrants can be similarly described. The point of this technological specification is that output comparisons are very informative about efforts. For example, if both agents are assigned high efforts, as on row four of the matrix in Table 2.4.1, then the event $q_1 \neq q_2$ is very unlikely.

The key factor for our comparison in the correlated case is the degree of inequality or status within the group. Figure 2.4.3 shows slices of the three-dimensional Pareto frontier for both regimes when the principal is constrained to receive various levels of promised utility $\overline{W}$. The solid lines show the frontier for the group regime, Program 2.3.3, and the hyphenated lines show the frontier for the relative performance regime, Program 2.2.2.

Starting from the upper left-hand panel, the principal's utility is $\overline{W} = -9.5$ (w in the graph). At symmetric $l$-weights, as on a 45-degree line, the program is indifferent between the two regimes. The reason, as alluded to above, is that in both regimes, both agents are recommended the low efforts so there is no incentive problem regardless of the regime. At asymmetric $l$-weights, however, the group regime dominates. Despite the large amount of transfers the principal is making to the group, the program still wants to transfer resources away from the low $l_1$-weight ("poor") to the high $l_2$-weight ("rich") agent. Consequently, the program wants the poor agent to work hard and the rich agent to work little. For these asymmetric effort levels, the group regime is effective because the rich agent can force the poor agent to work hard. In sum, allowing some noise in the observations, the group collective regime is more likely to be seen as one moves across small economies with higher and higher degrees of internal inequality.

As the principal's promised utility $\overline{W}$ increases, the relative performance regime begins to dominate for a broader range around the symmetric $l$-weights. This pattern is most evident in the second panel, where $\overline{W} = 6.0$, though the degree of dominance is strongest in the third $\overline{W} = 17$ panel. The relative performance regime strictly dominates at more or less symmetric weights because at higher levels of $\overline{W}$ both agents are required to work the high effort. Here relative performance is very effective, for the reasons discussed earlier. The group regime is susceptible to both agents simultaneously deviating.

Nevertheless, the group regime still dominates at relatively asymmetric levels of $l$-weights. The reason for this pattern is that the group regime is still much more effective at extracting wealth from a single agent than the relative performance regime. As above, the group regime can implement allocations that entail one agent working the high amount while receiving low consumption. At $\overline{W} = 6, 17,$ and $19$ surplus is transferred from this agent to the other agent and to the principal. In contrast the relative performance regime is limited, for incentive reasons, in its ability to extract wealth. Some consumption always needs, for incentive reasons, to be transferred to an agent working the high amount.

Accordingly, the higher is $\overline{W}$, the lower is consumption, and ceteris paribus, the greater is the range of utilities that deliver the group regime. Eventually, for a high enough levels of $\overline{W}$, the relative performance frontier vanishes and only the group regime can be used to extract enough resources to satisfy the principal's participation constraint. In this sense, the group regime is more likely the higher is $\overline{W}$, or equivalent, the lower is the wealth or status of local members relative to outsiders. However, the $\overline{W} = -9.5$ case provides an exception on the other extreme.

To summarize, the relative performance regime dominates at symmetric levels of utility. Output comparison of decentralized units of economic organization is very effective. For asymmetric distributions of utility, the group organization dominates. Thus, multi-agent organizations exist in order to extract wealth from some but not necessarily all members of the organization. Essentially the relatively "rich" agent is acting as a delegated enforcer. This effect becomes more acute as the wealth of the outsider increases.

**2.5 Pareto Problem with Choice of Organization**

We return to Example 2.4.2 and note that Figure 2.4.3 demonstrates that the Pareto frontier (the exterior of the relative performance and group regime frontiers) is not convex. This means that randomization over points along this frontier may improve welfare. Utilities obtainable from randomization are captured geometrically by taking the convex hull of the Pareto frontier described in Figure 2.4.3.[15] More formally, we incorporate this class of contracts by introducing notation that allows the principal not only to randomly assign consumptions, outputs, and efforts, but also to randomly assign the agents' organizational regime. To see how to combine the relative performance and group programs, we first modify the notation developed earlier. Let $p^r(c,q,e_{1\bullet},e_{2\bullet})$ denote the joint probability of a relative performance assignment and a consumption, output, effort assignment of $(c,q,e_{1\bullet},e_{2\bullet})$. If we interpret the choice variable $p(c,q,e_{1\bullet},e_{2\bullet})$ in Section 2.2 as the conditional probability of $(c,q,e_{1\bullet},e_{2\bullet})$ and let $p(r)$ denote the unconditional probability of assigning the agents the relative performance regime, then $p^r(c,q,e_{1\bullet},e_{2\bullet}) = p(c,q,e_{1\bullet},e_{2\bullet})p(r)$. Of course, the unconditional probability of the agents being assigned the relative performance regime is then $\sum_{c,q,e_{1\bullet},e_{2\bullet}} p^r(c,q,e_{1\bullet},e_{2\bullet})$. Similarly, we let $p^g(c_g,q,a_1,a_2,m)$ denote the joint probability of the agents' being assigned the group regime and the $(c_g,q,a_1,a_2,m)$ allocation.

In the combined program, the principal now chooses the vector $(p^r(c,q,e_{1\bullet},e_{2\bullet}), p^g(c_g,q,a_1,a_2,m))$, the joint probability of the organizational assignment along with associated allocations. To make the notation consistent between the two regimes we develop the combined program under the assumption that preferences are separable. As we discussed earlier, the problem and notation can be easily generalized to case of non-separable preferences.

Now that regimes may be assigned randomly, we need to modify the principal's participation constraint. It takes the form

$$(2.5.1) \quad \sum_{c,q,e_{1\bullet},e_{2\bullet}} p^r(c,q,e_{1\bullet},e_{2\bullet})W(q_1+q_2-c_1-c_2) + \sum_{c_g,q,a_1,a_2,m} p^g(c,q,a_1,a_2,m)W(q_1+q_2-c_g) \geq \overline{W}.$$

Proceeding in this way for objective functions and other constraints we can merge Programs 2.2.1 and 2.3.3 into

**Program 2.5.1.**

Maximize by choice of $p^r(c,q,e_{1\bullet},e_{2\bullet}) \geq 0$ and $p^g(c_g,q,a_1,a_2,m) \geq 0$ the objective function

$$(2.5.2) \quad \begin{aligned} &\textstyle\sum_i l_i\{ \sum_{c,q,e_{1\bullet},e_{2\bullet}} p^r(c,q,e_{1\bullet},e_{2\bullet})[U_i(c_i)+V_i(T_i-e_i)] \\ &+ \textstyle\sum_{c_g,q,a_1,a_2,m} p^g(c_g,q,a_1,a_2,m)[U_i(c_i(c_g,m))+V_i(T_i-e_i(e_g,m))]\} \end{aligned}$$

subject to participation constraint (2.5.1) above; with $p^r(c,q,e_{1\bullet},e_{2\bullet})$ satisfying the relative performance incentive constraints (2.2.3), (2.2.4), and the relative performance technology constraints (2.2.5); with $p^g(c_g,q,a_1,a_2,m)$ satisfying the group incentive constraints (2.3.7), and the group technology constraints (2.3.8); and subject to a probability constraint

$$(2.5.3) \quad \sum_{c,q,e_{1\cdot},e_{2\cdot}} p^r(c,q,e_{1\bullet},e_{2\bullet}) + \sum_{c_g,q,a_1,a_2,m_1} p^g(c_g,q,a_1,a_2,m) = 1.$$

As previously described, the solution to this problem is well defined given weights $(l_1, l_2)$ and promised utility $\overline{W}$. If the program is solved for the full range of Pareto weights and promised utility, we can trace out the entire Pareto frontier.[16] This makes more precise our earlier statement that the likelihood of a regime depends on internal inequality. That is, the likelihood here is the contract probability.

## 3. Collective Organizations as Information Monitors

This section analyzes a second class of prototypes. These are designed to study an alternative definition of a collective organization. Organizations are characterized in this section by the degree to which agents monitor each other through the joint operation of production technologiesThe types of organizations that are feasible in this model include sole proprietorships (no joint assignments), a single production unit (all joint assignments), and various combinations of these two extremes (some technologies jointly assigned and others not). As in Section 2, monitoring is an important motivating force for organizations, but in this section we rule out collusion and deliberately impose a different tradeoff between organizational structures.

**3.1 The Environment and Commodity Space of a Small Three-Person Economy**

Imagine that there are two agents, an outsider, and three technologies. As in Section 2, we denote by $e_{i\bullet} = (e_{i1}, e_{i2}, e_{i3})$ the vector of technology-specific efforts of agent *i*. The probability of the output vector $q = (q_1, q_2, q_3)$ across the three technologies is described by the function $p(q \mid a_1, a_2, a_3)$ where $a_i$ is the aggregate effort on technology *i*. The production function can be specified quite generally to incorporate complementarities in production and correlated returns, but for the examples below we assume that the return on technology *j* depends only on the total effort $a_j$ on technology *j* and that technology returns are uncorrelated and otherwise identical.

A given technology may be assigned in two distinct ways. If technology *j* is assigned only to agent *i*, then agent *i* alone can work it. In this case his effort $e_{ij}$ is private information, and thus there is a potential moral-hazard problem. Alternatively, technology *j* may be assigned jointly to agents one and two. In this case they observe each other's efforts and so by implementation arguments each may be induced to costlessly reveal each other's efforts to the outsider.[17] More properly, there exists an equilibrium of the direct revelation mechanism in which efforts can be revealed to outsiders.[18] Of course, this argument only works if there is no collusion against the outsider and, as previously noted, we rule this out in this section. Instead, we will focus here on a collective organization as one which achieves common internal information via joint operation of technologies.

There are numerous possible assignments of agents to technologies, including joint and individual assignments. But since the return of technology *j* is a function of total effort on that technology, $a_j = e_{1j} + e_{2j}$, and is independent here of the returns on other technologies, it does not matter whether agent *i* works technology *j*, per se. What matters is the *number* of private and group technologies worked by both agents. Table 3.1.1 lists the possible assignments.

Columns one and two list the number of technologies worked solely by agents one and two, respectively, while column three lists the number of technologies worked jointly. Row six, the (1, 1, 1) assignment, indicates that each agent works one technology by himself and one technology jointly. The assignment in the fourth row, (3, 0, 0), means that agent one works all three technologies by himself. The assignment (0, 0, 3) in the last row means that every technology is used in joint production. If we let D denote the set of assignments, and let $d \in D$ denote a particular assignment, that is, a row in Table 3.1.1, we can proceed as we did in Section 2.5 for the group and relative performance regimes.

As we assumed earlier, joint operation of technologies makes efforts full information. To make the choice between joint and individual production non-trivial, we impose two costs on joint production. First, to be consistent with the idea that the agents monitor by working together, we require that agent's efforts be supplied in equal amounts on jointly operated technologies. Second, we impose a utility cost from the organizational assignment. In the examples, we will assume that there is only disutility from jointly working a technology. The idea is that there are some diseconomies of scale or coordination costs associated with the number of projects that agents work together.[19]

Agent *i*'s preferences are denoted by $U_i(c_i, T_i - e_i) - g(d)$, where $g(d)$ represents the disutility from the organizational assignment. As in Section 2, the outsider's preferences are over the agents' surplus, and they are written $W(q_1 + q_2 + q_3 - c_1 - c_2)$.

Let $S = C \times Q \times E_{1\bullet} \times E_{2\bullet}$ be the cross product of the sets of feasible points of consumption, output, and effort. As before, we assume that each set contains a finite number of points. Further, each set $E_{i\bullet}$ contains vectors that indicate agent *i* may work zero effort on any or all of the three technologies. Specifically, to make assignments essential we assume that if agent *i* is not assigned to technology *j*, then he supplies zero effort on it, that is $e_{ij} = 0$. If he is assigned to technology *j*, then we require for $e_{i\bullet}$ to be feasible that $e_{ij} > 0$. This assumption, plus the requirement that $e_{1j} = e_{2j}$ for all jointly operated technologies *j*, means that each organizational assignment necessitates restricting *S* in a different way. We can capture this by creating the notation $S_d$ to indicate feasible grid points for organizational assignment *d*. The problem will then be to choose a probability distribution over vectors indexed by the organization dependent grids, that is, over $(S_{(0,3,0)}, S_{(1,2,0)}, \ldots, S_{(0,0,3)})$.

In the interest of brevity, we describe in detail the grid, $S_d$, and the incentive constraints, for only the d = (1,1,1) assignment. Later, we will indicate how to modify the grid and incentive constraints to handle other assignments. Further, without loss of generality, we assume that for the (1,1,1) assignment, agent one's technology is technology one, agent two's technology is technology two, and the group's technology is technology three.

The set of feasible grid points for the (1, 1, 1) assignment is

(3.1.1) $$S_{(1,1,1)} = \{(c, q, e_{1\bullet}, e_{2\bullet}) \in S \mid e_{11} > 0,\ e_{12} = 0,\ e_{21} = 0,\ e_{22} > 0,\ e_{13} = e_{23} > 0\}.$$

The conditions on $S_{(1,1,1)}$ require that agent one works positive effort on technology one, does not work technology two, and works an amount on technology three that is not only positive but equal to agent two's effort on that technology. The conditions similarly restrict agent two's feasible grid points. Other restrictions, such as maximum total effort per technology or maximum total effort by an agent, are easily incorporated into $S_d$ and will be done so in the numerical examples below.

Given the technology assignment, the choice variable for the program is $p^d(c, q, e_{1\bullet}, e_{2\bullet})$, the joint probability of an organizational assignment, consumption, outputs, and efforts. (Note that the notation implicitly assumes that $(c, q, e_{1\bullet}, e_{2\bullet})$ are the feasible grid points for assignment *d*.) Aside from the grid, that is, the set $S_d$, only the incentive constraints are affected by the organizational structure. Consequently, we describe them before writing out the entire program.

For the (1,1,1) assignment there are two sources of private information, agent one's action on technology one and agent two's action on technology two. The allocation $p^{(1,1,1)}(c, q, e_{1\bullet}, e_{2\bullet})$ must satisfy incentive constraints for agent one, that is, for each $e_{11}$, $e_{13}$, $e_{23}$,

$$
\begin{aligned}
&\sum_{c,q,e_{2\bullet}} p^{(1,1,1)}(c, q, e_{1\bullet}, e_{2\bullet})[U_1(c_1, T_1 - (e_{11} + e_{13})) - g((1,1,1))] \\
&\geq \sum_{c,q,e_{2\bullet}} p^{(1,1,1)}(c, q, e_{1\bullet}, e_{2\bullet}) \frac{p(q \mid \hat{e}_{11}, e_{22}, e_{13} + e_{23})}{p(q \mid e_{11}, e_{22}, e_{13} + e_{23})}[U_1(c_1, T_1 - (\hat{e}_{11} + e_{13})) - g((1,1,1))]
\end{aligned}
\tag{3.1.2}
$$

for all feasible deviations $\hat{e}_{11}$. The incentive constraint is designed to stop agent one from deviating on technology one. Since the other agent does not work technology one, total effort on technology one is $e_{11}$. Because the agents' efforts on technology three are public, agent one must follow the $e_{13}$ recommendation on technology three.

Agent two's incentive constraints are similar, taking the form that the allocation $p^{(1,1,1)}(c, q, e_{1\bullet}, e_{2\bullet})$ must satisfy for each $e_{22}$, $e_{13}$, $e_{23}$,

$$
\begin{aligned}
&\sum_{c,q,e_{1\bullet}} p^{(1,1,1)}(c, q, e_{1\bullet}, e_{2\bullet})[U_2(c_2, T_2 - (e_{22} + e_{23})) - g((1,1,1))] \\
&\geq \sum_{c,q,e_{1\bullet}} p^{(1,1,1)}(c, q, e_{1\bullet}, e_{2\bullet}) \frac{p(q \mid e_{11}, \hat{e}_{22}, e_{13} + e_{23})}{p(q \mid e_{11}, e_{22}, e_{13} + e_{23})}[U_2(c_2, T_2 - (\hat{e}_{22} + e_{23})) - g((1,1,1))]
\end{aligned}
\tag{3.1.3}
$$

for all feasible deviations $\hat{e}_{22}$.

Grids and incentive constraints could be written out for the other organizational structures. For example, the (3,0,0) technology assignment would require that $S_{(3,0,0)} = \{(c, q, e_{1\bullet}, e_{2\bullet}) \in S \mid e_{1j} > 0,\ e_{2j} = 0,\ j = 1,2,3\}$ and that in addition incentive constraints prevent agent one from deviating on any combination of the three technologies. The (0, 0, 3) assignment, that is, the case where all technologies are jointly operated, would require $S_{(0,0,3)} = \{(c, q, e_{1\bullet}, e_{2\bullet}) \in S \mid e_{1j} = e_{2j} > 0,\ j = 1,2,3\}$ and no incentive constraints. In the interest of brevity we do not present the various combinations explicitly.

With the domain and incentive constraints specified, we can now proceed to the description of the program. The Pareto problem is to maximize the weighted sum of agent's utilities by choosing $p^{d}(c,q,e_{1\bullet},e_{2\bullet})$ over each domain $S_d$. But as in Section 2.5 we jump immediately to the specification where assignment $d$ is chosen (at random) as well. Recall that $g(d)$ is the disutility accruing to each agent from the organizational assignment. Letting $(l_1, l_2)$ be the Pareto weights on the agents and $\overline{W}$ the reservation utility of the principal, the program is

**Program 3.1.1:**

Choose the $p^{d}(c,q,e_{1\bullet},e_{2\bullet}) \geq 0$ to maximize

(3.1.4) $$\sum_{d,c,q,e_{1\bullet},e_{2\bullet}} p^{d}(c,q,e_{1\bullet},e_{2\bullet})(l_1[U_1(c_1,T_1-e_1)-g(d)]+l_2[U_2(c_2,T_2-e_2)-g(d)])$$

subject to the organization-dependent incentive constraints like (3.1.2) and (3.1.3), the principal's participation constraint

(3.1.5) $$\sum_{d,c,q,e_{1\bullet},e_{2\bullet}} p^{d}(c,q,e_{1\bullet},e_{2\bullet})W(q_1+q_2+q_3-c_1-c_2) \geq \overline{W},$$

technology constraints

(3.1.6) $$\forall d,\bar{q},\bar{e}_{1\bullet},\bar{e}_{2\bullet},\ \sum_{c} p^{d}(c,\bar{q},\bar{e}_{1\bullet},\bar{e}_{2\bullet}) = p(\bar{q}|\bar{e}_{1\bullet},\bar{e}_{2\bullet})\sum_{c,q} p^{d}(c,q,\bar{e}_{1\bullet},\bar{e}_{2\bullet}),$$

and a probability measure constraint

(3.1.7) $$\sum_{d,c,q,e_{1\bullet},e_{2\bullet}} p^{d}(c,q,e_{1\bullet},e_{2\bullet}) = 1.$$

## 3.2 Theoretical and Empirical Implications of the Assignment Regimes

A given assignment, or organizational structure, implies a potentially limited insurance regime. When an agent works non-trivial efforts on a technology as a sole proprietor, he must be induced to carry out those efforts. In this respect, at least, much of the earlier discussion of the classic agency model applies here. There will not be in general full insurance for an agent *i* assigned nontrivial effort on individually operated technology $j$, that is, $c_i$ will move with $q_j$ in a manner discussed earlier, with the outputs associated with high efforts rewarded.

Still, many assignments with individual proprietorships also allow joint production on the remaining technologies. As there are no incentive problems on such technologies, outputs of those technologies will be fully insured by a risk-neutral principal. That is, holding returns on individually operated technologies fixed, consumptions will not vary with output produced on the jointly operated technologies. Indeed, we could identify a level of consumption for each of the agents that comes from the output of these joint technologies, but which is smoothed at mean values completely by the principal. In this respect, at least, there is full insurance within the group in the sense of the group regime of Section 2. But unlike the group regime of Section 2, the principal fully insures fluctuations in group output.

With respect to *total* consumption and *total* output produced by group members, that is, including on individually operated technologies, these interpretations change. First, there is not within-group full insurance. As long as there are individually operated technologies and incentive constraints bind, individual consumptions must comove with output on individually operated technologies. Second, for the same reason, the group's *total* output is not fully insured by the principal.

As $l$ -weights and the reservation utility $\overline{W}$ are varied, we would expect *a priori* the obvious impact: Poor agents work hard and receive less consumption while rich agent work less and receive more consumption. Exceptions occur, however, as inequality effects and incentive constraints interact. A poor agent working hard and receiving low levels of consumption on average from a given technology will be subject to much risk. This consumption risk hurts the agent. But consumption cannot be smoothed while keeping high effort incentive compatible. To avoid these effects, a relatively rich agent can be assigned as a monitor. While this augments the latter's disutility, the positive effect is to allow full insurance and remove incentive constraints on that technology. This effect can be augmented if both agents are poor relative to the principal as we show in examples below. Note, finally, that costly monitoring in the form of joint and equal effort in production can limit effort-sharing and efficiency in production within the set of agents constituting a group. For example, equal efforts need not satisfy the effort sharing rules (2.4.4) of the groups in Section 2, and the marginal product of effort on the various technologies need not be equated.

**3.3 Examples**

To illustrate the effect of Pareto weights and reservation utility on the optimal organizational structure, we present two numerical examples. Both examples use the following grid spaces

$$C = C_1 \times C_2 = [0.00, 0.02, 0.04, ..., 1.60] \times [0.00, 0.02, 0.04, ..., 1.60].$$

Consumptions thus can take on a large number of values, enough so that the grid allows us to capture the effect of risk aversion reasonably well. There are, however, only two outputs per each technology, at levels of 0 and 1. The set of feasible outputs is thus

$$Q = Q_1 \times Q_2 \times Q_3 = \{0,1\} \times \{0,1\} \times \{0,1\}.$$

Finally, efforts are restricted to sets of non-negative integers such that no more than 12 total units of effort can be taken by individual *i*. The set of efforts is thus

$$E_{i\bullet} = \{e_{i\bullet} \in (I \times I \times I) \big| \textstyle\sum_j e_{ij} \leq 12\},$$

where the set of technology specific effort is the set of integers $I = \{0,1,2,\ldots\}$.

Other restrictions depend on the assignment, as noted. As before, we only write the (1,1,1) assignment portion of the grid. Recalling that $S = C \times Q \times E_{1\bullet} \times E_{2\bullet}$, then

(3.4.1) $S_{(1,1,1)} = \{(c,q,e_{1\cdot},e_{2\cdot}) \in S \mid \forall j, \sum_i e_{ij} \le 8, e_{11}, e_{22} > 0, e_{12} = e_{21} = 0, e_{13} = e_{23} > 0\}.$

Note that for this example total effort on a single technology is restricted to be no more than 8 units of effort. This condition applies to all of the possible assignments. Assuming outputs are drawn independently over technologies, then the probability distribution of the output on each technology j is written in the obvious notation as

$$p(q_j = 0 \mid e_{1j} + e_{2j}) = 1 - ((e_{1j} + e_{2j} - 0.9)/50)^{0.2},$$
$$p(q_j = 1 \mid e_{1j} + e_{2j}) = ((e_{1j} + e_{2j} - 0.9)/50)^{0.2}.$$

Thus, the probability of the high output $q_j = 1$ on technology $j$ varies from 0.2885, if one unit of effort is supplied, to 0.7401, if the maximum feasible amount of effort, eight, is supplied. The probability distribution of the high output is concave in effort and there is much curvature at a total effort of two units. The joint distribution of the vector of technology outputs is easily calculated.

In the examples below the agents' preferences are identical. Preferences for agent *i* are

$$U(c_i, T_i - e_i) - g(d) = c_i^{\alpha} - k_e(e_{i1} + e_{i2} + e_{i3})^{\gamma} - k_g y,$$

where $\alpha$ is the degree of risk aversion in consumption, $k_e$ and $\gamma$ are the degrees of work disutility and aversion, respectively, and $k_g$ is the degree of group disutility, where *y* as the number of technologies jointly worked by the group in an assignment *d*. Several of these parameters will be varied in the examples which follow, but again most of the comparative statics exercises focus on inequality, changing the principal's utility $\overline{W}$ and the Pareto weights, $\lambda_1$ and $\lambda_2$, as before.

**Numerical Example 3.3.1**

Our first example in this section uses the following parameters:

$$\alpha = 0.2, \quad k_e = 0.05, \quad \gamma = 1.0, \quad k_g = 0.005, \quad \overline{W} = 0.6.$$

Table 3.3.1 lists the optimal technology assignment as a function of the Pareto weights, for the range $\lambda_1 \in [0.0, 0.5]$. Again, at $\lambda_1 = 0.0$ there is the most inequality in status, while at $\lambda_1 = 0.5$ the two agents are valued equally.

At $\lambda_1 = 0.5$ all production is joint. The entire economy is one large group. Here, unlike section 2, internal equality is a force for collective group regimes. Not shown are the consumption features of the optimal contract, but not surprisingly agents are fully insured because there are no hidden efforts. Quickly, however, as the Pareto weights are changed and we have an increase in local inequality, the economy switches to a (3,0,0) assignment and no groups. One reason the switch is so sharp is that at $g = 1$ disutility of effort is linear. Consequently, there is no loss to the objective function in assigning all effort to the low Pareto weight agent. Indeed, ceteris paribus the program prefers this effort assignment for the computed $l_1$ that are less than $l_2$. The issue in this case, then, is whether or not the program can efficiently implement the efforts, and for this parameterization it can.

**Numerical Example 3.3.2**

Other parameter values deliver other assignments and interesting monitoring behavior. Consider the parameterization

$$a = 0.5, \quad k_e = 0.01, \quad g = 2.0, \quad k_g = 0.02, \quad \overline{W} = 0.5.$$

The key difference from the previous example is that by setting $g = 2.0$ there is the increasing marginal disutility of effort. This along with higher work aversion makes it valuable to distribute effort more equally between the two agents.

Table 3.3.2 lists the optimal technology assignments as a function of the Pareto weight, $l_1$. At $l_1 = 0.5$, we observe the symmetric assignment (1,1,1), that is, individual proprietorship plus joint production. The allocation looks much as one might expect. Agents' contracts are identical and an agent’s consumption is solely a function of output on his privately worked technology. In this example, an agent receives higher consumption if he produces a higher level of output on his own technology. As $l_1 \to 0$, namely by $l_1 = 0.4$, the symmetric (1,1,1) assignment is replaced by the asymmetric (2,1,0) assignment. With each agent working entirely on his own, the entire economy consists of two sole proprietorships, though agent one’s proprietorship is larger, in the sense that agent one is working two technologies, not one.

Interestingly, a jointly operated technology reappears at yet more unequal Pareto weights. That is, as in Section 2 groups reappear for extreme internal inequality. At $l_1 = 0.1$, (the second row in Table 3.3.2), the optimum is a lottery over the (1,0,2) and (2,1,0) assignments with probabilities of 0.702 and 0.298, respectively. At $l_1 = 0.0$, only the (1,0,2) assignment is part of the optimum. In this assignment, agent one is the only agent working a technology on his own, but now he is being monitored on the other two technologies by agent two who consequently must also supply effort. The program has decided to use the second agent as a monitor and make him share work with agent one. As the first agent receives virtually zero weight in the objective function, consumptions are unequal, here more unequal than the distribution of work effort.

Another interesting parameter to vary is the principal's utility $\overline{W}$. Table 3.3.3 lists optima as a function of $\overline{W}$ holding $l_1 = 0.5$ fixed. Starting from the bottom of the table, at $\overline{W}$ equal to 1.75 or 1.5, the entire economy is one large production unit. All fluctuations in output are borne by the principal; but the agents bear the utility cost of working together, they work high levels of effort, and they receive low levels of consumption. The contract is structured this way in order to make large transfers to the principal. The level of $\overline{W}$ requires so much resource extraction from the agents that the consumption levels have to be low, too low to induce the necessary high efforts on individually worked technologies. Consequently, the program gives up on avoiding the utility cost from working group technologies and uses the agents to monitor each other. (Recall one more time we preclude collusion here.)

As $\overline{W}$ is lowered, and the status of the local residents improves, the number of jointly operated technologies declines. There are no exceptions in the example here, whereas in Section 2 this relationship held with exceptions. At $\overline{W} = 1.25$, there is a lottery over the (1,1,1) and (0,0,3) assignments with probabilities of 0.124 and 0.876, respectively. As $\overline{W}$ declines further, the solution again becomes degenerate and the symmetric (1,1,1) assignment is optimal. Interestingly, the program skips the (1,0,2) and (0,1,2) more limited group assignments. These assignments require too much asymmetry in effort for the program to find them appealing.

**4. Pareto Problems for Large Economies and a Partial Decentralization**

The contract regimes for the small 3-person economies of section 2 were written in the notation $p^r(e, \mathrm{g}, e_{1\cdot}, e_{2\cdot})$ and $p^g(c_g, q, a_1, a_2, m)$ for the relative performance and group regimes, respectively, of section 2, and $p^d(c, q, e_{1\cdot}, e_{2\cdot})$ for the possible d-assignments of section 3. As has been implicit in the discussion, though, local economies $k$ could vary in the preferences $U_{ki}(c_i) + V_{ki}(T_{ki} - e_i)$, technologies, $p_k(q \mid e_{1\bullet} + e_{2\bullet})$ and especially the Pareto weights $(l_{k1}, l_{k2})$ of local residents. Then contract regimes could vary across these local economies k varying in particular with the degree of local inequality. Indeed, with a continuum of economies of each type we can let $a_k$ be the relative number of economies of type $k$ and let $r_k$ be the relative Pareto weight of economy $k$. One can then write down a mechanism design problem similar to Programs 2.5.1 and 3.1.1, but here for the larger single economy. We would then maximize weighted sums of type utilities subject to the restrictions on the commodity space enumerated earlier, and subject to a single economy-wide resource constraint that the surplus when added across all small economies be no less than zero, that is, in effect, that the utility of a single risk neutral principal be at least zero. The solution, when reinterpreted, does allow interactions among the local economies, in particular, in the provision of insurance for local fluctuations among local economies of the same type. When there are nontrivial lotteries for an economy of a given type, putting mass among the various possible contractual regimes, convexifying the Pareto frontier, then the extended model predicts coexistence of regimes, as observed in data. In principle, transfers across economies of different types are also allowed, though the natural benchmark would be no transfers across types, or equivalently here, $\overline{W} = 0$.

In an earlier working paper, we have succeeded in a partial decentralization of the large economies in which each local economy interacts in larger economy-wide markets. The advantage of that decentralization is that it sets $\overline{W}$ to zero endogenously, unless, as in the second welfare theorem, there are wealth reallocations across economy types. We can also begin to talk about making the Pareto weights $(l_{k1}, l_{k2})$ endogenous, as determined by the endogenous value of individual endowments evaluated at equilibrium prices. It is this connection that underlies our interchangeable use of Pareto weights and wealth in our language. For work along this line see Prescott and Townsend (2000).

**5. Conclusion**

This paper uses mechanism design models to study collective economic organizations. In the tradition of Alchian and Demsetz (1972), monitoring plays a crucial role in these organizations. In the first class of models, assignments involving joint monitoring with collusion possibilities were identified as collective, multi-agent organizations. In the second class of models, the defining characteristic of a multi-agent organization was the joint operation of technology. As in Legros and Newman (1996), the distribution of wealth was found to play an important role in determining optimal organization structure. In the first model, internal inequality favored multi-agent organization. In the second model internal inequality favored single agent organization, but there were exceptions for extremes of internal wealth. The connection between the wealth distribution and multi-agent organizations also depended on the wealth of the local group members relative to outsiders. In both models, local poverty favored multi-agent organizations, but in the first model there were exceptions for extremes of local wealth. We also found that organizational structure can have testable implications for consumption and effort allocations and more generally that organizations and allocations are jointly determined.

Clearly, there are characteristics of collective organizations not in our models. For example, Alchian and Demsetz (1972) place great emphasis on the role of a boss or a centralized contractual party who is also a residual claimant. In our models, the exogenous outsider plays this role, but extensions to our models, such as the inclusion of a supervisory technology, could be used to study this characteristic. Grossman and Hart (1986) emphasize the role of residual control rights that emanate from incomplete contracts. Span of control models, as in Lucas (1978), emphasize limits on management technologies as organizations get large. Conceivably, these latter two characteristics can be added to our models. In general, we view our models as prototypes that can be built upon and altered to suit the application.

## Bibliography

Alchian, Armen A., and Harold Demsetz. "Production, Information Costs, and Economic Organization." *American Economic Review*, 62 (1972): 777-95.

Becker, Gary S. *A Treatise on the Family*. Cambridge, MA: Harvard University Press, 1981.

Benjamin, Dwayne. "Household Composition, Labor Markets, and Labor Demand: Testing for Separation in Agricultural Household Models." *Econometrica*, 60 (1992): 287-322.

Chiappori, Pierre-Andre. "Collective Labor Supply and Welfare." *Journal of Political Economy*, 100 (1992): 437-67.

Coase, R H. "The Nature of the Firm." *Economica*, New Series, 4 (1937): 386-405.

Cochrane, John H. "A Simple Test for Consumption Insurance." *Journal of Political Economy*, 99 (1991): 957-76.

Deaton, Angus. "On Risk, Insurance, and Intra-village Smoothing." Manuscript, 1993.

Feder, Gershon, Lawrence J. Lau, Justin Y. Lin and Luo Xiaopeng. "Agricultural Credit and Farm Performance in China." *Journal of Comparative Economics* 13 (1989): 508-26.

Gaynor, Martin and Paul Gertler. "Moral Hazard and Risk Spreading in Partnerships." *RAND Journal of Economics* 26 (1995): 591-613.

Gorman, W.M. "Community Preference Fields." *Econometrica* 21 (1953): 373-90.

Grossman, Sanford and Oliver Hart. "An Analysis of the Principal Agent Problem." *Econometrica* 51 (1983): 7-45.

Grossman, Sanford and Oliver Hart. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy* 94 (1986): 691-719.

Harris, Milton, and Robert Townsend. "Resource Allocation under Asymmetric Information." *Econometrica* 49 (1981): 33-69.

Hart, Oliver. *Firms, Contracts, and Financial Structure*. Oxford: Clarendon Press, 1995.

Hart, Oliver and Bengt Holmstrom. "The Theory of Contracts." In *Advances in Economic Theory: Fifth World Congress*. (Truman F. Bewley Ed.) Cambridge, England: Cambridge University Press, 1987: 71-155.

Holmstrom, Bengt. " Moral Hazard in Teams." *Bell Journal of Economics* (1982): 324-40.

Holmstrom, Bengt and Paul Milgrom. “Aggregation and Linearity in the Provision of Intertemporal Incentives.” *Econometrica* 55 (1987): 303-28.

Holmstrom, Bengt and Paul Milgrom. “Regulating Trade Among Agents.” *Journal of Institutional and Theoretical Economics* 146 (1990): 85-105.

Holmstrom, Bengt and Jean Tirole. “The Theory of the Firm.” In *Handbook of Industrial Organization*, Vol. I. (Richard Schmalensee and Robert D. Willig Eds.), Amsterdam, North-Holland, 1989: 61-133.

Itoh, Hideshi. “Incentives to Help in Multi-agent Situations.” *Econometrica* 59 (1991): 611-36.

Itoh, Hideshi. “Coalitions, Incentives, and Risk Sharing.” *Journal of Economic Theory* 60 (1993): 410-27.

Lang, Kevin and Peter-John Gordon. “Partnerships as Insurance Devices: Theory and Evidence.” *RAND Journal of Economics* 26 (1995): 614-29.

Legros, Patrick and Andrew F. Newman. “Wealth Effects, Distribution, and the Theory of Organization.” *Journal of Economic Theory* 70 (1996): 312-41.

Lucas, Robert E., Jr. “On the Size Distribution of Business Firms.” *Bell Journal of Economics*, 9 (1978): 508-523.

Mace, Barbara J. “Full Insurance in the Presence of Aggregate Uncertainty.” *Journal of Political Economy*, 99 (1991): 928-56.

Morduch, Jonathan, Stern, Hal. “Using Mixture Models to Address Unobserved Heterogeneity: An Examination of Gender Bias in Health Outcomes in Bangladesh. Manuscript (1994).

Townsend, Robert M. and Rolf A.E. Mueller. “Mechanism Design and Village Economies: Credit to Tenancy to Cropping Groups.” *Review of Economic Dynamics,* (1998): 119-172.

Prescott, Edward S. “Computing Moral Hazard Problems Using the Dantzig-Wolfe Decomposition Algorithm.” Manuscript, 1998.

Prescott, Edward S. and Robert M. Townsend. “ Firms as Clubs in Walrasian Markets with Private Information.” Manuscript, 2000.

Ramakrishnan, R.T.S. and A.V. Thakor. “Cooperation versus Competition in Agency.” *Journal of Law, Economics, & Organization* 7 (1991): 248-83.

Tirole, Jean. “Collusion and the Theory of Organizations.” In *Advances in Economic Theory Sixth World Congress,* Vol. II (Jean-Jacques Laffont Ed.), Cambridge, England: Cambridge University Press, 1992.

Townsend, Robert M. *The Medieval Village Economy: A Study of the Pareto Mapping in General Equilibrium Models*. Princeton, N.J.: Princeton University Press, 1993.

Townsend, Robert M. “Risk and Insurance in Village India.” *Econometrica*, 62 (1994): 539-91.

Udry, Christopher. “Gender, Agricultural Production, and the Theory of the Household.” *Journal of Political Economy*, 104 (1996): 1010-46.

[1] See Gaynor and Gertler (1995) and Lang and Gordon (1995) for discussions of law and medical partnerships.

[2] For papers using this computational method see Lehnert (1998), Lehnert, Ligon, and Townsend (1998), Phelan and Townsend (1991), and Prescott (1998).

[3] Indeed, correlation in returns is one force which pushes the assignment to individually operated technologies, even when operating both technologies is allowed. If an agent works both technologies then he could simultaneously shirk on both technologies, eliminating the principal's ability to infer efforts by output comparison. Assigning an agent to a single technology eliminates this strategy, assuming, of course, that an agent cannot sabotage production on the other technology.

[4] See Tirole (1992) for an analysis of models with collusion.

[5] With non-separable preferences consumption sharing rules take the more complicated form $c_i(c_g, e_1, e_2, m)$. No such similar simplification works for effort because with non-separable preferences, effort sharing rules depend on the returns to the group from working $e_g$. Consequently, the program would only be able to simplify the choice variable to $p(c_g, q, e_{1\bullet}, e_{2\bullet}, m)$ and the optimal internal distribution of efforts would need to be incorporated in the incentive constraints.

[6] In fact, if the previous substitutions were not made, then Program 2.3.2 would be identical to Program 2.2.1 except for the incentive constraints. Everything else, the grid space (excluding the $m$), the technology constraints, the participation constraint, and the probability measure constraints, would be unchanged.

[7] There are cases with uncorrelated returns where something akin to this effect can cause the principal to use relative performance and assign both agents to the same technology. One such case is where high efforts are much more informative about deviations than medium efforts. If it is desirable for both agents to work a medium effort, then assigning them both to work the same technology, so that total effort on that technology is high, may be better than having them work separate technologies. The reason is that the inference benefit from the high level effort can reduce and even eliminate any "free-riding" problem between the two agents.

[8] Our models above are not multi-period models. But again we imagine repeated sampling of the static agreement generating panel data.

[9] Indeed, note again that weights $m$ internal to the group, determining the full information allocation within the group, are not necessarily the weights $l = (l_1, l_2)$ used in the principal agent group problem.

[10] Again, continue to ignore momentarily the potential differences between $l$ and within-group weights $m$.

[11] At $l_1 = 0.0$, the optimal $m$ is equal to zero.

[12] The reason that it was not symmetric is that the optimal computed contract contained a lottery over $m_1 = 0.5$ and $m_1 = 0.53$ and consumption sharing rules for the latter weight are not symmetric. But an optimal symmetric contract also exists. Because the agents are symmetric in utilities, technologies, and Pareto weights, an allocation with $m_1 = 0.47$ and $m_1 = 0.5$ is also optimal. An ex ante symmetric allocation is then easily constructed by assigning equal probability to the first and second allocations.

[13] Remember, for the examples we are assuming that each agent can only work his own technology. Therefore, there is no effort sharing rule like in (2.4.2).

[14] Also note that each agent's unconditional probability distribution is virtually the same as in the earlier example. The slight difference was added to ensure that the likelihood ratio in both regimes' incentive constraints (2.2.3), (2.2.4), and (2.3.7) are well defined.

[15] Note that the frontier is not the convex hull of each of the slices (the panels) in Figure 2.4.3. Instead, it the convex hull in the three-dimensional space consisting of agent one's utility, agent two's utility, and the principal's utility.

[16] We have computed the previous example with explicit lotteries but do not show the results since the Pareto frontier is nearly identical to the convex hulls of each slice shown in the panel.

[17] We assume full observability of efforts only for simplicity. Weaker assumptions like observations of a correlated signal can be incorporated, which would allow incentive problems internal to organizations to be studied.

[18] We do not take up issues of unique implementation here.

[19] There is some anecdotal support for this assumption. In interviews that the first author conducted of cropping groups (farming partnerships) in rural India, respondents claimed that one limit on the size of the partnerships was an increasing difficulty in reaching decisions about what to do. A formal questionnaire administered by the second author to joint-liability groups of the BAAC, Thailand uncovered this management difficulty as well.