Incentive-Compatibility of MPP in Matching Problems

__Theorem 3__ When the men-proposing procedure (MPP) is followed, no coalition of men can make themselves better off by (jointly) misrepresenting their preferences.

__Corollary__(*) (restricting to coalitions of size one) When MPP is followed, it is a weakly dominant strategy for each man to reveal his true preference

The proof of Theorem 3 DOES require a little notation and is developed in the next 4 pages.

(*) The corollary is in: Dubins, L.E., and D.A. Freedman "Machiavelli and the Gale-Shapley algorithm", __American Mathematical Monthly__, 88, 485-94

See also: Gale, D., and M. Sotomayor "Ms. Machiavelli and the stable matching problem" __American Mathematical Monthly__, 92, 261-8

Proof of Thm 3

Let $M \equiv$ the set of men
$W \equiv$ " " " women

Consider an arbitrary profile of preferences $P = (P_i)_{i \in M \cup W}$, and let $P' = (P'_i)_{i \in M \cup W}$ be any other profile which is the same as P except that a coalition of men $M' \subset M$ misrepresent their true preferences in P (i.e., $P'_i = P_i$ for $i \in (M \setminus M') \cup W$ and each $i \in M'$ is viewed as switching from his true preference P_i to a misrepresentation P'_i)

Denote by μ the matching obtained via MPP under P
" " μ' " " " " " P'

Define $\tilde{M}$ to be the set of men who prefer their wives in μ' to their wives in μ
(in symbols, letting $\mu(i) \equiv$ "spouse of i in μ" and $\alpha P_i \beta \equiv$ "α is preferred to β in P_i", we have $\tilde{M} = \{i \in M : \mu'(i) P_i \mu(i)\}$)

<u>SUPPOSE</u> $M' \subset \tilde{M}$ (i.e., each member of the coalition M' is benefited when they jointly misrepresent $(P_i)_{i \in M'}$ by $(P'_i)_{i \in M'}$)
We shall show that this leads to a contradiction.

~~Notation $\mu(i)$ = spouse of i in μ~~
~~(contd) $\mu(i)$~~

Case 1 $\mu'(\tilde{M}) \neq \mu(\tilde{M})$

Let $w \in \mu'(\tilde{M}) \setminus \mu(\tilde{M})$ and let the husband of w in μ, μ' be denoted m, m'.
Since $w \in \mu'(\tilde{M})$, we have $m' \in \tilde{M}$ and so

$$w \, P_{m'} \, \mu(m')$$

This implies that, in the MPP under P, m' proposes to w (before he proposes to $\mu(m')$) and is rejected by w. But then

$$\boxed{m \, P_w \, m'}$$

On the other hand, $w \notin \mu(\tilde{M})$ implies $m \notin \tilde{M}$ and so

$$w \, P_m \, \mu'(m)$$

But, since $M' \subset \tilde{M}$ and $m \notin \tilde{M}$, we have $m \notin M'$ and so $P_m = P'_m$. Thus the display above may be rewritten

$$\boxed{w \, P'_m \, \mu'(m)}$$

The two boxed displays show that μ' is not stable under P', a contradiction to Theorem 1

Case 2 $\mu'(\tilde{M}) = \mu(\tilde{M}) = W^*$

Focus throughout on the MPP under P

First observe that each woman $w \in W^*$ has different husbands m, m' in μ, μ' since — by the definition of $\tilde{M}$ — each man in $\tilde{M}$ is better off in μ' than in μ, and so cannot have the same wife in the two matchings. Moreover observe that each $w \in W^*$ receives proposals from, and rejects, men in $\tilde{M}$ (her husband m' being one such man).

Let w (from now on) denote one of the LAST women in W^* to receive a proposal from a man in $\tilde{M}$. Since w rejects $m' \in \tilde{M}$ as we saw, she is not without a partner at the time of this proposal. Denote by m^* the partner and by $\bar{m} \in \tilde{M}$ the proposer.

If $\bar{m}$ is rejected by w, then $\bar{m}$ would go on to propose to $\mu(\bar{m}) \in W^*$ at a later date, contradicting that w receives (one of) the last proposals from $\tilde{M}$. So m^* is rejected by w in favor of $\bar{m}$ (and, in fact, $\bar{m} = m$) ~~though that is of no concern to anything~~. But then $m^* \notin \tilde{M}$, otherwise m^* would again propose to $\mu(m^*) \in W^*$ at a later date, contradicting the definition of w.

Since w is engaged to m^* after (~~same time~~) rejecting m', we have

$$\boxed{m^* P_w \, m'} \equiv \boxed{m^* P'_w \, \mu'(w)}$$

Since $m^* \notin \tilde{M}$, we get

$$\mu(m^*) \, P_{m^*} \, \mu'(m^*) \, , \text{ or } \mu(m^*) = \mu'(m^*)$$

But m^* was with w before being with $\mu(m^*)$, hence

$$w \, P_{m^*} \, \mu(m^*)$$

The last two displays yield

$$w \, P_{m^*} \, \mu'(m^*)$$

which (recalling that $P_{m^*} \equiv P'_{m^*}$ since $m^* \notin \tilde{M}$) may be rewritten

$$\boxed{w \, P'_{m^*} \, \mu'(m^*)}$$

The two boxed displays contradict that μ' is stable under P', and thus contradict Theorem 1

QED

The lattice structure of stable matchings

Fix $P = (P_i)_{i \in M \cup W}$.
Let μ and μ' be two stable matchings.
Define $\lambda \equiv \mu \vee_M \mu'$ (as a map from M to W)
by $\lambda(m) = \begin{cases} \mu(m) & \text{if } \mu(m)\, P_m\, \mu'(m) \\ \mu'(m) & \text{otherwise} \end{cases}$

Similarly define $\tilde{\lambda} \equiv \mu \wedge_M \mu'$ exactly as above replacing "$\mu(m)\, P_m\, \mu'(m)$" by "$\mu'(m)\, P_m\, \mu(m)$"
(In words: λ, $\tilde{\lambda}$ assigns to each man the better, worse of his spouses in μ, $\tilde{\mu}$)
Similarly define $\mu \vee_W \mu'$ and $\mu \wedge_W \mu'$

IIW 1 <u>Theorem</u> (Conway) Both $\mu \vee_M \mu'$ and $\mu \wedge_M \mu'$ are stable matchings. Moreover
$\mu \vee_M \mu' = \mu \wedge_W \mu'$ and $\mu \wedge_M \mu' = \mu \vee_W \mu'$

<u>Proof</u> Straightforward and left as an exercise.

(<u>Remark</u>: The lattice is distributive i.e.
$$\mu \vee_M (\mu' \wedge_M \mu'') = (\mu \vee_M \mu') \wedge_M (\mu \vee_M \mu'')$$
and
$$\mu \wedge_M (\mu' \vee_M \mu'') = (\mu \wedge_M \mu') \vee_M (\mu \wedge_M \mu'')$$
Indeed, <u>every</u> finite distributive lattice arises from a two-sided matching problem