

Deep Learning for Two-Sided Matching*

Sai Srivatsa Ravindranath^a, Zhe Feng^a, Shira Li^b, Jonathan Ma^b, Scott D. Kominers^c, and David C. Parkes^a

^aJohn A. Paulson School of Engineering and Applied Sciences, Harvard University

saisr,zhe_feng,parkes@g.harvard.edu

^bHarvard College

shirali@alumni.harvard.edu, jonathan.q.ma@gmail.com

^cHarvard Business School

kominers@fas.harvard.edu

April 27, 2022

Abstract

We initiate the use of a multi-layer neural network to model two-sided matching and to explore the design space between strategy-proofness and stability. It is well known that both properties cannot be achieved simultaneously but the efficient frontier in this design space is not understood. We show empirically that it is possible to achieve a good compromise between stability and strategy-proofness—substantially better than that achievable through a convex combination of deferred acceptance (stable and strategy-proof for only one side of the market) and randomized serial dictatorship (strategy-proof but not stable).

1 Introduction

Two-sided matching markets, such as Uber, Airbnb, stock markets, and dating apps, play a significant role in today’s world. As a result, there is a tremendous and rising interest to design better mechanisms for two-sided matching. The seminal work of Gale and Shapley [14] introduced a simple mechanism for stable matching in two-sided markets—*Deferred-acceptance* (DA)—which has since has been applied in doctor-hospital matching [25], school choice [3, 22, 2], and the matching of cadets to their branches of military service [30, 29]. DA is *stable*, i.e., no pair of agents mutually prefer each other to their DA partners. On the other hand, DA is not *strategy-proof* (SP); that is, under fully general preferences, it is always possible that some agent can mis-report her preferences to obtain a better matching than she would receive under the DA mechanism.¹ Another well-known mechanism, *random serial dictatorship* (RSD), is SP but not stable.² More generally, it is well-known that it is impossible to achieve both stability and strategy-proofness in two-sided matching [9, 24]. At the same time, there is little understanding of the nature of this tradeoff beyond point solutions such as

*This work is supported in part through an AWS Machine Learning Research Award.

¹As we discuss below, DA *is* strategy-proof for agents on one side of the market, but not for agents on both sides of the market simultaneously.

²Indeed, RSD is typically studied for one-sided assignment problems rather than for two-sided matching mechanisms. For two-sided matching, not only is RSD unstable, but it can also fail to match participants in a way that is better than receiving no match at all. We adopt RSD as a benchmark in this paper—despite its flaws—because we are not aware of more suitable SP mechanisms for two-sided matching.

DA and RSD—and we are not aware of any work that has attempted to map out the tradeoff more generally.

Inspired by the recent development of deep learning for optimal auction design [10, 8, 28, 23], we initiate the study of multi-player neural networks to model two-sided matching. We show how to use machine learning to characterize the frontier curve of the tradeoff between stability and strategy-proofness. The main challenges of applying neural networks to two-sided matching come from handling the ordinal preference inputs, and in identifying suitable differentiable surrogates for approximate strategy-proofness and stability.

For randomized matching mechanisms, the strongest SP concept is that of *ordinal strategy-proofness*. This aligns incentives with truthful reporting whatever the utility function of an agent. Ordinal SP is equivalent to a property of *first-order stochastic dominance* (FOSD) [11], which means that agents have a weakly higher chance of getting their top-ranked choices when they report their preferences truthfully. Our metric for SP quantifies the degree to which FOSD is violated. For this, we adopt an adversarial approach, seeking to augment the training data with suitable defeating mis-reports. We also provide a metric to quantify the degree to which stability is violated. We show that a loss function built from these quantities can be trained through SGD, and illustrate the use of the framework to identify the stability-strategyproofness frontier. We run simulations to validate the efficiency of our approach, demonstrating for different preference distributions that our approach can strike much better trade-offs between stability and strategyproofness than the convex combination of DA and RSD.

Related work. This work lies at the intersection of *two-sided matching* [27] and the role of *machine learning within economics* [6]. The matching mechanisms learned by our neural networks are randomized, approximately strategy-proof, and approximately stable. Budish et al. [7] and Mennle and Seuken [19, 20] discuss different notions of approximate strategy-proofness in the context of matching and allocation. In this work, we focus on ordinal SP and its analog of FOSD [11]. This is a strong and widely-used SP concept in the presence of ordinal preferences.

Classic results of Dubins and Freedman [9] and Roth [24] show that it is impossible to achieve both stability and strategy-proofness in two-sided matching simultaneously—although strategy-proofness *for one side of the market* is achieved by the Deferred Acceptance (DA) mechanism.³ Thus market designers have looked at mechanisms that relax one or both of these conditions. The *Random serial dictatorship* (RSD) mechanism [1] is SP but typically fails to produce stable outcomes (and indeed, it may even fail to be individually rational).⁴ On the other hand, Roth et al. [26] study the polytope of stable matchings and provide a wide class of stable matchings beyond the well-known DA outcome. The *stable improvement cycles* mechanism of Erdil and Ergin [12], meanwhile, achieves as much efficiency as possible on top of stability, but fails to be SP even for one side of the market. Finally, a series of results have shown that DA becomes SP for both sides of the market in certain large-market limit contexts; these results typically also require additional, structural assumptions on market participants’ preferences (see, e.g., [16, 17, 18]).

This work belongs to the emerging literature on *machine learning for economic design*. Narasimhan et al. [21] utilize support vector machines to search for good mechanisms among the weighted polytope mechanisms. Recently, many papers [10, 13, 15, 8, 28, 23] apply deep neural networks to optimal auction design and facility location problems. In this work, and inspired by the neural network

³Alcalde and Barberà [4] also showed the impossibility of individually rational, Pareto efficient, and SP allocation rules. Alva and Manjunath [5] extended this result to randomized matching contexts.

⁴The Top Trading Cycles (TTC) mechanism is likewise SP—but it effectively treats the market as one-sided, producing outcomes that do not reflect the other side’s preferences.

architecture proposed by Dütting et al. [10], we use neural networks to model the design of two-sided matching markets.

2 Preliminaries

Let W be a set of n workers and F a set of m firms, and suppose that each worker can be matched to at most one firm and each firm to at most one worker. A *matching* μ is a set of (worker, firm) pairs, with each worker and firm participating in at most one match. Let $\mathcal{B}$ denote the set of all *matchings*. If a worker or firm remains unmatched, we say that it is *matched to* $\perp$. If $(w, f) \in \mu$, then μ matches w to f , and we write $\mu(w) = f$ and $\mu(f) = w$. We write $(w, \perp) \in \mu$ (resp. $(\perp, f) \in \mu$) to denote that w (resp. f) is unmatched.

Each worker has a *strict preference order* $\succ_w$ over the set $\bar{F} = F \cup \{\perp\}$. Each firm has a strict preference order $\succ_f$ over the set $\bar{W} = W \cup \{\perp\}$. Worker w (firm f) prefers remaining unmatched to being matched with a firm (worker) that is ranked below $\perp$ (the agents ranked below $\perp$ are unacceptable). If worker w prefers firm f to f' then we represent this as $f \succ_w f'$, and similarly for the preferences of a firm. Let P denote the set of all preference profiles, with $\succ = (\succ_1, \dots, \succ_n, \succ_{n+1}, \dots, \succ_{n+m}) \in P$ denoting the preference profile that comprises all workers and firms.

A pair (w, f) forms a *blocking pair for matching* μ if w and f prefer each other to their partners in μ (or $\perp$ in the case that either or both are unmatched). A matching μ is *stable* if and only if there are no blocking pairs. A matching μ is *individually rational* (IR) if it is not blocked by any individual, i.e., there is no worker or firm that finds its partner unacceptable and prefers $\perp$.⁵

2.1 Randomized matchings

We work with *randomized matching mechanisms* g that map preference profiles $\succ$ to distributions on matchings, denoted $g(\succ) \in \Delta(\mathcal{B})$. This provides for differentiable mechanisms. Here, $\Delta(\mathcal{B})$ denotes the probability simplex on the set of matchings.

We write $r \in [0, 1]^{(n+1) \times (m+1)}$ to define the *marginal probability* $r_{wf} \geq 0$ with which worker w is matched with firm f , for each $w \in \bar{W}$ and each firm $f \in \bar{F}$. We require $\sum_{f' \in \bar{F}} r_{wf'} = 1$ for all $w \in \bar{W}$, and $\sum_{w' \in \bar{W}} r_{w'f} = 1$ for all $f \in \bar{F}$. For notational simplicity, we also write $g_{wf}(\succ)$ to denote the marginal probability of matching worker w (or $\perp$) and firm f (or $\perp$).

Theorem 1 (Birkhoff von-Neumann). *Given any randomized matching r , there exists a distribution on matchings, $\Delta(\mathcal{B})$, with marginal probabilities equal to r .*

The following definition is standard [7] and generalizes stability to randomized matchings.

Definition 2 (Ex ante justified envy). A randomized matching r causes *ex ante justified envy* if

- (1) some worker w prefers f over some (fractionally) matched firm f' (including $f' = \perp$) and firm f prefers w over some (fractionally) matched worker w' (including $w' = \perp$) (“ w has envy towards w' ” and “ f has envy towards f' ”), or
- (2) some worker w finds a (fractionally) matched $f' \in F$ unacceptable, i.e. $r_{wf'} > 0$ and $\perp \succ_w f'$, or some firm f finds a (fractionally) matched $w' \in W$ unacceptable, i.e. $r_{w'f} > 0$ and $\perp \succ_f w'$.

A randomized matching r is *ex ante stable* if and only if it does not cause any *ex ante justified envy*. Ex ante stability reduces to the standard concept of stability for a deterministic matching. Part (2) of the definition of ex ante justified envy captures the idea that a randomized matching

⁵Stability precludes empty matchings. For example, if a matching μ leaves a worker w and a firm f unmatched, where w finds f acceptable and f finds w acceptable, then (w, f) is a blocking pair to μ .

r should satisfy *individually rationality* (IR). That is, for any worker w , we should have $r_{wf'} = 0$ for all $f' \in F$ for which $\perp \succ_w f'$, and for any firm f , we should have $r_{w'f}$ for all $w' \in W$ for which $\perp \succ_f w'$.

To define strategy-proofness, say that $u_w : \bar{F} \rightarrow \mathbb{R}$ is a $\succ_w$ -utility for worker w when $u_w(f) > u_w(f')$ if and only if $f \succ_w f'$, for all $f, f' \in \bar{F}$. We similarly define a $\succ_f$ -utility for firm f . The following concept of ordinal SP is standard [11] and generalizes SP to randomized matchings.

Definition 3 (Ordinal strategy-proofness). A randomized matching mechanism g satisfies ordinal SP if and only if, for all agents $i \in W \cup F$, for any preference profile $\succ$, and any $\succ_i$ -utility for agent i , then for all reports $\succ'_i$, we have

$$\mathbf{E}_{\mu \sim g(\succ_i, \succ_{-i})}[u_i(\mu(i))] \geq \mathbf{E}_{\mu \sim g(\succ'_i, \succ_{-i})}[u_i(\mu(i))]. \quad (1)$$

By this definition, no worker or firm can improve its expected utility by misreporting their preference order, and for any utility function consistent with their preference order. For a deterministic mechanism, ordinal SP reduces to the requirement that no agent can improve its preference ranking for any misreport.

Erdil [11] characterizes the following *first-order stochastic dominance condition* as equivalent to ordinal SP (Definition 3); this condition provides a useful relaxation of SP for our learning framework.

Definition 4 (First Order Stochastic Dominance). A randomized matching mechanism g has *first order stochastic dominance* (FOSD) if and only if, the distribution on matchings for the true report *first order stochastically dominates* the distribution for a mis-report, e.g., for worker w , and for each $f' \in \bar{F}$ such that $f' \succ_w \perp$, and all reports of others $\succ_{-w}$, we have (and similarly for the roles of workers and firms transposed)

$$\sum_{f \in F: f \succ_w f'} g_{wf}(\succ_w, \succ_{-w}) \geq \sum_{f \in F: f \succ_w f'} g_{wf}(\succ'_w, \succ_{-w}). \quad (2)$$

Whether looking at its most preferred firm, its two most preferred firms, or so forth, worker w achieves a higher probability of matching on that firm or set of firms for its true report than for any mis-report, and whatever the reports of others.

Theorem 5 ([11]). *A two-sided matching mechanism is ordinal SP if and only if it satisfies FOSD.*

2.2 Deferred Acceptance and RSD

In this section, we consider two benchmark mechanisms: the stable but not SP *deferred-acceptance* (DA) mechanism—which is a deterministic mechanism—and the SP but not stable *randomized serial dictatorship* (RSD) mechanism. DA mechanisms are stable and ordinal SP for the proposing side of the market, but they are not ordinal SP for both sides.

Definition 6 (Deferred-acceptance (DA)). In worker-proposing deferred-acceptance (firm-proposing is defined analogously), each worker w maintains a list of acceptable firms ($f \succ_w \perp$) for which it has not had a proposal rejected (“remaining firms”). Repeat the following process until all proposals are accepted:

- $\forall w \in W$: w proposes to its best acceptable, remaining firm.
- $\forall f \in F$: f tentatively accepts its best proposal (if any), and rejects the rest.
- $\forall w \in W$: If w is rejected by firm f , it updates its list of acceptable firms to remove f .

Theorem 7 (see [27]). *DA is stable, but not FOSD strategy-proof.*

Definition 8 (Randomized serial dictatorship (RSD)). Sample a *priority order* π on the set $W \cup F$ uniformly at random, such that $\pi_1, \pi_2, \dots, \pi_{m+n}$ is a permutation on $W \cup F$ in decreasing order of priority. Proceed as follows:

- Initialize matching μ to the empty matching.
- In round $k = 1, \dots, m + n$:
 - If $\pi_k \in W \cup F$ is not yet matched in μ , then add to matching μ the match between π_k and its most preferred, unmatched agent, or $\perp$ if all remaining possibilities are unacceptable to π_k .

Theorem 9. *RSD satisfies FOSD—and thus is ordinal SP by Theorem 5—but is not stable.*

Proof. For ordinal SP, consider agent i in some position k in the order. The agent’s reported preference order has no effect on the choices of preceding agents, whether workers or firms (and even in regard to whether agent i itself is selected by an agent on the other side of the market). Reporting its true preference order ensures that it is matched with its most preferred agent of those remaining, in the event that it remains unmatched by the time position k is reached. Example 12 in Appendix A shows that RSD is not stable. $\square$

3 Two-Sided Matching as a Learning Problem

In this section, we show how to use deep learning to model a two-sided matching mechanism and then formulate two-sided matching as a learning problem.

3.1 Neural Network Architecture

We use a neural network to represent a two-sided matching mechanism. We write $\theta \in \mathbb{R}^d$ to denote the parameters of the network (d parameters). Let $g^\theta : P \rightarrow \Delta(\mathcal{B})$ denote the corresponding mechanism. This network provides a differentiable representation of a distribution on matchings for each preference profile.

We use a feed-forward neural network with $R = 4$ fully-connected hidden layers, $J = 256$ units in each layer, leaky ReLU activations, and a fully-connected output layer. See Figure 1. We represent a preference order at the input to the neural network by adopting an evenly spaced utility function. This is simply a representation choice and still allows the adoption of an FOSD-based metric for the study of approximately SP mechanisms. Given a preference profile $\succ$, the preference representation for a worker $w \in W$ and a firm $f \in F$ is given by $p_w^\succ = (p_{w1}^\succ, \dots, p_{wm}^\succ)$ and $q_f^\succ = (q_{1f}^\succ, \dots, q_{nf}^\succ)$ respectively. We also define $p_{w\perp}^\succ = 0$ and $q_{\perp f}^\succ = 0$. For example:

- For a preference order $\succ$ with $w_1 : f_1, f_2, \perp, f_3$, we have $p_{w1}^\succ = (\frac{2}{3}, \frac{1}{3}, -\frac{1}{3})$.
- For a preference order $\succ$ with $w_1 : f_1, f_2, \perp, f_3, f_4$, we have $p_{w1}^\succ = (\frac{2}{4}, \frac{1}{4}, -\frac{1}{4}, -\frac{2}{4})$.
- For a preference order $\succ$ with $f_1 : w_2, w_1, w_3$, we have $q_{f1}^\succ = (\frac{2}{3}, 1, \frac{1}{3})$.

For some worker w ’s utility for firm j and some firm f ’s utility for worker i , and with $\mathbf{1}(E)$ to indicate the indicator function for event E , we have:

$$p_{wj}^\succ = \frac{1}{m} \left(\mathbf{1}_{j \succ_w \perp} + \sum_{j'=1}^m (\mathbf{1}_{j \succ_w j'} - \mathbf{1}_{\perp \succ_w j'}) \right); \quad q_{if}^\succ = \frac{1}{n} \left(\mathbf{1}_{i \succ_f \perp} + \sum_{i'=1}^n (\mathbf{1}_{i \succ_f i'} - \mathbf{1}_{\perp \succ_f i'}) \right).$$

The vector $(p_{11}^\succ, \dots, p_{nm}^\succ, q_{11}^\succ, \dots, q_{nm}^\succ)$ constitutes the input to the network ($2 \times n \times m$ numbers).

The output of the network is a vector $r \in [0, 1]^{n \times m}$ with $\sum_{j=1}^m r_{wj} \leq 1$ and $\sum_{i=1}^n r_{if} \leq 1$ for every $w \in [n]$ and $f \in [m]$. This describes the marginal probabilities in a randomized matching for this input profile. The network first outputs two sets of scores $s \in \mathbb{R}^{(n+1) \times m}$ and $s' \in \mathbb{R}^{n \times (m+1)}$. We

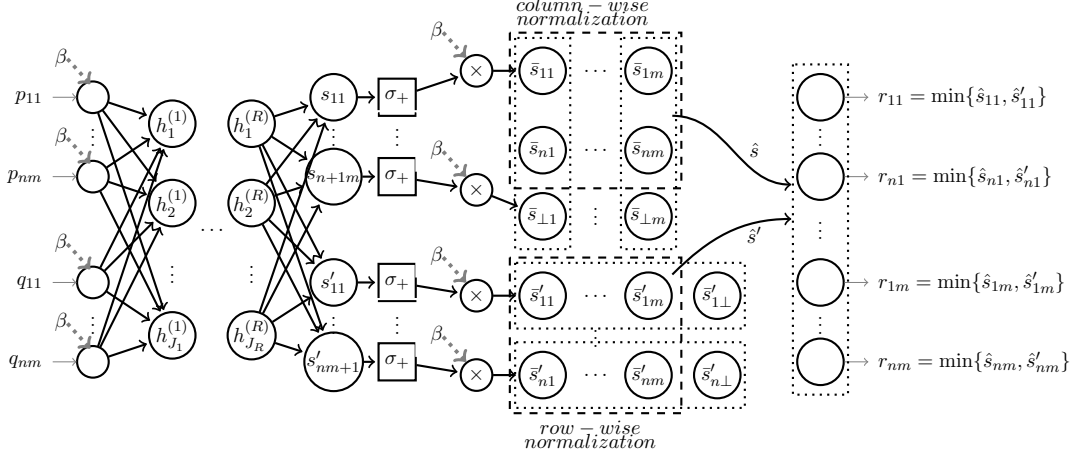


Figure 1: Matching network g for a set of n workers and m firms. Given inputs $p, q \in \mathbb{R}^{n \times m}$ the matching network is a feed-forward network with R hidden layers that uses softplus activation to generate non-negative scores and normalization to compute the randomized matching. We additionally generate a Boolean mask matrix, β , and multiply it with the score matrix before normalization to ensure IR by making the probability of matches that are unacceptable zero.

apply the *softplus function* (denoted by σ_+) element-wise to these scores, where $\sigma_+(x) = \ln(1 + e^x)$. To ensure IR, we first construct a Boolean mask variable β_{wf} , which is zero only when the match is unacceptable to one or both the worker and firm, i.e., when $\perp \succ_w f$ or $\perp \succ_f w$. We set $\beta_{n+1,f} = 1$ for $f \in F$ and $\beta_{w,m+1} = 1$ for $w \in W$. We multiply the scores s and s' element-wise with the corresponding Boolean mask variable to compute $\bar{s} \in \mathbb{R}_{\geq 0}^{(n+1) \times m}$ and $\bar{s}' \in \mathbb{R}_{\geq 0}^{n \times (m+1)}$.

For each $w \in \bar{W}$, we have $\bar{s}_{wf} = \beta_{wf} \ln(1 + e^{s_{wf}})$, for all $f \in F$. For each $f \in \bar{F}$, we have $\bar{s}'_{wf} = \beta_{wf} \ln(1 + e^{s'_{wf}})$, for all $w \in W$. We normalize $\bar{s}$ along the rows and $\bar{s}'$ along the columns to obtain *normalized scores*, $\hat{s}$ and $\hat{s}'$ respectively. The match probability r_{wf} , for worker $w \in W$ and firm $f \in F$, is computed as the minimum of the normalized scores: $r_{wf} = \min \left(\frac{\bar{s}_{wf}}{\sum_{f' \in \bar{F}} \bar{s}_{wf'}}, \frac{\bar{s}'_{wf}}{\sum_{w' \in \bar{W}} \bar{s}'_{w'f}} \right)$. We have $r_{wf} = 0$ whenever $\beta_{wf} = 0$, ensuring that every matching in the support of the distribution will be IR. Based on our construction, the allocation matrix r is always (weakly) doubly stochastic⁶. Budish et al., [7] shows any (weakly) doubly stochastic matrix can be decomposed to a convex combination of 0-1 (weakly) doubly stochastic matrices.

3.2 Formulation as a Learning Problem

We formulate a *loss function* $\mathcal{L}$ that is defined on training data $D = \{\succ^{(1)}, \dots, \succ^{(L)}\}$, with each preference profile $\succ$ in D sampled i.i.d. from a distribution on profiles. We allow for *correlated preferences*; e.g, workers may tend to agree that one of the firms is preferable to one of the other firms, and similarly for firms. The loss function captures a tradeoff between the degree with which stability and ordinal SP is violated. Recall that $g^\theta(\succ) \in [0, 1]^{n \times m}$ denotes the randomized matching. We write $g_{w\perp}^\theta(\succ) = 1 - \sum_{f=1}^m g_{wf}^\theta(\succ)$ and $g_{\perp f}^\theta(\succ) = 1 - \sum_{w=1}^n g_{wf}^\theta(\succ)$ to denote the probability of worker w and firm f being unmatched, respectively.

⁶This is a more general definition for doubly stochastic than is typical. Doubly stochastic is usually defined on a square matrix with the sum of rows and the sum of columns equal to 1.

Stability Violation. For worker w and firm f , we define the *stability violation* at profile $\succ$ as

$$\begin{aligned} stv_{wf}(g^\theta, \succ) = & \left(\sum_{w'=1}^n g_{w'f}^\theta(\succ) \cdot \max\{q_{wf}^\succ - q_{w'f}^\succ, 0\} + g_{\perp f}^\theta(\succ) \cdot \max\{q_{wf}^\succ, 0\} \right) \\ & \times \left(\sum_{f'=1}^m g_{wf'}^\theta(\succ) \cdot \max\{p_{wf}^\succ - p_{wf'}^\succ, 0\} + g_{w\perp}^\theta(\succ) \cdot \max\{p_{wf}^\succ, 0\} \right) \end{aligned} \quad (3)$$

This captures the first kind of *ex ante* justified envy in Definition 2, which is in regard to fractionally matched partners. We can omit the second kind of *ex ante* justified envy because the learned mechanisms satisfy IR through the use of the Boolean mask matrix (and thus have no violations of the second kind).

The average stability violation (or just stability violation) of mechanism g^θ on profile $\succ$ is $stv(g^\theta, \succ) = \frac{1}{2} \left(\frac{1}{m} + \frac{1}{n} \right) \sum_{w=1}^n \sum_{f=1}^m stv_{wf}(g^\theta, \succ)$. We can define the expected stability violation $STV(g^\theta) = \mathbb{E}_\succ stv(g^\theta, \succ)$. We also write $stv(g^\theta)$ to denote the average stability violation of a mechanism g^θ on the training data.

Theorem 10. A randomized matching mechanism g^θ is *ex ante stable up to zero-measure events* if and only if $STV(g^\theta) = 0$.

Proof. Since $stv(g^\theta, \succ) \geq 0$ then $STV(g^\theta) = \mathbb{E}_\succ stv(g^\theta, \succ) = 0$ if and only if $stv(g^\theta, \succ) = 0$ except on zero measure events. Moreover, $stv(g^\theta, \succ) = 0$ implies $stv_{wf}(g^\theta, \succ) = 0$ for all $w \in W$, all $f \in F$. This is equivalent to no justified envy. For firm f , this means $\forall w' \neq w, q_{wf}^\succ \leq q_{w'f}^\succ$ if $g_{w'f}^\theta > 0$ and $q_{wf}^\succ \leq 0$ if $g_{\perp f}^\theta > 0$. Then there is no justified envy for firm f . Analogously, there is no justified envy for worker w . If g^θ is *ex ante stable*, it trivially implies $STV(g^\theta) = 0$ by definition. $\square$

Regret. We turn now to quantifying the degree of approximation to ordinal SP. For a valuation profile $\succ \in P$ and a mechanism g^θ , the *regret* to a worker (firm) for submitting their truthful preference, for some valid utility function, is the maximum amount by which the worker (firm) could increase their expected utility through a mis-report, fixing the reports of others. Let $\succ_{-i} = (\succ_1, \dots, \succ_{i-1}, \succ_{i+1}, \dots, \succ_{n+m})$. The regret to worker w at preference order $\succ$ is

$$\text{regret}_w(g^\theta, \succ) = \max_{\succ'_w \in P} \left(\max_{f' \in F, f' \succ_w \perp} \sum_{f \in F: f \succ_w f'} g_{wf}^\theta(\succ'_w, \succ_{-w}) - g_{wf}^\theta(\succ_w, \succ_{-w}) \right). \quad (4)$$

The regret to a firm f at preference order $\succ$ is

$$\text{regret}_f(g^\theta, \succ) = \max_{\succ'_f \in P} \left(\max_{w' \in W, w' \succ_f \perp} \sum_{w \in W: w \succ_f w'} g_{wf}^\theta(\succ'_f, \succ_{-f}) - g_{wf}^\theta(\succ_f, \succ_{-f}) \right). \quad (5)$$

We define the *average regret* of mechanism g^θ on profile $\succ$ as

$$\text{regret}(g^\theta, \succ) = \frac{1}{2} \left(\frac{1}{n} \sum_{w \in W} \text{regret}_w(g^\theta, \succ) + \frac{1}{m} \sum_{f \in F} \text{regret}_f(g^\theta, \succ) \right). \quad (6)$$

We define the *expected regret* as $RGT(g^\theta) = \mathbb{E}_\succ \text{regret}(g^\theta, \succ)$. We can also write $rgt(g^\theta)$ to denote the average regret of the mechanism on training data.

Theorem 11. A randomized matching mechanism g^θ is *ordinal SP up to zero-measure events* if and only if $RGT(g^\theta) = 0$.

The proof is similar to Theorem 10 and it is deferred to Appendix B.

3.3 Training Procedure

For a mechanism parameterized as g^θ , the training problem that we formulate is as follows:

$$\min_{\theta} \lambda \cdot stv(g^\theta) + (1 - \lambda) \cdot rgt(g^\theta), \quad (7)$$

where $\lambda \in [0, 1]$ is a parameter that controls the tradeoff between approximate stability and approximate strategy-proofness. We make use of stochastic gradient descent to solve (7). The derivative of the degree of violation of stability with respect to network parameters is straightforward to calculate. The derivative of regret is complicated by the nested maximization in the definition of regret. Let $\hat{\succ}_i^{(\ell)}$ denote a *defeating preference* for agent i (a worker or firm) at preference profile $\succ^{(\ell)}$, that is a preference that reveals a violation in FOSD, or just $\succ_i^{(\ell)}$ in the case that there is no such mis-report (fixing the others' reports to be truthful). For the 4x4 two-sided matching problems that we study we can enumerate all possible misreports to compute the defeating preference. We can then compute the derivative of regret for agent i with respect to the network parameters by fixing the misreport to the defeating valuation and adopting truthful reports for the others.

For the case of RSD, we also need to measure the degree of IR violation of mechanisms (this necessarily zero for the other mechanisms). We define the IR violation at profile $\succ$ as

$$irv(g, \succ) = \frac{1}{2m} \sum_{w=1}^n \sum_{f=1}^m g_{wf}(\succ) \cdot (\max\{-q_{wf}, 0\}) + \frac{1}{2n} \sum_{w=1}^n \sum_{f=1}^m g_{wf}(\succ) \cdot (\max\{-p_{wf}, 0\}).$$

We extend the stability violation to include the average IR violation on test data when reporting the stability violation for RSD.

4 Experimental Results

We use the Adam optimizer to train our models for 50,000 mini-batch iterations, with mini-batches of size 1024 profiles. We report the results on a test set of 204,800 preference profiles. We use the *PyTorch* deep learning library, and all the experiments are run on a cluster of NVIDIA GPU cores. We only present the representative experiments here and refer the reader to see Appendix D for additional experiments.

We study both uncorrelated and correlated preference environments:

- For *uncorrelated preference orders*, for each worker or firm, we first sample uniformly at random from all preference orders, and then, with probability 0.2 (truncation probability), we choose at random a position at which to truncate this agent's preference order.
- For *correlated preference orders*, we first sample a preference profile as in the uncorrelated case. We also sample a single preference order on firms, and a single preference order on workers. For each agent, with probability, $p_{\text{corr}} > 0$, we replace that agent's preference order with the common preference order for its side of the market.

Specifically, we consider the following four matching environments:

- $n = 4$ workers and $m = 4$ firms with uncorrelated preferences.
- $n = 4$ workers and $m = 4$ firms with correlated preferences, and varying $p_{\text{corr}} = \{0.25, 0.5, 0.75\}$.

We compare the performance of the learned mechanisms, varying parameter λ between 0 and 1 to control the tradeoff between stability and strategy-proofness, with the best of worker- and firm-proposing DA (best in terms of average SP violation over the test data), and with RSD. We plot the resulting frontier on the stability violation ($stv(g^\theta)$) and SP violation ($rgt(g^\theta)$) in Figure 2. Because

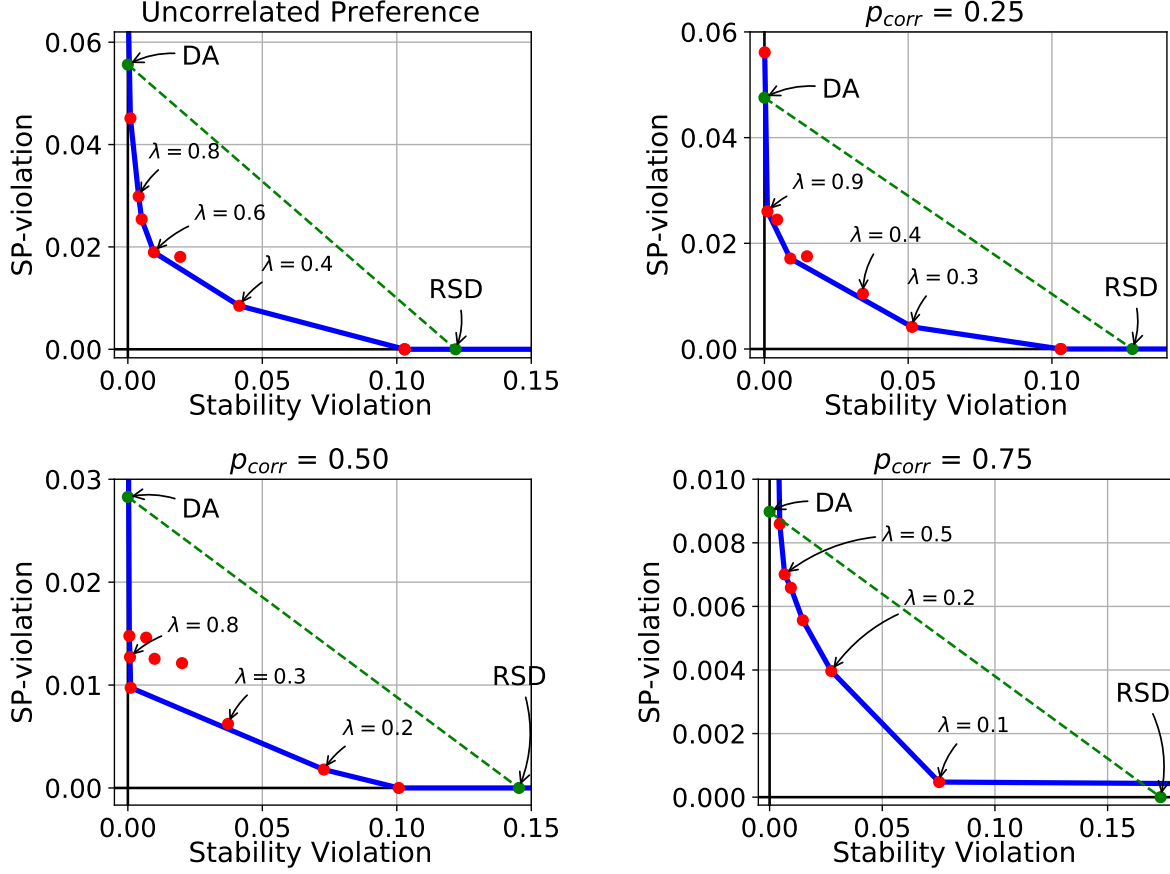


Figure 2: Comparing stability violation and strategy-proofness violation from the learned mechanisms for different choices of λ (red dots) with the best of worker- and firm-proposing DA, as well as RSD, in 4x4 two-sided matching, and considering uncorrelated preference orders as well as settings with increasing correlation ($p_{corr} \in \{0.25, 0.5, 0.75\}$). The stability violation for RSD includes IR violations.

the RSD mechanism does not guarantee IR, we include the IR violation of the RSD mechanism in the reported stability violation (none of the other mechanisms fail IR).

At $\lambda = 0.0$ we learn a mechanism with very low regret (≈ 0) but achieves poor stability. This performance is similar to that of RSD. For large values of λ we learn a mechanism that approximates DA. For intermediate values, we find interesting tradeoffs between SP and stability. Notably, for lower levels of correlations we see substantially better SP than DA for very little loss in stability. For higher levels of correlations, we see substantially better stability than RSD for very little loss in SP. Comparing the scale of the y-axes, we also see that increasing correlation tends to reduce the opportunity for strategic behavior across both the DA and learned mechanisms.

Figure 3 shows, in each environment, the expected welfare from the matchings, measured for the equi-spaced utility function. We compare it with the maximum of the expected welfare achieved by the worker- and firm- proposing DA mechanisms. As we increase λ , our learned mechanism achieves better welfare. This is an idealized view of welfare because it assumes truthful inputs (and we would expect welfare to be reduced to the extent that mechanisms are not SP). For choices of λ in the range 0.8 and higher that provide interesting opportunities for improving SP relative to DA, we also see good welfare (together with better SP!). For small values of λ the learned mechanisms have relatively low welfare compared to RSD. This is interesting and suggests that achieving IR together with SP (RSD is not IR!) is very challenging in two-sided markets.

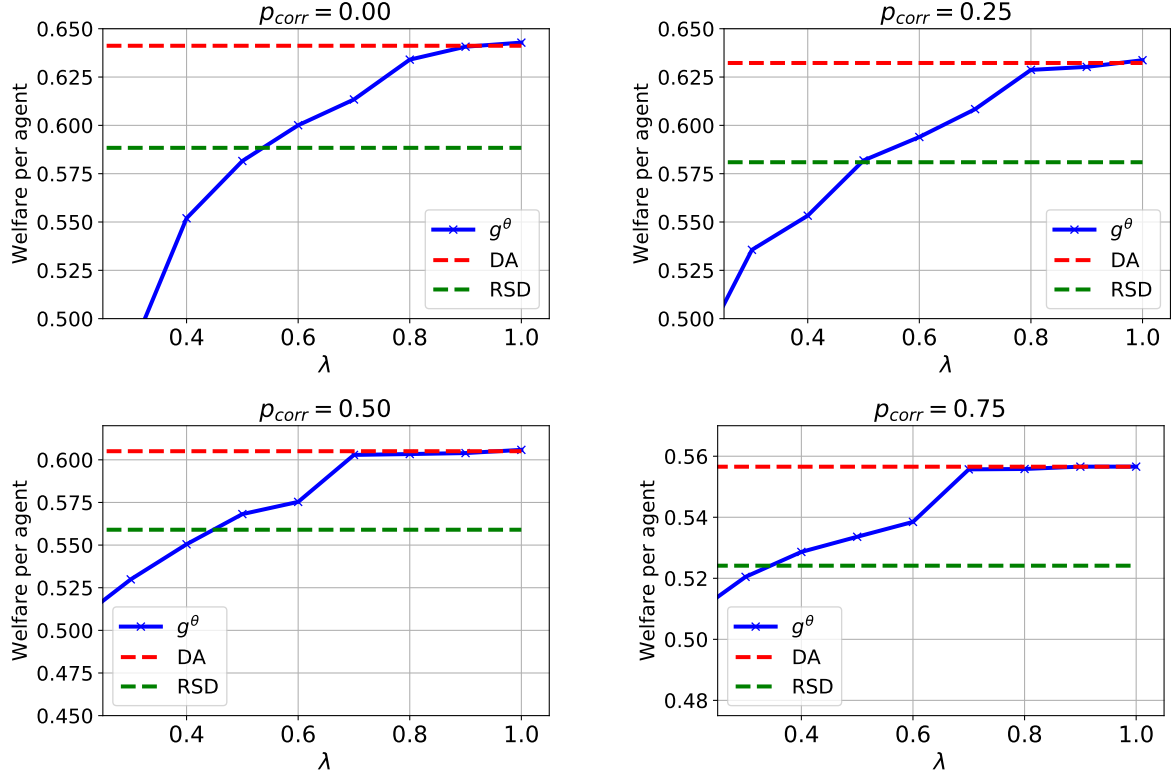


Figure 3: Comparing welfare per agent of the learned mechanisms for different values of the tradeoff parameter λ with the best of firm- and worker- proposing DA, and RSD, and considering uncorrelated preference orders as well as settings with increasing correlation ($p_{corr} \in \{0.25, 0.5, 0.75\}$).

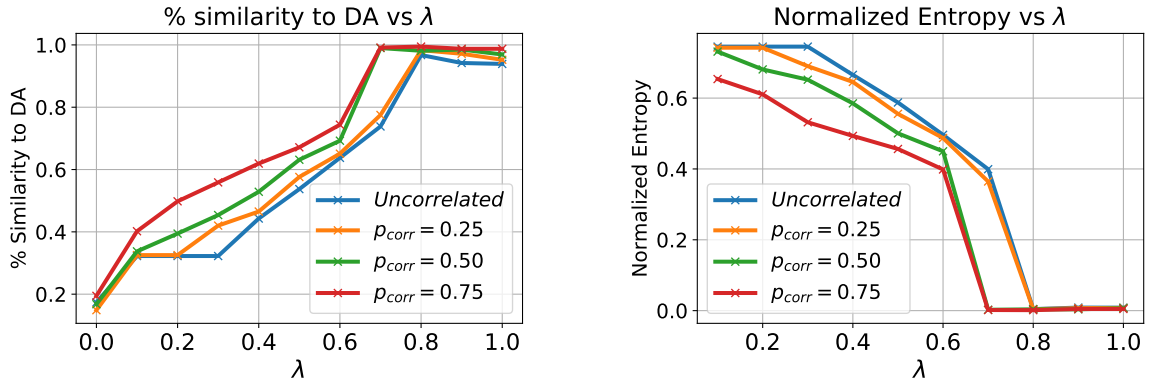


Figure 4: **Left:** Comparing the average instance-wise max similarity scores ($sim(g^\theta)$) of the learned mechanisms with worker- and firm-proposing DA. **Right:** Normalized entropy of the learned mechanisms for different values of the tradeoff parameter λ . Results are shown for uncorrelated preferences, as well as increasing correlation ($p_{corr} \in \{0.25, 0.5, 0.75\}$).

In understanding the learned mechanisms, we are also interested in their similarity to DA. Let w -DA and f -DA denote the worker- and firm-proposing DA, respectively. For a given preference profile, we compute the similarity score as

$$\text{sim}(g^\theta, \succ) = \max \left\{ \frac{\sum_{(w,f): g_{wf}^{w\text{-DA}}(\succ)=1} g_{wf}^\theta(\succ)}{\sum_{(w,f): g_{wf}^{w\text{-DA}}(\succ)=1} 1}, \frac{\sum_{(w,f): g_{wf}^{f\text{-DA}}(\succ)=1} g_{wf}^\theta(\succ)}{\sum_{(w,f): g_{wf}^{f\text{-DA}}(\succ)=1} 1} \right\}. \quad (8)$$

This calculates the agreement between the two mechanisms (normalized by the size of the DA matching), and taking the best of w -DA or f -DA. We denote the average similarity score on test data as $\text{sim}(g^\theta)$. Figure 4 shows this average similarity to DA as λ varies. As we increase λ , i.e. penalize stability violations more, the matchings corresponding to the learned mechanisms get increasingly close to that of the DA mechanisms (while also achieving better SP properties, as per Figure 2!). We also quantify the degree of randomness of the learned mechanism by computing the *normalized entropy per agent*, taking the expectation over all preference profiles. For a given profile $\succ$, we compute normalized entropy as (it is 0 for a deterministic mechanism):

$$H(\succ) = -\frac{1}{2n} \sum_{w \in W} \sum_{f \in \bar{F}} \frac{g_{wf}(\succ) \log_2 g_{wf}(\succ)}{\log_2 m} - \frac{1}{2m} \sum_{f \in F} \sum_{w \in \bar{W}} \frac{g_{wf}(\succ) \log_2 g_{wf}(\succ)}{\log_2 n}.$$

Figure 4 shows how the average normalized entropy changes with λ . As we increase λ , and the mechanisms come closer to the DA design, the learned mechanisms becomes less random.

5 Conclusion

These results suggest a new target for economic theory: are there mechanisms that are almost as stable as DA while attaining better SP properties than DA? This is very intriguing given the importance of DA to economic theory. Also of interest is that our results suggest that achieving IR together with SP (please recall that RSD is not IR!) is very challenging in two-sided markets; this is reflected in the lower welfare of our mechanisms on the part of the frontier that emphasizes SP. The experimental results also show that for larger values of λ (and thus with more attention to stability), the learned mechanisms have comparable welfare to DA and are relatively deterministic.

From a methodological perspective, we see a few opportunities for future research. Indeed, our current results are presented in relatively small, 4x4 markets. In the current training pipeline, we enumerate mis-reports in order to calculate regret and thus the degree of approximation to ordinal SP. This is the bottleneck to further scaling. It is interesting to use utility-improving utility gradients (input gradients) to find defeating mis-reports, in the spirit of Dütting et al. [10]. The new challenge here comes from needing to handle ordinal and thus discrete preference inputs. In addition, it is interesting to study whether we can find better notion of strategy-proofness violation to model approximately SP mechanism.

References

- [1] A. Abdulkadiroğlu and T. Sönmez. Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica*, 66(3):689–702, 1998.
- [2] A. Abdulkadiroğlu, P. A. Pathak, and A. E. Roth. Strategyproofness versus efficiency in matching with indifference: Redesigning the NYC high school match. *American Economic Review*, 99:1954–1978, 2009.
- [3] A. Abdulkadiroğlu and T. Sönmez. School choice: A mechanism design approach. *American Economic Review*, 93:729–747, 2003.
- [4] J. Alcalde and S. Barberà. Top dominance and the possibility of strategy-proof stable solutions to matching problems. *Economic Theory*, 4(3):417–435, 1994.
- [5] S. Alva and V. Manjunath. The impossibility of strategy-proof, Pareto efficient, and individually rational rules for fractional matching. *Games and Economic Behavior*, 119:15–29, 2020.
- [6] S. Athey. The Impact of Machine Learning on Economics. In *The Economics of Artificial Intelligence: An Agenda*, NBER Chapters, pages 507–547. National Bureau of Economic Research, Inc, 2018.
- [7] E. Budish, Y.-K. Che, F. Kojima, and P. Milgrom. Designing random allocation mechanisms: Theory and applications. *American Economic Review*, 103(2):585–623, April 2013.
- [8] M. J. Curry, P. Chiang, T. Goldstein, and J. P. Dickerson. Certifying strategyproof auction networks. *CoRR*, abs/2006.08742, 2020.
- [9] L. E. Dubins and D. A. Freedman. Machiavelli and the Gale-Shapley algorithm. *American Mathematical Monthly*, 88:485–494, 1981.
- [10] P. Duetting, Z. Feng, H. Narasimhan, D. C. Parkes, and S. S. Ravindranath. Optimal auctions through deep learning. In *Proceedings of the 36th International Conference on Machine Learning, ICML*, pages 1706–1715, 2019.
- [11] A. Erdil. Strategy-proof stochastic assignment. *Journal of Economic Theory*, 151:146 – 162, 2014.
- [12] A. Erdil and H. Ergin. What’s the matter with tie-breaking? Improving efficiency in school choice. *American Economic Review*, 98(3):669–89, 2008.
- [13] Z. Feng, H. Narasimhan, and D. C. Parkes. Deep learning for revenue-optimal auctions with budgets. In *Proceedings of the 17th Conference on Autonomous Agents and Multi-Agent Systems*, pages 354–362, 2018.
- [14] D. Gale and L. S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.
- [15] N. Golowich, H. Narasimhan, and D. C. Parkes. Deep learning for multi-facility location mechanism design. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 261–267, 2018.
- [16] N. Immorlica and M. Mahdian. Incentives in large random two-sided markets. *ACM Transactions on Economics and Computation*, 3:#14, 2015.

- [17] F. Kojima and P. A. Pathak. Incentives and stability in large two-sided matching markets. *American Economic Review*, 99:608–627, 2009.
- [18] S. Lee. Incentive compatibility of large centralized matching markets. *The Review of Economic Studies*, 84(1):444–463, 2016.
- [19] T. Mennle and S. Seuken. The pareto frontier for random mechanisms. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, EC ’16, page 769, New York, NY, USA, 2016. Association for Computing Machinery.
- [20] T. Mennle and S. Seuken. Partial strategyproofness: Relaxing strategyproofness for the random assignment problem. *Journal of Economic Theory*, 191:105–144, 2021.
- [21] H. Narasimhan, S. Agarwal, and D. C. Parkes. Automated mechanism design without money via machine learning. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, pages 433–439, 2016.
- [22] P. A. Pathak and T. Sönmez. Leveling the playing field: Sincere and sophisticated players in the Boston mechanism. *American Economic Review*, 98(4):1636–1652, 2008.
- [23] J. Rahme, S. Jelassi, J. Bruna, and S. M. Weinberg. A permutation-equivariant neural network architecture for auction design. *CoRR*, abs/2003.01497, 2020.
- [24] A. E. Roth. The economics of matching: Stability and incentives. *Mathematics of Operations Research*, 7(4):617–628, 1982.
- [25] A. E. Roth and E. Peranson. The redesign of the matching market for american physicians: Some engineering aspects of economic design. *American Economic Review*, 89(4):748–780, September 1999.
- [26] A. E. Roth, U. G. Rothblum, and J. H. V. Vate. Stable matchings, optimal assignments, and linear programming. *Mathematics of Operations Research*, 18(4):803–828, 1993.
- [27] A. E. Roth and M. Sotomayor. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*, volume 18 of *Econometric Society Monographs*. Cambridge University Press, 1990.
- [28] W. Shen, P. Tang, and S. Zuo. Automated mechanism design via neural networks. In *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems*, 2019.
- [29] T. Sönmez. Bidding for army career specialties: Improving the ROTC branching mechanism. *Journal of Political Economy*, 121:186–219, 2013.
- [30] T. Sönmez and T. B. Switzer. Matching with (branch-of-choice) contracts at United States Military Academy. *Econometrica*, 81:451–488, 2013.

Deep Learning for Two-Sided Matching

Appendix

A RSD is not Stable

In the following example, we show RSD mechanism is not stable.

Example 12. Consider $n = 3$ workers and $m = 3$ firms with the following preference orders:

$$\begin{aligned} w_1 : f_2, f_3, f_1, \perp & \quad f_1 : w_1, w_2, w_3, \perp \\ w_2 : f_2, f_1, f_3, \perp & \quad f_2 : w_2, w_3, w_1, \perp \\ w_3 : f_1, f_3, f_2, \perp & \quad f_3 : w_3, w_1, w_2, \perp \end{aligned}$$

The matching found by worker-proposing DA is $(w_1, f_3), (w_2, f_2), (w_3, f_1)$. This is a stable matching. If f_1 truncates and misreports its preference as $f_1 : w_1, w_2, \perp, w_3$, the matching found is $(w_1, f_1), (w_2, f_2), (w_3, f_3)$. Firm f_1 is matched with a more preferred worker, and hence the mechanism is not strategy-proof. Now consider the matching under RSD. The marginal matching probabilities r is given by:

$$r = \begin{pmatrix} \frac{11}{24} & \frac{1}{4} & \frac{7}{24} \\ \frac{1}{6} & \frac{3}{4} & \frac{1}{12} \\ \frac{3}{8} & 0 & \frac{5}{8} \end{pmatrix}$$

f_2 and w_2 are the most preferred options for w_2 and f_2 respectively and they would prefer to be matched with each other always rather than being fractionally matched with each other. Here (w_2, f_2) is a blocking pair and thus RSD is not stable.

B Proof of Theorem 11

Proof. Since $\text{regret}(g^\theta, \succ) \geq 0$ then $RGT(g^\theta) = \mathbb{E}_\succ \text{regret}(g^\theta, \succ) = 0$ if and only if $\text{regret}(g^\theta, \succ) = 0$ except on zero measure events. Moreover, $\text{regret}(g^\theta, \succ) = 0$ implies $\text{regret}_w(g^\theta, \succ) = 0$ for any worker w and $\text{regret}_f(g^\theta, \succ) = 0$ for any firm f . For each worker w , by definition of $\text{regret}_w(g^\theta, \succ)$ (Eq. (4)), $\sum_{f \in F: f \succ_w f'} g_{wf}(\succ_w, \succ_{-w}) \geq \sum_{f \in F: f \succ_w f'} g_{wf}(\succ'_w, \succ_{-w})$ and similar result holds for each firm f . Then g^θ is FOSD-SP. If g^θ is FOSD-SP, it is straightforward to show that $\text{regret}(g^\theta, \succ) = 0$. $\square$

C Training Details

For all our settings, we use a neural network with $R = 4$ hidden layers with 256 hidden nodes each. We use the leaky ReLU activation function at each of these layers. To train our neural network, we use the Adam Optimizer with decoupled weight delay regularization (implemented as AdamW optimizer in PyTorch) We set the learning rate to 0.005 for uncorrelated preferences setting and 0.002 when $p_{corr} = \{0.25, 0.5, 0.75\}$. The remaining hyperparameters of the optimizer are set to their default values. We sample a fresh minibatch of 1024 profiles and train our neural networks for a total of 50000 minibatch iterations. We reduce the learning rate by half once at 10000th iteration and once at 25000th iteration. We report our results on 204800 preference profiles.

For our training, we use a single Tesla V-100 GPU. For each setting, the neural network takes 4.6 hours to train.

D Additional Experiments

In addition to the experiments in Section 4, we also consider the following matching environments:

- $n = 4$ workers and $m = 4$ firms with correlated preference with $p_{\text{corr}} = 0.25$ and no truncation.
- $n = 4$ workers and $m = 4$ firms with correlated preference with $p_{\text{corr}} = 0.25$ and but preferences are truncated with $p_{\text{trunc}} = 0.5$

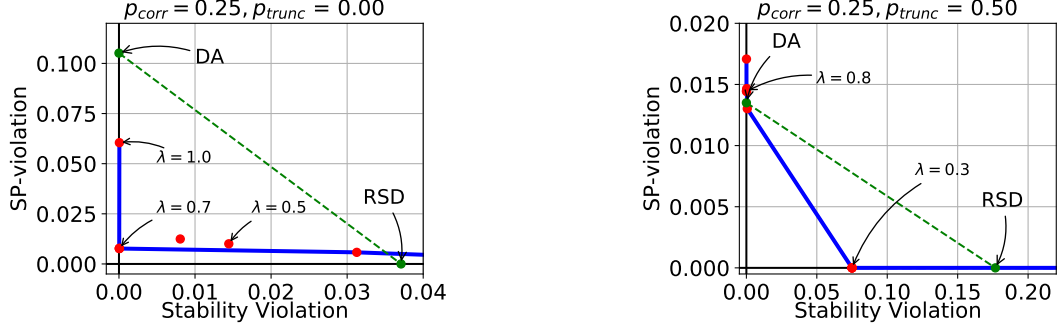


Figure 5: Comparing stability violation and strategy-proofness violation from the learned mechanisms for different choices of λ (red dots) with the best of worker- and firm-proposing DA, as well as RSD, in 4x4 two-sided matching, with correlated preference ($p_{\text{corr}} = 0.25$) for different values of truncation probability. The stability violation for RSD includes IR violations.

We plot the resulting frontier for both these settings in Figure 5.