SIAM Rev.
Vol. 39, No. 4, pp. 575–604, December 1997

OF STABLE MARRIAGES AND GRAPHS, AND STRATEGY AND POLYTOPES*

MICHEL BALINSKI† AND GUILLAUME RATIER‡

Abstract. This expository paper develops the principal known results (and some new ones) on the stable matchings of marriage games in the language of directed graphs. This both unifies and simplifies the presentation and renders it more symmetric. In addition, it yields a new algorithm and a new proof for the existence of stable matchings, new proofs for many known facts, and some new results (notably concerning players' strategies and the properties of the stable matching polytope).

Key words. marriage problem, stable matching, graph, polytope, game, two-sided market

AMS subject classifications. 90C27, 05C90, 52B10, 90D40

PII. S0036144595294515

1. Introduction. It now has been a third of a century since the publication of the short and limpid paper by Gale and Shapley [4]. The question it posed, and to some degree solved, has come to be recognized as being at the core of some fundamental questions in economics. This, together with its mathematical and algorithmic ramifications, has spawned a literature that by now includes at least three books and upwards of one hundred papers.

The *marriage problem*, or *marriage game*, involves two distinct sets of players, M and W, variously interpreted as firms and workers, hospitals and interns, buyers and sellers, students and universities, coeds and sororities, or just plain men and women. Each player of each set has a strict preference ordering over those members of the opposite set that she/he considers to be "acceptable." The analysis focuses on stable matchings, meaning pairings of players from opposite sets, or celibates, having the property that no two players not paired could both improve their individual preferences by being paired. In economics these situations are referred to as "two-sided markets." Instead of imagining a stock market where actors are alternately buyers or sellers depending on prices or other factors, think of labor markets where job seekers or workers occupy a different and distinct role from job offerers or firms.

Among its many practical applications, the model has served to study auctions and auctioneering, to assign interns to hospitals, and to review the evolving history of specific two-sided markets. Mathematically it turns out that stable matchings always exist, sometimes in profusion; that they form a distributive lattice; indeed, that every distributive lattice may be realized as the set of stable matchings of some marriage problem; and that the convex hull of the stable matchings—the *stable matching polytope*—is easily described by a set of readily interpretable linear inequalities.

The style of the mathematics, light and airy and sometimes quite delightful, stands in stark contrast with the awkwardness and weight of the notation, the formalities of the presentation of the problem, and its subsequent analysis. The bidding or auction-like procedure introduced by Gale and Shapley to establish the existence

*Received by the editors November 8, 1995; accepted for publication (in revised form) August 29, 1996.

http://www.siam.org/journals/sirev/39-4/29451.html

†C.N.R.S. and Laboratoire d'Econométrie de l'Ecole Polytechnique, Paris, France (balinski@poly.polytechnique.fr).

‡Laboratoire d'Econométrie de l'Ecole Polytechnique and Université de Paris I Panthéon-Sorbonne, Paris, France.

of stable matchings—where, as in ancient times, the men propose and the women dispose—may be neatly described over a glass of Burgundy, whereas even the definition of the preferences bogs down in lots of chalk and dust. By and large, the proofs are ad hoc combinatorics. The discovery of the stable matching polytope allowed new derivations of the main results via duality in linear programming (though the approach provides no new proof of the existence of solutions), but again the development seems unnecessarily technical and somewhat ponderous.

The aim of this paper is to propose another approach, via directed graphs. We contend that the presentation of the marriage game and the proofs of the various results are at once unified and simplified. A new algorithm for finding stable matchings, and so a new proof for the existence of solutions, naturally emerges. The problem, reasoning, and proofs are couched in terms of graphs. At the least this makes it natural to include the marriage problem in an elementary course on graph theory. We do not exhaustively prove every known result. Instead, we try to show that the main results and some new ones, including when it does and does not pay actors to misrepresent their preferences, are easily understood and proved in terms of graphs and so as pictures at an exhibition. A brief history of the problem, given at the end of the paper, attributes the known results to its originators.

The imagery of marriage—*par excellence* a two-sided affair involving well-identified opposing players—is typical in a profession that seeks to extract the essentials of a subject. It harbors, however, the danger of giving an appearance of triviality, as though the problem is no more than "academic" or nothing more than something of a puzzle. The reader should be aware that behind this convenient presentation of the meat of the matter lurk real questions of importance to society.

2. Stable matchings. A *marriage problem* or *marriage game* is specified by a directed graph defined over a grid as follows.

There are two distinct, finite sets of players, $M = \{m_1, m_2, \ldots, m_{|M|}\}$ (the "men") and $W = \{w_1, w_2, \ldots, w_{|W|}\}$ (the "women"), and each player has a strict complete preference order over those players of the opposite set whom she/he considers to be acceptable. A player prefers to be celibate than to be paired with someone unacceptable to him/her. This data is modelled as a graph with nodes corresponding to pairs of mutually acceptable players and arcs expressing the preferences.

Specifically, the *nodes* of the *marriage graph* Γ are the pairs (m, w), $m \in M$ and $w \in W$, for which w is acceptable to m and m to w. They are located on the $M \times W$ grid where each row corresponds to a man $m \in M$ and each column to a woman $w \in W$. The (directed) *arcs* of Γ, or ordered pairs of nodes, are of two types: a horizontal arc $\{(m, w_i), (m, w_j)\}$ expresses man m's preference for w_j over w_i (sometimes written $w_j >_m w_i$); symmetrically, a vertical arc $\{(m_i, w), (m_j, w)\}$ expresses woman w's preference for m_j over m_i (sometimes written $m_j >_w m_i$).

An example of a marriage game having four men and four women is given by the marriage graph Ψ; see Figure 1. The arcs implied by the transitivity of the preferences, such as $m_4 >_{w_3} m_1$, are omitted. *Throughout the paper arcs implied by transitivity will be omitted.*

A matching in a marriage game is a set of monogamous marriages among consenting players. Some players may remain celibate. In terms of our model, a *matching* μ in a marriage graph Γ is a set of nodes of Γ where no two are in the same *line* of Γ (meaning in the same row or the same column). An example of a matching δ in Ψ is the set of black square nodes in the rendition of Ψ; see Figure 2. The players m_4 and w_4 are celibates in δ.

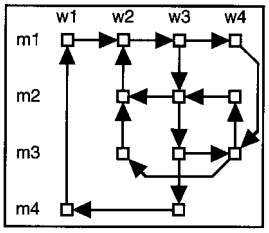

FIG. 1. Ψ.

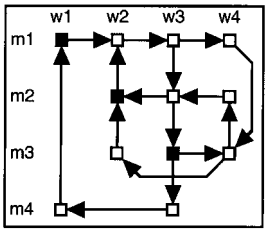

FIG. 2. *Matching δ in Ψ.*

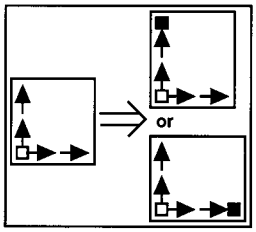

FIG. 3. *Stability.*

The fundamental idea in the analysis of a marriage game is that each player is a rational soul who seeks to optimize his/her preferences. But, as the saying goes, it takes two to tango. Thus, if $(m, w) \in \Gamma$ so that m and w are mutually acceptable but do not belong to a matching μ, then it must be that at least one of them is better off in μ: otherwise, if both were worse off or celibate, m and w acting together but in total disregard of all other players could agree to marry and so improve their situations. Accordingly, a matching μ in a marriage graph is *stable* if the implication in Figure 3 is verified.

The black squares in Figure 3 represent nodes in the matching μ, call it the "black" matching, and the white squares represent a same node not in the black matching. The definition says that if a pair (m, w) represented by the white square is not in the black matching (left graph), then the black matching either pairs woman w with a man she prefers to m (right top) or pairs man m with a woman he prefers to w (right bottom).

In short, a matching μ is *stable* if each node not in μ has a successor in Γ that is in μ. (In the language of graph theory a stable matching in a marriage graph is a "kernel" of the graph.) The matching δ in Ψ is not stable: the definition of stability is not satisfied for several nodes: (m_1, w_2), (m_1, w_4), (m_3, w_4), and (m_4, w_3).

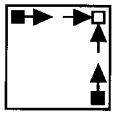

FIG. 4. *Blocking pair.*

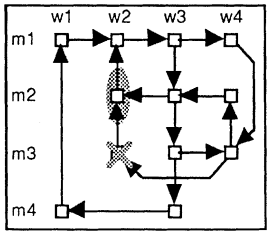

FIG. 5. Ψ^1.

FIG. 6. *Man-best node* (a)*; woman-best node* (b).

FIG. 7. *Man-worst node* (a)*; woman-worst node* (b).

In a marriage graph where no player is ever celibate, the case first studied by Gale and Shapley, stability is equivalent to the absence of a *blocking pair* (m, w), pictured in Figure 4. More generally, a man m and woman w not paired *block* a matching μ if they are both better off together than under μ (in the figure they both abandon black mates, but one or both could be celibate).

The first important question asks if stable matchings exist. Observe that in the marriage graph Ψ it is impossible for there to exist a stable matching μ that includes the node (m_3, w_2). For otherwise, if $(m_3, w_2) \in \mu$, the fact that μ is a matching implies that neither (m_2, w_2) nor any of its successors could belong to μ, thus violating stability for the pair (m_2, w_2). This happens because w_2 is the best choice for m_2, so all the successors of (m_2, w_2) are in the same line w_2. In other terms, since woman w_2 is man m_2's first choice, she has no earthly reason to accept any man who is less attractive to her than m_2 (such as m_3). This suggests dropping the node (m_3, w_2) to obtain the reduced graph Ψ^1 (see Figure 5), hunting for more nodes that cannot belong to any stable matching, dropping them, and continuing in a like manner. But is this legitimate?

Figures 6 and 7 define nodes in a marriage graph Γ that will be useful in the sequel. A *man-best node* means a node (m, w) where w is the woman preferred by m in Γ; a *woman-worst* node means a node (m, w) where m is the least preferred by w of the men in Γ, and similarly for the others.

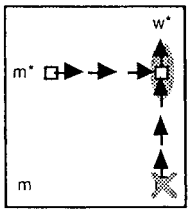

FIG. 8. *Man-best domination.*

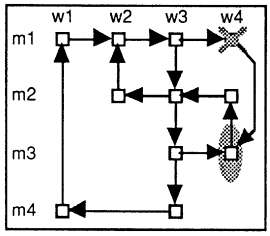

FIG. 9. Ψ^2.

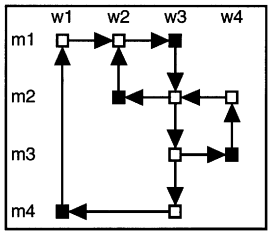

FIG. 10. *Man-best stable matching* μ_M *of* Ψ^2 *and* Ψ.

Two marriage graphs are said to be *equivalent* if they admit precisely the same set of stable matchings.

LEMMA 1. *Suppose* Γ *is a marriage graph that contains the configuration in Figure 8. Then the marriage graph* Γ' *obtained by deleting* (m, w^*) *from* Γ *is equivalent to* Γ.

Proof. Suppose μ is a stable matching in Γ. Then each node of Γ either belongs to μ or has a successor in μ, and this property carries over to Γ'. Since μ is a matching in Γ', it must be a stable matching in Γ'.

For the other way around, suppose μ' is a stable matching in Γ'. Then each node of Γ' either belongs to μ' or has a successor in μ', and this property holds for the nodes of Γ' that are in Γ. Thus, it only remains to show that the property holds for the deleted node (m, w^*) in Γ as well. But this is immediate, since (m, w^*) has a (man-best) successor (m^*, w^*) that satisfies the property. Since μ' a matching in Γ, it must be a stable matching in Γ. □

The graphs Ψ and Ψ^1 are thus indeed equivalent, and one can search for further man-best dominations. In Ψ^1, (m_1, w_4) can be deleted because it is dominated by the man-best node (m_3, w_4) to obtain Ψ^2, whose set of stable matchings is exactly the same as that of Ψ. The graph Ψ^2 (see Figure 9) contains no man-best domination, so no further such deletions can be made. However, in Ψ^2 the set of man-best nodes μ_M constitutes a stable matching (see Figure 10).

LEMMA 2. *The man-best nodes of a marriage graph that contains no man-best domination is a stable matching.*

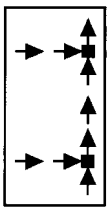

FIG. 11. *If man-best nodes are not matching.*

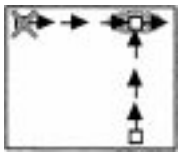

FIG. 12. *Woman-best domination.*

Proof. Two facts need to be established. First, man-best nodes are a matching. If they were not, then two such nodes would belong to the same column and the configuration in Figure 11 would be in the graph, implying the existence of a man-best domination, a contradiction. The second fact is that every node other than a man-best node has a man-best node as a successor. But this is immediate because any node is either a man-best node or has a successor in its row which is a man-best node. □

THEOREM 1. *Stable matchings exist in all marriage games.*

Proof. The deletion of nodes by Lemma 1 cannot go on forever, so they must converge to an equivalent marriage graph containing no man-best domination in a finite number of steps. But in this marriage graph at least one stable matching is evident. □

Man-best dominations have their natural counterparts in woman-best dominations, pictured in Figure 12. So, instead of obtaining an equivalent marriage graph where the man-best nodes constitute a stable matching, one could obtain an equivalent graph where the woman-best nodes constitute a stable matching. Or, even better, do both at once and in any order, since each deletion of a node yields an equivalent marriage graph. Each such deletion is called a *reduction*, and the graph that is obtained is called a *reduced marriage graph*. For example, in Ψ^2 (m_1, w_1) can be deleted because it is dominated by the woman-best node (m_1, w_2) to obtain the reduced graph Ψ^3. Then (m_4, w_3) can be deleted from Ψ^3 because it is dominated by the woman-best node (m_4, w_1) to obtain the reduced graph Ψ^4 (see Figures 13 and 14).

The marriage graph Ψ^4—equivalent to the original game Ψ—contains no man-best or woman-best dominations. So its man-best nodes μ_M (in black) constitute one stable matching of Ψ and its woman-best nodes μ_W (hatched) constitute another stable matching of Ψ. It happens to have one more stable matching (in gray), so it has three in all, as pictured in Figure 15.

3. Comparing stable matchings. A more elaborate example is needed to be able to illustrate the various qualitative and structural properties of stable matchings. Consider the eight man, nine woman problem Φ pictured in Figure 16. The figure contains the information necessary to reduce Φ to the equivalent marriage graph Φ',

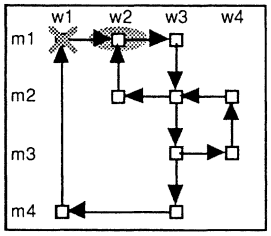

FIG. 13. Ψ^3.

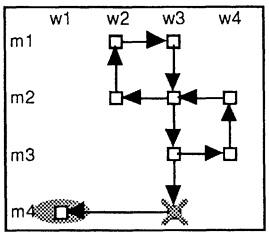

FIG. 14. Ψ^4.

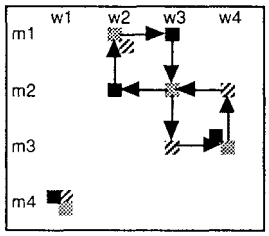

FIG. 15. *Stable matchings of* Ψ^4 *and* Ψ.

which is free of man-best and woman-best dominations, by successive reductions. In "reading" Φ, interpret a node covered with a gray horizontal (vertical) oval as either a woman-best node (man-best node) in Φ itself or in one of its equivalent reductions that is "used" in exploiting a domination; interpret a node covered with a gray cross as a node that is deleted either from Φ itself or from one of its equivalent reductions via a man- or woman-best domination. For example, (m_8, w_8) is both a man-best and woman-best node in Φ, so it is covered by both a horizontal and vertical gray oval, and all of its predecessors in row m_8 and column w_8 may immediately be deleted. However, (m_4, w_6) becomes a man-best node only after (m_4, w_7) is deleted at a previous step.

The graph shown in Figure 17, Φ', decomposes into three independent marriage graphs, $\Phi' = \Phi'_1 \cup \Phi'_2 \cup \Phi'_3$: Φ'_1, which includes the first three men and women; Φ'_2, which contains the fourth through seventh men and women; and Φ'_3, which contains the pair (m_8, w_8). The stable matchings of Φ' are "direct products" of the stable matchings of the components. The first component Φ'_1 is identical to the first component of Ψ^4, so it has precisely three stable matchings; the second component Φ'_2 also has precisely three stable matchings, and the last component Φ'_3 has exactly one stable matching. The woman w_9 remains celibate in all stable matchings.

There are nine stable matchings in all gotten by concatenation, including, of course, the man-best μ_M and woman-best μ_W stable matchings. They are all pictured

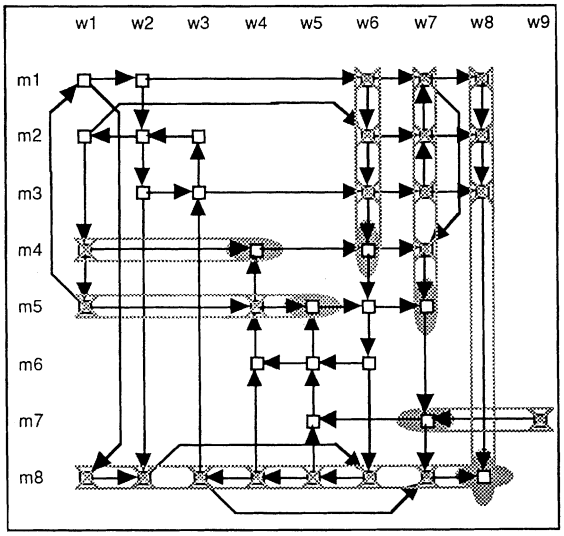

FIG. 16. Φ.

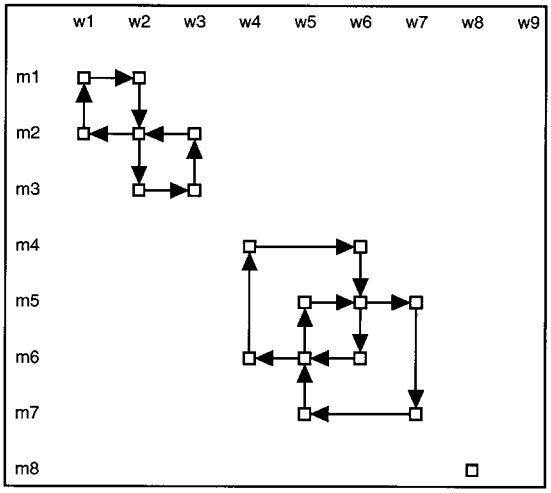

FIG. 17. Φ'.

in Figure 18. For example, the black squares of Φ'_1 together with the black circles of Φ'_2 and (m_8, w_8) constitute the man-best stable matching μ_M; the hatched squares of Φ'_1 together with the hatched circles of Φ'_2 and (m_8, w_8) constitute the woman-best stable matching μ_W.

To be able to refer to specific stable matchings of Φ, let μ_{xy}, where x and y take the names b for black, g for gray, or h for hatched, mean the stable matching formed by squares of color x of Φ'_1 together with circles of color y of Φ'_2 and (m_8, w_8). Thus, $\mu_{bb} = \mu_M$ and $\mu_{hh} = \mu_W$, and each of the nine stable matchings of Φ is represented by some μ_{xy}.

The marriage game Φ is sufficiently endowed in size and stable matchings to permit several immediate but interesting observations. To begin with, the man-best

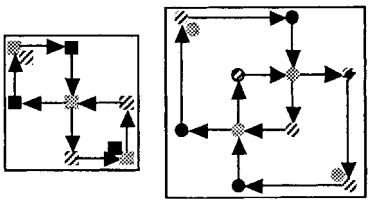

FIG. 18. *All stable matchings of* Φ_1' *and* Φ_2'.

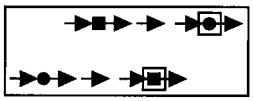

FIG. 19. $\mu \vee \mu'$.

stable matching is the *optimal* stable matching for the men; that is, there exists no stable matching in which any man is better off. It is the best solution from every man's point of view. Symmetrically, the woman-best stable matching is optimal for the women. On the other hand, the optimal solution for the men is the worst one for the women, and symmetrically. The war of the sexes is real! Indeed, this opposition of interests permeates stable matchings. Take two stable matchings of Φ, say μ_{hb} and μ_{gh}: $(m_2, w_3) \in \mu_{hb}$ whereas $(m_2, w_2) \in \mu_{gh}$, so m_2 is worse off in the first case than in the second, but then w_3 is better off in the first case than in the second. This opposition holds for any two stable matchings of Φ. It is also the case that the set of celibate players—the woman w_9 in Φ—is one and the same for all stable matchings.

Given any two stable matchings of a marriage game, μ and μ', consider the following construction. Assign each man to the woman he prefers among the (at most two) women with whom he is matched in μ and μ' and call this their *supremum*, $\mu \vee \mu'$. The picture definition—where the black square is a node of μ, the black circle is a node of μ', and an encompassing square is a node of $\mu \vee \mu'$—is shown in Figure 19. For example, take μ_{hb} and μ_{gh}, stable matchings of Φ. Their supremum yield the stable matching $\mu_{gb} = \mu_{hb} \vee \mu_{gh}$, which is at least as good for every man as either of the two stable matchings μ_{hb} and μ_{gh}. But "at least as good as" for the men translates into "at least as bad as" for the women. One could instead assign each man to the woman he least likes among the (at most two) women with whom he is matched in μ and μ' and call this their *infimum*, $\mu \wedge \mu'$. In our example the infimum yields the stable matching $\mu_{hh} = \mu_{hb} \wedge \mu_{gh}$, and it is at least as bad for every man as either of the two stable matchings μ_{hb} and μ_{gh} (so at least as good for the women).

Clearly, at least among some stable matchings of Φ, there is a collective *men-preferred ordering* which is in opposition to the collective *women-preferred ordering*. For any two different stable matchings μ and μ' of some marriage game Γ, $\mu >_M \mu'$ means that every man is at least as well off in μ as in μ' (and $>_W$ has the similar meaning for women). For example, $\mu_{gb} >_M \mu_{hb}$, whereas for women the preference is the opposite: $\mu_{gb} <_W \mu_{hb}$. Indeed, the collective men-preferred ordering of the nine stable matchings of the marriage game Φ may be described by an oriented graph $G(\Phi, >_M)$ and is pictured in Figure 20 together with the corresponding graphs $G(\Phi_1', >_M)$ and $G(\Phi_2', >_M)$. In $G(\Phi, >_M)$ the nodes are the stable matchings μ_{xy} and are represented by the "colors" black, gray, and hatched; each directed arc points upwards and means that the stable matching above is preferred by all the men to that below. Transitivity

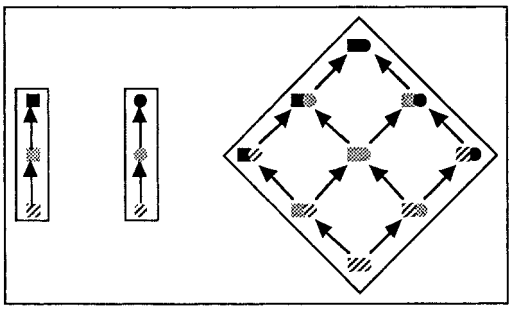

FIG. 20. $G(\Phi_1, >_M)$, $G(\Phi_2, >_M)$, $G(\Phi, >_M)$.

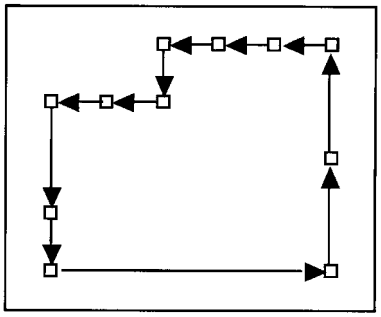

FIG. 21. *Principal circuit.*

is deduced and not pictured directly. It may be observed that the opposition of men and women is translated into the fact that the reverse ordering expresses the collective preferences of the women, $>_W$.

4. Properties of stable matchings. The observations made concerning the marriage game Φ are in fact truths that hold for all marriage games. They are easily established in the language of graphs.

A (reduced) marriage graph that is free of man-best and woman-best dominations is said to be *domination-free*. It may be characterized. Define a *principal circuit* of a marriage graph Γ (in which all arcs implied by transitivity have been suppressed) to be a simple directed cycle that expresses all of the preferences of a woman (or man) in Γ if it expresses any one of her (or his) preferences. Figure 21 is an example of a principal circuit (if Γ contains no other nodes in the three rows and three columns of arcs that express the preferences of three men and three women).

LEMMA 3. *A marriage graph* Γ *has a unique domination-free equivalent marriage graph* Γ^* *that is the union of isolated nodes and principal circuits.*

Proof. Γ^* has the property that every man-best node is also woman-worst, and symmetrically every woman-best node is also man-worst. A simple path-following argument—go from a man-best node (that is also woman-worst) to a woman-best node (that is also man-worst), from a woman-best to a man-best, etc.—shows that all woman-worst nodes of a domination-free marriage graph are also man-best, and similarly all man-worst nodes are also woman-best. Thus, Γ^* is the union of principal circuits and isolated nodes. (For an example, see Φ' in section 2.)

The man-best, woman-worst and the man-worst, woman-best nodes are clearly unique. Suppose a node (m, w) is eliminated at some stage of reduction because of

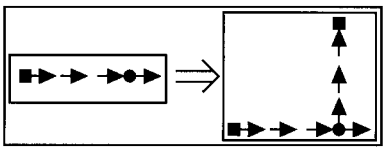

FIG. 22. *Opposing interests.*

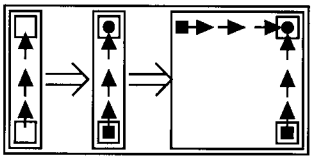

FIG. 23. $\mu \vee \mu'$ *a matching.*

a man-best domination. Then at every subsequent stage—and so in the domination-free graph that is obtained as well—there must be a man-best node (m^*, w) with $m^* >_w m$. Thus, independent of the order of reduction, (m, w) must be eliminated at some stage. □

THEOREM 2. *The unique man-best stable matching μ_M is optimal among the stable matchings for the men, and, symmetrically, μ_W is optimal for the women.*

Proof. Given a marriage graph Γ, obtain the equivalent domination-free graph Γ^*. In Γ^*, μ_M and μ_W are clearly unique and optimal since they are the preferred nodes, respectively, of the men (in each row) and the women (in each column). □

LEMMA 4. *If μ and μ' are stable matchings, $(m, w) \in \mu$, and m is better off in μ than in μ', then w is worse off in μ than in μ'.*

Proof. In Figure 22 the round node represents $(m, w) \in \mu$ and the square nodes belong to μ'. The hypothesis is pictured on the left and the conclusion on the right: the proof is the observation that, otherwise, μ' would not be stable. □

COROLLARY 1. *If μ and μ' are stable matchings, $\mu >_M \mu'$ if and only if $\mu <_W \mu'$.*

LEMMA 5. *A player who is married (respectively, celibate) in one stable matching is married (respectively, celibate) in all stable matchings.*

Proof. Theorem 2 and Lemma 4 imply that each player in every stable matching either has his/her best choice, worst choice, or a choice somewhere in between. Thus, once celibate always celibate and once paired always paired. □

LEMMA 6. *If μ and μ' are stable matchings, then their supremum and infimum, $\mu \vee \mu'$ and $\mu \wedge \mu'$, are stable matchings as well.*

Proof. The proof is essentially the same for the two cases, so consider the set of nodes $\mu \vee \mu'$. Two facts must be established: first, $\mu \vee \mu'$ is a matching; second, it is stable. If $\mu \vee \mu'$ (encompassing squares) were not a matching, then it would include two nodes corresponding to a same woman (as on the left in Figure 23). Two such nodes could not belong either to μ (black squares) or to μ' (black circles) since they are matchings (as in the middle of Figure 23), but the black circle node of the woman who is matched twice must, by the definition of supremum, be preferred by the corresponding man, so in his row a black square precedes the black circle (as on the right). But this implies that the set of black squares μ is not stable, which is a contradiction. If $\mu \vee \mu'$ were not stable then the picture on the left of Figure 24 obtains (perhaps with a black circle and black square interchanged), implying by the definition of supremum that μ' is unstable as shown on the right. □

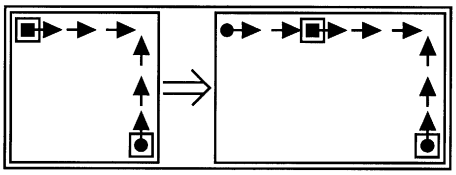

FIG. 24. *$\mu \vee \mu'$ stable.*

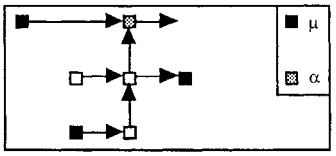

FIG. 25. *Man preferred by w among those who prefer her.*

Notice that due to Lemma 4, the supremum $\mu \vee \mu'$ has the equivalent definition: assign each woman to the man she least likes among the (at most two) men with whom she is matched in μ and μ'. Symmetrically, the infimum $\mu \wedge \mu'$ has the equivalent definition: assign each woman to the man she prefers among the (at most two) men with whom she is matched in μ and μ'.

THEOREM 3. *The set of stable matchings of a marriage game, with the partial order $>_M$, is a distributive lattice $\mathcal{L}(\Gamma)$.*

A *lattice* is a finite, partially ordered set $\mathcal{L}$ having the property that every pair of its elements x and y has a supremum (or least upper bound) written $x \vee y$ and an infimum (or greatest lower bound) written $x \wedge y$. A lattice is *distributive* if every three of its elements, x, y, and z, satisfy $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$ and $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$.

Proof. That the set of stable matchings is finite, partially ordered, and that every pair of its elements has a supremum and infimum is evident.

To see that the first distributive law holds, consider three stable matchings μ, μ', and μ^*. By definition $\mu \vee (\mu' \wedge \mu^*) <_M \mu \vee \mu'$ and $\mu \vee (\mu' \wedge \mu^*) <_M \mu \vee \mu*$, so $\mu \vee (\mu' \wedge \mu^*) <_M (\mu \vee \mu') \wedge (\mu \vee \mu^*)$. The same reasoning with the equivalent definitions yields $\mu \vee (\mu' \wedge \mu^*) >_M (\mu \vee \mu') \wedge (\mu \vee \mu^*)$ and completes the proof. The second law may be established similarly. □

Are there matchings better than μ_M? It is relatively easy to construct examples in which there exists a matching $\mu >_M \mu_M$, $\mu \neq \mu_M$, but of course μ is not stable.

LEMMA 7. *If μ is a stable matching in which every man is matched and every woman who is matched is preferred by at least one man to his mate in μ, then there exists a stable matching $\mu' \neq \mu$ for which $\mu' >_M \mu$.*

Proof. Given μ, a stable matching in the marriage graph Γ, choose in the column of each woman w who is matched the node (m, w) that corresponds to the man m she prefers among all those who prefer her to their current mates in μ. Call this set of nodes α; see Figure 25. The cardinality of both α and μ is $|M|$. The set of nodes $\alpha \cup \mu$ must contain a cycle C alternating between nodes of α and of μ.

To see this, consider the bipartite graph whose nodes are $M \cup W'$, where W' is the set of women matched in μ with (m, w) an edge if $(m, w) \in \alpha \cup \mu$. It has $|\alpha| + |\mu| = 2|M|$ edges, so it must contain a cycle of edges alternating between α and μ.

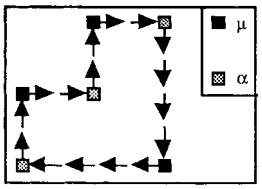

FIG. 26. C.

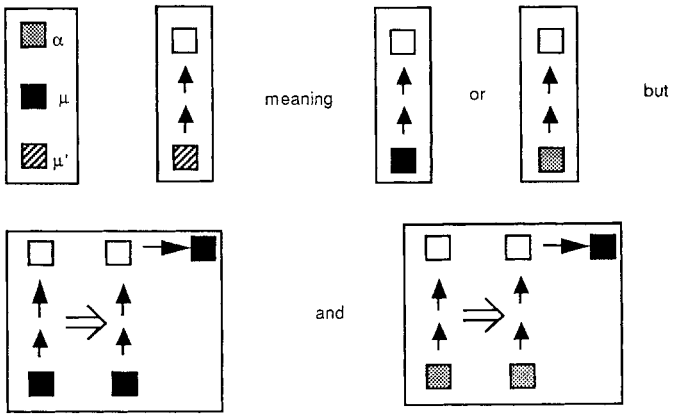

FIG. 27.

In C, every node of α is preceeded in its row by a node of μ (by definition) and succeeded in its column by a node of μ (since μ is stable), as pictured in Figure 26.

Define μ' to be μ except for the nodes within the cycle C, where those of α are substituted for those of μ:

$$(m, w) \in \mu' \text{ if either } (m, w) \in \alpha \cap C \text{ or } (m, w) \in \mu - C.$$

μ' is clearly a matching. To see that it is stable, suppose that some node (m, w) is not followed by a node of μ' in its column. Then it will be shown that it must be followed by a node of μ in its row and so (by construction) by a node of μ' in its row.

If there is no node of μ' in the column of the node (m, w) then it must be followed by a node of μ in its row, by the stability of μ. Otherwise, the implications of Figure 27 obtain the first by the stability of μ and the second by the definition of α. □

This yields the following theorem as an immediate corollary.

THEOREM 4. *There exists no matching μ (stable or not) with $\mu >_m \mu_M$ for all $m \in M$.*

Proof. If there are more men than women, some man must be celibate so there can be no such μ. Otherwise, suppose such a μ exists. μ cannot contain a woman not matched under μ_M—for that would contradict the optimality of μ_M—so the conditions of the previous lemma obtain, and there exists a stable matching $\mu' \neq \mu_M$ with $\mu' >_M \mu_M$, which is a contradiction. □

LEMMA 8. *If μ is a matching preferred to μ_M by the men $m \in M' \subset M$ but not by $m \notin M'$, then some (m, w) with $m \notin M'$ blocks μ.*

Proof. Either (a) the set of women $\mu(M')$ and $\mu_M(M')$ are the same, or (b) they are not.

(a) Let W' be the common set of women. Suppose $\mu_{M'}$ is the man-best stable matching in the marriage subgraph $\Gamma_{M'}$ of Γ defined on the nodes $M' \times W'$. Since

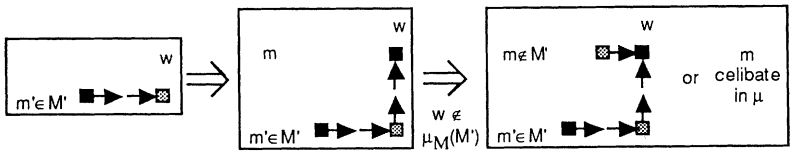

FIG. 28.

$\mu >_m \mu_M$ for all $m \in M'$, Theorem 4 implies $\mu_{M'} >_{M'} \mu_M$. Define μ' to be the matching that agrees with μ_M except that on M' it is $\mu_{M'}$. Therefore, $\mu' >_M \mu_M$, so μ' cannot be stable, and there must exist a blocking pair (m, w). The pair (m, w) cannot belong to $M' \times W'$, for this would mean $\mu_{M'}$ is not stable in $\Gamma_{M'}$. Nor can it belong to $M' \times (W - W')$ or $(M - M') \times (W - W')$, for this would mean μ_M is not stable in Γ. Therefore, it must belong to $(M - M') \times W'$, as claimed.

(b) If $\mu(M') \neq \mu_M(M')$, there must exist a woman $w \in \mu(M') - \mu_M(M')$ as pictured in Figure 28. w must be paired in μ_M, otherwise (m', w) blocks μ_M, and $m \notin M'$, since $w \notin \mu_M(M')$. So the pair (m, w) blocks μ as was to be shown. □

5. Finding stable matchings. The algorithm implicit in the reduction procedure that proves the existence of solutions may be interpreted as a generalization of the Gale–Shapley "men propose, women dispose" algorithm. In this section the algorithms are detailed and analyzed, once again in the language of graphs.

REDUCTION ALGORITHM.

For each row and column and in any order
  find in the row of m_i the man-best node (m_i, w),
    eliminate all nodes preceding (m_i, w) in w's column;
  find in the column of w_j the woman-best node (m, w_j),
    eliminate all nodes preceding (m, w_j) in m's row;
until every man-best node is woman-worst, and every woman-best is
  man-worst.

The set of man-best nodes is then the optimal matching μ_M of the men, and the set of woman-best nodes is that of the women μ_W.

One "step" eliminates a node. The work of each step is the same. This relies on describing the graph as a doubly-linked list and maintaining four lists of nodes: man-best and not woman-worst, woman-best and not man-worst (if either of these is nonempty a domination is identified), man-best and woman-worst, and woman-best and man-worst. Thus, the complexity is $O(n^2)$.

The Gale–Shapley "propose–dispose" algorithm has a neat description. To state it succinctly, take it as a rule of behavior that no man can propose twice to the same woman, so if he is rejected by a woman she simply disappears from his list of preferences. The algorithm is then this: each man proposes to his favorite, and each woman with at least one proposal says "maybe" to the best of the lot (in her estimation) and rejects the others. The algorithm stops when no man is rejected, for then each man either has one "maybe" in hand or he has been rejected by every woman acceptable to him: the set of "maybes" constitutes the stable matching μ_M. If, instead, the women do the proposing and the men the disposing, then the stable matching μ_W obtains.

This algorithm is readily explained in terms of graph reductions.

PROPOSE–DISPOSE ALGORITHM.

For each row find the man-best node, and call this the set β_M.

If two or more nodes of β_M are in a same column w,

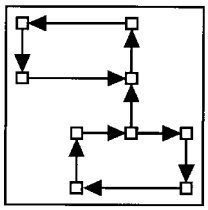

FIG. 29. *Propose–dispose reduced.*

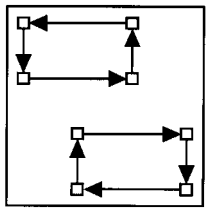

FIG. 30. *Domination-free; "fully" reduced.*

eliminate the one or more nodes of β_M that precede the best of them in w's column, and
redefine β_M by finding the new man-best node in the row of each eliminated node;
until no column has more than one node of β_M.

The set β_M of man-best nodes is then the optimal matching μ_M of the men.

The proof that this algorithm works is the same: each reduction produces an equivalent graph, and when the set of man-best nodes is a matching, it clearly must be a stable matching. The worst-case complexity analysis and estimate is also the same. But it is clear that when one or more nodes are to be eliminated in the propose–dispose algorithm, it takes no extra work to eliminate other dominated nodes and only hastens convergence. In addition, the algorithm only produces the man-optimal stable matching μ_M, and the usual prescription for producing the woman-optimal stable matching μ_W is to begin again, with the roles of men and women reversed. But with this graph reduction algorithm it is possible to continue from the equivalent graph in which the man-best nodes are a matching, doing the same thing with the roles of men and women interchanged, and so obtain a graph in which the woman-best nodes are also a matching. But the reduction algorithm does better because its final graph more clearly reveals the underlying structure of the problem. The following example makes the point. The two marriage graphs of Figures 29 and 30 are equivalent; the first is obtained via the propose–dispose procedure applied first as stated then with the roles of the men and women reversed, and the second is obtained via the full reduction algorithm.

On the other hand, it should not be assumed that the reduction algorithm eliminates all nodes that are unnecessary to obtaining stable matchings. The black node in the marriage graph in Figure 31 is a case in point: it may be eliminated to obtain an equivalent graph, but not by man-best or woman-best domination. Some new reason needs to be invoked.

It is also possible for there to be a node which belongs to no stable matching μ but whose elimination would change the set of stable matchings. This is the case of the black node in Figure 32.

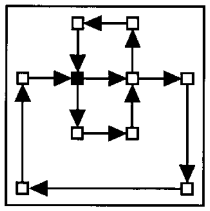

FIG. 31. *Undominated node that can be eliminated.*

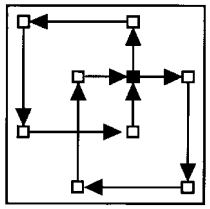

FIG. 32. *Necessary node but in no stable* μ.

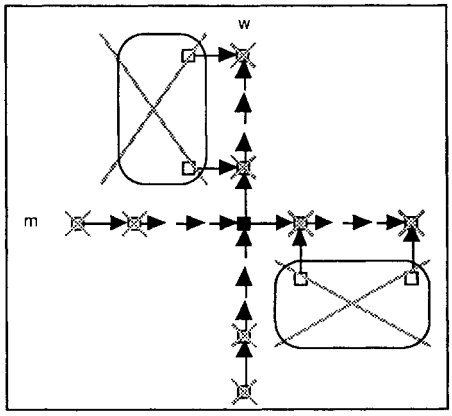

FIG. 33. *Elimination of nodes.*

The algorithm may be used for other purposes as well. Suppose one wished to find a stable matching that included the specific pair (m, w) or that included several specific pairs $(m_1, w_1), \ldots, (m_k, w_k)$. Such stable matchings may or may not exist. To find such stable matchings or to prove none exist, it suffices to take the graph obtained by the reduction algorithm and for each pair (m, w) that is to be included taken in turn, eliminate all nodes that could cause an instability, as pictured in Figure 33, and then apply the reduction algorithm to what is left. If at any stage one of the nodes to be included is eliminated, or if all the nodes of some man or woman matched in μ_M are eliminated, then there is no stable matching that includes the wish list; otherwise, all the stable matchings in the final marriage graph constitute stable matchings of the original graph that include the wish list.

6. Number of stable matchings. How many stable matchings can there be in a marriage game? At least one and, in general, several, judging from the two examples so far considered. There are examples that contain exactly one: take any game where

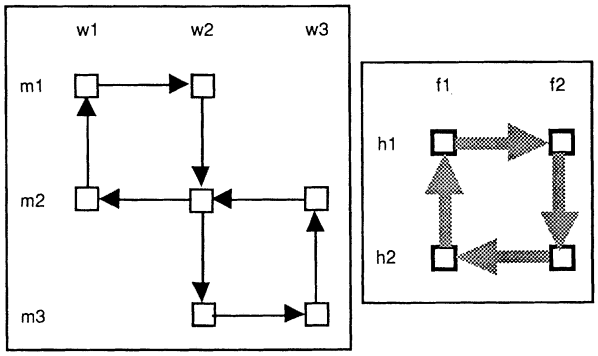

FIG. 34. Γ *and* Φ.

there is a matching that pairs each woman with the man she prefers to all others, and he simultaneously prefers her to all others. Can there be very many?

The following facts are immediately evident.

LEMMA 9. *Suppose a marriage graph* Γ *is the union of disjoint connected subgraphs* $\Gamma_1, \ldots, \Gamma_k$. *Then*

(1) *each* Γ_i *is itself a marriage graph*;

(2) μ, *a stable matching of* Γ, *implies that* $\mu_{|\Gamma_i}$, *its restriction to* Γ_i, *is a stable matching of* Γ_i; *and*

(3) μ, *a matching of* Γ, *and* $\mu_{|\Gamma_i}$, *its restriction to* Γ_i, *a stable matching of* Γ_i *for every* i, *implies that* μ *is stable in* Γ.

Thus, to obtain all stable matchings of a marriage graph Γ, it is sufficient to obtain all stable matchings of each of its connected subgraphs. In particular, a stable matching that contains only man-best and woman-best nodes can be obtained by taking either the man-best or the woman-best nodes of each connected component. If there are k such components that are not isolated nodes, this produces 2^k stable matchings, so a marriage game with $2k = |M| = |W|$ can realize that number. It is possible to do better.

Given two marriage graphs, Φ defined on $H \times F$ and Γ defined on $M \times W$, their *product* is the marriage graph $\Delta = \Phi * \Gamma$, where the set of men are all pairs (hm), $h \in H$ and $m \in M$, and the set of women are all pairs (fw), $f \in F$ and $w \in W$, so that Δ is defined on the grid of pairs (hm, fw). For each node (h, f) of Φ and (m, w) of Γ, there corresponds a node (hm, fw) of Δ. The preferences of each man hm over the set of women is the lexicographic preference induced by the preferences of h then m, which means if h prefers f to f' then hm prefers fw to $f'w'$ and if m prefers w to w' then hm prefers fw to fw'. The preferences of each woman fw over the set of men is defined analogously. In effect, each node of Φ is replaced by a copy of Γ, as may be seen in Figures 34 and 35. The arcs are differentiated to show their provenances.

Suppose that ϕ is a stable matching of Φ, for example $((h_1, f_2); (h_2, f_1))$. Then insert into each copy of Γ that corresponds to a node that belongs to ϕ some stable matching of Γ; say the black nodes that are apparent in Figure 35. It is simple to verify that this yields a stable matching for Δ.

THEOREM 5. *The maximum number of stable matchings is exponential.*

Proof. Let $N(\Gamma)$ be the number of stable matchings of a graph Γ. The above argument shows that $N(\Phi * \Gamma) \geq N(\Gamma)[N(\Gamma)]^{N(\Phi)}$. If N_k is the maximum number of stable matchings of a graph having k men, then $N_{pq} \geq Np[Nq]^p$. From $N_2 = 2$, one deduces that $N_{2^n} \geq 2[N_{2^{n-1}}]^2$ and, therefore, $N_k \geq 2^{k-1}$ when $k = 2^n$. □

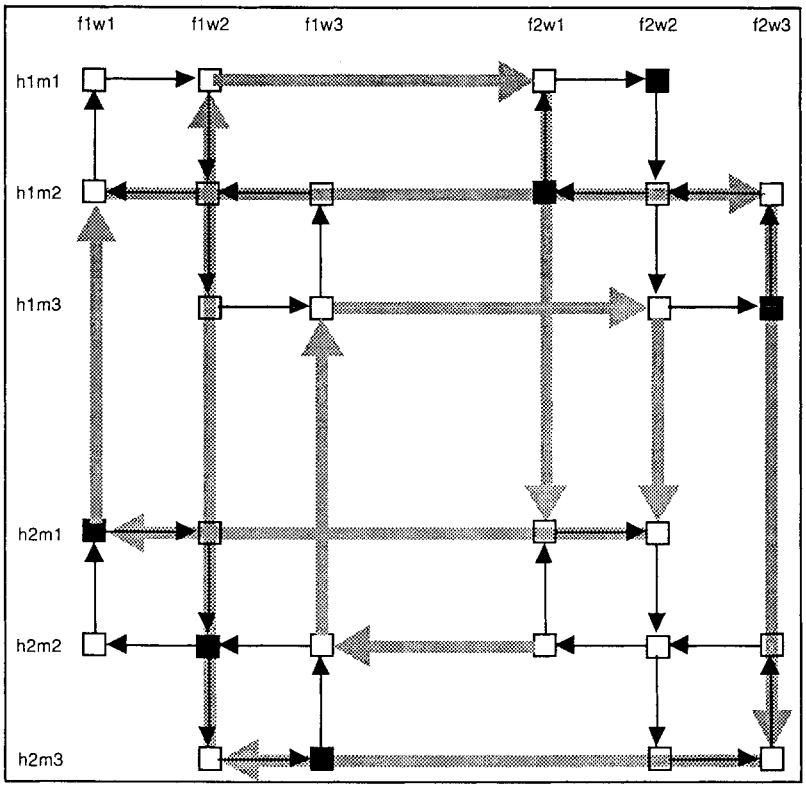

FIG. 35. $\Delta = \Phi * \Gamma$.

7. Strategy in marriage games. Suppose you were to participate in a marriage game—indeed, most of us do at one or another juncture of our lives, in seeking entrance to a university or in finding a job or in hunting for a mate—would you truthfully reveal your preferences? Or might you be tempted to artfully shade your true preferences in the hopes of thereby doing better? A player's true preferences are one thing; her announced preferences are quite another! What is announced constitutes a *strategy* that, together with the announced preferences or strategies of all the other participants, determines the outcomes of the game and the players' possible mates. It is, of course, supposed that the outcome of a game is a stable matching based on the announced preferences.

The four men, four women marriage game Ψ given in section 2 is a handy example: if everyone announces true preferences and, in particular, woman w_3 announces her true preferences ($m_1 <_{w3} m_2 <_{w3} m_3 <_{w3} m_4$), then she may end up matched with m_1, m_2, or m_3. On the other hand, if she announces only $m_3 <_{w3} m_4$, saying that she is unwilling to contract marriage with either m_1 or m_2, then she ensures that she can only be matched with m_3, a clear improvement. Symetrically, if m_1 announces $w_3 <_{m_1} w_4$ instead of his true preferences (thus, lying by saying he refuses to consider either w_1 or w_2), he can assure a marriage with w_3 instead of an uncertainty between w_2 and w_3. And these are only little fibs, false claims concerning willingness!

A *mechanism* is taken to mean the choice of one stable matching (such as the mechanism that yields the man-best stable matching). No mechanism exists for which

honesty is always the best policy. The example Ψ suffices to prove this. It has three stable matchings: black, hatched, and gray. If the mechanism yields the black matching, it contains (m_1, w_3); if it yields the gray, it contains (m_2, w_3): by changing her preferences, w_3 can force with profit the hatched matching which contains (m_3, w_3). If the mechanism yields the hatched matching, then m_1 can lie and enforce with profit the black matching.

Moreover, for any mechanism and any marriage game Γ that contains more than one stable matching, there always exists at least one player who can, with profit, cheat (if no other players cheat). Consider the equivalent marriage game Γ' that is domination-free and let μ be the matching picked out by the mechanism. Either $\mu(m) \neq \mu_M(m)$ for some man or $\mu(w) \neq \mu_W(w)$ for some woman. In the first case, man m can cheat by saying he is unwilling to marry any woman other than $\mu_M(m)$ and thus force whatever mechanism is being used to choose a stable matching, including $(m, \mu_M(m))$. In the second case, the woman can similarly cheat to enforce $(\mu_W(w), w)$. If, in particular, the man-best mechanism is used and the game has at least two stable solutions, then the same argument shows that at least one woman can cheat with profit if no other players cheat.

There is a plethora of propositions on the strategic problems of marriage games, most concerned with the asymmetric situation where it is supposed that the "men propose and women dispose" mechanism is used, meaning that the unique outcome of the game is the man-best matching μ_M (based on the announced preferences of the players). The following theorem shows that (if the women do not cheat) it never pays the men to announce strategies other than their true preferences.

THEOREM 6. *Suppose Γ expresses the true preferences of all players and that Γ' does too, except for a subset M' of the men who announce altered preferences. There is no stable matching μ' in Γ' that is preferred to μ_M by all of the men in M'.*

Proof. Suppose that the matching μ' is preferred to μ_M by the men M^*, where $M' \subset M^*$. Then by Lemma 8 there must exist a pair (m, w) with $m \notin M^*$ that blocks μ' in Γ . But the strategies of m and w announced in Γ' are identical to their true preferences given in Γ, so (m, w) is a blocking pair in Γ' as well. □

On the other hand, suppose that all of the women cheat in the same way; namely, every woman w opts for the strategy in which she declares that every man m ranked below $\mu_W(w)$ is unacceptable to her. It is obvious (look at the equivalent domination-free marriage graph) that the only stable matching that remains is the woman-best stable solution μ_W, so it clearly pays the women to do this (if more than one stable matching exists, otherwise no harm is done). Recall that by (the mirror image of) Theorem 4 there exists no (stable or unstable) matching $\mu >_W \mu_W$, so in fact these choices constitute a *strong equilibrium* for the women: no coalition of one or more women can do better by using any other strategy; moreover, this is true whatever mechanism is used. How each woman w is to know the identity of the man $\mu_W(w)$ is another question.

A more symmetric example concerns a coalition of one man and one woman.

THEOREM 7. *If (m, w) is a node of a domination-free marriage graph, then there exists a stable matching μ in which either man m and woman w are matched, $(m, w) \in \mu$, or they are both strictly better off than being matched.*

This may be interpreted as saying that if the "useless" information concerning preferences is discarded, then any one mutually acceptable pair (m, w) could together agree to falsify their preferences, with m saying that any woman less attractive than w is not acceptable and w saying that any man less attractive than m is not acceptable;

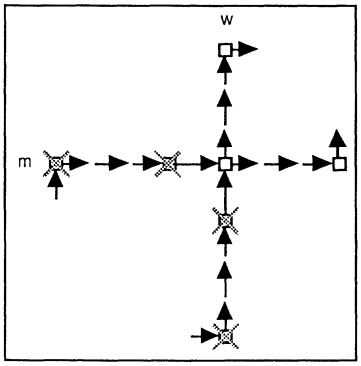

FIG. 36. Γ'.

they thereby assure themselves of at least each other if not better (assuming, of course, that the others are not so sly and do not cheat).

As a preliminary observation, note that if (m, w) is simultaneously a woman-best and man-worst node (or man-best and woman-worst node) in a marriage graph Γ, then man m is paired (woman w is paired) in every stable matching of Γ.

Proof. Suppose (m, w) is a node of a domination-free marriage graph Γ. Consider the marriage graph Γ' obtained by eliminating all nodes that precede (m, w) in Γ, as pictured in Figure 36 (the eliminated nodes are crossed out). The idea is to show that any stable matching μ of Γ' is also a stable matching of Γ. There are two possibilities to consider: either $(m, w) \in \mu$ or $(m, w) \notin \mu$.

Suppose $(m, w) \in \mu$. Since μ is stable, every node of Γ' has a successor in μ, so every node of Γ except those dropped have successors in μ. But those dropped have the successor (m, w), so the condition is true for all nodes of Γ, showing that μ is a stable matching of Γ.

Suppose then that $(m, w) \notin \mu$. It may be assumed that (m, w) has predecessors in both its row and its column (as pictured); otherwise, since Γ is domination-free, it would either be a man-best or a woman-best node, so either μ_M or μ_W would include (m, w) in a stable matching. Consider the set of woman-best nodes in Γ' except for that in the row of m. They are all man-worst in Γ' and, therefore, as noted above each such man is necessarily matched in Γ'. Thus, every man other than m is necessarily matched in Γ'. The symmetric argument shows that every woman other than w is necessarily matched in Γ'. But μ is a stable matching in Γ', and (m, w) must have a successor in its row or column, so either man m or woman w is matched by μ in Γ'. By parity, it is impossible for all individuals but one to be matched, so all must be matched. Since all are matched and $(m, w) \notin \mu$, m is matched with someone he prefers to w and w with someone she prefers to m. But this shows that μ is a stable matching in Γ because the nodes that were eliminated all have successors. □

A slight generalization of Theorem 6 shows that there are limits to ill-gotten profits in the symmetric case as well.

THEOREM 8. *Suppose* Γ *expresses the true preferences of all players and that* Γ' *does too, except for some coalition of men* M' *and women* W' *who announce altered preferences. There is no stable matching* μ' *in* Γ' *that is preferred by each member of the coalition to every stable matching of* Γ.

Proof. Suppose that the matching μ' is preferred to every stable matching of Γ by each member of the coalition. This means that, in Γ, $\mu'(m) >_m \mu_M(m)$ for every $m \in M'$ and $\mu'(w) >_w \mu_W(w)$ for every $w \in W'$. If either M' or W' is empty then

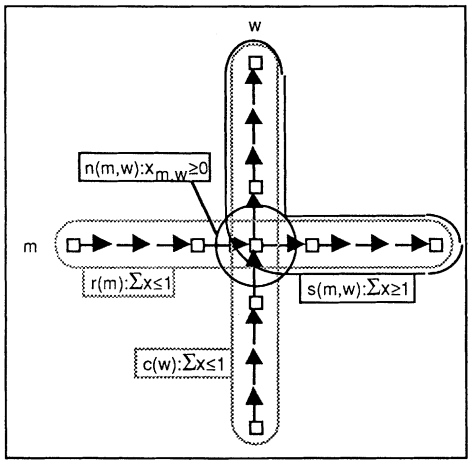

FIG. 37. $P(\Gamma)$.

Theorem 8 is the same as Theorem 6. So, suppose M' is nonempty. By Lemma 8 there must exist a pair (m, w) with $m \notin M'$ that blocks μ' in Γ. It is impossible for $w \in W'$, because this would imply that μ_M and μ_W are also blocked by (m, w). Therefore, $w \notin W'$, so man m and woman w have not altered their preferences and (m, w) blocks μ' in Γ' as well. □

8. The stable matching polytope. One of the principal themes of modern combinatorial optimization is the study of the convex hull of the set of feasible solutions to a problem. In terms of marriage games, the question is this: determine a set of linear inequalities and/or equations whose extreme points are precisely the stable matchings. Of course, this presupposes an algebraic model for representing stable matchings, so for expressing sets of nodes in a marriage graph.

Given a marriage graph Γ, associate with each node (m, w) a variable $x_{m,w}$ having values 0 or 1. A subset μ of the nodes is modelled by the vector $x = (x_{m,w})$, $(m, w) \in \Gamma$, via the relation $(m, w) \in \mu$ if and only if $x_{m,w} = 1$. In short, x is the characteristic vector of the set μ. By an abuse of language we will say that x *is* a stable matching when it is the characteristic vector of a stable matching, and we will also refer to the values of successor or predecessor nodes *of* $x_{m,w}$.

LEMMA 10. *A* 0,1*-valued vector* $x = (x_{m,w})$, $(m, w) \in \Gamma$, *is a stable matching if and only if it verifies the linear inequalities pictured in Figure* 37.

Proof. The inequalties associated with each man m and woman w say that no individual can be paired more than once; that is, x is a matching. Call these the *row* and the *column constraints* (referred to as $r(m)$ and $c(w)$). There are two inequalities associated with each acceptable pair (m, w). One says that $x_{m,w} \geq 0$ and is referred to as a *nonnegativity constraint* $(n(m, w))$; the other says that either $x_{m,w} = 1$ or the value of at least one of its successor nodes must be 1. This is precisely the condition for stability, so it is called a *stability constraint* (referred to as $s(m, w)$). □

THEOREM 9. *The extreme points of* $P(\Gamma)$ *are the stable matchings of* Γ.

It happens that the intuitively obvious inequalities—with the restriction that each $x_{w,m}$ must take on the value 0 or 1 dropped—define the matching polytope $P(\Gamma)$. In view of Lemma 10 it suffices to show that any noninteger valued $x \in P(\Gamma)$ is not an extreme point. A preliminary lemma is needed.

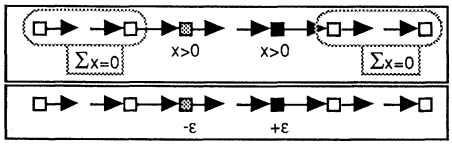

FIG. 38. *Definition of* x^ε.

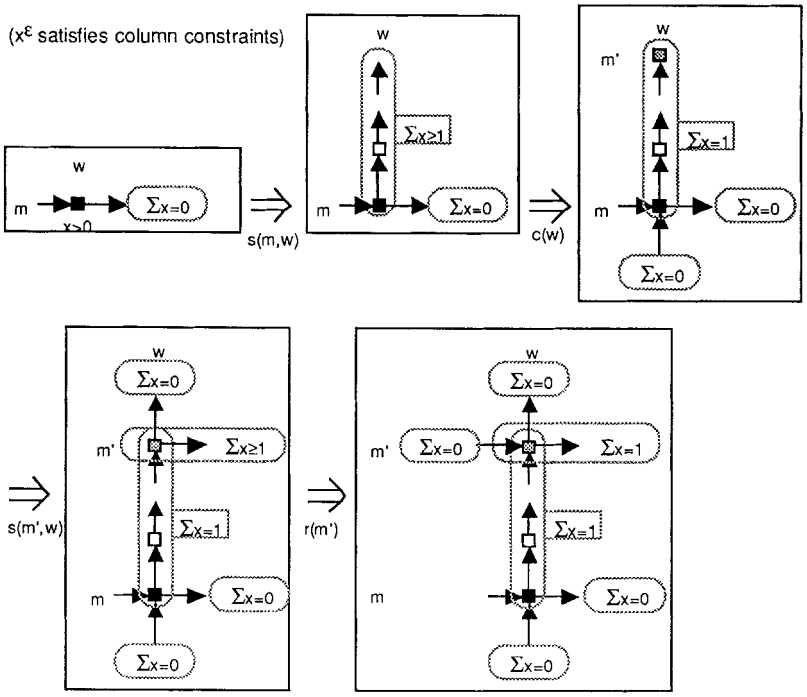

FIG. 39. x^ε *satisfies column constraints.*

LEMMA 11. *If* $x \in P(\Gamma)$, *then* $x^\varepsilon \in P(\Gamma)$ *for* $|\varepsilon| < \min\{x_{m,w} > 0\}$, *where*

$$x^\varepsilon_{m,w} = \begin{cases} x_{m,w} + \varepsilon & \text{if } x_{m,w} > 0 \text{ and all successor values in } m\text{'s row are } 0, \\ x_{m,w} - \varepsilon & \text{if } x_{m,w} > 0 \text{ and all predecessor values in } m\text{'s row are } 0, \\ x_{m,w} & \text{otherwise,} \end{cases}$$

as pictured in Figure 38.

Proof. That $x^\varepsilon \geq 0$ and that the row constraints are satisfied by x^ε is evident. The sequence of pictures in Figure 39 shows that the column constraints are satisfied by x^ε. Call the nodes for which $x_{m,w} > 0$ positive nodes of x. The last picture shows that the positive man-best (positive woman-best) node is a positive woman-worst (positive man-worst) node. Repeating, one sees that the positive nodes constitute a domination-free marriage graph. Moreover, the sum of the positive x's in the column must be exactly 1, and the same argument implies that in the new row the sum of the positive x's will also be 1. Repeating, new rows and columns are successively identified and none but the first can be repeated; so the sum of the positive x's is 1 in all lines.

This leaves the stability constraint to be verified for each (m, w). There are three possibilities: either (1) (m, w) is positive man-worst in its row or is worse; or

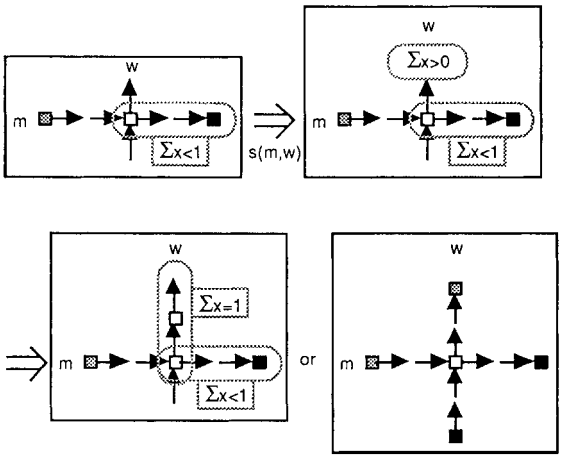

FIG. 40. *Case* (3).

(2) (m, w) is positive man-best in its row or is better; or (3) (m, w) is between the positive man-worst and positive man-best in its row.

In case (1), the successors of (m, w) in m's row sum to 1 in x and in x^ε as well, so the stability condition is automatically satisfied by x^ε. In case (2), the successors of (m, w) in w's column in x sum to 1 because of the stability constraint, so the condition carries over to x^ε, which again means that the corresponding stability constraint is satisfied. Finally, in case (3), the implications pictured in Figure 40 show that either (m, w) and its successors in w's column in x sum to 1, so the condition carries over to x^ε, or the woman-best and woman-worst nodes in w's column are, respectively, above and below (m, w), so the stability constraint necessarily remains satisfied in x^ε.

This completes the proof of Lemma 11. □

The proof of Theorem 9 is now immediate. Lemma 11 shows that if $x \in P(\Gamma)$ is not integer valued, then $x = \frac{1}{2}(x^\varepsilon + x^{-\varepsilon})$ for $x^\varepsilon \neq x^{-\varepsilon}$, which both belong to $P(\Gamma)$, so x is not an extreme point. If x is integer valued then $x = x^\varepsilon$.

The structure of the stable matching polytope $P(\Gamma)$ is in fact determined by the structure of the distributive lattice of the stable matchings of the marriage game Γ. To see this requires a bit of work and several concepts. There are alternative definitions of a face of a polytope. For our purposes, a *face* F of a polytope P defined by equations and/or inequalities is a maximal set of extreme points of P that all saturate the same set of inequalities. The *dimension* of a face is the dimension of the space spanned by its extreme points. An extreme point is a face of dimension 0. Two extreme points are *neighbors* if they are a face of dimension 1. In the sequel, the extreme point corresponding to a stable matching μ will be written $x(\mu)$.

THEOREM 10. *Two extreme points $x(\mu_-)$ and $x(\mu_+)$ of $P(\Gamma)$ are neighbors if and only if*

(i) *μ_- and μ_+ are comparable, say $\mu_+ >_M \mu_-$, and*

(ii) *there exist no two other stable matchings μ and μ' with $\mu_- = \mu \wedge \mu'$ and $\mu_+ = \mu \vee \mu'$.*

Proof. The proof exploits one basic idea: $x(\mu_-)$ and $x(\mu_+)$ are neighbors in $P(\Gamma)$ if and only if their midpoint is uniquely expressible as a convex combination of extreme points of $P(\Gamma)$ as $\frac{1}{2}[x(\mu_-) + x(\mu_+)]$.

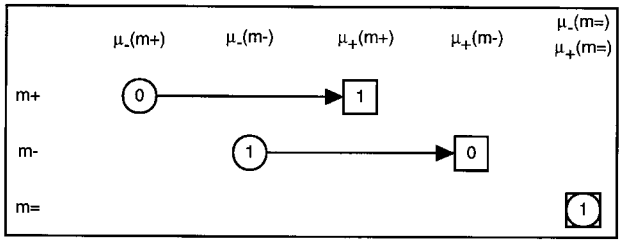

FIG. 41. $x(\mu)$.

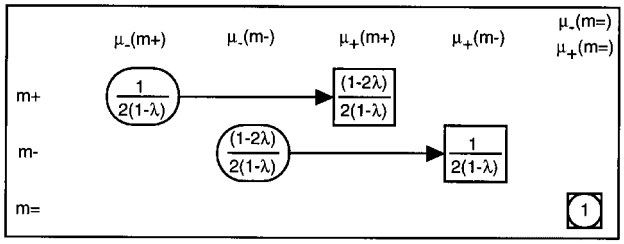

FIG. 42. y.

If μ_- and μ_+ are not comparable, then $x(\mu_-)+x(\mu_+) = x(\mu_- \wedge \mu_+)+x(\mu_- \vee \mu_+)$, so the midpoint of $x(\mu_-)$ and $x(\mu_+)$ is expressible as a convex combination of extreme points of $P(\Gamma)$ in more than one way, showing that condition (i) is necessary.

If μ_- and μ_+ are comparable and there exist two other stable matchings μ and μ' with $\mu_- = \mu \wedge \mu'$ and $\mu_+ = \mu \vee \mu'$, then $x(\mu_-) + x(\mu_+) = x(\mu) + x(\mu')$, which shows that condition (ii) is necessary.

To see that the conditions are sufficient, it will be shown that $x(\mu_-)$ and $x(\mu_+)$, not neighbors but μ_- and μ_+ comparable, imply that, in fact, there do exist two other stable matchings μ and μ' with $\mu_- = \mu \wedge \mu'$ and $\mu_+ = \mu \vee \mu'$, contradicting (ii). $x(\mu_-)$ and $x(\mu_+)$, not neighbors, imply that their midpoint may be expressed as at least two different convex combinations of the extreme points of $P(\Gamma)$; that is,

$$\frac{1}{2}[x(\mu_-) + x(\mu_+)] = \sum_{\mu} \lambda_\mu x(\mu), \text{ where } \sum_{\mu} \lambda_\mu = 1 \text{ and } \lambda_\mu \geq 0 \text{ for all } \mu,$$

with $0 < \lambda_\mu < 1$ for at least one stable matching other than μ_- and μ_+. Calling this different stable matching μ and $\lambda = \lambda_\mu > 0$, this means that

$$\frac{1}{2}[x(\mu_-) + x(\mu_+)] = \lambda x(\mu) + (1-\lambda)y, \text{ for } y \in P(\Gamma).$$

$y \in P(\Gamma)$ implies $y \geq 0$, so $(m, w) \in \mu$ only if either $(m, w) \in \mu_-$ or $(m, w) \in \mu_+$. But μ is different than μ_- and μ_+, so there are men of type m_- where μ agrees with μ_- but not μ_+, men of type m_+ where μ agrees with μ_+ but not μ_-, and men of type $m_=$ where μ agrees with both μ_- and μ_+: as seen in Figure 41, where the circles represent μ_-, the squares represent μ_+, and the 1's represent μ or the value of the corresponding $x(\mu)$.

The sum of the two values given in each row of y (see Figure 42) is exactly one, so applying the transformation of Lemma 11 twice (first with $\varepsilon = -\frac{(1-2\lambda)}{2(1-\lambda)}$ to obtain

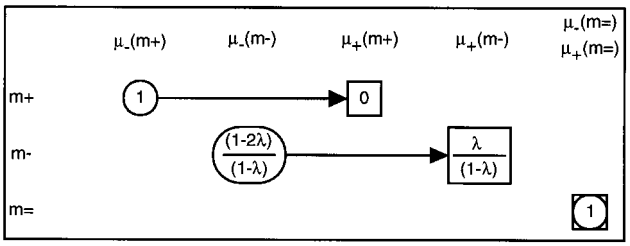

FIG. 43. y'.

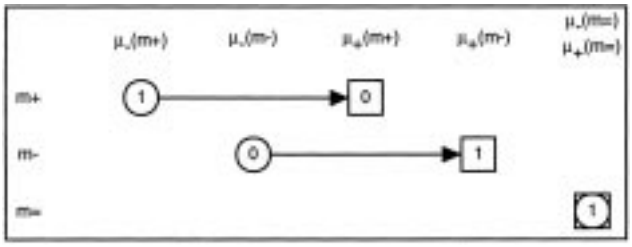

FIG. 44. $x(\mu')$.

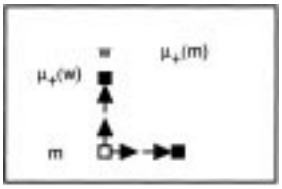

FIG. 45. μ_+ *does not saturate* (m, w) *stability.*

y' in $P(\Gamma)$ then with $\varepsilon = \frac{(1-2\lambda)}{2(1-\lambda)}$), obtains an integer-valued point in $P(\Gamma)$, which is therefore an extreme point $x(\mu')$. The corresponding values of y' and $x(\mu')$ are as pictured in Figures 43 and 44. Note that μ' is the "complement" of μ in $\mu_- \cup \mu_+$: if $(m, w) \in \mu$ and $(m, w) \in \mu_-$, then $(m, w') \in \mu'$ and $(m, w') \in \mu_+$. So $\mu_- = \mu \wedge \mu'$ and $\mu_+ = \mu \vee \mu'$. Thus, a stable matching μ' has been found which violates condition (ii). □

We now turn to the characterization of all of the faces of $P(\Gamma)$.

LEMMA 12. *If two not comparable stable marriages* μ *and* μ' *both saturate an inequality of* $P(\Gamma)$*, then so do and* $\mu_- = \mu \wedge \mu'$ *and* $\mu_+ = \mu \vee \mu'$.

Proof. This is obvious for the nonnegativity, the row, and the column constraints. To see the truth of the statement for the stability constraint associated with (m, w), suppose μ and μ' both saturate it but that μ_+ does not (see Figure 45). By the distributive lattice properties, $\mu_+(w) \leq_w \mu(w)$ and $\mu_+(w) \leq_w \mu'(w)$, and $\mu_+(m) = \mu(m)$ or $= \mu'(w)$, say $\mu_+(m) = \mu(m)$. But then $(m, \mu(m))$ and $(\mu(w), w)$ are successors of (m, w), contradicting the fact that μ saturates the stability constraint associated with (m, w). The argument is similiar for μ_-. □

Given two comparable stable matchings of a marriage graph Γ, $\mu_- <_M \mu_+$, the *interval* of nodes (m, w) in m's row between them are those (women) for which $\mu_-(m) \leq_m w \leq_m \mu_+(m)$, and the interval of nodes (m, w) in w's column between

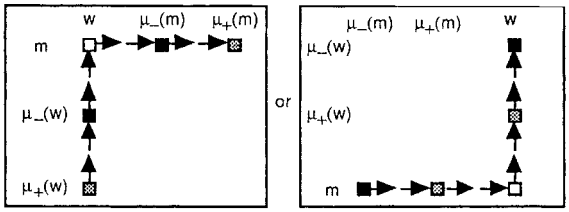

FIG. 46. *μ_+ and μ_- saturate $s(m,w)$.*

them are those (men) for which $\mu_+(w) \leq_w m \leq_w \mu_-(w)$. The subgraph of Γ *between* μ_- and μ_+, called $\Gamma(\mu_-, \mu_+)$, has as its nodes the intervals between μ_- and μ_+ in all lines of Γ. A simple picture shows that $\Gamma(\mu_-, \mu_+)$ is domination-free.

LEMMA 13. *The stable marriages μ of $\Gamma(\mu_-, \mu_+)$ are the stable marriages of Γ that satisfy $\mu_- \leq_M \mu \leq_M \mu_+$.*

Proof. It is clear that a stable marriage μ of Γ that satisfies the inequality is a stable matching of $\Gamma(\mu_-, \mu_+)$.

Suppose, then, that μ is a stable marriage of $\Gamma(\mu_-, \mu_+)$ and that (m,w) is a node of Γ. If $(m,w) \in \Gamma(\mu_-, \mu_+)$, it has a successor in μ. If $(m,w) \notin \Gamma(\mu_-, \mu_+)$, then it may be assumed that either $w <_m \mu_-(m)$ or $\mu_+(m) <_m w$. In the first case, μ, a stable marriage of $\Gamma(\mu_-, \mu_+)$, implies that $(m, \mu_-(m))$ has a successor in μ and so (m,w) does too. The second case is the same with the roles of the sexes interchanged. □

In particular, $\Gamma(\mu_M, \mu_W)$ is a domination-free marriage graph equivalent to Γ.

Given two comparable stable matchings of a marriage graph Γ, $\mu_- <_M \mu_+$, the *hypercube* of μ_- and μ_+, called $H^\Gamma(\mu_-, \mu_+)$, is the set of stable marriages μ of Γ for which $\mu_{|\Gamma_i}$, its restriction to each of Γ's connected subgraphs Γ_i, is either $\mu_{-|\Gamma_i}$ or $\mu_{+|\Gamma_i}$. Indeed, this set is a hypercube when viewed as a subdistributive lattice of the distributive lattice of stable marriages of Γ. To see this, designate the stable marriages of $H^\Gamma(\mu_-, \mu_+)$ by $\mu = \mu_{j_1,\ldots,j_k}$, where $j_i = 1$ if $\mu = \mu_+$ on Γ_i, and $j_i = 0$ if $\mu = \mu_-$ on Γ_i.

One should not jump to any rash conclusions: $H^\Gamma(\mu_-, \mu_+)$ is neither the set of stable matchings of $\Gamma(\mu_-, \mu_+)$ nor the set of stable matchings of $\Gamma(\mu_-, \mu_+)$ that use only man-best or woman-best nodes.

LEMMA 14. *If two comparable stable marriages μ_- and μ_+ both saturate an inequality of $P(\Gamma)$ then so does every $\mu \in H^\Gamma(\mu_+, \mu_-)$.*

Proof. If the inequality is a nonnegativity, a row, or a column constraint, the result is immediate. So suppose it is the stability constraint associated with the node (m,w). Either (1) (m,w) belongs to one of the connected components Γ_i of $\Gamma(\mu_+, \mu_-)$ or (2) (m,w) does not belong to the subgraph $\Gamma(\mu_+, \mu_-)$.

Case (1). Either $\mu = \mu_+$ or μ_- in Γ_i, say μ_+. Every man and every woman represented in Γ_i is matched and $\mu = \mu_+$ in Γ_i. Therefore, $x_{m,w}(\mu) = x_{m,w}(\mu_+)$ for every $(m,w) \in \Gamma_i$, showing that μ and μ_+ either both saturate the stability constraint of (m,w) or neither do.

Case (2). Suppose that $w <_m \mu_-(m) <_m \mu_+(m)$ and the stability constraint of (m,w) is saturated. The fact that μ_- and μ_+ both saturate the constraint means that $(\mu_+(w), w)$ and $(\mu_-(w), w)$ cannot be successors of (m,w), and Figure 46 obtains. But $\mu_- <_M \mu <_M \mu_+$, so the only successors of (m,w) that can take on the value 1 are in m's row between $(m, \mu_-(m))$ and $(m, \mu_+(m))$, showing that μ indeed saturates the constraint. □

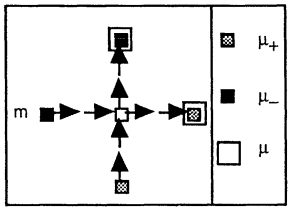

FIG. 47. *μ does not saturate $s(m, w)$.*

LEMMA 15. *$H^\Gamma(\mu_+, \mu_-)$ is a face of $P(\Gamma)$.*

Proof. Since the stable marriages of the hypercube $H^\Gamma(\mu_+, \mu_-)$ saturate the same set of inequalities, it is only necessary to show that any other stable marriage $\mu \notin H^\Gamma(\mu_+, \mu_-)$ fails to saturate some inequality of that set.

If $\mu(m)$ is different from $\mu_-(m)$ and from $\mu_+(m)$ for some one man m, then the nonnegativity constraint $x_{m,\mu(m)} \geq 0$ does the trick. So it may be assumed that $\mu(m) = \mu_-(m)$ or $\mu_+(m)$ for all men $m \in M$.

But μ is different from μ_- and μ_+, so μ must be different from μ_- and μ_+ on some connected component Γ_i of $\Gamma(\mu_+, \mu_-)$. Consider the equivalent domination-free subgraph of Γ_i, and let Γ_i' be one of its connected subgraphs where μ is different from μ_- and μ_+. In Γ_i' the nodes of μ_+ are man-best and woman-worst, the nodes of μ_- are man-worst and woman-best, and μ sometimes (at least once) agrees with μ_+ and sometimes (at least once) with μ_-. Take any node (m, w) of Γ_i' where $w = \mu(m) = \mu_+(m)$, and follow the path that alternates from the man-best node in man m's row to the woman-best node in woman $\mu_+(m)$'s column, . . . , until a cycle C^+ is formed: $\mu(m) = \mu_+(m)$ for every m in C^+. Similarly, taking any node (m, w) of Γ_i', where $w = \mu(m) = \mu_-(m)$, a cycle C^- is formed where μ and μ_- agree. Continue to do this, forming cycles of type C^+ and type C^- until all nodes $(m, \mu_+(m))$ and $(m, \mu_-(m))$ are included. Since Γ_i' is connected, a cycle of type C^+ must intersect with a cycle of type C^-, so the situation pictured in Figure 47 obtains. Observe that μ_- and μ_+ both saturate the stability constraint, whereas μ does not. □

THEOREM 11. *The faces F of $P(\Gamma)$ are the sets of stable marriages satisfying the following:*

(i) *for every pair of not comparable stable marriages μ and μ' in F, $\mu_- = \mu \wedge \mu'$ and $\mu_+ = \mu \vee \mu'$ are also in F;*

(ii) *for every pair of comparable stable marriages μ_-, $\mu_+ \in F$, $H^\Gamma(\mu_+, \mu_-)$ is in F.*

Proof. Previous lemmas have established the necessity of the conditions. To establish sufficiency, suppose the opposite: namely, the conditions are satisfied by F, but there exists another stable matching not in F that also saturates the inequalities saturated by all of the stable marriages of F. More specifically, within the class of all such stable matchings (assumed to be nonempty), take a $\mu^* \in F$ that is inferior to no other member of the class in the view of the men.

The first condition implies that F is a lattice. Let $\mu_+ \in F$ be the supremum of the elements of F and $\mu_- \in F$ be the infimum, so $\mu_- <_M \mu_+$. It must be that $\mu_- <_M \mu^* <_M \mu_+$; otherwise, for some man $m \in M$, either $\mu^*(m) <_m \mu_-(m)$ or $\mu_+(m) <_m \mu^*(m)$. If $\mu^*(m) <_m \mu_-(m)$, then $x_{m,\mu^*(m)}(\mu) = 0$ for all $\mu \in F$, whereas $x_{m,\mu^*(m)}(\mu^*) = 1$, contradicting the choice of μ^*. The second possibility is similarly dispatched.

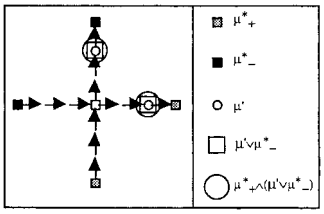

FIG. 48. *μ'' does not saturate $s(m, w)$.*

Let μ^*_+ be the infimum of the elements of F that are superior to μ^* and μ^*_- be the supremum of the elements of F that are inferior to μ^*, so $\mu^*_- <_M \mu^* <_M \mu^*_+$. It will be shown that for every man $m \in M, \mu^*(m)$ is equal to either $\mu^*_-(m)$ or $\mu^*_+(m)$.

Suppose that for some man $m \in M$, $\mu^*_-(m) <_m \mu^*(m) <_m \mu^*_+(m)$. Then there must exist a stable matching $\mu' \in F$ with $\mu'(m) = \mu^*(m)$; otherwise, $x_{m,\mu^*(m)} \geq 0$ would be saturated by all elements of F but not by μ^*. Thus, $\mu^*_-(m) <_m \mu'(m) <_m \mu^*_+(m)$. Define $\mu'' = \mu^*_+ \wedge (\mu' \vee \mu^*_-)$: by condition (i) it belongs to F; by definition $\mu^*_- <_M \mu'' <_M \mu^*_+$ and $\mu''(m) = \mu'(m) = \mu^*(m)$. But μ'' and μ^* are not comparable; otherwise, if, for example, $\mu'' >_M \mu^*$, then μ'' would have been involved in the defintion of μ^* and implied $\mu'' \geq_M \mu^*_+$, which is a contradiction.

Now define $\mu^{**} = \mu'' \vee \mu^*$. Clearly, $\mu^{**} <_M \mu^*_+$. By definition and the fact that μ'' and μ^* are not comparable, $\mu^{**} >_M \mu^*$. Furthermore, μ^{**} cannot belong to F, for if it did it would have been involved in the definition of μ^*_+ and implied $\mu^{**} \geq_M \mu^*_+$, which is again a contradiction. Finally, by Lemma 11, μ^{**} saturates the constraints that the stable matchings of F saturate, since both μ^* and μ'' do. Therefore, $\mu^*_- <_M \mu^* < M\mu^{**} < M\mu^*_+$, which contradicts the original choice of μ^*.

One concludes that, indeed, for every man m, $\mu^*(m)$ is either equal to $\mu^*_-(m)$ or to $\mu^*_+(m)$, though μ^* is neither μ^*_- nor μ^*_+. But this is exactly the situation encountered in the proof of Lemma 14, where it was shown that in this case there exists a stability constraint associated with some (m, w) in $\Gamma(\mu^*_-, \mu^*_+)$ that is saturated by μ^*_- and μ^*_+ but not by μ^*.

The line of argument is now familiar. There must exist a stable matching in F, call it once again $\mu' \in F$, that does not saturate the stability constraint of (m, w). Define (using again the same name) $\mu'' = \mu^*_+ \wedge (\mu' \vee \mu^*_-)$. By defintion, μ'' belongs to F, and $\mu^*_- <_M \mu'' <_M \mu^*_+$. Figure 48 shows that it too does not saturate the stability constraint of (m, w).

The stable matching $\mu^{**} = \mu'' \vee \mu^*$ does not saturate the stability constraint of (m, w) either, but it does saturate the inequalities saturated by all of the stable marriages of F and satisfies $\mu^*_- <_M \mu^* <_M \mu^{**} <_M \mu^*_+$. This final contradiction to the definition of μ^* completes the proof. □

The foregoing development yields the concluding observation.

THEOREM 12. *The faces of the matching polytope $P(\Gamma)$ are determined by the lattice of the stable marriages $\mathcal{L}(\Gamma)$.*

Proof. A face is nothing more than a sublattice of $\mathcal{L}(\Gamma)$ with the added property that for every pair of comparable stable marriages μ_- and μ_+ of F, the hypercube $H^\Gamma(\mu_+, \mu_-)$ belongs to F as well. This information is completely contained in the lattice $\mathcal{L}(\Gamma)$ itself. □

It has already been noted that every distributive lattice may be realized as the ordering of the stable marriages of some marriage game. But it is clear that there exist different marriage games having one and the same lattice $\mathcal{L}$ and, so, one and the same matching polytope P.

9. Remarks on the history of the marriage game. David Gale formulated the problem, postulated the existence of stable matchings, and with Lloyd Shapley established existence with the propose–dispose algorithm, so they naturally recognized the optimality of solutions for the men and for the women. Their joint paper [4] started it all. Unbeknownst to them, the same ideas had entered into practical use years before. The Association of American Medical Colleges introduced in 1951 what is now called the "national resident matching program" for assigning medical graduates to residencies in hospitals based only on the ordered preferences of graduates for hospitals and of hospitals for graduates: their program produces the hospital-optimal solution, not the graduates-optimal solution! For a detailed account of the history of the problem and its many applications in economics, as well as a careful attribution of who contributed what, the reader is referred to the basic reference, the comprehensive and prize-winning book of Roth and Sotomayor [17]. The references given below include the results we specifically develop.

Maffray [10] was the first to recognize that a stable matching is simply the kernel of a graph. The systematic use of this idea is our central tool and was used in the doctoral dissertation of Ratier [13] and his forthcoming paper [14].

The systematic opposition of the interests of men and women was pointed out in Knuth [9], whereas the fact that the stable matchings form a lattice seems to have first been noticed by John Conway (see [9]). Blair [1] proved that every finite distributive lattice may be represented by a set of stable matchings, but the marriage games he needed could be very big indeed; Gusfield et al. [8] showed how to do it with more modest sized marriage games. The fact that if a player is celibate in one stable matching then she is in every stable matching was first established by McVitie and Wilson [11]. The other properties of stable matchings in section 4 come from the papers of Gale and Sotomayor [5], [6].

Dubins and Freedman [3] showed that under the "men propose and women dispose" mechanism it pays no man or subset of the men to misrepresent their preferences. That a coalition of one man and one woman can profitably cheat was established with linear programming arguments in Ratier's thesis [13]. The "limitations" to the profits in cheating was proven by Demange, Gale, and Sotomayor [2].

The stable matching polytope was first established in the absence of celibates by Vande Vate [19] and then generalized to its present definition by Rothblum [18]. The unified treatment of the subject using linear programming is in the joint paper of Roth, Rothblum, and Vande Vate [16]. The results concerning extreme points and faces are by and large in Ratier's thesis [13] and his forthcoming paper [14].

The three books referred to in the introduction are Knuth's book [9], Gusfield and Irving's book [7], and Roth and Sotomayor's book [17]. The first two have a more computational-computer science flavor, whereas the last may be described as a book in theoretical and applied economics. The papers of Mongell and Roth [12] and of Roth [15] give interesting applications of the model to the analysis of the historical evolution of particular two-sided markets.

REFERENCES

[1] C. Blair, *Every finite distributive lattice is a set of stable matchings*, J. Combin. Theory Ser. A, 37 (1984), pp. 353–356.

[2] G. Demange, D. Gale, and M. Sotomayor, *A further note on the stable matching problem*, Discrete Appl. Math., 16 (1987), pp. 217–222.

[3] L. E. Dubins and D. A. Freedman, *Machiavelli and the Gale-Shapley algorithm*, Amer. Math. Monthly, 88 (1981), pp. 485–494.

[4] D. Gale and L. S. Shapley, *College admissions and the stability of marriage*, Amer. Math. Monthly, 69 (1962), pp. 9–15.
[5] D. Gale and M. Sotomayor, *Ms Machiavelli and the stable matching problem*, Amer. Math. Monthly, 92 (1985), pp. 261–268.
[6] D. Gale and M. Sotomayor, *Some remarks on the stable matching problem*, Discrete Appl. Math., 11 (1985), pp. 223–232.
[7] D. Gusfield and R. W. Irving, *The Stable Marriage Problem: Structure and Algorithms*, MIT Press, Cambridge, MA, 1989.
[8] D. Gusfield, R. Irving, P. Leather, and M. Saks, *Every finite distributive lattice is a set of stable matchings for a small stable marriage instance*, J. Combin. Theory Ser. A, 44 (1987), pp. 304–309.
[9] D. E. Knuth, *Mariages stables et leurs relations avec d'autres problèmes combinatoires,* Les Presses de l'Université de Montréal, Montréal, 1976.
[10] F. Maffray, *Kernels in perfect line-graphs*, J. Combin. Theory Ser. B, 55 (1992), pp. 1–8.
[11] D. G. McVitie and L. B. Wilson, *Stable marriage assignments for unequal sets*, BIT, 10 (1970), pp. 259–309.
[12] S. J. Mongell and A. E. Roth, *Sorority rush as a two-sided matching mechanism*, Amer. Economic Rev., 81 (1991), pp. 441–464.
[13] G. Ratier, *Les Mariages Stables: Graphes et Programmation Linéaire*, Doctoral thesis, Université de Paris 1, Paris, France, 1995.
[14] G. Ratier, *On the stable marriage polytope*, Discrete Math., to appear.
[15] A. E. Roth, *A natural experiment in the organization of entry level labor markets: Regional markets for new physicians and surgeons in the U.K.*, Amer. Economic Rev., 81 (1991), pp. 415–440.
[16] A. E. Roth, U. G. Rothblum, and J. H. Vande Vate, *Stable matchings, optimal assignments, and linear programming*, Math. Oper. Res., 18 (1993), pp. 803–828.
[17] A. E. Roth and M. A. O. Sotomayor, *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis,* Cambridge University Press, Cambridge, 1990.
[18] U. Rothblum, *Characterization of stable matchings as extreme points of a polytope*, Math. Programming, 54 (1992), pp. 57–67.
[19] J. H. Vande Vate, *Linear programming brings marital bliss*, Oper. Res. Lett., 8 (1989), pp. 147–153.