

Roth, A.E.,
"Stability and Polarization of Interests in
Job Matching,"
Econometrica, 52, 1984, 47-57.

The copyright to this article is held by the Econometric Society, <http://www.econometricsociety.org>. It may be downloaded, printed and reproduced only for personal or classroom use. Absolutely no downloading or copying may be done for, or on behalf of, any for-profit commercial firm or other commercial purpose without the explicit permission of the Econometric Society. For this purpose, contact Julie P. Gordon, Executive Director of the Society, at: jpg@northwestern.edu.

STABILITY AND POLARIZATION OF INTERESTS IN JOB MATCHING

BY ALVIN E. ROTH¹

A model of job-matching is considered, in which the set of employees hired by each firm, and the set of jobs accepted by each worker, are endogenously determined, as are the job descriptions settled on by each worker-firm pair. The set of outcomes that are in equilibrium, in the sense of being stable with respect to recontracting, is shown to be nonempty. It is shown that the interests of the firms and workers are polarized over the set of stable outcomes: There exists a firm-optimal stable outcome that is the best stable outcome for every firm and the worst for every worker, and a corresponding worker-optimal stable outcome that is best for every worker and worst for every firm. These results generalize and extend previous results for models of this type, and raise questions about the nature and underlying causes of such polarization of interests.

1. INTRODUCTION

IN A JUSTLY FAMOUS PAPER, Gale and Shapley [5] studied two-sided "marriage markets" whose stable (core) outcomes reflect a surprising coincidence of interest among agents on the same side of the market, and a corresponding conflict of interest between agents on opposite sides. Markets of this kind consist of two kinds of heterogeneous agents (e.g., firms and workers) each of whom has its own preferences over potential matches with agents of the other kind. A "market game" of this type serves to advance the common interests of firms and workers wishing to be matched, and to resolve the conflicting interests of firms, who are in competition for desirable workers, and of workers, who are in competition for desirable firms.

It is thus surprising to find this overall pattern of common and conflicting interests reversed when attention is confined to outcomes that are stable, in the sense that no pair of agents both prefer to be matched to each other rather than to their assigned matches. (The set of stable outcomes constitutes the core of this game, which can be identified with the competitive outcomes of a closely related market; cf. Kelso and Crawford [7].) Specifically, when agents in the simple symmetric model studied by Gale and Shapley are not indifferent between potential matches, there is a "firm-optimal" outcome in the core which every firm agrees is the best core outcome, and a corresponding "worker-optimal" core outcome which is best for every worker.² In addition, it is straightforward to

¹ It is a pleasure to acknowledge stimulating conversations on this topic with Charles Blair, Vincent Crawford, Zvi Ritz, and Francoise Schoumaker. This work has been supported by NSF Grants SOC78-09928 and SES-8207690.

² For example, consider the simple marriage game (whose agents are "men" and "women") in which all the men happen to rank the same woman as the most desirable match. Then these men disagree about what is the most desirable outcome of the game, since each man prefers an outcome which gives him his first choice (and hence gives other men lower choices, in this example). However when attention is confined to stable outcomes, the men all agree which is the best stable outcome. (In particular, in this example any stable outcome must match the most preferred woman with the man she ranks first, and so this woman is not a possible match for any of the other men, in a competitive marriage market whose outcomes are stable.)

show in this model that the optimal core outcome for one side of the market is the worst core outcome for every agent on the other side of the market (cf. Knuth [8]).³

This polarization of interests in the core has also been observed in a number of related models. Shapley and Shubik [16] and Thompson [17] considered transferable utility models in which all outcomes can be evaluated in terms of monetary outcomes, and Crawford and Knoer [2] considered a nontransferable utility model of a labor market in which salaries and job descriptions, as well as job assignments, are determined endogenously. (This last model is a generalization of Gale and Shapley's marriage market.) These models treat firms and workers in a symmetrical way, and assume that the number of employees required by each firm is fixed, and that firms have separable preferences over workers. In each of these models, there is again a core outcome optimal for firms, which is the worst for every worker, and a dual outcome that is the worst outcome in the core for every firm and the best for every worker.

Kelso and Crawford [7] generalized the model of Crawford and Knoer by dropping the assumption that firms have fixed needs for workers and separable preferences. In this model workers may work for only one firm, but firms may employ any number of workers, and each firm has preferences over *sets* of employees, salaries, and job descriptions. Kelso and Crawford showed that, when firms regard individual workers as substitutes in a certain sense, and when agents are not indifferent between matches in the core and other potential matches, there is an outcome in the core which is simultaneously the best core outcome for every firm. Since firms and workers are not symmetric in this model, Kelso and Crawford's treatment leaves open the question whether there exists a dual core outcome which is best for all workers.⁴

One result demonstrated below is that such a worker-optimal outcome does indeed exist, when agents have suitably strict preferences. In fact, we shall consider a more general, completely symmetric model, which determines endogenously not only the workforce of each firm but also the number and kind of jobs held by each worker, and see that in this model too, there are firm-optimal and worker-optimal stable outcomes. The stable outcome which is best for all the firms is shown to be worst for all the workers, and vice-versa.

³These optimal stable outcomes also reflect the two-sided nature of the market in a surprising way when the matching process is considered as a noncooperative game in which each agent's preferences are unknown to the other agents. In Roth [11] it was shown that no procedure exists for aggregating agents' stated preferences to produce a stable outcome in a way which gives every agent the incentive to reveal his preferences truthfully. However a procedure which aggregates stated preferences to produce the optimal stable outcome for one side of the market makes it a dominant strategy for agents on that side of the market to correctly reveal their preferences, so that the incentive to misrepresent can be confined to one side of the market. (This latter fact was independently observed by Dubins and Freedman [4] and has been extended to some related models by Ritz [10] and by Demange [3]. The nonexistence result was independently obtained by Bergstrom and Manning [1].)

⁴Kelso and Crawford [7] state "We have not considered the effects of reversing the roles of firms and workers . . . Because of the use made of our assumption that workers are indifferent about which other workers their firms hire, this seems to involve significantly greater difficulties . . ."

2. THE MODEL

The agents in this model consist of a finite set W of m workers, and a finite set F of n firms. For each worker-firm pair (i, j) , the finite set $X(i, j)$ represents the feasible job descriptions at which worker i could be employed by firm j . That is, the elements x_{ij} of $X(i, j)$ represent all the different feasible combinations of salary, working conditions, responsibilities, etc., which could be agreed upon by worker i and firm j . (The finiteness of $X(i, j)$ embodies the assumption that the elements of a job description can take on only discrete values; e.g., salary cannot be specified more precisely than to the nearest penny, hours to the nearest second, etc.) Two otherwise similar jobs which may be offered by different firms or held by different workers will be considered distinct, so that $X(i, j)$ is disjoint from $X(k, l)$ if $(i, j) \neq (k, l)$. Also associated with each agent k is an alternative u_k which corresponds to being unmatched (so u_k corresponds to unemployment for a worker k , or to having an unfilled position for a firm k).

For each firm j , let $X(j)$ denote the union of the sets $X(i, j)$ over all workers i , and for each worker i let $X(i)$ denote the union over all firms j of the sets $X(i, j)$. For each agent k , let $\bar{X}(k) = X(k) \cup \{u_k\}$, so that $\bar{X}(k)$ represents all the job descriptions possible for agent k , together with the possibility of remaining unmatched.

Firms have preferences over sets of employees, and workers have preferences over sets of jobs. For any firm j let $Y(j)$ be the set of all nonempty subsets of $X(j)$ that contain at most one element from any $X(i, j)$, and for any worker i let $Y(i)$ be the set of all nonempty subsets of $X(i)$ that contain at most one element of any $X(i, j)$. Each agent k has a complete and transitive binary preference relation defined on $\bar{Y}(k) = Y(k) \cup \{u_k\}$. If a and b are elements of $\bar{Y}(k)$, then $aP_k b$ means agent k (strictly) prefers a to b , and $aR_k b$ means agent k likes a at least as well as b . Agent k 's preferences will be called *strict* if $aR_k b$ implies that either $a = b$ or else $aP_k b$.

The polarization of interests on which this paper focuses depends on agents not being indifferent between assignments that they receive at different stable outcomes (see Roth [13]). It is simplest to suppose that agents are not indifferent between any two feasible assignments, i.e. that agents' preferences are strict. This can be regarded as the generic case. For example, if each agent's preferences correspond to a utility function whose value for each alternative is partly determined by a random error term drawn from a continuous probability distribution, then with probability one every agent has strict preferences.

So every agent k in this model (workers as well as firms) is assumed to have strict preferences over the set $\bar{Y}(k)$. For any agent k and any subset S of $\bar{X}(k)$, $C_k(S)$ is agent k 's most preferred element of $\bar{Y}(k)$ that is a subset of S , i.e. $C_k(S)$ is agent k 's choice set from S .

An outcome of this market specifies which workers are employed by each firm, which firms employ each worker, the job description for each worker-firm pair, and which workers and firms are unmatched. If E is the subset of employed workers at a given outcome, and E' the subset of firms who find at least one

employee, then the assignment of workers to firms is specified by a correspondence f from $E \cup E'$ to itself where $f(j)$ denotes the workers employed by firm j and $f(i)$ denotes the firms who employ worker i . The correspondence f takes E into E' and E' into E , and has the property that i is an element of $f(j)$ if and only if j is an element of $f(i)$. An outcome of the market is specified by an $(m+n)$ -vector x^f such that for each k in $E \cup E'$, x_k^f is the set of feasible job descriptions held by agent k , and for each k not in E or E' , $x_k^f = u_k$ indicates that agent k is unmatched. Some job description x_{ij} in $X(i, j)$ matching worker i and firm j is contained in x_i^f if and only if j is contained in $f(i)$ (and vice versa), in which case x_{ij} is also contained in x_j^f . At most one element of $X(i, j)$ is contained in any x_i^f or x_j^f .

An outcome x^f is *stable* if it is individually rational and if no outcome y^g exists with subsets S of workers and S' of firms with $S = g(S')$ and $S' = g(S)$, for which $y_k^g \subset C_k(x_k^f \cup y_k^g)P_k x_k^f$ for all agents k in $S \cup S'$.

Thus if x^f is a stable outcome, there is no coalition of firms and workers that can propose among themselves a set of job descriptions, different from those assigned to them at x^f , that every agent in the coalition would include in his most preferred feasible set of job descriptions drawn from the new proposals together with those already assigned to the agent. Since preferences are assumed to be strict, such a coalition, if one existed, would make all its members strictly better off. Note that a stable outcome x^f must have the property that $x_k^f = C_k(x_k^f)$ for each agent k (since otherwise it would be unstable via agent k and the empty set S (or S') of agents on the other side of the market).⁵

Without some further assumptions about the nature of firms' preferences, the set of stable outcomes could be empty. The two assumptions we adopt below, to ensure that stable outcomes always exist, both put some limits on the way in which an agent's preferences for a particular job description can depend on what other job descriptions are in its portfolio. The first of these, Pareto separability, requires that, if at some outcome x^f , a given job description x_{ij} in x_i^f and x_j^f is Pareto optimal in $X(i, j)$ (i.e., if there is no other element of $X(i, j)$ which both worker i and firm j prefer), then x_{ij} is Pareto optimal in $X(i, j)$ regardless of what other elements are contained in x_i^f or x_j^f . Thus it is meaningful to speak of a Pareto optimal subset $\hat{X}(i, j)$ of $X(i, j)$, independently of any specific outcome. This assumption can be stated formally as follows.⁶

⁵The set of stable outcomes is a subset of the core (defined by strict domination). An unstable outcome may be in the core as conventionally defined if the instability involves some but not all of the workers matched with a particular firm, since after the instability was resolved, any worker who had the same job description both before and after would be indifferent to the change. It should nevertheless be clear how such additional core points are unstable, if the rules allow firms to negotiate freely with individual workers. A fuller discussion of unstable points in the core of a game is given in Roth and Postlewaite [14].

⁶In the model of Kelso and Crawford [7] this assumption follows from the fact that, in that model, both firms and workers evaluate job descriptions in terms of a value parameterized by the salary associated with the job.

PARETO SEPARABILITY: For each i in W and j in F , the set $X(i, j)$ contains a Pareto optimal subset $\hat{X}(i, j) = \{\hat{x}_{ij}^1, \dots, \hat{x}_{ij}^p\}$. If x_i^f and y_i^f differ only in the job description of worker i at job j , and $\hat{x}_{ij}^q \in x_i^f$ and $\hat{x}_{ij}^r \in y_i^f$, then worker i prefers x_i^f to y_i^f if and only if $q > r$. If x_j^f and y_j^f differ only in the job description of worker i , and $\hat{x}_{ij}^q \in x_j^f$ and $\hat{x}_{ij}^r \in y_j^f$, then firm j prefers y_j^f to x_j^f if and only if $q > r$.

For an arbitrary job description \hat{x}_{ij}^s in $\hat{X}(i, j)$, we will sometimes refer to s as the *generalized salary* associated with the job description, since the worker prefers s to be high, and the firm prefers a low s . This is a “generalized” salary since it serves to parameterize the Pareto optimal elements of $\hat{X}(i, j)$, which may differ in dimensions other than actual salary. An outcome x^f will be called *pairwise Pareto optimal* if each x_{ij} in x^f and x_j^f is contained in $\hat{X}(i, j)$ for every i and j . Note that every stable outcome must be pairwise Pareto optimal, since otherwise it would be unstable via a worker-firm pair with a possible Pareto improvement.

The second assumption about preferences is that firms and workers regard each other more as substitutes than as complements, in the sense that if a worker, say, is a desirable employee at a given job description among a given group of co-workers, then he remains a desirable employee at that job description, even in a less desirable group of co-workers. Formally, the preferences of each agent k have the following property.⁷

SUBSTITUTABILITY: Let $x_k^f = C_k(x_k^f \cup y_k^g)$. Then for any x_{ij} in x_k^f , x_{ij} is contained in $C_k(y_k^g \cup \{x_{ij}\})$.

Note that this assumption about preferences comes into play only under very stringent conditions: agent k 's preference must be such that, given the chance, it would choose not to supplement x_k^f with any element of y_k^g . It is nevertheless quite a strong assumption, since it rules out even such complementarity as can arise through linkage in a budget constraint. However in the absence of some such assumption, the set of stable outcomes can be empty (cf. Kelso and Crawford [7]). Note also that preferences over sets of job descriptions have been defined entirely in terms of the job description(s) assigned to an agent. This involves an assumption that workers are indifferent to who their co-workers might be, and firms are indifferent to whether their employees moonlight at other jobs.⁸

⁷This property is a translation to the present model of the property introduced in Kelso and Crawford [7] called the gross substitutes assumption. Preferences which possess this property will be called substitutable preferences.

⁸Of course, this last assumption may be a good approximation even in situations where workers, say, may be highly sensitive as to the identity of their co-workers. If only a small part of the workforce is on the labor market at any given moment, then those workers who are not on the market may be regarded as a part of the firm at which they currently work. Thus, for example, a college professor might regard the faculty of various universities as relatively fixed when determining his preferences for different universities, in which case his preferences could be modeled without violating this assumption, even though his preferences for universities are influenced by their faculties.

3. POLARIZATION OF INTERESTS

In this section, all agents will be assumed to have substitutable, Pareto separable preferences. Before going on to the main results, it will be useful to establish the following two technical lemmas. The first of these is a direct consequence of the substitutability of the preferences.

LEMMA 1: *Let $x_{ij} \in C_k(x_k^f \cup y_k^g)$. Then $x_{ij} \in C_k(y_k^g \cup \{x_{ij}\})$.*

PROOF: Since the choice function C_k arises from a binary preference relation, $C_k(x_k^f \cup y_k^g) = C_k(C_k(x_k^f \cup y_k^g) \cup y_k^g)$. Using this more cumbersome expression, the lemma now follows directly from the definition of substitutability.

The second lemma states that if an individually rational, pairwise Pareto optimal outcome is not stable, the instability can always be traced to a coalition consisting of a single worker-firm pair.

LEMMA 2: *Pairwise Instability: An individually rational outcome x^f , such that $x_k^f = C_k(x_k^f)$ for all agents k , is stable if and only if no worker-firm pair (i, j) exists such that, for some job description \hat{x}_{ij}^g in $\hat{X}(i, j)$,*

$$\hat{x}_{ij}^g \in C_i(x_i^f \cup \{\hat{x}_{ij}^g\})P_i x_i^f$$

and

$$\hat{x}_{ij}^g \in C_j(x_j^f \cup \{\hat{x}_{ij}^g\})P_j x_j^f.$$

PROOF: If such an i, j and \hat{x}_{ij}^g exist, then x^f is unstable by definition (via $S = \{i\}$, $S' = \{j\}$, and $y_i^g = y_j^g = \{\hat{x}_{ij}^g\}$). If x^f is not pairwise Pareto optimal, then such an i, j , and \hat{x}_{ij}^g exist by definition. Conversely, suppose x^f is pairwise Pareto optimal, but unstable via a coalition S of workers and S' of firms and outcome y^g such that $g(S) = S'$, $g(S') = S$, and $y_k^g \subset C_k(x_k^f \cup y_k^g)P_k x_k^f$ for all k in $S \cup S'$. Since x^f is pairwise Pareto optimal, it follows that $g(k)$ is disjoint from $f(k)$ for all k in $S \cup S'$ (i.e., y^g doesn't have any worker-firm pairs in common with x^f). So let $y_i^g \cap y_j^g = x_{ij}$ for some i in S and j in S' . We may assume that $x_{ij} = \hat{x}_{ij}^g$ is Pareto optimal in $X(i, j)$ (otherwise let \hat{x}_{ij}^g be Pareto superior to x_{ij}). Then the fact that i and j have substitutable preferences implies via Lemma 1 that x^f is unstable via i, j and \hat{x}_{ij}^g , which completes the proof.

The following theorem establishes that the set of stable outcomes is nonempty for every configuration of agents' preferences.

THEOREM 1: *The set of stable outcomes is always nonempty.*

The proof will be constructive, by means of the following algorithm.⁹ Following the statement of the algorithm, it will be analyzed by means of five propositions, the last of which completes the proof of the theorem.

ALGORITHM: *Step 1* (a) Each firm j proposes its most preferred set $x_j(1)$ in $\bar{Y}(j)$.

(b) Each worker i accepts his choice set from u_i and the set $x_i(1) = \{x_{ij} \in x_j(1) \text{ for some } j\}$ of alternatives proposed to him, and rejects the rest.

⋮

Step k (a) Each firm j proposes its most preferred set $x_j(k)$ in $\bar{Y}(j)$ with the property that no element x_{ij} in $x_j(k)$ has been rejected at an earlier step.

(b) Each worker i accepts his choice set from the set of alternatives not yet rejected at step $k - 1$ together with those proposed in step k , and rejects the rest.

⋮

The algorithm terminates at any step t in which no rejections are issued, and results in the outcome x^f which matches each worker i with its current choice set $x_i^f = C_i(x_i(t) \cup \{u_i\})$ and each firm j with its latest proposal $x_j^f = x_j(t)$.

Analysis of the Algorithm

PROPOSITION 1: *Increasing generalized salaries: Firm j includes a job description $\hat{x}_{ij}^{s+1} \in \hat{X}(i, j)$ in its proposal $x_j(k)$ at step k only if \hat{x}_{ij}^s has been rejected at an earlier step.*

PROOF: The proposition follows from the assumption of Pareto separability, together with the requirement that each firm propose, at each step, its most preferred set of alternatives none of which have yet been rejected. A proposal $x_j(k)$ containing \hat{x}_{ij}^{s+1} would not meet this requirement unless \hat{x}_{ij}^s has previously been rejected.

PROPOSITION 2: *Offers remain open: For every firm j , if x_{ij} is contained in $x_j(k - 1)$ and is not rejected by worker i in step $k - 1$, then x_{ij} is contained in $x_j(k)$ also. Equivalently, $C_i(x_i(k - 1) \cup \{u_i\})$ is a subset of $x_i(k) \cup \{u_i\}$ for each worker i .*

PROOF: Note that $x_j(k - 1) = C_j(x_j(k - 1) \cup x_j(k))$, since $x_j(k - 1)$ is firm j 's choice set from all those feasible sets whose elements have not been rejected prior

⁹This algorithm, and its treatment here, have a strong family resemblance to those used by Gale and Shapley [5], Crawford and Knoer [2], Kelso and Crawford [7], and Jones [6]. Another kind of existence theorem for some related problems is found in a recent paper by Quinzii [9], in which a nonconstructive proof of the nonemptiness of the core is given for a general class of nontransferable utility games which includes both marriage markets and exchange economies with indivisibilities of the kind studied in Shapley and Scarf [15], Roth and Postlewaite [14], and Roth [12].

to step $k - 1$, while $x_j(k)$ is the choice set from the smaller class of feasible sets whose elements have not been rejected prior to step k . So substitutability implies that, if $x_{ij} \in x_j(k - 1)$, then $x_{ij} \in C_j(x_j(k) \cup \{x_{ij}\})$. So if x_{ij} in $x_j(k - 1)$ is not rejected at step $k - 1$, it must be contained in $x_j(k)$, since otherwise $C_j(x_j(k) \cup \{x_{ij}\}) P_j x_j(k)$, violating the requirement that $x_j(k)$ is the most preferred set whose elements haven't yet been rejected.

PROPOSITION 3: *Rejections are final: If x_{ij} is rejected at step k (i.e., $x_{ij} \in x_i(k)$ but $x_{ij} \notin C_i(x_i(k) \cup \{u_i\})$) then for any $p \geq k$, $x_{ij} \notin C_i(x_i(p) \cup \{u_i\} \cup \{x_{ij}\})$.*

PROOF: Suppose the proposition is false, and let $p > k$ be the *first* step at which $x_{ij} \in C_i(x_i(p) \cup \{u_i\} \cup \{x_{ij}\})$. Since $x_i(p) \cup \{u_i\}$ contains $C_i(x_i(p - 1) \cup \{u_i\})$ by Proposition 2, substitutability implies $x_{ij} \in C_i(C_i(x_i(p - 1) \cup \{u_i\}) \cup \{x_{ij}\})$, and thus $x_{ij} \in C_i(x_i(p - 1) \cup \{u_i\} \cup \{x_{ij}\})$, which contradicts the definition of p and completes the proof of the proposition.

PROPOSITION 4: *Termination: The algorithm terminates at some (finite) step t , and the final outcome x^f is individually rational and pairwise Pareto optimal.*

PROOF: Since no worker rejects the same job description more than once, the finiteness of the sets $X(i, j)$ of job descriptions insures termination in a finite number of steps. Individual rationality follows from the fact that for each agent k , x_k^f is a choice set from which u_k could have been chosen (and indeed, x_k^f can equal u_k). Pairwise Pareto optimality follows from the fact that a firm j can only propose a Pareto dominated outcome x_{ij} in $X(i, j)$ after the outcome \hat{x}_{ij}^q which Pareto dominates it has been rejected. But if i rejected \hat{x}_{ij}^q then i will also reject x_{ij} , since workers' choice sets become no less desirable as the algorithm progresses. So only pairwise Pareto optimal outcomes are ever accepted.

PROPOSITION 5: *Stability: The outcome x^f resulting from the algorithm is stable.*

PROOF: Suppose there exists a worker i , a firm j , and a job description \hat{x}_{ij}^q such that $\hat{x}_{ij}^q \in C_j(x_i^f \cup \{\hat{x}_{ij}^q\}) P_j x_i^f$. Then firm j must have proposed \hat{x}_{ij}^q to worker i and had it rejected as some step $k < t$. So $\hat{x}_{ij}^q \notin C_i(x_i^f \cup \{\hat{x}_{ij}^q\})$, and so (by Lemma 2) x^f cannot be unstable.

This completes the proof of Theorem 1.

Note that although firms and workers are treated symmetrically in this model, the above algorithm treats them asymmetrically, since firms propose, and workers accept or reject. The outcome produced by the algorithm is shown below to be the firm-optimal stable outcome. The corresponding worker-optimal stable outcome results from the corresponding algorithm, in which workers propose and firms accept or reject.

THEOREM 2: *There exist firm-optimal and worker-optimal stable outcomes. Specifically, the stable outcome x^f produced by the above algorithm is firm-optimal in the sense that, for each firm j and every stable outcome y^g , $x_j^f = C_j(x_j^f \cup y_j^g)$.*

Note that the theorem claims more than that there exists a stable outcome which all firms like at least as well as any other: it makes the additional claim that $x_j^f = C_j(x_j^f \cup y_j^g)$ for all firms j and stable outcomes y^g . This means that if firms were free to choose from the union of job descriptions they are assigned at x^f and y^g , no firm would want to supplement its assignment at x^f . It is not apparent that the second claim should follow from the first.

PROOF OF THEOREM 2: Call x_{ij} in $X(i, j)$ *possible* if there exists a stable outcome y^g such that $x_{ij} \in y_i^g \cap y_j^g$. That is, a job description x_{ij} is possible if there is some stable outcome at which worker i is employed by firm j at job description x_{ij} . To prove the theorem, it will be sufficient to establish the claim that in the course of the algorithm described above, no possible alternative x_{ij} is ever rejected by any worker i . Because of the fact that firms propose at each step their most preferred set among the alternatives that have not yet been rejected, this is equivalent to showing that the stable outcome produced by the algorithm assigns to each firm j its choice set from the set of possible job descriptions.

The proof is by induction, with the inductive assumption that no possible job description has been rejected up to step $r - 1$ of the algorithm. We show that no possible job description is rejected at step r .

Let $x(r)$ be the $(m + n)$ -vector whose components are the set of alternatives $x_k(r)$ proposed to each worker k in step r , and proposed by each firm k . Let x_{ij} be rejected in part (b) of step r ; that is $x_{ij} \in x_i(r) \cap x_j(r)$, but $x_{ij} \notin C_i(x_i(r) \cup \{u_i\})$.

Suppose that, contrary to the claim, x_{ij} is in fact possible. Then there exists a stable outcome y^g such that $x_{ij} \in y_i^g \cap y_j^g$. The numbered paragraphs that follow each begin with a statement to be proved, and follow with its proof.

(1) The above assumptions imply that $C_i(x_i(r) \cup \{u_i\}) = C_i(x_i(r))$. Otherwise, $C_i(x_i(r) \cup \{u_i\}) = u_i$, and the individual rationality of y^g then implies $y_i^g = C_i(y_i^g \cup C_i(x_i(r) \cup \{u_i\}))$, which by substitutability implies $x_{ij} \in C_i(C_i(x_i(r) \cup \{u_i\}) \cup \{x_{ij}\})$, which contradicts the assumption that $x_{ij} \notin C_i(x_i(r) \cup \{u_i\})$. (This last follows from the fact that C_i is a choice function, so if $C_i(S) = x$ and $x \subset T \subset S$, then $C_i(T) = x$.)

(2) There must be a firm h and job description x_{ih} such that $x_{ih} \in C_i(x_i(r))$ and $x_{ih} \notin y_i^g$. Otherwise, $C_i(x_i(r))$ is a strict subset of y_i^g , so $C_i(y_i^g \cup C_i(x_i(r))) = y_i^g$, and substitutability implies $x_{ij} \in C_i(C_i(x_i(r) \cup \{x_{ij}\}))$, which contradicts the assumption that $x_{ij} \notin C_i(x_i(r))$, by the argument at the close of the previous paragraph. Let $H = \{h \in F \mid x_{ih} \in C_i(x_i(r)) \text{ and } x_{ih} \notin y_i^g\}$ be the nonempty set of such firms.

(3) For each firm h in H , $x_{ih} \in C_h(y_h^g \cup \{x_{ih}\})$. To see this, note first that $C_h(x_h(r) \cup y_h^g) = x_h(r)$, since $x_h(r)$ is h 's choice set from the set of alternatives

not rejected prior to step r , and since no element of y^g can have been rejected prior to step r , by the inductive assumption and the stability of y^g . Substitutability thus implies $x_{ih} \in C_h(y_h^g \cup \{x_{ih}\})$.

(4) There exists a firm h in H such that $x_{ih} \in C_i(y_i^g \cup \{x_{ih}\})$. If not, then $y_i^g = C_i(y_i^g \cup C_i(x_i(r)))$, so substitutability implies $x_{ij} \in C_i(C_i(x_i(r)) \cup \{x_{ij}\})$, which contradicts the assumption that $x_{ij} \notin C_i(x_i(r))$, as in paragraph (1).

But paragraphs (3) and (4) establish that y^g is unstable via worker i and the firm h and job-description x_{ih} identified in paragraph (4). This contradicts the assumption that x_{ij} is possible, and completes the proof.

The next theorem establishes that the optimal stable outcome for one side of the market is the worst stable outcome for the other side of the market.

THEOREM 3: *Every firm likes any stable outcome at least as well as the worker-optimal stable outcome, and every worker likes any stable outcome at least as well as the firm optimal stable outcome.*

PROOF: By symmetry, it is sufficient to prove the second part of the theorem: if x^f is the firm-optimal stable outcome, then for any other stable outcome y^g , $y_i^g R_i x_i^f$ for every worker i . Recall that $x_j^f = C_j(x_j^f \cup y_j^g)$ for every firm j and stable y^g .

Suppose the theorem is false: then there exists a stable outcome y^g and a worker i such that $x_i^f P_i y_i^g$. Thus $C_i(x_i^f \cup y_i^g) P_i y_i^g$, so there exists some firm j in $f(i)$ for which some x_{ij} in $X(i, j)$ is contained in $C_i(x_i^f \cup y_i^g)$ and $x_i^f \cap x_j^f$ but not in y_i^g or y_j^g . Substitutability implies $x_{ij} \in C_j(y_j^g \cup \{x_{ij}\})$, and Lemma 1 implies $x_{ij} \in C_i(y_i^g \cup \{x_{ij}\})$. Since preferences are strict, this implies y^g is unstable with respect to i, j , and x_{ij} . This contradiction completes the proof.

These results show that the polarization of interests that occurs in the matching models referred to in the introduction does not depend on many of the special assumptions made in those models. This makes it all the more surprising that the literature contains virtually no comparable results concerning polarization of interests in the core or in the set of equilibria of other kinds of markets. The cause and extent of the polarization of interests in bilateral matching models thus remains something of a mystery. An outline of what is known about the detailed conflict and coincidence of workers' and firms' interests throughout the set of stable outcomes (and not just at the firm and worker-optimal outcomes) is contained in Roth [13], but many open questions remain.

University of Pittsburgh

Manuscript received September, 1982; revision received March, 1983.

REFERENCES

- [1] BERGSTROM, THEODORE, AND RICHARD MANNING: "Can Courtship be Cheatproof?" mimeo (personal communication), 1982.
- [2] CRAWFORD, VINCENT P., AND ELSIE MARIE KNOER: "Job Matching with Heterogeneous Firms and Workers," *Econometrica*, 49(1981), 437-450.
- [3] DEMANGE, GABRIELLE: "Strategyproofness in the Assignment Market Game," mimeo, Laboratoire D'Econometrie De L'Ecole Polytechnique, Paris, 1982.
- [4] DUBINS, L. E., AND D. A. FREEDMAN: "Machiavelli and the Gale-Shapley Algorithm," *American Mathematical Monthly*, 88(1981), 485-494.
- [5] GALE, DAVID, AND LLOYD SHAPLEY: "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, 69(1962), 9-15.
- [6] JONES, PHILIP C.: "A Polynomial Time Market Mechanism," mimeo, Department of Industrial Engineering, Northwestern University, Evanston, Illinois, 1982.
- [7] KELSO, A. S., JR., AND V. P. CRAWFORD: "Job Matching, Coalition Formation, and Gross Substitutes," *Econometrica*, 50(1982), 1483-1504.
- [8] KNUTH, DONALD E.: *Marriages Stables*. Montreal: Les Presses de l'Universite de Montreal: Montreal, 1976.
- [9] QUINZII, MARTINE: "Core and Competitive Equilibria with Indivisibilities," mimeo, Laboratoire D'Econometrie De L'Ecole Polytechnique, Paris, 1982.
- [10] RITZ, ZVI: "Incentives and Stability in Some Two-Sided Economic and Social Models," mimeo, Department of Business Administration, University of Illinois, Urbana-Champaign, 1982.
- [11] ROTH, ALVIN E.: "The Economics of Matching: Stability and Incentives," *Mathematics of Operations Research*, 7(1982), 617-628.
- [12] ———: "Incentive Compatibility in a Market with Indivisible Goods," *Economics Letters*, 9(1982), 127-132.
- [13] ———: "Conflict and Coincidence of Interest in Job Matching: Some New Results and Open Questions," mimeo, Department of Economics, University of Pittsburgh, 1982.
- [14] ROTH, ALVIN E., AND ANDREW POSTLEWAITE: "Weak Versus Strong Domination in a Market with Indivisible Goods," *Journal of Mathematical Economics*, 4(1977), 131-137.
- [15] SHAPLEY, LLOYD S., AND HERBERT SCARF: "On Cores and Indivisibility," *Journal of Mathematical Economics*, 1(1974), 23-28.
- [16] SHAPLEY, LLOYD S., AND MARTIN SHUBIK: "The Assignment Game I: The Core," *International Journal of Game Theory*, 1(1972), 111-130.
- [17] THOMPSON, GERALD L.: "Computing the Core of a Market Game," in *Extremal Methods and Systems Analysis, Lecture Notes in Economics and Mathematical Systems #174*, ed. by A. V. Fiacco and K. O. Kortanek. Berlin: Springer, 1980, pp. 312-334.

