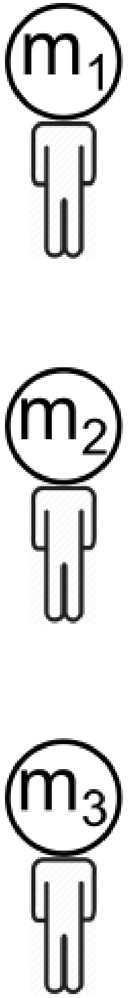
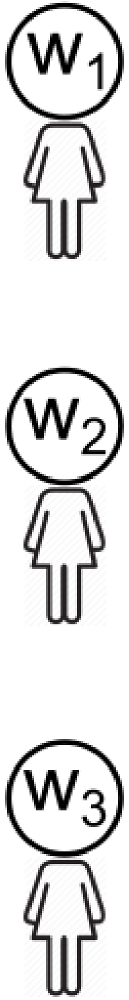
Lecture 18

Manipulation (Contd.) and Fairness in the Stable Matching Problem

Apr 09, 2026 | Rohit Vaish

Stable Matching Problem





m_1: $w_1 > w_2 > w_3$

m_2: $w_2 > w_1 > w_3$

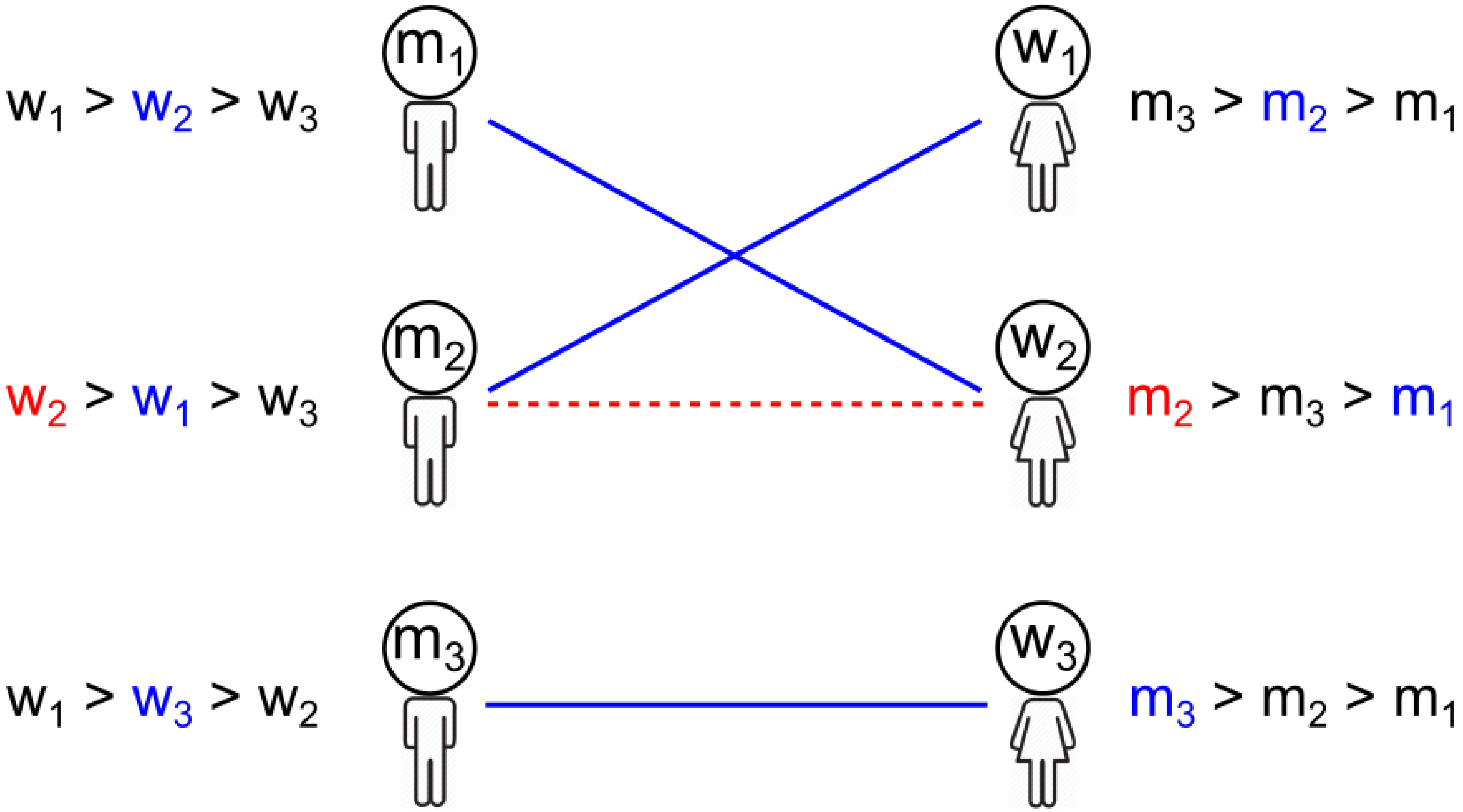
m_3: $w_1 > w_3 > w_2$

w_1: $m_3 > m_2 > m_1$

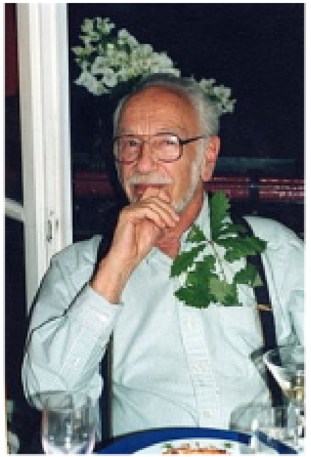
w_2: $m_2 > m_3 > m_1$

w_3: $m_3 > m_2 > m_1$

Stable Matching Problem



A matching is stable if there is no blocking pair.

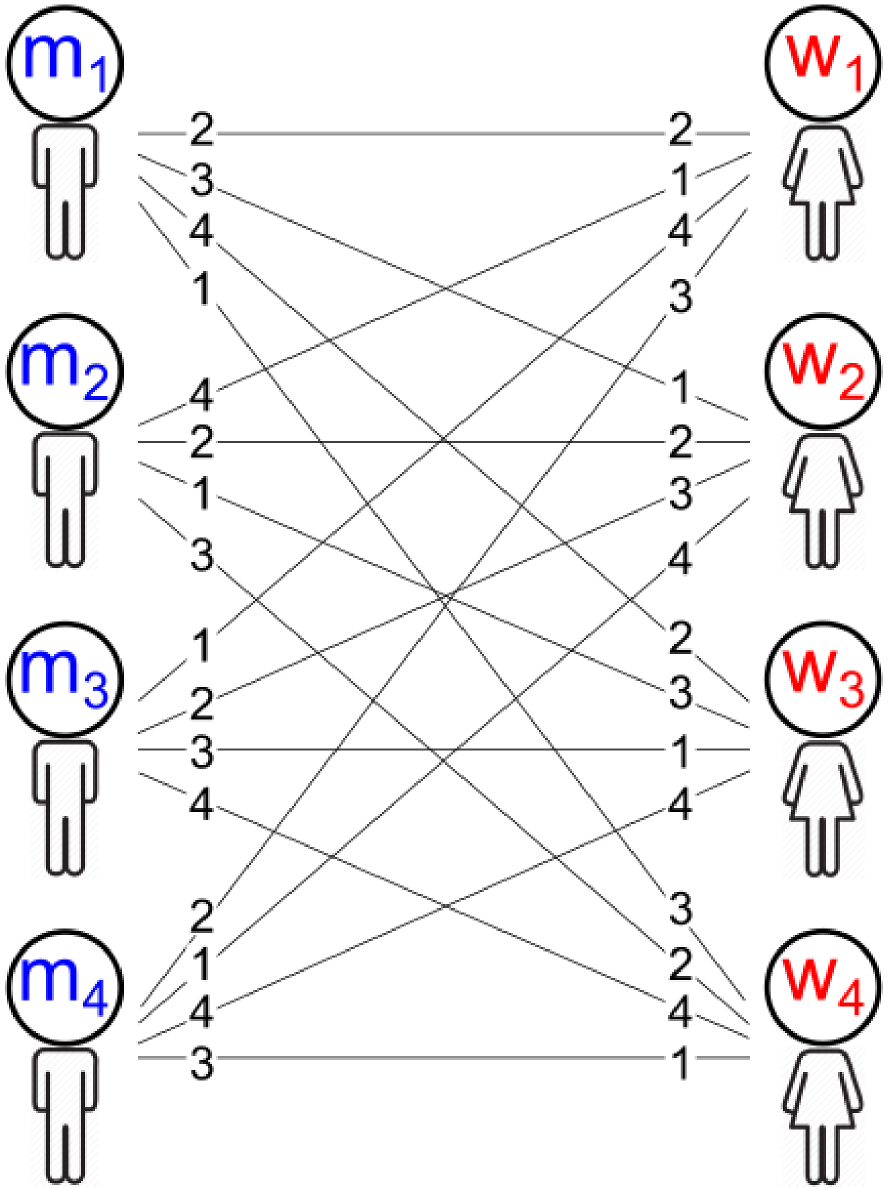


COLLEGE ADMISSIONS AND THE STABILITY OF MARRIAGE

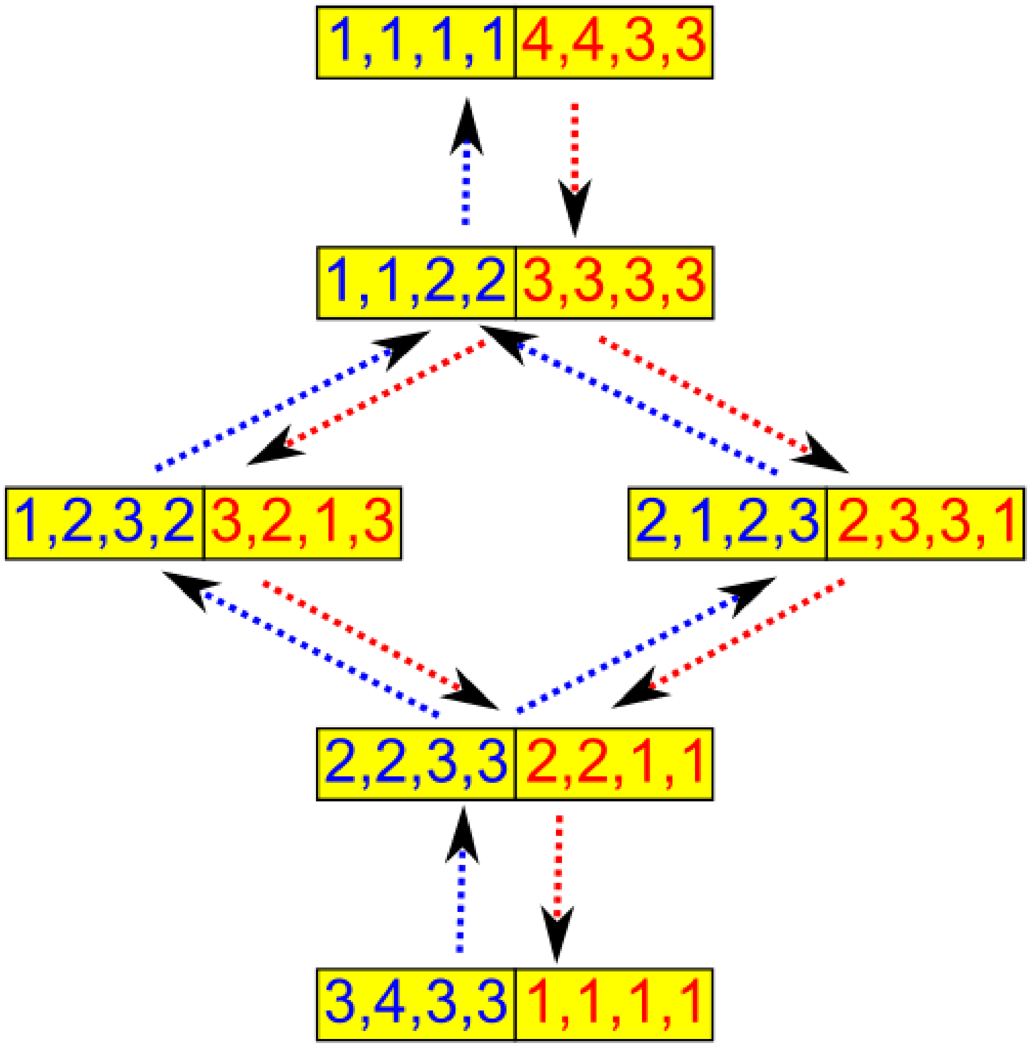
D. GALE* AND L. S. SHAPLEY, Brown University and the RAND Corporation

Source: *The American Mathematical Monthly*, Jan., 1962, Vol. 69, No. 1 (Jan., 1962), pp. 9–15

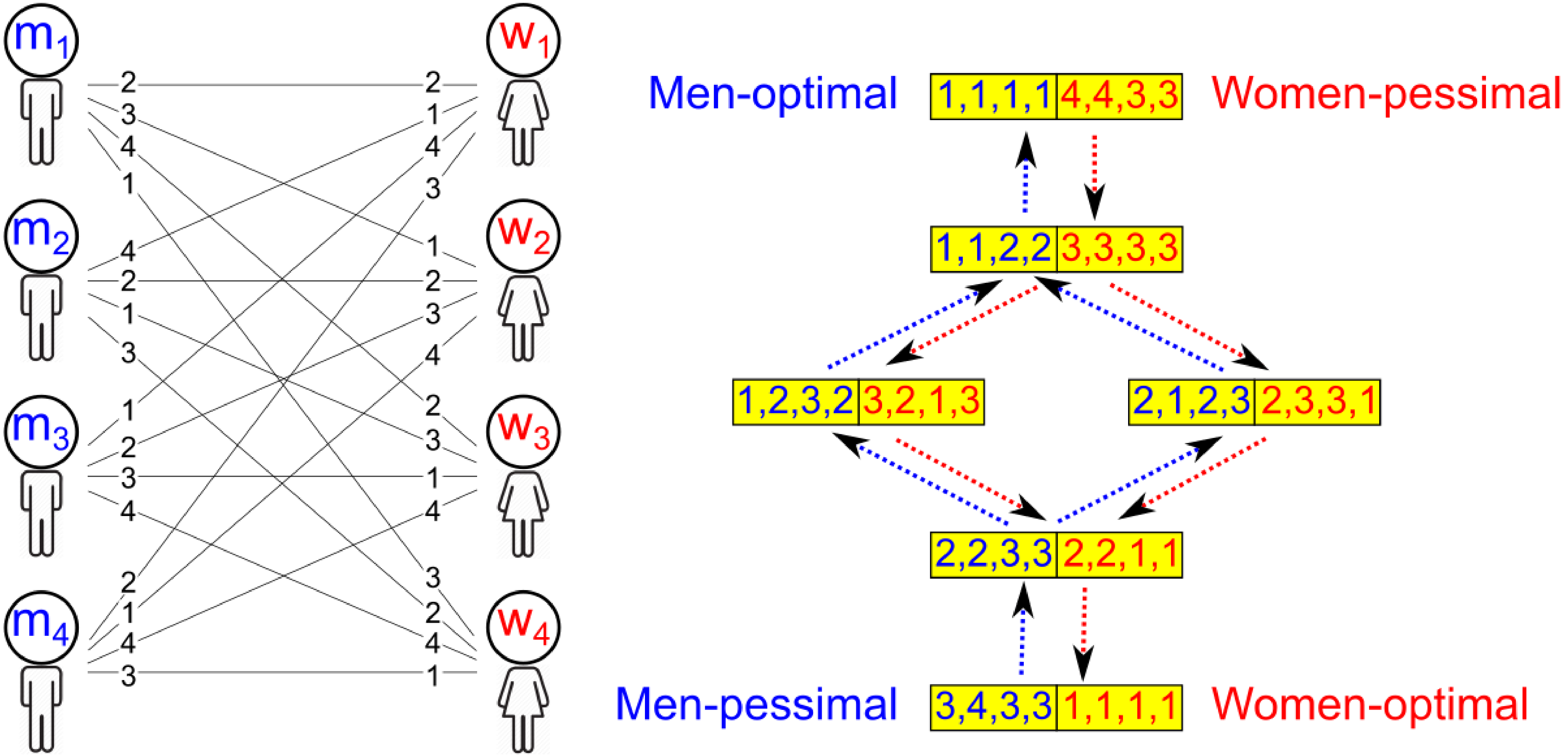
Given any preference profile, a stable matching for that profile always exists and can be computed in polynomial time.



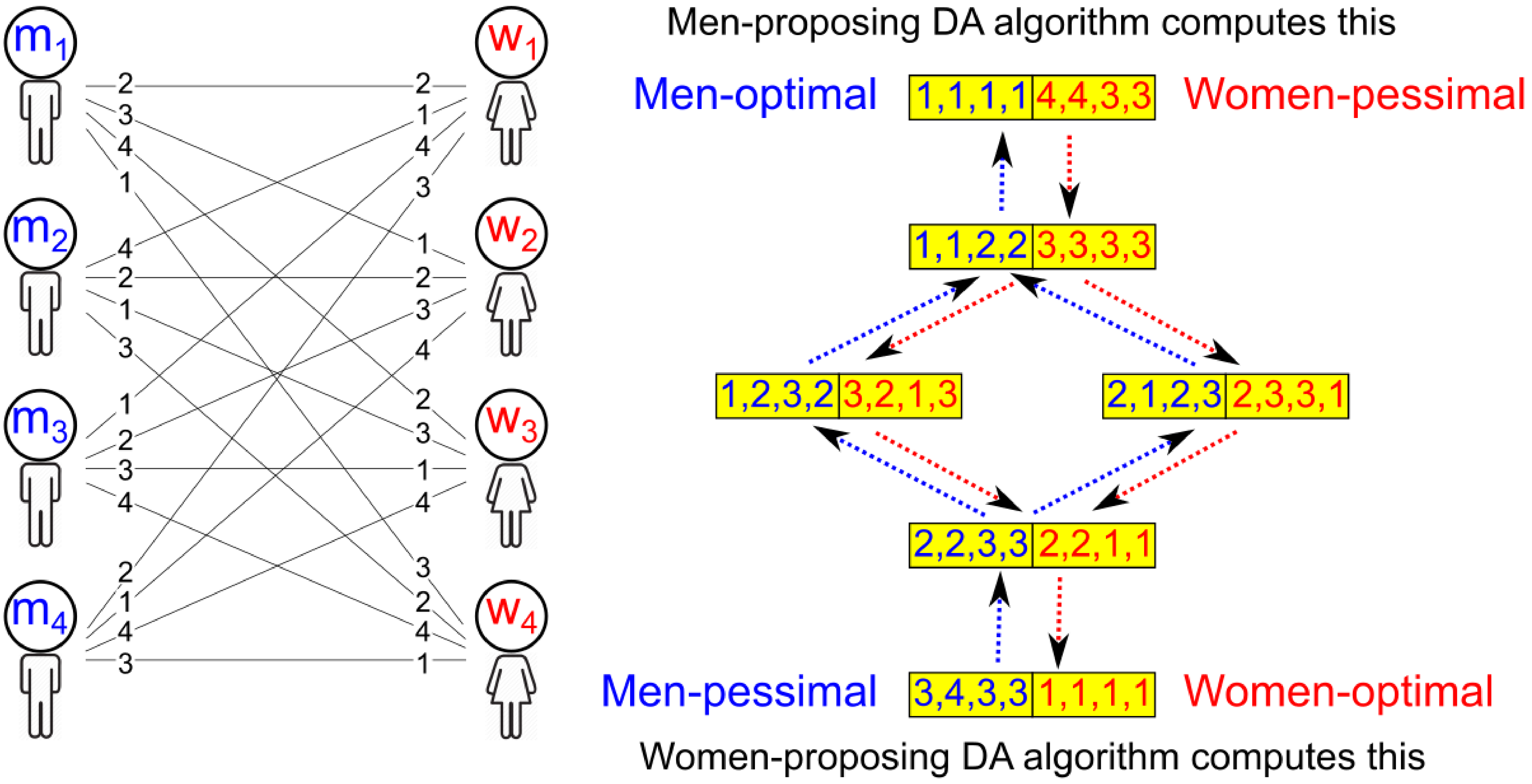
m_1
2 3 4 1
m_2
4 2 1 3
m_3
1 2 3 4
m_4
2 1 4 3
w_1
2 1 4 3
w_2
1 2 3 4
w_3
2 3 1 4
w_4
3 2 4 1



1,1,1,1 4,4,3,3
1,1,2,2 3,3,3,3
1,2,3,2 3,2,1,3
2,1,2,3 2,3,3,1
2,2,3,3 2,2,1,1
3,4,3,3 1,1,1,1



m1
m2
m3
m4
w1
w2
w3
w4
Men-optimal
1,1,1,1
4,4,3,3
Women-pessimal
1,1,2,2
3,3,3,3
1,2,3,2
3,2,1,3
2,1,2,3
2,3,3,1
2,2,3,3
2,2,1,1
Men-pessimal
3,4,3,3
1,1,1,1
Women-optimal


m1
m2
m3
m4
w1
w2
w3
w4
Men-proposing DA algorithm computes this
Men-optimal
1,1,1,1 4,4,3,3
Women-pessimal
1,1,2,2 3,3,3,3
1,2,3,2 3,2,1,3
2,1,2,3 2,3,3,1
2,2,3,3 2,2,1,1
Men-pessimal
3,4,3,3 1,1,1,1
Women-optimal
Women-proposing DA algorithm computes this

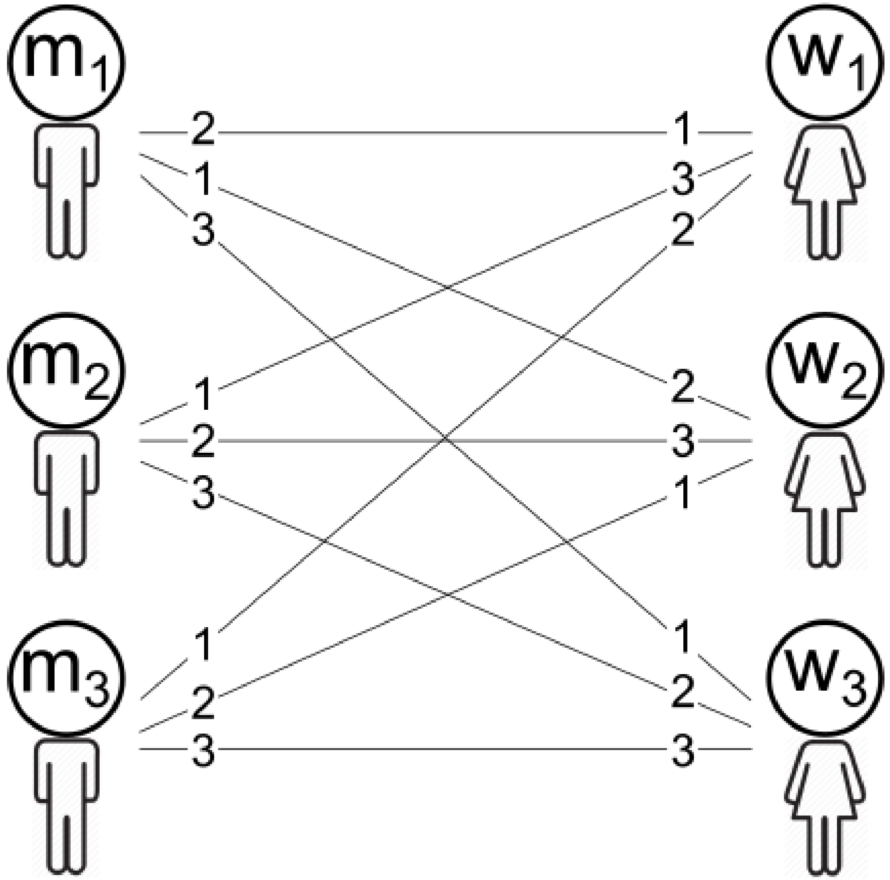
Can an agent get a better partner under DA algorithm by misreporting his/her preferences?

[Dubins and Freedman, *Amer. Math. Mon.* 1981; Roth, *MOR* 1982]

DA algorithm is not strategyproof.

[Dubins and Freedman, *Amer. Math. Mon.* 1981; Roth, *MOR* 1982]

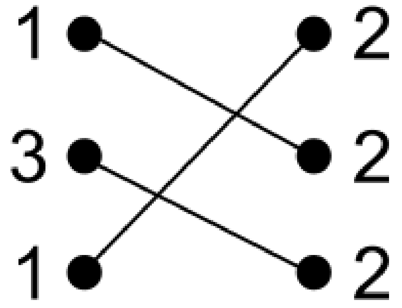
DA algorithm is not strategyproof.

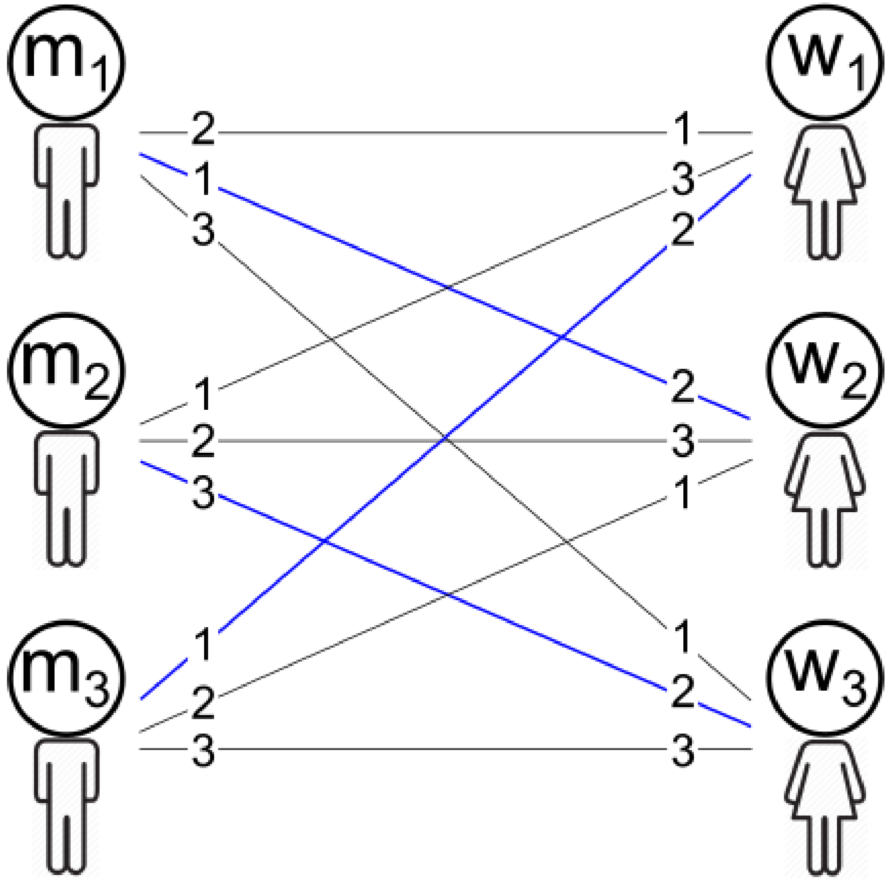


[Dubins and Freedman, *Amer. Math. Mon.* 1981; Roth, *MOR* 1982]

DA algorithm is not strategyproof.

Outcome for true prefs



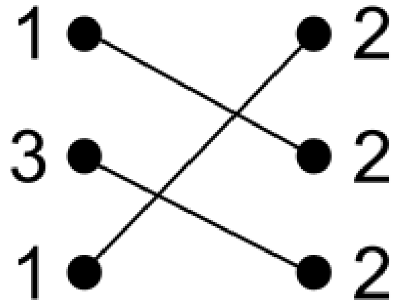


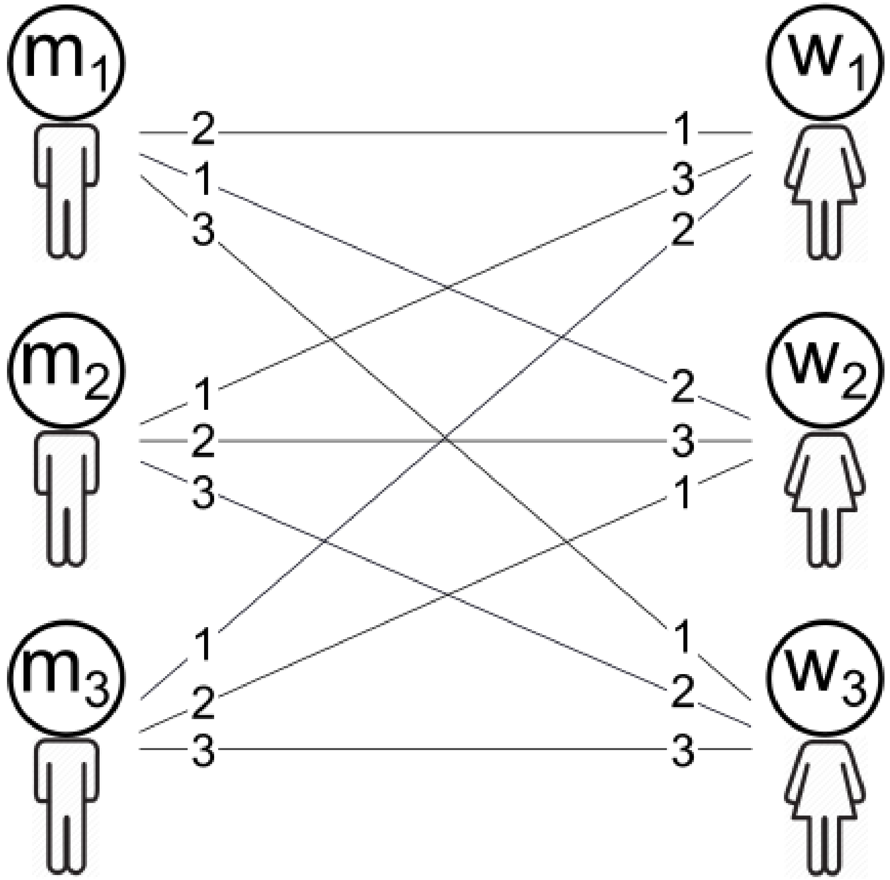
[Dubins and Freedman, *Amer. Math. Mon.* 1981; Roth, *MOR* 1982]

DA algorithm is not strategyproof.

Can I get a better partner by misreporting my preferences?

Outcome for true prefs

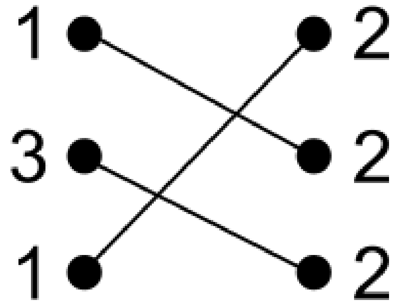


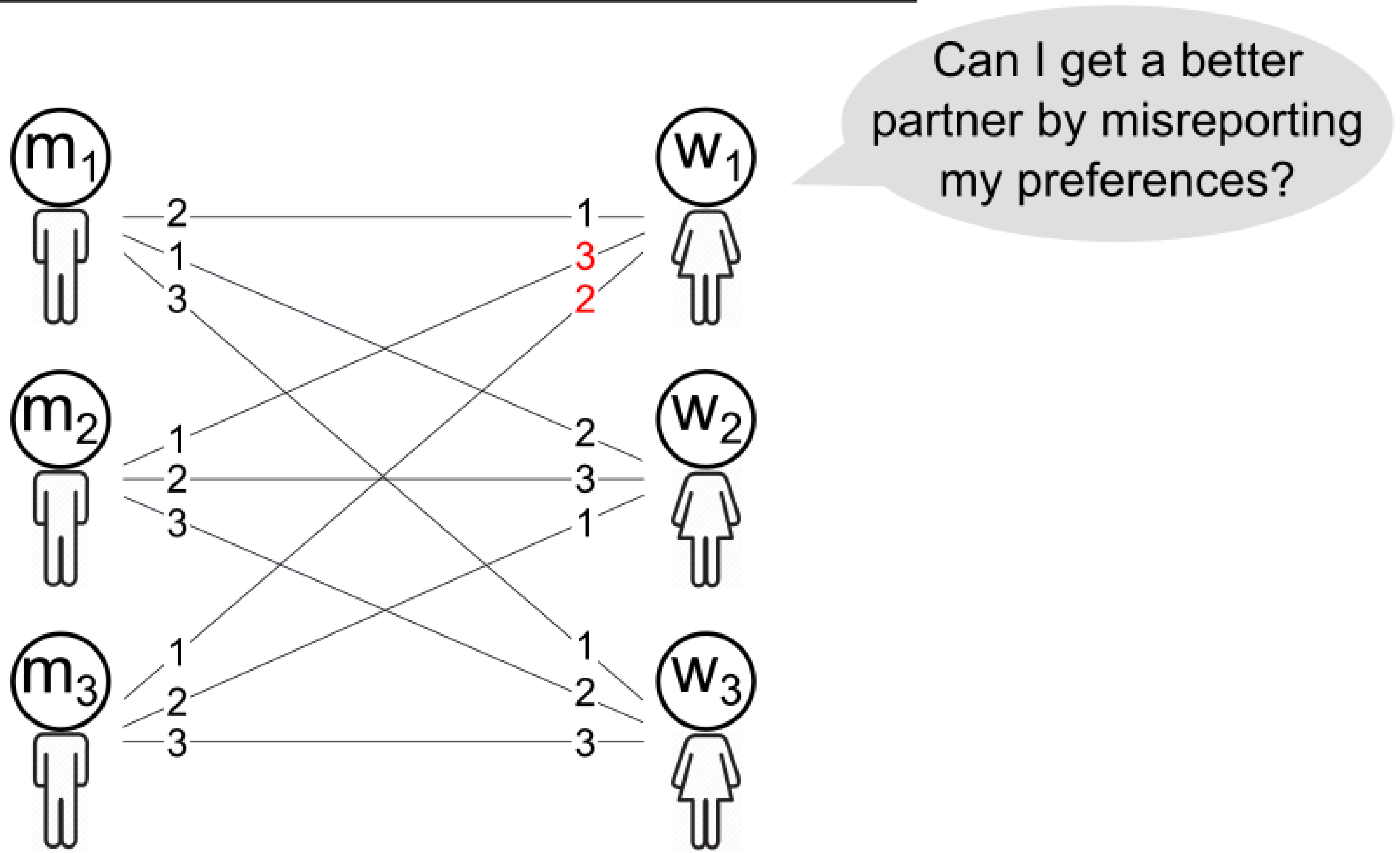


[Dubins and Freedman, *Amer. Math. Mon.* 1981; Roth, *MOR* 1982]

DA algorithm is not strategyproof.

Outcome for true prefs

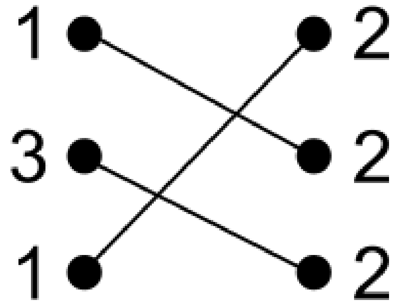


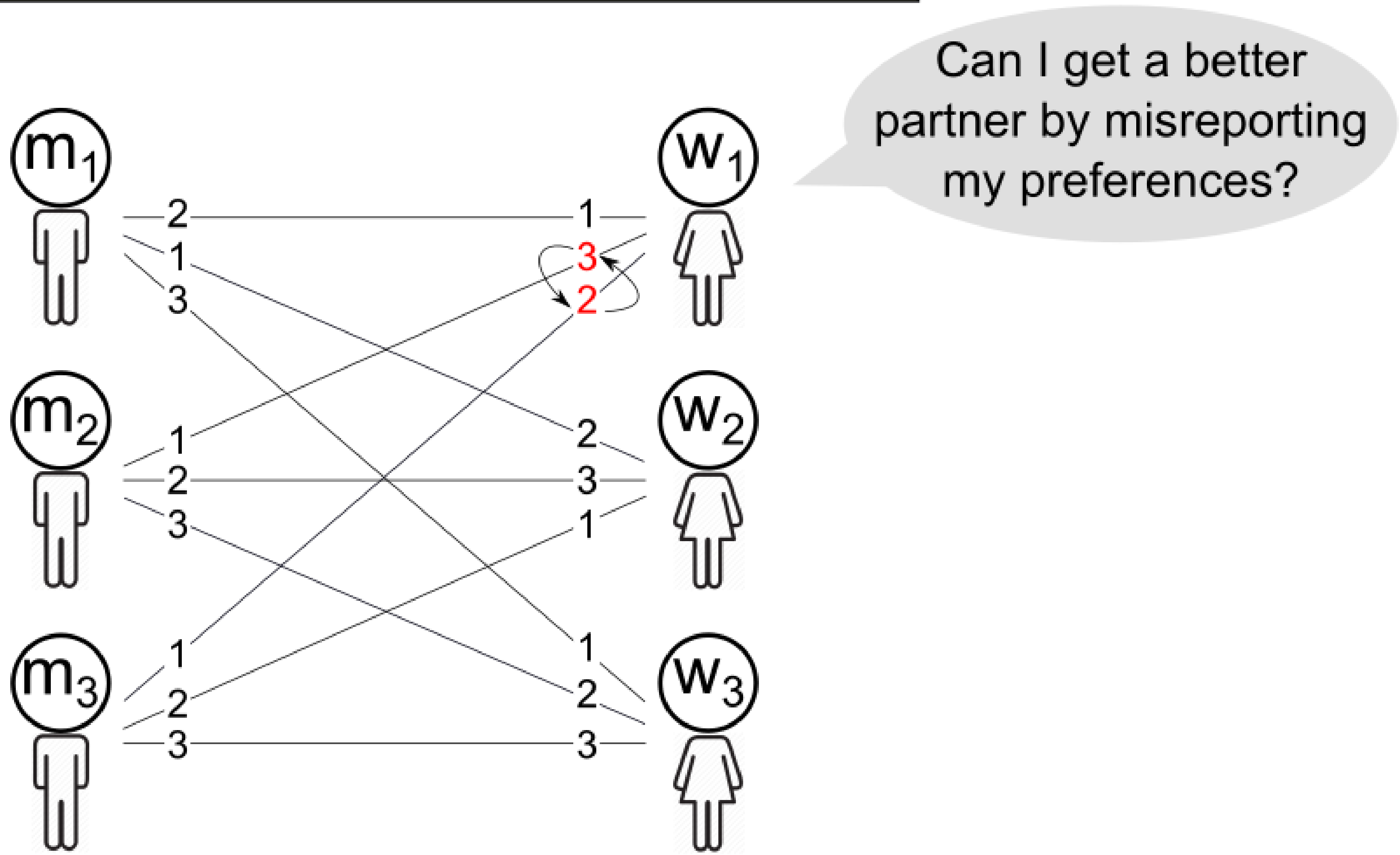


[Dubins and Freedman, *Amer. Math. Mon.* 1981; Roth, *MOR* 1982]

DA algorithm is not strategyproof.

Outcome for true prefs

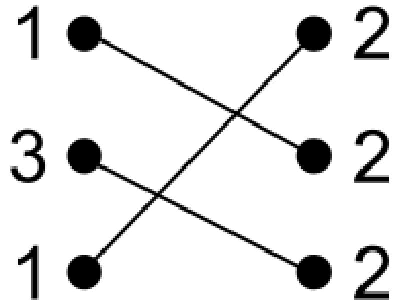


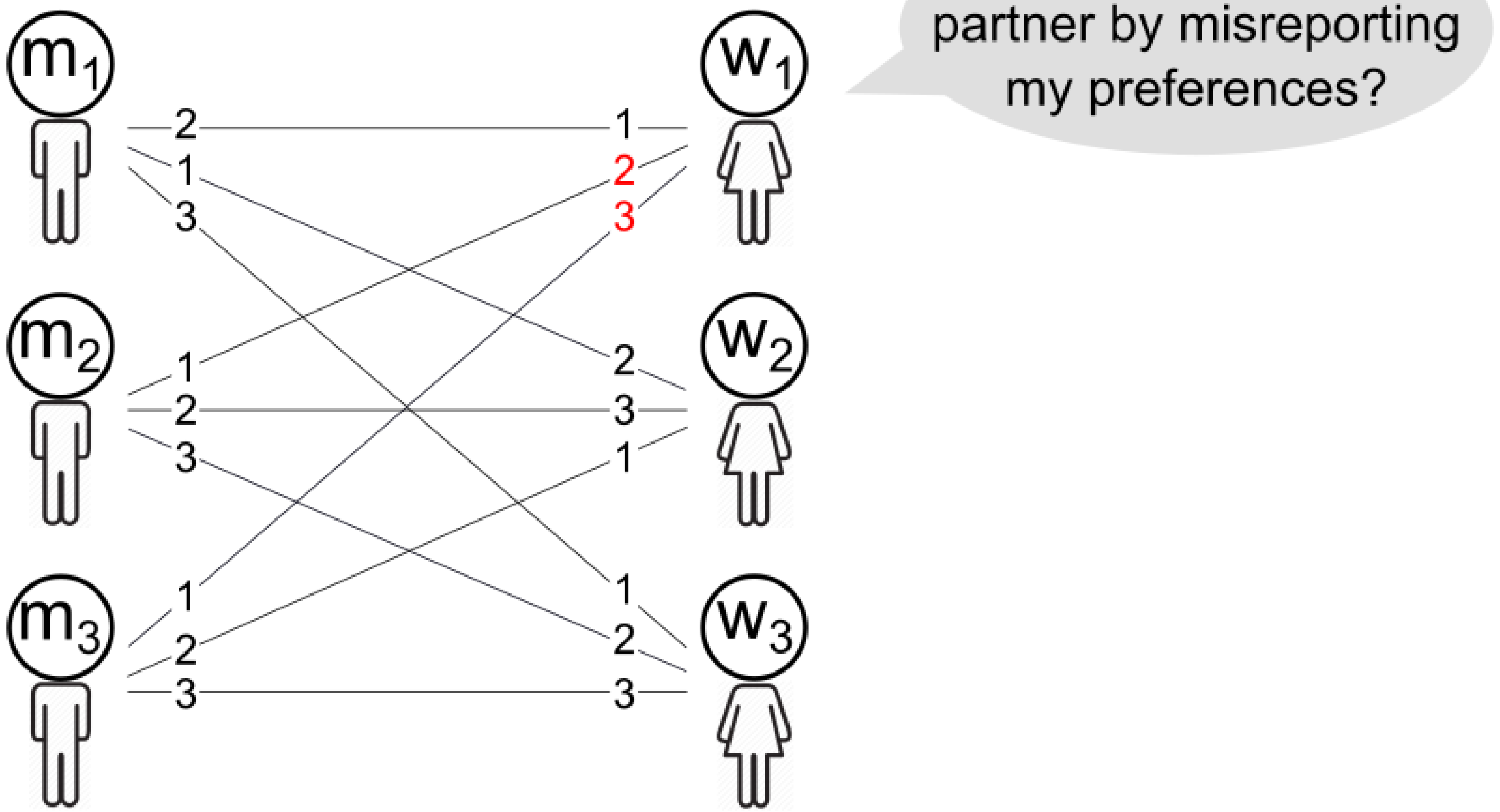


[Dubins and Freedman, *Amer. Math. Mon.* 1981; Roth, *MOR* 1982]

DA algorithm is not strategyproof.

Outcome for true prefs

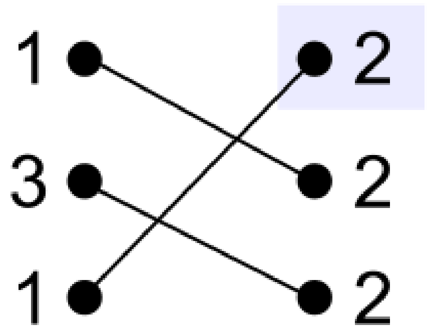


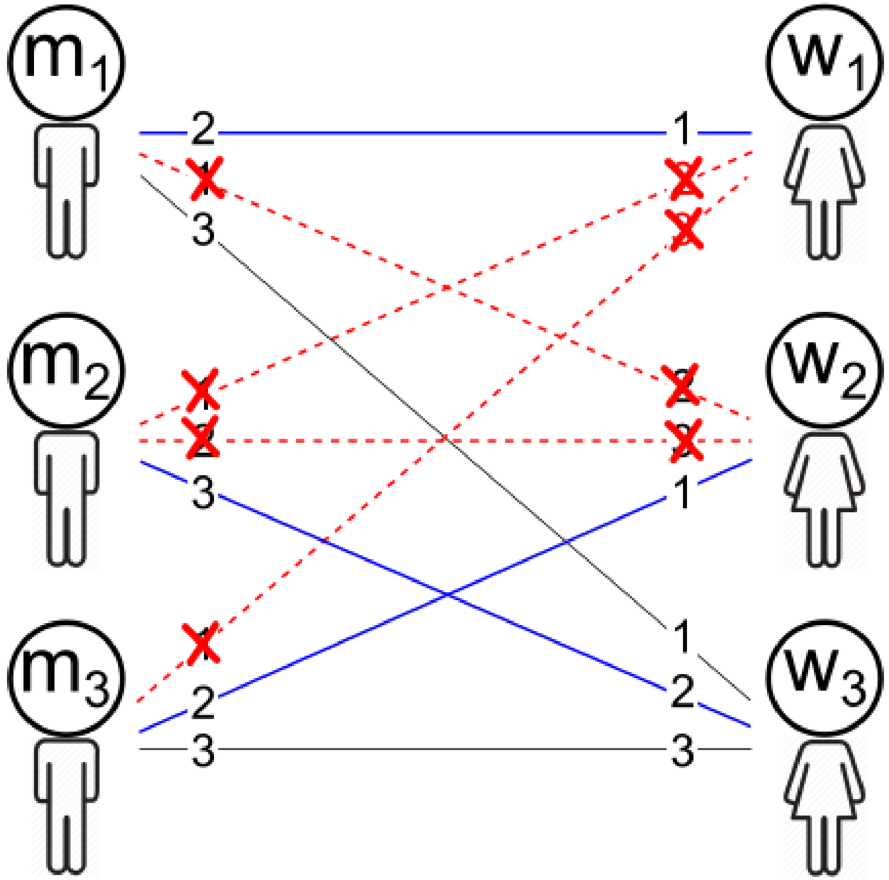


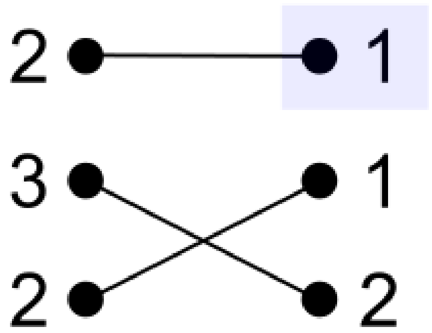
[Dubins and Freedman, *Amer. Math. Mon.* 1981; Roth, *MOR* 1982]

DA algorithm is not strategyproof.

Outcome for true prefs







Any luck for the men?

[Dubins and Freedman, *Amer. Math. Mon.* 1981; Roth, *MOR* 1982]

DA algorithm is strategyproof for the men.

So, men can't cheat in the men-proposing DA algorithm
but there exists an oppourtunity for a woman to manipulate.

Is it possible to efficiently compute a beneficial misreport
whenever one exists?

[Teo, Sethuraman and Tan, *Manag. Sci.* 2001; Vaish and Garg, *IJCAI* 2017]

An optimal manipulation for a woman can be computed in polynomial time.

[Teo, Sethuraman and Tan, *Manag. Sci.* 2001; Vaish and Garg, *IJCAI* 2017]

An optimal manipulation for a woman can be computed in polynomial time.

We will use a structural result about optimal manipulation strategies.

[Vaish and Garg, *IJCAI* 2017]

Any optimal manipulation for a woman can also be achieved by an "inconspicuous" misreport.

[Vaish and Garg, *IJCAI* 2017]

Any optimal manipulation for a woman can also be achieved by an "inconspicuous" misreport.

True list of woman w: $m_1 > m_2 > m_3 > m_4 > \boxed{m_5} > m_6 > m_7 > m_8$

[Vaish and Garg, *IJCAI* 2017]

Any optimal manipulation for a woman can also be achieved by an "inconspicuous" misreport.

True list of woman w: $m_1 > m_2 > m_3 > m_4 > \boxed{m_5} > m_6 > m_7 > m_8$

An optimal manipulation: $\boxed{m_2} > m_4 > m_1 > m_8 > m_6 > m_3 > m_5 > m_7$

[Vaish and Garg, *IJCAI* 2017]

Any optimal manipulation for a woman can also be achieved by an "inconspicuous" misreport.

True list of woman w: $m_1 > m_2 > m_3 > m_4 > \boxed{m_5} > m_6 > m_7 > m_8$

An optimal manipulation: $\boxed{m_2} > m_4 > m_1 > m_8 > m_6 > m_3 > m_5 > m_7$

An *inconspicuous* misreport that is also optimal for w: $m_1 > \boxed{m_2} > m_6 > m_3 > m_4 > m_5 > m_7 > m_8$

[Vaish and Garg, *IJCAI* 2017]

Any optimal manipulation for a woman can also be achieved by an "inconspicuous" misreport.

True list of woman w: $m_1 > m_2 > m_3 > m_4 > \boxed{m_5} > m_6 > m_7 > m_8$

An optimal manipulation: $\boxed{m_2} > m_4 > m_1 > m_8 > m_6 > m_3 > m_5 > m_7$

An *inconspicuous* misreport that is also optimal for w: $m_1 > \boxed{m_2} > m_6 > m_3 > m_4 > m_5 > m_7 > m_8$

[Teo, Sethuraman and Tan, *Manag. Sci.* 2001; Vaish and Garg, *IJCAI* 2017]

An optimal manipulation for a woman can be computed in polynomial time.

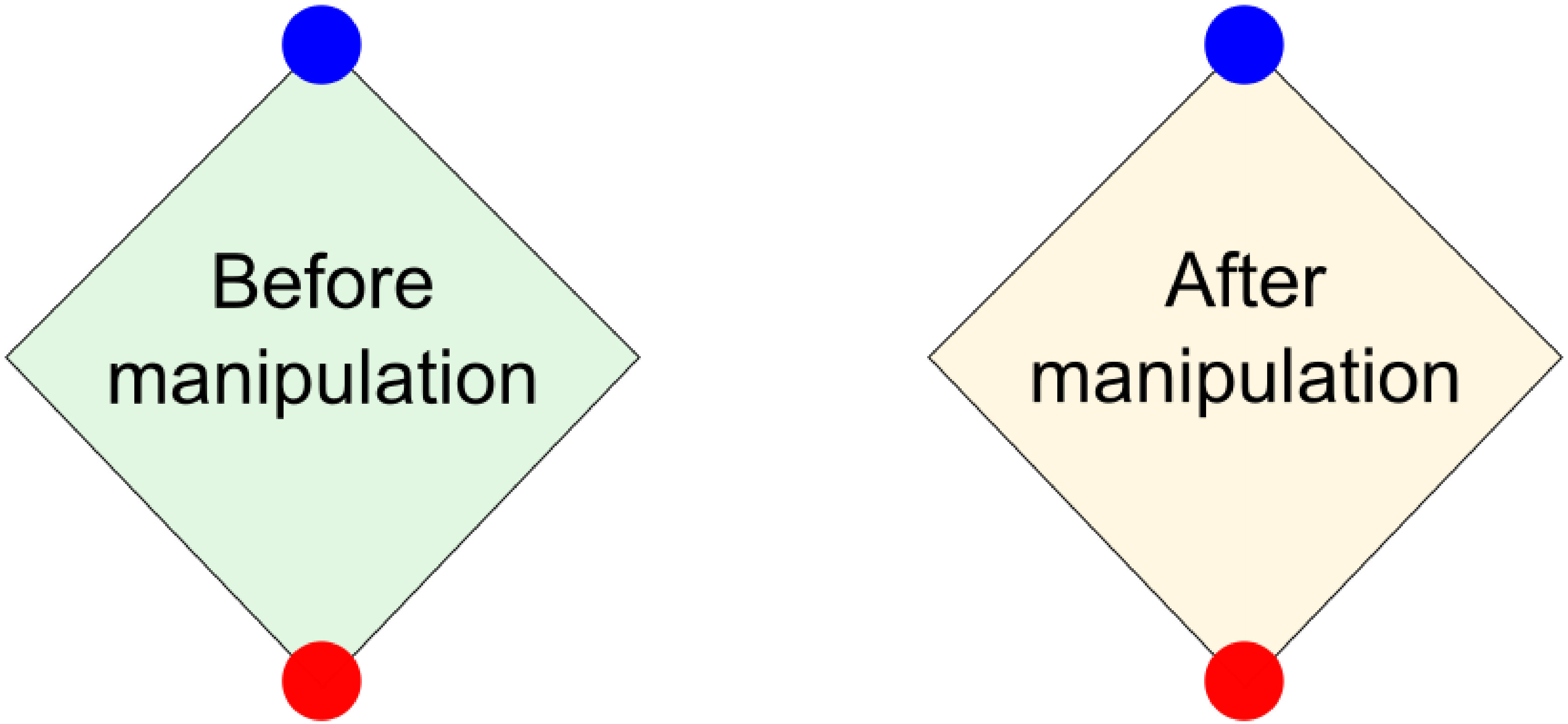
But what about stability (w.r.t. true preferences)?

[Vaish and Garg, *IJCAI* 2017]

The DA matching after optimal manipulation by a woman is stable with respect to the true preferences.

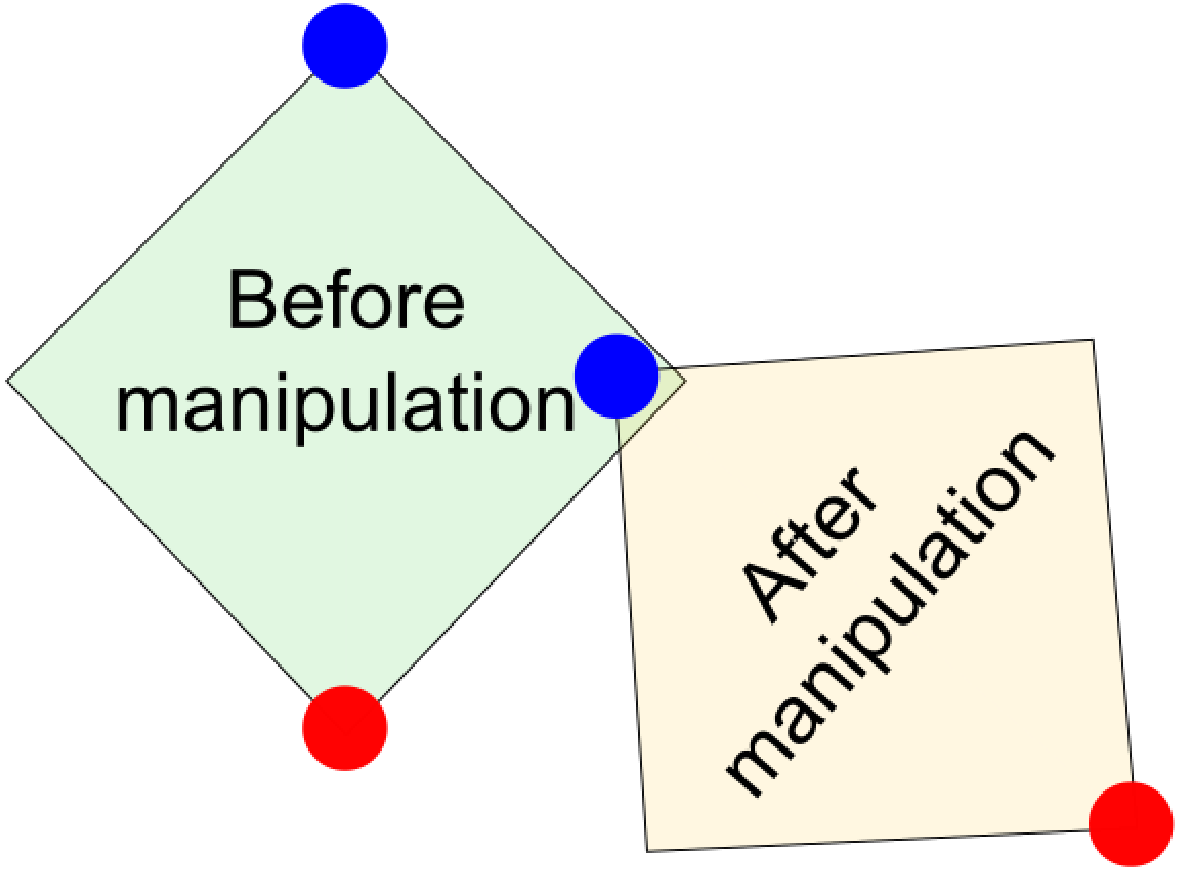
[Vaish and Garg, *IJCAI* 2017]

The DA matching after optimal manipulation by a woman is stable with respect to the true preferences.



[Vaish and Garg, *IJCAI* 2017]

The DA matching after optimal manipulation by a woman is stable with respect to the true preferences.



[Vaish and Garg, *IJCAI* 2017]

The DA matching after optimal manipulation by a woman is stable with respect to the true preferences.

Stable marriages are manipulable,
but optimally manipulated marriages are stable.

DA fails strategyproofness---too bad!

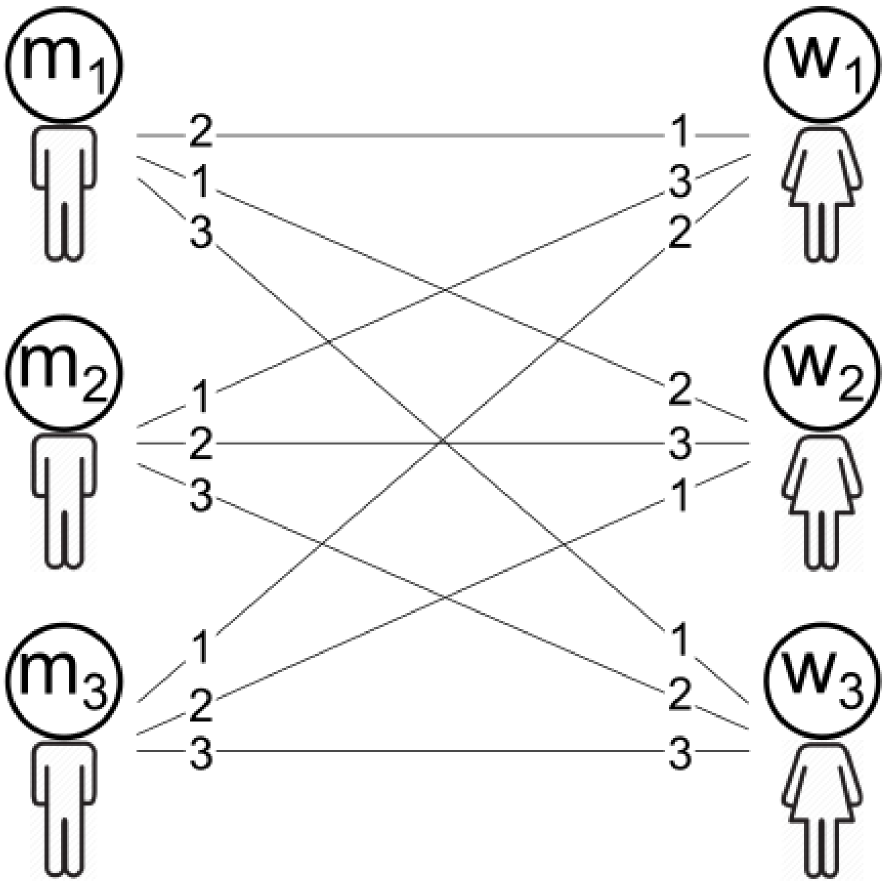
Let's think of a different stable matching algorithm that is truthful.

[Roth, *MOR* 1982]

No stable matching procedure can be strategyproof.

[Roth, *MOR* 1982]

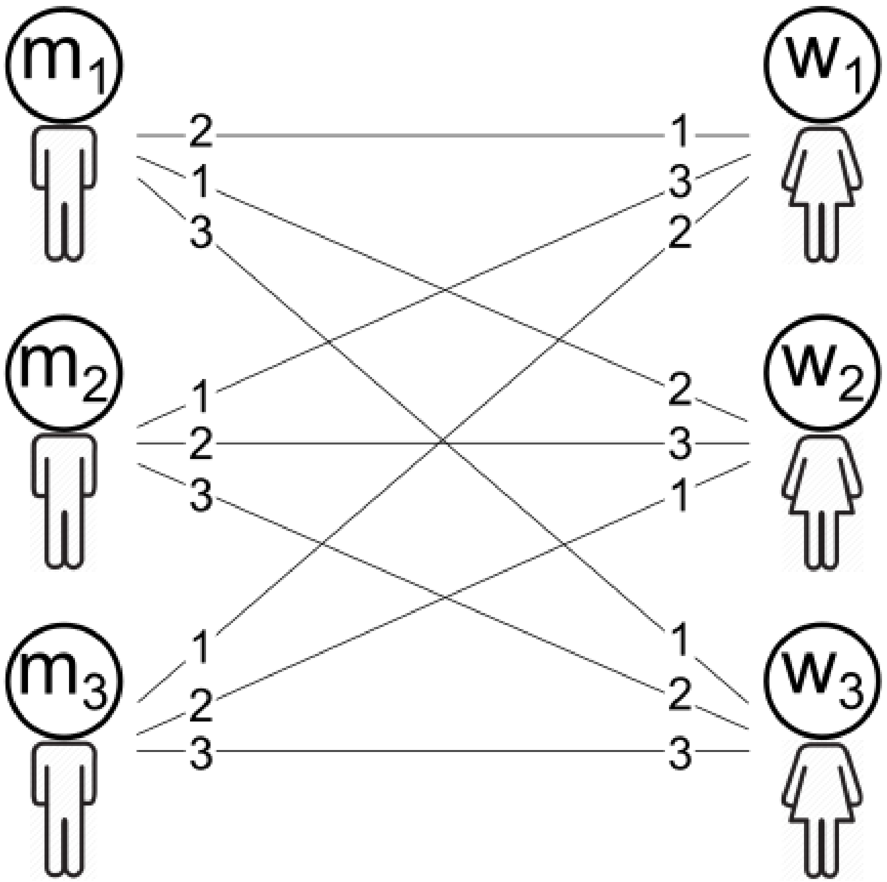
No stable matching procedure can be strategyproof.



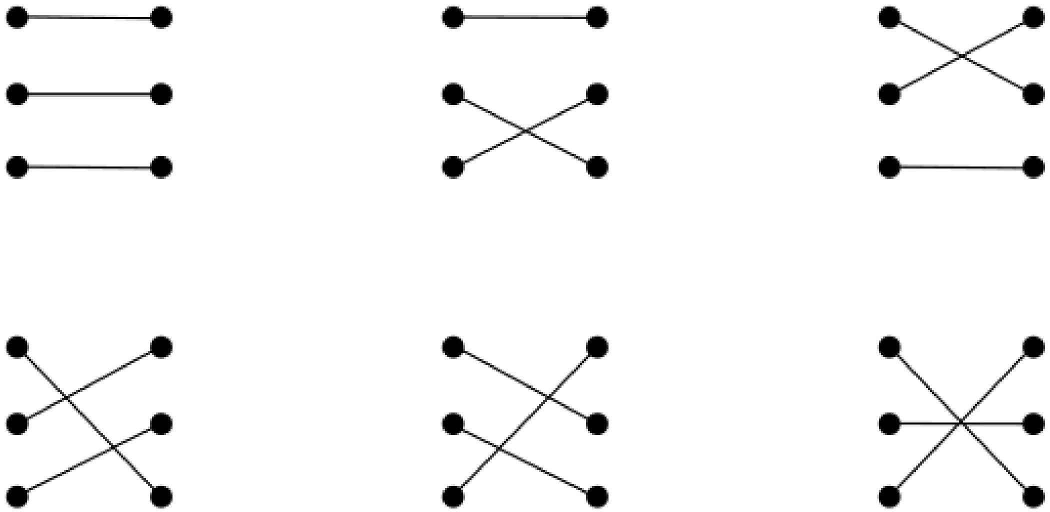
Instance I_0

[Roth, *MOR* 1982]

No stable matching procedure can be strategyproof.

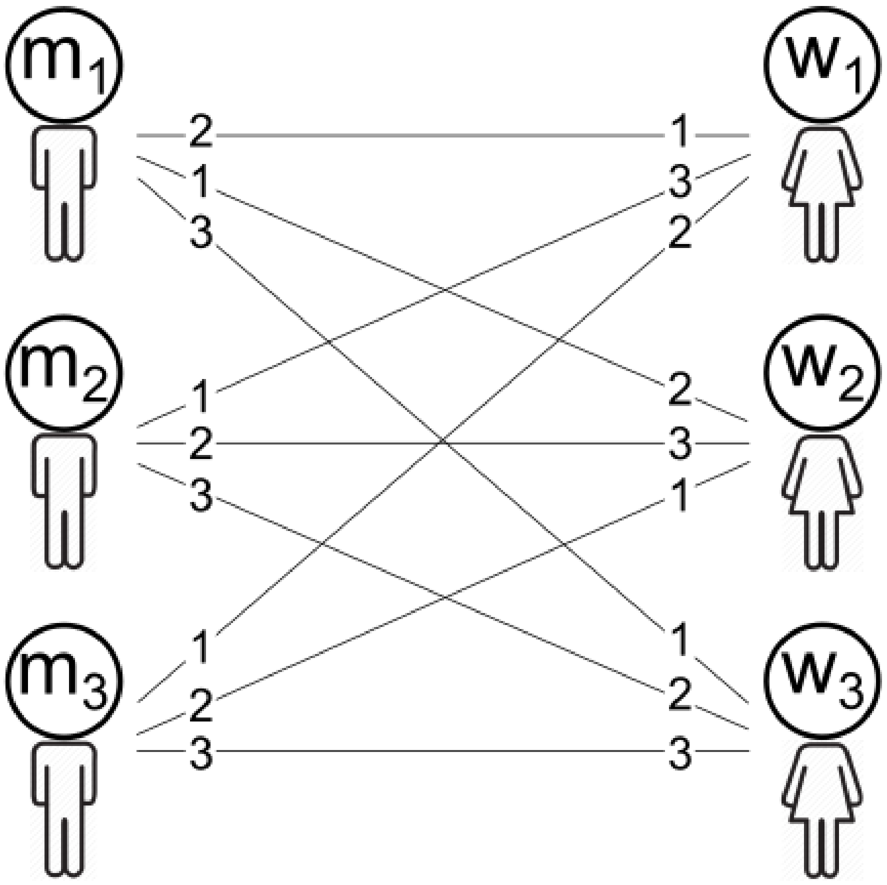


Instance I_0

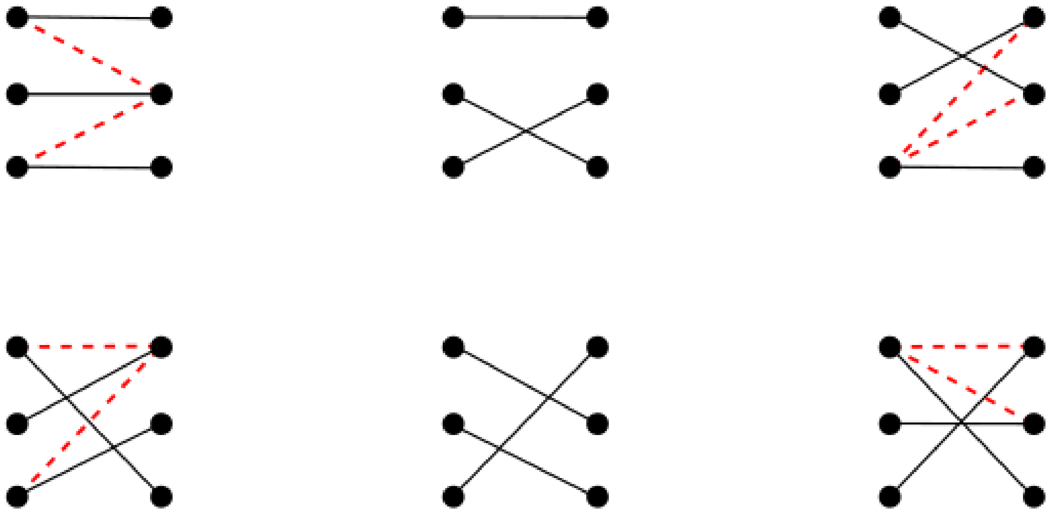


[Roth, *MOR* 1982]

No stable matching procedure can be strategyproof.

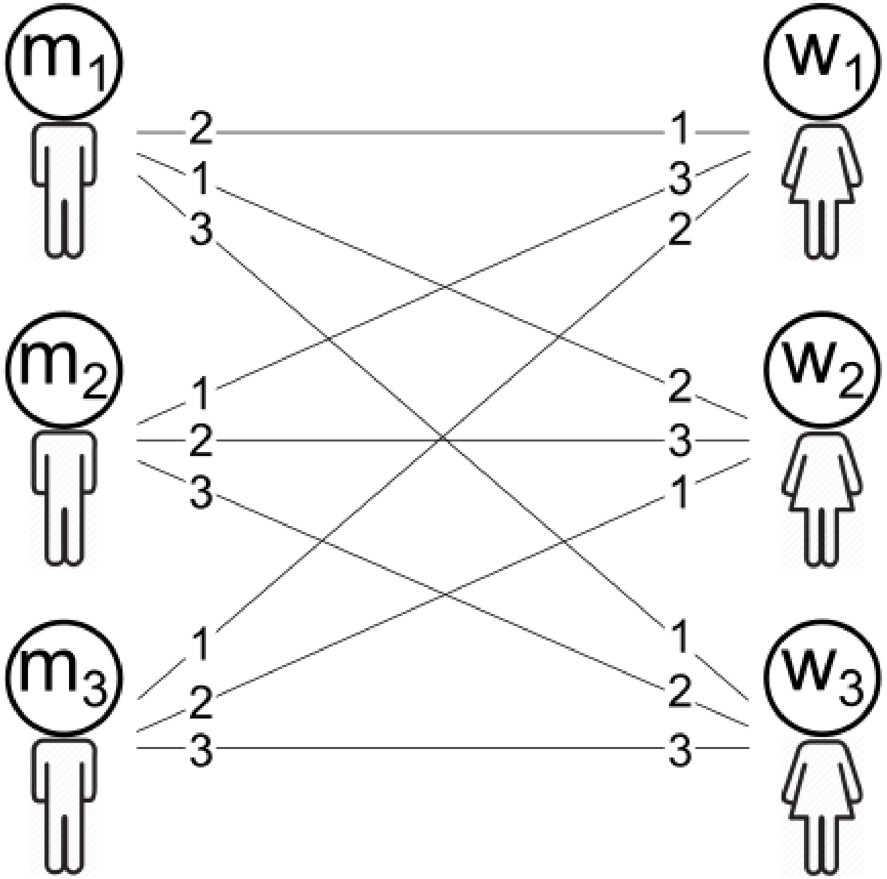


Instance I_0

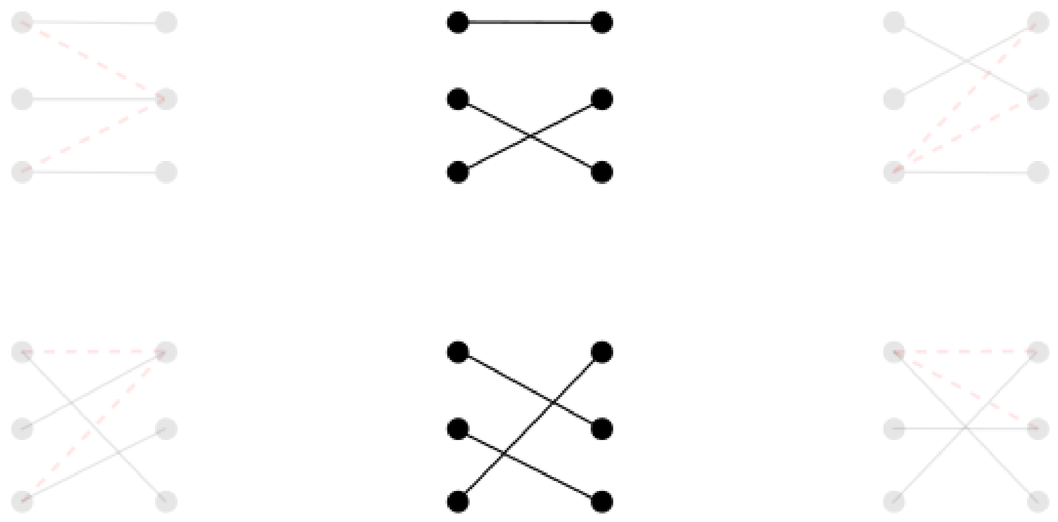


[Roth, *MOR* 1982]

No stable matching procedure can be strategyproof.

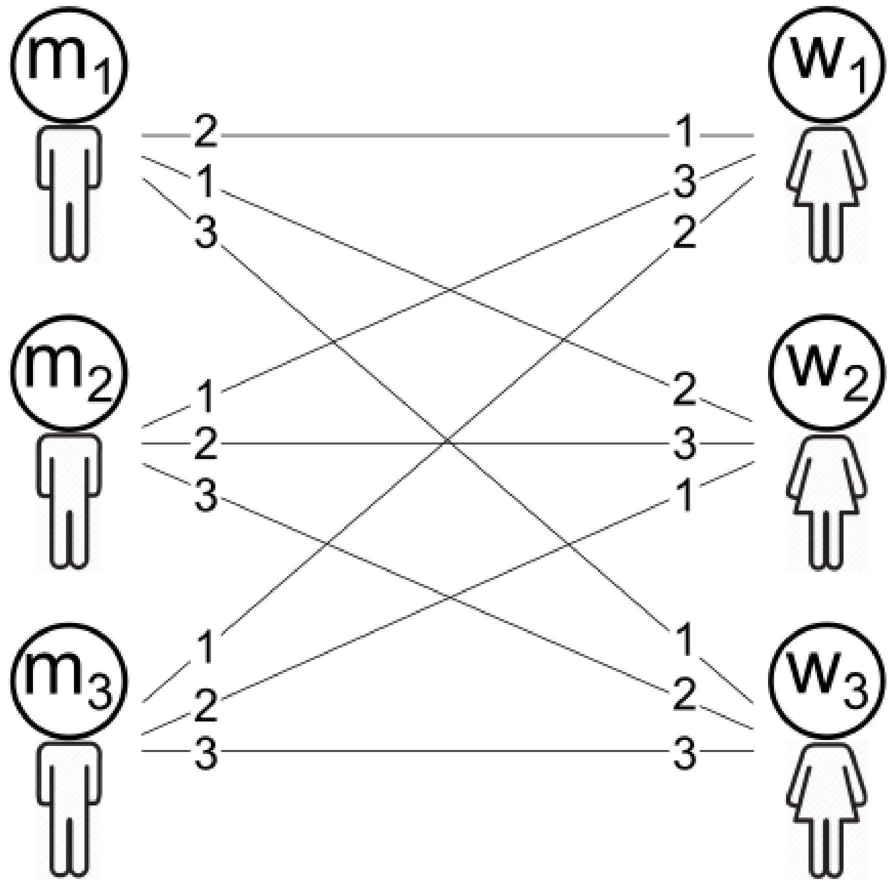


Instance I_0

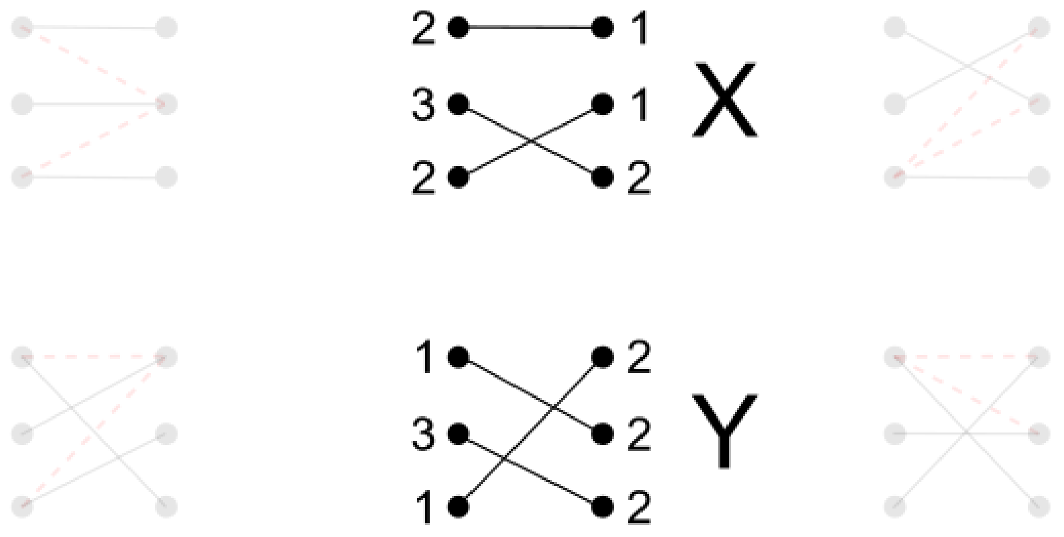


[Roth, *MOR* 1982]

No stable matching procedure can be strategyproof.

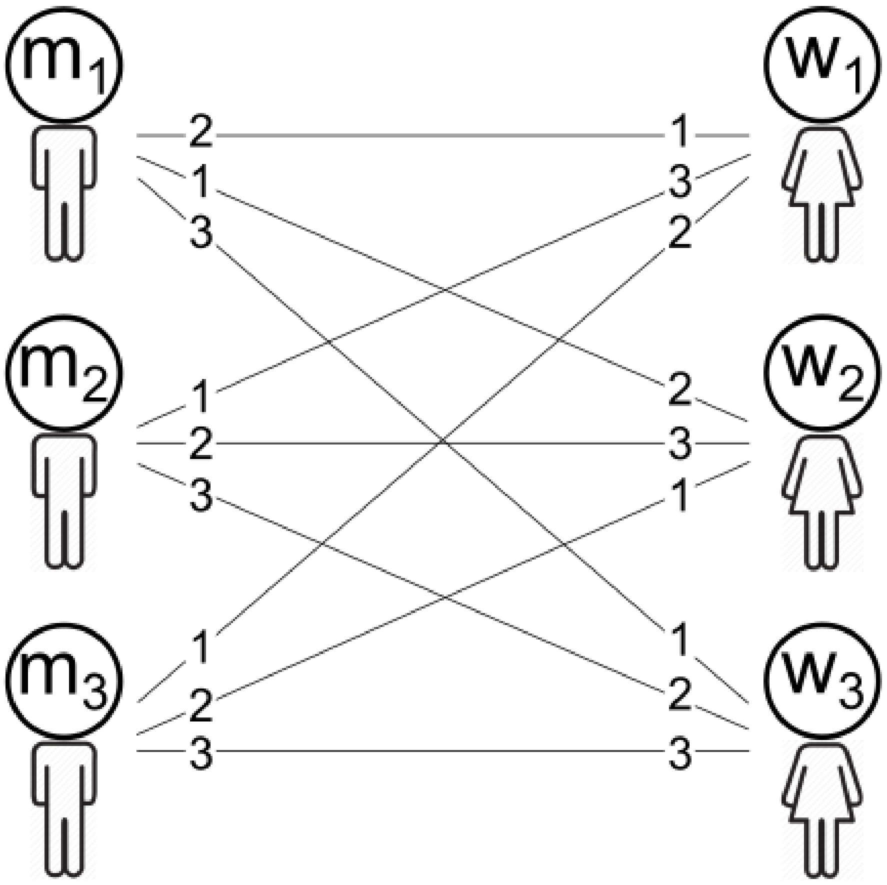


Instance I_0



[Roth, *MOR* 1982]

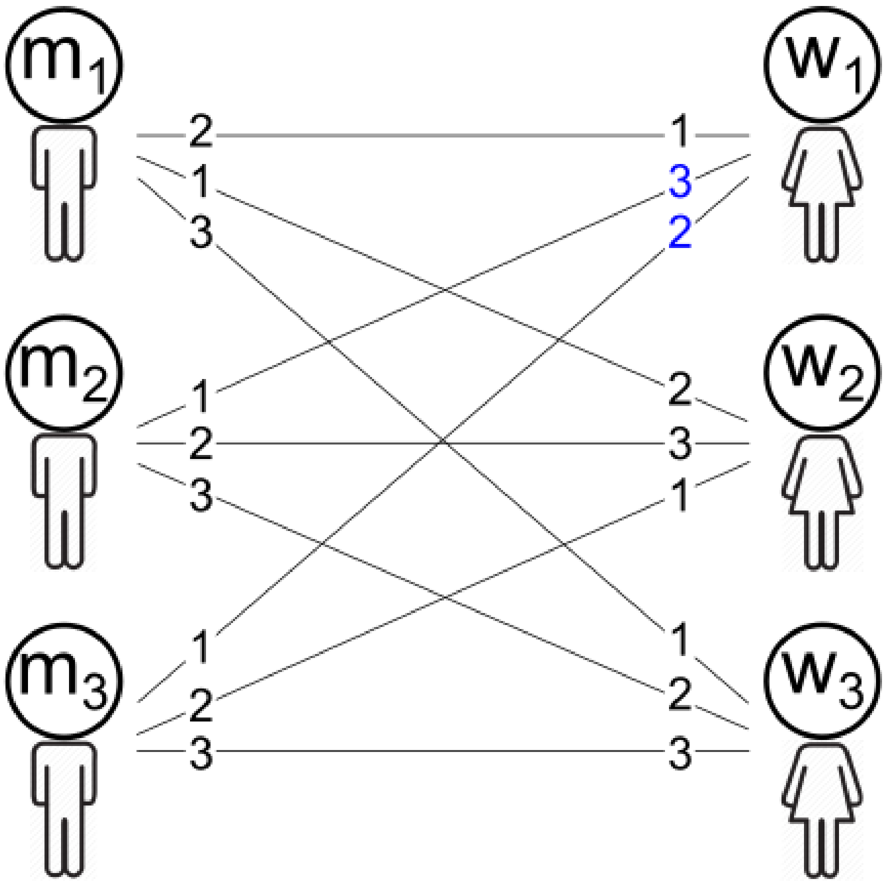
No stable matching procedure can be strategyproof.



Instance I_0

[Roth, *MOR* 1982]

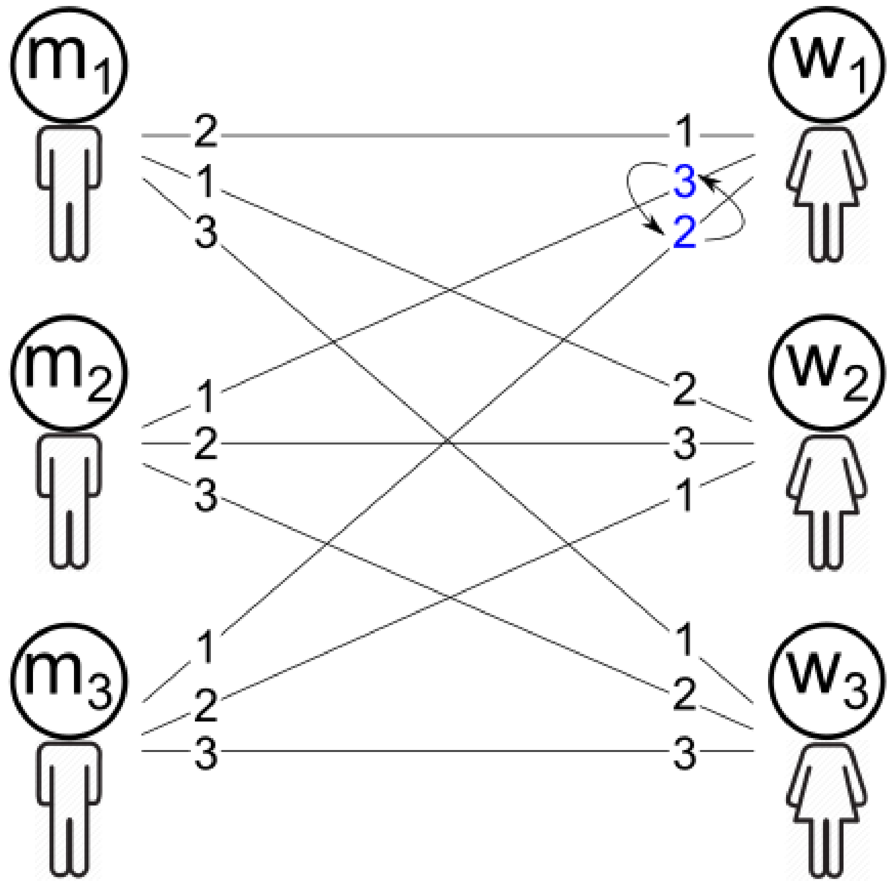
No stable matching procedure can be strategyproof.



Instance I_0

[Roth, *MOR* 1982]

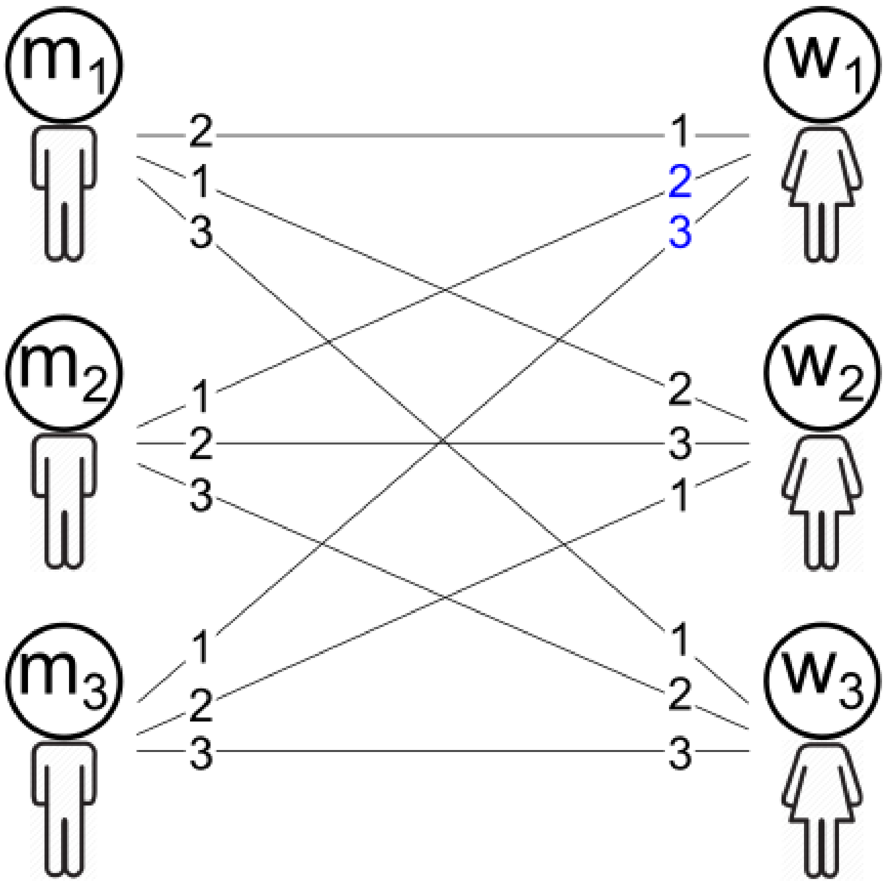
No stable matching procedure can be strategyproof.



Instance I_0

[Roth, *MOR* 1982]

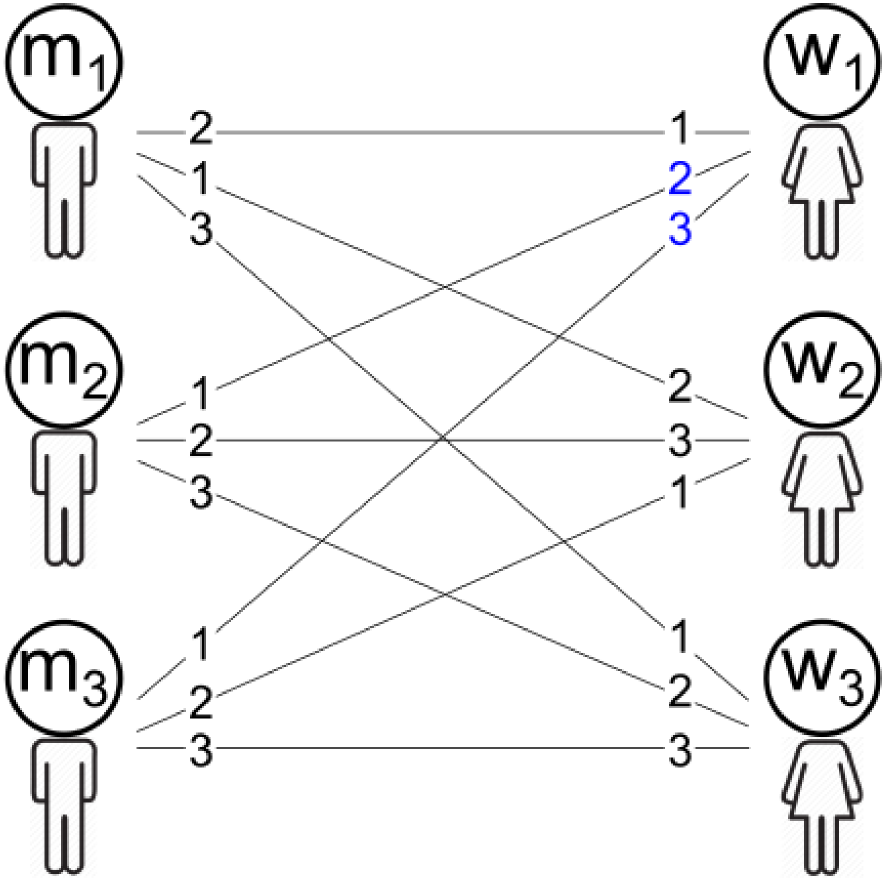
No stable matching procedure can be strategyproof.



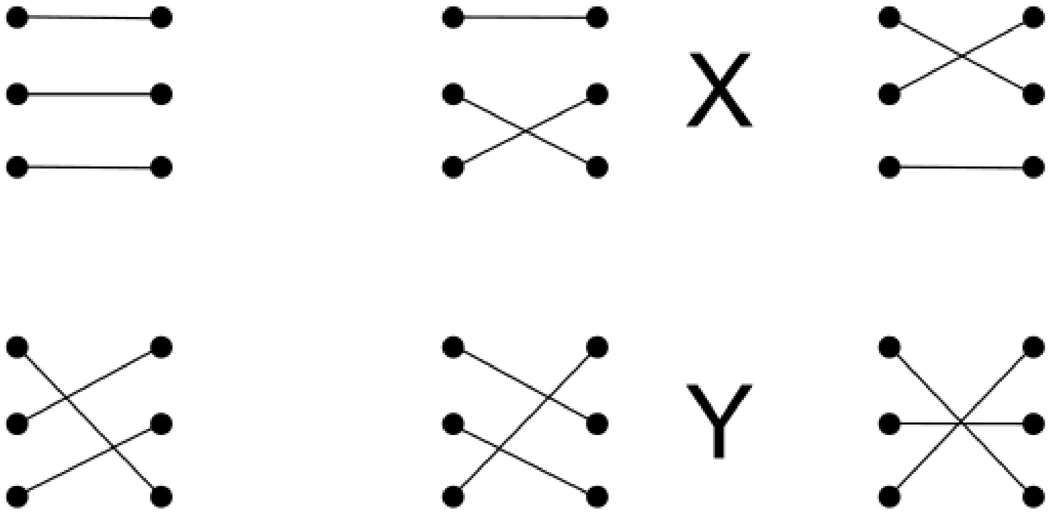
Instance I_1

[Roth, *MOR* 1982]

No stable matching procedure can be strategyproof.

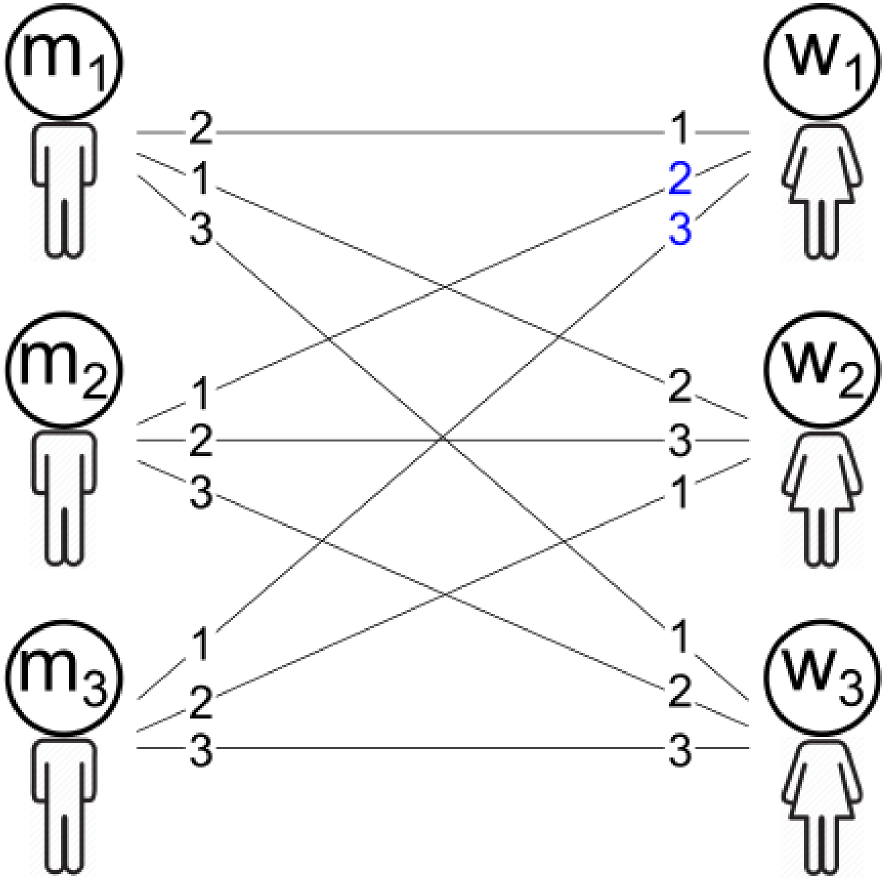


Instance I_1

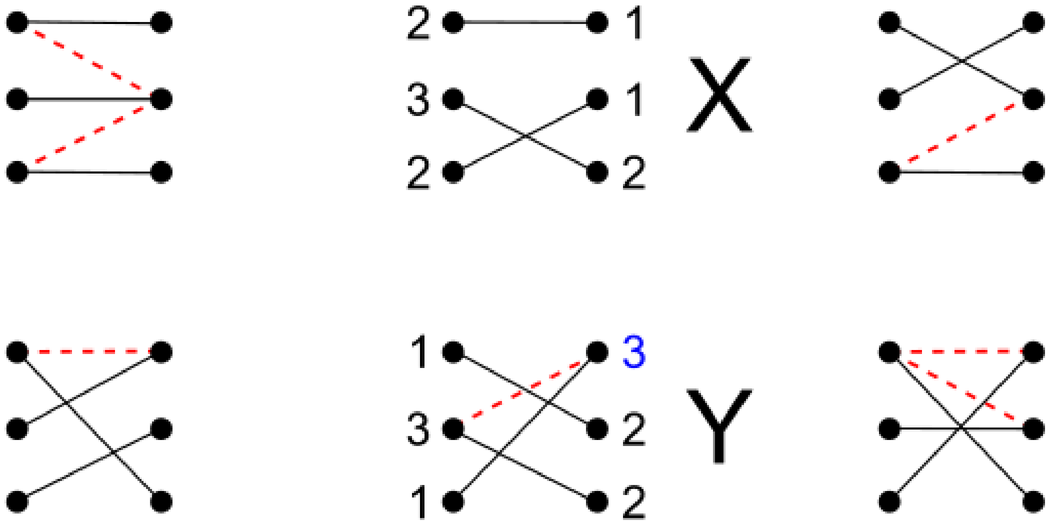


[Roth, *MOR* 1982]

No stable matching procedure can be strategyproof.

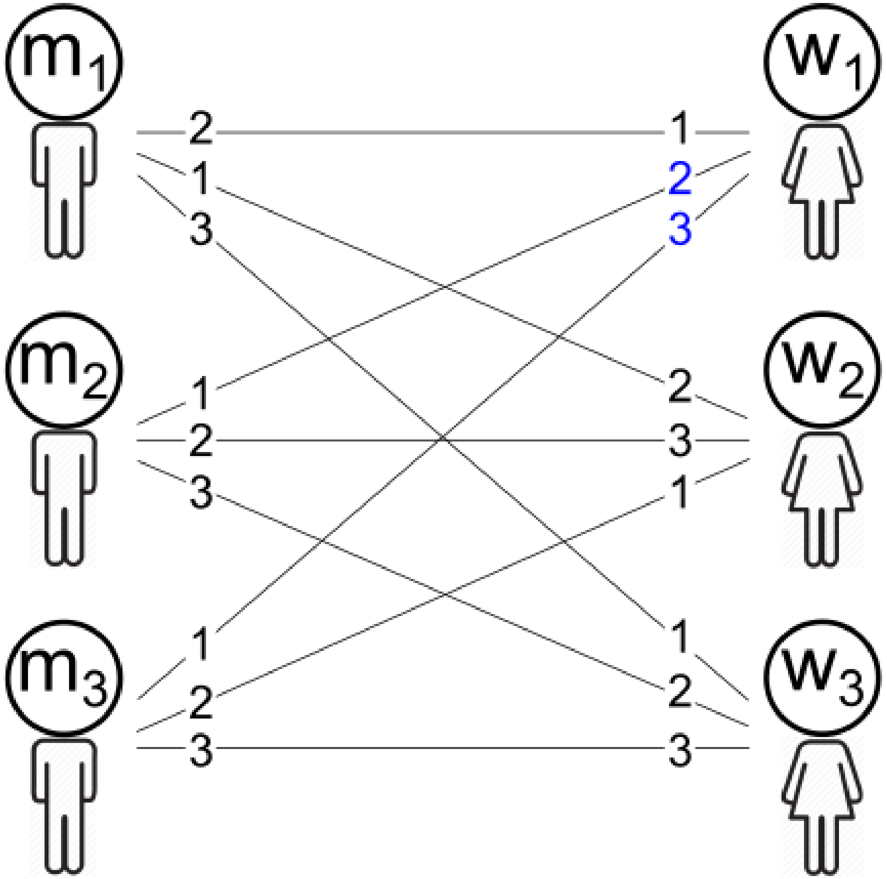


Instance I_1

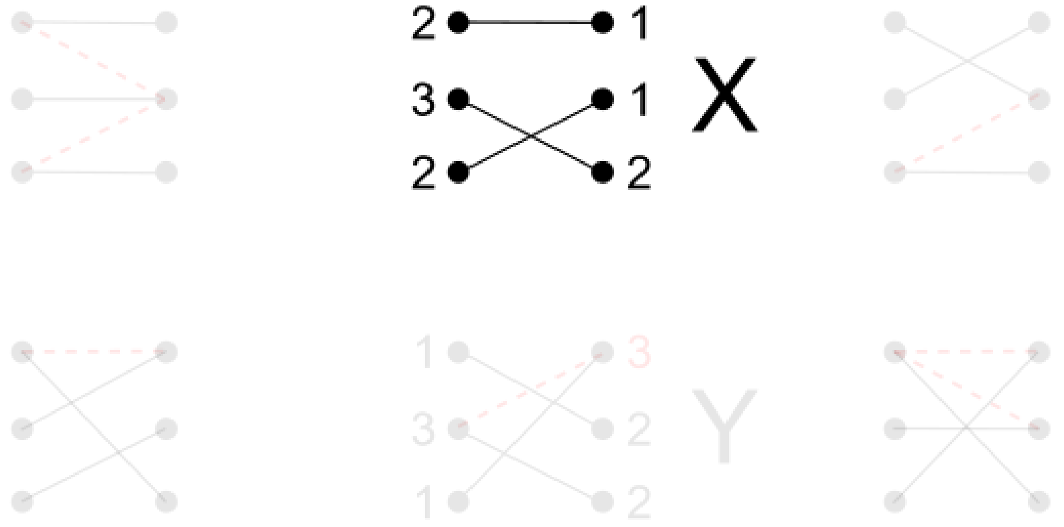


[Roth, *MOR* 1982]

No stable matching procedure can be strategyproof.

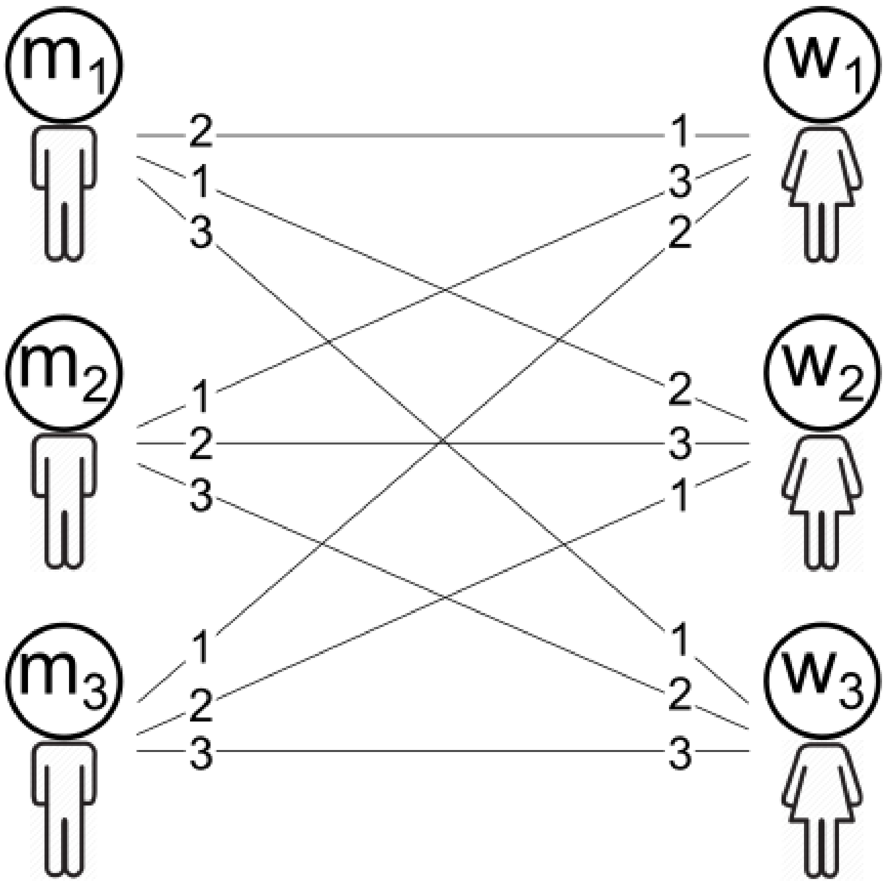


Instance I_1



[Roth, *MOR* 1982]

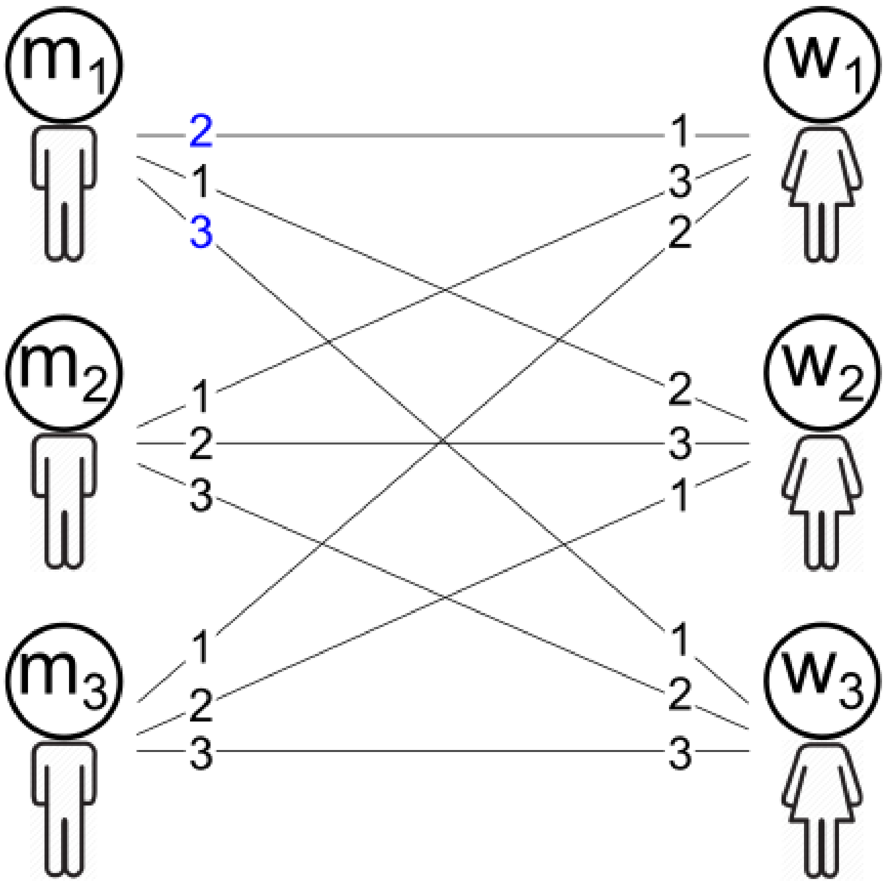
No stable matching procedure can be strategyproof.



Instance I_0

[Roth, *MOR* 1982]

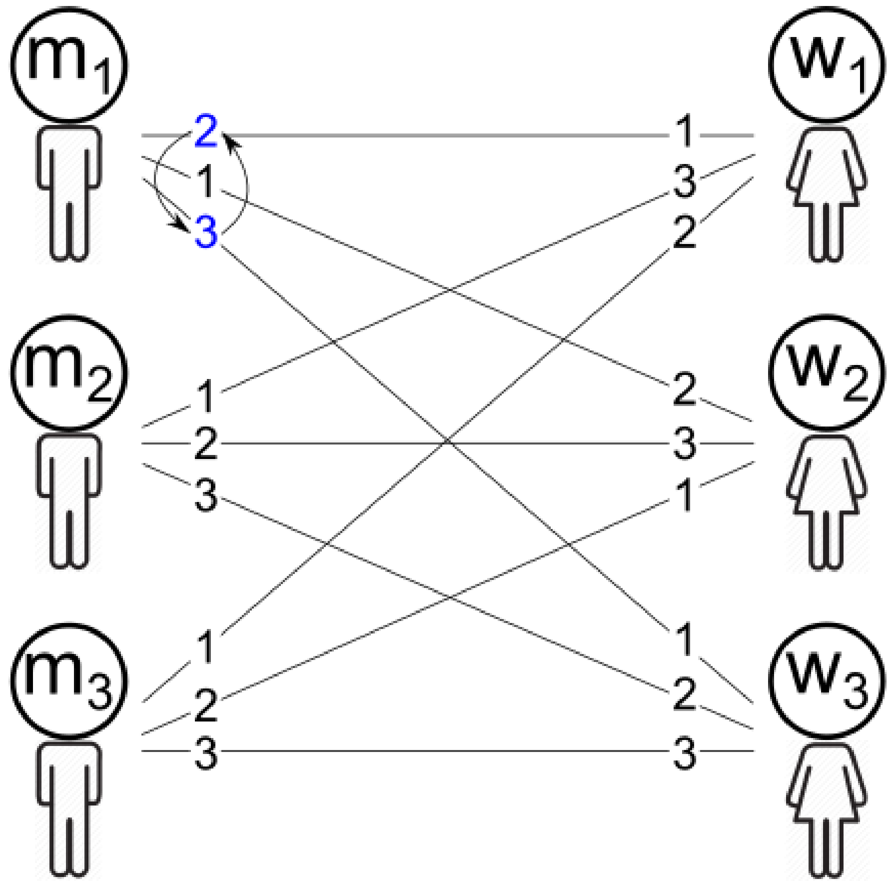
No stable matching procedure can be strategyproof.



Instance I_0

[Roth, *MOR* 1982]

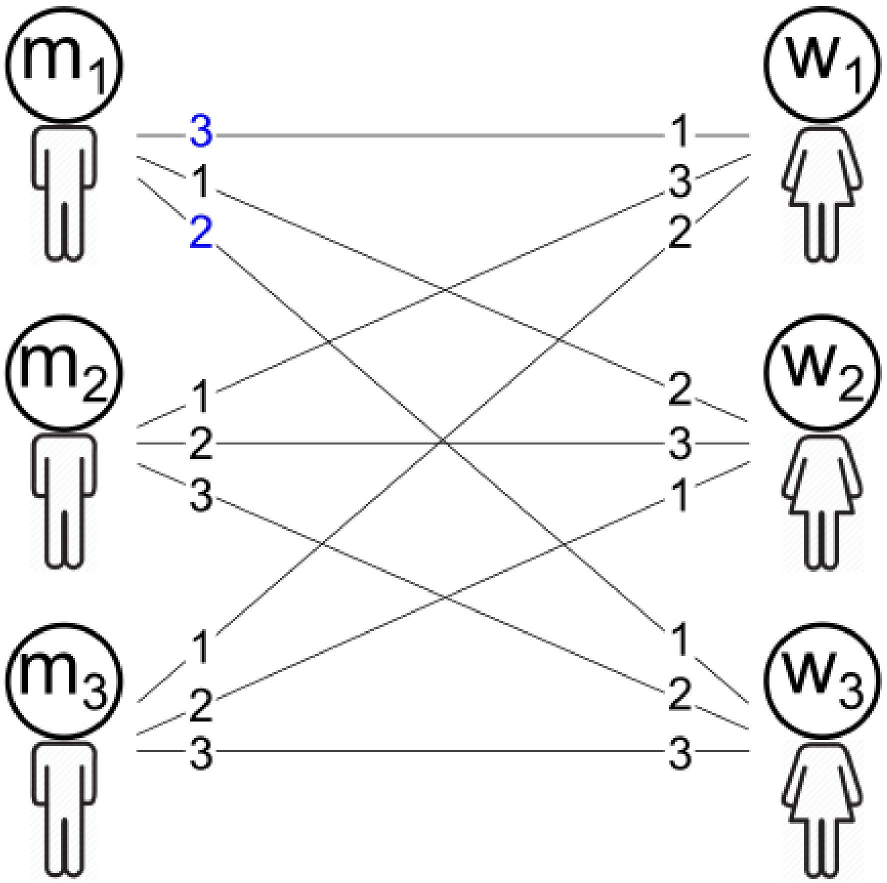
No stable matching procedure can be strategyproof.



Instance I_0

[Roth, *MOR* 1982]

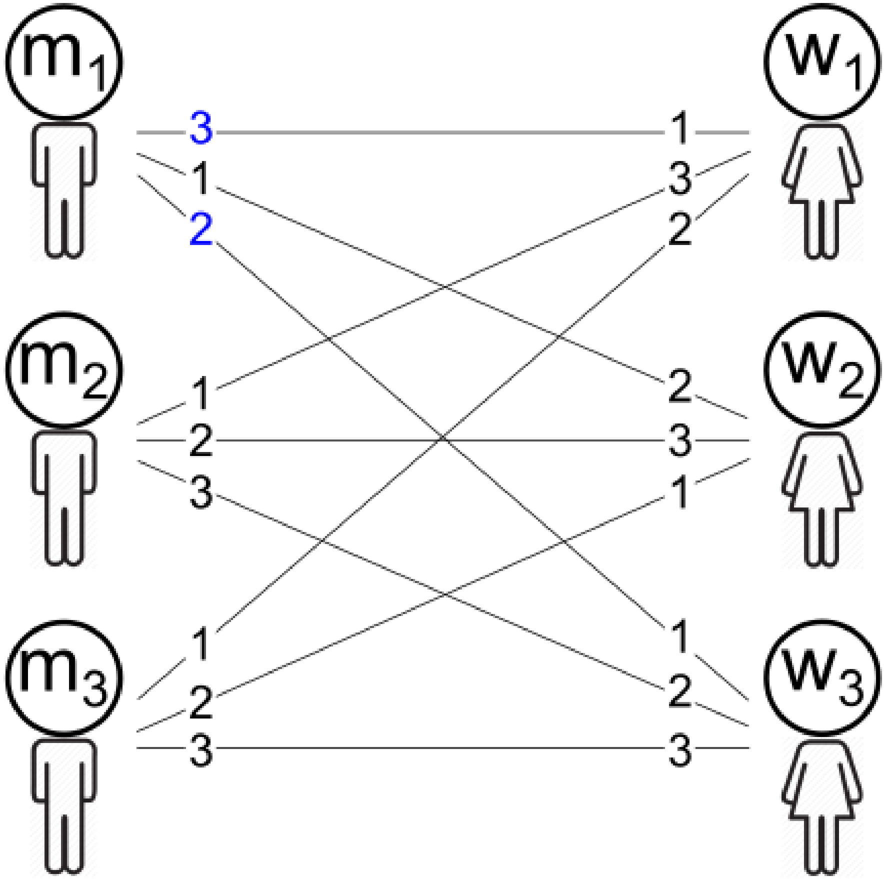
No stable matching procedure can be strategyproof.



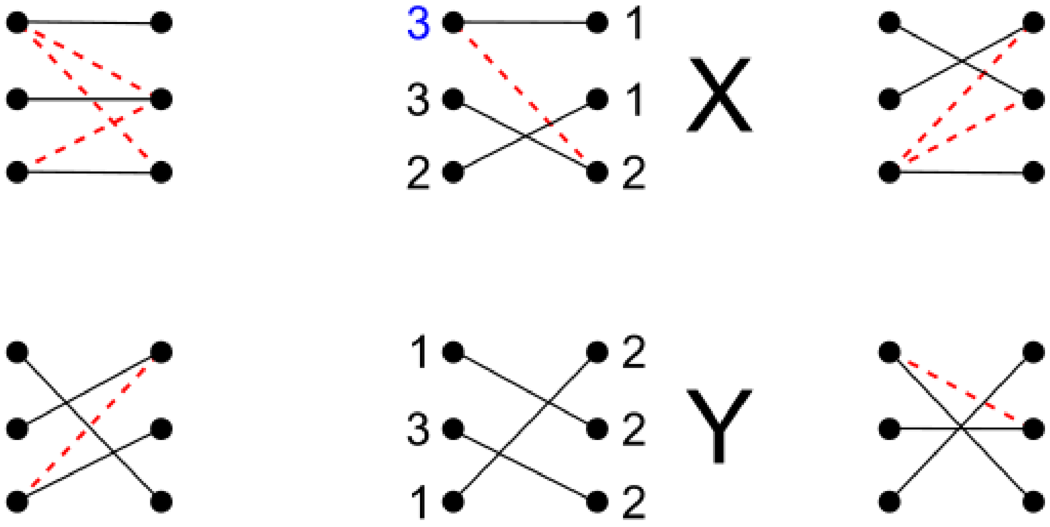
Instance I_2

[Roth, *MOR* 1982]

No stable matching procedure can be strategyproof.

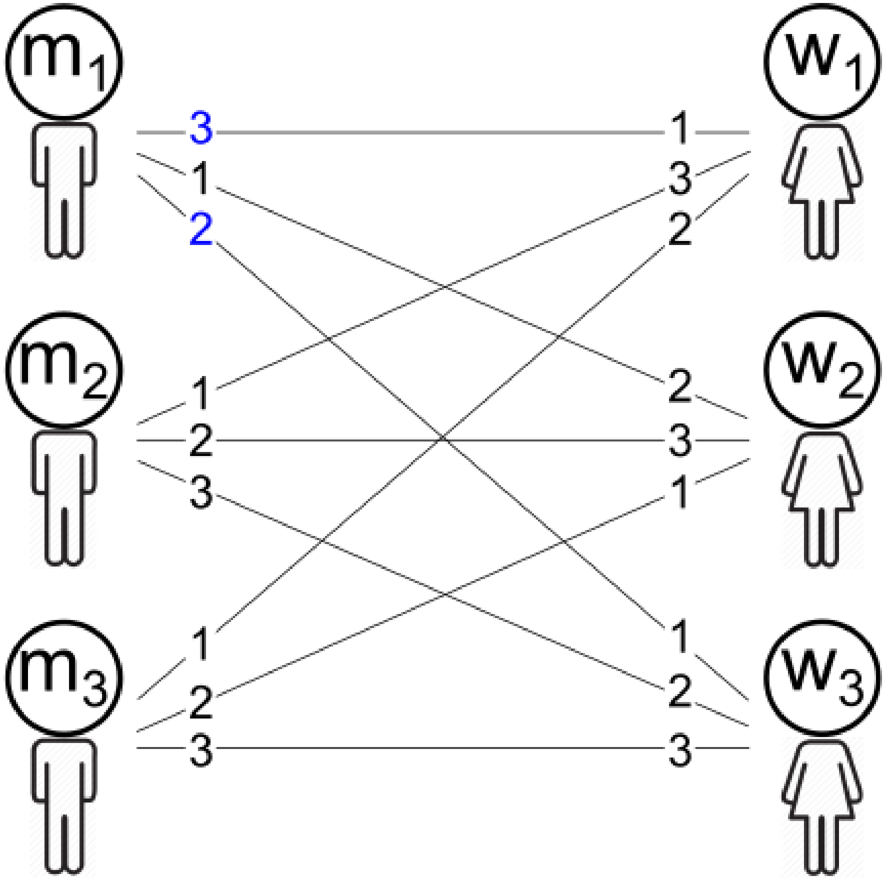


Instance I_2

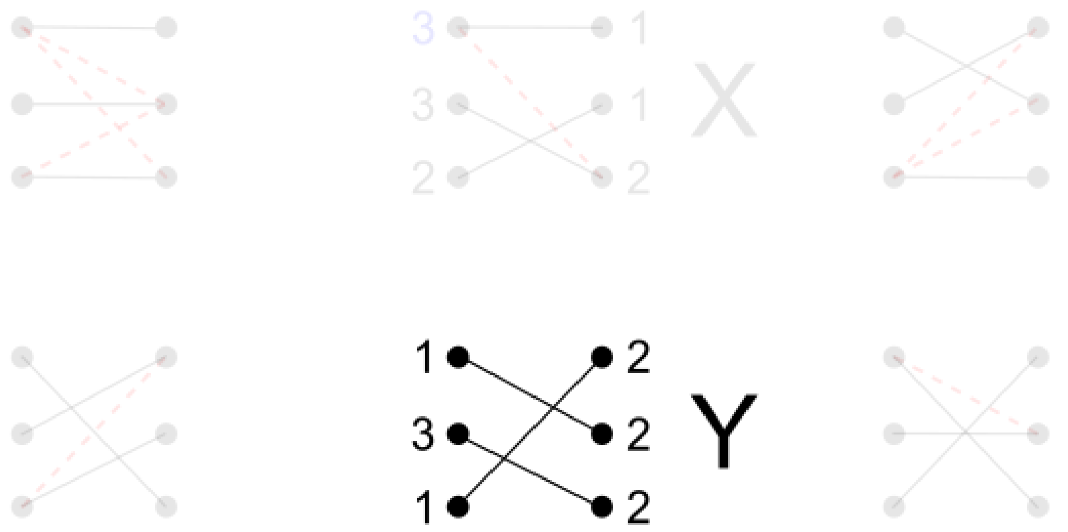


[Roth, *MOR* 1982]

No stable matching procedure can be strategyproof.

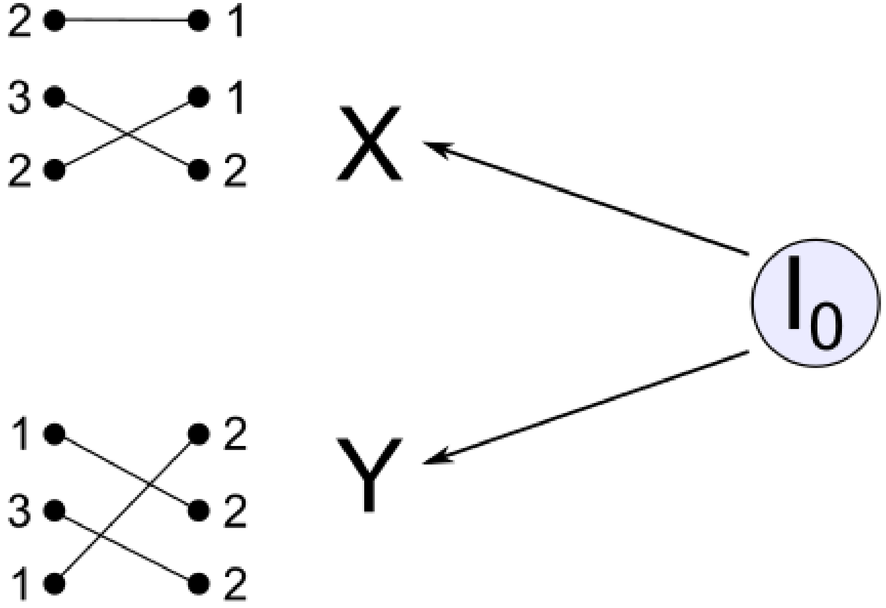


Instance I_2



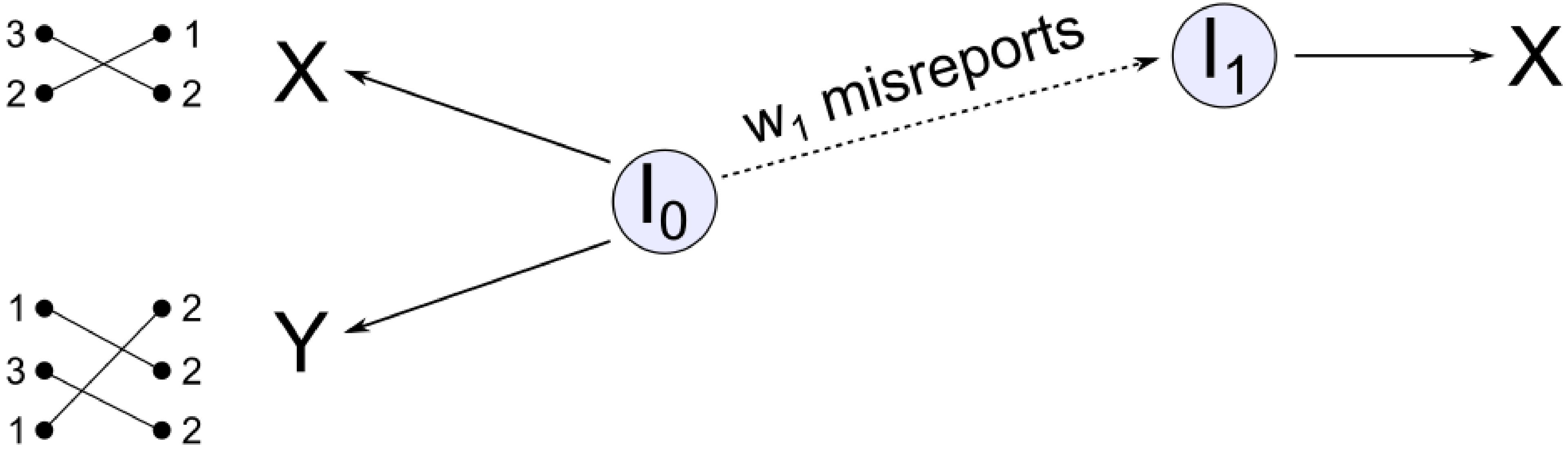
[Roth, *MOR* 1982]

No stable matching procedure can be strategyproof.



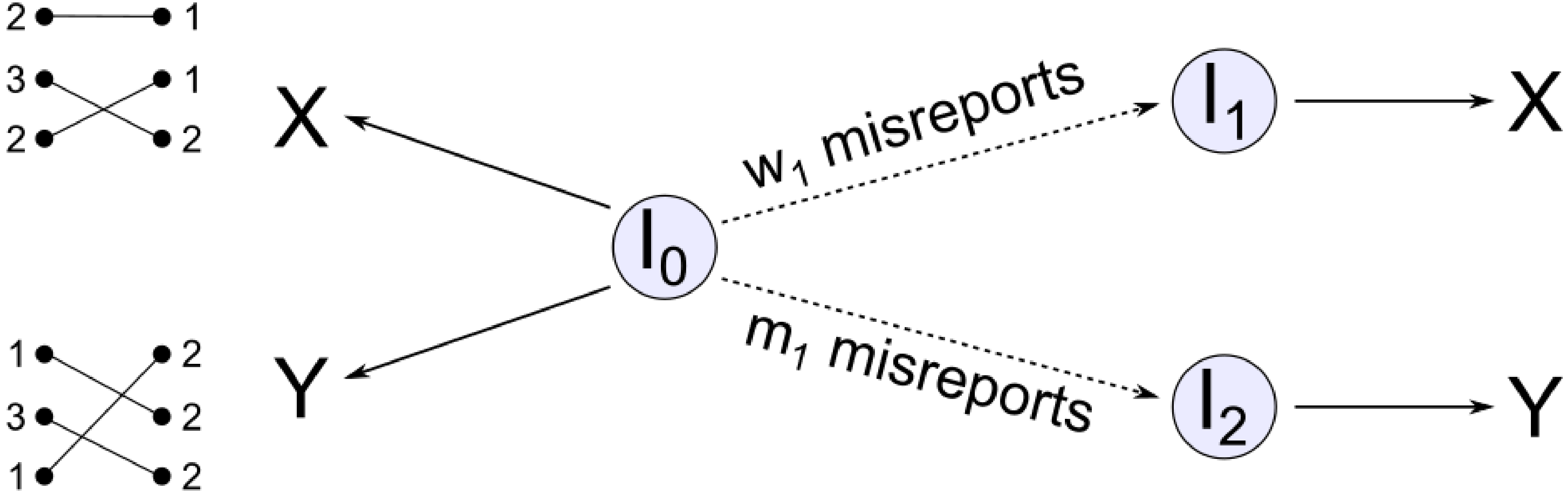
[Roth, *MOR* 1982]

No stable matching procedure can be strategyproof.



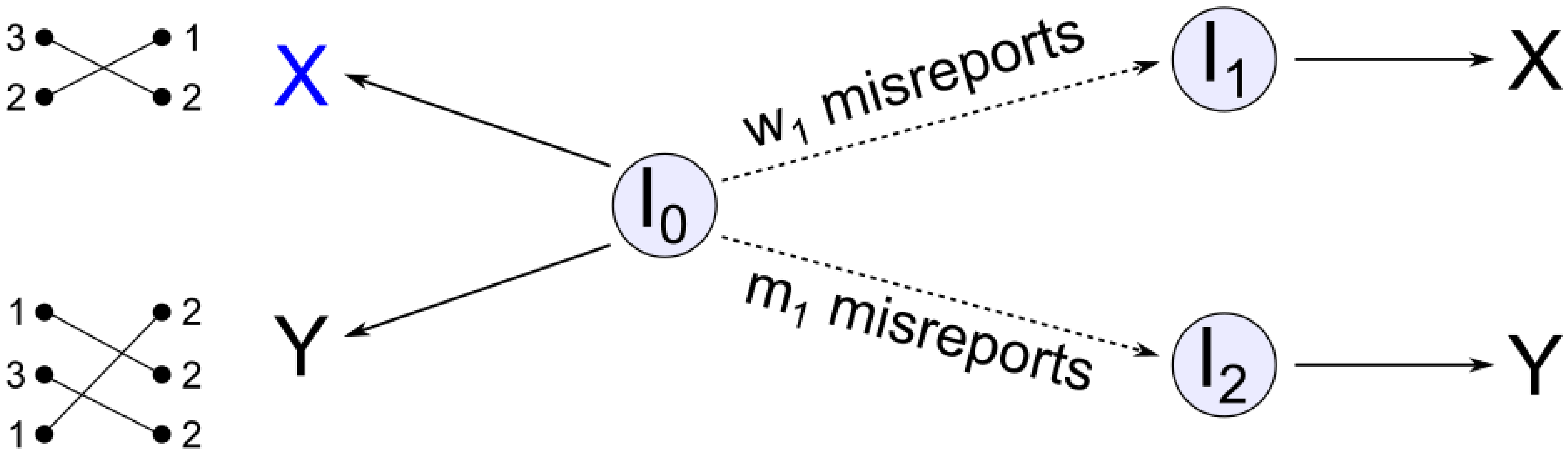
[Roth, *MOR* 1982]

No stable matching procedure can be strategyproof.



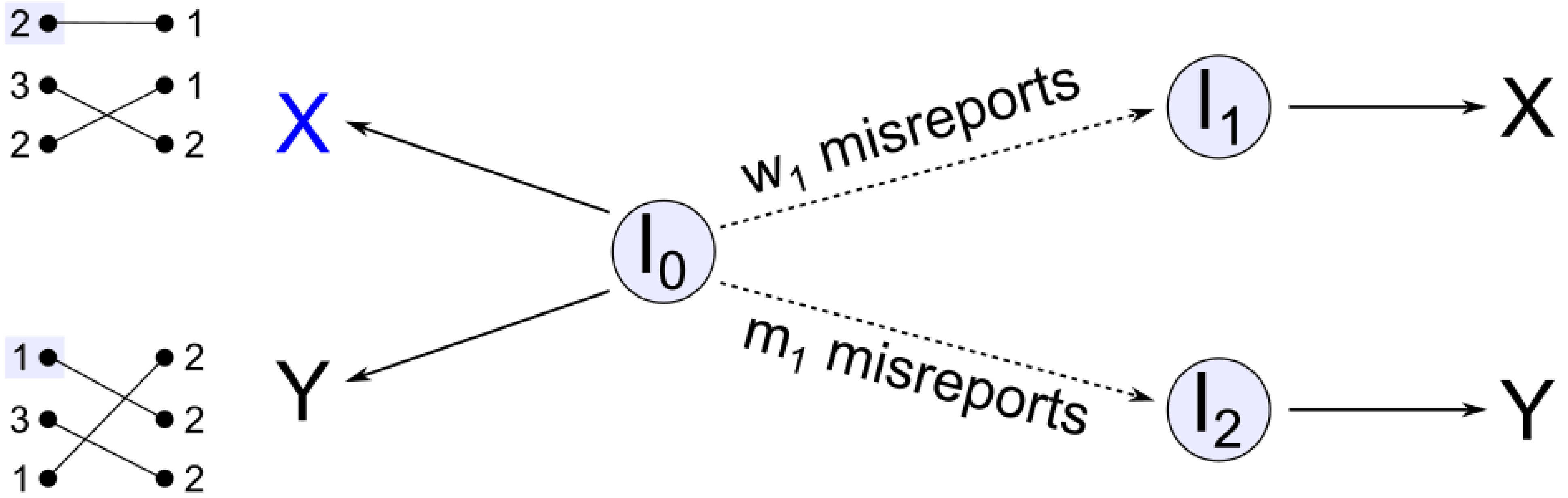
[Roth, *MOR* 1982]

No stable matching procedure can be strategyproof.



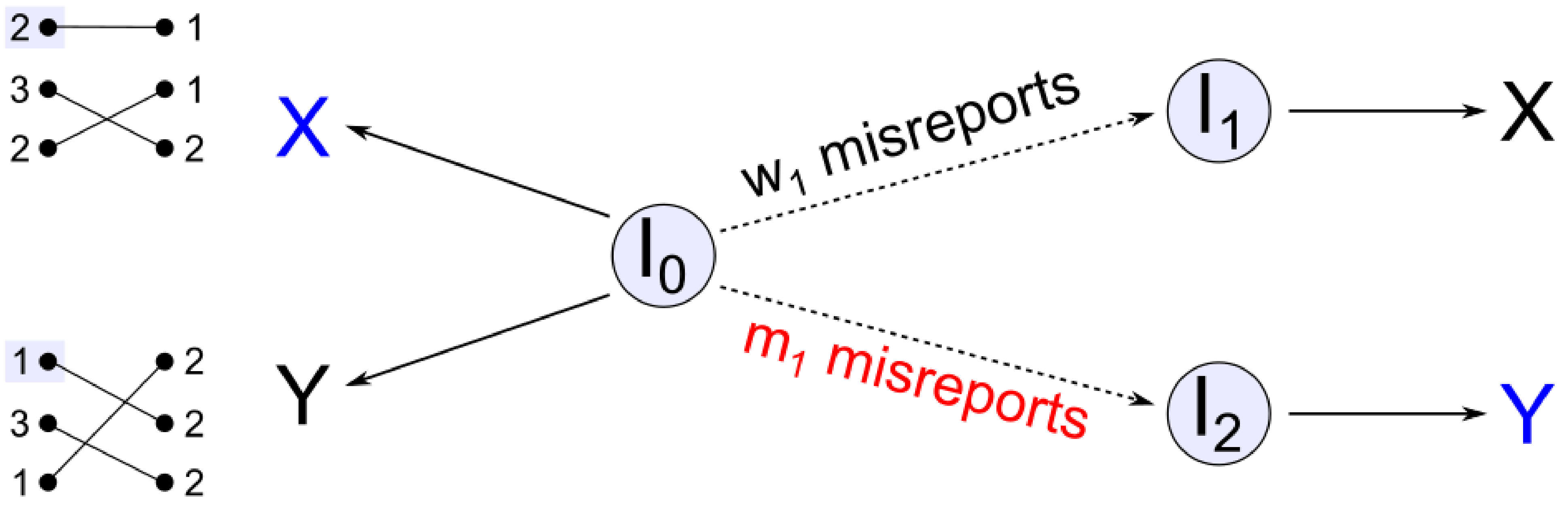
[Roth, *MOR* 1982]

No stable matching procedure can be strategyproof.



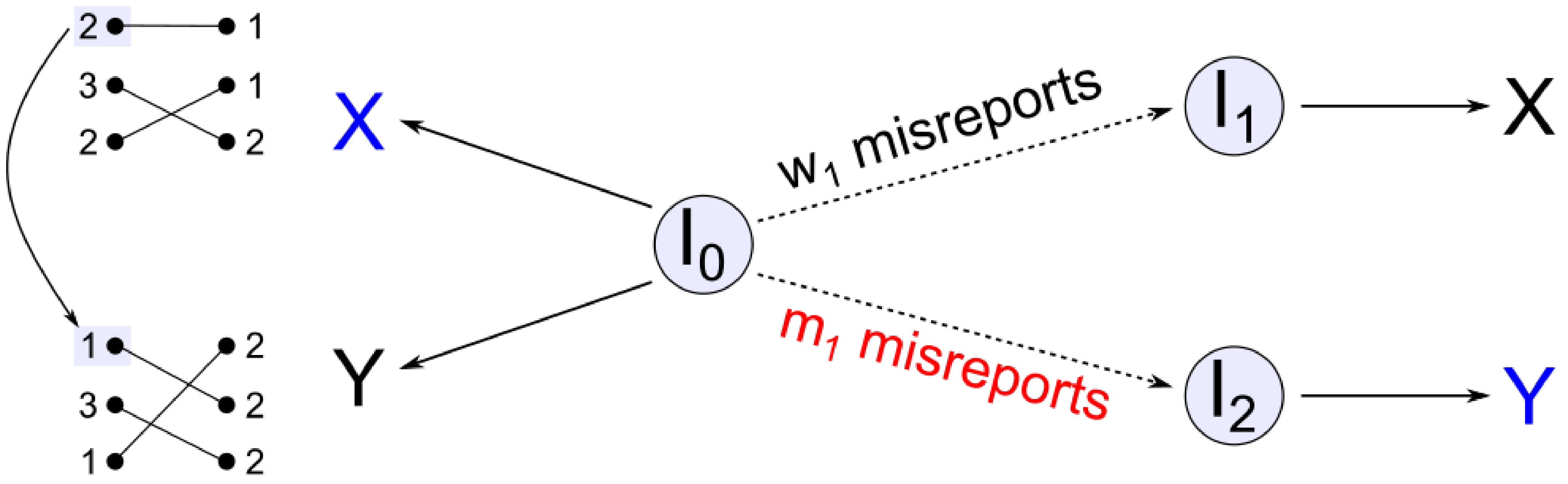
[Roth, *MOR* 1982]

No stable matching procedure can be strategyproof.



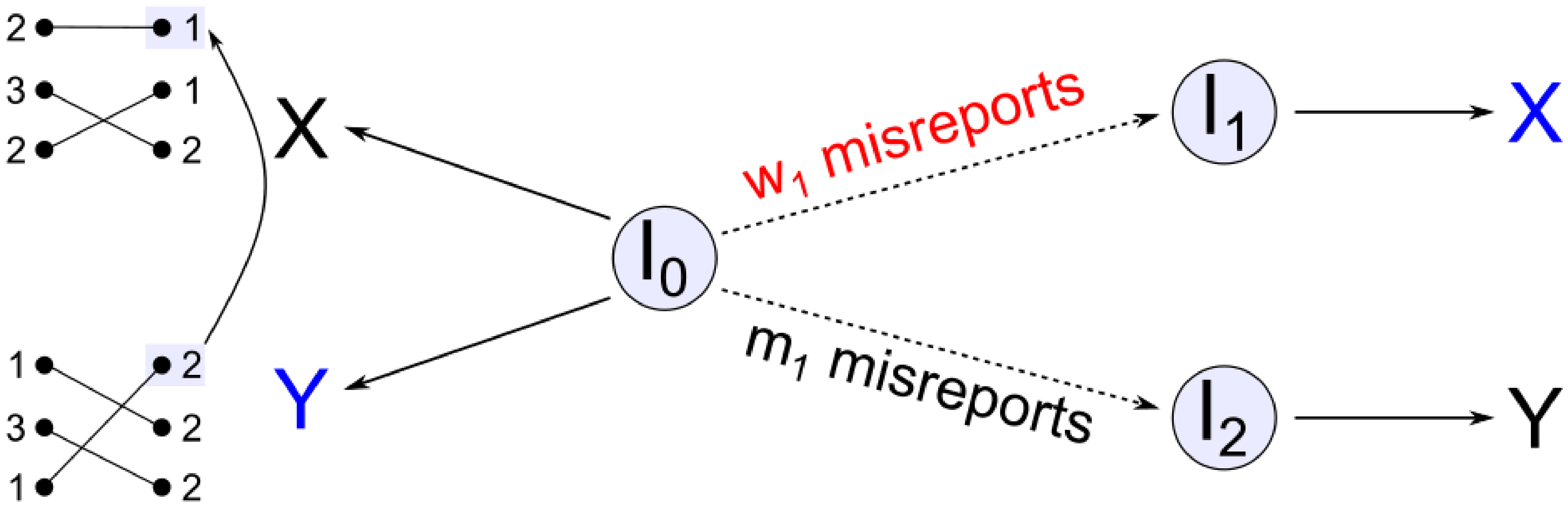
[Roth, *MOR* 1982]

No stable matching procedure can be strategyproof.



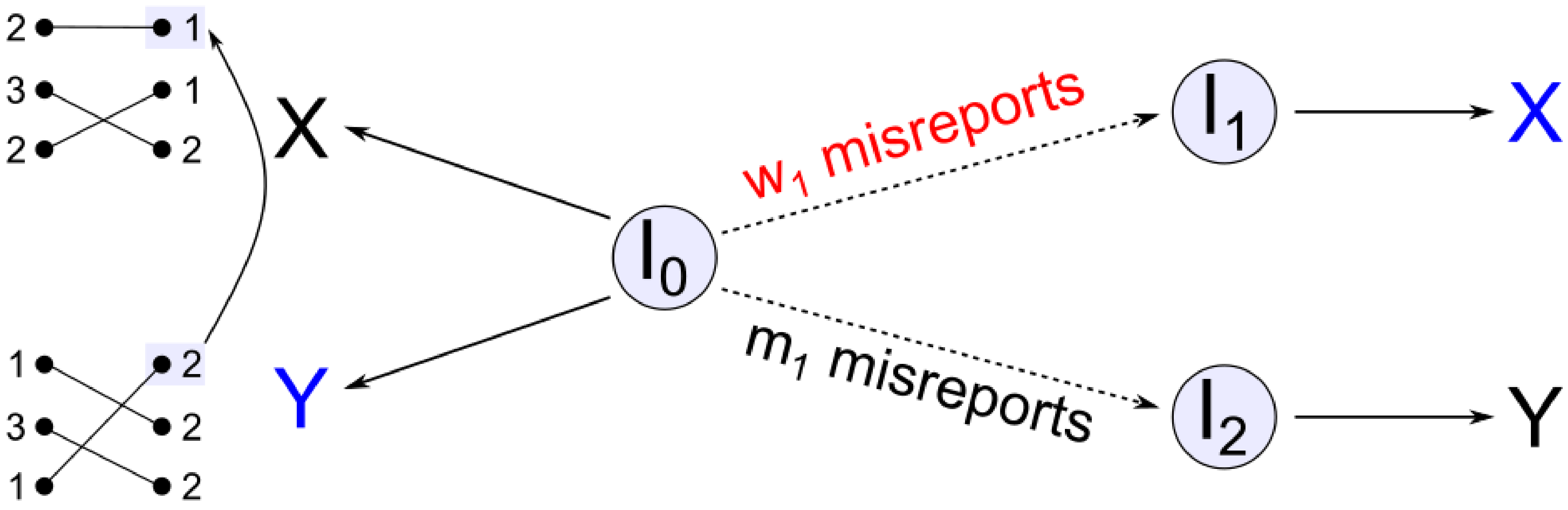
[Roth, *MOR* 1982]

No stable matching procedure can be strategyproof.



[Roth, *MOR* 1982]

No stable matching procedure can be strategyproof.



RECAP

RECAP

DA is strategyproof for the proposing side (men)
but can be manipulated by the proposed-to side (women).

RECAP

DA is strategyproof for the proposing side (men)
but can be manipulated by the proposed-to side (women).

An optimal manipulation strategy is "inconspicuous" w/o loss of generality
and thus can be efficiently computed.

RECAP

DA is strategyproof for the proposing side (men)
but can be manipulated by the proposed-to side (women).

An optimal manipulation strategy is "inconspicuous" w/o loss of generality
and thus can be efficiently computed.

Optimal manipulation is stability-preserving (w.r.t. true preferences).

RECAP

DA is strategyproof for the proposing side (men)
but can be manipulated by the proposed-to side (women).

An optimal manipulation strategy is "inconspicuous" w/o loss of generality
and thus can be efficiently computed.

Optimal manipulation is stability-preserving (w.r.t. true preferences).

No stable matching procedure is strategyproof for all agents.

Quiz

Quiz

Identify the optimal manipulation strategy for each woman.

$m_1: w_2 > w_1 > w_3 > w_4 > w_5$

$m_2: w_4 > w_1 > w_5 > w_3 > w_2$

$m_3: w_3 > w_1 > w_4 > w_5 > w_2$

$m_4: w_1 > w_3 > w_4 > w_5 > w_2$

$m_5: w_1 > w_5 > w_4 > w_3 > w_2$

$w_1: m_1 > m_2 > m_3 > m_4 > m_5$

$w_2: m_3 > m_2 > m_4 > m_5 > m_1$

$w_3: m_4 > m_2 > m_1 > m_5 > m_3$

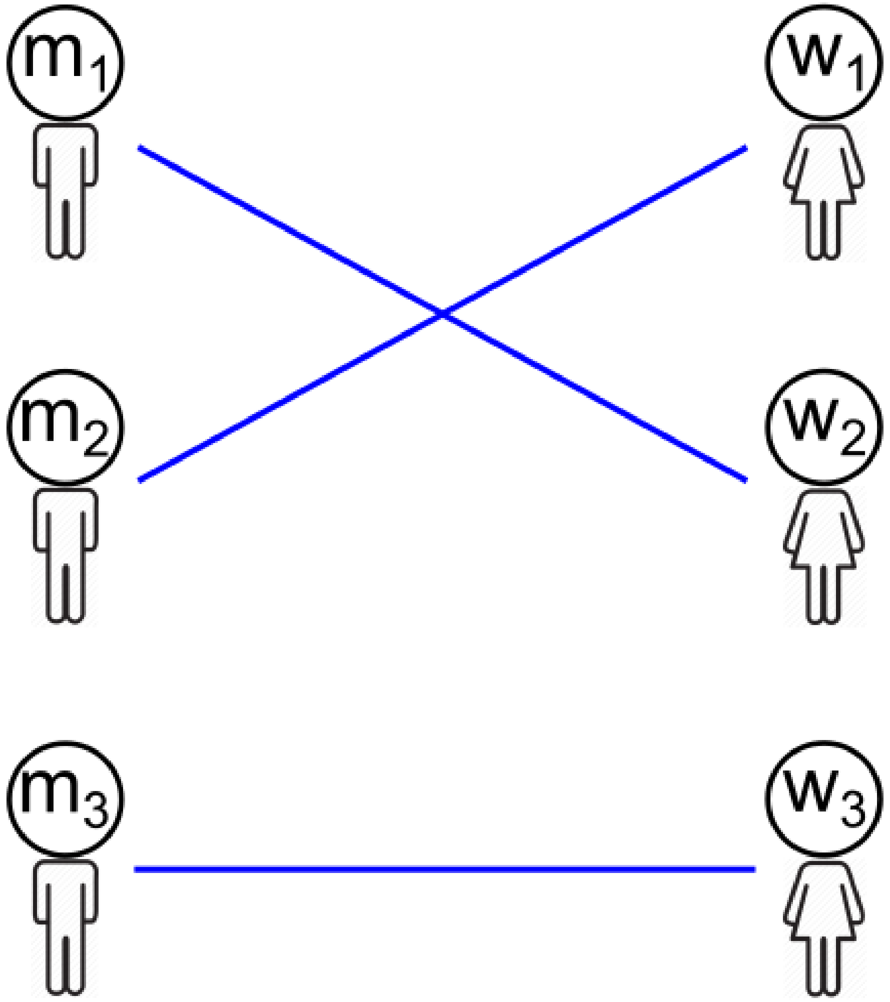
$w_4: m_5 > m_1 > m_4 > m_3 > m_2$

$w_5: m_1 > m_2 > m_3 > m_4 > m_5$

Goal for Today

Understanding the structure of the set of stable matchings through linear programming.

(This will guide us towards fair stable matchings.)



$$
P = \begin{array}{c} \\ m_1 \\ m_2 \\ m_3 \end{array}
\begin{array}{c} \begin{array}{ccc} w_1 & w_2 & w_3 \end{array} \\
\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array}
$$

Fractional Stable Matching

Fractional Stable Matching

Any non-negative nxn matrix X satisfying the following:

Fractional Stable Matching

Any non-negative nxn matrix X satisfying the following:

$X_{i,j} \geq 0$ for all $i \in [n]$ and $j \in [n]$

Fractional Stable Matching

Any non-negative nxn matrix X satisfying the following:

$X_{i,j} \geq 0$ for all $i \in [n]$ and $j \in [n]$

$\sum_j X_{i,j} = 1$ for all $i \in [n]$ *Every man is fully matched.*

Fractional Stable Matching

Any non-negative nxn matrix X satisfying the following:

$X_{i,j} \geq 0$ for all $i \in [n]$ and $j \in [n]$

$\sum_j X_{i,j} = 1$ for all $i \in [n]$ *Every man is fully matched.*

$\sum_i X_{i,j} = 1$ for all $j \in [n]$ *Every woman is fully matched.*

Fractional Stable Matching

Any non-negative nxn matrix X satisfying the following:

$$X_{i,j} \geq 0 \text{ for all } i \in [n] \text{ and } j \in [n]$$

$\sum_j X_{i,j} = 1 \text{ for all } i \in [n]$ *Every man is fully matched.*

$\sum_i X_{i,j} = 1 \text{ for all } j \in [n]$ *Every woman is fully matched.*

$$X_{i,j} + \sum_{k: w_k \succ_{m_i} w_j} X_{i,k} + \sum_{k: m_k \succ_{w_j} m_i} X_{k,j} \geq 1 \text{ for all } i \in [n] \text{ and } j \in [n]$$ *Stability*

Fractional Stable Matching

Any non-negative nxn matrix X satisfying the following:

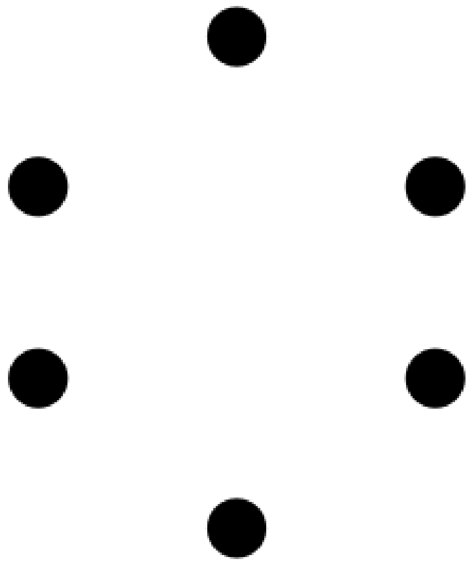
$X_{i,j} \geq 0$ for all $i \in [n]$ and $j \in [n]$

$\sum_j X_{i,j} = 1$ for all $i \in [n]$ *Every man is fully matched.*

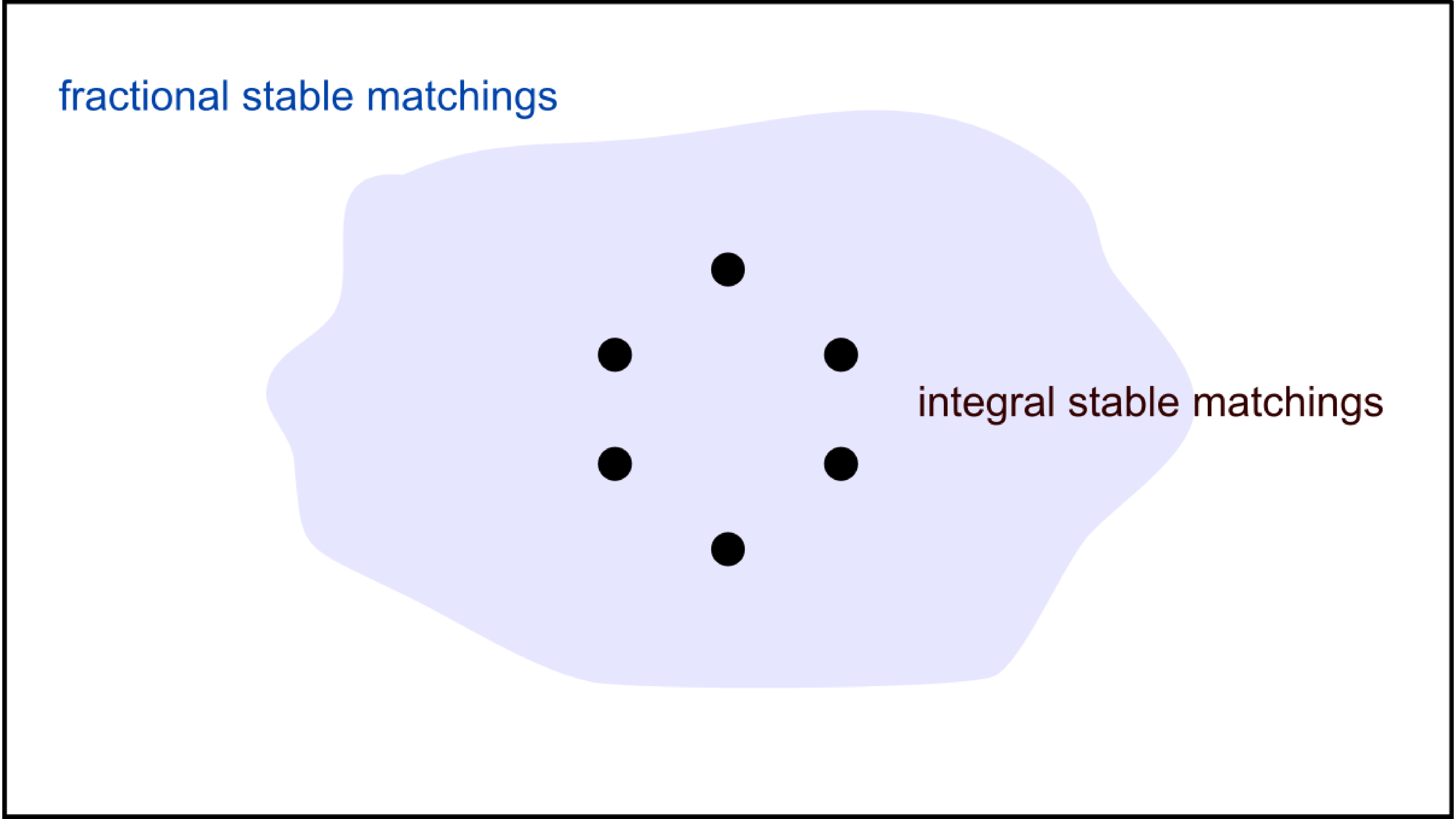
$\sum_i X_{i,j} = 1$ for all $j \in [n]$ *Every woman is fully matched.*

$$X_{i,j} + \sum_{k:w_k \succ_{m_i} w_j} X_{i,k} + \sum_{k:m_k \succ_{w_j} m_i} X_{k,j} \geq 1 \text{ for all } i \in [n] \text{ and } j \in [n]$$ *Stability*

Any integral stable matching is also a fractional stable matching.



integral stable matchings


fractional stable matchings
integral stable matchings