The Roommate Problem with Assigned Roles

Dev Kinger
Indian Statistical Institute, Delhi

January 4, 2026

Abstract

This paper introduces the Roommate Problem with Assigned Roles, extending classical matching theory to environments in which successful pair formation requires the endogenous assignment of asymmetric functional roles, specifically a Leader and a Follower. Our analysis focuses on Role Dominance, a preference structure under which agents strictly prioritize securing their preferred functional role over matching with their preferred partner. We show that introducing endogenous role assignment fundamentally reshapes the structural conditions governing stability relative to the classical Stable Roommates Problem. While the flexibility to assign roles provides a mechanism to mitigate standard cyclic instabilities, it does not eliminate friction altogether. Rather, new impediments emerge from the scarcity of preferred roles and the interaction between role imbalance and residual preference cycles. To address these structural challenges, we develop a constructive algorithmic framework designed to identify a stable matching whenever one exists. We outline a procedure that sequentially satisfies mutual partner affinities subject to strict role constraints, offering a systematic solution for achieving stability in hierarchical matching markets.

1 Introduction

Matching theory has traditionally been organized around two distinct structural paradigms: the Stable Marriage Problem (SMP) and the Stable Roommates Problem (SRP). In the classical Two-Sided model (SMP), the population is exogenously partitioned into two disjoint sets—such as men and women or workers and firms—where matching can only occur across the divide. Conversely, in the One-Sided model (SRP), agents form a single pool where any individual can match with any other.

In both frameworks, the central equilibrium concept is *stability*. A matching is deemed stable if it admits no *blocking pair*—a pair of agents who are not currently matched with each other but would mutually prefer to be. In standard theory, this deviation is purely a function of partner identity: two agents defect from their current partners simply because they prefer each other. However, these frameworks leave a significant theoretical gap. They do not account for environments where agents are drawn from a single pool of peers yet must assume distinct, asymmetric functional roles within the match.

This paper introduces the Roommate Problem with Assigned Roles, a theoretical framework that bridges the structural gap between one-sided and two-sided matching models. In this setting, agents originate from a single pool and form pairs endogenously. However, unlike the standard Stable Roommates Problem where the match is symmetric, our model imposes an internal hierarchy: every formed pair must assign distinct functional roles to its participants. Throughout this paper, we denote these roles as a *Leader* and a *Follower*. While these labels are used for conceptual clarity, they generalize to any binary, asymmetric division of labor. The fundamental feasibility constraint of our model is strict: every valid match must consist of exactly one Leader and one Follower.

Within this economy, agents possess preferences over two separable dimensions: the identity of their partner and the specific role they occupy within the pair. Consequently, an agent does not simply evaluate a partner in isolation; rather, they evaluate a "bundle" of attributes comprising both the partner's identity and the assigned role. To illustrate how this multidimensionality mimics real-world constraints, consider two distinct professional settings: a surgical team and an aviation crew.

In the operating room, a procedure strictly requires one *Lead Surgeon* and one *Assisting Surgeon*. Two senior doctors may possess high mutual professional regard and desire to collaborate (partner preference). However, if both strictly prefer to Lead, they cannot form a viable partnership despite their mutual affinity, as neither is willing to accept the subordinate role. Similarly, in the cockpit, flight safety protocols demand a clear hierarchy between the *Pilot in Command* and the *Co-pilot*. Two aviators may be perfectly compatible in terms of skill and temperament, yet a stable crew cannot form if they are unable to agree on who commands the aircraft. In both cases, the compatibility of the pair is insufficient on its own. Thus, stability in our framework requires a double coincidence of wants: agents must simultaneously agree on *who* they match with and *what* role each performs.

This introduction of endogenous role assignment fundamentally alters the geometry of stability. A deviation is no longer just about swapping partners; it is about finding a better partner-role combination. A blocking pair may arise from two agents currently assigned the same role (e.g., two "Leaders") who decide to match, provided one is willing to switch to the Follower role. Alternatively, an existing pair might remain together but block the matching by swapping roles internally.

We analyze this problem under the assumption of separable preferences. This creates a bifurcation in the theoretical framework based on which dimension an agent prioritizes. If agents are *Partner Dominant*—caring more about who they are with than what they do—the problem structurally collapses back into the classical Stable Roommates Problem. To capture the distinct theoretical challenges of hierarchical matching, we focus our analysis specifically on the regime of Role Dominance.

Under Role Dominance, an agent's preference for a specific hierarchical position (Leader or Follower) lexicographically dominates their preference for a partner. This constraint effectively erects an endogenous "Wall" through the population. The wall partitions agents into two distinct groups: those who successfully secure their preferred role, and those who are forced to accept their non-preferred role to make a match feasible.

If there is a surplus of agents desiring the Leader role, some must become Followers; if there is a surplus of agents desiring the Follower role, some must become Leaders. In either case, agents on the dissatisfied side of the wall constantly seek to "cross over." They actively search for partners who would allow them to assume their preferred role, creating a persistent tension at the boundary. Stability, therefore, depends on whether these dissatisfied agents can find a willing partner to facilitate their jump to the preferred side of the wall.

Motivated by these structural insights, the primary objective of this paper is to provide a constructive solution to this problem. We aim to move beyond theoretical existence results and develop a concrete algorithmic framework. We propose an algorithm designed to navigate the trade-off between partner affinity and role availability, capable of producing a stable matching whenever one exists under the strict constraints of role dominance.

2 The Model

In this section, we formalize the framework of the Roommate Problem with Assigned Roles. We define the market structure, the nature of agent preferences, and the concept of stability. Our goal is to translate the intuitive constraints introduced in the previous section—specifically the necessity of role assignment and the hierarchy of preferences—into a rigorous theoretical setting.

2.1 Preliminaries

We consider a market consisting of a finite set of agents $N = \{1, 2, \dots, 2n\}$, where $n \geq 1$. The assumption of an even population size is made to facilitate the possibility of a perfect matching, where every agent is successfully paired.

In the classical Stable Roommates Problem, an outcome for an agent is simply a partner $j \in N$. In our framework, an outcome is a tuple consisting of a partner and a functional role. Let $\mathcal{R} = \{L, F\}$ be the set of available roles, where L denotes the *Leader* and F denotes the *Follower*.

2.2 Preferences

Each agent $i \in N$ holds strict preferences $\succ_i$ over the set of possible outcomes, denoted by $\Omega_i = \mathcal{R} \times (N \setminus \{i\})$. A generic element $(r, j) \in \Omega_i$ represents the bundle where agent i holds role r while matched with partner j.

To maintain tractability while capturing the essential trade-offs of the environment, we first restrict the structure of these preferences.

Assumption 1 (Separability)**.** *Agents possess separable preferences. Formally, for all $i \in N$, the preference relation $\succ_i$ satisfies the following consistency conditions:*

$$\forall j, j' \in N \setminus \{i\}, \forall r, r' \in \mathcal{R} : \quad (r, j) \succ_i (r, j') \implies (r', j) \succ_i (r', j') \qquad (i)$$

$$\forall r, r' \in \mathcal{R}, \forall j, j' \in N \setminus \{i\} : \quad (r, j) \succ_i (r', j) \implies (r, j') \succ_i (r', j') \qquad (ii)$$

Assumption 1 imposes independence between partner identity and role in agents' evaluations of matched outcomes. Condition (i) requires that an agent's ranking of potential partners be invariant to the role they occupy: if agent i prefers partner j to j' under some role r, the same ordering must hold under an alternative role r'. Similarly, condition (ii) imposes the symmetric restriction across partners, requiring that an agent's ranking of roles be invariant to the identity of the partner. Separability therefore allows each preference relation $\succ_i$ to be represented by two strict marginal orderings: a partner preference $\succ_i^P$ defined over $N \setminus \{i\}$, and a role preference $\succ_i^R$ defined over the set of roles $\mathcal{R}$. While agents ultimately rank complete partner–role bundles, these marginal orderings are well-defined and consistent across dimensions.

Although, separability allows us to rank partners and roles independently, it does not specify the relative weight an agent assigns to these dimensions. To capture the hierarchical nature of the market, we impose a second restriction.

Assumption 2 (Role Dominance)**.** *Agents lexicographically prioritize their role assignment over their partner's identity. Formally, for any agent $i \in N$, roles $r, r' \in \mathcal{R}$, and partners $j, k \in N \setminus \{i\}$:*

$$r \succ_i^R r' \implies (r, j) \succ_i (r', k)$$

Note that we explicitly rule out the alternative regime of *Partner Dominance*. If agents prioritized partners over roles, the problem would structurally collapse back into the standard Stable Roommates Problem, yielding no new insights. By assuming Role Dominance, we ensure the population effectively bifurcates into those seeking the Leader role and those seeking the Follower role, generating the structural "Wall" described in the introduction.

2.3 Matching and Stability

In our framework, a feasible outcome must specify both a partner for each agent and a functional role within the pair. For any role $r \in \mathcal{R}$, let $\bar{r}$ denote the unique complementary role (i.e., if $r = L$, then $\bar{r} = F$).

Definition 1 (Matching)**.** A matching is a function $\mu : N \to N \times \mathcal{R}$ satisfying the following consistency condition for all agents $i, j \in N$ and roles $r \in \mathcal{R}$:

$$\mu(i) = (j, r) \iff j \neq i \text{ and } \mu(j) = (i, \bar{r}).$$

This definition ensures that any feasible outcome consists of disjoint pairs of agents, with each pair assigned exactly one Leader and one Follower. Hence, a matching jointly determines both the partner structure and the role assignment.

We restrict attention to individually rational matchings, where every agent weakly prefers their assigned outcome to remaining unmatched. Given this, instability can arise only through profitable pairwise deviations.

Definition 2 (Blocking Pair)**.** A pair of agents $\{i, j\}$ **blocks** the matching μ if there exists a feasible role assignment $r \in \mathcal{R}$ such that both agents strictly prefer the new partnership to their current assignments:

$$(j, r) \succ_i \mu(i) \quad \text{and} \quad (i, \bar{r}) \succ_j \mu(j).$$

A blocking pair represents a feasible alternative partnership in which both agents strictly prefer the proposed match and role assignment to their current assignments under μ. Importantly, the deviation requires agreement on both the identity of the partner and the allocation of complementary roles.

Definition 3 (Stability)**.** A matching μ is said to be **stable** if it admits no blocking pairs.

3 Stability Analysis: Motivating Examples

In this section, we use specific examples to illustrate the mechanics of our model. We show that the existence of a stable matching depends on the structure of partner preferences—specifically, on the presence and length of preference cycles.

We present two cases. The first demonstrates how role assignment can resolve the instability issues common in the classical Stable Roommates Problem. The second demonstrates that when preference cycles conflict, the scarcity of preferred roles makes a stable matching impossible.

3.1 Example 1: Existence of Stable Outcomes

Consider a market where all agents possess *Role Dominant* preferences and strictly prefer to be Leaders (L). We examine two distinct configurations of partner preferences.

Case A: The Single Odd Cycle Consider a market with four agents $N = \{1, 2, 3, 4\}$. The preferences over partners are given by:

Agent 1	Agent 2	Agent 3	Agent 4
2	3	1	1
3	1	2	3
4	4	4	2

Here, the preferences of agents $\{1, 2, 3\}$ form a classic odd cycle ($1 \to 2 \to 3 \to 1$), while agent 4 is outside the cycle but desires agent 1. In the standard Stable Roommates Problem (without roles), this configuration has no stable matching. However, under Role Dominance, the following outcome is stable:

$$\mu = \{(1_L, 4_F), (3_L, 2_F)\}$$

Stability relies on two factors. First, the agents assigned the Follower role (2 and 4) are "compensated" with their highest-ranked partners: Agent 2 is matched with 3 (her first choice), and Agent 4 is matched with 1 (his first choice). Consequently, no Follower can improve their utility by matching with a different Leader, as they are already paired with their ideal partners. Second, a coordinated deviation between the two dissatisfied Followers is impossible. While both 2 and 4 would prefer to be Leaders, any match between them would require one to accept the Follower role. Since both strictly prioritize the Leader role, they cannot agree on a viable division of labor, effectively preventing a *mutiny* of the misassigned agents.

Case B: The Even Cycle Consider a market with six agents where preferences form a perfect even cycle $1 \to 2 \to 3 \to 4 \to 1$, with agents 5 and 6 mutually preferring each other.

Agent 1	Agent 2	Agent 3	Agent 4	Agent 5	Agent 6
2	3	4	1	1	2
3	4	1	2	6	5
4	1	2	3	2	1
5	5	5	5	3	3
6	6	6	6	4	4

In this scenario, a stable matching exists and is given by:

$$\mu = \{(2_L, 1_F), (4_L, 3_F), (5_L, 6_F)\}$$

The even length of the cycle allows for a perfect alternation of roles. We can assign roles (L, F) sequentially along the cycle edges without generating a conflict at the closure of the loop. Every agent who accepts the Follower role is matched with their preferred partner in the cycle. Again, the matching is stable because no two followers can form a mutiny and get simultaneously better off and no follower would want to leave their assigned leader.

Based on these cases, we derive two key structural observations regarding stability under Role Dominance.

Remark 1. Stability in hierarchical markets requires that agents who are assigned their non-preferred role must be "compensated" with high-utility partners. Specifically, a misassigned agent will not block a matching if they are paired with their best attainable partner.

Remark 2. Even cycles are stabilizing because they support a perfect alternation of roles, ensuring that every agent accepting a non-preferred role is successfully matched with a complementary partner. In contrast, odd cycles create a structural conflict: the alternating pattern cannot be closed, leaving at least one agent unable to secure the partner type required for stability.

3.2 Example 2: Non-Existence of Stable Outcomes

We now consider a case where the preference structure is too complex to be resolved by role assignment. Consider a market with six agents composed of two disjoint odd cycles.

Preferences: Agents form two independent cycles: $1 \to 2 \to 3 \to 1$ and $4 \to 5 \to 6 \to 4$. All agents strictly prefer to be Leaders (L).

Agent 1	Agent 2	Agent 3	Agent 4	Agent 5	Agent 6
2	3	1	5	6	4
3	1	2	6	4	5
4	4	4	1	1	1
5	5	5	2	2	2
6	6	6	3	3	3

In this environment, **no stable matching exists**.

The presence of two disjoint odd cycles creates an unresolvable conflict. To stabilize the first cycle $\{1, 2, 3\}$, one agent must accept the Follower role. As established in Example 1, this agent requires a "compensating" partner from outside the cycle or a specific alignment within it. However, the second cycle $\{4, 5, 6\}$ faces the exact same constraint. With all agents preferring to be Leaders, there is no way to assign roles such that the "misassigned" agents in both cycles are simultaneously satisfied. Any proposed matching will inevitably leave at least one agent with both a non-preferred role and a sub-optimal partner, creating a valid blocking pair and allowing a follower-follower mutiny.

This leads us to the fundamental necessary condition for stability in our model.

Theorem 1 (**Existence of Stable Outcomes**). *A complete stable matching μ always exists unless both of the following conditions hold:*

1. *The population is partitioned into unequal sets of role demand:*

$$|N_L| \neq |N_F|$$

2. *(Tan 1991) The preference relation over partners contains strictly more than one disjoint odd cycle. That is, there exist at least two disjoint sets of agents $C_1, C_2 \subseteq N$ such that each forms a cycle:*

$$a_1 \succ_2 a_2 \succ_3 \cdots \succ_k a_1$$

where k is odd.

4 Role Scarcity and the Structure of Instability

We analyze the structural impediments to stability that arise when the population is partitioned into unequal sets of role demand. Let N_L and N_F denote the sets of agents who strictly prefer the Leader and Follower roles, respectively.

Without loss of generality, assume that the demand for the Leader role is at least as high as the demand for the Follower role:

$$|N_L| \geq |N_F|$$

Given a matching μ, we define the *residual set* $R(\mu)$ as the subset of role-preferring agents who are unable to secure their preferred role. Since $|N_L| \geq |N_F|$, this set consists of agents who prefer to be Leaders but are assigned the Follower role:

$$R(\mu) := \{i \in N_L \mid \mu(i) \in N \text{ and agent } i \text{ acts as Follower}\}$$

The following lemmas establish that any instability involving a role switch must originate from this residual set, and that this set is internally inherently unstable.

Lemma 1. *If a blocking pair (i, j) exists in which agent i strictly improves their utility by switching from the Follower role to the Leader role, then i must belong to the residual set $R(\mu)$.*

Proof. For agent i to strictly benefit from switching roles ($F \to L$), two conditions must hold: i must prefer the Leader role ($i \in N_L$) and currently be assigned the Follower role. By definition, the set of agents satisfying these exact conditions is the residual set $R(\mu)$. Therefore, $i \in R(\mu)$. □

Lemma 2. *For any matching μ where $|N_L| \geq |N_F|$, every agent in the residual set $R(\mu)$ strictly prefers the Leader role:*

$$\forall i \in R(\mu), \quad L \succ_i F$$

Proof. By construction, $R(\mu) \subseteq N_L$. The set consists precisely of those agents who strictly prefer the Leader role (L) but are currently assigned the Follower role (F). Consequently, all agents within $R(\mu)$ share the identical direction of desired improvement. This uniformity renders internal resolution impossible; any matching formed exclusively by agents within $R(\mu)$ necessitates that at least one agent accepts the non-preferred role. □

5 Conclusion

This paper introduces the *Roommate Problem with Assigned Roles*, extending the classical matching framework to environments where agents must endogenously determine both a partner and a functional division of labor. We demonstrate that while role flexibility can resolve cyclic instabilities inherent in the standard model—specifically within even cycles—it introduces a new structural friction driven by role scarcity. Our analysis identifies the *Residual Set* $\mathcal{R}(\mu)$ as the locus of this instability; because agents in this set share identical role preferences, they can resolve their dissatisfaction through internal trade making them the primary source of instability. Consequently, we conclude that stability in hierarchical markets is strictly conditional on the market's ability to "compensate" these residual agents with high-utility partners. Future research will build upon these structural conditions to develop algorithmic solutions for identifying stable outcomes in general markets.

References

Gale, D., & Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1), 9–15.

Gusfield, D., & Irving, R. W. (1989). *The Stable Marriage Problem: Structure and Algorithms*. MIT Press.

Irving, R. W. (1985). An efficient algorithm for the stable roommates problem. *Journal of Algorithms*, 6(4), 577–595.

Mishra, D. (n.d.). *Lecture Notes on Mechanism Design*. Indian Statistical Institute, Delhi. `https://www.isid.ac.in/~dmishra/gmdoc/mdnotes.pdf`

Tan, J. J. M. (1991). A necessary and sufficient condition for the existence of a complete stable matching. *Journal of Algorithms*, 12(1), 154–178.