The Center for Mathematical Studies in Economics and Management Science
Northwestern University, Evanston, Illinois 60208

Discussion Paper No. 767

SUSTAINABLE MATCHING PLANS WITH ADVERSE SELECTION

by

Roger B. Myerson

March 1988

Abstract. Economies with adverse selection are studied as matching problems. Feasible stationary matching plans are defined by market-clearing equations and informational incentive constraints. When the matching market is contestable, alternative matching systems are characterized by viable prospectus mappings. A feasible plan is sustainable if it can, with a suitably constructed waiting list, inhibit defections to all alternative matching systems. Competitively sustainable plans are shown to always exist, although they may be Pareto-dominated by other sustainable plans. Representatively sustainable plans are also defined, and are shown to always exist and be Pareto-efficient.

Acknowledgements. The author is indebted to Douglas Gale and Larry Jones for helpful comments and suggestions. Support for this research by N.S.F. grant SES-8605619 is gratefully acknowledged.

(This paper is a substantially revised version of DP #721, "Dynamic Matching Problems with Incentive Constraints.")

SUSTAINABLE MATCHING PLANS WITH ADVERSE SELECTION

1. Introduction.

Rothschild and Stiglitz [1976] and Spence [1973] showed that fundamental difficulties may arise when we try to extend traditional notions of market equilibrium to economies in which individuals have private information that is relevant but unobservable to their trading partners. In such economies, adverse selection or informational incentive constraints may hinder individuals' efforts to identify mutually beneficial opportunities to trade with each other. Rothschild and Stiglitz showed simple examples of such economies in which no market equilibrium seemed to exist. The difficulty seems to be that, in an economy with informational incentive constraints, the stabilizing factors that sustain equilibria in free markets may be different from those that had been studied before the seminal work of Rothschild and Stiglitz and Spence. The goal of this paper is to systematically explore some factors that may play a role in sustaining equilibria of economic systems with informational incentive constraints.

Since general existence theorems are known for Nash equilibria of noncooperative games, some work in response to the Rothschild-Stiglitz paradox has naturally focused on modelling the market as a noncooperative game. Dasgupta and Maskin [1986] have shown that Nash equilibria in randomized strategies exist for a simple game that naturally corresponds to the rules of the market that Rothschild and Stiglitz studied. However, Hellwig [1986a] has shown that there are many different game models that can be just as naturally studied, as models of markets with adverse selection, and that the equilibria (or stable equilibria) of these various game models may be very different. The details of who makes offers and when, after what observations, can significantly alter the solution to the game. Such sensitivity is not be a bad thing when the details in question correspond to measurable parameters of the real situations that we want to study. But in the real world where time is continuous and potentially infinite, agents often have an enormous richness of bargaining and marketing options that may go beyond the scope of any tractable game model. Such considerations may motivate the approach of this paper, in which we to try to work with models of markets that abstract away from the noncooperative-game structure of bargaining between agents.

Wilson [1977], Riley [1979], and Miyazaki [1977] have studied equilibrium concepts for markets with adverse selection that rely on general market responses to any deviation from the given equilibrium. Certainly, the ability of established firms to respond to entrants or deviants (and to directly or indirectly punish them for deviating) can be an important stabilizing factor in a market. However, there are many situations in which small-scale entry or deviation may be undetectable. In such situations, the temptation to make unpunished entries or deviations from an established market system may erode and ultimately destroy it, if it requires such responsiveness to be supported as an equilibrium. Thus, in this paper we try to develop solution concepts that are based on the opposite assumption: that alternatives to an established or incumbent market system may be costlessly offered at any time, and that the established market system is truly sustainable only if it can inhibit such alternatives without threatening to actively change in response to them. This assumption has been called <u>contestability</u> by Baumol, Panzar, and Willig [1986]. A contestable free-market economy, even if centralized, is different from a Soviet-style economy in which alternative market systems are forcibly prohibited or restricted. The use of the term <u>sustainability</u> in this paper may be viewed as an extension of Baumol, Panzar, and Willig's use of the same term.

The essential problem in any economy is to match suppliers of goods and services with their potential users and consumers. A market can be viewed as a system for matching individuals and arranging trades among them. Thus, the basic conceptual approach of this paper is to view an economy as a dynamic matching problem. We consider an abstract model in which individuals of various types arrive or are _born_ into the economy at some given rates. Each individual's private information is his own _type_, which he knows from birth. After arriving in the economy, each individual waits to be matched for some period of time, and finally exits from the economy in some exit configuration.

An exit configuration may be interpreted as a description of a set of individuals who are trading with each other and of all the net trades between these individuals. We assume that an exit configuration includes a description of all the trades that are made by all the individuals belonging to it. In this sense, the term "exit" is indeed appropriate, because an individual has no further economic activities to be determined once he is assigned to an exit configuration.

Part of the difficulty of thinking about economies with adverse selection is that the concept of a commodity becomes more complicated and less analytically fruitful when individuals care about with whom they trade, as well as what they trade. When an insurance company sells some number of identical insurance policies, we cannot determine the impact of these sales on its expected profit without specifying how many policies were bought by low-risk individuals and how many were bought by high-risk individuals. That is, we cannot simply add up quantities bought or sold, for each commodity, and assume that the total net trade is all that matters. For this reason, we find it simpler here to use a model that abstracts away from the structure of net trades, subsuming them into the definition of an exit configuration, and instead emphasizes the way that individuals of various types are matched for trade with each other.

In this paper, we consider only matching problems in which the birth (or arrival) rate of individuals of each type is constant over an infinite time horizon. We assume that individuals have no cost of waiting to be matched, and use a zero discount rate. Throughout this paper, we consider only matching plans that are <u>stationary</u>, in the sense that an individual who is born at any point in time would expect the same treatment as an individual with the same type who is born at any other point in time.

A <u>feasible</u> matching plan is is defined, in Section 3, to be any plan for matching or allocating all individuals into exit configurations that does not require any individual to reveal information that is not in his own best interest. That is, a feasible matching plan must satisfy the informational incentive constraints that arise because each individual has an option to lie about his type. A <u>sustainable</u> matching plan is defined, in Section 6, as a feasible matching plan that can inhibit defections to alternative matching systems, when the assumption of contestable markets is added.

There is a wide literature on bargaining in economic models of search and matching; see, for example, Butters [1984], Diamond [1982], Mortensen [1982], Rubinstein and Wollinsky [1985], and Gale [1986]. These papers generally assume that, following some exogenously given search process, individuals meet each other in pairs and then each pair plays a given bargaining game that results either in an agreement to trade or an impasse, in which case they separate and search again for new trading partners. In contrast, we do not here assume any specific rules for search, meeting, and bargaining. Instead, we assume only that these rules should lead to outcomes that are sustainable. Our basic concept of sustainability expresses the idea that we should not expect to observe rules of search and trading from which individual could be easily persuaded to switch to an alternative market system. Thus, in a sustainable market system, whenever individuals meet at random, they should have essentially no latitude for bargaining, in the sense that no bargaining game could offer them all expected utility payoffs that are higher than what they expect to get in the market.

To be able to focus clearly on the role of informational incentive constraints in matching problems, we ignore here all other kinds of incentive constraints. That is, we assume that the only problematic incentive constraints are the informational incentive constraints involved in getting individuals to reveal their private information. To illustrate the kind of moral hazard or strategic incentive constraints that we are ignoring here, consider a matching plan that randomly assigns individuals to exit configurations, some of which offer higher payoffs than others. Then an individual who is assigned to one exit configuration might have an incentive to refuse this assignment, if he can, and reenter the matching system in hopes of getting a better assignment on the second pass. Also, after an individual has accepted an assignment to an exit configuration that involves him in some net trades, he might still have an incentive to reenter the market to try to make additional trades, if he can (as in the model of Gale [1986], and also in some of the models studied by Jaynes [1978]). In this paper, we rule out both of these kinds of manipulation by assuming that, after an individual has voluntarily entered a matching system, he can be committed to accept the exit configuration into which it assigns him as mandatory and final. This is clearly a limiting assumption, which future research should seek to relax.

The key to understanding the concept of sustainability that is developed here is contained in the phrase <u>adverse selection</u>. We may use this phrase in a broad sense to refer to the problem that arises when individuals care about the types of individual with whom they trade, but cannot trust other individuals to honestly reveal their types without giving them some incentive to do so. However, in its original application in the insurance industry, the phrase has been used more specifically to refer to the fact that anyone who invites others to trade with him in some new market (e.g., the market for a new kind of insurance policy) must take into account the possibility that most of the individuals who accept his invitation might be bad types with whom such trades are less profitable or even unprofitable. Adverse selection in this narrow sense is then a factor that tends to stabilize and sustain established market systems, because it makes it harder to create alternative market systems that can guarantee their viability and profitability. Thus, we shall see that the set of sustainable matching plans for economies with adverse selection is always nonempty and may be quite large.

In the context of noncooperative games with incomplete information, it is well known that adverse expectations may sustain a very large set of sequential equilibria (see, for example, Fudenberg and Maskin [1986]). In effect, people may be deterred from deviating from an equilibrium by the expectation that other people would believe that anybody who observably deviates from the equilibrium must have the unobservable characteristics of the worst possible type of individual (the type with whom nobody would want to do business). Hellwig [1986b] and Gale [1987] have emphasized the role that such inhibitive adverse expectations can play in determining the sequential equilibria of specific noncooperative-game models of markets with adverse selection. In this paper, we extend this approach by considering a broader general class of deviations or alternative matching systems, and we prove the general existence of sustainable plans which can inhibit defections to all alternatives in this class by expectations of such adverse selection by alternative markets.

The way that people would behave and the matches that they would get in an alternative matching system generally depend on the characteristics of the input population that enters the alternative. The <u>characteristics of the input population</u> that are relevant here are the relative numbers of each type of individual in the population that enters the alternative, and the payoffs that individuals of each type would expect if they returned to the established matching system. (We assume that individuals who enter an alternative matching system can costlessly return to the established matching system, if they are not matched in the alternative.) In Section 4, we define the <u>prospectus</u> of an alternative matching system to be a mapping that specifies the possible expected gains from entering the alternative and the probability of exiting through the alternative, for each type of individual, as a function of the characteristics of the alternative's input population. <u>Viable</u> prospectus mappings are defined in Section 4, taking into account the fact that an alternative system must implement an incentive-compatible matching plan, for any given inputs.

As defined in Section 5, a vector of input-characteristics is <u>inhibitive</u> if, given such inputs, no alternative matching system with a viable prospectus can guarantee that it will not simply return to the established system all individuals who enter it, so that no one would expect to gain by entering the alternative. We apply this concept of inhibitiveness both to alternatives that recruit self-selected inputs and to short-term alternatives that get inputs selected at random from the available population.

Any alternative matching system that operates on the fringe of an established matching system must, at any point in time, recruit its inputs from the population that is currently available and waiting to be matched in the established system. There are two ways that an established system can influence the characteristics of the waiting population: by waiting-time differentials across types, and by discrimination between individuals of the same type. In Section 6, we define a sustainable matching plan to be a feasible plan that could, in these ways, allow the creation of a waiting population that is inhibitive to all alternatives, in the sense of Section 5.

In a dynamic economy, the relative numbers of different types of individuals who are available and waiting to trade at any point in time are usually considered minor variables, often left undetermined in analysis of microeconomic models. For example, when we say that there is equilibrium in the market for suburban homes, we normally mean that the rates of arrival of buyers and sellers into this market are equal; we do not mean that the number of sellers who offering their homes for sale is equal to the number of buyers who are looking for homes at any point in time. (See Rubinstein [1986] for a recent reexamination of this point.) In markets with adverse selection, however, these numbers may be important endogenous variables and determinants of sustainability, as was emphasized in the analysis of Butters [1984]. If different types have different expected waiting times before they are finally matched for trading, the relative numbers of each type in the population that is available for trade at any point in time may differ substantially from the relative numbers of each type that exogenously enter the market over any interval of time. For example, if there are nine high-productivity workers born for every low-productivity worker in every generation, but low-productivity workers search nine times longer than good workers to find a job, then there would be equal numbers of high-productivity and low-productivity workers offering their labor at any point in time in a stationary steady state. Thus, intertype waiting-time differentials allow an established market system to create an adversely selected population of individuals who are waiting at any point in time, and this adverse selection may in turn inhibit entry into any alternative market system.

By "discrimination among individuals of the same type," we mean a policy of the following sort: among the set of individuals who all share some given type characteristics (say, the low-productivity workers), most may be matched for trade quickly and given favorable terms of trade, but a few may be offered less favorable terms of trade and may be asked to wait a long period of time before accepting these terms. Such discrimination clearly requires a centrally registered waiting list, to prevent discriminated individuals from reapplying for the more favorable terms of trade. But with such discrimination, an established market system can create a queue of waiting individuals in which the individuals of some types (the low-productivity workers) would be much more eager to enter an alternative market than the individuals of the same type who were matched for trade more quickly (and who form a majority within any generation). Thus, discrimination may make it harder for an alternatives to select only individuals of some types (high-productivity workers) without selecting any individuals of other types (low-productivity workers). That is, such discrimination may prevent alternative market systems from making a favorable (opposite of adverse) selection from the available population, at the expense of the established market system.

We say that a matching plan is <u>competitively sustainable</u> if it can be sustained using only intertype waiting-time differentials, without any intratype discrimination (which requires centralization, as we remarked above). We say that a plan is <u>representatively sustainable</u> if it can be sustained only using intratype discrimination, without any difference between the expected waiting times (on arrival into the market) of individuals of different types, so that the distribution of types in the population at any point in time is the same of the distribution of types in the generation that enters the market over any interval of time. The main existence theorems of this paper are that both competitively sustainable matching plans and representatively sustainable matching plans always exist.

The main welfare results of this paper are that competitively sustainable plans may be Pareto-inefficient, but representatively sustainable plans are always Pareto-efficient within the class of feasible matching plans (subject to incentive constraints). That is, in economies with adverse selection, Pareto-efficiency is compatible with free contestability of markets, but Pareto-efficiency may not be compatible with decentralization.

<u>2. An example.</u>

Let us now consider a specific example, to illustrate the kind of difficulties that Rothschild and Stiglitz found in defining equilibrium for markets with informational incentive constraints. There are three types of individuals in this economy: high-productivity workers, low-productivity workers, and employers. In each generation, there are equal numbers of low-productivity workers and employers, and there are nine times as many high-productivity workers as employers. Each employer can hire up to fifty workers for up to 40 hours each. An employer gets $30 profit per hour from each high-productivity worker and $20 profit per hour from each low-productivity worker that he hires. Every worker has 40 hours of labor to sell, which he can sell to at most one employer. Each high-productivity worker has a personal reservation price of $25 per hour for his labor, and each low-productivity worker has a reservation price of $5 per hour for his labor. Each worker knows his own type, but employers cannot distinguish between the two types of workers when they are hired (except by offering a choice of employment terms such that the two types would make different decisions). Everyone is risk neutral, and money is freely transferable.

There are many possible matching plans that could be implemented in this economy, but two plans stand out as the most promising candidates for being equilibria. In both of these plans, the employers (who are effectively in excess supply in this economy) get zero expected profits. In the first of these plans, which we may call the <u>standard pooling plan</u>, all workers are hired for 40 hours at a wage of $29 per hour, which is the expected productivity of a randomly sampled worker in each generation. In the second of these plans, which we may call the <u>standard separating plan</u>, each low-productivity worker is hired for 40 hours at a wage of $20 per hour, but each high-productivity worker is hired for 24 hours at a wage of $30 per hour and must have 16 hours of unemployment. It is straightforward to check that, because the high-productivity workers value their time more highly than the low-productivity workers, neither type of worker could gain by imitating the other type in this separating plan.

The standard separating plan is obviously very wasteful. In the Rothschild-Stiglitz theory, it is not an equilibrium because an employer with excess capacity could offer to hire additional workers full-time wage of $28.50 per hour, which would attract all of the workers in the market and (because 90% of all workers have high productivity) would give the employer an expected net profit on each worker hired.

On the other hand, the standard pooling plan also fails the test for a Rothschild-Stiglitz equilibrium, because an employer could offer a wage of $29.50 per hour for 38 hours, requiring that the remaining 2 hours be unemployed. Relative to the full-time wage of $29.00 per hour, this offer would be better for the high-productivity workers, but it would be worse for the low-productivity workers. Thus, when all other employers are implementing the standard pooling plan, the employer who makes this offer of $29.50 per hour for 38 hours would expect to attract only the high-productivity workers and make a positive profit.

Thus, it seems that the standard pooling plan and the standard separating plan each create opportunities for employers to gain by deviating from the supposed plan. In fact, no matching plan can satisfy the criteria for a Rothschild-Stiglitz equilibrium in this example. However, the general existence theorems of this paper imply that there must be sustainable plans, as defined in Section 6, for this economy when it is viewed as a dynamic matching problem. In fact, the standard pooling plan is the unique representatively sustainable plan, the standard separating plan is the unique competitively sustainable plan, and there are many other sustainable plans.

3. Dynamic matching problems and feasible matching plans.

Consider a large stationary economy into which a new generation of individuals is born (or arrives) every day. Each individual is born into the economy with a fixed type which he knows as his private information. Let N denote the nonempty finite set of different types for the individuals in the economy. For the example in Section 2, we may let $N = \{1,2,3\}$, where type 1 denotes the high-productivity workers, type 2 denotes the low-productivity workers, and type 3 denotes the employers.

Although each individual in the economy knows his own type, he cannot necessarily identify the types of others. Thus, an employer may be uncertain as to which of his potential workers are high-productivity workers and which are low-productivity workers. We let J be a subset of $N \times N$ which represents the type pairs that may be problematic to verify. That is, $(j,i) \in J$ iff $i \neq j$ and a type-i individual could imitate a type-j individual if he were given any incentive to do so. For the above example, J might equal $\{(1,2),\ (2,1)\}$, if the different types of workers can imitate each other but cannot pretend to be employers (who own factories).

We assume that there is no aggregate uncertainty in the economy, so that everyone knows the relative number of each type in every generation. For each i in N, let ρ_i denote the rate at which type-i individuals are born in this economy, per unit of time. We assume that, for each i, ρ_i is strictly positive and constant over time. With continuous time, we can always define our unit of time so that the aggregate birth rate is one; that is

$$\Sigma_{i \in N}\, \rho_i = 1. \tag{3.1}$$

For our example, we could let $\rho_1 = 9/11$, $\rho_2 = 1/11$, $\rho_3 = 1/11$. When we normalize birth rates to sum to one, then ρ_i can also be interpreted as the proportion of type-i individuals in the generation that is born during any interval of time.

In this simple model, an individual may search or wait over a period of time, and then he exits from the economy as a part of some coalition. A coalition consists of a specified number of individuals of each type. When individuals exit together in a coalition, they may also may choose to make some net trades themselves, and they may perform other nontrade activities. An exit configuration is a pair consisting of a coalition and a feasible vector of net trades and other activities for the members of the coalition. For example, a coalition might consist of nine high-productivity workers, one low-productivity worker, and one employer. An exit configuration might consist of this coalition together with a specification that "each high-productivity worker sells 24 hours of labor to the employer for a total payment of $720, and each low-productivity worker sells 40 hours of labor to the employer for a total of $800."

We let E denote the set of all possible exit configurations. In developing the technical definitions and results of this paper, we shall assume that E is a finite set.

For any e in E, and any i in N, we let $r_i(e)$ denote the number of type-i individuals belonging to the coalition in the exit configuration e. Clearly, we must require

$$r_i(e) \geq 0, \quad \forall i \in N, \quad \forall e \in E.$$

For the exit configuration described above, we would have $r_1(e) = 9$, $r_2(e) = 1$, and $r_3(e) = 1$. For every e in E, there must exist at least one type i in N such that $r_i(e) > 0$.

An individual's payoff in this economy is completely determined by the configuration in which he exits, that is, by the coalition that he joins and by the trades and activities that are implemented by the members of the coalition. We do allow that an individual's payoff may depend on the types and activities of all the members of his coalition, but we assume that there are no externalities imposed on him by individuals outside of his coalition. We assume that there are no costs of waiting or searching before an individual joins a coalition. For any e in E and i in N, we let $u_i(e)$ denote the expected payoff, measured in some von Neumann-Morgenstern utility scale, that a type-i individual would get from exiting in the configuration e, if all the members of the coalition in e are honest about their types. For any e in E and any (j,i) in J, we let $\hat{u}_i(e,j)$ denote the expected payoff to a type-i individual if he pretended to be a type-j individual, while everyone else was being honest about their types, and exited as a part of an ostensible configuration e (which actually contained $r_j(e) - 1$ type-j individuals and $r_i(e) + 1$ type-i individuals, because of his misrepresentation).

For each type i, we assume that there exists a <u>solitary exit configuration</u> $\bar{e}_i$ such that $r_i(\bar{e}_i) = 1$, $u_i(\bar{e}_i) = 0$, and, for every $j \neq i$, $r_j(\bar{e}_i) = 0$, and $u_j(\bar{e}_i, i) = 0$. Here $\bar{e}_i$ represents the exit configuration in which a type-i individual must exit alone, without trading with anyone else. In effect, we are normalizing our utility scales so that an individual who exits alone (with no trading partners) gets a payoff of zero.

To fully represent the example in Section 2 as such a model, it may seem that we would need an infinite set of exit configurations. This is because, for any nonnegative integers m_1 and m_2 (not both zero) and any real numbers τ_1, τ_2, γ_1, and γ_2 such that

$$m_1 + m_2 \leq 50, \quad 0 \leq \tau_1 \leq 40, \quad 0 \leq \tau_2 \leq 40, \quad 0 \leq \gamma_1 \leq 1200, \quad 0 \leq \gamma_2 \leq 1200,$$

there should be an exit configuration e representing a coalition containing one employer, m_1 high-productivity workers, and m_2 low-productivity workers, where each high-productivity is working τ_1 hours for a total wage bill of γ_1 dollars, and each low-productivity worker is working τ_2 hours for a total wage bill of γ_2 dollars. For such an exit configuration e, we would have

$$r_1(e) = m_1, \quad r_2(e) = m_2, \quad r_3(e) = 1,$$
$$u_1(e) = \gamma_1 - 25\tau_1, \quad u_2(e) = \gamma_2 - 5\tau_2,$$
$$u_3(e) = m_1(30\tau_1 - \gamma_1) + m_2(20\tau_2 - \gamma_2),$$
$$\hat{u}_1(e,2) = \gamma_2 - 25\tau_2, \quad \hat{u}_2(e,1) = \gamma_1 - 5\tau_1.$$

In addition, we must include the solitary exit configurations $\bar{e}_1$, $\bar{e}_2$, and $\bar{e}_3$. Although this requires an infinite set of exit configurations, it is a compact, so that this problem can be approximated arbitrarily closely by models with finite sets of exit configurations. In fact, it would suffice to include exit configurations for the extreme points of the convex set of $(m_1,m_2,\tau_1,\tau_2,\gamma_1,\gamma_2)$ vectors described above, and the three solitary exit configurations.

The structures $(N, J, E, (\rho_i, r_i, u_i, \hat{u}_i, \bar{e}_i)_{i\in N})$ completely specify the general model of the <u>dynamic matching problem</u> to be studied in this paper.

Given a dynamic matching problem, we define a <u>matching plan</u> to be any function μ that assigns a nonnegative number $\mu(e)$ to every exit configuration e, where $\mu(e)$ represents the rate at which instances of the exit configuration e are to occur in the plan μ. In this paper, we consider only stationary matching plans, in which these rates are constant over time. With birth rates normalized to sum to 1, $\mu(e)$ is can be interpreted as the number of e exit configurations that occur per birth, during any interval of time. For example, in a marriage matching system, if e denotes a marriage and every individual exits in a marriage with one other individual, then $\mu(e)$ must equal

1/2, since there would be one marriage for every two births. In general, the set of all matching plans may be denoted by $\mathbb{R}_+^E$.

For any type i and any matching plan μ, we define the following functions:

$$R_i(\mu) = \Sigma_{e\in E}\ r_i(e)\ \mu(e)$$

$$V_i(\mu) = \Sigma_{e\in E}\ u_i(e)\ r_i(e)\ \mu(e)$$

and, for any type j such that $(j,i,) \in J$, we let

$$\hat{V}_i(\mu,j) = \Sigma_{e\in E}\ \hat{u}_i(e,j)\ r_i(e)\ \mu(e).$$

Notice that these three functions are all linear in μ. For any μ in M and any i in N such that $R_i(\mu) > 0$, we define

$$U_i(\mu) = V_i(\mu)/R_i(\mu).$$

Similarly, for any μ in M and any (j,i) in J such that $R_j(\mu) > 0$, we define

$$\hat{U}_i(\mu,j) = \hat{V}_i(\mu,j)/R_j(\mu).$$

To interpret these functions, notice first that $R_i(\mu)$ is the rate at which type-i individuals are being matched, per unit time, in the matching plan μ. The expected payoff to an individual of type i is $U_i(\mu)$ if everyone is honest about their types as they participate in the matching plan μ. On the other hand, if a type-i individual pretended that his type was j then his expected payoff would be $\hat{U}_i(\mu,j)$, if everyone else participated honestly in the plan μ. Since $U_i(\mu)$ and $\hat{U}_i(\mu,j)$ are nonlinear in μ, it will often be more convenient to work with the functions $V_i(\mu) = U_i(\mu)\ R_i(\mu)$ and $\hat{V}_i(\mu,j) = \hat{U}_i(\mu,j)\ R_j(\mu)$, which are linear.

We say that a matching plan μ is <u>feasible</u> iff

(3.2) $$R_i(\mu) = \rho_i, \quad \forall i \in N,$$

and

(3.3) $$U_i(\mu) \geq \hat{U}_i(\mu,j), \quad \forall (j,i) \in J.$$

Notice that, when (3.2) is satisfied, (3.3) is equivalent to

$$(3.3') \qquad V_i(\mu)/\rho_i \geq \hat{V}_i(\mu,j)/\rho_j, \quad \forall(j,i) \in J,$$

so the set of feasible matching plans is defined by a finite collection of linear inequalities in μ.

Condition (3.2) asserts that μ should clear the market, creating exit opportunities for type-i individuals at the same rate that new type-i individuals arrive in the economy. Condition (3.3) lists the <u>informational incentive constraints</u> for this economy, which assert that no individual of any type i should expect to gain in the plan μ by pretending that his type is some other j that he can imitate. Thus, if μ satisfies (3.2) and (3.3), then μ could be implemented by a centralized matching system which asked every individual to report his type and then assigned each individual to a randomly determined exit configuration, so that his probability of exiting in configuration e would be $\mu(e)r_i(e)/R_i(\mu)$ if he reported that his type was i. Condition (3.3) implies that it would be a Nash equilibrium for all individuals to report their types honestly in such a matching system, and condition (3.2) asserts that this matching system would actually match everyone. Conversely, under weak assumptions about the structure of E, one can guarantee that any matching plan that could be implemented by any market system must satisfy the constraints (3.3), by standard revelation-principle arguments.

To avoid questions of moral hazard, we assume that, after an individual enters a matching system, he can be asked to make a commitment to accept, as mandatory and final, the exit configuration into which the system assigns him. If there are no alternatives to the given established matching system, then each type-i individual can be made to choose between making this commitment and exiting alone in $\bar{e}_i$, where he gets $u_i(\bar{e}_i) = 0$. Thus, when there are no alternative matching systems, all individuals would be willing to make the commitment to a matching plan μ if it satisfies the following <u>individual-rationality</u> or <u>participation</u> constraints:

(3.4) $$U_i(\mu) \geq 0, \quad \forall i \in N.$$

Since individuals may be assigned to solitary exit configurations under the matching plan μ as well, the generalized revelation principle also implies that, without loss of generality, we can assume that everyone does make the commitment to be matched according to μ when it satisfies (3.4). (This justifies the equality in (3.2)).

We now consider the more difficult problem of characterizing the subset of the feasible matching plans that are sustainable when markets are freely contestable. That is, we assume from now on that there is nothing to prevent an entrepreneur from trying to organize some alternative matching system. Under this assumption, an established matching system may need to offer individuals much more than the simple nonnegative expected payoffs stipulated in (3.4), to prevent them from defecting to an alternative matching system. Of course, it would be easy to inhibit defections to alternative matching systems if each individual believed that no one else would ever enter an alternative matching system (and such beliefs could even be supported in a noncooperative Nash equilibrium). So we will rule out such trivial equilibria by assuming that positive numbers of individuals will always at least enter any alternative matching system.

In Sections 4 and 5, we develop a model to describe alternative matching systems and how defections to them can be inhibited. Then, in Section 6, we combine this model with the concept of a feasible matching plans for an established system, to develop the general definition of sustainability.

4. Viable alternative matching plans and prospectus mappings.

An alternative matching plan has both an advantage and a disadvantage relative to the established matching plan. The advantage of an alternative is that it need not match everyone who enters it, because it can direct some individuals to return to the established matching plan. By (3.2), the established matching plan is supposed to be able to match everyone into an exit configuration (which may, of course, be a solitary configuration $\bar{e}_i$). The disadvantage of an alternative is that it cannot guarantee that everyone who was born during any given time interval will enter the alternative; the alternative may get some adverse selection out of a given generation. Thus, the matching plan that is implemented by an alternative matching system may have to depend on the characteristics of the population that enters it.

To describe the input population that an alternative matching plan has to work with, we need two vectors, an allocation vector in $\mathbb{R}^N$ and a distribution vector in $\Delta(N)$, where

$$\Delta(N) = \{q \in \mathbb{R}^N | \ \Sigma_{i\in N}\ q_i = 1 \ \text{ and } \ q_j \geq 0 \ \ \forall j \in N\}.$$

If we say that a pair (w,q) in $\mathbb{R}^N \times \Delta(N)$ represents the characteristics of an alternative's input population, where $w = (w_i)_{i\in N}$ is the allocation vector and $q = (q_i)_{i\in N}$ is the distribution vector, then we mean that, for each i in N, q_i is the proportion of type-i individuals in the input population, and w_i is the expected payoff that the type-i individuals in this input population would get if they were returned by the alternative to the established matching system. We may refer to any such a pair (w,q) in $\mathbb{R}^N \times \Delta(N)$ as an input-characteristics vector, or as an input vector for short.

There are some technical difficulties that arise when we consider input populations in which some types are not represented. If no type-j individuals are supposed to be entering an alternative, then it is not clear how to define the expected payoff that a type-j individual would get if he did enter the alternative, nor is it clear how to define the expected payoff that a type-i individual would get if he imitated a type-j individual after entering the alternative matching system. We solve this technical problem by analyzing first the case where all type enter in strictly positive proportions, and then extending our analysis to the case where some proportions are zero by upper-semicontinuity. Let $\Delta^{o}(N)$ denote the set of all distributions over N in which every type is has a positive proportion; that is,

$$\Delta^{o}(N) = \{q \in \mathbb{R}^{N} | \ \ \Sigma_{i\in N} q_i = 1 \ \text{ and } \ q_j > 0 \ \ \forall j \in N\}.$$

Suppose that (w,q) is any input vector in $\mathbb{R}^{N} \times \Delta^{o}(N)$, representing the allocation and distribution vectors of an input population. We say that η is a <u>viable alternative matching plan</u> with the input vector (w,q) iff $\eta \in \mathbb{R}^{E}_{+}$,

(4.1) $\qquad R_i(\eta)/q_i \le 1, \quad \forall i \in N,$

(4.2) $\qquad (V_i(\eta) - R_i(\eta)\ w_i)/q_i \ge 0, \quad \forall i \in N,$

and

(4.3) $\qquad (V_i(\eta) - R_i(\eta)\ w_i)/q_i \ge (\hat{V}_i(\eta,j) - R_j(\eta)\ w_i)/q_j, \quad \forall (j,i) \in J.$

Given any such alternative matching plan $\eta = (\eta(e))_{e\in E}$, each number $\eta(e)$ may be interpreted as ratio of the number of e-configured coalitions that come out of the alternative to the number of individuals who enter the alternative. Equivalently, when we think of the alternative as operating over some short interval of time, $\eta(e)$ may be interpreted the rate at which e-configured coalitions would exit the alternative system, in a time scale chosen so that the rate at which individuals enter into the alternative matching system is one. Notice that the division by q_i and q_j in these formulas (essential only in (4.3)) requires us to assume that the distribution vector q is in $\Delta^{o}(N)$.

Condition (4.1) represents the requirement that the rate $R_i(\eta)$ at which type-i individuals are matched by the alternative matching plan cannot exceed the rate q_i at which they are entering the alternative. To interpret condition (4.2), notice that

$$w_i + (V_i(\eta) - R_i(\eta)\ w_i)/q_i$$
$$= \Sigma_{e\in E}\ u_i(e)\ r_i(e)\ \eta(e)/q_i + (1 - \Sigma_{e\in E}\ r_i(e)\ \eta(e)/q_i)\ w_i.$$

Thus, $(V_i(\eta) - R_i(\eta)\ w_i)/q_i$ is the expected gain, over his payoff w_i in the established matching plan, that a type-i individual would get when he enters the alternative matching plan, taking into account the probability $(1 - \Sigma_{e\in E}\ r_i(e)\ \eta(e)/q_i)$ that he will be returned to the established matching system. So condition (4.2) asserts that, when any type-i individual enters the alternative, he should not expect to lose, relative to the payoff w_i that he could get in the established matching system, because he has the option of staying in the established matching system. Condition (4.3) asserts the expected gain over w_i for each type-i individual, when he enters the alternative, should be not less than the expected gain over w_i that he could get by pretending (in the alternative matching system) to be a different type j that he can imitate.

We now must distinguish between an alternative matching <u>system</u> and an alternative matching <u>plan</u>, like η above. The alternative matching system is an institution or game whose outcome may depend on the characteristics of the individuals who enter it. Thus, the plan η that an alternative system implements may depend on the input characteristics (w,q). That is, an alternative matching system may be thought of as a mapping from $\mathbb{R}^N \times \Delta(N)$ into $\mathbb{R}^E_+$, that always selects viable plans (at least in the case where we have defined them, when q is in $\Delta^o(N)$). However, for the purposes of individuals' decisions about whether to enter an alternative matching system or stay with the established system, all that matters about an alternative matching plan is the expected gain and the probability of being matched (that is, the probability of not being returned to the established system) that it offers, for all types of individuals. Thus, we may imagine that an alternative matching system might offer some prospectus that describes how these expected gains and probabilities might depend on the characteristics of the input population that the alternative draws.

So we define a <u>prospectus</u> mapping to be any upper-semicontinuous correspondence from $\mathbb{R}^N \times \Delta(N)$ to nonempty subsets of $\mathbb{R}^N \times [0,1]^N$. ($[0,1]^N$ is the set of vectors, indexed on N, whose components are all between 0 and 1.) The requirement that prospectus mappings should be upper-semicontinuous is a natural technical condition. We say that a prospectus mapping $F:\mathbb{R}^N \times \Delta(N) \twoheadrightarrow \mathbb{R}^N \times [0,1]^N$ is <u>viable</u> iff, for every (w,q) in $\mathbb{R}^N \times \Delta^o(N)$ and every (y,s) in $F(w,q)$, there exists some η that is a viable alternative matching plan with the input vector (w,q), such that

$$y_i = (V_i(\eta) - R_i(\eta)\ w_i)/q_i, \quad \text{and} \quad s_i = R_i(\eta)/q_i, \quad \forall i \in N.$$

(Here $y = (y_i)_{i\in N}$ and $s = (s_i)_{i\in N}$.) Thus, for any prospectus mapping F, if $(y,s) \in F(w,q)$ then, for every type i, y_i could be the expected gain over w_i that a type-i individual would get from entering the alternative, and s_i could be the probability that a type-i individual would be matched (not returned) by the alternative, if (w,q) were the allocation and distribution vectors of its input population. For F to be viable, these expected gains and probabilities of being matched must be actually achievable by some viable alternative plan, at least with any input-characteristics vector for which we know how to define viable plans.

For the example in Section 2, a particularly simple viable prospectus, of some theoretical interest, can be constructed as follows. Consider an alternative matching system in which each worker and employer who enters is asked whether he would accept terms of employment in which every worker works 38 hours at a wage of \$29.50 per hour. Suppose that any employer who accepts these terms and who is assigned a worker is then given priority for getting more workers until his capacity of 50 workers is reached. Workers and employers who reject these terms, or who cannot be matched because of excess supply on their side of the market, are returned to the established matching plan. Notice that the net payoff for a high-productivity worker under these terms would be $38 \times (29.50 - 25) = 171$, and the net payoff for a low-productivity worker would be $38 \times (29.50 - 5) = 931$. Under these terms, an employer would make a net profit of $50 \times 38 \times (30 - 29.50) = 950$ from hiring 50 high-productivity workers, but $50 \times 38 \times (20 - 29.50) = -18050$ from hiring 50 low-productivity workers. When F denotes the prospectus mapping for this alternative, for any (w,q) in $\mathbb{R}^N_+ \times \Delta^o(N)$, $(y,s) \in F(w,q)$ iff at least one of the following conditions is satisfied:

(4.4) $171 \geq w_1, \quad 931 \leq w_2, \quad 950 \geq w_3,$

$s_1 = \min\{1,\ 50q_3/q_1\}, \quad s_2 = 0, \quad s_3 = \min\{1,\ q_1/50q_3\},$

$y_1 = s_1(171 - w_1), \quad y_2 = 0, \quad \text{and} \quad y_3 = s_3(950 - w_3);$

(4.4) $171 \geq w_1, \quad 931 \geq w_2, \quad (950q_1 - 18050q_2)/(q_1 + q_2) \geq w_3,$

$s_1 = s_2 = \min\{1,\ 50q_3/(q_1 + q_2)\}, \quad s_3 = \min(1,\ (q_1 + q_2)/50q_3,$

$y_1 = s_1(171 - w_1), \quad y_2 = s_2(931 - w_2), \quad \text{and}$

$y_3 = s_3((950q_1 - 18050q_2)/(q_1 + q_2) - w_3);$

(4.5) $171 \leq w_1 \quad \text{and} \quad s = q = \underset{\sim}{0};$

(4.6) $\quad 171 \geq w_1, \quad 931 \geq w_2, \quad (950q_1 - 18050q_2)/(q_1 + q_2) \leq w_3,$ and

$s = w \quad = \underset{\sim}{0}.$

5. Inhibitive allocations and input vectors.

To sustain the established matching system, we need to be able to explain why no individuals should exit under any alternative matching systems. To formulate such explanations, we need to distinguish two different assumptions about how individuals might enter into an alternative matching system.

We may say that an alternative matching system gets random inputs if, for any individual who is available to be matched at the point in time when the alternative matching system is offered, the probability of his entering the alternative is independent of his type. Thus, with random inputs, the population that enters an alternative matching system must be an unbiased sample drawn from the overall population that is currently available to be matched in the established system. For example, in search-and-bargaining models like that of Gale [1986], the bargaining games that form at any point in time are assumed to have random inputs from the available population.

On the other hand, we may say that an alternative matching system gets self-selected inputs iff individuals enter it only after making some positive decision to do so, and individuals of some types may be more likely to make this decision than other types. In the model of Gale [1987], for example, all markets are like specialty stores that have self-selected inputs. With self-selected inputs, the distribution of types in the population that enters an alternative system may differ from the distribution of types currently available in the established system. In particular, if an alternative matching system was expected to offer zero net gains over the established plan to all types of individuals, so that all individuals would be indifferent between entering the alternative and not entering it, then, for each type of individual, any probability of choosing to enter the alternative could be rationally justified. That is, any distribution of entering types can be justified with self-selected inputs, if expected gains from entering are zero for all types.

Let us first analyze the problem of inhibiting alternative matching systems that get self-selected inputs. The allocation vector w, which describes the payoffs that available individuals of each type would get in the established matching system, must be determined by the operation of the established system before an alternative prospectus is announced. However, with self-selected inputs, the distribution of types actually entering an alternative is an endogenous variable to be determined after the alternative prospectus is announced. Thus, we define an <u>inhibitive allocation vector</u> to be any w in $\mathbb{R}^N$ such that, for every viable prospectus mapping F, there exists some q in $\Delta(N)$ such that

$$(\underset{\sim}{0},\underset{\sim}{0}) \in F(w,q).$$

(Here $\underset{\sim}{0}$ denotes the zero vector in $\mathbb{R}^N$.) That is, w is an inhibitive allocation vector iff, for any alternative matching system, there is some conjecture about the distribution of types that would choose to enter this alternative such that its prospectus allows the expectation that no individuals of any type will gain from entering into the alternative and that everyone who enters it will ultimately be returned back to the established matching system.

Let us now analyze the problem of inhibiting alternative matching systems that get random inputs. Notice that random inputs are reasonable to assume only if entry into the alternative can be guaranteed to be completely costless to individuals. Intuitively, such an assumption is questionable if an alternative requires that individuals who enter must wait for some extended time before being matched or returned to the established system. Thus, in this paper we only consider random inputs for alternative matching systems that operate with a very short time horizon. Such short-term alternatives must draw their inputs from the overall population of individuals that are available to be matched at any point in time.

If different types have different expected waiting times in the established matching system, then the distribution of types in population of individuals that are available to be matched at any point in time may be any vector q in $\Delta(N)$, not necessarily equal to ρ (which is the distribution of types in the generation born during any interval of time). Even in a marriage market where males and females have equal birth rates and marry only once, the numbers of never-married males and females may be unequal. For example, if all marriages were between forty-year-old males and twenty-year-old females, then there would be twice as many males as females waiting to be matched at any point in time, in the steady state. (To achieve this inequality there would have to be some period of time during which some females never married. But this could be a purely transient phenomenon restricted to one generation. Following a long tradition in economics, we ignore here such transient phenomena and concentrate on the steady state.)

Notice, however, that both the allocation vector w and the distribution q of types in the population of individuals that are available to be matched must be determined by the operation of the established matching system before the prospectus is announced for any alternative matching system with random inputs. Thus, we say that (w,q) is an <u>inhibitive input vector</u> iff $w \in \mathbb{R}^N$, $q \in \Delta(N)$, and, for every viable prospectus mapping F,

$(\underset{\sim}{0},\underset{\sim}{0}) \in F(w,q)$.

So suppose that the established matching system operates so that, in the population of individuals who are available to be matched at any point in time, the allocation of expected payoffs (from the established system) and the distribution of types form an inhibitive input vector. Then any viable prospectus for an alternative matching system with random inputs must allow the expectation that the no individuals of any type will gain by entering the alternative and that everyone who does enter it will return back to the established matching system.

As we have remarked, for alternatives that get random inputs, the entire input characteristics vector (w,q) must be determined by the established system before the alternative announces its prospectus. On the other hand, for alternatives that get self-selected inputs, only w must be determined by the established system before the prospective is announced, because q may be a function of the prospectus. Thus, it might seem that random inputs would induce a stronger definition of inhibitiveness than self-selected inputs. In fact, Theorem 1 asserts that our two definitions of inhibitiveness are essentially equivalent. An allocation vector is inhibitive of alternatives that get self-selected inputs if and only if it is part of some input vector that is inhibitive of alternatives that get random inputs.

<u>Theorem 1.</u> An allocation w in $\mathbb{R}^N$ is inhibitive if and only if there exists some q in $\Delta(N)$ such that (w,q) is an inhibitive input vector.

Section 8 contains the proof of all theorems and lemmas in this paper.

We now introduce some technical definitions and lemmas related to these concepts of inhibitiveness.

As an alternative matching plan, the zero vector in $\mathbb{R}^E$ (defined by $\underset{\sim}{0}(e) = 0$ for every e in E) represents the plan to match no one and to return everyone to the established matching system. Notice that the zero vector in $\mathbb{R}^E$ is always a viable alternative matching plan, satisfying (4.1)-(4.3) for any (w,q) in $\mathbb{R}^N \times \Delta^o(N)$. We say that the input vector (w,q) is <u>strongly inhibitive</u> iff $q \in \Delta^o(N)$ and there is no viable alternative matching plan with the input vector (w,q), other than the zero vector in $\mathbb{R}^E$. It is straightforward to check that the set of strongly inhibitive input vectors is a relatively open subset of $\mathbb{R}^N \times \Delta(N)$. Its closure is the inhibitive set.

<u>Lemma 1.</u> An input vector (w,q) is inhibitive if and only if there exists some sequence of strongly inhibitive input vectors $\{(w^k,q^k)\}_{k=1}^{\infty}$ such that $\lim_{k\to\infty} (w^k,q^k) = (w,q)$. Furthermore, if (w,q) is an inhibitive input vector, then $w_i \geq 0$ for every type i.

It may be of interest to consider what happens if we drop the assumption that individuals of different types can have different expected waiting times in the established matching system. That is, suppose that, in the established matching system, expected waiting times at birth must be the same for all types. Then distributions other than ρ cannot occur in the input vectors for alternative matching systems with random inputs. That is, the population entering an alternative matching system with random inputs must be, by type, representative of the generation that is born in the economy over any interval of time. So we may say that an allocation vector w in $\mathbb{R}^N$ is <u>representatively inhibitive</u> iff there exists some sequence of vectors $\{w^k\}_{k=1}^{\infty}$ in $\mathbb{R}^N$ such that (w^k,ρ) is strongly inhibitive for every k, and $\lim_{k\to\infty} w^k = w$. Obviously, by Lemma 1, if w is representatively inhibitive then (w,ρ) is an inhibitive input vector.

The following lemma may be useful for identifying strongly inhibitive input vectors.

Lemma 2. An input vector (w,q) in $\mathbb{R}^N \times \Delta^o(N)$ is strongly inhibitive if and only if there exist numbers $\lambda(i)$ for all i in N and $\alpha(j|i)$ for all (j,i) in $N \times N$ such that

$\lambda(i) \geq 0, \quad \forall i \in N,$

$\alpha(j|i) \geq 0$, and if $(j,i) \notin J$ then $\alpha(j|i) = 0, \quad \forall i \in N, \quad \forall j \in N,$

and, for every exit configuration e in E,

$$\Sigma_{i\in N}\, r_i(e)\ ((\lambda(i) + \Sigma_{j\in N}\, \alpha(j|i))\ u_i(e) - \Sigma_{j\in N}\, \alpha(i|j)\ \hat{u}_j(e,i))/q_i$$
$$< \Sigma_{i\in N}\, r_i(e)\ ((\lambda(i) + \Sigma_{j\in N}\, \alpha(j|i))\ w_i - \Sigma_{j\in N}\, \alpha(i|j)\ w_j)/q_i.$$

The formulas in Lemma 2 can be interpreted in terms of the virtual utility concept of Myerson [1984a, 1984b]. They assert that a strongly inhibitive allocation is one which corresponds to an allocation of virtual utility that could not be blocked by the individuals in any exit configuration, if payoffs were in transferable virtual utility. With this interpretation of inhibitiveness, the set of sustainable plans, as defined in Section 6, can be interpreted as a generalized inner core (see Shubik [1982]) for games with incomplete information.

In our definition of an input vector, we have implicitly assumed that all of the type-i individuals who might enter an alternative matching system would have the same expected payoff (denoted by w_i) if they returned to the established system. The more general case, in which type-i individuals with a range of expected payoffs from the established system might be expected to enter an alternative matching system, remains as a problem for future research. However, there is a simpler related issue which we can and must consider now.

Suppose that, among the people entering an alternative matching system, there is an individual of type i who would have an expected payoff of x_i from returning to the established matching system. Suppose, however, that the behavior of everyone else in this alternative system is based on the belief that the population entering the alternative system is as described by the input-characteristics vector (w,q), and suppose that all the type-i individuals who have a return payoffs of w_i are choosing to return to the established system with probability 1. If $x_i > w_i$, then the type-i individual whose return payoff is x_i should also return to the established system with probability 1. To see why this is so, let s_i denote the probability that this individual will exit in the alternative (not return to the established system), and let z_i denote his conditionally expected payoff if he exits in the alternative, when he uses his optimal strategy for interacting with other individuals in the alternative matching system. We must have $(z_i - x_i)s_i \geq 0$, because this individual has the option to simply return to the established system and get x_i. On the other hand, whatever this individual chooses to do, the other type-i individuals with return payoff w_i could do it too. Because they are electing to return to the established system with probability 1, which gives them no expected gains from the alternative, we must have $(z_i - w_i)s_i \leq 0$. These two inequalities imply that $(x_i - w_i)s_i \leq 0$. So the probability s_i must equal zero, because $x_i > w_i$. On the other hand, without further information about the structure of the alternative matching system, we could say nothing about s_i if x_i were less than w_i.

Thus, if all type-i individuals with expected allocation w_i in the established system would be deterred from exiting through the alternative system by the belief that its input-characteristics vector would be (w,q), then all type-i individuals with expected allocations that are higher than w_i in the established system would also be deterred from exiting through the alternative system by this belief.

6. Sustainable matching plans.

When a new-born individual of type i first enters an established matching system that is implementing the matching plan μ, his expected payoff from exiting in the established system is $U_i(\mu)$. However, this fact does not necessarily imply that every individual of type i who enters an alternative matching system must have an expected payoff of $U_i(\mu)$ from returning to the established system. The established matching system could discriminate between different individuals of the same type. Suppose that some individuals might be put into a waiting list or queue after they learn how this discrimination is to be applied in their particular cases, and individuals in this waiting list have the option of entering an alternative matching system while they are waiting. Suppose also that, at any point in time, virtually all of the individuals of any given type who are available to be matched are in this waiting list (because the length of time that it takes the established system to match the individuals who are not put into the waiting list is very short or infinitesimal in comparison with the waiting time for individuals who are put in the waiting list), and alternative systems cannot distinguish new-born individuals from individuals on the waiting list. Then, alternative systems will recruit their inputs from the established system's waiting list, in which the type-i individuals might expect a payoff different from $U_i(\mu)$ in the established system.

Now, for each type i, let w_i denote the expected payoff for type-i individuals who are in the waiting list, and let q_i denote the proportion of type-i individuals in the waiting list, at any point in time. Suppose that (w,q) is an inhibitive input vector. Then, any alternative matching system that recruits randomly from the waiting list must allow in its prospectus that every individual from the waiting list who enters the alternative will, with probability one, return to the established system. But, at any point in time, there also must exist at least a small number of new-born individuals, who have not yet learned how the established system may discriminate in their cases and have not yet learned whether they will be put in the waiting list (although they do know their types, of course). Such new type-i individuals must have an expected payoff $U_i(\mu)$ in the established matching system, if it is implementing the matching plan μ. To guarantee that no one will exit through the alternative, we must also be able to show that such new-born individuals would also have no incentive to enter the alternative system and exit through it. By the result cited at the end of Section 5, new-born individuals would also be deterred from exiting through the alternative if $U_i(\mu) \geq w_i$ for every type i. That is, if inputs drawn randomly from the waiting list are inhibitive to alternative matching systems, and if the individuals in the waiting list do not expect higher payoffs than new-born individuals of the same type, then all individuals, new-born and wait-listed, can be rationally deterred from exiting through any alternative matching system.

Thus, we say that a matching plan μ is <u>sustainable</u> iff μ is a feasible matching plan (that is, $\mu \in \mathbb{R}_+^E$ and μ satisfies (3.2) and (3.3)) and there exists some inhibitive allocation vector w such that

(6.1) $$U_i(\mu) \geq w_i, \quad \forall i \in N.$$

So a feasible matching plan is sustainable if there is an inhibitive allocation vector in which individuals of each type get a payoff that is not more than the expected payoff for new-born individuals of the same type under this plan.

In this definition of sustainability, the inequalities in (6.1) are the only connection between the sustainable matching plan and the inhibitive allocation or input vector that sustains it. The weakness of this connection is due to two assumptions that we have made: (1) different types may have different expected waiting times before being matched, so that the distribution vector q in the waiting list is not necessarily equal to ρ; and (2) the established system can discriminate between individuals who are put on the waiting list and those of the same type who are matched immediately, so that the allocations w_i for individuals in the waiting list are not necessarily equal to $U_i(\mu)$. It may be of interest to see how the definition of sustainability is changed when either of these assumptions is dropped.

If assumption (1) above is dropped, so that all individuals have the same expected waiting time at birth, then only representatively inhibitive allocation vectors should be considered, because the distribution vector in the waiting list must be $q = \rho$. So we say that μ is a <u>representatively sustainable</u> plan iff it is a feasible matching plan and there exists some representatively inhibitive allocation w such that (6.1) is satisfied.

Assumption (2) above only makes sense if the established matching system can keep some centralized file on all individuals. If, on the other hand, the established matching system (which we have till now treated as some kind of black box) is a decentralized competitive market, then there is nothing to prevent an individual who has been asked to wait from going to another part of the market and presenting himself as a new-born individual. (A decentralized matching system could prevent this only by putting some kind of indelible mark on an individual when he is asked to wait; but such a mark could also be used by alternative matching systems to distinguish wait-listed individuals from new-born individuals of the same type. This would invalidate the assumption, which we used at the end of Section 5, that all individuals of a given type have the same strategic options when they enter an alternative system.) So, in a decentralized competitive matching system, any individual who is waiting to be matched at a given point in time must have the same expected payoff as a new-born individual of the same type. Thus, we say that μ is a competitively sustainable plan iff μ is a feasible matching plan and there exists some inhibitive allocation vector w such that

(6.2) $$U_i(\mu) = w_i, \quad \forall i \in N.$$

With these definitions, we can now state the main results of this paper: two existence theorems, a welfare theorem, and an equivalence theorem.

Theorem 2. The set of competitively sustainable plans is nonempty.

Theorem 3. The set of representatively sustainable plans is nonempty.

Theorem 4. Any representatively sustainable plan is weakly Pareto-efficient within the set of feasible matching plans.

Theorem 5. If $J = \emptyset$ (so that no type can imitate any other) then any sustainable plan is both competitively and representatively sustainable.

By Theorems 4 and 5, if there are no informational incentive constraints then all sustainable plans must be Pareto-efficient. Indeed, by Lemma 2, if $J = \emptyset$ then all sustainable plans are in the inner core.

In the above development of our concept of sustainability, we considered only alternatives that get random inputs. Of course, if a waiting list deters individuals from exiting through alternatives that get random imputs, then it can also deter them from exiting through alternatives that get self-selected inputs, because a self-selected input population can have the same type distribution as a random sample from the waiting list. However, if we add the assumption that all alternatives get self-selected inputs, then the interpretation of our concept of sustainability can be somewhat simplified.

If the established system is implementing a competitively sustainable plan μ and all alternatives get self-selected inputs, then there is no need to have a manipulated waiting list at all. This is because, for any inhibitive (w,q) that satisfies (6.2), the distribution of types in the self-selected population that enters any alternative could equal q.

More generally, if all alternatives get self-selected inputs and the established system is implementing a sustainable plan μ that satisfies (6.1) for some inhibitive input vector (w,q), then it suffices to suppose that, for every type i, w_i is the lowest payoff that is expected in the established system by some type-i individuals who are waiting and available at any point in time. Any alternative can then get an inhibitive self-selected input population in which q is the distribution of types and, for each type i, the entering type-i individuals all expect this lowest payoff w_i in the established system. Such individuals would be the most eager (among all type-i individuals) to consider alternatives to the established system; and if they do not choose to exit through an alternative then no one will.

7. Sustainable plans in the example.

Let us see how these solution concepts apply in the context of the example from Section 2.

The input vector $((120, 600, 0), (5/11, 5/11, 1/11))$ is inhibitive because, for any positive ε,

$$((120 + \varepsilon, 600 + \varepsilon, \varepsilon), (5/11, 5/11, 1/11))$$

satisfies the conditions of Lemma 2 with

$$\lambda(1) = 25/44, \quad \lambda(2) = 15/44, \quad \lambda(3) = 1/11, \quad \alpha(1|2) = 5/44, \quad \alpha(2|1) = 0.$$

That is, the payoff allocation generated by the standard separating plan is inhibitive of all alternative matching systems when they get equal numbers of high-productivity (type-1) and low-productivity (type-2) workers. Thus, any feasible matching plan that satisfies

$$U_1(\mu) \geq 120, \quad U_2(\mu) \geq 600, \quad U_3(\mu) \geq 0$$

is sustainable. Such matching plans include the standard separating plan (for which $(U_1(\mu), U_2(\mu), U_3(\mu)) = (120, 600, 0))$, the standard pooling plan (for which $(U_1(\mu), U_2(\mu), U_3(\mu)) = (160, 960, 0))$, and the pooling plan in which all workers are hired full-time at a wage of \$28 per hour (for which $(U_1(\mu), U_2(\mu), U_3(\mu)) = (120, 920, 400))$. Among these sustainable matching plans, the only competitively sustainable matching plan is the standard separating plan.

To understand why the standard separating plan is competitively sustainable, and why other feasible plans are not, recall how in Section 2 we showed that neither the standard separating plan nor the standard pooling plan could be a Rothschild-Stiglitz equilibrium. The standard separating plan is not a Rothschild-Stiglitz equilibrium because a pooling alternative could attract all workers and give employers a positive expected profit, under the assumption that the large majority of all workers are type 1. But this assumption does not apply when our solution concept is competitive sustainability, because the proportion q_1 of type-1 workers in the input population for alternative matching systems can be less than ρ_1. On the other hand, the standard pooling plan is not a Rothschild-Stiglitz equilibrium because it could be perturbed slightly (higher hourly wage, shorter hours) to create an alternative that would attract only the type-1 workers, who generate positive profit for employers. This kind of argument continues to hold when our solution concept is competitive sustainability. To see why, notice that (6.2) and (3.3) imply that $U_1(\mu) = w_1$ and $\hat{U}_2(\mu,1) \leq w_2$. Thus, if type-1 workers are more profitable for employers in a feasible plan μ, then the terms of employment for type-1 workers in μ can be perturbed to create terms that are better than w_1 for type-1 workers and worse than w_2 for type-2 workers. So with inputs (w,q), an alternative offering these terms can match each unreturned employer with 50 profitable type-1 workers, so that he too is better off than in the established system, where he cannot expect more than 10 workers. This argument can be extended to show that neither type of worker can generate positive profits for employers in competitively sustainable plans.

For an example of a feasible plan that is not sustainable at all, even though it is Pareto-efficient within the set of feasible matching plans, consider the pooling plan in which all workers are hired full-time at a wage of \$27 per hour (10 workers per employer). For this matching plan, $(U_1(\mu), U_2(\mu), U_3(\mu)) = (80, 880, 800)$. To see why this allocation vector, and every allocation vector w that is less than or equal to it in all components, is not inhibitive, consider an alternative system that operates as follows. First, every worker is asked whether he would prefer to work full-time (40 hours) at a wage of \$19.40 per hour or to work for 23 hours at a wage of \$29.00 per hour, or return to the established system. With reservation wages of \$25 per hour, the high-productivity workers would all choose the \$29.00 option (notice that $23 \times (29.00 - 25) = 92 > 80 \geq w_1$). With reservation wages of \$5 per hour, the low-productivity workers would choose either the \$19.40 wage or return to the established system (depending on whether w_2 is greater or less than $40 \times (19.40 - 5) = 576$, which is greater than $23 \times (29.00 - 5) = 552$). The workers that have not been returned are then matched (under the employment terms that they chose) with employers, fifty workers to each employer, until either all employers or all workers have been matched; excess workers or employers are compelled to return to the established system. (The determination of which workers are compelled to return, if any, is made independently of their choice of employment terms.) So each employer who is not returned exits with a full complement of fifty workers. From each high-productivity worker he makes $23 \times (30 - 29) = 23$, and from each low-productivity worker he makes $40 \times (20 - 19.40) = 24$, so his profit is at least $50 \times 23 = 1150 > 800 \geq w_3$. So, the prospectus for this alternative system will specify an exit probability equal to 1 for at least one type, at every input vector (w,q) such that $w_1 < 92$ and $w_3 < 1150$. Thus, (80, 880, 800) and any vector that is less than or equal to it in every component, is not an inhibitive allocation.

Another allocation vector which is not inhibitive is (160, 960, 0). For any q in $\Delta^o(N)$, the viable prospectus mapping discussed at the end of Section 4 has nonzero values for any input vector an open neighborhood of ((160, 960, 0), q) (because only case (4.4) applies here). On the other hand, the allocation vector (160, 600, 0) is representatively inhibitive, because, for any positive ε, the input vector

$$((160 + \varepsilon,\ 600 + \varepsilon,\ \varepsilon),\ (9/11,\ 1/11,\ 1/11))$$

satisfies the conditions of Lemma 2 with

$$\lambda(1) = 10/11,\quad \lambda(2) = 0,\quad \lambda(3) = 1/11,\quad \alpha(1|2) = 1/11,\quad \alpha(2|1) = 0.$$

Thus, the standard pooling plan, which gives payoff $40 \times (29 - 20) = 160$ to the high-productivity workers, payoff $40 \times (29 - 5) = 960$ to the low-productivity workers, and expected payoff 0 to the employers, is not competitively sustainable but is representatively sustainable. In fact, it is the unique representatively sustainable plan for this example.

It may be helpful to see why the argument in Section 2 against the standard pooling plan does not apply when our solution concept is representative sustainability. The alternative ($29.50 per hour for 38 hours) that was supposed to overturn the standard pooling plan fails to do so when the expected payoff in the established system for low-productivity workers is $w_2 = 600 < 38 \times (29.50 - 5) = 931$, so that the proposed alternative would not be able to induce low-productivity workers to voluntarily return to the established system. On the other hand, the argument in Section 2 against the standard separating plan (that a pooling plan could be better for all workers and employers) continues to apply when our solution concept is representative sustainability, because representativeness of the input population implies that an alternative will recruit nine times more high-productivity workers than low-productivity workers. To see why no matching plan that gives positive profits to employers can be representatively sustainable, notice that such a plan could be overturned by a viable alternative that gave workers slightly higher wages and attracted employers by promising them to either match them with 50 workers (90% of whom are type-1, by representativeness) or return them to the established system (where they cannot expect more than 10 workers).

It may be surprising that (160, 600, 0) is an inhibitive allocation while (160, 960, 0) is not inhibitive. After all, the two allocations differ only in that type 2 gets a lower payoff in the inhibitive allocation, which makes type-2 individuals more eager to exit through alternative systems. But type-2 individuals are the low-productivity workers, so greater eagerness of type-2 individuals to exit through alternative systems in turn makes it harder for an alternative system to assure employers that they are only being matched with high-productivity workers. Thus, lowering the expected payoff to the low-productivity workers can make an allocation inhibitive.

To understand better how a centralized matching system that implements the standard pooling plan could make a waiting list in which individuals get the representatively inhibitive allocation (160, 600, 0), consider the following perturbation of the standard pooling plan. Let ε and δ be positive numbers, where ε is very small. Each worker is asked to fill out an application form stating whether he is a high-productivity worker or a low-productivity worker. If he says that he is a high-productivity worker then, with probability $1 - 2\varepsilon$, he is assigned immediately to an exit configuration in which he works 40 hours at a wage of \$29 per hour; with probability ε, he is assigned immediately to an exit configuration in which he works 23 hours at a wage of \$30 per hour; and, with probability ε, he is asked to wait some period of time δ after which he will be assigned to an exit configuration in which he works 40 hours at a wage of \$29 per hour. If he says that he is a low-productivity worker then, with probability $1 - \varepsilon$, he is assigned immediately to an exit configuration in which he works 40 hours at a wage of \$29 per hour; and, with probability ε, he is asked to wait a period of time δ

after which he will be assigned to an exit configuration in which he works 40 hours for a wage of \$20 per hour. It is straightforward to check that each type of worker gets a strictly higher expected payoff from being honest than from lying about his type in his application form. Notice, however, that among the individuals who are asked to wait, the high-productivity workers have an expected payoff of $40 \times (29 - 25) = 160$, and the low-productivity workers have an expected payoff of $40 \times (20 - 5) = 600$. Notice also that each type of worker has the same expected waiting time $\varepsilon\delta$ when he enters this system. Thus (if employers also have the same expected waiting time), the allocation and distribution vectors that represent the population available in the waiting list at any point in time are

$$w = (160, 600, 0) \quad \text{and} \quad q = \rho = (9/11, 1/11, 1/11).$$

As ε goes to zero, the matching plan implemented by this system converges to the standard pooling plan.

Baumol, Panzar, and Willig [1986] have argued that sustainability against short-term entry by potential competitors (alternatives) is a sufficient condition to guarantee Pareto-efficiency of economic systems. This conclusion does not hold in dynamic matching problems with adverse selection. However, we have seen that representatively sustainable matching plans always exist and are Pareto-efficient. So Baumol, Panzar, and Willig's argument can be extended to the markets with adverse selection if alternatives can be assured that their inputs have the same distribution of types as the population that arrives in the economy during any period of time. If all types get the same expected waiting time (on arrival) in the established system, then alternatives that get random inputs can be assured such a representative type distribution.

On the other hand, Pareto-efficiency may be incompatible with decentralization. In our example, the unique competitively sustainable matching plan is strictly Pareto-dominated by other feasible matching plans. It is competitively sustained by the input vector ((120, 600, 0), (5/11, 5/11, 1/11)). To create such a distribution vector (with equal numbers of high- and low-productivity workers) in the population that is available at any point in time, low-productivity workers must expect to wait and search nine times longer than the high-productivity workers (because the ratio of birth rates is $\rho_1/\rho_2 = 9$). In a decentralized market, such a waiting-time inequality could occur with a separating plan if most employers only offered the terms of employment that attract high-productivity workers, so that low-productivity workers would have to search longer find their terms of employment. (In a model with small waiting costs, the standard separating plan could be perturbed slightly so that low-productivity workers would strictly prefer their terms over the terms that the high-productivity workers take.) In general, then, our analysis suggests that competitive markets with adverse selection may be sustained in a Pareto-inefficient equilibrium by a kind of Gresham's Law, according to which the good types are only briefly available but the bad types circulate widely.

8. Proofs.

We begin with some preliminary definitions and observations.

Suppose that we are given a dynamic matching problem, as generally defined in Section 3. Finiteness of the sets N and E implies that we can select some number B such that

$$|u_i(e)| \leq B \quad \text{and} \quad |\hat{u}_i(e,j)| \leq B, \quad \forall i \in N, \quad \forall e \in E, \quad \forall j \in N.$$

This in turn implies that, for any matching plan η, (8.1) $\quad |V_i(\eta)| \le B|R_i(\eta)|$ and $|\hat{V}_i(\eta,j)| \le B|R_j(\eta)|$, $\forall i \in N$, $\forall j \in N$.

For any (w,q) in $\mathbb{R}^N \times \Delta^o(N)$, let $H(w,q)$ denote the set of all (y,s) in $\mathbb{R}_+^N \times [0,1]^N$ for which there exists some viable alternative matching plan η, satisfying (4.1)-(4.3), such that

(8.2) $\quad y_i = (V_i(\eta) - R_i(\eta)\, w_i)/q_i$ and $s_i = R_i(\eta)/q_i$, $\forall i \in N$.

It is straightforward to check that H is upper-semicontinuous on its domain, and that $H(w,q)$ is always a nonempty convex set. H satisfies the following boundedness property:

(8.3) $\quad$ if $(y,s) \in H(w,q)$ then $|y_i + s_i w_i| \le Bs_i$,

because there is some η such that $V_i(\eta) = q_i(y_i + s_i w_i)$ and $R_i(\eta) = q_i s_i$. So, given $(y,s) \in H(w,q)$,

$$\text{if } w_i \ge 0 \text{ then } 0 \le y_i \le B,$$

and

$$\text{if } w_i > B \text{ then } s_i = 0.$$

(The participation constraint (4.2) guarantees that $y_i \ge 0$.) As defined in Section 4, a viable prospectus mapping is any upper-semicontinuous correspondence $F:\mathbb{R}^N \times \Delta(N) \twoheadrightarrow \mathbb{R}_+^N \times [0,1]^N$ such that, for any (w,q) in $\mathbb{R}^N \times \Delta^o(N)$, $\varnothing \ne F(w,q) \subseteq H(w,q)$.

Let $\bar{H}:\mathbb{R}^N \times \Delta(N) \twoheadrightarrow \mathbb{R}_+^N \times [0,1]^N$ be the minimal upper-semicontinuous extension of H to all of $\mathbb{R}^N \times \Delta(N)$. Then $\bar{H}$ is a viable prospectus mapping.

<u>Proof of Lemma 1</u>.

An input vector (w,q) in $\mathbb{R}^N \times \Delta^o(N)$ is strongly inhibitive iff $H(w,q) = \{(\underset{\sim}{0},\underset{\sim}{0})\}$. (Here $\underset{\sim}{0}$ is the zero vector in $\mathbb{R}^N$.) Let Θ denote the closure in $\mathbb{R}^N \times \Delta(N)$ of the set of all strongly inhibitive input vectors. We need to show that Θ is the set of all inhibitive input vectors.

Suppose first that $(w,q) \in \Theta$, so that (w,q) is a limit of some sequence of strongly inhibitive input vectors $\{(w^k,q^k)\}_{k=1}^{\infty}$. Then for any viable prospectus F, $F(w^k,q^k) = \{(\underset{\sim}{0},\underset{\sim}{0})\}$ for every k, and so, by upper-semicontinuity of F, $(\underset{\sim}{0},\underset{\sim}{0}) \in F(w,q)$. Thus, if $(w,q) \in \Theta$ then (w,q) is inhibitive.

If $\hat{\eta}$ is a viable alternative matching plan with inputs (w,q) and if β is some positive scalar such that $\beta R_i(\hat{\eta}) \le q_i$ for every type i, then $\eta = \beta\hat{\eta}$ is also a viable alternative with inputs (w,q). (This holds because (4.1) is the only constraint in the definition of a viable alternative plan that is not linearly homogeneous in η.) So, given any input vector (w,q) in $\mathbb{R}^N \times \Delta^o(N)$, if (w,q) is not strongly inhibitive, so that there exists a nonzero plan that is a viable alternative with (w,q), then there exists some viable alternative η such that $R_i(\eta) = q_i$ for at least one type i. (Let $\beta = \min_{i\in N} q_i/R_i(\eta)$.) Thus, if (w,q) is not strongly inhibitive then there exists some (y,s) in H(w,q) such that $s_i = 1$ for at least one type i.

Define the mapping $G:\mathbb{R}^N \times \Delta(N) \twoheadrightarrow \mathbb{R}_+^N \times [0,1]^N$ so that, for every (w,q) in $\mathbb{R}^N \times \Delta(N)$,

if $(w,q) \in \Theta$ then $G(w,q) = \bar{H}(w,q)$,

and

if $(w,q) \notin \Theta$ then $G(w,q) = \{(y,s) \in \bar{H}(w,q) |\ \Sigma_{i\in N}\, s_i \ge 1\}$.

It is straightforward to show that G is an upper-semicontinuous correspondence, and that, for any (w,q), G(w,q) is a nonempty subset of $\bar{H}(w,q)$. So G is a viable prospectus mapping. Thus, if (w,q) is inhibitive then $(\underset{\sim}{0},\underset{\sim}{0}) \in G(w,q)$. But $(\underset{\sim}{0},\underset{\sim}{0}) \in G(w,q)$ if and only if $(w,q) \in \Theta$. So if (w,q) is inhibitive then (w,q) must be in Θ. So Θ is indeed the set of all inhibitive input vectors.

No allocation vector with negative components could be strongly inhibitive, because an alternative plan that lets all individuals with negative allocations exit in solitary configurations would be viable. Thus, if (w,q) is inhibitive then every component w_i must be nonnegative. Q.E.D.

In the proof of Lemma 1, we have also shown that any input vector (w,q) in $\mathbb{R}^N \times \Delta(N)$ is inhibitive if and only if $(\underset{\sim}{0},\underset{\sim}{0}) \in G(w,q)$. For this reason, we may refer to G(•) as the universal prospectus for our dynamic matching problem.

Proof of Theorem 1.

If there exists some q such that (w,q) is an inhibitive input vector, then w satisfies the definition of an inhibitive allocation vector with this distribution vector q, for every viable prospectus.

Conversely, if w is an inhibitive allocation vector, then there exists some q such that $(\underset{\sim}{0},\underset{\sim}{0}) \in G(w,q)$, because G is a viable prospectus. But then, by the remark following the proof of Lemma 1, (w,q) must be an inhibitive allocation. Q.E.D.

Proof of Theorem 2.

We can naturally extend the correspondence H to the domain $\mathbb{R}^N \times \mathbb{R}^N_{++}$ so that $(y,s) \in H(w,q)$ iff there exists some η in $\mathbb{R}^E_+$ satisfying (4.1)-(4.3) and (8.2). ($\mathbb{R}^N_{++}$ is the orthant of vectors in $\mathbb{R}^N$ in which all components are strictly positive.) With this extended definition of H, for any $(w,\hat{q})$ in $\mathbb{R}^N \times \Delta^o(N)$ and any positive scalar β, $H(w,\beta\hat{q}) = H(w,\hat{q})$.

Let us use this extended definition of H to also define G for any vector in $\mathbb{R}^N \times \mathbb{R}^N_{++}$. That is, for any $(w,\hat{q})$ in $\mathbb{R}^N \times \Delta^o(N)$ and any positive scalar β, we let $G(w,\beta\hat{q}) = G(w,\hat{q})$.

Let ε be a small positive number. Let n denote the number of types in the set N. (That is, $n = |N|$.) We define a function $\Phi_\varepsilon:[0,B]^N \times [0,1]^N \to \mathbb{R}^N_+ \times \mathbb{R}^N_{++}$ so that $(w,q) = \Phi_\varepsilon(y,s)$ iff, for every i in N,

$$q_i = \rho_i(1+\varepsilon)/(s_i+\varepsilon) \quad \text{and} \quad w_i = y_i/\varepsilon + (s_i+\varepsilon)nB.$$

Because Φ_ε is continuous and G upper-semicontinuously maps vectors in $\mathbb{R}^N_+ \times \mathbb{R}^N_{++}$ to nonempty convex subsets of $[0,B]^N \times [0,1]^N$, the Kakutani fixed-point theorem implies that we can select (y,s) so that $(y,s) \in G(\Phi_\varepsilon(y,s))$. Let $(w,q) = \Phi_\varepsilon(y,s)$, where (y,s) is this fixed point of $G(\Phi_\varepsilon(\bullet))$. Applying the definitions of G and Φ_ε, we can select a plan η in $\mathbb{R}^E_+$ that satisfies the viability constraints with (w,q) and such that, for every type i,

$$q_i = \rho_i(1+\varepsilon)/((R_i(\eta)/q_i)+\varepsilon) \quad \text{and}$$

$$w_i = (V_i(\eta) - w_i\, R_i(\eta))/(\varepsilon q_i) + (R_i(\eta)/q_i + \varepsilon)nB.$$

These two equations imply by straightforward algebra that, for every i,

$$\text{(8.4)} \qquad \varepsilon q_i = \rho_i(1+\varepsilon) - R_i(\eta) \quad \text{and} \quad w_i = V_i(\eta)/(\rho_i(1+\varepsilon)) + \varepsilon nB.$$

The above equations also imply that

$$(V_i(\eta) - w_iR_i(\eta))/q_i = \varepsilon V_i(\eta)/(\rho_i(1+\varepsilon)) - \varepsilon nBR_i(\eta)/q_i, \quad \forall i \in N.$$

But then, using (4.2),

$$0 \le \Sigma_{i\in N}\,(V_i(\eta) - w_iR_i(\eta))/q_i = \Sigma_{i\in N}\,(\varepsilon V_i(\eta)/(\rho_i(1+\varepsilon)) - \varepsilon nBR_i(\eta)/q_i)$$
$$\le n\varepsilon B/(1+\varepsilon) - n\varepsilon B\,\Sigma_{i\in N}\,R_i(\eta)/q_i.$$

This implies that $\Sigma_{i\in N}\, s_i = \Sigma_{i\in N}\, R_i(\eta)/q_i \le 1/(1+\varepsilon) < 1$. Let $\hat{q}$ be the distribution vector in $\Delta^o(N)$ defined by $\hat{q}_i = q_i/\Sigma_{j\in N}\, q_j$ for every i. Then, by definition of G, $(w,\hat{q})$ is an inhibitive input vector, because

$$G(w,q) = G(w,\hat{q}) \ne \{(y,s) \in H(w,\hat{q}) \mid \ \Sigma_{i\in N}\, s_i \ge 1\}.$$

Furthermore, using (4.3) and (8.4), for any (j,i) in J,

$$0 \le (V_i(\eta) - w_iR_i(\eta))/q_i - (\hat{V}_i(\eta,j) - w_iR_j(\eta))/q_j$$
$$= \varepsilon V_i(\eta)/(\rho_i(1+\varepsilon)) - \varepsilon nBR_i(\eta)/q_i$$
$$- \hat{V}_i(\eta,j)/q_j + (V_i(\eta)/(\rho_i(1+\varepsilon)) + \varepsilon nB)R_j(\eta)/q_j$$

$$= (\varepsilon + R_j(\eta)/q_j)(V_i(\eta)/(\rho_i(1 + \varepsilon))) - \hat{V}_i(\eta,j)/q_j$$
$$- \varepsilon nB(R_i(\eta)/q_i - R_j(\eta)/q_j)$$
$$= (\rho_j/q_j)(V_i(\eta)/\rho_i - \hat{V}_i(\eta,j)/\rho_j) - \varepsilon nB(R_i(\eta)/q_i - R_j(\eta)/q_j).$$

Thus,

$$V_i(\eta)/\rho_i - \hat{V}_i(\eta,j)/\rho_j \geq (\varepsilon nB)(R_i(\eta)q_j/(\rho_j q_i) - R_j(\eta)/\rho_j).$$

Rewriting the first equation in (8.4) as $R_i(\eta) = \rho_i + \varepsilon(\rho_i - q_i)$ and applying (4.1), we conclude that

$$R_i(\eta) \leq \rho_i, \quad \forall i \in N.$$

Thus,

$$V_i(\eta)/\rho_i - \hat{V}_i(\eta,j)/\rho_j \geq -\varepsilon nB. \tag{8.5}$$

Now let μ be the matching plan such that, for every type i, $\mu(\bar{e}_i) = \eta(\bar{e}_i) + (\rho_i - R_i(\eta))$, and $\mu(e) = \eta(e)$ for every exit configuration e other than the solitary exit configurations. Then

$$V_i(\mu) = V_i(\eta), \quad \hat{V}_i(\mu,j) = \hat{V}_i(\eta,j), \quad R_i(\mu) = \rho_i, \quad \forall i \in N, \quad \forall j \in N.$$

By (8.4), $U_i(\mu)/(1 + \varepsilon) = V_i(\mu)/(\rho_i(1 + \varepsilon)) = w_i - \varepsilon nB$, so

$$|U_i(\mu) - w_i| = |U_i(\mu)\varepsilon/(1 + \varepsilon) - \varepsilon nB)| \leq \varepsilon(1 + n)B.$$

By (8.5), $U_i(\mu) - \hat{U}_i(\mu,j) = V_i(\eta)/\rho_i - \hat{V}_i(\eta,j)/\rho_j \geq -\varepsilon nB.$

Now consider a sequence of values for ε that go to zero, and let $\bar{w}$, $\bar{q}$, and $\bar{\mu}$ be limits of convergent subsequences of the w, $\hat{q}$, and μ vectors constructed from these values of ε as above. (Such a limit $\bar{\mu}$ exists because, for each ε, μ is in the compact subset of $\mathbb{R}_+^E$ in which

$$\Sigma_{e \in E} (\Sigma_{i \in N} r_i(e)) \mu(e) \leq \Sigma_{i \in N} \rho_i = 1.)$$

Then $(\bar{w},\bar{q})$ is inhibitive, $U_i(\bar{\mu}) = \bar{w}_i$ for every i, and $U_i(\bar{\mu}) \geq \hat{U}_i(\bar{\mu},j)$ for every (j,i) in J. Thus, $\bar{\mu}$ is a competitively sustainable plan. Q.E.D

<u>Proof of Theorem 3</u>.

Let Θ^* denote the set of all representatively inhibitive allocation vectors in $\mathbb{R}^N$. Let $\Gamma:\mathbb{R}^N \twoheadrightarrow \mathbb{R}_+^N \times [0,1]^N$ be defined so that

$$\Gamma(w) = H(w,\rho) \quad \text{if} \quad w \in \Theta^*,$$

$$\Gamma(w) = \{(y,s) \in H(w,\rho) \mid \textstyle\sum_{i\in N} s_i \geq 1\} \quad \text{if} \quad w \in \mathbb{R}_+^N \quad \text{but} \quad w \notin \Theta^*,$$

$$\Gamma(w) = \{(y,s) \in H(w,\rho) \mid \forall i, \text{ if } w_i < 0 \text{ then } s_i = 1\}, \quad \text{if} \quad w \notin \mathbb{R}_+^N.$$

If $w \notin \Theta^*$ then (w,ρ) is not strongly inhibitive and so there does exist some (y,s) in $H(w,\rho)$ such that $\sum_{i\in N} s_i \geq 1$ (as we showed in the proof of Lemma 1). Also, if $w \notin \mathbb{R}_+^N$ then the plan $\bar{\eta}$, that is defined so that $\bar{\eta}(\bar{e}_i) = \rho_i$ for every i such that $w_i < 0$, and $\bar{\eta}(e) = 0$ for every other e in E, is viable with inputs (w,ρ). Thus, for every w, $\Gamma(w)$ is a nonempty convex set. Γ is upper-semicontinuous because Θ^* and $\mathbb{R}_+^N$ are closed, $\Theta^* \subseteq \mathbb{R}_+^N$, and H is upper-semicontinuous.

Now let $\Psi:[-1, B+1]^N \twoheadrightarrow [-1, B+1]^N$ be the correspondence defined so that $x \in \Psi(w)$ iff there exists some (y,s) in $\Gamma(w)$ such that

$$x_i = w_i + s_i - 1/(n+1), \quad \forall i \in N.$$

(Notice that s_i must equal zero here if $w_i > B$, by (8.3); whereas s_i must equal one if $w_i < 0$, by definition of Γ.) By the Kakutani fixed-point theorem, there exists some w such that $w \in \Psi(w)$.

This fixed point w must be a representatively inhibitive allocation in Θ^*, because there is some (y,s) in $\Gamma(w)$ such that $s_i = 1/(n+1)$ for every i, and so $\sum_{i\in N} s_i < 1$. Furthermore, by definition of H, there must exist some η such that η is a viable alternative plan with inputs (w,ρ) and $R_i(\eta)/\rho_i = 1/(n+1)$ for every type i. Now let $\mu = (n+1)\eta$. Then μ is also a viable alternative plan with inputs (w,ρ), and $R_i(\mu) = \rho_i$ for every type i. But then the viability constraint (4.3) implies the informational incentive constraint (3.3) in the definition of feasibility. So μ is a feasible plan. Furthermore, the viability constraint (4.2) implies the sustainability constraint (6.1). So μ is a representatively sustainable plan.

Q.E.D.

Proof of Theorem 4.

Suppose that μ is a feasible plan that is Pareto-inefficient. Then there exists some other feasible plan ν such that $U_i(\nu) > U_i(\mu)$ for every i in N. But then, for any $\hat{w}$ such that $\hat{w}_i \leq U_i(\nu)$ for every i, ν is a viable alternative plan with input vector $(\hat{w},\rho)$, and so $(\hat{w},\rho)$ is not strongly inhibitive.

Thus, if w is any vector such that $w_i \leq U_i(\mu) < U_i(\nu)$ for every i, then (w,ρ) cannot be the limit of any sequence of strongly inhibitive input vectors of the form $(\hat{w},\rho)$, and so w cannot be representatively inhibitive. So the Pareto-inefficient plan μ cannot be representatively sustainable. Q.E.D.

Proof of Theorem 5.

Suppose $J = \emptyset$. If $q_i > 0$ for every i, then there exists a nonzero matching plan that is viable with inputs (w,q) iff there exists a nonzero matching plan η such that $V_i(\eta) \geq R_i(\eta)\, w_i$ for every type i. (Even if this η violates (4.1), some positive multiple of η would satisfy (4.1).) Thus, if (w,q) is any strongly inhibitive input vector, then so is $(\hat{w},\rho)$, for any $\hat{w}$ such that $\hat{w}_i \geq w_i$ for every type i. Thus, if $U_i(\mu) \geq w_i$ for every i and w is an inhibitive allocation vector, then $(U_i(\mu))_{i\in N}$ is a representatively inhibitive allocation vector. Q.E.D.

Proof of Lemma 2.

Given any input vector (w,q) in $\mathbb{R}^N \times \Delta^o(N)$ and any positive number ε, consider the following linear programming problem:

$$\text{maximize } \sum_{e\in E} \varepsilon\eta(e) \text{ subject to } \eta \in \mathbb{R}^E_+,$$

$$(V_i(\eta) - w_i R_i(\eta))/q_i \geq 0, \quad \forall i \in N,$$

$$(V_i(\eta) - w_i R_i(\eta))/q_i - (\hat{V}_i(\eta,j) - w_i R_j(\eta))/q_j \geq 0, \quad \forall (j,i) \in J.$$

By substituting in the definitions of $V_i(\eta)$, $R_i(\eta)$ and $\hat{V}_i(\eta,j)$, it can be shown that the dual to this linear programming problem is:

$$\text{minimize } 0 \text{ subject to } \lambda \in \mathbb{R}^E_+, \quad \alpha \in \mathbb{R}^J_+,$$

$$\sum_{i\in N} \lambda(i)(w_i - u_i(e))r_i(e)/q_i$$
$$+ \sum_{(j,i)\in J} \alpha(j|i)((w_i - u_i(e))r_i(e)/q_i - (w_i - \hat{u}_i(e,j))r_j(e)/q_j) \geq \varepsilon, \quad \forall e \in E.$$

(We write the (j,i)-component of α as "$\alpha(j|i)$" to emphasize that this is the dual variable corresponding to the constraint that "an individual should not have an incentive to claim to be type j, given that his actual type is i.") For any positive ε, the input vector (w,q) is strongly inhibitive if and only if the first linear programming problem has an optimal solution with a value of zero. By the duality theorem of linear programming, the first linear program has an optimal solution with a value of zero if and only if the dual problem has a feasible solution that satisfies its constraints. But the dual constraint for each e can be rewritten

$$\sum_{i\in N} ((\lambda(i) + \sum_{j\in N} \alpha(j|i))w_i - \sum_{j\in N} \alpha(i|j)w_j)r_i(e)/q_i$$
$$\geq \sum_{i\in N} ((\lambda(i) + \sum_{j\in N} \alpha(j|i))u_i(e) - \sum_{j\in N} \alpha(i|j)u_j(e,i))r_i(e)/q_i + \varepsilon,$$

when we adopt the convention that $\alpha(j|i) = 0$ if $(j,i) \notin J$. Thus, the constraints in Lemma 2 can be satisfied if and only if (w,q) is strongly inhibitive. Q.E.D.

REFERENCES

W. J. Baumol, J. C. Panzar, and R. D. Willig, [1986] "On the Theory of Perfectly-Contestable Markets," in New Developments in the Analysis of Market Structure, edited by J. E. Stiglitz and G. F. Mathewson, Cambridge: M.I.T. Press.

G. Butters [1984], "Equilibrium Price Distributions in a Random Meetings Market," Princeton University discussion paper.

P. Dasgupta and E. Maskin [1986], "The Existence of Equilibrium in Discontinuous Economic Games, 2: Applications," Review of Economic Studies 53, 27-41.

P. A. Diamond [1982], "Wage Determination and Efficiency in Search Equilibrium," Review of Economic Studies 49, 217-227.

D. Fudenberg and E. Maskin [1986], "The Folk Theorem in Repeated Games with Discounting and Incomplete Information," Econometrica 54, 533-554.

D. Gale [1986], "Bargaining and Competition Part I: Characterization" and "Bargaining and Competition Part II: Existence," Econometrica 54, 785-806, 807-818.

D. Gale [1987], "A Walrasian Theory of Markets with Adverse Selection," University of Pittsburgh discussion paper.

M. Hellwig [1986a], "A Note on the Specification of Inter-Firm Communication in Insurance Markets with Adverse Selection," University of Bonn discussion paper.

M. Hellwig [1986b], "Some Recent Developments in the Theory of Competition in Markets with Adverse Selection," University of Bonn discussion paper.

G. D. Jaynes [1978], "Equilibria in Monopolistically Competitive Insurance Markets," Journal of Economic Theory 19, 394-422.

H. Miyazaki [1977], "The Rat Race and Internal Labor Markets," Bell Journal of Economics 8, 394-418.

D. T. Mortensen [1982], "The Matching Process as a Noncooperative Bargaining Game," in Economics of Information and Uncertainty, edited by J. McCall, Chicago: University of Chicago Press.

R. B. Myerson [1984a], "Two-Person Bargaining Problems with Incomplete Information," Econometrica 52, 461-487.

R. B. Myerson [1984b], "Cooperative Games with Incomplete Information," International Journal of Game Theory 13, 69-96.

J. G. Riley [1979], "Informational Equilibrium," Econometrica 47, 331-359.

M. Rothschild and J. Stiglitz [1976], "Equilibrium in Competitive Insurance Markets," Quarterly Journal of Economics 90, 629-650.

A. Rubinstein [1986], "Competitive Equilibrium in a Market with Decentralized Trade and Strategic Behaviour: an Introduction," London School of Economics discussion paper.

A. Rubinstein and A. Wolinsky [1985], "Equilibrium in a Market with Sequential Bargaining," Econometrica 53, 1133-1150.

M. Shubik [1982], Game Theory in the Social Sciences, Cambridge: M.I.T. Press.

M. Spence [1973], "Job Market Signaling," Quarterly Journal of Economics 87, 355-374.

C. Wilson [1977], "A Model of Insurance Markets with Incomplete Information," Journal of Economic Theory 16, 167-207.