MATHEMATICS OF OPERATIONS RESEARCH
Vol 10, No. 3, August 1985
Printed in U.S.A.

CONFLICT AND COINCIDENCE OF INTEREST IN JOB MATCHING: SOME NEW RESULTS AND OPEN QUESTIONS*†

ALVIN E. ROTH

University of Pittsburgh

The game-theoretic solution to certain job assignment problems allows the interests of agents on the same side of the market (e.g. firms or workers) to be simultaneously maximized. This is shown to follow from the lattice structure of the set of stable outcomes. However it is shown that in a more general class of problems the optimality results persist, but the lattice structures do not. Thus this paper raises as many questions as it answers.

1. Introduction. Game theory concerns situations involving conflicting interests of multiple decision makers. Because multiple objectives cannot in general be simultaneously maximized, the optimization results found in other areas of operations research are largely absent from game theory, and the theoretical analysis of games focuses not on identifying "optimal" outcomes, but primarily on identifying outcomes that are in some appropriate sense stable. It is therefore noteworthy that in an important class of games associated with two-sided markets, the differing objectives of (competing) agents on the same side of the market can be simultaneously maximized on the set of stable outcomes. In such market games, therefore, an optimal stable outcome can be identified for each side of the market. This kind of optimization has not been observed in any other nontrivial class of games.

This paper explores the underlying structure which allows such optimization to occur. It will be shown that, in all but one of the models in which such optimal stable outcomes have been shown to exist, this optimality is closely related to the algebraic structure of the set of stable outcomes. Specifically, in these models the preferences of the agents impose in a natural way a lattice structure on the set of stable outcomes, and the optimal stable outcomes for the two sides of the market are the maximum and minimum elements of the lattice. However in the remaining model, which is also the most general, it will be shown that this natural lattice structure no longer occurs. There thus remains an open problem: to identify some property of these models which explains the existence of optimal stable outcomes in the most general case.

2. Overview of the literature. The first model of the kind considered here was introduced by Gale and Shapley (1962), who studied two-sided "marriage markets" consisting 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.

Viewed as an institution designed to achieve a mutually satisfactory matching of firms and workers, any market of this type serves to advance the common interests

*Received April 5, 1983; revised March 7, 1984.
AMS 1980 subject classification. Primary: 90D40. Secondary: 90D45.
OR/MS Index 1978 subject classification. Primary: 231 Games/group decisions.
Key Words. Job assignment, job matching, market games.
†This work has been supported by grants from the National Science Foundation and the Office of Naval Research, and by Fellowships from the John Simon Guggenheim Memorial Foundation and the Alfred P. Sloan Foundation. It has also benefitted from stimulating conversation and correspondence with Lou Billera, Charles Blair, and Zvi Ritz.

0364-765X/85/1003/0379$01.25

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 that this overall pattern of common and conflicting interests is reversed when attention is confined to the set of 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.) Specifically, when agents in the simple symmetric model studied by Gale and Shapley are not indifferent between potential matches, there exists 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.[1] In addition, it is straightforward to show in this model that the optimal core outcome for one side of the market is the worst outcome in the core for every agent on the other side of the market (cf. Knuth 1976).[2]

Firm and worker-optimal stable outcomes have also been shown to exist in a number of related models. Shapley and Shubik (1972) and Thompson (1980) considered transferable-utility models in which all outcomes can be evaluated in terms of monetary payoffs, and Crawford and Knoer (1981) 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.

Kelso and Crawford (1982) 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 that is simultaneously the best core outcome for every firm. Since firms and workers are not symmetric in this model, Kelso and Crawford's treatment left open the question of whether there exists a dual core outcome which is best for all workers.

One result demonstrated in Roth (1984) was that such a worker-optimal outcome does indeed exist, when agents have suitably strict preferences. It was shown that there are firm-optimal and worker-optimal stable outcomes in a more general, completely symmetric model, in which workers may take more than one job, and each worker has preferences over sets of jobs.

The principal purpose of this paper is to explore the reasons behind, and the precise extent of, the surprising pattern of common and conflicting interests exhibited by the various models discussed above. To this end we shall consider three increasingly general models of job matching.

The results for these three models can be briefly described as follows. The first

[1] 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 an achievable match for any of the other men, in a competitive marriage market whose outcomes are stable.)

[2] These 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. See Roth (1982a) for a discussion of these matters.

model considered is a straightforward generalization of the model of Crawford and Knoer.[3] It will be shown that, in this model, choices between different stable outcomes turn out to have a remarkable "consensus" property. If all the firms, say, are allowed to choose their most preferred workers from the workers assigned to them at two different stable outcomes, then no worker is selected by two different firms, and so the choice results in a feasible outcome, which is itself stable. This consensus property directly implies the existence of firm-optimal and worker-optimal outcomes, and imparts to the set of stable outcomes a lattice structure, in which the common interests of the firms turn out to be directly opposed to the common interests of the workers, in the sense that all firms prefer one stable outcome to another if and only if all workers have the opposite preference. This opposition of interests in turn implies that the optimal stable outcome for one side of the market will be the worst stable outcome for every agent on the other side of the market. (These results extend the similar results for Gale and Shapley's marriage market reported in Knuth (1976). The same phenomenon occurs in the transferable utility model of Shapley and Shubik (1972).)

The second model considered is a straightforward generalization of the asymmetric model of Kelso and Crawford. The consensus property of the previous model continues to hold for choices made by firms, and this implies the existence of the firm-optimal stable outcome identified by Kelso and Crawford.

In further investigating the second model, however, we will see that the consensus property fails to hold for choices made by workers, and so cannot be used to explain why a worker-optimal stable outcome nevertheless exists. The common interests of the firms are no longer opposed to those of the workers throughout the stable set in this model, and so this can no longer be used to explain why, nevertheless, the optimal stable outcome for one side of the market still turns out to be the worst for every agent on the other side of the market. Enough of the structure found in the first model remains, however, so that it might appear that only a few pieces are missing from a complete explanation of the polarization of interests in this case also.

However in a version of the symmetric model of Roth (1984b), it will be shown that no vestige remains of the elaborate structure observed in the previous two models, but firm- and worker-optimal stable outcomes nevertheless exist, and it continues to be the case that the optimal stable outcome for one side of the market is the worst stable outcome for agents on the other side.

3. **Three models of job matching.** In each of the models considered here there will be a bipartite graph (N, X) whose set N of nodes is the union of a set W (of workers) indexed by $i = 1, \ldots, m$ and a set F (of firms) indexed by $j = 1, \ldots, n$. The finite set of arcs $X(i,j) = \{x_{ij}^1, \ldots, x_{ij}^{p(i,j)}\}$ connecting worker i and firm j represents the job descriptions[4] at which i might be employed by j, and the set X of arcs is the union over all (i,j) in $W \times F$ of $X(i,j)$. Denote by $X(i) = \bigcup_{j \in F} X(i,j)$ and $X(j) = \bigcup_{i \in W} X(i,j)$ the arcs incident to a node $i \in W$ or $j \in F$, respectively. The three models to be considered differ in what constitutes a feasible outcome.

In Model I, a feasible outcome is a subset of f of X such that for every k in N, $|f \cap X(k)| < 1$; i.e. every agent is assigned at most one job.[5] For any agent k and

[3] The first of these models is also studied in Ritz (1982), in a form similar to that studied here.

[4] For an arbitrary job description x_{ij}^s in $X(i,j)$, we will sometimes refer to s as the *generalized salary* associated with the job description, since the worker will later be assumed to prefer s to be high, and the firm will later be assumed to prefer a low s. This is a "generalized" salary since it serves to parameterize the elements of $X(i,j)$, which may differ in dimensions other than actual salary.

[5] It would not change the results if every agent k could be assigned some number q_k of jobs different from 1, so long as q_k is fixed and preferences over jobs are separable in the sense that the desirability of a job is not influenced by the set of jobs already assigned. However this other model is not equivalent to the one studied here: see Roth (1985a).

feasible outcome f, the feasible assignment to k at f is $f_k = f \cap X(k)$, the job description assigned to agent k. (If $f_k = \emptyset$ then agent k is unmatched at the outcome f.) The set of possible feasible assignments for an agent k in N is thus $A(k) = \{f_k \mid f_k \subset X(k) \text{ and } |f_k| = 1, \text{ or } f_k = \emptyset\}$.

In Model II, a feasible outcome is a subset f of X such that for every i in W, $|f \cap X(i)| \leq 1$; i.e. every worker is assigned at most one job, but firms may be assigned sets of workers. The feasible assignment at f for an agent k is $f_k = f \cap X(k)$, so the set of feasible assignments for a worker i in W is $A(i) = \{f_i \mid f_i \subset X(i) \text{ and } |f_i| = 1, \text{ or } f_i = \emptyset\}$ as in Model I, while the set of feasible assignments for a firm j in F is $A(j) = \{f_j \subset X(j) \mid |f_j \cap X(i,j)| \leq 1 \text{ for all } i \in W\}$. That is, a feasible assignment for a firm j is a set of job descriptions, one for each worker employed by the firm.[6]

In Model III, a feasible outcome is a subset f of X such that for all (i, j) in $W \times F$, $|f \cap X(i,j)| \leq 1$; i.e. workers may have more than one job and firms more than one employee, but no worker has more than one job description at any firm. For any agent k the feasible assignment at f is $f_k = f \cap X(k)$, so the set of feasible assignments for a worker i is $A(i) = \{f_i \subset X(i) \mid |f_i \cap X(i,j)| \leq 1 \text{ for all } j \in F\}$ and for a firm j, $A(j) = \{f_j \subset X(j) \mid |f_j \cap X(i,j)| \leq 1 \text{ for all } i \in W\}$.

In each of the models, every agent k has a strict preference—i.e. a rank-ordering—over his (finite) set $A(k)$ of feasible assignments.[7] For any elements f_k and g_k of $A(k)$, let $f_k P_k g_k$ denote that agent k prefers f_k to g_k, and for any subset S of $X(k)$, let $C_k(S)$ denote agent k's most preferred element of $A(k)$ that is a subset of S, i.e. $C_k(S)$ is agent k's most preferred choice from among all feasible subsets of S.

A feasible outcome f is defined to be *stable* if $f_k = C_k(f_k)$ for all k in N, and if no worker-firm pair (i, j) and job description $x_{ij} \in X(i,j)$ exist such that

(i) $x_{ij} \in C_i(f_i \cup \{x_{ij}\})$ and $C_i(f_i \cup \{x_{ij}\}) P_i f_i$; and

(ii) $x_{ij} \in C_j(f_j \cup \{x_{ij}\})$ and $C_j(f_j \cup \{x_{ij}\}) P_j f_j$.

The requirement that $f_k = C_k(f_k)$ implies that no agent can improve his assignment by declining some job description assigned to him, while (i) and (ii) imply that no worker i and firm j can agree on a job description that would improve both their assignments.[8]

Without further conditions on agents' preferences, the set of stable outcomes can sometimes be empty in Models II and III (but not in Model I). In what follows, we will therefore assume that agents' preferences obey the following two conditions.[9]

Generalized salary condition. For any (i, j) in $W \times F$ let f and g be feasible outcomes such that f_i differs from g_i and f_j differs from g_j only in the job description of worker i at firm j, with $x_{ij}^q \in f_i \cap f_j$ and $x_{ij}^r \in g_i \cap g_j$. Then $f_i P_i g_i$ and $g_j P_j f_j$ if and only if $q > r$.

Note that this condition is satisfied trivially in any example in which there is only one job description at which any worker-firm pair can be matched, i.e. if $X(i, j)$ contains at most one element for any i and j. The polarization of interests discussed here, which was first observed in such examples, therefore does not depend on this condition, which serves to make the set of Pareto optimal contracts between a given firm and worker independent of the job assignments agreed to by the firm and other workers, or the worker and other firms.

[6] There is no loss of generality in assuming that a worker can have no more than one job description at a given firm, since the set of job descriptions can be enlarged to include any feasible combination of jobs as a single job description.

[7] An agent's preferences are called strict if he is not indifferent between any two distinct assignments.

[8] In Roth (1985b) the precise relationship of this definition of stability to the core is discussed for a closely related model.

[9] These conditions were introduced in the model of Kelso and Crawford (1982).

Substitutability condition[10]. For any k in N, let $f_k = C_k(f_k \cup g_k)$. Then any x_{ij} in f_k is contained in $C_k(g_k \cup \{x_{ij}\})$.

This latter assumption states that workers and firms regard each other more as substitutes than as complements, in the sense that if a worker, say, is a desirable employee at job description x_{ij} in a firm's feasible assignment f_j, then he remains a desirable employee at that job description, even in a less desirable group of co-workers. Note that this condition is trivially satisfied in Model I and by workers' preferences in Model II, since feasible assignments and choice sets there are always singletons. Substitutability has the following immediate implication.

LEMMA 1. *For substitutable preferences, let* $x_{ij} \in C_j(f_j \cup g_j)$. *Then* $x_{ij} \in C_j(g_j \cup \{x_{ij}\})$.

PROOF. Since the choice function C_j arises from a binary preference relation, $C_j(f_j \cup g_j) = C_j(C_j(f_j \cup g_j) \cup g_j)$. Using this more cumbersome expression, the lemma now follows directly from the definition of substitutability.

4. The polarization of interests. The following theorem sets the stage for the subsequent investigation of the detailed structure of the set of stable outcomes.

THEOREM 1 (Roth 1984). *In Model* III,

(a) *The set of stable outcomes is nonempty.*

(b) *There exist a firm-optimal stable outcome which every firm likes at least as well as any stable outcome, and a worker-optimal stable outcome which every worker likes as well as any stable outcome.*[11]

(c) *The optimal stable outcome for one side of the market is the worst stable outcome for every agent on the other side of the market.*

Since Model III generalizes Model II which in turn generalizes Model I, Theorem 1 applies to all three models.[12] However we will see that the structure of the set of stable outcomes does not generalize from Model I to III in a straightforward way.

4.1. *Polarization of interests in Model* I. In the introduction, we referred to a consensus property for choices between different stable outcomes. This can be formally stated as follows.

THEOREM 2 (Consensus property for firms). *In Model* I *let f and g be stable outcomes. There is a feasible outcome h defined by $h_j = C_j(f_j \cup g_j)$ for all j in F, and $h_i = h_j$ for every i in W such that $h_j \subset X(i, j)$ for some j in F, and $h_i = \emptyset$ for all other i in W. Furthermore, this outcome h is itself stable.*

This theorem will be proved in the analysis of Model II, where it also holds. (Note that the fact that h is feasible means that no two firms choose the same worker.) Theorem 2 and the symmetry between firms and workers in Model I immediately imply that there is a symmetric consensus property for workers.

[10] Preferences which possess this property will be called substitutable preferences.

[11] In Roth (1984a) firm and worker-optimal stable outcomes were shown to have a stronger property. Let f be, say, the firm-optimal stable outcome, and let g be any other stable outcome. Then for any firm j it follows not only that $f_j R_j g_j$, but also $f_j = C_j(f_j \cup g_j)$. So every firm j is so satisfied by its optimal stable assignment that it wouldn't choose to modify it by including any elements from its assignment at any other stable outcome.

[12] As presented here Model III is not a formal generalization of Models II and I, since the three models have different sets of feasible outcomes. However the analysis of the set of stable outcomes would be unaffected if the models were rewritten with a common set of feasible outcomes, but with agents in the first two models having preferences which make individually irrational those assignments that are presently infeasible. Certain other kinds of feasibility constraints can also be incorporated into the model in this way.

THEOREM 3 (Consensus property for workers). *In Model* I *let f and g be stable outcomes. There is a feasible stable outcome h defined by $h_i = C_i(f_i \cup g_i)$ for all i in W, $h_j = h_i$ for every j in F such that $h_i \subset X(i, j)$ for some i in W, and $h_j = \emptyset$ for all other j in F.*

Here the fact that h is feasible means that no two workers choose the same firm.

Given the nonemptiness of the set of stable outcomes,[13] these two theorems imply the existence of firm-optimal and worker-optimal stable outcomes, respectively. (That is, part (b) of Theorem 1 follows in Model I from Theorems 2 and 3.) Every firm, for example, likes the outcome h constructed in Theorem 2 at least as well as both f and g. So, since there can be only finitely many stable outcomes, the firm-optimal stable outcome can be constructed by an iterative process of applying Theorem 2 to its "output" h and any stable outcome which some firm prefers to h, until no further such outcomes can be found.

The next two theorems establish that the common interest of the firms is opposed to that of the workers throughout the set of stable outcomes.

THEOREM 4. *In Model* I *let f and g be stable outcomes. If $f_iP_ig_i$ or $f_i = g_i$ for all workers i, then $g_jP_jf_j$ or $g_j = f_j$ for all firms j.*

This theorem will be proved in the analysis of Model II. The symmetry between firms and workers in Model I immediately implies the converse of Theorem 4.

THEOREM 5. *In Model* I *let f and g be stable outcomes. If $g_jP_jf_j$ or $g_j = f_j$ for all firms j, then $f_iP_ig_i$ or $f_i = g_i$ for all workers i.*

Part (c) of Theorem 1 follows immediately in Model I from Theorems 4 and 5.

The common interests of each set of agents define a partial order on the set of stable outcomes (e.g., f is at least as great as g in the ordering defined by the common interests of the firms if and only if $f_jP_jg_j$ or $f_j = g_j$ for all firms j). The properties of the set of stable outcomes can be formulated in lattice-theoretic terms (cf. Birkhoff 1973).

THEOREM 6. *In Model* I,

(a) *The set of stable outcomes is a complete distributive lattice under the partial order of the firms' common interests, and also under the partial order of the workers' common interests.*

(b) *The lattice under one partial order is the dual of the lattice under the other partial order.*

(c) *In either lattice, for any two stable outcomes f and g, the greatest lower bound $f \wedge g$ and the least upper bound $f \vee g$ both give every agent on one side of the market his preferred choice of the two outcomes, and every agent on the other side of the market his less preferred choice.*

PROOF. Theorem 2 establishes that the set of stable outcomes is a join semilattice under the partial order of the firms, and characterizes the least upper bounds (joins), $h = f \vee g$ for any pair of stable outcomes f and g. Theorem 3 establishes the same result for the partial order of the workers. Theorems 4 and 5 establish that the two partial orders are duals, so the two semilattices are duals, and constitute a lattice under either partial order. The finiteness of the lattice insures that a greatest lower bound and a least upper bound exist for every set of stable outcomes, so the lattice is complete.

[13] In a recent paper by Quinzii (1984) a nonconstructive proof of the nonemptiness of the core is given for a general class of nonside-payment games which includes both marriage markets and exchange economies with indivisibilities of the kind studied in Shapley and Scarf (1974), Roth and Postlewaite (1977), and Roth (1982b).

To show that the lattice is distributive, it suffices to show $f \vee (g \wedge h) = (f \vee g) \wedge (f \vee h)$ for all stable outcomes f, g, h. But for any firm j, $(f \vee (g \wedge h))_j = (g \wedge h)_j$ if and only if $g_jP_jf_j$ *and* $h_jP_jf_j$. Otherwise the expression on the left equals f_j. Similarly $((f \vee g) \wedge (f \vee h))_j = (g \wedge h)_j$ if and only if $g_jP_jf_j$ *and* $h_jP_jf_j$, and otherwise the expression on the left equals f_j. So the lattice is distributive, since this is true for all firms j. This completes the proof.

Since Blair (1982) has shown that every distributive lattice corresponds to the set of stable outcomes for some marriage market, there is no more restrictive lattice structure that can be used to describe stable outcomes here.[14]

4.2. *Polarization of interests in Model* II. The consensus property for choices by firms between stable outcomes continues to hold for this model. Since that property does not continue to hold for Model III, it is proved below.

THEOREM 7 (Consensus property for firms). *In Model* II *let f and g be stable outcomes. There is a feasible outcome h defined by $h_j = C_j(f_j \cup g_j)$ for all j in F; $h_i = h_j \cap X(i,j)$ for every i in W such that $h_j \cap X(i,j)$ is nonempty for some j in F; and $h_i = \emptyset$ for all other i in W. Furthermore h is stable.*

PROOF OF THEOREM 7. First we show that h is a well-defined, feasible outcome. Suppose not: then two different firms must choose the same worker; i.e., there exist a worker i and different firms j and k such that h_j contains f_i and h_k contains g_i. Worker i must prefer one of these assignments to the other (since workers have strict preferences): without loss of generality, let $f_iP_ig_i$. But firms have strict preferences, so $h_jP_jg_j$ (since $g_j \neq h_j$), and these preferences are substitutable, so f_i is contained in $C_j(g_j \cup f_i)$ (via Lemma 1, since h_j contains f_i). Therefore g would be unstable via i, j and f_i, contrary to the assumption. So h must be well defined and feasible.

Next we show that h is stable. Since f and g are stable, h is individually rational,[15] and so $h_k = C_k(h_k)$ trivially for all workers k and by construction of h for all firms k. Suppose h were not stable. Then there is a worker i, a firm j, and a job description x_{ij} in $X(i,j)$ such that $x_{ij}P_ih_i$ and $x_{ij} \in C_j(h_j \cup \{x_{ij}\})P_jh_j$. (So x_{ij} is not an element of either f_j or g_j.) But h_i equals either f_i or g_i. Without loss of generality, let $h_i = g_i$. Denote $C_j(C_j(h_j \cup \{x_{ij}\}) \cup g_j)$ by v_j. Then $v_jP_jh_j$, and so v_j is not a subset of $h_j = C_j(h_j)$, so $x_{ij} \in v_j$. But $v_j = C_j(v_j \cup g_j)$, so substitutability implies $x_{ij} \in C_j(g_j \cup \{x_{ij}\})$. So g is unstable via i, j, and x_{ij}, which provides the contradiction needed to complete the proof.

However Theorem 3 does not extend to this model; i.e., the consensus property for choices by workers between stable outcomes which held in Model I no longer holds in Model II. As the following counterexample shows, the problem is not that different workers may choose the same firm, since this is not infeasible in this model. Instead, the problem is that the feasible outcome h resulting from worker choices between two stable outcomes may not itself be stable.

Counterexample II-1. No Consensus Property for Workers. Consider the case of six workers, $W = \{1,2,3,4,5,6\}$, and five firms, $F = \{1',2',3',4',5'\}$. There is exactly one feasible job description for each worker-firm pair, so $X(i,j) = \{(i,j)\}$ for all i in W and j in F, so we may regard workers' preferences as being defined over firms, and firms' preferences as being defined over sets of workers. It will be sufficient for our purposes to specify only the first two or three elements in each agent's preference ordering, which may be extended in any way consistent with the substitutability of the

[14] However when agents can be indifferent between distinct job descriptions they receive at stable outcomes, it is straightforward to construct an example showing that there need be no lattice structure, no consensus property, and no optimal stable outcome for either set of agents.

[15] That is, for every agent k, either $h_k = \emptyset$ or $h_kP_k\emptyset$.

TABLE 1

	Firm				
Outcome	$1'$	$2'$	$3'$	$4'$	$5'$
f	$\{1\}$	$\{2\}$	$\{3\}$	$\{4,6\}$	$\{5\}$
g	$\{4\}$	$\{1,3\}$	$\{2\}$	$\{5\}$	$\{6\}$
h		$\{1,3\}$	$\{2\}$	$\{4,6\}$	$\{5\}$

firms' preferences. The preferences of the workers are as follows:[16] $2'P_1 1'P_1 \dots$; $1'P_2 3'P_2 2'P_2 \dots$; $1'P_3 2'P_3 3'P_3 \dots$; $4'P_4 1'P_4 \dots$; $1'P_5 5'P_5 4'P_5 \dots$; $1'P_6 4'P_6 5'P_6 \dots$. The preferences of the firms are given by: $\{4\}P_{1'}\{1\}P_{1'}\{2,3,5,6\}P_{1'} \dots$; $\{2\}P_{2'}\{1,3\}P_{2'} \dots$; $\{3\}P_{3'}\{2\}P_{3'} \dots$; $\{5\}P_{4'}\{4,6\}P_{4'} \dots$; $\{6\}P_{5'}\{5\}P_{5'} \dots$.

Let f and g be the stable outcomes in which the workers matched to each firm are given by Table 1.

(That is, in our more general formal notation, $f_i = \{(i,i')\}$ for $i = 1, \dots, 5$, $f_6 = \{(6,4')\}$, $f_{j'} = \{(j,j')\}$ for $j = 1', 2', 3', 5'$, $f_{4'} = \{(4,4'),(6,4')\}$, etc.)

It is easily verified that f and g are both stable. The outcome h is defined by giving each worker i his choice between f and g: that is $h_i = C_i(f_i \cup g_i)$ for each i in W. Note that firm $1'$ is left unmatched at h (i.e., $h_{1'} = \emptyset$).

To see that the consensus property for workers' choices fails here, note that h is unstable, via workers $\{2,3,5,6\}$ and firm $1'$. Note also that, unlike the case in Model I, the outcome h produced by giving each worker i the assignment f_i or g_i which he prefers cannot be equivalently defined by giving each firm j the assignment f_j or g_j which he likes least; here that fails to produce even a feasible outcome, since firms $1'$ and $2'$ would both include worker 1 in their least preferred choice. (Note that in Counterexample II-1, the worker-optimal stable outcome is the outcome which gives every worker his first choice firm, and the firm-optimal stable outcome gives every firm its first choice assignment of workers.)

We can now consider the extent to which the opposition of the common interests of firms and workers, studied in Theorems 4 and 5, continues to hold in Model II. The phenomenon captured in Theorem 4 persists here: if all workers like one stable outcome at least as well as another, then firms have the reverse preference. This will be proved below, since it will no longer be so in Model III. However the phenomenon captured in Theorem 5 no longer persists: all firms may prefer one stable outcome to another, and some workers may agree. This will be demonstrated in Counterexample II-2.

THEOREM 9. *In Model* II *if f and g are stable outcomes such that $f_iP_ig_i$ or $f_i = g_i$ for all workers i, then $g_jP_jf_j$ or $g_j = f_j$ for all firms j.*

PROOF OF THEOREM 9. Suppose the theorem is false: then $f_jP_jg_j$ for some firm j. Denote by S the subset of workers i such that $f_i \subset C_j(f_j \cup g_j)$. If $f_i = g_i$ for all $i \in S$, then $C_j(f_j \cup g_j)$ is a strict subset of g_j, so g is unstable since $g_j \neq C_j(g_j)$. Otherwise, there is a worker i in S such that $f_i \neq g_i$, so $f_iP_ig_i$. Since firm j has substitutable preferences, Lemma 1 implies $f_i \subset C_j(g_j \cup f_i)$, so g is unstable via i, j and f_i. The contradiction completes the proof.

The following counterexample shows that Theorem 5 does not extend to this model.

Counterexample II-2. Consider the case of two workers, $W = \{1,2\}$, and one firm, $F = \{1'\}$. For workers $i = 1,2$, the set of feasible job descriptions is given by $X(i,1') = \{(i,\$s) \mid s = 1, \dots, 10\}$. The preferences of worker $i = 1,2$ are given by

[16] That is, worker 1 for example likes firm $2'$ most of all, followed by $1'$, and so on.

$(i,\$s)P_i(i,\$t)$ if and only if $s > t$; and $(i,\$1)P_i\emptyset$. So worker i prefers higher salaries to lower salaries, and employment to unemployment.

The firm's preference relation is given by

(a) $\{(1,\$s_1),(2,\$s_2)\}P_{1'}\{1,\$t_1),(2,\$t_2)\}$ if $s_1 + s_2 < t_1 + t_2$ or if $s_1 + s_2 = t_1 + t_2$ and $s_1 > t_1$,

(b) $\{(1,\$10),(2,\$10)\}P_{1'}\{(i,\$s)\}$ for any $(i,\$s)$ in $X(i,1')$, $i = 1,2$,

(c) $\{(1,\$s)\}P_{1'}\{2,\$t)\}$ if and only if $s \leqslant t$,

(d) $\{(2,\$10)\}P_{1'}\emptyset$.

So firm 1′ prefers to hire both workers, at as low a total payroll as possible (with a tie-breaking preference for paying worker 1 more than worker 2). If he can't hire both workers even at top dollar, he prefers to hire one worker at the lowest salary possible, with a tie-breaking preference for worker 1, and he prefers to hire one worker, even worker 2 at the highest feasible salary, to remaining unmatched.

It is clear that every outcome which matches both workers to the firm is stable. Let f be the stable outcome $f = \{(1,\$5),(2,\$5)\}$, which matches both workers to the firm at a salary of \$5 each. Let g be the stable outcome $g = \{(1,\$6),(2,\$1)\}$, which matches both workers to the firm at salaries of \$6 for worker 1 and \$1 for worker 2.

Then the only firm prefers g to f, and so does worker 1. That is, every firm prefers g to f, but it is not the case that every worker likes f at least as well as g. So the preferences of workers and firms are not opposed in the same way as in Model I. This completes the counterexample.

The properties of the set of stable outcomes can be formulated in lattice-theoretic terms as follows.

THEOREM 10. *In Model* II,

(i) *The set of stable outcomes is a complete lattice under the partial order of the firms' common interests.*

(ii) *For any two stable outcomes f and g, the least upper bound $f \vee g$ gives every firm its choice from its employees at f and g.*

PROOF. Theorem 7 establishes that the set of stable outcomes is a join semilattice under the partial order of the firms, with the required least upper bound. It is complete, since it is finite. To prove the theorem it will therefore be sufficient to demonstrate that every subset of stable outcomes has a greatest lower bound. Let S be such a subset, and let $L(S)$ be the set of stable outcomes g such that for all f in S, $f_jP_jg_j$ or $f_j = g_j$ for all firms j. (That is, $L(S)$ is the set of all lower bounds for S under the partial order of the firms.) By Theorem 1, $L(S)$ is nonempty, since it contains the worker-optimal stable outcome. Therefore, since the join semilattice is complete, $L(S)$ has a least upper bound h, which is therefore the greatest lower bound for S. This completes the proof.

As in Model I, the firm and worker-optimal stable outcomes in Model II are the maximum and minimum elements of a lattice ordered by the firms' common interests. However only parts of the lattice structure, and only parts of its explanatory power for Theorem 1, have survived the generalization to Model II. These issues become even clearer in Model III, where workers as well as firms have substitutable preferences.

5. **Polarization of interests in Model III: Some open questions.** The consensus property which has supported the lattice structure observed in Models I and II is absent from both sides of the market in Model III. Since Model III generalizes Model II, it is immediate from Counterexample II-1 that the consensus property fails to hold for workers, and since (unlike Model II) Model III treats workers and firms symmetrically, the property also fails to hold for firms in Model III. Similarly it follows from Counterexample II-2 that the common interests of workers and firms are not in

general opposed throughout the set of stable outcomes, in either direction in Model III.

This leaves us without any natural explanation of why, in the most general of the matching models considered, firm and worker-optimal stable outcomes exist, and why the optimal stable outcome for one side of the market is the worst stable outcome for all agents on the other side. That is, although the most general model can be shown to have these properties, only in the more specialized models can these properties be explained in terms of properties of the entire set of stable outcomes, which in turn reflect the preferences of the agents, which constitute the data of the model.

Two avenues of further investigation suggest themselves. The first would involve an effort to determine if any lattice properties similar to those studied here survive the generalization to Model III. Note well that what has been shown here is only that the consensus property and opposition of interests do not generalize to Model III. It remains an open question whether the set of stable outcomes might nevertheless always be a lattice, with some suitably defined meet and join. If, for any preferences of the agents, the set of stable outcomes is a lattice, and if the meet and join have a natural interpretation in terms of the choices of the agents (comparable to the interpretation allowed by the consensus property), then a general explanation of the existence of optimal stable outcomes could perhaps be constructed along roughly the same lines as for the more specialized cases.

Alternatively, it may be necessary to explore quite different kinds of structural properties of the set of stable outcomes. For example, the bipartite nature of the matching problem makes it possible to speculate that the set of stable outcomes might possess some matroid properties that would allow the existence of optimal stable outcomes to be explained in terms of the kind of optimization results associated with matroids.

One reason it seems important to understand the cause and extent of the polarization of interests among agents on different sides of a matching problem is that, when such polarization exists, the particular institutions used to resolve a given matching problem may have differential welfare implications for agents on opposite sides of the market. In Roth (1984a) I consider from this point of view the succession of institutional procedures by which graduating medical students in the United States have obtained employment by hospitals as interns and residents. The organization of that labor market underwent considerable turmoil for at least 30 years prior to 1951, at which time a procedure was instituted that employs an algorithm which turns out to yield the hospital-optimal stable outcome. With small modifications, that procedure continues to be used to organize the market for interns and residents today.

Note added in proof: A recent paper by Blair (1984) shows that, for a class of markets differing from Model III only in subtle respects, a lattice structure exists with respect to an appropriate partial order.

References

Birkhoff, Garrett. (1973). *Lattice Theory*. Amer. Math. Soc. Colloq. Publ., 25, Providence, R.I., 3rd Ed.

Blair, Charles. (1982). Every Finite Distributive Lattice is a Set of Stable Matchings. *J. Combin. Theory* (forthcoming).

———. (1984). Stable Matching with Multiple Partners and Lattice Structure. mimeo, University of Illinois.

Crawford, Vincent P. and Knoer, Elsie Marie. (1981). Job Matching with Heterogeneous Firms and Workers. *Econometrica* **49** 437–450.

Gale, David and Shapley, Lloyd. (1962). College Admissions and the Stability of Marriage. *Amer. Math. Monthly* **69** 9–15.

Kelso, A. S., Jr. and Crawford, V. P. (1982). Job Matching, Coalition Formation, and Gross Substitutes. *Econometrica* **50** 1483–1504.

Knuth, Donald E. (1976). *Marriages Stables*. Les Presses de L'Universite de Montreal, Montreal.

Quinzii, Martine. (1984). Core and Competitive Equilibria with Indivisibilities. *Internat. J. Game Theory* **13** 41–60.

Ritz, Zvi. (1982). Incentives and Stability in Some Two-sided Economic and Social Models. mimeo, Department of Business Administration, University of Illinois, Urbana-Champaign.

Roth, Alvin E. (1982a). The Economics of Matching: Stability and Incentives. *Math. Oper. Res.* **7** 617–628.

———. (1982b). Incentive Compatibility in a Market with Indivisible Goods. *Econom. Lett.* **9** 127–132.

———. (1984a). The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory. *J. Polit. Econ.* **92** 991–1016.

———. (1984b). Stability and Polarization of Interests in Job Matching. *Econometrica* **52** 47–57.

———. (1985a). The College Admissions Problem Is Not Equivalent to the Marriage Problem. *J. Economic Theory* forthcoming.

———. (1985b). Common and Conflicting Interests in Two-Sided Matching Markets. *European Economic Rev.* forthcoming.

Shapley, Lloyd S. and Scarf, Herbert. (1974). On Cores and Indivisibility. *J. Math. Econom.* **1** 23–28.

——— and Shubik, Martin. (1972). The Assignment Game. I. The Core. *Internat. J. Game Theory* **1** 111–130.

Thompson, Gerald L. (1980). Computing the Core of a Market Game. *Extremal Methods and Systems Analysis.* A. V. Fiacco and K. O. Kortanek eds., Springer, Berlin, *Lecture Notes in Economics and Mathematical Systems* #174, 312–334.

DEPARTMENT OF ECONOMICS, UNIVERSITY OF PITTSBURGH, PITTSBURGH, PENNSYLVANIA 15260