An Elementary Proof of the Lone Wolf Theorem*

Andrei Ciupan
Department of Economics
Harvard University

John William Hatfield
McCombs School of Business
University of Texas at Austin

Scott Duke Kominers
Society of Fellows
Harvard University

August 31, 2016

Abstract

In this paper, we provide a new, elementary proof of the Lone Wolf Theorem, which shows that any agent who is unmatched in some stable marriage matching is unmatched in all stable marriage matchings. Along the way, we provide a new proof that the market equilibration process of Blum and Rothblum (2002) always produces a stable matching under which the last agent to enter receives his or her best stable matching partner.

JEL Classification: C71, C78, D47

Keywords: Marriage matching, Lone wolf theorem, Market entry

*The authors thank Ravi Jagadeesan and Alvin E. Roth for helpful comments. Kominers gratefully acknowledges the support of National Science Foundation grant SES-1459912 and the Ng Fund of the Harvard Center of Mathematical Sciences and Applications. Any comments or suggestions are welcome and may be emailed to aciupan@hbs.edu, john.hatfield@utexas.edu, and kominers@fas.harvard.edu.

1 Introduction

One of the classic results in matching theory is the "Lone Wolf Theorem," which in its simplest form states that any agent who is unmatched at some stable marriage matching is unmatched in all stable marriage matchings.[1] The earliest version of the Lone Wolf Theorem for one-to-one matching was published by McVitie and Wilson (1970) just eight years after Gale and Shapley's seminal 1962 paper (see also Gale and Sotomayor (1985a)); generalizations have since been obtained for many-to-one matching (Roth (1984, 1986)), many-to-many matching (Alkan (2002); Klijn and Yazıcı (2014)), matching with affirmative action constraints (Kojima (2012)), roommate markets (Klaus and Klijn (2010)), matching with contracts (Hatfield and Milgrom (2005); Jagadeesan (2016)), supply chain matching (Hatfield and Kominers (2012)), and matching in trading networks (Jagadeesan et al. (2016a)).

The Lone Wolf Theorem and its generalizations are central to the now-standard strategy for deriving (one-sided) strategy-proofness results for deferred acceptance (see Hatfield and Milgrom (2005) and Jagadeesan et al. (2016b)). Moreover, many-to-one matching generalizations of the Lone Wolf Theorem, typically called the "Rural Hospitals Theorem," have had real-world policy significance: The Rural Hospitals Theorem implies that switching stable matching mechanisms can not on its own improve the allocation of doctors to rural hospitals; this observation was important in the U.S. National Resident Matching Program redesign (see, e.g., Roth (1984, 1986, 2002, 2008)), and has led to some forms of direct affirmative action for rural hospitals in Japan (see, e.g., Kamada and Kojima (2011, 2012)).

The standard proof of the Lone Wolf Theorem uses the lattice structure of stable matchings (Knuth (1976); Roth and Sotomayor (1990)): a given stable matching is compared, say, to the man-optimal stable matching (which exists because the lattice structure of the set of stable matchings). As every man weakly prefers the man-optimal stable matching to all other

[1] An equivalent alternate formulation is that *the set of matched agents is invariant across stable marriage matchings*.

stable matchings, any man matched under the original matching must also be matched under the man-optimal stable matching. Moreover, every woman weakly disprefers the man-optimal stable matching to the original matching (this is another consequence of lattice structure), so any woman matched under the man-optimal stable matching must also be matched under the original matching. But since each agent has at most one partner, we can only have *both*

1. that all men who are matched under the original matching are matched under the man-optimal stable matching *and*

2. that all women who are matched under the man-optimal stable matching are matched under the original matching

if the same sets of men and women are matched under each.[2]

Here, we give a new proof of the Lone Wolf Theorem. Our argument is "elementary," in the sense that it does not use any structural properties of the set of stable outcomes. Instead of appealing to lattice structure, we consider a process due to Blum and Rothblum (2002) in which agents enter sequentially and the market re-equilibrates to a stable match after each entry (see also Blum et al. (1997)). As Blum and Rothblum (2002) showed, sequential market entry favors the "last agent standing"—the last agent to enter the market obtains his or her best partner achievable across *all* stable matchings. We give a new, self-contained proof of the "last agent standing" result.[3] We then use that result to prove the Lone Wolf Theorem by noting that if some agent i is unmatched at some stable matching, then that matching remains stable when i is removed from the market. If we then allow i to reenter the market and re-equilibrate to a stable matching, the "last agent standing" result implies that i obtains his best partner across all stable matchings. We then show that if i obtains a partner in this way, then the matching we started with (under which i was unmatched) could not have been stable.

[2] Balinski and Ratier (1998) and Klaus and Klijn (2010) gave alternative arguments for the Lone Wolf Theorem based on preference graph constructions.

[3] The Blum and Rothblum (2002) proof relied upon structural results on the set of stable matchings, including the Lone Wolf Theorem itself. Our approach, by contrast, makes use of a preference cycle structure recently identified by Ciupan (2016).

The remainder of this paper is organized as follows: Section 2 introduces the classic marriage model of Gale and Shapley (1962). Section 3 describes the process of Blum and Rothblum (2002) that models market re-equilibration after the entry of an agent. In Section 4 we introduce the "last agent standing" result, and use that result to derive the Lone Wolf Theorem.

2 Model

There are finite sets M and W of *men* and *women*; hence the set of *agents* is $I \equiv M \cup W$. Each man $m \in M$ is endowed with a strict *preference ordering* $\succ_m$ over $W \cup \{\emptyset\}$, where $\emptyset$ is the *outside option*. Similarly, each woman $w \in W$ is endowed with a strict *preference ordering* $\succ_w$ over $M \cup \{\emptyset\}$. We write $i \succcurlyeq_j k$ if $i \succ_j k$ or $i = k$.

A *matching for* $J \subseteq I$ is a map $\mu : J \to J \cup \{\emptyset\}$ from the set of men and women in J to the set of men and women in J plus the outside option such that:

1. Under μ, each man $m \in J \cap M$ is matched to a woman in J or the outside option, i.e., $\mu(m) \in (W \cap J) \cup \{\emptyset\}$, and each woman $w \in J \cap W$ is matched to a man in J or the outside option, i.e., $\mu(w) \in (M \cap J) \cup \{\emptyset\}$.

2. If a man m is matched to a woman w, then w is matched to m (and vice versa), i.e., if $\mu(i) = j \in J$, then $\mu(j) = i$.

We say that agent $i \in I$ is *unmatched under* μ if $\mu(i) = \emptyset$. Throughout, when we introduce a matching μ without explicitly referring to an associated set of agents, we mean that μ is a matching for I.

A matching μ for J is *individually rational* if, for all $i \in J$, we have that $\mu(i) \succcurlyeq_i \emptyset$. A matching μ for J is *unblocked* if there does not exist a *blocking pair* $(m, w) \in (M \cap J) \times (W \cap J)$ such that $w \succ_m \mu(m)$ and $m \succ_w \mu(w)$. A matching μ for J is *stable* if it is both individually rational and unblocked. We say that $i \in I \cup \{\emptyset\}$ is *achievable* for j if there exists a stable matching μ such that $\mu(j) = i$.

3 Market Equilibration Following Entry

We now describe a *market equilibration process* due to Blum and Rothblum (2002) that defines how the market evolves after entry.[4] Formally, for any agent $i \in I \setminus J$, starting with a matching μ for J, we use the following algorithm to find a new matching ν for $J \cup \{i\}$ that we call the *market equilibration outcome after entry by i at μ.*

Step 0: Set $\mu_0 \equiv \mu$ and $i_1 \equiv i$.

Step t: We construct a new matching μ_t as follows: Let $K \equiv \{j \in J \cup \{i\} : i_t \succ_j \mu_t(j)\} \cup \{\emptyset\}$ be the set of agents who prefer i_t to their partner under μ_t (along with the outside option, $\emptyset$). We let j_t be i_t's most-preferred partner in K, i.e.,

$$j_t = \max_{\succ_{i_t}} K = \max_{\succ_{i_t}} (\{j \in J \cup \{i\} : i_t \succ_j \mu_t(j)\} \cup \{\emptyset\}).^{5}$$

- If $j_t \neq \emptyset$, then we match i_t to j_t, setting $\mu_t(i_t) = j_t$ and $\mu_t(j_t) = i_t$.
  - If $\mu_{t-1}(j_t) \neq \emptyset$—$j_t$ is matched under μ_{t-1}—then that partner is set to be unmatched under μ_t, i.e., $\mu_t(\mu_{t-1}(j_t)) = \emptyset$. We keep all other agents' matches unchanged, i.e., $\mu_t(j) = \mu_{t-1}(j)$ for all $j \in ((J \cup \{i\}) \setminus \{i_t, j_t, \mu_{t-1}(j_t)\})$. Next, we let $i_{t+1} = \mu_{t-1}(j_t)$ and proceed to Step $t + 1$.
  - If $\mu_{t-1}(j_t) = \emptyset$—$j_t$ is unmatched under μ_{t-1}—then we set $\mu_t(j) \equiv \mu_{t-1}(j)$ for all $j \in ((J \cup \{i\}) \setminus \{i_t, j_t\})$. Finally, we set $\nu \equiv \mu_t$ and end the algorithm.
- If $j_t = \emptyset$, then we set $\mu_t(i_t) = \emptyset$ and set $\mu_t(j) \equiv \mu_{t-1}(j)$ for all $j \in ((J \cup \{i\}) \setminus \{i_t\})$. Finally, we set $\nu \equiv \mu_t$ and end the algorithm.[6]

Extending the language typically used to describe deferred acceptance (Gale and Shapley (1962)), if agent i_t becomes matched to agent $j_t \neq \emptyset$ in Step t of the market equilibration

[4] A similar market equilibration process featuring two-sided proposals was studied by Dworczak (2016).

[5] Here, we use $\max_{\succ_{i_t}}$ to denote maximization with respect to the preference relation $\succ_{i_t}$.

[6] It is immediate that the output of the market equilibration process is a matching, as μ_t is a matching at each Step t.

process, then we say that i_t *(successfully) proposed* to j_t in Step t, and i_t was *held* by j_t. If an agent k is held by j at Step t, and some agent $k' \neq k$ is held by j at Step $t' > t$, then we say that j *held* k' *over* k. Similarly, if j is unmatched under μ_t, and some agent $k' \neq \emptyset$ is held by j at Step $t' > t$, then we say that j *held* k' *over* $\emptyset$. Over the course of the market equilibration process, each agent on the opposite side of the market from i holds weakly more-preferred partners;[7] hence, if j holds k' over k, then we must have $k' \succ_j k$.

4 Results

Blum and Rothblum (2002) showed that if we start with an empty market and add agents sequentially using the market equilibration process, then we obtain a stable outcome. Moreover, the stable outcome obtained in this way favors the "last agent standing," in that the last agent to enter obtains his or her most-preferred achievable partner.

Theorem 1 (Blum and Rothblum (2002))**.** *Let μ be any stable matching for $I \setminus \{i\}$, and let ν be the market equilibration outcome after entry by i at μ. Then ν is stable and $\nu(i)$ is i's most-preferred achievable partner.*[8]

Blum and Rothblum (2002) proved Theorem 1 by combining powerful results on the structure of stable outcomes: the lattice structure theorem of Gale and Sotomayor (1985b) (see also Knuth (1976)) and the Lone Wolf Theorem of McVitie and Wilson (1970). By contrast, we prove Theorem 1 via a new, elementary argument that uses neither lattice structure nor the Lone Wolf Theorem.

Having shown that Theorem 1 does not itself depend on the Lone Wolf Theorem, we next show that Theorem 1 leads to a very short proof of the Lone Wolf result.

Theorem 2 (Lone Wolf Theorem – McVitie and Wilson (1970))**.** *If there exists a stable*

[7] This fact implies that the market equilibration process actually terminates: When i_t successfully proposes to an agent j_t, j_t becomes *strictly* better off than under μ_{t-1}; this can only happen finitely many times, as j_t's partner never becomes less preferred, and there are only finitely many agents in the market.

[8] We prove this result in the Appendix.

matching for I under which agent $i \in I$ is unmatched, then i is unmatched under every stable matching for I.

Proof. Suppose that ν is a stable matching (for I) under which $\nu(i) = \emptyset$. We define μ by $\mu(j) = \nu(j)$ for all $j \in J \equiv I \smallsetminus \{i\}$; we observe that μ is a matching for J, as no agent $j \in J$ was matched to i under ν. As ν is stable, μ must be, as well:

- The individual rationality of μ follows immediately from the individual rationality of ν.
- If there were a pair $(m, w) \in (M \cap J) \times (W \cap J)$ blocking μ, then (m, w) would also block ν, as we would have $w \succ_m \mu(m) = \nu(m)$ and $m \succ_w \mu(w) = \nu(w)$.

Now we let ν' be the market equilibration outcome after entry by i at μ.

Claim. *Agent i is unmatched under ν', that is, $\nu'(i) = \emptyset$.*

Proof. We suppose the contrary, i.e., that $\nu'(i) \neq \emptyset$, seeking a contradiction. If $\nu'(i) \neq \emptyset$, then in particular, i proposed to $\nu'(i)$ at some step of the market equilibration process starting at μ, and $\nu'(i)$ held i over $\mu(\nu'(i))$.[9] But the preceding observations mean that $\nu'(i) \succ_i \emptyset$ and $i \succ_{\nu'(i)} \mu(\nu'(i)) = \nu(\nu'(i))$; this contradicts the hypothesis that ν is unblocked. □

By Theorem 1, we know that $\nu'(i) = \emptyset$ is i's most-preferred achievable partner. Thus, we see that i is unmatched under every stable matching. □

Appendix – Proof of Theorem 1

Without loss of generality, we assume that $i \in M$. We let μ be any stable matching for $I \smallsetminus \{i\}$, and let ν be the market equilibration outcome after entry by i at μ. By construction, ν is a matching for I.

First Part. *The matching ν is stable.*

[9] Recall that we say that j holds k' over k if there is some step t in the market equilibration process at which k is matched to j and some later step t' at which k' is matchted to j.

Proof. Individual rationality of ν is immediate, as μ is individually rational, and agents only propose to and hold partners they prefer to $\emptyset$. To see that ν is unblocked, we suppose that there is a pair $(m, w) \in (M \cap (J \cup \{i\})) \times (W \cap (J \cup \{i\})) = M \times W$ such that $w \succ_m \nu(m)$ and $m \succ_w \nu(w)$.

As $i \in M$, we know that each woman holds weakly improving partners throughout the market equilibration process after entry by i at μ; hence in particular, we have $\nu(w) \succcurlyeq_w \mu(w)$. Thus, given our hypothesis that $m \succ_w \nu(w)$, we have

$$m \succ_w \nu(w) \succcurlyeq_w \mu(w). \tag{1}$$

If m was never active during the market equilibration process,[10] then we would have $\nu(m) = \mu(m)$. In that case, combining (1) and the hypothesis that $w \succ_m \nu(m) = \mu(m)$ would imply that (m, w) blocks μ—a contradiction. Thus, there must be some step of the market equilibration process at which m was active; we let t be the last such step. As $\nu(m) \neq w$, we know that at Step t of the market equilibration process, either

1. w was not in K, as she was already assigned to a man she prefers to m (i.e., $\mu_t(w) = \mu_{t-1}(w) \succ_w m$), or

2. w was not optimal for m among women in K, so that m (successfully) proposed to some woman $\mu_t(w)$ he prefers to w, i.e., $\mu_t(m) \succ_n w$.

In the former case, we have an immediate contradiction, as $\nu(w) \succcurlyeq_w \mu_t(w)$, so we would have $\nu(w) \succcurlyeq_w \mu_t(w) = \mu_{t-1}(w) \succ_w m$, contradicting the hypothesis that $m \succ_w \nu(w)$. In the latter case, as t is the last step in which m is active, we must have $\nu(m) = \mu_t(m)$—but again, this contradicts our hypothesis, as $\nu(m) = \mu_t(m) \succ_m w$ and we have assumed that $w \succ_m \nu(m)$. □

Second Part. *The agent $\nu(i)$ is i's most-preferred achievable partner.*

[10] We say that a man k is *active* during the market equilibration process if there is some t such that $k = i_t$ in the market equilibration process.

Proof. We suppose the contrary, i.e., that there is some woman $w^1 \succ_i \nu(i)$ who is achievable for i. Now, we let ν' be a stable matching (for I) in which $\nu'(i) = w^1$.

Now, we must have

$$\nu(w^1) \succ_{w^1} \nu'(w^1) = i, \tag{2}$$

or else (given that $w^1 \succ_i \nu(i)$), (i, w^1) would block ν (which we have already shown is stable). We let $m^1 = \nu(w^1)$. By construction, we must have

$$\nu'(m^1) \succ_{m^1} \nu(m^1) = w^1, \tag{3}$$

as otherwise (given (2)), (m^1, w^1) would block ν'. We let $w^2 = \nu'(m^1)$. Again, we must have

$$\nu(w^2) \succ_{w^2} \nu'(w^2) = m^1,$$

or else (given (3)), (m^1, w^2) would block ν. Iterating this logic, we obtain a sequence of women and men $m^0, w^1, m^1, w^2, m^2, \ldots$ such that $m^0 = i$ and for each $r \geq 1$, we have

$$m^r = \nu(w^r) \succ_{w^r} \nu'(w^r) = m^{r-1} \tag{4}$$

$$w^{r+1} = \nu'(m^r) \succ_{m^r} \nu(m^r) = w^r. \tag{5}$$

Claim. *For all $r \geq 1$, m^r must be matched under μ.*

Proof. If $\mu(m^r) = \nu(m^r)$, then the claim is immediate. Otherwise, if $\mu(m^r) \neq \nu(m^r)$, we know that m^r successfully proposed to w^r at some step t of the market equilibration process after entry by i. But then, m^r was active at Step t, so there must have been an earlier step $t' < t$ at which some woman held some other man over m^r; we let t'' be the earliest such step. Then we must have

$$\emptyset \neq \mu_{t''-1}(m^r) = \cdots = \mu_0(m^r) = \mu(m^r),$$

so, in particular, we have the claim. □

But now, as there are finitely many men and women, the sequence must cross itself. That is, there must be some r such that $m^r = m^{r'}$ for some $r' \leq r$; we let s be the minimal such r.

Claim. *We have* $m^s = m^0$.

Proof. If $m^s = m^r$ for some r with $s > r > 0$, then we have

$$\nu'(\nu(m^s)) = m^{s-1}, \tag{6}$$

$$\nu'(\nu(m^r)) = m^{r-1} \tag{7}$$

by construction. But then (6) and (7) together contradict the minimality assumption in our choice of s, as the hypothesis that $m^s = m^r$ implies that $\nu'(\nu(m^s)) = \nu'(\nu(m^r))$. □

The two preceding claims together imply that $m^s = m^0 = i$ is matched under μ, which is impossible, as μ is a matching for $I \setminus \{i\}$. □

References

Alkan, A. (2002). A class of multipartner matching markets with a strong lattice structure. *Economic Theory 19*(4), 737–746.

Balinski, M. and G. Ratier (1998). Graphs and marriages. *American Mathematical Monthly 105*(5), 430–445.

Blum, Y., A. E. Roth, and U. G. Rothblum (1997). Vacancy chains and equilibration in senior-level labor markets. *Journal of Economic Theory 76*(2), 362–411.

Blum, Y. and U. G. Rothblum (2002). "Timing is everything" and marital bliss. *Journal of Economic Theory 103*(2), 429–443.

Ciupan, A. (2016). Preference cycles in stable matchings. In preparation.

Dworczak, P. (2016). Deferred acceptance with compensation chains. Working paper.

Gale, D. and L. S. Shapley (1962). College admissions and the stability of marriage. *American Mathematical Monthly 69*, 9–15.

Gale, D. and M. Sotomayor (1985a). Ms. Machiavelli and the stable matching problem. *American Mathematical Monthly 92*(4), 261–268.

Gale, D. and M. A. O. Sotomayor (1985b). Some remarks on the stable matching problem. *Discrete Applied Mathematics 11*(3), 223–232.

Hatfield, J. W. and S. D. Kominers (2012). Matching in Networks with Bilateral Contracts. *American Economic Journal: Microeconomics 4*(1), 176–208.

Hatfield, J. W. and P. Milgrom (2005). Matching with contracts. *American Economic Review 95*, 913–935.

Jagadeesan, R. (2016). Lattice structure and the rural hospitals theorem in matching with complementarities. In preparation.

Jagadeesan, R., S. D. Kominers, and R. Rheingans-Yoo (2016a). Lone wolves in the core. In preparation.

Jagadeesan, R., S. D. Kominers, and R. Rheingans-Yoo (2016b). Strategy-proofness of worker-optimal matching with continuously transferable utility. Working paper.

Kamada, Y. and F. Kojima (2011). Stability and strategy-proofness for matching with constraints: A problem in the Japanese medical matching and its solution. *American Economic Review Papers & Proceedings.*

Kamada, Y. and F. Kojima (2012). Stability and strategy-proofness for matching with constraints: A problem in the Japanese medical match and its solution. *American Economic Review 102*(3), 366–370.

Klaus, B. and F. Klijn (2010). Smith and Rawls share a room: Stability and medians. *Social Choice and Welfare 35*(4), 647–667.

Klijn, F. and A. Yazıcı (2014). A many-to-many 'rural hospital theorem'. *Journal of Mathematical Economics 54*, 63–73.

Knuth, D. E. (1976). *Marriages Stables.* Les Presses de L'Université de Montréal.

Kojima, F. (2012). The "rural hospital theorem" revisited. *International Journal of Economic Theory 8*(1), 67–76.

McVitie, D. G. and L. B. Wilson (1970). Stable marriage assignment for unequal sets. *BIT Numerical Mathematics 10*, 295–309.

Roth, A. E. (1984). The evolution of the labor market for medical interns and residents: A case study in game theory. *Journal of Political Economy 92*, 991–1016.

Roth, A. E. (1986). On the allocation of residents to rural hospitals: A general property of two-sided matching markets. *Econometrica 54*, 425–427.

Roth, A. E. (2002). The economist as engineer: Game theory, experimental economics and computation as tools of design economics. *Econometrica 70*, 1341–1378.

Roth, A. E. (2008). Deferred acceptance algorithms: History, theory, practice, and open questions. *International Journal of Game Theory 36*, 537–569.

Roth, A. E. and M. A. O. Sotomayor (1990). *Two-sided matching: A study in game-theoretic modeling and analysis.* Cambridge University Press.