RATIONALITY
& COMPETITION
CRC TRR 190

Reciprocal Preferences in Matching Markets

Timm Opitz (Max Planck Institute for Innovation and Competition, LMU Munich)

Christoph Schwaiger (LMU Munich)

Discussion Paper No. 388

February 20, 2023

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Georg Weizsäcker, Ph.D., Humboldt University Berlin, 10117 Berlin, Germany
info@rationality-and-competition.de

Reciprocal Preferences in Matching Markets

Timm Opitz[†] Christoph Schwaiger[‡]

This version: February 2023

Abstract

Agents with *reciprocal preferences* prefer to be matched to a partner who also likes to collaborate with them. In this paper, we introduce and formalize reciprocal preferences, apply them to matching markets, and analyze the implications for mechanism design. Formally, the preferences of an agent can depend on the preferences of potential partners and there is incomplete information about the partners' preferences. We find that there is no stable mechanism in standard two-sided markets. Observing the final allocation of the mechanism enables agents to learn about each other's preferences, leading to instability. However, in a school choice setting with one side of the market being non-strategic, modified versions of the deferred acceptance mechanism can achieve stability. These results provide insights into non-standard preferences in matching markets, and their implications for efficient information and mechanism design.

Keywords: *Market Design, Matching, Reciprocal Preferences, Non-standard Preferences, Gale-Shapley Deferred Acceptance Mechanism, Incomplete Information*

JEL Codes: *C78, D47, D82, D83, D91*

[†] Max Planck Institute for Innovation and Competition, Munich, Germany and Department of Economics, Ludwig-Maximilians-Universität (LMU), Munich, Germany; email: timm.opitz@ip.mpg.de

[‡] Department of Economics, Ludwig-Maximilians-Universität (LMU), Munich, Germany; email: christoph.schwaiger@econ.lmu.de

We would like to thank Marina Chugunova, Florian Englmaier, Svenja Friess, David Heller, Elisabeth Hofmeister, Dorothea Kübler, Matthias Lang, Vincent Meissner, Cristina Rujan, Klaus Schmidt, Marco Schwarz, Roland Strausz, Erina Ytsma, and seminar audiences at ESA Meeting 2021 (virtual), LMU Munich, and the Max Planck Institute for Innvoation and Competition for their helpful thoughts and comments. Financial support by Deutsche Forschungsgemeinschaft through CRC TRR 190 (project number 280092119), the GRK 1928, and the Max Planck Institute for Innovation and Competition is gratefully acknowledged.

1 Introduction

Standard matching theory assumes that agents do not care about the preferences of their potential partners. In contrast, we study the observation that individuals *like to be liked*. For example, school principals “want to run a school where people want to be there”, and hence “take into account [...] that one kid wants to go there more than another kid”.[1] The same holds true in the labor market—employees prefer to work for firms that favor them, and may even reconsider a job offer after learning that they were not the first-choice candidate.[2] Conversely, an employer may look for a worker who prefers to work for them rather than for another company. Agents who prefer to be matched to a partner who likes to be matched with them are defined as having *reciprocal preferences*.

In this paper, we introduce reciprocal preferences, apply them to matching markets, and analyze implications for mechanism design. We resort to the standard marriage model (Gale & Shapley, 1962), where agents are one-to-one matched. We augment the setting by allowing reciprocal preferences on one side of the market. These agents care about the preferences of the agents on the other side but they do not know these preferences perfectly. We analyze standard two-sided markets and school choice settings, in which one side of the market is non-strategic. In both settings, we investigate the implications of reciprocal preferences on stability when the Deferred Acceptance (DA) mechanism is applied (Gale & Shapley, 1962). The DA mechanism plays a key role in two-sided matching markets because it achieves stability under standard assumptions −no participant benefits from breaking up the formed match, which implies Pareto efficiency. Moreover, we generalize and analyze stability in a broad class of matching mechanisms.

We derive three main results. First, stability of the DA mechanism in two-sided matching markets ceases to hold when agents care about others' preferences over themselves without perfectly knowing these preferences. The DA mechanism achieves stability in a standard setting under complete information. Under incomplete information, preference misrepresentations of agents may lead to instability. In our setting with incomplete information and reciprocal preferences, instability arises as well. However, this instability is not due to strategic play but to an updating about the preferences of other agents in the market. We test stability under two different degrees of information revelation. Under the more restrictive notion, agents only observe

[1] David M. Herszenhorn, Council Members See Flaws in School-Admissions Plan New York Times, Nov. 19, 2004, https://www.nytimes.com/2004/11/19/education/council-members-see-flaws-in-schooladmissions-plan.html, accessed 01/31/2022).

[2] See https://www.forbes.com/sites/lizryan/2018/01/20/im-the-second-choice-candidate-should-i-still-take-the-job, accessed 01/31/2022, for an example on the perceived importance of being the most preferred candidate.

the final matches. Under the less restrictive notion, they also learn their matched partner's type. Under both stability notions, agents might infer the true preferences (and thus the types) of other agents, which causes instability.

Second, we show that there is no alternative mechanism that guarantees a stable matching in two-sided markets with (one-sided) reciprocal preferences and uncertainty about true preferences. Stability cannot be achieved because agents update about each others' preferences by seeing the final matching.

Third, modified versions of the DA mechanism achieve stability in a school choice setting where one side of the market must state its true preferences based on laws and regulations, which is the case with schools. We show that the standard DA mechanism does not prevent instability in school choice settings when students with reciprocal preferences face uncertainty about the school's preferences. Alternative mechanisms can resolve the problem of uncertainty, which can be either a sequential variant of the DA mechanism where students learn schools preferences or a variant of the DA mechanism that allows students to state their complete reciprocal preference profile.

These results help us to understand why certain matching markets do not work satisfactorily. If reciprocal preferences are strong and standard matching mechanisms do not consider these, then involved parties may be reluctant to adopt a centralized matching mechanism. They may instead prefer decentralized markets because they allow them to learn the preferences of others.[3] Agents with reciprocal preferences may also have incentives to modify an existing matching mechanism. To attract especially interested candidates, universities introduce early admissions (Avery & Levin, 2010), give a bonus to students who rank the institution well, or even only accept these candidates (Chiu & Weng, 2009).[4] While individually rational, these modifications may prevent the desired functioning of matching mechanisms. If a mechanism does not achieve the goal of finding the best possible partner and modifications are not feasible, then it may even fail completely after some time (McKinney, Niederle, & Roth, 2005).

We derive important policy implications for the efficient design of markets when agents have reciprocal preferences by combining our results of two-sided mechanisms and the school choice setting. The feasibility to achieve stable allocations in a school choice setting implies that it may be advantageous to move a standard two-sided market closer to a school choice setting. Students

[3] For example, Gundlach (2021) documents that only three percent of parents of kindergarten-age children in Germany approve the use of an algorithm alone when deciding on the allocation of daycare places.

[4] The idea that less interested candidates will receive a deduction has been formally incorporated in the centralized mechanism for high school admission in Taiwan (Dur et al., 2022).

in a school choice setting can be helped if the (binding) admission criteria are known before the mechanism takes place and if they have information about the past success criteria. Given that schools are non-strategic, this eliminates uncertainty about their true preferences and helps the other market side to submit their ranking. In centralized matching mechanisms where one side is an institution connected to the market designer, it can be desirable to force one side of the market to state their true preferences. Additionally, we point to interesting trade-offs in the information design of matching mechanisms. The communication of the final matching can cause allocations to break apart that would otherwise be stable without this information.

Our theoretical framework extends to a broader range of applications than just reciprocal preferences. We start from the main assumption that the type of potential partners is uncertain, and agents care about these type. Specifically, agents with reciprocal preferences prefer partners who rank them favorably. However, the framework applies to any situation where the type of a potential partners is important to an agent, and these types differ in their preferences. This does not even require that an agent cares about how they are ranked by their potential partners. For example, an applicant learns about the employer by observing who else got invited for an interview.

Our findings connect to two strands of the literature of matching mechanisms. First, we contribute to an emerging literature of non-standard preferences in matching markets by introducing reciprocal preferences. Fernandez (2020) highlights that regret aversion may induce truth-telling for both market sides in the DA mechanism. Other studies emphasize how non-standard preferences prevent the desired functioning of allocation mechanisms. These studies show that costly information acquisition about one's preferences leads to higher acceptance rates of early offers, despite not being more desirable (Grenet, He, & Kübler, 2022), and that expectation-based loss aversion can lead to non-truthful preference submissions (Dreyfuss, Heffetz, & Rabin, 2022; Meisner & von Wangenheim, 2023). Meanwhile, Antler (2015) finds that a slight modification of the standard DA mechanism preserves stability when the agents' preferences are directly affected by the reported preferences of others. In his model, agents have perfect information about others' preferences, while outsider observers do not. Because agents have image concerns, they prefer to be matched with a partner who ranked them highly. In contrast, our analysis assumes that agents care about the unknown *true* preferences and the types of their potential partners, rather than the stated preferences in an environment of complete information. In complementary work (Opitz & Schwaiger, 2023), we validate the assumption of reciprocal preferences experimentally, and show that agents indeed care about the preferences of their partners.

Second, we contribute to the literature of incomplete information in matching markets by studying an environment without perfect knowledge about the preferences of the potential partners. Roth (1989) introduces uncertainty about the preferences of other players. Roth (1989) analyzes stability under the assumption that all preferences were to become common knowledge and finds that no mechanism is stable with respect to the true preferences. If an agent has enough information about the preferences of other players, then stating non-truthful preferences can be optimal. An agent misrepresents their preferences to reject a candidate that then applies somewhere else, which starts a process of new applications and rejections in the market. This can lead to a match with a partner who is preferred over the one the agent rejected. Due to uncertainty, this process can either lead to a more preferred candidate for the agent or fail and cause instability. Fernandez, Rudov, and Yariv (2022) show that the results of Roth (1989) hold with only minimal uncertainty on the proposing side of the market in a DA mechanism. In contrast, in our model, instability arises from learning about the preferences of others, which influences the expected utility of being matched with an agent. This implies that our findings are robust to the market side that faces uncertainty and hold irrespective of strategic play.

Most closely related to our work is the literature of incomplete information and interdependent preferences, where match utilities depend on the type of the agent one is matched with. Hence, in contrast to Roth (1989), players' types affect their desirability and not just their preference reporting strategies. Both Chakraborty, Citanna, and Ostrovsky (2010) and Liu et al. (2014) build on the idea that agents can draw insights from the actions of other players. In Chakraborty, Citanna, and Ostrovsky (2010), each school receives a signal about the quality of the students on the other market side before submitting their preferences to the mechanism. By observing the final matching, a school can learn about the signals that other schools received and update their belief about the quality of a student. Anticipating that rematching is possible, the school is tempted to strategically misrepresent its signal, which can lead to instability. Liu et al. (2014) analyze a setting with transferable utility in which firms and workers are already matched. A firm only knows the quality of their matched worker. Firms can still update their information about the quality of other workers by observing rematching (or absence of rematching) in the market. In our model, types of agents differ in their preference profiles and that other agents care about these. Hence, agents may update their beliefs about the underlying types by only observing the final matching. Our stability results do not require anticipated rematching, nor do preferences of the own market side affect the desirability of other agents.

The rest of this paper is structured as follows. Section 2 outlines an illustrative example that provides intuitions for the formulation of reciprocal preferences and their consequences in matching markets. In Section 3, we set up the formal matching market and introduce reciprocal preferences. Section 4 analyzes the implications of reciprocal preferences in the DA mechanism, we then generalize our findings to a broader class of mechanisms in Section 5. In Section 6, we extend our analysis to a school choice setting with one non-strategic market side. Finally, Section 7 discusses the implications of our findings and concludes.

2 Illustrative Example

In this section, we provide intuitions for the consequences of reciprocal preferences on stability in matching markets through an illustrative example before presenting the main theoretical analysis. The example shows that the DA mechanism leads to an unstable allocation.

We consider a small job market with three firms (A, B, C), three workers (I, II, III), and a DA mechanism to match them one-to-one. All firms (A, B, C), as well as workers II and III have standard preferences. Both workers II and III prefer to only work for either of the firms over being unmatched. Worker II wants to work only for firm A, and worker III wants to be matched only with firm C. Worker I has reciprocal preferences: she cares how she is ranked by a firm.[5] She prefers working for firm A over firm B if firm A ranks her first. If firm A ranks her second, then she prefers firm B over firm A. This means that her preference list is given by $A_1 \succ B \succ A_2$. The indices denote the true rank assigned to her by the respective firm. The (reciprocal) preferences of workers are common knowledge. Although the type of a firm is private knowledge, every agent knows the distribution of firms' types.

Acquiring information about the true preferences of firms is challenging in practice. If the firms' preferences were perfectly observable, then the preference list of worker I would reduce to the standard case. Worker I prefers $A \succ B$ if firm A ranks her first, while she prefers $B \succ A$ if firm A ranks her second. In reality, potential employees typically only have limited information about the exact demands of a firm and face uncertainty about the characteristics of the competing applicants. Moreover, employers may not be interested in truthfully revealing their preferences so that they can give each applicant the impression that they are a preferred candidate. Hence, we allow for uncertainty about the firms' preferences.

We incorporate uncertainty about the firms' preferences as follows (we refer to the different realizations of preferences as *types*). A firm knows its own realized type, but the other agents do

[5] We refer to institutions (firms/schools) as "they/them", to individuals (workers/students) as “she/he”.

not. In this example, firm A has two possible types denoted by a superscript A^1, A^2. Firm A^1 considers only worker I and II as potential employees, and has preferences of $I \succ II$. When being of type A^2, it only considers workers III and I, and prefers $III \succ I$. The probability of firm A being of type A^1 is p. Firm B only wants to be matched with worker I and firm C only wants to be matched with worker III. We summarize the information on the matching market in Example 1. In addition, we assume that worker I has a higher expected utility of being matched with firm B than taking the lottery of being matched with firm A without knowing the type of firm A ($I : u(B) > p \cdot u(A_1) + (1 - p) \cdot u(A_2)$).[6]

Example 1

Proposer / Firm		Receiver / Worker
$A^1 : I \succ II$	with (p)	$I : A_1 \succ B \succ A_2$
$A^2 : III \succ I$	with $(1 - p)$	$II : A$
$B : I$		$III : C$
$C : III$		

Given:

$I : u(B) > p \cdot u(A_1) + (1 - p) \cdot u(A_2)$

Given their knowledge about the matching market and the mechanism, workers can infer the type of a firm after observing the final matching. For example, if firm A is matched with worker II, then agents can infer that firm A is of type A^1 because type A^2 does not consider worker II as a relevant candidate.

To illustrate the main intuitions, we first derive the optimal strategy of worker I in the DA mechanism, and then show that the outcome is unstable. Except for worker I, all agents in the matching market have standard preferences and will state these truthfully.[7] Given that all agents except worker I submit true preferences to the mechanism, both types of firm A, as well as firm B, will always make an offer to worker I during the DA mechanism (firm A^2 will always be rejected by worker III, and will therefore make an offer to worker I). Worker I states her preferences based on her expected utility. Assuming that her utility of being matched with firm B is higher than the lottery of being matched with types A^1 or A^2, then she states B ($\succ A$).

[6] We sometimes only denote the preferences that a firm or worker has over being unmatched (e.g., the preferences $A : I \succ II \succ A \succ III$ can be denoted as $A : I \succ II$).

[7] We will show later that it is a weakly dominant strategy for proposers to state true preferences in a DA, even if receivers have reciprocal preferences. Because workers II and III only consider working for one firm, it is also a weakly dominant strategy for them to state their true preferences.

Given that worker I states $B \succ A$, the type of firm A will be revealed through the final matching. If type A^1 is realized, then worker I is matched with firm B, and firm A is matched with worker II. Through observing the match of worker II and firm A, worker I can infer that firm A is of type A^1. Therefore, worker I wants to be matched with firm A. Given that firm A and worker I want to be matched to each other mutually, the matching is unstable. This happens because information about the type of a firm is revealed through the mechanism and the resulting final matching. We call this notion *immediate stability.*

3 Model

Overview. We consider a market with workers on one side of the market, firms on the other side, and a mechanism to match them one-to-one. Our matching setup differs from the standard model in Gale and Shapley (1962) in two aspects. First, the realization of firms' types and preferences is private information. Second, workers have reciprocal preferences, and therefore they prefer partners who also like to be matched with them. Agents may care about the fundamental preferences of others according to belief-based and preference-based motives. Being a preferred candidate can be a signal about the match-specific value that the firm may be better informed about (Avery & Levin, 2010; Lee & Niederle, 2015). Second, workers may enjoy interacting with a firm that likes them (Montoya & Insko, 2008; Montoya & Horton, 2012), even if preferences do not signal differences in relationship productivity. Hence, we allow for more general preference profiles than in a standard setting.

Set-up. Two disjoint finite sets of agents are one-to-one matched, following the marriage model of Gale and Shapley (1962). We consider firms $F = \{A, B, C, ...\}$ and workers $W = \{I, II, III, ...\}$. We sometimes denote an arbitrary firm with f and an arbitrary worker with w. Firms have strict and complete preferences over workers. Firm f's preferences are represented by an ordered list $P(f)$ on the set of $W \cup \{f\}$. A firm's preferences are private knowledge of the firm. Equivalently, we may think about the type (instead of the preferences) of firms being private information. The finite set of all possible types for a given firm is given by T_i. We denote the type of a firm by a superscript (e.g., $T_A = \{A^1, A^2, A^3, ...\}$). The type of a firm f is independently drawn from a distribution g_f known to each agent in the market. Workers have reciprocal preferences over firms. The preference of a worker w is represented by an ordered list $P(w)$ on the set of all possible types of all firms $T_A \cup T_B \cup T_C \cup ... \cup \{w\}$. For example, a firm's preferences might be $P(f) = II, I, f, III$. A worker's preference might be $P(w) = A^1, B^1, A^2, w, B^2$ denoting that a worker w prefers being matched with firm type A^1 over firm type B^1 over firm type A^2 over being unmatched f and over being matched with firm type B^2. $\boldsymbol{P} = \{P(A), P(B), ..., P(I), P(II)...\}$ denotes the set of all preferences. We also use the notation $i \succ j$ to state that i is preferred over j, both for firms and workers.

We examine the idea that a worker cares about being ranked well by the other market side. The true rank of worker w in the preference list of a firm is denoted by τ. Formally, we impose that for any firm f, a worker prefers $f_\tau \succeq f_{\tau+1}$. For example, worker I weakly prefers to be the first choice of A than the second choice of A. By defining the match utility of a worker w as $u_w(f_\tau)$, we generalize that if $i < j$ ($i, j \in \mathbb{N}$), then $u_w(f_i) \geq u_w(f_j)$. Being ranked better by firm A weakly increases the utility from being matched with firm A. This implies that we rule out cases in which workers like to be an undesirable alternative by a potential partner.[8]

In a matching market, each agent faces the decision about what preferences order $Q(f)$ (or $Q(w)$) to state given a mechanism h. $\boldsymbol{Q} = \{Q(A), Q(B), ..., Q(I), Q(II)...\}$ is the set of all stated preferences by firms and workers. Following our theoretical set-up, a matching market is described by the agents in the matching market, their possible types and probabilities of realisation for every type, and their preference profile (including reciprocal preferences).

A matching mechanism h takes the stated preference profiles $\boldsymbol{Q}$ and then maps them into a matching μ. A matching μ is a one-to-one correspondence. After the matching mechanism takes place, every firm f of the market is either matched with a worker denoted by $\mu(f) = w$ or is matched with itself $\mu(f) = f$. The same applies to workers.

Stability. In a setting of incomplete information, the stability criterion has to specify under which circumstances an agent would like to rematch. We use a standard framework of expected utility to define stability. An agent wants to rematch if their expected utility of rematching is higher than their expected utility of staying with their current match. Utility is non-transferable, and agents rematch if their (expected) utility from rematching is larger than their (expected) utility from their current matching. Defining stability in terms of expected utility corresponds to the concept of *Bayesian stability* in Bikhchandani (2017).[9]

[8] This may neglect potential behavioral mechanisms where a worse rank leads to higher desirability. For example, a worse rank may increase the desire to work with this party to convince it about one's quality. Alternatively, a better rank may be a signal of the worse quality of the other party [e.g., "*I don't care to belong to any club that will have me as a member.*" (Groucho Marx)]. Although theoretically possible, we consider these mechanisms to be secondary to a preference for a partner who likes to be matched with one.

[9] Other papers consider different stability concepts; for example, the idea that a matching is not stable if there is a positive probability to profit from rematching (Lazarova & Dimitrov, 2017), when there is no probability that the rematching leads to a worse outcome (ex-ante stability in Bikhchandani, 2017), or whenever there is scope for a mutually beneficial outcome given that transfer payments are possible (Liu et al., 2014).

Definition 1. *Bayesian stability: A matching μ is Bayesian blocked by any worker-firm pair (w, f) that are not matched, but the expected utility of worker w and the utility of firm f increase by matching with each other. A matching μ is Bayesian blocked by an agent (any firm f or worker w) if the agent prefers to be unmatched to their current match in (expected) utility. A mechanism is Bayesian stable if at least one of its equilibria is not blocked by any individual or any pair of agents for every realization of all possible type realizations T_i.*

Our main stability concept evaluates *Bayesian stability* directly after the mechanism determines the matching, and the matching becomes public, which we call *immediate stability.* Once the matches are public, Bayesian updating about the types of other agents is possible. If this updating process leads to a blocking individual or blocking pair, then the matching is not *immediate stable.* Agents only infer information about types and preferences from the observed matching, which is in contrast to (for example) Roth (1989) where preferences become public knowledge.

Definition 2. *Immediate stability: A matching is immediate stable if it is Bayesian stable after the outcome of the matching mechanism is known to the agents.*

This definition of stability allows us to characterize matching mechanisms that lead to stable outcomes. A mechanism guarantees stability if at least one of its equilibria is always Bayesian stable. This implies that the mechanism leads to stable outcomes for every realization of the firms' types. Hence, a mechanism is considered to be stable when every possible outcome of the given equilibrium is stable. It is important to note that we implicitly define stability together with the underlying mechanism because of potentially different degrees of information revelation. While specific outcomes may be stable with one mechanism, they are not stable with another.[10]

For our supplementary stability concept, we assume that a player learns the type of their matched agent after the matching. For example, the type of the partner could be revealed through interaction. Therefore, this constitutes another natural point to evaluate stability. If an agent wants to rematch after seeing the final matching and learning the true type of their partner, then the matching is not *ex-post stable.* Hence, *immediate stability* and *ex-post stability* are evaluated at different points in time (see Figure 1) but rely on the same stability concept (*Bayesian stability*). At both points in time, workers update their beliefs about the firms' types.

[10] For illustration, let us assume that a mechanism randomly pairs workers and firms without taking the stated preferences into account. This mechanism may lead to the same outcome as a DA mechanism. In the former cases, workers cannot infer something from the outcome about firms' true preferences, while they can in a DA mechanism.

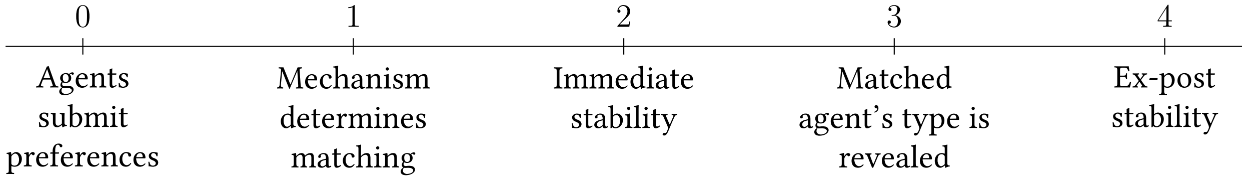
0
1
2
3
4
Agents submit preferences
Mechanism determines matching
Immediate stability
Matched agent's type is revealed
Ex-post stability

Figure 1: Timeline

Definition 3. *Ex-post stability: A matching is ex-post stable if it is Bayesian stable after the outcome of the matching mechanism is known to the agents and workers learn the true type of the firm that they are matched with.*

There is a connection between both stability notions in that ex-post stability implies immediate stability. The proof of Proposition 1 is delegated to the Appendix (see Section A.1). The intuition is that an immediate outcome can result in one or more ex-post outcomes. If all of the ex-post outcomes are stable, then the immediate outcome must also be stable because a mechanism is stable if at least one of its equilibria is always stable with respect to the true preferences. Meanwhile, immediate stability does not guarantee ex-post stability.[11] Imagine that there is one ex-post unstable outcome that only materializes with a small probability. In this situation, it can still be the case that the outcome is immediate stable. Therefore, our main results focus on the more restrictive concept of *immediate instability*, any results related to *ex-post stability* are deferred to the Appendix.

Proposition 1. *In a matching market with reciprocal preferences, ex-post stability is a sufficient condition for immediate stability.*

Strategy-proofness. The concept of strategy-proofness generally carries over from the complete information setting. A mechanism is strategy-proof if it is optimal to report preferences truthfully for every agent and for all strategies of other agents. In our setting, strategy-proofness for a standard mechanism such as the DA mechanism is only defined for agents without reciprocal preferences because workers with reciprocal preferences cannot state their full preferences profile.

Definition 4. *Strategy-proofness: A mechanism is strategy-proof if truthful preference revelation (or simply truth-telling) is a weakly dominant strategy for every agent.*

[11] The same correspondence holds between the concept of ex-post stability and the complete information stability concept of Roth (1989), where all agents' types become public knowledge. In a matching market with reciprocal preferences, stability according to Roth (1989) is a sufficient condition for ex-post stability.

4 Two-Sided Matching Markets: Deferred Acceptance Mechanism

The DA mechanism (Gale & Shapley, 1962) plays a key role in two-sided markets because it achieves stable matching outcomes in a standard complete information setting. Agents on both sides of the market (proposers and receivers) submit a rank-ordered preference list about the acceptable agents on the other side of the market to the mechanism. In the DA, stability is compatible with truth-telling for the proposing side of the mechanism. Under complete information, the equilibrium misrepresentation of preferences by the receiving side still results in stable outcomes (Roth, 1984). Given that there is no stable and strategy-proof mechanism for both sides of the market (Roth, 1982), the DA mechanism comes close to achieving the optimal outcome.

Due to these desirable properties, we first analyze reciprocal preferences in the standard DA mechanism before generalizing our main findings to a broader class of mechanisms. We start with an analysis where workers are on the receiving side of the DA mechanism. We then extend the analysis to the case where workers are on the proposing side. Both sides of the market must submit a standard preference list to the mechanism in which each acceptable agent of the other market side appears only once. Therefore, also an agent with reciprocal preferences must submit a standard preference list. A worker with reciprocal preferences $A_1 \succ B \succ A_2$ has to decide whether to state $A \succ B$ or $B \succ A$, independent on whether they are on the proposing or receiving side of the mechanism.

In our setting, truth-telling remains a weakly dominant strategy for firms on the proposing side of the DA mechanism. This follows immediately from the results in the standard setting of the DA mechanism. While the setting with reciprocal preferences may change the strategy of other agents, the mechanism remains the same. By the definition of a weakly dominant strategy, the potential adjustments of other players will not change the optimality of stating true preferences for a firm.[12]

Proposition 2. *Truth-telling is a weakly dominant strategy for firms with standard preferences in a one-to-one-matching under the DA mechanism.*

Proof. Suppose not. Then there must be another strategy of the firm which does better for at least

[12] Following standard assumptions, we exclude that agents can break up their match assigned by the DA mechanism and rematch with other agents. In Appendix A.2.1, we relax this assumption and assume that agents anticipate the possibility to rematch. This leads to strategic misrepresentation of preferences by proposers in the DA mechanism.

one set of played strategies by the other players. But this contradicts the fact that, given any set of strategies by the other players, truth-telling is a weakly dominant strategy in a DA. □

Meanwhile, workers on the receiving side may have an incentive to misrepresent their preferences. This directly follows from the fact that our framework encompasses the standard case, in which truth-telling is not a weakly dominant strategy for receivers.

In the next step, we analyze the effect of reciprocal preferences on stability. As shown through Example 1, the DA mechanism ceases to be immediate stable if agents on the receiving side have reciprocal preferences. We additionally show that there are matching markets where no strategy for a worker leads to an immediate stable matching. This means that a receiver with reciprocal preferences cannot choose any strategy that guarantees stability, even though all of the other agents state true preferences. In the case of standard preferences, an outcome is always stable if every agent states their true preferences (Gale & Shapley, 1962). A similar result cannot be established in the case with reciprocal preferences.

We demonstrate that worker I cannot choose any strategy that guarantees an immediate stable outcome by analyzing the same market as in Example 1. Depending on the stated preferences of worker I, worker I can be matched with firm A, firm B or remain unmatched. Every preference submission by worker I has to result in one of these outcomes. We show that each of these three outcomes can be immediate unstable. We have already shown that being matched with firm B (e.g., by stating $I : B$ or $I : B \succ A$) is immediate unstable because worker I infers the type of firm A after seeing the final matching. Worker I wants to be matched with firm A after learning that the firm is of type A^1. If worker I is matched with firm A (e.g., by stating $A \succ B$ ($\succ I$) or by just stating A ($\succ I$)), then the matching will always be immediate unstable. Worker I does not learn the type of A and forms a blocking pair with firm B because she prefers being matched with firm B over firm A if she does not know the type of firm A ($I : u(B) > p \cdot u(A_1) + (1-p) \cdot u(A_2)$). In the last case where worker I remains unmatched (by not stating a firm), she forms a blocking pair with either of the two firms. This shows that worker I cannot submit any preference profile that guarantees stability due to the information revelation by seeing the final matching.[13]

Proposition 3. *The DA mechanism with reciprocal preferences is not always immediate stable. There are markets where there is no strategy for a worker that leads to an immediate stable outcome if all the other players state true preferences.*

Instability not only arises with strategic play, but is a direct consequence of the updating about

[13] In Appendix A.2.2, we generalize this result to the concept of ex-post stability.

others' preferences. Therefore, reciprocal preferences can even rationalize instability in strategy-proof mechanisms without relying on untruthful preference submission through behavioral biases and misperceptions of the mechanism.

Instability also arises when the agents with reciprocal preferences are on the proposing side and the agents with types are on the receiving side of the DA mechanism. Our main findings of instability due to reciprocal preferences apply irrespective of the market side that faces uncertainty. This implies that information about which side of the market has reciprocal preferences cannot solve the problem from the designer's perspective. Our stability results do not rely on the uninformed market side being on the receiving side of the DA mechanism (e.g., compared to Fernandez, Rudov, & Yariv, 2022).

The following Example 2 shows that the DA mechanism is neither immediate nor ex-post stable with reciprocal preferences on the proposing and uncertainty on the receiving side.

Example 2

Proposer / Worker	Receiver / Firm	
$I \quad : A_1 \succ B \succ A_2$	$A^1 : I$	with (p)
$II \quad : II$	$A^2 : II \succ I \succ III$	with $(1-p)$
$III : A$	$B \ : I$	

Given:

$I : u(B) > p \cdot u(A_1) + (1-p) \cdot u(A_2)$

Worker I has to decide about stating $A \succ B$ or $B \succ A$. If she states $A \succ B$, then she will be matched with firm A; and if she states $B \succ A$, then she will be matched with firm B. We assume that worker I prefers to be matched with firm B rather than being matched with firm A without knowing the type $[u(B) > p \cdot u(A_1) + (1-p) \cdot u(A_2)]$. Hence, she states $B \succ A$ and is matched with firm B. However, worker I will infer the type of firm A through seeing the final matching. If firm A is of type A^2, then it is matched with worker III, while firm A remains unmatched if it is of type A^1. In the case of firm A^1, worker I and firm A want to rematch. The matching is immediate and ex-post unstable.

Regarding agents' incentives, we show that proposers do not have a weakly dominant strategy in this setting. While in a standard setting of uncertainty, truthful preference revelation is a weakly dominant strategy for the proposing side, there is not one in the case of reciprocal preferences. The optimal strategy of a proposer with reciprocal preferences depends on the behavior of other market participants. We show this in a simple example (Appendix A.2.3). Therefore, strategic considerations may even play a role for the proposing side in an environment where proposers have reciprocal preferences.

5 Stability in Two-Sided Matching Markets

This section analyzes whether an alternative mechanism can remedy the observed instability under the DA mechanism. We generalize the findings and obtain an impossibility result: there is no mechanism that is immediate or ex-post stable. We first demonstrate that no mechanism always achieves immediate stability. For this, it is sufficient to show that a matching market exists for which no mechanism can achieve immediate stability.

Proposition 4. *There is no mechanism that is always immediate stable for every matching market with reciprocal preferences.*

We demonstrate this by showing one matching market where no mechanism can achieve immediate stability.

Proof. Let h be a matching mechanism. The matching mechanism h selects a matching μ for the stated preference profiles $\boldsymbol{Q}$. A matching μ is a one-to-one correspondence and denotes with whom agents are matched. We show that no mechanism h can select an immediate stable matching according to the true preferences in the following matching market.

The market consists of firm A and three workers (I, II, III). Firm A can have two different types (A^1, A^2). Worker I has reciprocal preferences, while workers II and III have standard preferences. The preference profiles P are defined in Example 3. Worker I prefers to be unmatched rather than being matched with firm A given their prior about firm A's type in this example $(I : u(I) > p \cdot u(A_1) + (1-p) \cdot u(A_2))$.

Example 3

Firm		Worker
$A^1 : I \succ II \succ A \succ III$	with (p)	$I \quad : A_1 \succ I \succ A_2$
$A^2 : III \succ I \succ A \succ II$	with $(1-p)$	$II \quad : A \succ II$
		$III : III$

Given:

$I : u(I) > p \cdot u(A_1) + (1-p) \cdot u(A_2)$

The proof shows that no mechanism can select a match for both types A^1 and A^2 of firm A that is always immediate stable. Every possible matching is either immediate unstable because the firm's type is revealed and a blocking pair emerges or because one of the firm types has an incentive to mimic the other, which leads to immediate instability.

We proceed in three steps to establish that no mechanism is immediately stable. First, we exclude from the set of potentially stable matchings those that are immediate unstable without any information update and given worker I's prior about the firm's type. Second, we narrow down the set of immediate stable matchings by eliminating all matchings where the information updating about the firm's type through seeing the final matching leads to instability. Third, we show that the remaining set of possibly stable matching cannot be reached due to strategic play of the firm and a resulting pooling equilibrium that is always immediate unstable.

Step 1: To reduce the set of potentially stable matchings, we first exclude all matchings that are unstable given the preferences of agents and the workers' priors about firm A's type. A matching can only be immediate stable if firm A^1 is matched with worker I or II, and type A^2 is matched with worker I or remains unmatched. A matching is immediate unstable in all the other cases. Worker III prefers to be unmatched regardless of firm A's type, while firm A prefers to be unmatched if it is of type A^2 and gets matched to worker II and firm A^1 prefers to be matched with worker II if being unmatched.

Step 2: We further narrow down the set of immediate stable matchings by considering the information updating of workers through seeing the final matching. We start on the basis of the remaining possible, stable matchings after Step 1 and show that firm A^1 cannot be matched with worker II. If there is a positive probability that both type's of firm A are matched to worker II, then the matching is immediate unstable because firm A^2 prefers being unmatched over being matched with firm A^2. Hence, if the mechanism is immediate stable it can only match firm A^1 with worker II. However, if firm A^1 is matched with worker II, worker I can infer the type of firm A^1. The matching is immediate unstable because firm A^1 and worker I form a blocking pair. Therefore, given a immediate stable mechanism, firm A^1 cannot be matched with worker II and can only be matched with worker I.

Step 3: We show that every remaining possible, stable matching cannot be reached due to the strategic play of firm type A^2. For every matching that has a positive probability of firm A^2 being unmatched and firm A^1 being matched with worker I for sure, type A^2 has an incentive to mimic type A^1. This would result in the matching $\mu(A^1) = I$ and $\mu(A^2) = I$, which is not immediate stable by assumption because worker I prefers to be unmatched compared to being matched with firm A without knowing the type of firm A ($I : u(I) > p \cdot u(A_1) + (1 - p) \cdot u(A_2)$). This implies that every possible matching is immediate unstable. □

To demonstrate that there is no mechanism that always leads to an ex-post stable matching, we build on Proposition 1.

Proposition 5. *In a setting with reciprocal preferences, no mechanism always leads to an ex-post stable matching.*

Proof. Proposition 1 states that ex-post stability is a sufficient condition for immediate stability in a matching market with reciprocal preferences. We show in Proposition 4 that there is no immediate stable mechanism. By law of contraposition, a matching cannot be ex-post stable if it is immediate unstable. Therefore, we can conclude that there is no ex-post stable mechanism. □

6 School Choice Markets

In the school choice setting, only one side of the market consists of strategic agents (Abdulkadiroğlu & Sönmez, 2003). The other side has priorities that are not subject to misrepresentations. This framework especially applies to public institutions (e.g., schools), which have priorities over other agents (e.g., students). These priorities must be truthfully reported to the mechanism. For example, a school may be required by law to prioritize its applicants according to specific rules (e.g., scores in entrance exams or distance to the students' residence). Understanding the effects of reciprocal preferences in school choice is crucial in its own right but it can also provide insights for effective mechanism design when comparing outcomes with those of standard two-sided markets. We will show that excluding strategic considerations for one side of the market, combined with choosing the right mechanism, resolves the instability associated with reciprocal preferences.

Like the standard two-sided market, reciprocal preferences only affect stability in a school choice setting when there is uncertainty about the priorities. One may think about (at least) three different reasons for uncertainty in the school choice setting. First, the rules on how priorities are formed might not be publicly available.Second, the rules are publicly available, but costly to find and process. Third, even when rules are common knowledge, an applicant lacks information about the characteristics of other applicants. Therefore, she still faces uncertainty about the final priorities.

Under uncertainty about the priorities, the DA mechanism results in unstable outcomes. The uncertainty about schools' priorities can still result in ex-post suboptimal decisions of the applicants. The intuition is that applicants have to make a choice under uncertainty about the true priorities because these are not revealed ex-ante. Given that individuals care about their true ranking in the preference lists, this can result in unstable matchings. We show this through Example 4, where applicant I has to decide between stating $A \succ B$ or $B \succ A$. Given that uncertainty applicant I prefers being matched with school B rather than being matched with school A given her prior about the schools' type [$I : u(B) > p \cdot u(A_1) + (1-p) \cdot u(A_2)$], she states $B \succ A$. Applicant I learns whether school A is of type A_1 or A_2 after seeing the final matching. If firm A is unmatched worker I knows that it is of type A^1 and they build a blocking pair. When school A is of type A^1, the outcome is immediate and ex-post unstable.[14]

Example 4

Proposer / School		Receiver / Applicant
$A^1 : I$	with (p)	$I \quad : A_1 \succ B \succ A_2$
$A^2 : II \succ I \succ III$	with $(1-p)$	$II \quad : II$
$B : I$		$III : A$

Given:

$I : u(B) > p \cdot u(A_1) + (1-p) \cdot u(A_2)$

We present two simple remedies to overcome this inefficiency: first by inducing information revelation before applicants submit their preference list, and second by incorporating reciprocal preferences into the ranking. We consider a simple sequential variant of the DA mechanism to induce information revelation. In this Two-Stage Deferred Acceptance (TSDA) mechanism, the market side with uncertain preferences first submits their rankings publicly to the mechanism. The side with reciprocal preferences then submits their preference ranking, incorporating the information from the first step. With these preferences, we run a standard DA mechanism. Therefore, the only difference between a standard DA and the TSDA mechanism is the preference submission's timing and observability.

The TSDA mechanism solves the information asymmetries without any caveats in the school choice problem. Institutions state their priorities truthfully, which results in a situation of complete information. Reciprocal preferences can affect the preference order but every applicant can rank schools unambiguously by incorporating the information about the schools' preferences. Truth-telling then means submitting those preferences. If applicants are on the receiving side of

[14] In Appendix A.2.4, we show that this does not depend on whether the applicants are on the receiving or proposing side of the algorithm.

the algorithm, they might still misrepresent their preferences in equilibrium as in the standard DA mechanism but the outcome is stable (see Roth, 1984). If applicants are on the proposing side, then they do not have any incentive to misrepresent their preferences. Hence, a TSDA mechanism with the applicants as the proposers leads to truth-telling and stability.

A variant of the standard DA mechanism that allows individuals to submit their complete preferences order (even if they have reciprocal preferences) also leads to stability. In the standard DA, participants submit a rank-ordered list in which all of the acceptable partners are listed once. Instead, the preferences lists of our modified version of the DA mechanism can include reciprocal preferences. This means that agents can submit their full contingent preference lists to the mechanism (e.g., $I : A_1 \succ B \succ A_2$ instead of having to decide whether to rank A over B, or vice versa). We call this version the Reciprocal References Deferred Acceptance mechanism (RPDA).

The mechanism then uses the (simultaneously) stated preferences of all agents and creates a standard preference ranking for every agent. For applicant I who submitted the preference profile $A_1 \succ B \succ A_2$ to the mechanism, the mechanism assesses whether applicant I prefers school A over B based on the stated preferences of these schools. If school A stated applicant I as their highest preference, then the mechanism assigns the preference profile $A \succ B$ to agent I. If school A did not state applicant I as their highest preference, then the mechanism uses the ranking $B \succ A$. With these rank-ordered lists, a standard DA mechanism takes place. Implicitly, the mechanism takes the stated preferences as the true preferences. While the reciprocal preferences of agents are based on the true preferences, the mechanism determines the final rankings by using the stated preference information.

The RPDA mechanism is stable when applicants are on the proposing side of the algorithm (and schools with standard preferences on the receiving side). The proposing side has a weakly dominant strategy to state true preferences and the receiving side submits their (true) priorities.[15] Applicants do not have to take into account any strategic considerations when stating their complete preference profile. Like the TSDA mechanism, the RPDA mechanism is strategy-proof and stable when individuals are on the proposing side. However, when applicants are on the receiving side of the mechanism, uncertainty and preference misrepresentations can lead to

[15] Stating true preferences is a weakly dominant strategy for proposers in the standard DA mechanism (Roth, 1982). Therefore, an agent cannot do better than stating (reciprocal) preferences in the RPDA mechanism that correspond to these standard preferences. In the school choice setting, schools are non-strategic and reveal their priorities truthfully. Given that applicants state their true reciprocal preferences, the standard preference ranking produced by the RPDA mechanism precisely reflects the applicants' preferences under complete information. Therefore, stating true reciprocal preferences is a weakly dominant strategy in the RPDA mechanism.

instability. As in standard DA, the RPDA mechanism is not strategy-proof for the receiving side. By misrepresenting their preferences, receivers might be able to implement the receiver optimal stable matching (instead of the proposer optimal matching). Due to the uncertainty, these misrepresentation can result in immediate and ex-post unstable outcomes, even in the absence of reciprocal preferences.

Proposition 6. *The DA mechanism does not achieve stability in a school choice setting where applicants have reciprocal preferences. In contrast, a sequential variant of the DA (TSDA) mechanism and a variant in which agents can submit their reciprocal preferences to the mechanism (RPDA) can achieve stability.*

Therefore, the TSDA or the RPDA mechanisms may be perceived as solutions to situations with reciprocal preferences in standard two-sided matching markets with strategic players on both sides of the market. We demonstrate in Section 5 that there are no mechanisms that achieve *Bayesian stability* (both after agents observe the mechanism's outcome and once agents learn the type of their partner). This implies that neither the TSDA nor the RPDA mechanism can guarantee *Bayesian stability.*

In contrast to the DA mechanism, neither TSDA and RPDA mechanisms are strategy-proof if we consider a standard setting with strategic firms on the proposing side (Proof in Appendix A.2.5). The underlying idea of the TSDA and RPDA mechanism is that workers can react to the types and preferences of firms. This also implies that firms start to send favorable signals to applicants strategically. For example, despite preferring a very talented applicant who is extremely unlikely to join a (mediocre) firm in any case, this firm may use its top signal for an applicant whose decision could be influenced by it. One can even show that there are markets in which the DA mechanism achieves stable outcomes in undominated strategies that imply truthful reporting for firms, while the TSDA mechanism does not (see Appendix A.2.6).

7 Discussion and Conclusion

In this paper, we motivate, formalize, and analyze the effects of reciprocal preferences in matching mechanisms. Reciprocal preferences allow for the possibility that preferences of the other market side influence an agent's preferences. We provide three main results when agents care about others' preferences without perfectly knowing these. First, we show that the DA mechanism ceases to be stable when agents observe the final allocation of the mechanism. Even if agents do not strategically misrepresent their preferences, the final matching can be unstable. Second, we demonstrate that no mechanism always leads to a stable matching when (at least) one of the agents has reciprocal preferences. Third, when extending our analysis to a school choice setting, we show that variants of the DA mechanism achieve stability.

We assume one-sided reciprocal preferences throughout this paper. In principle, this can be extended to consider reciprocal preferences on both sides of the market. Once we do so, we need to define how the reciprocal preferences of worker I correspond to firm A's preferences if these are also reciprocal to avoid a recursive problem. If firm A's preference order is $I_1 \succ II_1 \succ I_2 \succ II_2$, then it is unclear whether worker I is ranked better or worse than worker II in the preference list of firm A (and hence how worker I's preferences should respond to these). One reasonable possibility to solve this problem is to assume that every reciprocal preference order must have a general structure, such that being ranked the same by every agent on the other market side induces the same ranking. Given the reciprocal preferences of firm A, this *general preference order* would be $I \succ II$, because the firm prefers $I \succ II$ when being ranked first by both workers, and when being ranked second by each worker. Once we assume that there exits such a *general preference order* and that reciprocal preferences are based on it, the model becomes tractable again.

Understanding the effects of reciprocal preferences is critical for understanding why certain matching markets do not perform satisfactorily. First, reciprocal preferences provide a rationale for why agents may prefer decentralized matching over centralized (algorithmic) matching mechanisms. Decentralized markets in which agents interact allow them to learn the other sides' preferences, which makes it hard to establish a centralized mechanism if they care about others' preferences. Second, modifications of the mechanism by participants can be a consequence of reciprocal preferences. Agents may add screening tools to attract especially interested candidates, such as early admission (Avery & Levin, 2010), or may even base their own ranking directly on other agents' preferences (Chiu & Weng, 2009; Dur et al., 2022). Third, the results help us to understand observed instability in centralized matching markets. Given that true preferences are largely unobservable, theory can help evaluate the reasons for instability and allow for a better design of matching mechanisms. Policy conclusions in response to observed instability in a mechanism depend crucially on whether the cause of instability is a strategic misrepresentation of preferences or are reciprocal preferences per se.

This paper informs market designers about what information should be disclosed. We show that information about others' preferences is crucial for stability. While the market designer has arguably little influence on the agents' learning preferences of others in individual conversations, they can guide information flows by revealing all final matches to the agents (or even the submitted preference rankings). This is not only crucial for the effect of reciprocal preferences to come into play but it also matters if agents have other non-standard preferences, such as regret that crucially depends on counterfactual outcomes.

Our findings on stability in school choice also inform policy where institutions are strategic agents and forcing them to act non-strategic might be impossible. In this case, one may oblige schools to determine preferences based on objective and transparent criteria. The need for schools to determine their ranking based on these criteria increases the credibility that the ranking corresponds to the actual preferences. Ranking applicants according to some pre-specified criteria where compliance can be (partially) verified by the market designer mitigates the scope of submitting non-truthful preferences and enhances stability. In that sense, designing rules for a matching market that make it more similar to a school choice setting can increase stability and welfare.

This paper highlights that standard matching mechanisms do not function as desired when agents care about others' preferences without perfectly knowing them. This is true not only for reciprocal preferences but also for more generally situations in which agents have type-dependent preferences, with the types themselves being defined by the preferences of the other agents. While this paper focuses on reciprocal preferences, investigating different classes of type-dependent preferences in matching markets remains for future research. Natural extensions of our current analysis include considering additional information structures about initial preferences, other information sets when evaluating stability, and different stability concepts under uncertainty. While we derive our (negative) results on stability through the theoretical analysis of markets, we have a limited understanding of how often instability actually occurs in markets when using standard mechanisms. Therefore, in complementary work (Opitz & Schwaiger, 2023), we show the relevance of reciprocal preferences for (in)stability through a laboratory experiment and validate our underlying theoretical assumption.

References

Abdulkadiroğlu, A., & Sönmez, T. (2003). School choice: A mechanism design approach. *American Economic Review*, *93*(3), 729–747.

Antler, Y. (2015). Two-sided matching with endogenous preferences. *American Economic Journal: Microeconomics*, *7*(3), 241–58.

Avery, C., & Levin, J. (2010). Early admissions at selective colleges. *American Economic Review*, *100*(5), 2125–2156.

Bikhchandani, S. (2017). Stability with one-sided incomplete information. *Journal of Economic Theory*, *168*, 372–399.

Chakraborty, A., Citanna, A., & Ostrovsky, M. (2010). Two-sided matching with interdependent values. *Journal of Economic Theory*, *145*(1), 85–105.

Chiu, Y. S., & Weng, W. (2009). Endogenous preferential treatment in centralized admissions. *RAND Journal of Economics*, *40*(2), 258–282.

Dreyfuss, B., Heffetz, O., & Rabin, M. (2022). Expectations-based loss aversion may help explain seemingly dominated choices in strategy-proof mechanisms. *American Economic Journal: Microeconomics*, *14*(4), 515–55.

Dur, U., Pathak, P. A., Song, F., & Sönmez, T. (2022). Deduction dilemmas: The Taiwan assignment mechanism. *American Economic Journal: Microeconomics*, *14*(1), 164–185.

Fernandez, M. (2020). Deferred acceptance and regret-free truth-telling. *Working Paper.*

Fernandez, M., Rudov, K., & Yariv, L. (2022). Centralized matching with incomplete information. *American Economic Review: Insights*, *4*(1), 18–33.

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

Grenet, J., He, Y., & Kübler, D. (2022). Preference discovery in university admissions: The case for dynamic multioffer mechanisms. *Journal of Political Economy*, *130*(6), 1427–1476.

Gundlach, J. (2021). *Per Algorithmus zum Kitaplatz? Potenziale und Erfolgsfaktoren für eine bessere Kitaplatzvergabe mithilfe von algorithmischen Systemen.* Bertelsmann Stiftung.

Lazarova, E., & Dimitrov, D. (2017). Paths to stability in two-sided matching under uncertainty. *International Journal of Game Theory*, *46*(1), 29–49.

Lee, S., & Niederle, M. (2015). Propose with a rose? Signaling in internet dating markets. *Experimental Economics*, *18*(4), 731–755.

Liu, Q., Mailath, G. J., Postlewaite, A., & Samuelson, L. (2014). Stable matching with incomplete information. *Econometrica*, *82*(2), 541–587.

McKinney, C. N., Niederle, M., & Roth, A. E. (2005). The collapse of a medical labor clearinghouse (and why such failures are rare). *American Economic Review*, *95*(3), 878–889.

Meisner, V., & von Wangenheim, J. (2023). Loss aversion in strategy-proof school-choice mechanisms. *Journal of Economic Theory*, *207*, 105588.

Montoya, R. M., & Horton, R. S. (2012). The reciprocity of liking effect. In M. Paludi (Ed.), *The Psychology of Love* (pp. 39–57). Santa Barbara, CA: Praeger.

Montoya, R. M., & Insko, C. A. (2008). Toward a more complete understanding of the reciprocity of liking effect. *European Journal of Social Psychology*, *38*(3), 477–498.

Opitz, T., & Schwaiger, C. (2023). Everyone likes to be liked – Experimental evidence for reciprocal preferences in matching markets. *CRC TRR 190 Discussion Paper No. 366.*

Roth, A. E. (1982). The economics of matching: Stability and incentives. *Mathematics of Operations Research*, *7*(4), 617–628.

Roth, A. E. (1984). Misrepresentation and stability in the marriage problem. *Journal of Economic Theory*, *34*(2), 383–387.

Roth, A. E. (1989). Two-sided matching with incomplete information about others' preferences. *Games and Economic Behavior*, *1*(2), 191–209.

A Appendix

A.1 Proof of Proposition 1

In a matching market with reciprocal preferences, ex-post stability is a sufficient condition for immediate stability.

Proof. Assume a matching is immediate unstable but all the possible resulting outcomes are ex-post stable. A matching is immediate unstable if it is blocked by a pair or an individual after the mechanism took place, and every worker updates her belief about the types of the firms through seeing the final matching. When evaluating ex-post stability, workers learn the true type of their matched firm. Therefore, a change in stability between immediate and ex-post stability can only be attributed to workers receiving information about their matched partner's type. There is no information update for firms.

If a matching is immediate unstable because it is blocked by a pair, it must be the case that a worker (e.g. I) prefers another firm (e.g. B) over her matched firm (e.g. A). This means that the expected utility of being matched with firm B is higher than the expected utility of being matched with firm A. At the same time, if the matching is ex-post stable, she prefers to stay with her matched firm (A) over the other firm (B) for every possible realization of firm A's type. It cannot be the case that being matched with any type of firm A is better than the expected utility of being matched with firm B (ex-post stability), but the expected utility of being matched with firm A is lower than the expected utility of being matched with firm B (immediate instability).

The same logic applies for a worker that prefers being unmatched over her current match in terms of immediate stability. For a worker who is unmatched after the mechanism takes place, the information set is the same when evaluating immediate and ex-post stability. Hence, the matching cannot be immediate stable and ex-post unstable. □

A.2 Proofs of Additional Statements

A.2.1 Statement 1: Truth-telling is not a weakly dominant strategy for firms with an anticipated rematching stage in a DA mechanism with reciprocal preferences.

:

Example A.1

Proposer / Firm		Receiver / Worker
$A^1 : I \succ II$	with (p)	$I \quad : A_1 \succ I \succ A_2$
$A^2 : III \succ I$	with $(1-p)$	$II \quad : B \succ A \succ C$
$B \ : III \succ II$		$III : C \succ A \succ B$
$C \ : II \succ III$		

Given:

$I : u(A_1) \cdot p + u(A_2) \cdot (1-p) > u(I)$

Truth-telling is a weakly dominant strategy for proposers in a DA if receivers have reciprocal preferences (see Proposition 2). Example A.1 shows that this is not the case if players can rematch after the matching and anticipate this. We show that a firm can improve its expected utility by deviating from truth-telling.

Proof. Truth-telling is a weakly dominant strategy if the outcome is never worse than any other strategy, given every strategy of all other players. We show that another strategy than truth-telling is better for firm A^2 given a rematching stage and given strategies of the other players. Assume that all proposers and workers I and II state true preferences. For receiver I, we assume that she states $A \succ I$. Independent of firm A's type, worker I will always be matched with firm A, if both types of firm A state true preferences. However, the matching of the other firms and workers depends on the state preferemces of firm A. In the event of A^1, firm B is matched with worker III, and firm C is matched with worker II. If firm A is of type A^2, firm B is matched with worker II and firm C is matched with worker III. After observing the matching, worker I infers the type of firm A and breaks up the matching if she knows that she is matched with firm A^2. For this reason, firm A always states $I \succ II$. Worker I cannot tell which type she is matched with and will not break up the match with firm A. Hence, firm A^2 is mimicing the strategy of type A^1 and not stating true prefernces. Because we are checking for a weakly dominant strategy, we do not have to check whether the played strategies are an equilibrium. □

A.2.2 Statement 2: There are matching markets that are always ex-post unstable for any possible strategy a player can choose in the DA mechanism if all the other players state true preferences.

Example A.2

Proposer / Firm		Receiver / Worker
$A^1 : I \succ II$	with (p)	$I \; : A_1 \succ I \succ A_2$
$A^2 : II \succ I$	with $(1-p)$	$II : II$

Given:

$I : u(I) < p \cdot u(A_1) + (1-p) \cdot u(A_2)$

Proof. The DA is not ex-post stable in our setting with reciprocal preferences. Example A.2 shows that no set of strategies always results in an ex-post stable matching. Firm A has two possible types, A^1 and A^2 and worker I has reciprocal preferences and would like to be matched when firm A ranks her first but wishes to be unmatched if the firm is of type A^2. We assume that the expected utility from being matched with firm A without knowing its type is higher for worker I than being unmatched $[p \cdot u(A_1) + (1-p) \cdot u(A_2) > u(I)]$. We show that both possibilities to submit preferences ($A \succ I$ or $I \succ A$) lead to ex-post unstable outcomes. If worker I decides to match with firm A (by stating $A \succ I$), the match is immediate stable. However, once she learns the type of firm A and it turns out to be of type A^2, she prefers to be unmatched, and the matching is unstable. If worker I decides to remain unmatched initially (by stating $I \succ A$), she does not learn the type of firm A. However, the expected utility of being matched with firm A without knowing its type is higher than remaining unmatched. Firm A and worker I want to match, the matching is immediate and ex-post unstable. □

A.2.3 Statement 3: Proposers with reciprocal preferences do not have a weakly dominant strategy in the DA mechanism.

Example A.3 shows that proposer I with reciprocal preferences does not have a weakly dominant strategy for preference reporting when the DA is applied.

Example A.3

Proposer / Worker	Receiver / Firm	
$I \ : A_1 \succ B \succ A_2$	$A^1 : I$	with (p)
$II : A$	$A^2 : II \succ I$	with $(1-p)$
	$B \ : I$	

Given:

$I : u(B) > p \cdot u(A_1) + (1-p) \cdot u(A_2)$

Proof. Given that proposer I prefers to be matched with firm B over being matched with firm A without knowing firm A's type [$I : u(B) > p \cdot u(A_1) + (1-p) \cdot u(A_2)$], the optimal strategy of applicant I depends on the strategy of other players. Here, the decision to state $A \succ B$ or $B \succ A$ depends on proposer II. If proposer II states her true preferences, proposer I states $A \succ B$ because she does not incur the risk be matched with A^2. However, if applicant II states $II : II$, it is optimal for applicant I to state $I : B \succ A$ because she is matched with type A^2 with probability $(1-p)$ if when stating $I : A \succ B$. This proves that there is no dominant strategy for applicant I. □

A.2.4 Statement 4: The DA mechanism is not immediate and ex-post stable in the school choice setting, regardless of whether applicants are on the receiving or proposing side.

In Section 6, we show that the DA is neither immediate nor ex-post stable in a school choice setting, with applicants on the receiving side. Example A.4 shows that this also holds when applicants are on the proposing side.

Example A.4

Proposer / Applicant	Receiver / School	
$I \quad : A_1 \succ B \succ A_2$	$A^1 : I$	with (p)
$II \quad : II$	$A^2 : II \succ I \succ III$	with $(1-p)$
$III : A$	$B \ : I$	

Given:

$I : u(B) > p \cdot u(A_1) + (1-p) \cdot u(A_2)$

Proof. Applicant I decides whether to state $A \succ B$ or $B \succ A$ and prefers being matched with school B over being matched with school A without an update about its type. Accordingly, applicant I states $B \succ A$. However, applicant I will infer the type of school A after the matching. Only if school A is of type A^1, school A is unmatched, and the matching is immediate and ex-post unstable. □

A.2.5 Statement 5: With reciprocal preferences, neither the TSDA nor the RPDA is strategy-proof for firms in a standard two-sided matching market with firms on the proposing side.

We show with Example A.5 that neither the TSDA nor the RPDA is strategy-proof for firms in a standard two-sided matching market if firms are on the proposing side.

Example A.5

Proposer / Firm		Receiver / Worker
$A^1 : I \succ II$	with (p)	$I \ : A_1 \succ I \succ A_2$
$A^2 : II \succ I$	with $(1-p)$	$II : B \succ II$
$B \ : II \succ I$		

Given:

$I : p \cdot u(A_1) + (1-p) \cdot u(A_2) > u(I)$

Proof.

TSDA: In equilibrium, firm A^1 states true preferences. However, firm A^2 will mimic the strategy of firm A^1. Worker I cannot infer the type of firm A but will state firm A as her first choice due to the utility condition $[p \cdot u(A_1) + (1-p) \cdot u(A_2) > u(I)]$. Firm B and worker II state true preferences. No player has an incentive to deviate from their strategy. Stating true preferences for player A^2 is not an equilibrium because worker I can infer the type of firm A^2. Worker I would prefer to stay unmatched, resulting in no match for firm A^1.

RPDA: The same logic applies to the RPDA. Assume that both receivers state their true, complete reciprocal preferences. If both types of firm A state true preferences then type A^2 is unmatched. Therefore, type A^2 will misrepresent their preferences and state $I \succ II$. Firm A will always be matched with worker I.

Hence, both for the TSDA and RPDA truth-telling is no weakly dominant strategy. □

A.2.6 Statement 6: There are markets where the DA mechanism achieves stable outcomes in undominated strategies that imply truthful reporting for firms, while the TSDA does not.

We show one exemplary matching market (Example A.6) where the DA mechanism achieves stable outcomes in undominated strategies that imply truthful reporting for firms, while the TSDA does not.

Example A.6

Proposer / Firm		Receiver / Worker
$A^1 : I \succ II$	with (p)	$I\ : B \succ A$
$A^2 : II \succ I$	with (q)	$II : A \succ B$
$A^3 : I$	with $(1-p-q)$	
$B^1 : II \succ I$	with (w)	
$B^2 : I \succ II$	with $(1-w)$	

Given:

$$I : u(A) > \frac{p}{1-q} \cdot u(B) + \frac{1-p-q}{1-p} \cdot u(I)$$

$$II : u(B) > \frac{p}{1-q} \cdot u(A) + \frac{1-p-q}{1-p} \cdot u(II)$$

$$A^1 : u(II) < w \cdot u(II) + (1-w) \cdot u(A)$$

Proof.

DA: Stating true preferences is an equilibrium for both sides of the market. As shown before, truth-telling is a weakly dominant strategy for proposers in a DA mechanism. Due to the assumption on the utility of worker I and worker II truth-telling is also optimal for the workers [$I : u(A) > \frac{p}{1-q} \cdot u(B) + \frac{1-p-q}{1-p} \cdot u(I)$ and $II : u(B) > \frac{p}{1-q} \cdot u(A) + \frac{1-p-q}{1-p} \cdot u(II)$]. Receivers have no incentive to state truncated preferences because the expected utility of misrepresenting preferences is lower than the expected utility of stating true preferences. Given that all players state true preferences, the matching is stable.

TSDA: If firms state their true preferences, applicants can infer the firms' types and strategically misrepresent their preferences to ensure their preferred matching. Due to the sequential game the misrepresentation by the workers can be prevented if proposers truncate their preference list before the receivers do. Hence, there is a trade of for firms. If all firms state true preferences the receivers will misrepresent preferences in some cases and ensure their preferred matching. If firm A^1 misrepresents its preferences (truncating) it ensures a better match in some cases while it risk being unmatched in other cases. Given the assumptions on the utilities $[A^1 : u(II) < w \cdot u(II) + (1 - w) \cdot u(A)$, firm A^1 truncates his preferences if being of type A^1 while firm B does not. In the illustrated market, if firm A is of type A^1, it will truncate its list by only stating I as potential partner to prevent receivers from misrepresenting their preferences, who wish to do whenever confronted with types A^1 and B^1. However, it involves the risk of being unmatched in the event of B^2. In that case, the matching is immediate and ex-post unstable. □