CHAPTER 1.2

1 ~~23~~. BÖHM-BAWERK'S HORSE MARKET

In this celebrated model, the traded goods are identical (indivisible) horses and (i) each participant wants to consume at most one horse, (ii) each participant owns either one horse or no horse. Thus, the participants are partitioned into (potential) *sellers* who own a horse (hence are not interested in buying one) and (potential) *buyers* who own none. Moreover, agent i's preferences are described by a single nonnegative number, representing agent i's willingness to pay for a horse (if she is buyer), or equivalently her reservation price for selling her horse (if she is seller).

[9] Consider a feasible allocation $((q_i, t_i),\ i \in N)$ that is Pareto inferior to another feasible allocation $((q_i', t_i'),\ i \in N)$: $u_i(q_i) + t_i \leq u_i(q_i') + t_i'$ for all i with at least one strict inequality. Summing up and taking $\Sigma_N\, t_i = \Sigma_N\, t_i'$ into account (a consequence of feasibility), we get $\Sigma_N\, u_i(q_i) < \Sigma_N\, u_i(q_i')$. Conversely, if $Z = ((q_i, t_i,\ i \in N)$ is a feasible allocation and $\Sigma_N\, u_i(q_i) < \Sigma_N\, u_i(q_i')$ for some feasible allocation $(q_i',\ i \in N)$ of the nonmonetary goods, then Z is Pareto inferior to Z', given by $t_i' = t_i + (u_i(q_i) - u_i(q_i')) + \delta/n$, where $\delta = \Sigma_N(u_i(q_i') - u_i(q_i))$.

A useful computational device is to order the buyers by decreasing utilities (for a horse) and the sellers by increasing utilities:

$$u_1 \geq u_2 \geq \cdots \geq u_m, \qquad v_i \leq v_2 \leq \cdots \leq v_n, \tag{1}$$

where $u_j = u_j(h)$ is buyer j's utility for a horse, and similarly, $v_i = u_i(h)$ for seller i. We also denote by b_j the buyer with valuation u_j and by s_i the seller with valuation v_i. Consider the following example:

$$\text{5 buyers:} \quad 10 \geq 8 \geq 7 \geq 4 \geq 3, \qquad \text{6 sellers:} \quad 1 \leq 3 \leq 3 \leq 5 \leq 6 \leq 8. \tag{2}$$

Compute first the efficient allocations, namely, the optimal assignment of the 6 horses among the 11 agents. An assignment is described by the set of active buyers (who end up with a horse) and a set (of the same cardinality) of active sellers (who end up with no horse). Of course, an inactive buyer ends up with no horse and an inactive seller stays with her horse. In an efficient assignment, if buyer b_{j_1} is active while buyer b_{j_2} is not, we must have $u_{j_1} \geq u_{j_2}$ (otherwise a transfer of one horse from j_1 to j_2 enhances the joint utility); thus the q^* active buyers in an efficient assignment must have the q^* highest utilities among buyers, and similarly, the q^* active sellers must have the q^* lowest utilities among sellers. To determine the efficient number q^* in our example, we compute the best feasible surplus, namely, the largest difference in joint utility between a feasible allocation and the initial allocation (this is enough in view of the general characterization of efficiency under transferable utility):

best surplus	with one active buyer:	$10 - 1 = 9$,
	with two active buyers:	$9 + (8 - 3) = 14$,
	with three active buyers:	$14 + (7 - 3) = 18$,
	with four active buyers:	$16 + (4 - 5) = 15$,
	with five active buyers:	$17 + (3 - 6) = 14$.

Clearly, $q^* = 3$ and the general formula for q^* is as follows (given the inequalities (1)). The optimal number of "trades" is the largest number q^* such that $v_{q^*} < u_{q^*}$ and $u_{q^*+1} < v_{q^*+1}$ if such a number exists (with the convention $u_{m+1} = v_0 = -\infty$ and $v_{m+1} = u_0 = +\infty$; for instance, if $u_1 < v_1$, then $q^* = 0$). If such a number does not exist, there exists at least one index such that $v_q = u_q$; in this case, the optimal number of trades can be anywhere between the largest q such that $v_q < u_q$ and the largest q such that $v_q = u_q$ (whether or not a seller and a buyer with the same utility for a horse do trade is a matter of social indifference).

Among Pareto optimal allocations, we may restrict our attention to those where trade is voluntary (this is a simple consequence of private ownership: no one can be forced to trade), namely, an active buyer pays no more than his utility and an active seller receives at least her utility. This may entail a nonuniform transaction price: in our example, the buyers may pay respectively 6, 2, and 5 and the sellers may be paid respectively 4, 5, and 4. The competitive equilibrium price rules out nonuniform transactions by definition.

Define a *price p* to be *competitive* if at that price, total demand equals total supply. In our example, any price p between 4 and 5 is competitive because (the first) three buyers and (the first) three sellers want to trade. At price $p = 4$, exactly three sellers want to buy, the first three buyers definitely want to trade, while b_4 is indifferent between buying or not; we still call $p = 4$ a competitive price because one of the possible demands equals the unique supply.

The general formula is as follows (see Exercise 2.2). Let q^* be the largest index such that $v_{q^*} \leq u_{q^*}$ (remember our convention $u_{m+1} = v_0 = -\infty$ and $v_{n+1} = u_0 = +\infty$). Then the competitive prices cover the interval $[\sup(u_{q^*+1}, v_{q^*}), \inf(u_{q^*}, v_{q^*+1})]$.

Thus the competitive price is unique when there is an index q such that $v_q = u_q$ or when $u_{q^*} = u_{q^*+1}$ or when $v_{q^*} = v_{q^*+1}$. A justly famous example is the gloves market[10]: we have m identical buyers with common utility u and n identical sellers with common utility v, and $u > v$ (interpretation: each buyer wants a left-hand glove to match the right-hand glove he owns; each seller owns a left-hand glove *and* does not wear gloves). When there are fewer buyers than sellers ($m < n$), then $q^* = m$ and $v_{q^*+1} = v_{q^*} = v$, so the competitive price is v and buyers reap all the net surplus; symmetrically, with a shortage of sellers with respect to buyers ($n < m$), the competitive price is u ($q^* = n$, $u_{q^*} = u_{q^*+1}$), and the buyers get no surplus whatsoever. We come back to the gloves market in Section 7.7 (see Example 7.12) to compare this competitive allocation with the less extreme allocation resulting from the Shapley formula.

To each competitive price corresponds a set of *competitive allocations*, characterized by two properties:

(a) Each transfer of money t_i is the price of whatever horse agent i

[10] See Shapley and Shubik [1969a].

owns initially minus the price of the horse she consumes in the allocation in question (so in a Böhm–Bawerk market, $t_i = 0$ for an inactive agent, $t_i = p$ for an active seller, and $t_i = -p$ for an active buyer).

(b) An agent's own allocation maximizes her preferences over the set of all consumption vectors she can afford by selling her initial endowment at the competitive price (in this example, this means $u_j \leq p$ for an inactive buyer, $u_j \geq p$ for an active buyer, $v_i \leq p$ for an active seller, and $v_i \geq p$ for an inactive seller). Finally, we shall call *competitive set* the set of all competitive allocations for all possible competitive prices.

Turning to core stability, we find that all transactions must take place at the same price, implying that the core allocations coincide with the competitive allocations. We check this general fact in the numerical example (2). A core allocation must be Pareto optimal; hence, we know that the first three buyers (and only them) and the first three sellers (and only them) are active. Let p_j be the price paid by b_j and r_i that received by s_i ($1 \leq i, j \leq 3$). Suppose $p_j > r_i$ for some b_j and some s_i. Then the pair $\{b_j, s_i\}$ has an objection: they are both made better off by trading at price $(p_j + r_i)/2$ (b_j pays less and s_i earns more). Thus $p_j \leq r_i$ for all $i, j, 1 \leq i, j \leq 3$. However, feasibility implies $\Sigma_j p_j = \Sigma_i r_i$; hence, all six numbers p_i, r_j are equal to a common value p. It remains to check that p is between 4 and 5. If $p > 5$, then s_4 can form an objection with anyone of the active buyers (say b_1). Remember that s_4 is inactive by Pareto optimality. By trading with b_1 at price $(p + 5)/2$, s_4 does make a positive profit *and* gives a better deal to b_1. Thus $p > 5$ is impossible, and $p < 4$ is similarly ruled out. We have shown a particular case of a general fact:

In a Böhm–Bawerk's market, the competitive set and the core are nonempty and coincide. (3)

2 ~~3.2.~~ House Barter

This model was originally introduced by Shapley and Scarf [1974]. We have n agents, and n houses (all different). An agent can consume only one house (she can live in one place at a time) and cannot be homeless. Each agent ranks the houses from his first choice to his last choice, without ever being indifferent between two houses.[3] Initially, each agent owns a house (so there is a one-to-one assignment of houses to agents). Agents can exchange houses (within a given coalition, all rearrangements of their initial houses are feasible), but there is no medium of exchange such as money.

Remarkably, the house barter model *always* yields a unique core allocation (a rare occurrence when the domain of individual preferences is so large!).

Here is an example with four agents. Initially, agent i owns house h_i. The four individual preferences are as follows:

	u_1	u_2	u_3	u_4
top	h_3	h_4	h_1	h_3
	h_2	h_1	h_4	h_2
	h_4	h_2	h_3	h_1
bottom	h_1	h_3	h_2	h_4

. (1)

Agents 1 and 3 can swap houses and end up with their top choice. Therefore (a) the initial allocation is not Pareto optimal, and (b) any allocation in the core must give h_3 to agent 1 and h_1 to agent 3. For instance, the Pareto optimal allocation

$$h_2 \text{ to } 1, \qquad h_4 \text{ to } 2, \qquad h_1 \text{ to } 3, \qquad h_3 \text{ to } 4$$

is "blocked" by an objection of coalition $\{1,3\}$ that is able to guarantee, of its own resources, a better house to agent 1 and the same house to agent 3 (remember that an objection need not improve the welfare of every coalition member, but only of some members without deteriorating that of any member).

Now there are only two allocations where 1 and 3 swap houses, namely, the one where 2 and 4 keep their initial house, and the one where they swap. As the swap is Pareto improving, we are left with a unique candidate for a core allocation, namely,

$$h_3 \text{ to } 1, \qquad h_4 \text{ to } 2, \qquad h_1 \text{ to } 3, \qquad h_2 \text{ to } 4$$

It is easy to check that this allocation is not threatened by any other coalition.

[3] This assumption of strict preferences is important. When we allow indifferences in an agent's preferences, only some of the results stated in this section survive; see Exercise 3.4.

Lemma 3.1. The Top Trading Cycle Algorithm. The following algorithm defines the unique core outcome of the house trading model among the agents of $N = \{1, 2, \dots, n\}$.

For each agent $i \in N$, *let* $\sigma(i) \in N$ *be the owner of agent i's top house. Find a cycle of the mapping* σ *that is, a sequence* $i_1, \dots, i_K$ *such that* $\sigma(i_k) = i_{k+1}$, $k = 1, \dots, K-1$ *and* $\sigma(i_K) = i_1$ *and exchange houses along a cycle (so agent* i_k *gets the house owned by agent* i_{k+1}, *with the convention* $i_{K+1} = i_1$*). Note that the cycle may be of length* 1 *(when an agent's top choice is the house she initially owns), in which case this agent will simply not trade. For any mapping* σ, *there exists at least one such cycle, called a "top trading cycle."*

Call N_1 *the union of all top trading cycles of* σ *and let all agents of* N_1 *get their top choice. Repeat the operation within* $N \setminus N_1$ *(*$\sigma_1(i)$ *is the owner of agent i's top house within those owned by the agents in* $N \setminus N_1$, *and exchange houses along all top trading cycles of* σ_1*). Call* N_2 *the union of all top trading cycles of* σ_1 *and repeat the operation within* $N \setminus (N_1 \cup N_2)$, *and so on.*

Proof. Consider a top trading cycle of σ. The agents in this cycle can all achieve their top choice without the help of anyone outside the cycle; thus, a core outcome must serve N_1 precisely as in the algorithm. Given that N_1 is served in this way, a coalition in $N \setminus N_1$ cannot be better than when everyone gets her first choice in $N \setminus N_1$; on the other hand, a cycle of σ_1 in $N \setminus N_1$ can, as a coalition, guarantee his top choice in $N \setminus N_1$ to each of its members. Hence, in a core outcome, the agents of N_2 are served as in the algorithm (they get their top choice in $N \setminus N_1$), and so on. Thus the algorithm defines the only conceivable core outcome.

It remains to check that this outcome, call it z^*, is indeed in the core. Consider a coalition S and an allocation z_S feasible by exchanges within S, and guaranteeing to all in S as least the same level of welfare as that awarded by z^*. If the intersection of S with N_1 is not empty, it must consist of the union of some cycles of σ (otherwise the top choice of someone in $S \cap N_1$ belongs to an agent outside S); and z_S coincides with z^* on $S \cap N_1$. Repeating this argument we find that $S \cap N_2$ consists of the (possibly empty) union of cycles of σ_2, and z_S coincides with z^* on $S \cap N_2$, and so on. Q.E.D.

Notice that a looser algorithm following an arbitrary sequence of Pareto improving trading cycles (that may allocate second choice or worse to some agents) would not always reach the core. This can be seen in the simple example (1). Say that agents 1 and 2 exchange houses first (a Pareto improving move). The resulting allocation is

$$h_1 \to 2, \quad h_2 \to 1, \quad h_3 \to 3, \quad h_4 \to 4,$$

from which any sequence of Pareto improving trades (e.g., 2 and 4 trade houses, then 3 and 4 trade) results in the allocation

$$h_1 \to 3, \quad h_2 \to 1, \quad h_3 \to 4, \quad h_4 \to 2,$$

not the core allocation! Thus agent 1 made a mistake (from which agent 4 benefited) by accepting to trade for a less than top house with agent 2.

Remark 3.1

An additional feature of the core allocation achieved by the top trading cycle algorithm is this: any other allocation is blocked by an objection of a coalition using the very allocation that this coalition gets in the core. See Exercise 3.1.

The top trading cycle algorithm may take up to n rounds (where the first round consists of finding N_1 and exchanging houses along the cycles of N_1, the second round does the same for N_2, and so on). Yet the number of rounds does not reflect the intensity of trading in the core. Quite the contrary, we may have a single round where everyone trades (when σ has a single cycle comprising all agents in N), whereas we have n rounds only if the initial allocation is Pareto optimal and no trade takes place.[4] The successive rounds correspond, in fact, to the relative value of houses in the competitive equilibrium.

Lemma 3.2. Let T be the number of rounds in the algorithm of Lemma 3.1, so that $N_1 \cup N_2 \cup \cdots \cup N_T$ is a partition of N. Then a competitive equilibrium price obtains as follows. Pick a decreasing sequence $q_1 > q_2 > \cdots > q_T$ and let the price of all houses in N_t be q_t, for all $t = 1, \ldots, T$. Moreover, the core outcome is also the unique competitive equilibrium outcome.

Proof. There is certain oddity in talking about competitive price in a context where there is actually no medium of exchange. Yet we may create a fictitious unit of account (fiat money), call prices in this unit, and derive a (very real) budget constraint for each market participant. Consider the core allocation z^* and a price vector as described in the

[4] See Exercise 3.2.

Lemma. An agent N_t owns a house that is worth q_t, and thus she can afford any house in $N_t, N_{t+1}, \ldots, N_T$. Allocation z^* gives her top choice in $N_t \cup N_{t+1} \cup \cdots \cup N_T$, as required by the competitive equilibrium (maximization of her utility within the limits of her budget constraint).

Check now that a competitive equilibrium allocation must be precisely the core allocation z^*. Let p be a competitive price and z a corresponding competitive allocation. In z, houses are exchanged along certain trading cycles $M_1, M_2, \ldots, M_S$: the agents in M_s exchange houses in circular fashion (i_k gets the house owned by i_{k+1}). Ultimately, no money changes hands; hence, the price of all houses within a given cycle must be the same: all houses traded in M_s have the same price q_s. Without loss of generality, assume $q_1 \geq q_2 \geq \cdots \geq q_S$. Then the agents in M_1 get their top choice (so M_1 is a top trading cycle), those in M_2 get their top choice among the houses owned by the agents in $N \setminus M_1$ (clearly, if $q_1 = q_2$, M_2 is a top trading cycle in N as well; and even if $q_1 > q_2$, M_2 may be a top trading cycle in N), those in M_3 get their top choice among houses owned in $N \setminus (M_1 \cup M_2)$ and so on. The conscientious reader will check that this algorithm picks precisely the core outcome z^*, because N_1 is, for instance, the union $M_1 \cup M_2 \cup M_3$, N_2 is the union $M_4 \cup M_5$, and so on. The minor variations in the competitive price do not change the final allocation. Q.E.D.

The miraculous coincidence of the core and/or competitive equilibrium set into a single outcome is not the end of the story. Look at the direct mechanism that mechanically computes the outcome z^* from the reports of individual preferences. This mechanism is strategy-proof in the strong sense that even a joint misreport by a coalition cannot be profitable.

Lemma 3.3. The direct revelation mechanism implementing the unique core allocation is nonmanipulable: no coalition can gain (*that is, make at least one coalition member better off and none worse off*) *by jointly misreporting preferences.*

Proof. For the true preference profile, denote, as in Lemma 3.2, by $N_1, \ldots, N_T$ the partition of N resulting from the top trading cycle algorithm. Let S be a manipulating coalition and let t^*, $1 \leq t^* \leq t$, be the smallest index such that $S \cap N_{t^*} \neq \emptyset$. Clearly, when the agents in N^*, $N^* = \bigcup_{t=1}^{t^*-1} N_t$, report truthfully, they get their core house no matter what the agents outside N^* report (this is clear from the definition of the algorithm). Therefore, when S misreport, but $N \setminus S$ is truthful, the best house that an agent in S can hope for is her top choice among those houses not owned initially by an agent in N^*; but this is precisely what all agents in $S \cap N_{t^*}$ get by reporting truthfully! Thus, if the misreport by S is profitable to S, the agents in $S \cap N_{t^*}$ actually get no more and no less than their core house. We can now repeat the argument to show that the agents in $S \cap N_{t^*+1}$ get their core house as well, and so on. In the end, the manipulation does not strictly improve the welfare of anyone in S. Q.E.D.

Remark 3.2

Another source of potential manipulation by coalitions consists of exchanging houses *prior* to the implementation of the core mechanism. Here is an example with three agents and three houses. Initially, agent i owns house h_i, and their preferences are as follows:

u_1	u_2	u_3
h_3	h_1	h_1
	h_2	h_2
	h_3	h_3

With truthful report of preferences and of initial property rights, the first top trading cycle involves 1 and 3 and the resulting core allocation is

$$h_1 \to 3, \qquad h_2 \to 2, \qquad h_3 \to 1.$$

However, if agents 1 and 2 trade houses (not always a Pareto improving move) before showing up to play the core mechanism, the resulting allocation is

$$h_1 \to 2, \qquad h_2 \to 3, \qquad h_3 \to 1;$$

hence, agent 2 strictly benefits and agent 1 does not lose. As it turns out, such a manipulation can never be strictly profitable for all members of the coalition (trading houses prior to the mechanism). See Exercise 3.3. Notice also that the coalition (12) could achieve the same outcome by (i) misreporting their preferences, and (ii) exchanging houses ex post. (Exercise: prove this claim.)

To conclude this section, we discuss the robustness of the house barter model in two directions. First, allow the individual preferences to exhibit indifferences. Then a fair amount of the above results are preserved, in particular:

(a) *There is at least one competitive allocation and the competitive set is obtained by running the top trading cycle algorithms (exploring all the options created by indifferences).*
(b) *The core is a subset, possibly strict and possibly empty, of the set of competitive allocations. All core allocations yield the same utility distributions.*

Property (a) is straightforward (its proof follows the argument in the proof of Lemma 3.2). Property (b) (due to Wako [1991]; its proof is the subject of Exercise 3.5) is more surprising: we normally expect the core to be a superset of the competitive set! Here is a three-agent example showing that the core may be empty:

u_1	u_2	u_3
h_2	h_1, h_3	h_2
h_1, h_3	h_2	h_1
		h_3

Initially, agent i owns h_i. Using top trading cycles, we find two competitive allocations, namely,

$$z: \quad h_1 \to 2, \quad h_2 \to 1, \quad h_3 \to 3,$$

and

$$z': \quad h_1 \to 1, \quad h_2 \to 3, \quad h_3 \to 2.$$

Now z is blocked by an objection of $\{2,3\}$ and z' is blocked by an objection of $\{1,2\}$. Examples where the core contains a unique allocation and the competitive set contains at least one more allocation are given in Exercise 3.4.

The second way to generalize the house barter model is more radical. Suppose that we allow the agents to trade two kinds of goods, say, houses and cars. Assume that each agent initially owns one car and one house and consumes exactly one car and one house. Suppose, moreover, that individual preferences are separable between houses and cars (which car I own does not affect my ranking of the various houses and vice versa). Then it is not clear whether the core of the economy will be empty or not (no systematic result is available yet). Of course, if we allow nonseparable preferences, the core may be empty. (Exercise: give an example.)

3.3. The Marriage Market

This important model[5] is another instance where the core concept is more successful than the competitive equilibrium. Recall from Section 2.6 that the competitive concept collapses when a good is produced under increasing returns to scale; in that case, the core is a reasonable substitute, which, however, may cut a large subset of the Pareto optimal frontier (and may easily be empty, too; see Section 2.8). By contrast, in the marriage market, the core is always nonempty and often small; moreover, its end points (in a sense to be made precise soon) are easy to compute.

We have a set M of men and a set W of women. Each man (and each woman) has strict preferences over his (her) potential spouses. Marriages are exclusively monogamous and heterosexual.[6] We assume for simplicity that there is the same number of men and of women. This is of no consequence from a technical standpoint: all key results are maintained when M and W are of different size, and we view "remaining single" as an option that each agent ranks among the set of his potential mates (so that some mates are less desirable than the celibacy).[7]

A *matching* is a one-to-one pairing of each man to a woman. The core stability property follows naturally from the fact that each individual owns himself or herself. Say that a certain matching is proposed; if a certain pair of one man and one woman prefers one another over their proposed mate, then the matching in question is not in the core. Call a matching *stable* if no pair of one man and one woman can object as above; this is enough for core stability because of the assumption of strict preferences. In particular, pairwise stability implies Pareto optimality.[8]

We can interpret the marriage market as an exchange economy in which (a) each man and each woman owns a personalized indivisible good, (b) each man (resp. woman) wants to consume at most one of the goods initially held by women (resp. by men) and derives no utility from

[5] For a survey of its practical applications to several job-matching problems, see Roth and Sotomayor [1990].

[6] Other examples are a set of firms and a set of computer specialists (when each firm needs exactly one computer specialist), or a set of pianists and a set of violinists when there is a market for piano–violin duos (but not for trios).

[7] See Exercise 3.10 for more discussion. Of course, in realistic models of marriage, the celibacy is always an important option; e.g., there is a market for solo pianists or solo violinists.

[8] If the matching σ is Pareto inferior to the matching σ', there is a man m preferring $w' = \sigma'(m)$ over $w = \sigma(m)$. Then w' prefers $m = \sigma'(w')$ over $m' = \sigma(w')$ (remember that w' is never indifferent between any two men); hence the pair (m, w') blocks the matching σ.

the goods held by men (resp. women). There is no medium of exchange such as money.[9]

To discover a core stable matching, a simple heuristic can be helpful. Start with an arbitrary matching. If we find a blocking pair, match them and match their former mates together. Continue until, hopefully, we reach a stable outcome. Here is an example with four men, four women, and the following preferences:

m_1	m_2	m_3	m_4
w_4	w_4	w_3	w_4
*	w_3	w_1	w_1
*	*	*	*
*	*	*	*

,

w_1	w_2	w_3	w_4
m_2	*	m_1	m_1
m_1	*	m_2	*
m_4	*	*	*
m_3	*	*	*

. (2)

Start from the matching (m_i, w_i), $i = 1,2,3,4$. Here (m_1, w_4) object; so we marry them and (forcibly) marry w_1 to m_4. Now (m_2, w_3) have an objection; we marry them, as well as m_3 to w_2. Hence

$$(m_1, w_4) \qquad (m_2, w_3) \qquad (m_3, w_2) \qquad (m_4, w_1).$$

Check that this is a stable matching: for instance, although w_2 may prefer any other man to m_3, she cannot convince any of these to switch. In fact, this matching is the only matching in the core (out of 24 possible matchings; one can check directly that there is no other core matching, but it is much easier to apply Theorem 3.1 by checking that the M-optimal and the W-optimal matchings coincide).

Observe that this core matching can be interpreted as a competitive equilibrium. Set the same price for a man and a woman if they are married and choose these prices as follows:

$$p_{m_3} = p_{w_2} < p_{m_4} = p_{w_1} < p_{m_2} = p_{w_3} < p_{m_1} = p_{w_4}.$$

Then check that the price of any woman (resp. man) preferred by man m (resp. woman w) to his (her) core mate is higher than his (her) own price, so that no one can afford to buy a preferred mate (and everyone can afford to buy his/her current mate).

However, this interpretation of the core matching as a competitive allocation is often impossible. A simple example with three men and

[9] When we add money to the picture, the bilateral assignment model of Section 3.4 obtains.

three women is as follows:

m_1	m_2	m_3
w_1	w_1	w_2
*	w_2	w_3
*	w_3	w_1

,

w_1	w_2	w_3
m_3	m_2	m_3
m_1	*	*
m_2	*	*

. (3)

Here the unique stable matching is

$$(m_1, w_1) \quad (m_2, w_2) \quad (m_3, w_3).$$

(Indeed, any matching containing (m_2, w_1) is blocked by (m_1, w_1) and any matching containing (m_3, w_1) is blocked by (m_3, w_3); thus, we must have (m_1, w_1) in the core. In the market reduced to m_2, m_3, w_2, w_3, the matching (m_2, w_2) (m_3, w_3) gives everyone his/her top choice.) Yet if there exists a competitive price system, it must satisfy

$p_{m_i} = p_{w_i}$ for $i = 1, 2, 3$

(so that married agents can afford each other),

$p_{w_1} < p_{m_3}$ so w_1 cannot afford m_3,

$p_{m_3} < p_{w_2}$ so m_3 cannot afford w_2,

$p_{m_2} < p_{w_1}$ so m_2 cannot afford w_1.

These equalities and inequalities are inconsistent.

Our last example, before stating some general results about the marriage market, is meant to illustrate (i) that the core may contain several different matchings, and (ii) that the naive algorithm used above may not converge to a stable matching:

m_1	m_2	m_3
w_2	w_1	w_1
w_1	w_3	w_2
w_3	w_2	w_3

,

w_1	w_2	w_3
m_1	m_3	m_1
m_3	m_1	m_3
m_2	m_2	m_2

. (4)

Start with the initial matching

$$(m_1, w_1) \quad (m_2, w_2) \quad (m_3, w_3).$$

There are objecting pairs, namely, (m_1, w_2) and (m_3, w_2). If we choose to satisfy (m_1, w_2), we reach

$$(m_1, w_2) \quad (m_2, w_1) \quad (m_3, w_3),$$

where, again, two objecting pairs emerge: (m_3, w_2) and (m_3, w_1). Say that we choose (m_3, w_2) so as to reach

$$(m_1, w_3) \quad (m_2, w_1) \quad (m_3, w_2).$$

Here, again, we have a choice of objecting pairs; say that we choose (m_3, w_1) (the other objecting pair is (m_1, w_1). We are now matching

$$(m_1, w_3) \quad (m_2, w_2) \quad (m_3, w_1),$$

from which the objection by (m_1,w_1) brings us back (after four steps) to the original matching! On the other hand, by making different choices of objecting pairs, we reach quickly the two stable matchings of this market, namely,

$$(m_1, w_2) \quad (m_2, w_3) \quad (m_3, w_1) \quad \text{(via the } (m_1, w_2) \text{ objection, then the } (w_1, m_3) \text{ objection),} \tag{5}$$

$$(m_1, w_1) \quad (m_2, w_3) \quad (m_3, w_2) \quad \text{(via the } (m_3, w_2) \text{ objection).} \tag{6}$$

Notice that every man and every woman has a different mate in these two matchings, and that all men prefer (5) to (6), whereas all women prefer (6) to (5). This feature is quite general.

Theorem 3.1. (Gale and Shapley [1962]). In any marriage market with strict preferences:

(a) There is at least one stable matching; the core is never empty.

(b) There is a stable matching, called the M-optimal matching, where every man gets the best of all his core mates (there is no stable matching where he is matched to a preferred woman) and every woman gets her worst core mate. There is a stable matching, called the W-optimal matching, where every woman gets the best of all her core mates and every man gets his worst core mate. The core contains a single matching if and only if the M-optimal matching and the W-optimal matching coincide.

(c) The M-optimal matching is computed by means of the Gale–Shapley algorithm where men propose (defined below).

(d) A statement symmetrical to (c) holds for the W-optimal matching.

Definition 3.1. The Gale–Shapley Algorithm where Men Propose

STEP 1

Each man proposes to his first-choice woman. If a woman receives exactly one proposal, this man is called her engagee. If a woman receives more than one proposal, she keeps the proposer she likes best as her engagee and rejects the others. Men are now partitioned into engaged or rejected. The algorithm stops if all men are engaged otherwise, we go to the next step.

STEP 2

All rejected men propose now to their second-best choice. Each woman receiving new proposals keeps as her engagee the man she likes best among current proposer(s) and possibly former engagee, and rejects the others. (Thus a man previously engaged may now be rejected.) The algorithm stops if all men are engaged; otherwise, we go to the next step.

STEP 3

All rejected men propose now to their next choice, and women update their engagements according to new proposals (if any).

This continues until all men are finally engaged, at which point the engagement pattern turns into the final matching. Since each man proposes to any woman only once, the algorithm must stop after finitely many steps. The proof of Theorem 3.1 is in Appendix 3.1.

An example is the three-men, three-women market (4). If men propose, w_2 receives an offer from m_1 while w_1 receives two offers and keeps m_3. Next round, m_2 offers to w_3, and the algorithm stops on matching (5). If women propose, m_1 receives two offers, keeps w_1 and rejects w_3, while m_3 receives an offer from w_2. Next, w_3 offers to m_3, who still keeps w_2. After this second rebuttal, w_3 finally offers to m_2, and the algorithm stops on matching (6). A more complicated example of the Gate–Shapley algorithm is given in Exercise 3.6.

Remark 3.3

Many more interesting properties of stable matching are covered in Roth and Sotomayor [1990]. For instance, the M-optimal matching is shown to be weakly Pareto optimal from the men's point of view; see Exercise 3.7. More importantly, the core possesses a lattice structure by means of the following "supremum" and "infimum" operations. If μ and μ' are two stable matchings, construct the matching $\mu \vee \mu'$ as follows; match every man m to whomever he prefers from his mate in μ and his mate in μ'. If μ and μ' are two arbitrary matchings, this construction may not yield a one-to-one matching from M onto W, but if μ and μ' are both stable, it does. Moreover, $\mu \vee \mu'$ is a stable matching as well, and it matches every woman with whomever she likes least among her mate in μ and her mate in μ'; see Exercise 3.8.

The core of a marriage market is easy to estimate, because the M-optimal and W-optimal matchings are its two bounds (they give utility bounds for each agent). Moreover, the M-optimal matching can be used as a direct mechanism to implement a stable matching. Although this mechanism is not strategy-proof, its strategic properties are still very strong.

Lemma 3.4. In the direct mechanism where agents report their preferences after which the M-optimal matching is implemented, truthtelling is a dominant strategy for the men, although not for the women. Assuming that men report truthfully, the optimal manipulation of the women yields a stable matching.

The proof is omitted (see Roth and Sotomayor [1990].

Consider, for example, the M-optimal mechanism in economy (4). If all report truthfully, the matching (5) results. A (small) manipulation by w_1 will be enough to bring about the matching (6), a strict improvement for both w_1 and w_2. Woman w_1 reports $m_1 > m_2 > m_3$, and the Gale–Shapley algorithm works as follows:

STEP 1: w_1 receives offers from m_2 and m_3 and (untruthfully) keeps m_2; w_2 has an offer from m_1.

STEP 2: m_3 offers to w_2, who therefore keeps that offer and rejects m_1.

STEP 3: m_1 offers to w_1, who keeps that offer and rejects m_2.

STEP 4: m_2 offers to w_3, and the matching (6) is reached.

Thus the gap between the M-optimal and the W-optimal matchings serves as an upper bound on the possible extent of strategic manipulation in these two direct mechanisms (implementing respectively the M-optimal and the W-optimal matchings).