School Choice and Loss Aversion

Vincent Meisner (TU Berlin)
Jonas von Wangenheim (FU Berlin)

Discussion Paper No. 208

December 5, 2019

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, University of Munich, 80539 Munich, Germany
+49 (89) 2180 3405 | info@rationality-and-competition.de

School choice and loss aversion*

Vincent Meisner† Jonas von Wangenheim‡

December 2, 2019

Abstract

Extensive evidence suggests that participants in the direct student-proposing deferred-acceptance mechanism (DSPDA) play dominated strategies. In particular, students with low priority tend to misrepresent their preferences for popular schools. To explain the observed data, we introduce expectation-based loss aversion into a school-choice setting and characterize choice-acclimating personal equilibria in DSPDA. Truthful equilibria can fail to exist, and DSPDA might implement unstable and more inefficient allocations in both small and large markets. Specifically, it discriminates against students who are more loss averse or less overconfident than their peers, and amplifies already existing (or perceived) discrimination. To level the playing field, we propose serial dictatorship mechanisms as a strategyproof and stable alternative that is robust to these biases.

JEL-Classification: C78 D78, D82, D81, D91.
Keywords: Market design, Matching, School choice, Reference-dependent preferences, Loss aversion, Deferred acceptance.

PRELIMINARY, COMMENTS WELCOME

*We thank Georgy Artemov, Inácio Bó, Rustamdjan Hakimov, Fabian Herweg, Peter Katuščák, Dorothea Kübler, Takeshi Murooka, Roland Strausz, Georg Weizsäcker, and seminar participants in Berlin and Munich, as well as at CED'19, ESEM'19, MIP'19 and VfS'19 for useful comments and suggestions. Financial support by Deutsche Forschungsgemeinschaft through CRC TRR 190 is gratefully acknowledged.

†Technical University Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany, vincent.meisner@tu-berlin.de.

‡Freie Universität Berlin, jonas.wangenheim@fu-berlin.de.

1 Introduction

The direct student-proposing deferred-acceptance mechanism (DSPDA) offers a celebrated solution to the problem of matching prospective students to schools. It is strategyproof, (constrained) efficient, and leads to the student-optimal stable allocation.[1] Consequently, this mechanism is implemented in many existing school choice programs.[2] In DSPDA, students can maximize the probability to get into their most preferred school without hurting their chances of admission to other schools. Unfortunately, growing evidence from both the field and the lab suggests that students with low priority tend to conceal preferences for popular schools and fake preferences for district schools despite the dominance of the truthful strategy. Hence, potentially none of the desired properties are obtained.

We identify expectation-based loss aversion (EBLA, Kőszegi and Rabin (2006, 2007)) as a possible explanation for this puzzle. In our framework, the preference report is a channel to manipulate expectations about the matching outcome and these beliefs become a stochastic reference point to which final match outcomes are compared. Because students with low priority are likely to be rejected by popular schools, not ranking such schools highly is not very costly in terms of the expected match utility. More importantly, reporting such a ranking mitigates disappointment and not even trying to get into these schools shields off disappointment completely. We characterize the rank-ordered lists (ROLs) that are rationalizable as a choice-acclimating equilibrium (CPE) in DSPDA, and provide testable predictions. This theoretical foundation of commonly observed deviations is the first contribution of this paper.

As a second contribution, we show that these misrepresentations have negative impact on stability and efficiency in equilibrium and, importantly, these effects do not vanish as markets grow large. We consider a setting in which heterogeneously loss-averse students compete for scarce seats at elite schools, and show that in choice-acclimating Bayesian Nash equilibria (CBNE) instable and inefficient allocations can emerge with non-negligible probabilities. More specifically, loss-averse students may abstain from applying to elite schools if they are pessimistic about their admission chances, while potentially weaker students are accepted just because their lower degrees of loss aversion or higher degrees of confidence lead to submitting a preference for elite schools. Thus, we contribute a novel argument to the active debate on whether such deviations matter.[3]

Third, we delineate how social segregation can arise if certain characteristics are

[1] Strategy-proofness is desirable, because reporting the true preferences dominates misreporting them, no matter what other participants report. Hence, the cost of strategizing is eliminated such that less sophisticated players are not given a disadvantage. Stable allocations are considered fair because no student envies another student that is considered worse by her school.

[2] For instance, Pathak and Sönmez (2013) provide many examples.

[3] This depends very much on the origin of these deviations. For instance, for standard preferences, Artemov et al. (2017) derive robust equilibria, allowing for mistakes whose impact on payoff vanishes as market size grows large. In contrast to our model, only a negligible fraction of these mistakes are payoff-relevant in their model.

correlated with demographics. Interestingly, reference-dependent preferences open the door for biased beliefs as an important determinant of optimal ROLs, although they play no role in the standard model with a dominant strategy. We establish that DSPDA favors students who are less loss-averse or more overconfident. Indeed, evidence suggests, for example, that overconfidence is more (Barber and Odean, 2001; Niederle and Vesterlund, 2007) and loss aversion less (Karle et al., 2019) pronounced among men compared to women. Moreover, DSPDA augments the disadvantage for students who are already (or perceive to be) marginalized when discrimination distorts priority scores, because EBLA incentivizes such students to shy away from ranking better schools in the first place.[4] In that sense, DSPDA does not "level the playing field" entirely, voiding one of the crucial advantages prominently named by Pathak and Sönmez (2008). Our model highlights a flaw in the empirical strategy to identify preferences reported to DSPDA as true. Regarding affirmative action policy, this insight is important because the observation that certain students do not apply to certain schools does not necessarily mean that they prefer other schools.

Finally, we not only point out weaknesses in DSPDA, but also investigate how alternative mechanisms might remedy them. Under a regularity condition on school preferences, we suggest sequential school-proposing deferred-acceptance as an alternative to foster truthful behavior on the student side. If school preferences are homogenous, this mechanism collapses to a very simple serial dictatorship mechanism. Crucially, a remedy mechanism necessarily has to be sequential as we show that no static mechanism can improve upon DSPDA. Letting students choose sequentially allows (i) to manipulate the informational environment by revealing previous students' choices, and (ii) to shrink the choice set of students selecting later and to incentivize reporting true preferences over this set.

In our model, students privately observe their match values for each school and also privately learn their individual degree of loss aversion. Moreover, they receive a signal about their relative priorities compared to the other students at each school. For some of our results, we consider a particular form of this signal, i.e., each student is endowed with a one-dimensional score and schools simply prefer students with higher scores.[5] Generally, given beliefs about the other students' priorities and strategies, a student's preference report corresponds to a lottery over DSPDA match outcomes. For instance, by swapping two schools' ranks in the reported ROL, match probability mass is shifted from one school to the other. With respect to the drawn match values, truthful reporting is a dominant strategy and, thus, induces a lottery that first-order stochastically dominates any lottery induced by

[4]Through differences in perceived discrimination, our model can resolve apparently contradictory findings in the data. While Shorrer and Sóvágó (2017) document that students with better socioeconomic background are more likely to deviate, Chen and Pereyra (2019) make the opposite observation. Higher social status may lead to a more pessimistic belief about getting a tuition waiver, but cause a more optimistic belief about getting into an elite school.

[5]For instance, the score may represent the result of a general assessment test, such as the SAT or GRE. In many countries and cities, all schools use the same centralized score to rank students. See Fack et al. (2019, Table 1)

any other ROL. Following the CPE-framework by Kőszegi and Rabin (2007), the chosen outcome lottery constitutes the reference point. That is, students compare their school match with any alternative school, and each pairwise comparison is weighted by the actual match probability determined by the report. A loss-aversion parameter determines to what extend losses are weighted stronger than gains.

Kőszegi and Rabin (2007) already proved that CPE allows for a preference for stochastically dominated lotteries. That is, a sufficiently loss-averse student may prefer to be matched with school x with certainty over being matched with the same school x with probability $(1-\epsilon)$ and being matched with an even better school y with probability $\epsilon > 0$. Intuitively, the mere possibility of getting into y makes the realization of the more likely outcome x more painful. Not listing y abandons all hope so that this school does not enter the stochastic reference point and disappointment is avoided. Although a match to any school is ex-post preferred over the outside option, extremely loss-averse students may even completely stay out of the matching market for the same reason. While moderately loss-averse students never misrepresent their preferences, only sufficiently optimistic dominantly loss-averse students report their preferences truthfully, and students with lower priority signals may truncate or perturb their preferences. This result matches experimental, survey, and field data suggesting that, in contrast to high-priority students, low-priority students are prone to deviations from the dominant strategy.[6] In contrast to other applications of CPE, where assuming "no dominance of gain-loss utility" restricts attention to moderate loss aversion, we allow for dominant loss aversion as well, which is in line with experimental evidence.[7]

We draw on the extensive literature on matching mechanisms, but depart from the standard framework where preferences appear to be very general as they only need to be ordinal. However, they are independent of mechanisms and reports and thus cannot account for endogenous reference points. In their seminal paper, Gale and Shapley (1962) set up the one-to-one matching problem and introduce the deferred-acceptance mechanism as a solution to find stable matchings. They prove that such matchings exist with respect to any preference profile. They also establish that their mechanism, with truthful input, implements the stable allocation optimal for the proposing side, and Dubins and Freedman (1981) and Roth (1982) show that it is also strategyproof for this side. In addition to that, Roth (1982) also proves that it is not strategyproof for the receiving side.[8] Balinski and Sönmez (1999) show that DSPDA is constrained efficient in the sense that no

[6]Basteck and Mantovani (2018) suggest that lower cognitive ability may drive this observation. However, it is also found in experiments where priorities and preferences are induced. That is, the same individuals play a dominant or dominated strategy depending on their assigned score. Moreover, Hassidim et al. (2017b) observe the same pattern in a population in which even the lower tail comes from the top of the ability distribution of the general population. Controlling for cognitive limitations, Shorrer and Sóvágó (2017) and Artemov et al. (2017) find a causal relationship between admission selectivity and dominated choices.

[7]See, for instance, Sprenger (2015) and the reference we provide after Lemma 2.

[8]In fact, he establishes that no stable mechanism exists that is strategyproof for both sides of the market.

other fair mechanism Pareto-dominates it. Our model introduces a fundamentally different structure of incentives and questions all of these classical insights.

Combinations of behavioral theory and matching are still relatively rare. To the best of our knowledge, the first paper to consider non-standard preferences in matching is by Antler (2015) whose agents' preferences are directly affected by the reported preferences of others. Fernandez (2018) studies anticipated regret in deferred acceptance. The point that EBLA can help explain misrepresentations in DSPDA was recently and independently raised by Dreyfuss et al. (2019). Alongside with various differences in modeling choices, they focus on the individual decision problem and use empirical strategies to identify loss aversion in existing experimental data. In contrast, we take a deeper theoretical approach by deriving characterization results on rationalizable ROLs, analyzing strategic interaction, and evaluating remedy mechanisms. We discuss the distinction to our paper more carefully in Section A.II. Since many behavioral biases distort beliefs which are irrelevant with a dominant strategy, we hope to ignite a literature on behavioral matching with reference-dependent preferences where beliefs are decisive.

Hassidim et al. (2017a) gather stylized facts about the pervasive misrepresentation of preferences in truthful mechanisms. Similar to Rees-Jones (2018) and Chen and Pereyra (2019) who analyze survey data,[9] they find that "misrepresentation rates are higher in weaker segments of markets" and increase "when applicants expect to face stronger competition". In field data, misrepresentations are hard to identify since the true preferences are subjective and private. However, Hassidim et al. (2017b), Shorrer and Sóvágó (2017) and Artemov et al. (2017) exploit objective rankings in their data to expose "obvious misrepresentations",[10] and they all find the same pattern. For instance, Shorrer and Sóvágó (2017) discover that a non-negligible fraction of these misrepresentations are costly, leaving over $3,000 on the table on average.

Truthfulness is easier to detect in the lab where preferences are imposed by the experimenter. Recently, the intentions behind matching experiments have shifted. While the pioneers Chen and Sönmez (2006) have focused on a comparison of different mechanisms, more recently researchers are investigating patterns in preference manipulations. Hakimov and Kübler (2019) provide a well-structured overview over the current state of experimental research on matching markets. They document that rates of truthfulness in DSPDA seem to depend on multiple factors which should not impede the dominance of the strategy and vary widely between studies, e.g., only 38 % in the first rounds of Bó and Hakimov (forthcoming) and 88 % in the zero-information treatment of Pais and Pintér (2008). Rather than rooted in behavioral theory, the experimental studies are descriptive.

[9]They study the National Resident Matching Program and the Mexico City high school match, respectively.

[10]They study the Israeli Psychology Master's Match, Hungarian college admission, and Australian college admissions, respectively. Naturally, all students should prefer a school with scholarship over the same school without scholarship, but the authors record that students forgo tuition waivers and no-strings-attached stipends.

For instance, Chen and Sönmez (2006) introduced the district-school bias and the small-school bias, which capture the tendency that safe district schools are ranked higher and small schools are ranked lower. We offer a theory to explain this pattern.

A natural explanation for dominated play is that participants simply fail to identify the dominant strategy. Indeed, Ding and Schotter (2017) find that participants see incorrect advice as more convincing than advice to play truthfully. Li (2017) formalizes "obvious strategyproofness" (OSP) as a stronger concept for mechanisms in which the strategyproofness is more apparent, and provides experimental evidence that OSP mechanisms indeed induce more truthfulness. However, the OSP mechanism he tests is also robust to EBLA. Hence, it is still an open question whether the non-truthful play in DSPDA and more truthful play in the remedy mechanism are due to a behavioral bias (EBLA) or a cognitive limitation (difficulties to verify the dominance). Our predictions are able to explain the most common deviations documented by Li (2017). We discuss the differences between the two concepts in Section A.III.

Since Kahneman and Tversky (1979), loss aversion has been recognized as an integral part of human preferences. Based on their insights, Kőszegi and Rabin (2006) developed EBLA and subsequently, in 2007, introduced and analyzed CPE, the equilibrium concept we employ. EBLA is supported by evidence from the field, such as Crawford and Meng (2011) or Pope and Schweitzer (2011), and from the lab, such as Abeler et al. (2011) and Ericson and Fuster (2011). However, also evidence contradicting EBLA exists, see, e.g. Heffetz and List (2014) or Gneezy et al. (2017). However, Heffetz (2018) mends the conflicting evidence by introducing an extra treatment that allows expectations to "sink in". EBLA has been applied to a variety of economic models.[11]

2 The model

Players: We consider finite sets of students, $\mathcal{I} := \{A, B, \dots\}$,[12] and schools, $\mathcal{S} := \{1, \dots, m\}$. Each school $s \in \mathcal{S}$ has a capacity of $q_s \in \mathbb{N}$ seats for students. If we want to allow for students to remain unmatched, we can think of school m as a safe outside option with unlimited capacity.

Preferences: Each student $i \in \mathcal{I}$ draws a type $\theta_i = (\mathbf{v}_i, \mathbf{w}_i, \lambda_i)$, where each entry of vector $\mathbf{v}_i = (v_{i,s})_{s\in\mathcal{S}}$ represents the payoff student i receives from being matched with corresponding school s.[13] Similarly, each element of vector $\mathbf{w}_i = (w_{i,s})_{s\in\mathcal{S}}$

[11] Such as moral hazard (Herweg et al., 2010), monopoly pricing (Herweg and Mierendorff, 2013; Heidhues and Kőszegi, 2014; Carbajal and Ely, 2016), pricing with competition (Heidhues and Kőszegi, 2008; Karle and Peitz, 2014), auctions (Lange and Ratan, 2010; Rosato, 2014; von Wangenheim, 2017), bargaining (Rosato, 2017), and labor markets (Eliaz and Spiegler, 2014).

[12] More than 26 students can be accommodated by continuing the list with $AA, BB, \dots, AAA, \dots$ and so on.

[13] In order to evaluate reference-dependent utility, we must rely on cardinal utilities. Yet, our main results will not depend on the cardinal ranking.

represents the payoff school s receives from being matched with student i.[14] Let $(\mathbf{v}_i, \mathbf{w}_i)$ be distributed over a compact subset of $\mathbb{R}^m \times \mathbb{R}^m$ for all $i \in \mathcal{I}$. We explain the loss-aversion parameter $\lambda_i \geq 1$ in its own section later, it is discretely distributed over a finite set $\{\lambda^1, \lambda^2, \dots\}$.

For some results, we consider the following relevant special case:

Assumption 1 (Homogeneous school preferences)**.** $w_{i,s} = \omega_i \quad \forall s \in \mathcal{S}$ *and* ω_i *is uniformly distributed*[15] *on* $[0, 1]$.

The ordinal preference over schools corresponding to type θ_i is captured by a rank-ordered list (ROL) ν_i. Formally, a ROL is a permutation of set $\mathcal{S}$, where a ROL $(s_1, s_2, \dots, s_m)$ is interpreted as school s_1 being most preferred, s_m least preferred, and s_k having k-th highest preference.[16] We call $\mathfrak{S}(\mathcal{S})$ the set of all such permutations.

Mechanism: Our results refer to the direct student-proposing deferred-acceptance algorithm (DSPDA) defined (with its properties) in the appendix. We assume that schools always report their true preferences over students.[17] Formally, a reporting strategy for student i is a mapping $\sigma_i : \Theta_i \to \mathfrak{S}(\mathcal{S})$ from types into ROLs. In particular, we are interested in when the truthful strategy,

$$\sigma_i^*(\theta_i) = \nu_i \quad \forall i, \theta_i, \tag{1}$$

which fully reveals the true ROL, is optimal.

Information: The rules of the mechanism are fully understood. Schools know their preferences over students. Students know their own type, schools' capacities, and the distributions of other students' types.

Loss aversion: While schools always non-strategically report their true ROLs, each student reports the preferences maximizing her expected utility. Students are expectation-based loss averse in the sense of Kőszegi and Rabin (2006, 2007). Hence, in addition to classical match utility $v_{i,s}$ the student perceives gains and losses when comparing the realized match utility to her reference utility. For the specification of gain-loss utility we follow most of the literature by assuming a linear gain loss function with a kink at zero. More specifically, let

$$u(\theta_i, s|r) = v_{i,s} + \begin{cases} \eta(v_{i,s} - v_{i,r}) & \text{if } v_{i,s} \geq v_{i,r}, \\ \eta\lambda_i(v_{i,s} - v_{i,r}) & \text{if } v_{i,s} < v_{i,r}, \end{cases} \tag{2}$$

[14] It is not crucial that students learn their true priorities $\mathbf{w}_i$. It suffices that they receive a signal inducing a belief about their priority at each school relative to the other students.

[15] Given iid draws from continuous distributions, this is without loss of generality. If ω_i is distributed with cdf $\Phi(\omega) \neq \omega$, we can relabel the score to be $\omega' := \Phi(\omega)$ which is uniformly distributed for any Φ.

[16] Ties in the ROL may be broken arbitrarily. With continuous type distributions indifferences occur with probability zero and do not affect any result in this paper.

[17] This assumption distinguishes school choice where local laws determine schools' priorities from the college admission problem where colleges are strategic actors, see, e.g., Chen and Sönmez (2006).

denote student i's ex-post utility from being matched with school s, when school $r \in \mathcal{S}$ is her reference match. The parameter $\lambda_i > 1$ captures the individual degree of loss aversion, whereas $\eta \geq 0$ is the general weight assigned to gain-loss utility.[18] Let $\Lambda_i = \lambda_i \eta - \eta$ be the loss dominance, and we call students with $\Lambda_i \leq 1$ moderately loss averse and students with $\Lambda_i > 1$ dominantly loss averse.

Given θ_i, a belief about θ_{-i}, and all other students' reporting strategies σ_{-i}, each report $\sigma_i(\theta_i)$ completing the strategy profile $\sigma := (\sigma_i, \sigma_{-i})$ corresponds to a distribution $F_i = (f_{i,s})_{s \in \mathcal{S}}$, where $f_{i,s}$ denotes the probability with which i expects to be matched with school s. Given θ_i, σ_{-i} and beliefs about θ_{-i}, we say a lottery is feasible for student i if there exists a report that induces it, and let $\mathcal{F}_i$ be the set of feasible lotteries. We will provide more details on the origin of this distribution in Section 3.1.1. The expected utility from a lottery F_i evaluated with respect to some reference lottery $G = (g_s)_{s \in \mathcal{S}}$ is then

$$\mathcal{U}_i(\theta_i, F_i | G) = \sum_{s \in \mathcal{S}} f_{i,s} \left(\sum_{r \in \mathcal{S}} u(\theta_i, s|r) g_r \right). \tag{3}$$

Equilibrium: Given some σ_{-i}, a strategy σ_i is a choice-acclimating personal equilibrium (CPE) for student i if, for all $\theta_i \in \Theta_i$ the corresponding distribution F_i satisfies

$$U_i(\theta_i, F_i) := \mathcal{U}_i(\theta_i, F_i | F_i) \geq \mathcal{U}_i(\theta_i, F_i' | F_i') := U_i(\theta_i, F_i') \quad \forall F_i' \in \mathcal{F}_i. \tag{4}$$

That is, we assume expectation-based loss aversion (EBLA) according to Kőszegi and Rabin (2007, Section IV), where the reference point is determined by the actual belief over the own matching outcome. In CPE, strategies maximize expected utility given that the corresponding beliefs determines both the reference lottery and the outcome lottery. For the strategic interaction we say a strategy profile σ is a choice-acclimating Bayesian Nash equilibrium (CBNE), if every $\sigma_i \in \sigma$ is a CPE given σ_{-i} for all $i \in \mathcal{I}$.

3 Analysis

In Section 3.1, we consider the individual decision problem of a single student, taking as given her type and the other students' strategy profile σ_{-i}. After some preliminary analysis, we start with an example which provides the main intuitions of how students with EBLA manipulate their ROLs and why the truthful strategy may no longer be optimal. We then characterize CPE in DSPDA, and find that our theoretical predictions can explain patterns observed in the data. In Section 3.2, we investigate the game theoretical problem of strategic interaction and analyze CBNE when schools have homogeneous preferences. In particular, we highlight how DSPDA misallocates school seats in equilibrium by favoring less loss-averse (and more optimistic) students. In Section 3.3, we propose alternative mechanisms

[18] Because it turns out that only parameter Λ_i drives behavior, all our results continue to hold if parameters η_i are individual.

to remedy the harmful strategic behavior that arises due to loss aversion. In general, proofs are relegated to the appendix, Section A.IV.

3.1 The individual decision problem

As we consider the individual problem of some student i by fixing her type θ_i and the other students' strategy profile σ_{-i}, it is convenient to drop the student's indices i and also, without loss of generality, relabel schools such that $v_1 > v_2 > \cdots > v_m$.

3.1.1 Match probabilities and acceptability

By the nature of DSPDA, student i is rejected by school s if at some step of the algorithm more than q_s students with higher priority than i apply to school s. Hence, student i is matched to the k-th ranked school of her ROL if the capacities of all schools she ranked above are filled by students that these schools individually prefer over student i. Given the others' ROLs, we define a student as acceptable if and only if she obtains a seat at school s when ranking it first.

Lemma 1. *DSPDA assigns a student to her highest-ranked school at which she is acceptable.*

The probability of being acceptable at a school depends on the strategies of other students and on the schools' preferences over students, but not on the submitted ROL by the student herself. The submitted ROL does, however, establish which of the schools at which she is acceptable is ranked first, and hence constitutes the student's match. Therefore, the submitted ROL determines the match outcome distribution F, and selecting a ROL effectively corresponds to choosing a lottery over match outcomes.

More precisely, a student's beliefs about other students' types and strategies leads to probabilities p_s of being acceptable at school s.[19] Since the student is matched with her highest-ranked school at which she is acceptable, a reported ROL $(s_1, s_2, ..., s_m)$ leads to to a lottery with match probabilities $F = (f_1, ..., f_m)$, where f_{s_k} is the joint probability that the student is acceptable at school s_k but not acceptable at schools $s_1, ..., s_{k-1}$. Importantly, the acceptability probabilities are usually not independent, even when types are independent draws. Recall that reporting the true ROL is a dominant strategy under standard preferences when $\eta = 0$. Hence, choosing to report any other ROL corresponds to choosing a first-order stochastically dominated lottery, which can be optimal as we will demonstrate.

[19] Nothing in the analysis of this section relies on the presumption that beliefs are correct. The student could be overoptimistic about her priority, hold wrong beliefs about other student's preferences or draw wrong inference on other student's used strategies. Importantly, she will choose a reporting strategy maximizing her expected utility given the beliefs she holds about acceptability at each school.

3.1.2 Outside options and truncated lists

In many existing implementations of DSPDA, it is allowed to submit incomplete ROLs and sometimes participants are even restricted to such truncations. In contrast, experimentalists often force participants to submit full rankings. We can include the possibility to drop a school, i.e., not listing it in the ROL, by enriching set $\mathcal{S}$ with an outside option such that remaining unmatched corresponds to being matched to a fictional school m with unlimited capacity and normalized $v_m = 0$. Ranking a school after the outside option corresponds to dropping it from the ROL. Although such strategies are dominated with standard preferences, truncated ROLs are ubiquitous in data on DSPDA. We are able to explain prevalent dropping strategies in Corollary 2.

Depending on the environment, remaining unmatched is not always the outside option. Clearly, there can be an actual school with a large enough capacity that it never rejects any student. Moreover, σ_{-i} could be such that less than q_s students apply to school s for any type realization such that our student i is acceptable with probability one. Alternatively, many school choice programs prohibit "district schools" to reject students from their district. Our results on ROL truncation also apply to the "district school bias" prevalent in the data.

For the individual problem, we define the most preferred school with probability of acceptability equal to 1 as the (de facto) outside option. Evidently, the student will never be matched with any school ranked below the outside option. Hence, not listing a subset of schools or listing them in any arbitrary order below the outside option is equivalent in the sense that any such list induces the same match probabilities. To identify all equivalent lists as one, we shall henceforth work with the convention that dropping a subset of schools is achieved by first listing all dropped schools with index larger outside option k behind the outside option in increasing order and then listing all dropped schools with index smaller k in decreasing order.

3.1.3 Payoffs

For any ROL resulting in lottery $F = (f_1, ..., f_m)$, we can rewrite the expected utility in (3) as

$$\begin{aligned} U_i(\cdot, F) &= \sum_{s\in\mathcal{S}} f_s \left(\sum_{r\in\mathcal{S}} u(\theta_i, c|r) g_r \right) \\ &= \sum_{s=1}^{m} f_s v_s + \sum_{1\le s\le r\le m} f_s f_r \eta(v_s - v_r) + \sum_{1\le r\le s\le m} f_s f_r \lambda\eta(v_s - v_r) \\ &= \underbrace{\sum_{s=1}^{m} f_s v_s}_{\text{classical utility}} \underbrace{- \Lambda \sum_{1\le s\le r\le m} f_s f_r (v_s - v_r)}_{\text{gain-loss utility}} . \end{aligned} \tag{5}$$

Since losses are weighted stronger than gains expected gain-loss utility always enters negatively. The difference $(v_s - v_r)$ is by convention positive for each $r >$

s. One can think of the expected gain-loss term as the cost of uncertainty. It is proportional to the loss dominance Λ and the average distance between two realizations. An equal weight on gains and losses, $\lambda = 1$ results in $\Lambda = 0$ such that students only maximize classical utility. If $\Lambda > 1$, gain-loss utility dominates match utility which will become central soon.

3.1.4 Example

The following example illustrates the tradeoff between the gains from classical utility and the losses from expected reference-dependent utility, which provides the incentives to misrepresent true preferences. It foreshadows our characterization results on which ROLs EBLA can rationalize and provides intuition for comparative statics in a student's loss dominance parameter and her priority. Intuitively, increasing Λ augments the relative weight of gain-loss utility over match utility. Hence, reducing the exposure to sensations of loss by taming expectations becomes a central motif.

Example 1. *There are three students, $\mathcal{I} = \{A, B, C\}$, and two schools with a single seat such that one student will remain unmatched. By treating the outside option as a third school with unconstrained capacity, we obtain $\mathcal{S} = \{1, 2, 3\}$ with capacities $q_1 = q_2 = 1, q_3 = 3$. Suppose that all students prefer a school seat over being unmatched and that school 1 is expected to be the more popular school,*

$$Pr(v_{i,1} > v_{i,2} > v_{i,3}) = (1 - \epsilon) \quad \textit{and} \quad Pr(v_{i,2} > v_{i,1} > v_{i,3}) = \epsilon \quad \forall i \in \mathcal{I}.$$

Schools' preferences are determined by a single score which each student independently draws from a uniform distribution on $[0, 1]$, i.e., Assumption 1 holds. We take the perspective of student A with preferences $v_1 > v_2 > v_3$ and score ω. Suppose she believes the other two students are truthful, playing σ^*_{-A}. Table 1 provides the distribution of acceptability probabilities for $\omega = 1/4$ and $\varepsilon = 1/10$.

Acceptability	at 1	not at 1
at 2	$\omega^2 = 10/160$	$2\omega(1-\omega)(1-\varepsilon) = 57/160$
not at 2	$2\omega(1-\omega)\varepsilon = 3/160$	$(1-\omega)^2 = 90/160$

Table 1: Acceptability probabilites for $\omega = 1/4$ and $\epsilon = 1/20$ and $\omega = 1/4$.

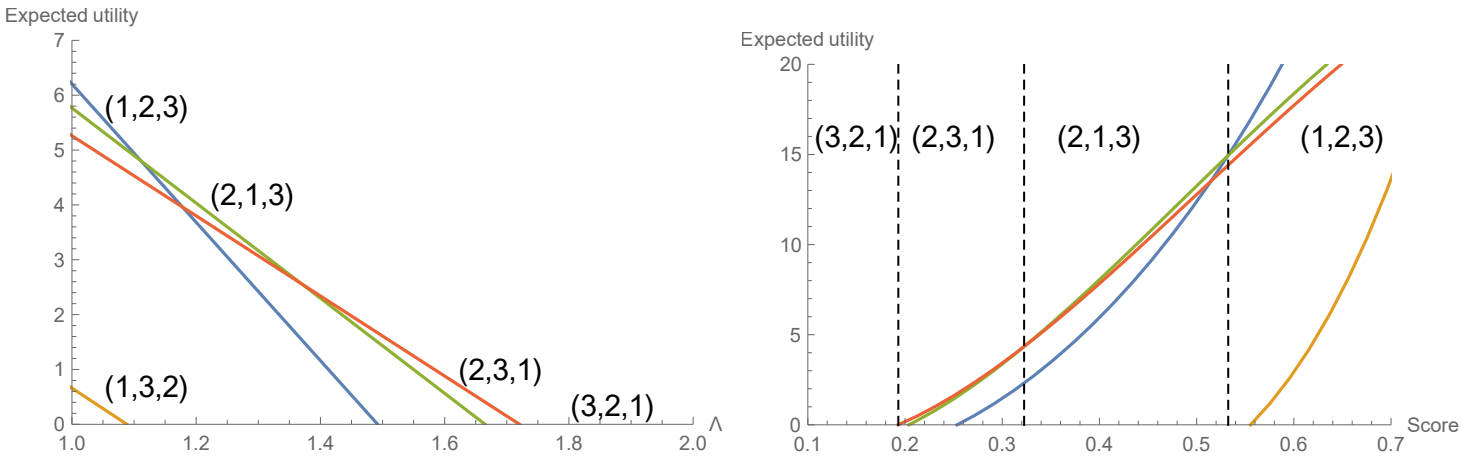
Evidently, the student is only acceptable at both schools if she has the highest score, and acceptable at neither school if she has the lowest score. She is acceptable at only one of the schools if she has the second highest score and the student with highest score prefers the other school. Note that the acceptability probabilities are interdependent, even though preferences and scores are drawn independently.

From the acceptability probabilities, the student can infer the lottery over match outcomes for any possible ROL. For instance, the true ROL, $(1, 2, 3)$, leads to a match with school 1 if and only if the student is acceptable there, with school 2 if and only if she is acceptable there but not at school 1, and to no match if and only if she is unacceptable at both schools. Table 2 presents match probabilities for all ROLs.

ROL	f_1	f_2	f_3
1,2,3	$13/160$	$57/160$	$90/160$
2,1,3	$3/160$	$67/160$	$90/160$
2,3,1	0	$67/160$	$93/160$
1,3,2	$13/160$	0	$147/160$
3,2,1	0	0	1
3,1,2	0	0	1

Table 2: All possible ROLs of the example and the corresponding lotteries for $\epsilon = 1/20$ and $\omega = 1/4$.

We see that flipping 1 and 2 in the ranking shifts a probability mass of $10/160$ (the probability of being acceptable at both schools) from school 1 to 2, which decreases classical utility but also the cost of uncertainty. Similarly, dropping the last ranked school simply shifts match probability mass from this school to the outside option. We will exploit this structure for several of our results. Trivially, ROLs listing the outside option first induce identical degenerate lotteries and are therefore regarded as equivalent.



(a) Expected utility as a function of Λ with $\omega = \frac{1}{4}$,

(b) Expected utility as a function of ω with $\Lambda = \frac{3}{2}$.

Figure 1: The expected utilities induced by every ROL as a function of (a) Λ and (b) ω, setting $v_1 = 100, v_2 = 30, v_3 = 0$ and $\epsilon = 1/20$.

Given the lotteries, we can calculate expected utilities for any Λ and select the optimal ROL. Figure 1 illustrates the expected utilities induced by different ROLs. Figure 1a demonstrates that for sufficiently small Λ the student always reports truthfully, as the lottery corresponding to the true ROL first-order stochastically dominates every other lottery and the positive effects on match utility dominate the cost of uncertainty. As we increase Λ, preferred schools are optimally ranked lower ultimately culminating in submitting an empty ROL when the perceived cost of uncertainty is sufficiently high.[20] *Notably, any optimal manipulation involves a*

[20]The fact that abstaining from the mechanism by choosing a dominated outside option is

flipping (or dropping) of the most preferred option – ROL $(1, 3, 2)$ *is never optimal. Intuitively, shifting probability mass from the extreme towards the expectation will most strongly diminish uncertainty, and is therefore most likely to exceed the losses in match utility. This insight will be generalized in our characterization result in in Proposition 1. From Figure 1b, we learn that students tend to become more truthful as their scores increase and they become more optimistic. Hence, in particular the disadvantaged students are prone to untruthful reporting.*

A large Λ *by itself does not lead to profitable deviations from the true ROL. If students had full information about students' and schools' preferences, they could infer their match outcome for each ROL from* σ_{-i} *by backwards induction, as their acceptability is only determined by the strategy of students with higher scores. Hence, there is no uncertainty and students have no cost of being truthful such that DSPDA implements the student-optimal stable matching.*

3.1.5 Characterization of optimal ROLs

It is well-known that DSPDA is strategyproof for students and implements the student-optimal stable allocation with standard preferences. However, as we have seen in Example 1, the dominance of the truthful strategy does not necessarily carry over to a truthful CPE if loss aversion is sufficiently strong. We show that, for any $\Lambda > 1$, a sufficiently pessimistic student will misrepresent her preferences. Conversely, Masatlioglu and Raymond (2016, Proposition 1) show that CPE respects first-order stochastic dominance if $\Lambda \leq 1$.

Lemma 2. *The truthful strategy* σ_i^* *is a CPE in DSPDA for student* i *for all* $(\mathbf{v}_i, \mathbf{w}_i)$ *and all possible beliefs if and only if* $\Lambda_i \leq 1$.

While many applied papers restrict attention to $\Lambda \leq 1$, "no dominance of gain-loss utility",[21] we allow (all or only some) students to be dominantly loss averse. There is substantial evidence that a large fraction of the population is indeed dominantly loss averse, and $\Lambda > 1$ also matches the conventional wisdom that "losses loom about twice as large as gains".[22] While the possible preference for first-order stochastically dominated lotteries that comes with this assumption may appear counterintuitive, it is observable.[23] More importantly, excluding such preferences

reminiscent of the "uncertainty effect" documented by Gneezy et al. (2006).

[21] This assumption was introduced by Herweg et al. (2010) as $\lambda \leq 2$ with fixed $\eta = 1$, and later picked up in various forms by, among many, Herweg and Mierendorff (2013), Herweg (2013), Karle and Peitz (2014), or Rosato (2014). Rather than based on evidence, the main reason why it is imposed seems to be that it makes problems well-behaved.

[22] While this rule of thumb originates from studies on riskless choices, it also seems to apply when risk is involved, see Tversky and Kahneman (1992), Gill and Prowse (2012), Sprenger (2015) or Karle et al. (2015). In our setting, it corresponds to $\frac{1+\eta\lambda}{1+\eta} \approx 2$, which implies $\Lambda \approx 1 + \eta > 1$.

[23] See the discussion around Proposition 7 by Kőszegi and Rabin (2007). While the "uncertainty effect" found by Gneezy et al. (2006) provides evidence in this direction, Rydval et al. (2009) suggest it cannot be replicated. In the context of choice bracketing, Tversky and Kahneman (1981) and Rabin and Weizsäcker (2009) provide experimental evidence that people can have a preference for dominated lotteries.

would preclude us from explaining our phenomenon in which students indeed do choose dominated lotteries.

As behavior in DSPDA under moderate loss aversion is standard, we consider dominant loss aversion for the remainder of this section, and provide novel characterizations of non-truthful CPE. At first glance, a full characterization appears quite arduous because of the multitude of ROLs available. Our results enable us to reduce substantially the set of ROLs that are candidates for a CPE. Our predictions are testable and consistent with pervasive deviations from the truthful strategy.

Based on the following lemma, Proposition 1 allows us to restrict attention to ROLs with certain properties when searching for a best-responding ROL given a type, beliefs, and others' reporting strategies. If all ROLs correspond to different lotteries over match outcomes, the proposition holds for any optimal ROL. However, if some ROLs correspond to identical lotteries (and therefore identical expected utility), it is possible that a student is indifferent between multiple ROLs out of which at least one will satisfy the properties of the lemma.[24]

Lemma 3. *If a strictly optimal ROL ranks school l after school n for $l < n$, it ranks the schools $1, ..., l-1, l$ in decreasing order.*

Table 3 shows all possible ROLs for a setting with three schools and the option to remain unmatched ("school 4"). The bold numbers are the listed schools and schools ranked after 4 can be interpreted as "dropped from the ranking". The darkly shaded ROLs are redundant in the sense that they are either equivalent to another ROL listing only one school or another ROL dropping all schools. The light shaded ROLs are the ones never strictly optimal as characterized by Lemma 3. For instance, $(1, 3, 2, 4)$ is not optimal as it ranks 2 after 3 but 1 before 2. Intuitively, if the student were willing to reduce risk by shifting probability mass from school 2 to 3, i.e., $(1, 3, 2, 4) \succ_i (1, 2, 3, 4)$, then she would be a forteriori willing to shift probability mass from the more extreme school 1 downwards, i.e., $(3, 1, 2, 4) \succ_i (1, 3, 2, 4)$, so $(1, 3, 2, 4)$ can never be strictly optimal.

An immediate consequence of Lemma 3 is the following characterization of the structure of any optimal ROL. We define a ROL as CPE-rationalizable if it can possibly result in a CPE, i.e., if there exist types such that this ROL is optimal. Only a comparably small set of ROLs is CPE-rationalizable. Indeed, while for m schools the number of ROLs is $m!$ (or $\sum_{i=1}^{m}(m-i)!\binom{m-1}{i-1} = \sum_{i=1}^{m} {}^{(m-1)!}/_{(i-1)!}$ non-redundant ROLs when m is an outside option), the number of ROLs as described in Proposition 1 is just 2^{m-1}.

[24]For this reason, we render ROLs equivalent for which only the ranking after the outside option differs. Identical lotteries can also arise if a subset of schools together constitute an outside option, making any permutation of schools ranked after them meaningless. Similarly, the ranking of two schools at which the student is never acceptable does not matter. There are no equivalent ROLs if for any subset of schools where acceptability is not certain the probability of being acceptable at precisely these schools is strictly between zero and one.

Full ROL	**Drop one**	**Drop two**	**Empty ROL**
1,2,3,4	**1,2**,4,3	**1**,4,3,2	4,3,2,1
2,1,3,4	**2,1**,4,3	**2**,4,3,1	4,1,2,3
3,1,2,4	**3,1**,4,2	**3**,4,2,1	4,2,1,3
1,3,2,4	**1,3**,4,2	1,4,2,3	4,3,1,2
2,3,1,4	**2,3**,4,1	2,4,1,3	4,1,3,2
3,2,1,4	**3,2**,4,1	3,4,1,2	4,3,2,1

Table 3: All possible permutations with three schools and an outside option. The darkly shaded ROLs are redundant. By Proposition 1, the lightly (and darkly!) shaded ROLs are never strictly optimal.

Proposition 1. *A strictly optimal ROL which ranks k first must rank schools $1, \ldots, k-1$ in decreasing order and schools $k+1, \ldots, m$ in increasing order.*

This characterization result is mainly a more intuitive reformulation of the preceding lemma. In particular, it implies that any manipulation of the ROL will concern the most preferred schools, a testable prediction. As a first impression of our theory's predictive power, we briefly consider the experiment by Li (2017, treatment SP-RSD). Here, each participant is privately endowed with a priority score, an integer between 1 and 10, and is informed about how all participants commonly value each of four prizes, random draws from a discrete distribution on [\$0,\$1.25]. Next, participants simultaneously submit a ROL to a direct serial dictatorship mechanism which processes the ROLs in decreasing order of the priority scores and ties are broken randomly. Essentially, this setting is a special case of our analysis of DSPDA.

Priority	1	2	3	4	5	6	7	8	9	10	ALL
1234	61.1%	57.1%	58.8%	67.7%	55.2%	79.0%	74.4%	85.7%	84.3%	91.3%	71.0%
other	38.9%	42.9%	41.3%	32.3%	44.8%	21.0%	25.6%	14.3%	15.7%	8.8%	29.0%
2134	1.1%	1.2%	3.8%	**6.5%**	**12.1%**	**8.1%**	**10.3%**	**7.1%**	**5.7%**	**1.3%**	**5.3%**
3214	6.7%	6.0%	**7.5%**	4.8%	3.4%	0.0%	0.0%	1.8%	0.0%	0.0%	3.2%
4321	**17.8%**	**8.3%**	3.8%	4.8%	1.7%	3.2%	1.3%	0.0%	2.9%	0.0%	4.9%
CPE	91.1%	77.4%	77.5%	88.7%	75.9%	91.9%	87.2%	98.2%	95.7%	93.8%	87.5%

Table 4: The first two rows indicate the shares of truthful and manipulated ROLs, respectively. The next three rows show the most common misrepresentations and their fraction of all submitted ROLs (most common in bold face). The final row states the fraction of CPE-rationalizable ROLs of all submitted ROLs. The columns represent each priority score with the last one being an aggregation over all scores.

Table 4 documents several noteworthy observations regarding our theoretical results. Table 5 in the appendix provides more details. While the standard theory can explain 71% of the ROLs (first row, last column), our theory can explain 87.5% of the reported ROLs (last row, last column). More importantly, the most common misrepresentations for each priority score (in bold face) are indeed all

CPE-rationalizable. Moreover, the rates of these misrepresentations move according to the intuitions suggested by our model, ROL $(4, 3, 2, 1)$ is most common among low scores, ROL $(3, 2, 1, 4)$ among lower intermediate scores, and ROL $(2, 1, 3, 4)$ among higher intermediate scores. As suggested by Example 1, high scores are more likely to submit truthful ROLs. Although this rate does not increase monotonically, there is a clear trend. The prediction relying only on the induced values as exogenous preferences, truthful reporting as a dominant strategy, fares especially bad for low scores. From Proposition 1, we can immediately deduce when students prefer to be truthful, Corollary 1 and Proposition 2.

Corollary 1. *The true ROL is optimal if and only if it is optimal to rank school 1 first.*

This insight helps us to provide necessary and sufficient conditions on the loss parameter which determine whether a manipulation of the true ROL is profitable. Based only on exogenous fundamentals, Proposition 2 gives precise bounds on when DSPDA is incentive-compatible for loss-averse students. These bounds are strict in the sense that for any $p_1 \in [\frac{1-1/\Lambda}{2}, 1 - 1/\Lambda]$ the answer to whether truthfulness is optimal depends on other acceptability probabilities and also the cardinal utilities.

Proposition 2. *Let p_1 be the probability that the student is acceptable at her most preferred school.*

1. *If $p_1 > 1 - 1/\Lambda$, the true ROL is optimal for any such θ_i.*
2. *If $p_1 < \frac{1-1/\Lambda}{2}$, the true ROL is not optimal for any such θ_i.*

The proposition immediately implies that under Assumption 1 sufficiently high scores report truthfully whereas sufficiently low types misrepresent whenever seats at their preferred school are scarce. This result is in line with the evidence suggesting a causal relationship between priority and truthfulness mentioned in our introduction and Table 4.

An important implication of the result is that students' beliefs are crucial. That is, one of the advantages of strategyproof mechanisms, namely, the irrelevance of priors, vanishes. Importantly, we have made no assumptions on whether the beliefs determining the acceptability probabilities are correct. Consequently, EBLA is a channel which renders other well-documented biases distorting the beliefs decisive. For instance, an overconfident student is more likely to be truthful as she overestimates her chances of getting into her favorite school. Hence, overconfidence and loss aversion countervail each other in terms of incentive compatibility. Indeed, Rees-Jones and Skowronek (2018) find that overconfident[25] participants are more likely to be truthful. Without our theory, this observation may appear counterintuitive as this bias usually steers behavior away from the rational

[25]In their online experiment, participants completed a test on logical reasoning ability and afterwards estimated the percentage of other participants they outperformed. They deem a participant overconfident if they overestimated their percentile rank.

unbiased benchmark.

In a similar vein, our results have relevant ramifications for affirmative action policies. For ease of exposition, suppose some school evaluates students according to a one-dimensional score ω, but students of a certain demographic are discriminated in the sense that their score is reduced to $\omega' = a\omega + b$ with $a < 1$ and $b < 0$. Consequently, a discriminated student is more pessimistic about her acceptability, and, hence, there are scores ω for which discriminated students do not reveal their top preference for this school, while students of other demographics with the same score would list it. Therefore, DSPDA aggravates the discrimination by discouraging the respective demographic from revealing true preferences. For this insight, it is irrelevant whether this discrimination is real or only perceived. Thus, the reasoning that such discrimination is inconsequential because marginalized students don't rank discriminating schools in DSPDA is inherently flawed in models incorporating EBLA.

Truncated lists are prevalent in the data, but they obviously constitute misrepresentations when dropped schools are preferred over the outside option. Since constraining the ROLs to a fixed number of schools can destroy both the strategy-proofness and stability of DSPDA, economists advocate against such restrictions. While prohibiting complete ROLs introduces strategic motifs into DSPDA with standard preferences, such motifs are already present in our setting and students may voluntarily choose to truncate their ROLs.

Because it is never optimal to list an undesirable school, i.e., one that is considered worse as the outside option, we consider only desirable schools. By dropping such a school from the list, a student simply forgoes the chance of being assigned to this school in favor of matching with the outside option. We now characterize optimal truncated ROLs. Proposition 1 implies that a student will never truncate her true ROL from below. If an incomplete ROL is optimal, it is optimal to drop the most preferred rather than the least preferred schools.

Corollary 2. *It is never optimal to drop some desirable school k, but list some school $\ell < k$.*

In light of Proposition 1, the intuition is straightforward. A student only listing a single school $s < (m-1)$ can shift match probability from the outside option to a preferred outcome by adding school $(m-1)$ to her ROL. In some sense, this addition provides a free insurance which delivers gains over the outside option and simultaneously reduces losses compared to school s. If the student was so loss-averse that she prefers to forgo such insurance, she would prefer to drop $s < (m-1)$ as well. As a result, if a singleton ROL is optimal, it only lists the least-preferred (desirable) school. The same intuition carries over to longer truncated lists, implying that the true ROL is never truncated in a strict[26] optimum.

[26]Note that a truncated true ROL could be optimal if it is equivalent to the true ROL. For instance, a top (or overconfident) student may assign probability 1 to being admitted to one of her first three choices such that the the order of the schools listed afterwards is irrelevant.

The extreme form of truncation is submitting an empty ROL and thereby essentially abstaining from the school-choice program. Indeed, such behavior can be optimal. If Λ is very large, almost all such types submit an empty ROL to obtain the outside option with certainty. Even for reasonable Λ, very pessimistic students might optimally stay out of the matching market. Notably, such students may prefer a certain match with the outside option although $\sum_{s=1}^{m-1} q_s \geq |\mathcal{I}|$, i.e., although a complete ROL would assign them to some desirable school with certainty, but it is ex-ante uncertain which one. The reason is that the uncertainty of the match outcome introduces scope for loss utility. Straightforwardly, no student abstains from the matching market if there is some school s with $q_s \geq |\mathcal{I}|$, because this school essentially becomes the outside option. Similarly, a school that is listed by no other student becomes a de-facto outside option.

3.2 Strategic interaction

In this section, we investigate strategic interaction and the structure of choice-acclimating Bayesian Nash equilibria (CBNE). For discrete types and homogeneous school preferences, we derive an essentially unique equilibrium in pure strategies. In addition, we rationalize the prevalent district school bias as an equilibrium phenomenon in a setting with district and elite schools. We provide bounds for the social cost of this bias and show that it persists as the market grows large.

To establish general existence of a CBNE, note that a CBNE is just a standard Bayesian Nash Equilibrium, where individual utilities over actions are given by utility function (5).[27] Equilibrium existence is then implied by Theorem 1 in Milgrom and Weber (1985).[28] In general, strategic interaction between loss-averse agents is difficult to analyze and has only been sparsely studied by now. To better understand the structure of symmetric CBNE, we assume for the remainder of this section that the students' type spaces are finite and schools have homogeneous preferences over students. Moreover, suppose further that all schools break ties between students in the same publicly known deterministic way.

A key observation in the analysis of strategic interaction with homogeneous schools preferences is the fact that a student's match outcome will only be affected by the behavior of other students with a higher score. Indeed, by the nature of DSPDA, a student will only be rejected by a school she proposes to at some stage if this school also has an offer from a student with a higher score. Hence, intuitively, the existence and the structure of a CBNE follows by an iterative argument where each student chooses her optimal ROL according to the rational beliefs she holds over

[27] This interpretation subtly involves some view on the interpretation of mixed strategies. We follow, e.g., Rubinstein (1991) in his interpretation that we should either regard mixed strategies "as the distribution of the pure choices in the population" or as "a plan of action which is dependent on private information which is not specified in the model." In both interpretations the player knows his own choice of (pure strategy) action when forming her reference point. In this interpretation we depart from Dato et al. (2017) who assume that the uncertainty of a mixed strategy realizes only *after* the player chose it and formed her reference point.

[28] More concretely, note that compact metric spaces are complete and separable and that utility functions are measurable for the induced Borel σ-algebra.

submitted ROLs by students of higher scores. We say an equilibrium is essentially unique if it is unique with respect to a public known rule determining how students decide when indifferent between multiple ROLs.

Proposition 3. *With homogeneous school preferences, there exists an essentially unique CBNE in pure strategies.*

Elite schools and the district-school bias

We now employ a simplified setting to derive the district-school bias introduced by Chen and Sönmez (2006). Hakimov and Kübler (2019) state the phenomenon as "the district school (or safe school) is ranked higher in the reported list than in the true preferences" and document how prevalent it is in a wide range of matching experiments.

Suppose for the sake of this example that there is a set $\mathcal{E} \subset \mathcal{S}$ elite schools where each school from this set is unambiguously preferred by each student over some (possibly type-dependent) safe outside option which can be thought of as the local district school where the student has top priority. To simplify, we assume that all elite schools induce the same utility to a student. We can then normalize that each elite school induces a match utility of $v > 0$ whereas the safe outside option induces a utility of zero.

Suppose further that for each student i a score ω_i is independently drawn from a common continuous distribution with compact support. By parameterizing the score to the respective quantile of the distribution, we can assume without loss of generality that scores are drawn uniformly from the unit interval. Let a student's loss dominance Λ_i be independently drawn from a common distribution with discrete support $\{\Lambda^0, \Lambda^1, \Lambda^2, \ldots, \Lambda^l\}$. Since truthful reporting is a dominant strategy for any $\Lambda < 1$, we can combine all loss dominance parameters in $[0, 1]$ into Λ^0 and assume, without loss of generality, $\Lambda^0 = 0$ and $\Lambda^1 > 1$. The following lemma shows that we can, without loss of generality, focus on the case of only one elite school.

Lemma 4. *For any belief on the distribution of acceptability probabilities of elite schools it is either a best response to apply to all elite schools or to apply to no elite school.*

Hence, in the following we will not think of a set $\mathcal{E}$ of elite schools, but rather of one elite school with joint capacity $q = \sum_{s \in \mathcal{E}} q_s$, which is assumed to be smaller than $|\mathcal{I}|$, the number of students.

In CBNE, a representative student's decision whether to apply for the elite school depends on her acceptability probability f at the elite school, which is given by rational beliefs about the probability that less than q students of higher score apply to the elite score. Hence, the acceptability probability is a function $f(\omega)$ which is weakly increasing in her score ω. Again, by (5), listing the elite school

before the outside option is optimal for any $\omega > 0$ if and only if

$$f(\omega)v - \Lambda f(\omega)(1 - f(\omega))v \geq 0 \iff \Lambda \leq \frac{1}{1 - f(\omega).} \tag{6}$$

Consequently, for any score $\omega \in (0, 1)$, there is a cutoff $\overline{\Lambda}(\omega) = \frac{1}{1-f(\omega)}$ such that applying to the elite school is a best response if and only $\Lambda \leq \overline{\Lambda}(\omega)$. Due to the monotonicity in the cutoff structure, the CBNE can again be determined iteratively. Students with the highest loss dominance Λ^l have the highest cutoff score $\overline{\omega}(\Lambda^l)$ below which they abstain from listing the elite school in their ROL. Anticipating this behavior, any student with loss dominance Λ^{l-1} can infer her score cutoff below which $\overline{\omega}(\Lambda^{l-1})$ she drops the elite school, and so on.

Lemma 5. *In the elite school problem, there is an essentially unique CBNE in which a student with loss dominance Λ applies to the elite school if and only if her score is above some cutoff score $\overline{\omega}(\Lambda) \in (0, 1)$, which is increasing in Λ.*

The red curve in Figure 2 illustrates the threshold in Λ for each score ω above which a student does not apply to the elite school in equilibrium, when the distribution of Λ approaches a uniform distribution on $[0, 3]$. The blue curve indicates the threshold of a best response when all other students are truthful and apply. Hence, the difference between the two curves illustrates exactly the effect of strategic interaction. Indeed, the fact that some higher-priority students do not apply due to their strong loss dominance makes a student more optimistic such that she is more likely to apply herself.

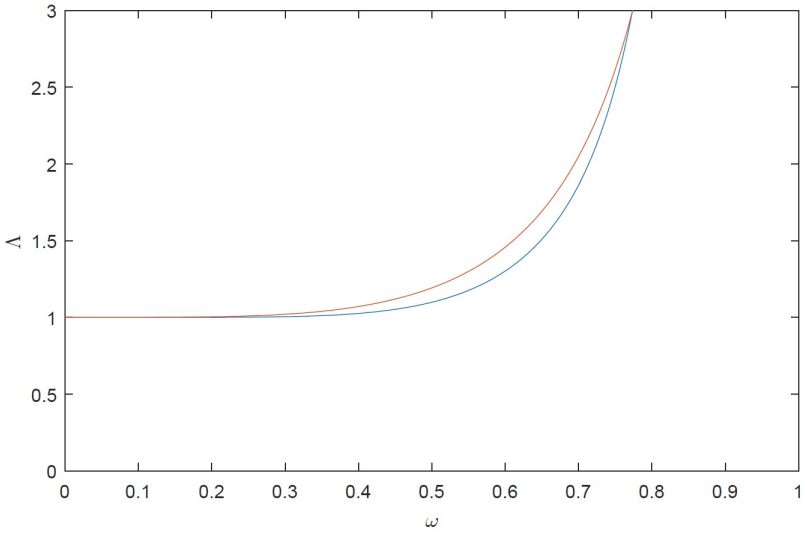


Figure 2: Truthfulness in the elite-school problem with $\Lambda \sim U[0, 3]$, $q = 3$, $n = 10$.

Next, we analyze the cost of inefficiency that arises from misreporting. In particular, we show that misrepresentations remain substantial and consequential as we let the number of students go to infinity. The fact that there are inefficiencies in the small markets in experiments is not necessarily alarming. Unfortunately,

misrepresentations also seem to be costly in large real-world markets, where we are only able to identify a lower bound on these costs through obvious misrepresentations. Artemov et al. (2017) and Hassidim et al. (2017b) find that 1 - 20% and 2 - 8% of obvious misrepresentations are ex-post costly, respectively, and Shorrer and Sóvágó (2017) further estimate that the 12 - 19% costly obvious misrepresentations amount to $3,000 - $3,500 on average (unconditionally $347 - $738 per misrepresentation).

Suppose for the sake of simplicity that there are only two types of loss aversion. A share α of students has loss dominance $\Lambda^H > 1$ whereas share $(1-\alpha)$ has loss dominance parameter $\Lambda^L < 1$ such that they always report truthfully. Obviously, as we increase the number of students in our setting while holding capacity q fixed, a growing share of students manipulate their reported preferences and abstain from applying to the elite school, as a limited capacity leaves only the very best students with a substantial acceptability probability. As most of the deviations are inconsequential in the sense that the manipulating students are very unlikely to be acceptable at the school anyway, low truthfulness rates do not necessarily decrease efficiency. A more meaningful measure for inefficiencies is the probability that the match outcome is unstable and the share of students suffering from justified envy relative to the capacity of the elite school. As the necessary score to be acceptable becomes more predictable when the number of students is large one might think that misrepresentations become unsubstantial in large markets. The following proposition shows that this is not the case.

Proposition 4. *Let q be the capacity of the elite school and α the share of students with loss dominace $\Lambda^H > 1$. Denote with $Y_q \sim \Gamma(q,1)$ a gamma-distributed random variable with shape q and rate 1, and with G_q its cdf. As n goes to infinity,*

1. *the probability of an unstable allocation is weakly above $\alpha\left(1-\frac{1}{\Lambda^H}\right)$ with equality only for $\alpha = 1$ or $q = 1$, and*
2. *the expected number of students exposed to justified envy is weakly above*

$$\alpha\left(1-\frac{1}{\Lambda^H}\right)\left(\mathbb{E}\left[Y_q | Y_q \geq G_q^{-1}(1/\Lambda^H)\right] - G_q^{-1}(1/\Lambda^H)\right),$$

and equality, again, only for $\alpha = 1$

In particular, the expected number of students who suffer from justified envy is bounded away from zero. Intuitively, there are two countervailing effects as n grows larger. First, the confidence intervals around the necessary cutoff score to be acceptable become smaller. Second, the number of students with a score in any interval becomes larger. The proposition shows that both effects are mainly offsetting each other such that the expected number of students approaches a constant above zero. Moreover, note that we are considering a stylized model where students are fully informed about their own score and uncertainty only stems from uncertainty about other students' scores. In reality, another source of uncertainty

may concern how the school evaluates abilities which is independent of market size. In this sense, our result provides a lower bound for costs of uncertainty.

3.3 Possible remedies

We have seen that under students' EBLA DSPDA may not implement the student optimal stable allocation, and that truthful reporting may no longer be optimal. This insight motivates the obvious question whether there are other matching mechanisms to achieve these goals. We first consider static mechanisms and then move on to two sequential mechanisms, sequential student-receiving (school-proposing) deferred acceptance and serial dictatorship.

Static mechanisms

If we restrict to static matching mechanisms, we can provide a negative result. A static mechanism, as formally defined in the Appendix A.I, is any mechanism which asks students about their preferences only once without providing feedback on other students' preferences.

Proposition 5. *For any distribution of preferences, a static mechanism that generates the student optimal stable allocation as CBNE for all realization of preferences exists if and only if the DSPDA is truthful.*

Hence, if DSPDA fails, there is no hope for remedies in the class of static mechanisms. Formally, the result is an immediate implication of the revelation principle for static mechanisms. Intuitively, if a student prefers to avoid the ex-ante risk that comes with the implementation of the student optimal matching, she will not reveal her preferences under any such mechanism.

Sequential school-proposing DA

Since uncertainty is the source of loss-averse students' deviations, the use of sequential mechanisms may mitigate this problem. A sequential mechanism enables feedback between different rounds, and hence has the ability to alter beliefs before asking for reports. At first sight, it may seem surprising that the sequential use of information enables us to go beyond what is achievable with static mechanisms, as it seems to violate the fundamental insight of the revelation principle that any sequential mechanism has a static direct equivalent Myerson (1979). In settings with dynamic information and expectation-based loss averse-agents, however, the revelation principle does not apply. As loss-averse agents evaluate outcomes with respect to beliefs, information endogenously affects their preferences over alternatives.[29]

For a dynamic equilibrium concept in the context of EBLA, we follow Rosato (2014) in his straightforward extension of a CPE to dynamic situations. At each decision node of an extensive form game, a student correctly anticipates her choices at any point in the future. Based on the induced beliefs and using backward induction, she selects the lottery most-preferred under the static CPE at every

[29] This point has been raised in the context of auctions, see von Wangenheim (2017).

decision node, with the reference point at each choice being her beliefs about final match outcomes conditional on the information available at that stage.

At any stage of the mechanism, new information can be revealed that alter the beliefs about the final match outcome, which depends on the student's behavior in future stages. Let $\mathcal{F}_{i,k}$ be the set of feasible lotteries given θ_i, σ_{-i} and the beliefs about θ_{-i} conditional on the information available at node k, and let $F_{i,k}$ be the lottery corresponding to some σ_i. Given some σ_{-i}, a strategy σ_i is an SCPE if, at any decision node k, it selects a lottery $F_{i,k}$ such that

$$U_i(\theta_i, F_{i,k}) \geq U_i(\theta_i, F'_{i,k}) \quad \forall \theta_i \in \Theta_i,\ \forall F'_{i,k} \in \mathcal{F}_{i,k}.$$

Accordingly, we call a strategy profile where each player's strategy is a SCPE given other players' strategies a sequential choice-acclimating Bayesian Nash equilibrium (SCBNE). We are now equipped with the tools to analyze the incentives of loss-averse students in sequential mechanisms, and turn to the evaluation of the sequential student-receiving (school-proposing) deferred-acceptance mechanism.

Definition 1 (Sequential student-receiving DA, SSRDA)**.**

$t = 1$ *All schools offer their most-preferred student a seat. All students may temporarily accept one of their offers (if they have one), and reject all other schools.*

$t > 1$ *All schools that have temporarily unfilled seats make an offer to the highest-ranked student that has not yet rejected them yet. All students may tentatively accept one of their new offers (if they have one), and reject their current match (if they have one).*

End *The process terminates after the first step without rejections.*

It is well known that DA is not truthful for the receiving side. Strategically rejecting an acceptable school may trigger other students to be matched with that school who may then leave capacities at more preferred schools. In practice, however, and in contrast to the proposing side, strategizing on the receiving side seems to play no major role.

Under complete information, it is known that the receiving side in DA cannot gain from manipulations if and only if the the student-optimal and the school-optimal stable match coincide.[30] Intuitively, the stable match is unique if preferences are sufficiently aligned between both sides of the market. For our model of incomplete information, where students' preferences are drawn randomly, it is convenient to think of alignment as a condition on schools' payoffs $w_{i,s}$ with respect to the drawn values $v_{i,s}$ or vice versa. The following definition conveys an appropriate extension of this notion for our setting with incomplete information.

[30]The reason is that under DSPDA truth telling leads to the school-optimal stable match. If this does not agree with the student-optimal stable match, students can force it by coordinating to only accept the school they would receive under their preferred stable match outcome.

Assumption 2 (Aligned preferences)**.** *Preferences are aligned if for all realization of preferences*

$$w_{i,s} \geq w_{i,s'} \iff v_{i,s} \geq v_{i,s'} \quad \forall s, s' \in \mathcal{S} \text{ and } \forall i \in \mathcal{I}.$$

Under Assumption 2, we may think of the schools' match values as a (not necessarily deterministic) function of students' match values. A possible interpretation would be that, for instance, students who prefer law schools over economics schools also score better in characteristics that are important to law schools.

Lemma 6. *If preference are aligned, there is a unique stable match for all realizations of preferences.*

Hence, if preferences are aligned, for no realization of preferences students can gain from strategizing in SSRDA. The same remains to hold under EBLA.

Proposition 6. *If preferences are aligned, the truthful strategy profile is a SCBNE in SSRDA, and the unique stable allocation is implemented for all preference realizations.*

Intuitively, a student with an offer can obtain a seat with certainty if she accepts. A strategic rejection of an acceptable school may in general trigger an offer from a more preferred school, but comes at the risk of receiving a worse outcome. Since the loss parameter Λ can be interpreted as a cost parameter for uncertainty, such strategic considerations become even less appealing for larger Λ. Hence, loss aversion tends to mitigate the incentives for such misreporting. This effect is related to the logic in Fernandez (2018), who identifies anticipated regret for the case where manipulations don't pay off as a possible explanation for the observed truthful behavior in SSRDA. The following example shows that SSRDA can produce match outcomes in equilibrium that are strictly preferred by students to the outcome under DSPDA.

Example 2. *Consider the elite-school problem under Assumption 1 with two students and $q = 1$. The only stable matching in this problem is that the student with higher score is assigned to the elite school whereas the lower-score student is matched with a district school. SSRDA implements this allocation. Indeed, both schools propose to the stronger student, she accepts at the elite school, leaving the district school for the lower-score student. Under DSPDA, however, students report their preferences truthfully only if* (6)*. Consequently, if the score of both students is below $1 - 1/\Lambda$, both students misrepresent their preferences, attend the district school, and the match outcome is neither stable nor student optimal.*

Serial dictatorship

Under Assumption 1, SSRDA simplifies considerably. When all schools have the same preferences, all schools approach the same student in the first step. Then, this student is aware that she has the highest score among all students and is immediately accepted at the school she selects. All other schools are rejected and

apply to the second-highest-score student who is then aware that she is now the highest-score student of the unmatched population and that she is assigned to her selected school with certainty, and so on. In short, SSRDA simply becomes serial dictatorship in which homogeneous priority scores determine the order. Since homogeneous school preferences constitute an example of aligned preferences, the following corollary is an immediate consequence of Proposition 6.

Corollary 3. *If schools have homogeneous preferences over students, the truthful strategy is an SCBNE in SSRDA, and the unique optimal stable allocation obtains.*

Li (2017) compares the outcome of DSPDA with the outcome of a sequential serial dictatorship mechanism as induced by SSRDA in a lab experiment. He finds that while in DSPDA 36 % of games do not end in the stable outcome as induced by the dominant strategy, this rate drops to 7 % under SSRDA. He explains this finding by the fact that –in contrast to SSRDA in this setting– DSPDA is not obviously strategyproof (OSP). A mechanism is OSP if for the equilibrium strategy the worst outcome is still weakly better than the best possible outcome from any alternative strategy. Hence, dominance in an OSP mechanism may be easier to detect by agents with cognitive limitations.

Ashlagi and Gonczarowski (2018) show in their Example 1 that the stable serial dictatorship mechanism is in general OSP when the proposing side has homogeneous preferences. However, they show that for general preferences it is impossible to construct an OSP mechanism which generates stable match outcomes. In particular they identify acyclicality in preferences in the sense of Ergin (2002) as a regularity condition that enables implementation of stable matchings with an OSP mechanism.[31]

In Appendix A.III, we build on Example 2 in Ashlagi and Gonczarowski (2018) to demonstrate that a mechanism that is OSP in implementing the student optimal stable allocation when preferences are acyclical may fail to induce stability when students are loss averse. This example sets the two concepts apart and provides a testable prediction which of the two biases induces observed manipulations. Our model conveys students' loss aversion as an alternative explanation for the observed differences, which has nothing to do with cognitive limitation, but comes from the optimal choice when students suffer from this behavioral bias.[32]

4 Conclusion

We have identified a possible reason why students play dominated strategies in the strategyproof direct student-proposing deferred-acceptance mechanism (DSPDA).

[31] In our context, a preference profile for schools over students is cyclical if there are three students A, B, C and two schools $1, 2$ such that $A \succ_1 B \succ_1 C \succ_2 A$, and it is acyclical if it is not cyclical.

[32] In reality both, cognitive limitations and behavioral biases may play an important role for the observed patterns. This is confounded by the fact that in Li (2017) the rate of misrepresenting slightly declines with learning, but remains well above the level of misrepresentations in the SSRDA.

The truthful equilibrium in dominant strategies is not a choice-acclimating personal equilibrium if some students are dominantly expectation-based loss-averse (EBLA), which is suggested by the data. Loss aversion is among the behavioral biases that have been replicated repeatedly in numerous experimental and field studies. The notion that students forgo small chances to get into preferred schools to avoid disappointment is therefore plausible. Indeed, the costly deviations from the dominant truthful strategy are most pervasive among low- and intermediate-priority students who desire to get into competitive programs. Our theoretical predictions fit this pattern in experimental and field data and also provide a formalized framework for the pervasive district-school and small-school biases.

The extensive evidence for dominated play in DSPDA calls into question the identification strategy to treat reported preferences as truthful. Regarding affirmative action this insight is important, because the observation that people of certain demographics do not reveal a preference for certain schools in DSPDA does not imply that they do not want to go there. In fact, we show that groups who are discriminated or perceive to be, indeed, are more likely to misrepresent their preferences. Moreover, we show that DSPDA in conjunction with EBLA discriminates against loss-averse and underconfident students. Since the data suggests that these characteristics are correlated with gender, DSPDA indirectly but inherently implements an imbalanced allocation and amplifies discrimination against already marginalized groups. Importantly, this misallocation problem does not vanish as markets grow large.

We have discussed remedy mechanisms and our theory suggests that sequential mechanisms should outperform static mechanisms in terms of truthfulness. Under conditions on the preferences, a sequential serial dictatorship mechanism delivers the unique stable allocation in dominant strategies and in a truthful choice-acclimating equilibrium, i.e., it succeeds where the celebrated direct student-proposing deferred-acceptance mechanism fails. Indeed, Li (2017) showed that this mechanism outperforms its static version. While he attributes this to obvious strategyproofness, we suggest that reference-dependent preferences drive this difference. That is, it is not obvious whether behavior is driven by a behavior bias (loss aversion) or a cognitive impairment (difficulties to understand the dominance of a strategy). We see the presence of at least some students with non-standard preferences as an undeniable fact and, hence, our paper provides first steps into understanding the former, but more experimental work is needed to disentangle the two.

Appendix

A.I Omitted definitions of Section 2

DSPDA is defined as follows: After all students report their ROLs,

$t = 1$ All students apply to the top-ranked school of their submitted ROL. Each school rejects the least-ranked students in excess of its capacity and tempo-

rarily holds the others.

$t > 1$ All students who were rejected in step $(t-1)$ apply to the highest-ranked school of their submitted ROL that has not rejected them yet. Each school rejects the lowest-ranked students in excess of its capacity from the pool of current applicants. Those who are not rejected are temporarily held.

End The process terminates after the first step without rejections.

Importantly, the mechanism is direct and then all steps are executed mechanically based on the reported ROL.

Properties of mechanisms: An allocation M is a many-to-one mapping from $\mathcal{I}$ to $\mathcal{S}$ such that $M(i) = s$ denotes that student i is matched to school s and $M^{-1}(s) = \{i : M(i) = s\}$ lists the students matched to s. Feasibility requires $|M^{-1}(s)| \leq q_s$. Let $\mathcal{M}$ be the set of all feasible allocations. An allocation rule is a function $\alpha : \mathfrak{S}(\mathcal{S})^n \to \mathcal{M}$, mapping profiles of ROLs into matchings. An allocation rule is strategyproof if

$$v_{i,\alpha(\nu)[i]} \geq v_{i,\alpha(\nu_i',\nu_{-i})[i]} \quad \forall\, i \in \mathcal{I}, \forall \theta. \tag{7}$$

Mechanisms that induce a strategyproof allocation rule with standard utility have the feature that it is a dominant strategy to report preferences truthfully with standard utility. However, the dominance of the truthful strategy may not carry over to settings with non-standard utility as our paper demonstrates.

An allocation M is stable if there are no pair i, s such that

$$v_{i,s} > v_{i,M(i)} \quad \text{and} \quad w_{s,i} > w_{s,i'} \quad \text{for some } i' \in M^{-1}(s), \tag{8}$$

i.e., no student i prefers another school s over her match, while this school prefers i over at least one of her matched students. A student-optimal stable matching is a stable matching M such that

$$v_{i,M(i)} \geq v_{i,M'(i)} \quad \text{for any stable matching } M'. \tag{9}$$

A school-optimal stable matching is defined accordingly.

A static matching mechanism consists of reporting spaces $R = \times_{i\in\mathcal{I}} R_i$ for each student i and an allocation function o, mapping reported profiles $\mathbf{r} = (r_i)_{i\in\mathcal{I}} \in R$ into allocations.

A.II Relation to Dreyfuss et al. (2019)

Similar to our paper, Dreyfuss et al. (2019) find that EBLA can explain non-truthful ROLs observed in the data. In their reduced form dynamic framework à la Kőszegi and Rabin (2009), students enter the decision problem with a reference point given by the outside option, whereas in our decision problem students already anticipate the choices ahead of them, which is reflected in their reference point. Moreover, Dreyfuss et al. (2019) consider an extra period where uncertainty resolves which gives rise to additional gain-loss utilities. The essential intuition

how students use manipulations to shield off potential disappointment is, however, similar in both models.

In our setup, we take the stylized approach that gains and losses are assigned when comparing to the value of other potential outcomes (narrow bracketing). Dreyfuss et al. (2019) take the opposite approach as they consider each school in a separate consumption dimension and assign gains and losses separately for each school. The reality is certainly somewhere in between, as schools may be comparable in some aspects but not in others. We chose this modelling approach to draw a clear comparison to the existing experimental literature, where stakes are simply money, and values are hence fully comparable between schools.[33]

The uncertainty in Dreyfuss et al. (2019) stems from iid. shocks on how individual schools asses a student's abilities with respect to exogenuously given school standards. This reduced form approach has two implications. First, it leaves no scope for strategic interaction between students. Second, it implies that acceptability probabilities are independent between schools, which not the case in our model, not even under Assumption 1 and independently drawn scores.

From the first theoretic insight that under EBLA there is scope for strategic misrepresentations, both papers proceed quite complementarily. While Dreyfuss et al. (2019) comprehensively reevaluate the experimental data in Li (2017) in the light of loss aversion, we delve deeper into the theoretical implications of loss aversion, and analyze the set of rationalizable strategies, strategic interaction, and evaluate alternative mechanisms under loss aversion.

A.III OSP versus EBLA

This section illustrates the distinction of the notion of robustness against EBLA and the concept of OSP. We start with the observation that robustness against EBLA does not imply that a mechanism is OSP. By Proposition 2, a student with EBLA preferences will report truthfully in DSPDA whenever the probability p_1 of being acceptable at her preferred school is sufficiently large. This condition certainly does not imply that DSPDA is obviously strategy proof. Indeed, whenever acceptance probabilities are non-degenerate in the sense that $p_1 \notin \{0, 1\}$ and there is some other school m with positive acceptance probability conditional on being not acceptable at school 1, truth telling will not always induce the most preferred match, whereas any list $(m, 1, ...)$ will do so with positive probability.

Building on Example 2 in Ashlagi and Gonczarowski (2018), we now provide an example of acyclical preferences and an OSP mechanism which always implements the student optimal stable matching, but fails to do so if students exhibit EBLA. There are two students, $I = \{A, B\}$ and two schools, $S = \{1, 2\}$. School 1 prefers student A over B, whereas school 2 prefers student B over A. Conversely, student A prefers school 1 over school 2 with probability $(1 - \epsilon)$, and student B prefers school 2 over school A with probability of $(1 - \epsilon)$ for some small $\epsilon > 0$. Note first

[33] Dreyfuss et al. (2019) take a similar approach in their empirical section when they analyze existing experiments.

that the DSPDA is not obviously strategy proof. Indeed, if, for instance, student 1 prefers school B truth telling is not obviously dominant as it may result in a match with A whereas listing A before B may result in a match with B with positive probability.

Ashlagi and Gonczarowski (2018, Figure 2) propose the following OSP sequential mechanism to obtain truth telling. First, student A is asked whether she prefers 1 or 2. If she prefers 1, she is assigned to 1 and B is assigned to 2. If she prefers 2, B is asked for her preferences which then determine the match outcome. Because B determines the match with certainty whenever she is asked, revealing her true preferences is an obviously dominant strategy (and a SCPE at this final decision node). If A prefers school 1, deviating yields her a lottery over $v_{A,1}$ and $v_{A,2}$ instead of a certain payoff $v_{A,1} > v_{A,2}$ such that the truth is both an SCPE and an obviously dominant strategy.

If A prefers school 2, misrepresenting yields her a sure payoff of $v_{A,1}$ and being truthful yields a lottery with payoff $pv_{A,2} + (1-p)v_{A,2}$. Because even the worst lottery outcome from being truthful is weakly better than the best (only) outcome from deviating, the truth is an obviously dominant strategy, making the mechanism OSP. However, for any $\Lambda > 1$, truthtelling is by Proposition 2 not a SCPE for student A if $\varepsilon < 1 - 1/\Lambda$. As a result, even OSP mechanisms can fail to have a truthful SCPE.

A.IV Proofs

Proof of Lemma 1. The claim follows immediately from the fact the DSPDA is strategyproof for every student preference. Take an arbitrary ROL ν_i for some student i and let s be the highest ranked acceptable school in ν_i.

Suppose that under ν_i the student is matched with s' ranked before s. But then, since she is unacceptable at s', she would prefer ROL ν_i over the true ROL whenever s' is her most preferred school, a contradiction to strategyproofness.

Suppose otherwise that under ν_i she is matched with s'' ranked behind s. But then, if her true ROL was ν_i she would prefer s to s'' which could be achieved by a manipulation which ranks s first, again a contradiction to strategyproofness. □

Proof of Lemma 2. Sufficiency follows from Masatlioglu and Raymond (2016, Proposition 1). For necessity, consider a student $\Lambda_i > 1$. Suppose there is one other student, and they compete for a single seat at some school which is preferred over some safe outside option, whose utility is normalized to zero. The school has homogeneous preferences over students, and scores are independently drawn from $U[0,1]$. Hence, the student's score ω is her acceptability probability at the school. For a match utility $v > 0$ with the school by (5) applying is a CPE for the student if and only if

$$\omega v - \Lambda\omega(1-\omega)v \geq 0.$$

Hence, truthfulness not a CPE if $\omega < 1 - 1/\Lambda$. □

Proof of Lemma 3. We start with a practical lemma which identifies when flipping two neighboring schools in a ROL is profitable. Consider two otherwise identical ROLs swapping two arbitrary adjacently ranked schools $x < y$, i.e., two ROLs $(..., x, y, ...)$ and $(..., y, x, ...)$. Let the former induce lottery $F = (f_s)_{s\in\mathcal{S}}$ and the latter induce lottery $\underline{F} = (\underline{f}_s)_{s\in\mathcal{S}}$, and let ε denote the probability of being acceptable at x and y but at no school which is ranked above x.

Lemma 7. *$U(\cdot, F) \geq U(\cdot, \underline{F})$ if and only if*

$$\frac{\varepsilon}{\Lambda} \geq \varepsilon \left(-\sum_{s=1}^{x} f_s + \varepsilon + \sum_{s=x+1}^{y-1} f_s \, \frac{v_x + v_y - 2v_s}{v_x - v_y} + \sum_{s=y}^{m} f_s \right)$$

with equality only in the case of indifference.

Proof of Lemma 7. We start by rewriting the expression for expected utility in (3).

$$\begin{aligned}
\sum_{s=1}^{m} f_s v_s - \Lambda \sum_{1\leq s\leq r\leq m} f_s f_r (v_s - v_r) &= \sum_{s=1}^{m} f_s v_s - \Lambda \left(\sum_{1\leq s\leq r\leq m} f_s f_r v_s - \sum_{1\leq s\leq r\leq m} f_s f_r v_r \right) \\
&= \sum_{s=1}^{m} f_s v_k - \Lambda \left(\sum_{1\leq s\leq r\leq m} f_s f_r v_s - \sum_{1\leq r\leq s\leq m} f_s f_r v_s \right) \\
&= \sum_{s=1}^{m} f_s v_k - \Lambda \sum_{s=1}^{m} f_s v_s \left(\sum_{r=c+1}^{m} f_r - \sum_{r=1}^{c-1} f_r \right)
\end{aligned}$$

Next, note that for the matching probabilities $\underline{f}_s$ of ROL $(..., y, x, ...)$, it must be that $f_s = \underline{f}_s$ for $s \neq x, y$ and $\underline{f}_x = f_x - \varepsilon$, $\underline{f}_y = f_y + \varepsilon$. This implies $U(\cdot, F) - U(\cdot, \underline{F}) \geq 0$ if and only if

$$0 \leq \sum_{s=1}^{m} f_s v_c \left(1 - \Lambda \left(\sum_{r=c+1}^{m} f_r - \sum_{r=1}^{c-1} f_r \right) \right) - \sum_{s=1}^{m} \underline{f}_s v_c \left(1 - \Lambda \left(\sum_{r=c+1}^{m} \underline{f}_r - \sum_{r=1}^{c-1} \underline{f}_r \right) \right)$$

$$\begin{aligned}
&= \varepsilon(v_x - v_y) - \Lambda \left[\sum_{s=1}^{m} f_s v_s \left(\sum_{r>c} f_r - \sum_{r<c} f_r \right) - \sum_{s=1}^{m} \underline{f}_s v_s \left(\sum_{r>c} \underline{f}_r - \sum_{r<c} \underline{f}_r \right) \right] \\
&= \varepsilon(v_x - v_y) - \Lambda \Bigg[\sum_{c\neq x, c\neq y} f_s v_s \left(\sum_{r>c} (f_r - \underline{f}_r) - \sum_{r<c} (f_r - \underline{f}_r) \right) \\
&\qquad + v_x \left(f_x \left(\sum_{r>x} f_r - \sum_{r<x} f_r \right) - (f_x - \varepsilon) \left(\sum_{r>x} \underline{f}_r - \sum_{r<x} \underline{f}_r \right) \right) \\
&\qquad + v_y \left(f_y \left(\sum_{r>y} f_r - \sum_{r<y} f_r \right) - (f_y + \varepsilon) \left(\sum_{r>y} \underline{f}_r - \sum_{r<y} \underline{f}_r \right) \right) \Bigg] \\
&= \varepsilon(v_x - v_y) - \Lambda \left[\sum_{c<x} f_s v_s \left(-\sum_{r<c} 0 + \sum_{r=c+1}^{x-1} 0 + \varepsilon + \sum_{r=x+1}^{y-1} 0 - \varepsilon + \sum_{r>y} 0 \right) \right.
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{s=x+1}^{y-1} f_s v_s \left(-\sum_{r<x} 0 - \varepsilon - \sum_{r=x+1}^{c-1} 0 + \sum_{r=c+1}^{y-1} 0 - \varepsilon + \sum_{r>y} 0 \right) \\
&+\sum_{s=y+1}^{m} f_s v_s \left(-\sum_{r<x} 0 - \varepsilon - \sum_{r=x+1}^{y-1} 0 + \varepsilon - \sum_{r=y+1}^{c-1} 0 + \sum_{r>c} 0 \right) \\
&+v_x \left(f_x \left(\sum_{r\neq y} 0 - \varepsilon \right) + \varepsilon \left(\sum_{r>x, r\neq y} f_r - \sum_{r<x} f_r + f_y + \varepsilon \right) \right) \\
&+v_y \left(f_y \left(\sum_{r\neq x} 0 - \varepsilon \right) - \varepsilon \left(\sum_{r>y} f_r - \sum_{r<y, r\neq x} f_r - (f_x - \varepsilon) \right) \right) \Bigg]
\end{aligned}
$$

$$
\begin{aligned}
&= \varepsilon(v_x - v_y) - \Lambda \Bigg[\sum_{s=x+1}^{y-1} f_s v_s \left(-2\varepsilon\right) + v_x \varepsilon \left(-f_x + \sum_{r>x, r\neq y} f_r - \sum_{r<x} f_r + f_y + \varepsilon \right) \\
&\qquad - v_y \varepsilon \left(f_y + \sum_{r>y} f_r - \sum_{r<y, r\neq x} f_r - f_x + \varepsilon \right) \Bigg] \\
&= \varepsilon(v_x - v_y) - \Lambda\varepsilon \Bigg[- \sum_{s=x+1}^{y-1} f_s 2v_s + (v_x - v_y) \left(-f_x + f_y + \varepsilon - \sum_{r<x} f_r + \sum_{r>y} f_r \right) \\
&\qquad + (v_x + v_y) \left(\sum_{r=x+1}^{y-1} f_r \right) \Bigg]
\end{aligned}
$$

dividing by $(v_x - v_y) > 0$ yields

$$
\varepsilon - \Lambda\varepsilon \left[\sum_{s=x+1}^{y-1} f_s \frac{v_x + v_y - 2v_s}{v_x - v_y} - f_x + f_y + \varepsilon - \sum_{c<x} f_s + \sum_{c>y} f_s \right] \geq 0,
$$

which yields the result:

$$
\frac{\varepsilon}{\Lambda} \geq \varepsilon \left(\sum_{s=y}^{m} f_s + \varepsilon - \sum_{s=1}^{x} f_s + \sum_{s=x+1}^{y-1} f_s \, \frac{v_x + v_y - 2v_s}{v_x - v_y} \right)
$$

For the second statement, replace all inequalities with equalities. □

We now prove the proposition by contradiction. Suppose there exists a ROL with some $1 \leq k < l < n \leq m$ for which l is listed behind n but k is listed before l and which is strictly preferred to all lists where k is ranked behind l and l behind n. Let n be the least preferred school, i.e., the one with the highest index, for which such a triple exists. For given n and l let k be the lowest-index school, i.e., the most preferred one, satisfying the requirement.

Since k is ranked before l, the optimal ROL is of one of the following forms:

i) $(..., k, ..., n, ..., l, ...)$

ii) $(..., n, ..., k, ..., l, ...)$

We make first considerations for both cases.

i) Since, by assumption, k is the lowest-index school ranked before l, the list must be increasing from k, and eventually decreasing (possibly at l) to a number above k. Call $\overline{x}$ the first school where the list starting from k has decreased. Now, by choosing $\underline{x}$ appropriately in the list between k and $\overline{x}$, we obtain in the optimal ROL a sequence $(..., \underline{x}, \underline{y}, ..., \overline{y}, \overline{x}, ...)$ (with possibly $\underline{y} = \overline{y}$), which is increasing from $\underline{x}$ to $\overline{y}$ and satisfies $\underline{x} < \overline{x} < \underline{y} \leq \overline{y}$.

ii) Since, by assumption, n is the highest-index school for which there exists l and k with l behind n but k before l, the list must be decreasing from n, but eventually increasing (possibly immediately after k) to a number below n. Call $\underline{y}$ the first school after n where the list is increasing. Now, by choosing $\overline{y}$ appropriately in the list between n and $\underline{y}$, we obtain in the optimal ROL a sequence $(..., \overline{y}, \overline{x}, ..., \underline{x}, \underline{y}, ...)$ (with possibly $\overline{x} = \underline{x}$), which is decreasing from $\overline{y}$ to $\underline{x}$ and satisfies $\underline{x} \leq \overline{x} < \underline{y} < \overline{y}$.

The next steps are identical for both cases.

Let f_s be the matching probabilities as induced by the optimal ROL, and let $\overline{f}_s$ be the matching probabilites as induced by the (otherwise identical) ROL that flips $\overline{x}$ and $\overline{y}$.

By the rules of DSPDA, we obtain $f_s = \overline{f}_s$ for all $s \neq \overline{x}, \overline{y}$, and

$$\overline{f}_{\overline{x}} = f_{\overline{x}} + \overline{\varepsilon} \quad \text{and} \qquad \overline{f}_{\overline{y}} = f_{\overline{y}} + \overline{\varepsilon}, \tag{10}$$

where $\overline{\varepsilon}$ is the probability that the student is acceptable at $\overline{x}$ and $\overline{y}$, but not acceptable at any school ranked before $\overline{x}$ and $\overline{y}$ in the optimal ROL.

Since the student is not indifferent, $\varepsilon > 0$ such that the strict optimality of the optimal ROL together with Lemma 7 imply

$$\frac{1}{\Lambda} > -\sum_{s=1}^{\underline{x}} f_s + \underline{\varepsilon} + \sum_{s=\underline{x}+1}^{\underline{y}-1} f_s \, \frac{v_{\underline{x}} + v_{\underline{y}} - 2v_s}{v_{\underline{x}} - v_{\underline{y}}} + \sum_{s=\underline{y}}^{m} f_s$$

Similarly, Lemma 7 implies

$$\begin{aligned} \frac{1}{\Lambda} &< -\sum_{s=1}^{\overline{x}} \overline{f}_s + \overline{\varepsilon} + \sum_{s=\overline{x}+1}^{\overline{y}-1} \overline{f}_s \, \frac{v_{\overline{x}} + v_{\overline{y}} - 2v_s}{v_{\overline{x}} - v_{\overline{y}}} + \sum_{s=\overline{y}}^{m} \overline{f}_s \\ &= -\sum_{s=1}^{\overline{x}} f_s + \sum_{s=\overline{x}+1}^{\overline{y}-1} f_s \, \frac{v_{\overline{x}} + v_{\overline{y}} - 2v_s}{v_{\overline{x}} - v_{\overline{y}}} + \sum_{s=\overline{y}}^{m} f_s - \overline{\varepsilon}. \end{aligned}$$

Hence,

$$-\sum_{s=1}^{\underline{x}} f_s + \underline{\varepsilon} + \sum_{s=\underline{x}+1}^{\underline{y}-1} f_s \frac{v_{\underline{x}} + v_{\underline{y}} - 2v_s}{v_{\underline{x}} - v_{\underline{y}}} + \sum_{s=\underline{y}}^{m} f_s < -\sum_{s=1}^{\overline{x}} f_s + \sum_{s=\overline{x}+1}^{\overline{y}-1} f_s \frac{v_{\overline{x}} + v_{\overline{y}} - 2v_s}{v_{\overline{x}} - v_{\overline{y}}} + \sum_{s=\overline{y}}^{m} f_s - \overline{\varepsilon},$$

which can be rearranged to

$$\underline{\varepsilon} + \overline{\varepsilon} + \sum_{s=\underline{x}+1}^{y} f_s \alpha_s + \sum_{s=\overline{x}+1}^{\underline{y}-1} f_s \beta_s + \sum_{\underline{y}}^{\overline{y}-1} f_s \gamma_s < 0.$$

This yields a contradiction as each summand on the left hand side is positive. Indeed, we have

$$\alpha_s \equiv 1 + \frac{v_{\underline{x}} + v_{\underline{y}} - 2v_s}{v_{\underline{x}} - v_{\underline{y}}} = \frac{2(v_{\underline{x}} - v_s)}{v_{\underline{x}} - v_{\underline{y}}} > 0,$$

because $v_{\underline{x}} > v_s$ as all $s > \underline{x}$ in the sum and $v_{\underline{x}} > v_{\underline{y}}$ by definition. Next,

$$\gamma_s \equiv 1 + \frac{2v_s - v_{\overline{x}} - v_{\overline{y}}}{v_{\overline{x}} - v_{\overline{y}}} = \frac{2(v_s - v_{\overline{y}})}{v_{\overline{x}} - v_{\overline{y}}} > 0,$$

because $v_s > v_{\overline{y}}$ as all $s < \overline{y}$ in the sum and $v_{\overline{x}} > v_{\overline{y}}$ by definition. Finally,

$$\begin{aligned}
\beta_s &\equiv \frac{v_{\underline{x}} + v_{\underline{y}} - 2v_s}{v_{\underline{x}} - v_{\underline{y}}} + \frac{2v_s - v_{\overline{x}} - v_{\overline{y}}}{v_{\overline{x}} - v_{\overline{y}}} \\
&= \frac{(v_{\overline{x}} - v_{\overline{y}})(v_{\underline{x}} + v_{\underline{y}} - 2v_s) + (v_{\underline{x}} - v_{\underline{y}})(2v_s - v_{\overline{x}} - v_{\overline{y}})}{(v_{\underline{x}} - v_{\underline{y}})(v_{\overline{x}} - v_{\overline{y}})} \\
&= \frac{2(v_s(v_{\underline{x}} - v_{\overline{x}} + v_{\overline{y}} - v_{\underline{y}}) + v_{\overline{x}} v_{\underline{y}} - v_{\underline{x}} v_{\overline{y}})}{(v_{\underline{x}} - v_{\underline{y}})(v_{\overline{x}} - v_{\overline{y}})} = \frac{num}{denom},
\end{aligned}$$

which is positive if the numerator is positive as the denominator is positive by definition. Because $v_{\underline{x}} > v_{\overline{x}}$ by definition and $v_{\underline{y}} < v_s$ for all s in the sum, the added term below is negative. Hence,

$$\begin{aligned}
num &> 2(v_s(v_{\underline{x}} - v_{\overline{x}} + v_{\overline{y}} - v_{\underline{y}}) + v_{\overline{x}} v_{\underline{y}} - v_{\underline{x}} v_{\overline{y}} + (v_{\underline{x}} - v_{\overline{x}})(v_{\underline{y}} - v_s)) = \\
&= 2((v_{\underline{x}} - v_s)(v_{\underline{y}} - v_{\overline{y}})) > 0
\end{aligned}$$

because $v_{\underline{x}} > v_s$ and $v_{\underline{y}} > v_{\overline{y}}$ by assumption. □

Proof of Proposition 1. The part on the decreasing order is just Lemma 3. For the increasing part, suppose otherwise that for some $k < l < n$ any optimal ROL is of form $(k, ..., n, ..., l, ...)$. This is a contradiction of Lemma 3 as for such n and l there is an optimal ROL which ranks k behind l. □

Proof of Proposition 2. 1. Suppose, by way of contradiction, that truth telling is suboptimal. By Corollary 1, this implies that school 1 is not ranked first. Let k be the school ranked in front of 1. Since the optimal ROL $(..., k, 1, ...)$ must be preferred to the list $(..., 1, k, ...)$ where we flip school 1 and k, Lemma 7 implies

$$\frac{\varepsilon}{\Lambda} < \varepsilon \left(-\sum_{s=1}^{1} f_s + \varepsilon + \sum_{s=2}^{k-1} f_s \frac{v_1 + v_k - 2v_s}{v_1 - v_k} + \sum_{s=k}^{m} f_s \right),$$

where f_s are the matching probabilities as induced by ROL $(..., 1, k, ...)$, and ε the probability that the student is acceptable at 1 and k but no higher ranked school. Since by full support we have $\varepsilon > 0$ and further $\frac{v_1+v_k-2v_s}{v_1-v_k} = 1 - 2\frac{v_s-v_k}{v_1-v_k} < 1$ this implies

$$\frac{1}{\Lambda} < -f_1 + \varepsilon + \sum_{s=2}^{m} f_s \leq -f_1 + f_1 + (1 - f_1) = 1 - f_1 \leq 1 - p_1,$$

hence $p_1 < 1 - \frac{1}{\Lambda}$, a contradiction.

2. Suppose, by way of contradiction that, truth telling is optimal. Hence $U((1, 2, ..., m) \geq U((2, 1, ..., m))$, and by Lemma 7

$$\frac{1}{\Lambda} \geq -f_1 + \varepsilon + \sum_{s=2}^{m} f_s = -f_1 + \varepsilon + (1 - f_1) > 1 - 2f_1 = 1 - 2p_1,$$

which can be rearranged to $p_1 > 0.5\left(1 - \frac{1}{\Lambda}\right)$, a contradiction. □

Proof of Proposition 3. The fact that all schools use the same deterministic tie-breaking rule for students of same score can be interpreted as an assumption that no two students share the same score. The existence of a pure strategy equilibrium then follows iteratively. Start with the student with the highest possible score. Since she has priority over all other students she infers that she will be accepted at any school and submits (according to the fixed tie-breaking rule) a ROL that lists her most preferred school (which certainly depends on $\mathbf{v}_i$) first. Next, consider the student and her preference profiles that exhibit the second highest possible score. If this is the same student she knows again that no other student has a higher score and will report the same ROL. If this is another student she infers correctly the probability that another student has a higher score and the probability distribution over her submitted ROLs. From that she infers correctly the distribution over her acceptability probabilities and picks her best response ROL, depending on her type. Continuing that procedure iteratively through all possible scores gives us a pure strategy CBNE which is unique when fixing how indifference is broken at each decision node. □

Proof of Lemma 4. By (5), a ROL which lists any subset of elite schools above the outside option induces an expected utility of $fv - \Lambda f(1 - f)v$, where f is the probability of being acceptable at at least one elite school of the subset. Since the utility is a U-shaped function in f, it is maximized by either maximizing or minimizing f. Hence, by either listing all or none of the elite schools above the outside option. □

Proof of Lemma 5. Suppose there are $|\mathcal{I}| = n$ students. The probability that there are less than q among $n - 1$ students with a score above w,

$$P(\omega) := \sum_{k=0}^{q-1} \binom{n-1}{k} (1 - \omega)^k \omega^{n-1-k},$$

is continuously and monotonically increasing in ω from 0 to 1. Since $1 - 1/\Lambda^l \in (0,1)$, there is by the intermediate value theorem a unique $\overline{\omega}(\Lambda^l)$ such that $P(\overline{\omega}(\Lambda^l)) = 1 - 1/\Lambda^l$. Hence, for any $\Lambda \leq \Lambda^l$ and any $\omega \geq \overline{\omega}(\Lambda^l)$ we have $\Lambda \leq \frac{1}{1-P(\omega)}$ meaning that applying to the elite is a best response for all types with $\omega \geq \overline{\omega}(\Lambda^l)$, even if all other students apply as well. As a student of type Λ^l infers that all students of score $\omega \geq \overline{\omega}(\Lambda^l)$ apply she infers that for score $\omega \geq \overline{\omega}(\Lambda^l)$ she has acceptability probability $f(\omega) = P(\omega)$ and by construction applies if and only if her score satisfies $\omega \geq \overline{\omega}(\Lambda^l)$. Next, a student of score $\omega < \overline{\omega}(\Lambda^l)$ infers that she is acceptable if there are less than q other students with either score above $\overline{\omega}(\Lambda^l)$ or score in $[\omega, \overline{\omega}(\Lambda^l)]$ and $\Lambda \neq \Lambda^l$. Again, this probability is strictly and continuously increasing in ω which implies a unique cutoff $\overline{\omega}(\Lambda^{l-1})$ such that $f(\overline{\omega}(\Lambda^{l-1})) = \frac{1}{1-\Lambda^{l-1}}$. Hence, truthful reporting for type Λ^{l-1} is optimal if and only if $\omega > \overline{\omega}\Lambda^{l-1}$. Proceeding this manner iteratively we obtain an essentially unique CBNE. □

Proof of Proposition 4. 1. A student with Λ^H and score ω applies in a setting of n students by Equation 5 if and only if her acceptability probability $F_n(\omega)$ satisfies $F_n(\omega)v - \Lambda^H F_n(\omega)(1 - F_n(\omega))v \geq 0$, i.e. if and only if $F_n(\omega) \geq 1 - \frac{1}{\Lambda^H}$. She is acceptable if and only if the the q-th highest score of $n-1$ other students is below ω. Hence, F_n is the cdf of the q-th highest order statistic of $n-1$ draws from the uniform distribution on $[0,1]$ (i.e. F_n describes a beta-distribution with parameters $Beta(\omega, n-q, q)$). We define as $\overline{\omega}_n = F_n^{-1}(1 - 1/\Lambda^H)$ as the cutoff score below which a student with Λ^H does not apply.

Next, note that justified envy occurs if and only if the q-th highest of n scores is below $\overline{\omega}_n$ and at least one of the q students with the highest score has a score below $\overline{\omega}$ and is of type Λ^H, as such a student doesn't apply but would be acceptable. For $\alpha = 1$ this is obviously the case if and only if the q-th highest score is below $\overline{\omega}$. For $\alpha < 1$ it is sufficient but not necessary that the q-th highest score is below $\overline{\omega}$ and is of type Λ^H. Since the q-th highest of n scores is below $\overline{\omega}_n$ if and only if q-th highest of the $n-1$ scores is below $\overline{\omega}_N$ and an n-th additional score is below $\overline{\omega}_n$, we obtain $\alpha F_n(\overline{\omega}_n)\overline{\omega}_N$ as a lower bound for the probability of justified envy, with equality only for $\alpha = 1$. Since for $n \to \infty$ we have $\overline{\omega}_n \to 1$ we obtain that the probability for a stable allocation for $n \to \infty$ is bounded from below by $\alpha F_n(\overline{\omega}_n) = \alpha(1 - \frac{1}{\Lambda})$.

2. We derive a closed form formula for the ex-ante risk of a student to suffer from justified envy. Again, this is the case if she is of type Λ^H, her score is below cutoff $\overline{\omega}$ but above the q-th highest of $n-1$ other scores. Hence, the ex-ante probability for the student to suffer from justified envy is given by

$$\begin{aligned}\mathbb{P}_{\text{envy}} &= \alpha \int_0^{\overline{\omega}} f_n(x)(\overline{\omega}_n - x)dx \\ &= \alpha F_n(\overline{\omega}_n)\Big(\overline{\omega}_n - \mathbb{E}[X_{q,n-1} | X_{q,n-1} \leq \overline{\omega}_n]\Big),\end{aligned}$$

where f_n is the density of F_n and $X_{q,n-1}$ is the q-th highest order statistic of $n-1$

draws. Hence, the expected number of people suffering from justified envy is

$$\alpha(1-1/\Lambda^H)n\Big(\overline{\omega}_n - \mathbb{E}[X_{q,n-1}|X_{q,n-1}\leq\overline{\omega}_n]\Big)$$

Next we see that

$$\begin{aligned}&\lim_{n\to\infty} n\Big(\overline{\omega}_n - \mathbb{E}[X_{q,n-1}|X_{q,n-1}\leq\overline{\omega}_n]\Big)\\ =&\lim_{n\to\infty} n\Big(\mathbb{E}[1-X_{q,n-1}|1-X_{q,n-1}\geq F^{-1}_{1-X_{q,n-1}}(1/\Lambda)] - F^{-1}_{1-X_{q,n-1}}(1/\Lambda)\Big),\end{aligned}$$

where $F^{-1}_{1-X_{q,n-1}}$ is the quantile function of $1-X_{q,n-1}$.

Now, since $1-X_{q,n-1}$ is distributed according to $Beta(q,n-q)$ and it holds that $\lim_{n\to\infty} nBeta(q,n) = Gamma(q,1)$ (where equality means equality in distribution), we obtain

$$\begin{aligned}&\lim_{n\to\infty} n\Big(\overline{\omega}_n - \mathbb{E}[X_{q,n-1}|X_{q,n-1}\leq\overline{\omega}_n]\Big)\\ =&\mathbb{E}[Y_q|Y_q\geq G_q^{-1}(1/\Lambda)] - G_q^{-1}(1/\Lambda),\end{aligned}$$

where G_q^{-1} is the quantile function of $Gamma(q,1)$. □

Proof of Proposition 5. Obviously, if the DSPDA is truthful, it implements the student optimal stable allocations for all realizations of preferences. For the converse, take any static mechanism (R,o) that implements the student optimal outcome as CBNE. More precisely, for each student i there exists a strategy $\sigma_i : \mathfrak{S}(\mathcal{S}) \to R_i$ such that the joint strategy profile is a CBNE given o. Consequently, for the associated direct mechanism $\Big(\prod \mathfrak{S}(\mathcal{S}), o\circ(\sigma_1,...,\sigma_n)\Big)$ truthfulness is by construction a CBNE, and it implements the student optimal stable allocation. Hence, DSPDA is truthful. □

Proof of Lemma 6. Suppose the stable match is not unique. Then the school-optimal (student-pessimal) stable match M_{sp} is different to the student-optimal stable match M_{so}. For convenience of notation we rename for some representative student i_n the assigned schools as $s_n = M_{so}(i_n)$ and $\tilde{s}_n = M_{sp}(i_n)$. By Lemma 1 in Roth (1986) the number of vacant seats at each school is the same under any stable match. Hence, any student i_1 who is placed differently under both matches necessarily replaces a student i_2 at school s_1 under M_{so} compared to M_{sp}. Then, necessarily i_2 replaces some student i_3 at school s_2, and so on. Since the number of students is finite, continuing iteratively this cascade necessarily generates a cycle where $i_n = i_1$ for some n. Since every student weakly prefers her match under the student-optimal outcome and preferences are by assumption strict we have $v_{i,s_i} > v_{i,s_{i-1}}$ for $i\in\{2,...,n\}$. By aligned preferences this implies $w_{i,s_i} > w_{i,s_{i-1}}$ for $i\in\{2,...,n\}$. Since, contrary, each school prefers the students under match

M_{sp} we obtain analogously $w_{i,s_{i-1}} > w_{i-1,s_{i-1}}$ for all $i \in \{2, ..., n\}$. Together we obtain

$$w_{i_1,s_1} < w_{i_2,s_1} < w_{i_2,s_2} < w_{i_3,s_2} < ... < w_{i_n,s_n} = w_{i_1,s_1},$$

a contradiction. □

Proof of Proposition 6. Fix some student and suppose that all other students behave truthfully in the mechanism. We start by showing that a unique change in her the decision of accepting or rejecting a unique offer from some school at some stage during the mechanism when she doesn't currently hold an offer (while fixing the strategy for all other stages) does not change the probability of receiving an offer from a more preferred school. If the student rejects rather than accepts a school then all students will receive weakly more proposals. Indeed, the rejected school will potentially send an additional proposal to another student in the next step. If she rejects the proposal the school will offer the seat yet to another student the subsequent round. If she accepts this proposal in favor of another school then this other school will offer the seat in the next round. By iterating this argument we can conclude due to the initial rejection all students—including the initial student—obtain weakly more proposals. Hence, a rejection can only weakly increase the probability of receiving an offer from a more preferred school. However, if this increase were strict and the student accepted all proposals by more preferred schools then the rejection could generate a stable match outcome which were weakly preferred by all students. Since preferences are aligned this contradicts the uniqueness of a stable match as derived in Lemma 6. Hence, the decision of the student to accept or reject a school can only affect acceptability for schools that she prefers less.

Next, we iterate the above argument, and show that accepting or rejecting some school k at some stage when currently holding or simultaneously receiving an offer from some school ℓ does not change the probability of receiving an offer from any school preferred to $\min\{\ell, k\}$. Note that rejecting ℓ when it proposes and accepting k when it proposes at a later stage induces the same match outcome as accepting ℓ at first, but only rejecting it in favor of k when an offer from k occurs, since both strategies induce the same cascade of school proposals and students are truthful. Further, by the above finding, rejecting both schools instead does not change the probability of proposals from schools preferred to k. From here we apply the above finding once more and obtain the same proposal probabilities for schools preferred to ℓ when accepting ℓ and rejecting k. Hence, the result that accepting or rejecting k in favor of ℓ does not change proposal probabilities from any school preferred to $\min\{\ell, k\}$.

Now, suppose that truthfulness is not an SCPE for the student. Going backwards in the decision tree of the student, take a decision node where truthfulness is suboptimal but such that it is optimal in all possible future decision nodes. We call k the school that offers a seat to the student at that decision node. Let $F = (f_1, ..., f_n)$ be the lottery over match outcomes if the student rejects k, and let $\tilde{F} = (\tilde{f}_1, ..., \tilde{f}_n)$ be the respective lottery if she accepts. We distinguish three

cases:

1. The student currently holds an offer from a school $\ell < k$, hence rejecting would be truthful. Since, by assumption, truthfulness is optimal in all subsequent decision nodes and the probability of proposals from schools $i < \ell$ does not depend on the decision about k, we have $f_i = \tilde{f}_i$ for $i < \ell$, and $f_\ell = \sum_{i=\ell}^{k} \tilde{f}_i$, while $f_i = \tilde{f}_i = 0$ for $i > k$. Hence, according to (5), we have

$$\begin{aligned}
U(\cdot, F) &= \sum_{i=1}^{\ell} f_i v_i - \Lambda \sum_{1 \le i \le j \le \ell} f_i f_j (v_i - v_j) \\
&= \sum_{i=1}^{\ell-1} \tilde{f}_i v_i + \sum_{i=\ell}^{k} \tilde{f}_i v_\ell - \Lambda \left(\sum_{1 \le i \le j \le \ell-1} \tilde{f}_i \tilde{f}_j (v_i - v_j) + \sum_{1 \le i \le \ell-1} \left(\sum_{j=\ell}^{k} \tilde{f}_j \right) \tilde{f}_i (v_i - v_\ell) \right) \\
&\ge \sum_{i=1}^{k} \tilde{f}_i v_i - \Lambda \left(\sum_{1 \le i \le j \le \ell-1} \tilde{f}_i \tilde{f}_j (v_i - v_j) + \sum_{\ell \le j \le k,\ 1 \le i \le j} \tilde{f}_i \tilde{f}_j (v_i - v_j) \right) \\
&= U(\cdot, \tilde{F}),
\end{aligned}$$

and truthfulness is optimal, a contradiction.

2. The student currently holds no offer from a school preferred to k, and v_k is above the utility of the outside option, hence accepting would be truthful. Again, since the decision doesn't change the match probability with more preferred schools we have $f_i = \tilde{f}_i$ for all $i < k$. Moreover, accepting gives a certain payoff of v_k in case no better offer occurs, whereas rejecting leaves the student with some payoff $v_{k+1}, ..., v_n$ in that case. Hence,

$$\begin{aligned}
U(\cdot, F) &= \sum_{i=1}^{n} f_i v_i - \Lambda \sum_{1 \le i \le j \le n} f_i f_j (v_i - v_j) \\
&\le \sum_{i=1}^{k-1} f_i v_i + \sum_{i=k+1}^{n} f_i v_k - \Lambda \left(\sum_{1 \le i \le j \le k-1} f_i f_j (v_i - v_j) + \sum_{1 \le i \le k-1} \left(\sum_{j=k+1}^{n} f_j \right) f_i (v_i - v_k) \right) \\
&= \sum_{i=1}^{k-1} \tilde{f}_i v_i + \tilde{f}_k v_k - \Lambda \sum_{1 \le i \le j \le k} f_i f_j (v_i - v_j) \\
&= U(\cdot, \tilde{F}),
\end{aligned}$$

and truthfulness is optimal, a contradiction.

3. The student currently holds no offer from a school preferred to k, and v_k is below the utility of the certain outside option, hence rejecting would be truthful. If there is yet with certainty an acceptable to come then accepting and rejecting gives rise to the same lottery, hence truthfulness is optimal. Otherwise, expected utility is below the utility obtained from receiving any acceptable offer with certainty. By going backwards through the decision tree take the first node where an offer with utility above the outside option was received. (This must exist, as the student has a certain outside option.) By presumption, this offer was rejected, as it is preferred

to k. However, rejecting this offer is not in line with CPE behavior as we showed in (ii) that in such circumstances accepting was optimal, a contradiction. □

A.V Data from Li (2017)

	PRIORITY SCORES																			
ROLs	1		2		3		4		5		6		7		8		9		10		ALL	
1234	55	61.1%	48	57.1%	47	58.8%	42	67.7%	32	55.2%	49	79.0%	58	74.4%	48	85.7%	59	84.3%	73	91.3%	511	71.0%
1243	1	1.1%	1	1.2%	1	1.3%	0	0.0%	0	0.0%	1	1.6%	1	1.3%	0	0.0%	1	1.4%	0	0.0%	6	0.8%
1324	2	2.2%	3	3.6%	2	2.5%	1	1.6%	2	3.4%	0	0.0%	1	1.3%	0	0.0%	1	1.4%	0	0.0%	12	1.7%
1342	1	1.1%	0	0.0%	0	0.0%	0	0.0%	1	1.7%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	1	1.3%	3	0.4%
1423	0	0.0%	1	1.2%	0	0.0%	1	1.6%	1	1.7%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	3	0.4%
1432	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	1	1.3%	1	0.1%
2134	1	1.1%	1	1.2%	3	3.8%	**4**	**6.5%**	**7**	**12.1%**	**5**	**8.1%**	**8**	**10.3%**	**4**	**7.1%**	**4**	**5.7%**	1	1.3%	**38**	**5.3%**
2143	0	0.0%	1	1.2%	3	3.8%	0	0.0%	1	1.7%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	1	1.3%	6	0.8%
2314	1	1.1%	2	2.4%	2	2.5%	1	1.6%	2	3.4%	1	1.6%	0	0.0%	2	3.6%	1	1.4%	1	1.3%	13	1.8%
2341	0	0.0%	0	0.0%	0	0.0%	2	3.2%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	1	1.4%	0	0.0%	3	0.4%
2413	0	0.0%	1	1.2%	2	2.5%	0	0.0%	0	0.0%	0	0.0%	2	2.6%	0	0.0%	0	0.0%	0	0.0%	5	0.7%
2431	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	1	1.6%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	1	0.1%
3124	1	1.1%	2	2.4%	2	2.5%	1	1.6%	3	5.2%	0	0.0%	4	5.1%	0	0.0%	1	1.4%	0	0.0%	14	1.9%
3142	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
3214	6	6.7%	5	6.0%	**6**	**7.5%**	3	4.8%	2	3.4%	0	0.0%	0	0.0%	1	1.8%	0	0.0%	0	0.0%	23	3.2%
3241	0	0.0%	0	0.0%	1	1.3%	0	0.0%	0	0.0%	0	0.0%	1	1.3%	0	0.0%	0	0.0%	0	0.0%	2	0.3%
3412	0	0.0%	0	0.0%	1	1.3%	0	0.0%	2	3.4%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	3	0.4%
3421	3	3.3%	2	2.4%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	5	0.7%
4123	1	1.1%	2	2.4%	1	1.3%	0	0.0%	1	1.7%	2	3.2%	1	1.3%	0	0.0%	0	0.0%	1	1.3%	9	1.3%
4132	0	0.0%	1	1.2%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	1	0.1%
4213	1	1.1%	1	1.2%	0	0.0%	1	1.6%	3	5.2%	1	1.6%	1	1.3%	0	0.0%	0	0.0%	0	0.0%	8	1.1%
4231	1	1.1%	2	2.4%	2	2.5%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	1	1.8%	0	0.0%	0	0.0%	6	0.8%
4312	0	0.0%	4	4.8%	4	5.0%	3	4.8%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	1	1.3%	12	1.7%
4321	**16**	**17.8%**	**7**	**8.3%**	3	3.8%	3	4.8%	1	1.7%	2	3.2%	1	1.3%	0	0.0%	2	2.9%	0	0.0%	35	4.9%
Total	90	100.0%	84	100.0%	80	100.0%	62	100.0%	58	100.0%	62	100.0%	78	100.0%	56	100.0%	70	100.0%	80	100.0%	720	100.0%
misrep'	35	38.9%	36	42.9%	33	41.3%	20	32.3%	26	44.8%	13	21.0%	20	25.6%	8	14.3%	11	15.7%	7	8.8%	209	29.0%
CPE	82	91.1%	65	77.4%	62	77.5%	55	88.7%	44	75.9%	57	91.9%	68	87.2%	55	98.2%	67	95.7%	75	93.8%	630	87.5%

Table 5: Absolute and relative frequency of all ROLs for each priority score in the experiment by Li (2017). The CPE-rationalizable ROLs are marked, and the frequencies of the most common misrepresentations for each priority score are in bold.

References

Abeler, J., Falk, A., Goette, L., Huffman, D., 2011. Reference points and effort provision. American Economic Review 101 (2), 470–92.

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

Artemov, G., Che, Y.-K., He, Y., 2017. Strategic 'mistakes': Implications for market design research. Mimeo, University of Melbourne.

Ashlagi, I., Gonczarowski, Y. A., 2018. Stable matching mechanisms are not obviously strategy-proof. Journal of Economic Theory 177, 405–425.

Balinski, M., Sönmez, T., 1999. A tale of two mechanisms: student placement. Journal of Economic theory 84 (1), 73–94.

Barber, B. M., Odean, T., 02 2001. Boys will be Boys: Gender, Overconfidence, and Common Stock Investment. The Quarterly Journal of Economics 116 (1), 261–292.

Basteck, C., Mantovani, M., 2018. Cognitive ability and games of school choice. Games and Economic Behavior 109, 156–183.

Bó, I., Hakimov, R., forthcoming. Iterative versus standard deferred acceptance: Experimental evidence. The Economic Journal.

Carbajal, J. C., Ely, J. C., 2016. A model of price discrimination under loss aversion and state-contingent reference points. Theoretical Economics 11 (2), 455–485.

Chen, L., Pereyra, J. S., 2019. Self-selection in school choice. Games and Economic Behavior.

Chen, Y., Sönmez, T., 2006. School choice: an experimental study. Journal of Economic theory 127 (1), 202–231.

Crawford, V. P., Meng, J., 2011. New york city cab drivers' labor supply revisited: Reference-dependent preferences with rational-expectations targets for hours and income. American Economic Review 101 (5), 1912–32.

Dato, S., Grunewald, A., Müller, D., Strack, P., 2017. Expectation-based loss aversion and strategic interaction. Games and Economic Behavior 104, 681–705.

Ding, T., Schotter, A., 2017. Matching and chatting: An experimental study of the impact of network communication on school-matching mechanisms. Games and Economic Behavior 103, 94–115.

Dreyfuss, B., Heffetz, O., Rabin, M., 2019. Expectations-based loss aversion may help explain seemingly dominated choices in strategy-proof mechanisms. Available at SSRN 3381244.

Dubins, L. E., Freedman, D. A., 1981. Machiavelli and the gale-shapley algorithm. The American Mathematical Monthly 88 (7), 485–494.

Eliaz, K., Spiegler, R., 2014. Reference dependence and labor market fluctuations. NBER macroeconomics annual 28 (1), 159–200.

Ergin, H. I., 2002. Efficient resource allocation on the basis of priorities. Econometrica 70 (6), 2489–2497.

Ericson, K. M. M., Fuster, A., 2011. Expectations as endowments: Evidence on reference-dependent preferences from exchange and valuation experiments. The Quar-

terly Journal of Economics 126 (4), 1879–1907.

Fack, G., Grenet, J., He, Y., 2019. Beyond truth-telling: Preference estimation with centralized school choice and college admissions. American Economic Review 109 (4), 1486–1529.

Fernandez, M., 2018. Deferred acceptance and regret-free truthtelling: A characterization result.

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

Gill, D., Prowse, V., 2012. A structural analysis of disappointment aversion in a real effort competition. American Economic Review 102 (1), 469–503.

Gneezy, U., Goette, L., Sprenger, C., Zimmermann, F., 2017. The limits of expectations-based reference dependence. Journal of the European Economic Association 15 (4), 861–876.

Gneezy, U., List, J. A., Wu, G., 2006. The uncertainty effect: When a risky prospect is valued less than its worst possible outcome. The Quarterly Journal of Economics 121 (4), 1283–1309.

Hakimov, R., Kübler, D., 2019. Experiments on matching markets: A survey. Tech. rep., WZB Discussion Paper.

Hassidim, A., Marciano, D., Romm, A., Shorrer, R. I., 2017a. The mechanism is truthful, why aren't you? American Economic Review 107 (5), 220–24.

Hassidim, A., Romm, A., Shorrer, R. I., 2017b. Redesigning the israeli psychology master's match. American Economic Review 107 (5), 205–09.

Heffetz, O., June 2018. Are reference points merely lagged beliefs over probabilities? Working Paper 24721, National Bureau of Economic Research.

Heffetz, O., List, J. A., 2014. Is the endowment effect an expectations effect? Journal of the European Economic Association 12 (5), 1396–1422.

Heidhues, P., Kőszegi, B., 2008. Competition and price variation when consumers are loss averse. American Economic Review 98 (4), 1245–68.

Heidhues, P., Kőszegi, B., 2014. Regular prices and sales. Theoretical Economics 9 (1), 217–251.

Herweg, F., 2013. The expectation-based loss-averse newsvendor. Economics Letters 120 (3), 429–432.

Herweg, F., Mierendorff, K., 2013. Uncertain demand, consumer loss aversion, and flat-rate tariffs. Journal of the European Economic Association 11 (2), 399–432.

Herweg, F., Müller, D., Weinschenk, P., 2010. Binary payment schemes: Moral hazard and loss aversion. American Economic Review 100 (5), 2451–77.

Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk. Econometrica 47 (2), 263–291.

Karle, H., Engelmann, D., Peitz, M., 2019. Student performance and loss aversion. Tech. rep., working paper.

Karle, H., Kirchsteiger, G., Peitz, M., April 2015. Loss aversion and consumption choice: Theory and experimental evidence. American Economic Journal: Microeconomics 7 (2), 101–20.

URL http://www.aeaweb.org/articles?id=10.1257/mic.20130104

Karle, H., Peitz, M., 2014. Competition under consumer loss aversion. The RAND Journal of Economics 45 (1), 1–31.

Kőszegi, B., Rabin, M., 2006. A model of reference-dependent preferences. The Quarterly Journal of Economics 121 (4), 1133–1165.

Kőszegi, B., Rabin, M., 2007. Reference-dependent risk attitudes. American Economic Review 97 (4), 1047–1073.

Kőszegi, B., Rabin, M., 2009. Reference-dependent consumption plans. American Economic Review 99 (3), 909–36.

Lange, A., Ratan, A., 2010. Multi-dimensional reference-dependent preferences in sealed-bid auctions–how (most) laboratory experiments differ from the field. Games and Economic Behavior 68 (2), 634–645.

Li, S., 2017. Obviously strategy-proof mechanisms. American Economic Review 107 (11), 3257–87.

Masatlioglu, Y., Raymond, C., 2016. A behavioral analysis of stochastic reference dependence. American Economic Review 106 (9), 2760–82.

Milgrom, P. R., Weber, R. J., 1985. Distributional strategies for games with incomplete information. Mathematics of operations research 10 (4), 619–632.

Myerson, R. B., 1979. Incentive compatibility and the bargaining problem. Econometrica: journal of the Econometric Society, 61–73.

Niederle, M., Vesterlund, L., 08 2007. Do Women Shy Away From Competition? Do Men Compete Too Much? The Quarterly Journal of Economics 122 (3), 1067–1101.

Pais, J., Pintér, Á., 2008. School choice and information: An experimental study on matching mechanisms. Games and Economic Behavior 64 (1), 303–328.

Pathak, P. A., Sönmez, T., 2008. Leveling the playing field: Sincere and sophisticated players in the boston mechanism. American Economic Review 98 (4), 1636–52.

Pathak, P. A., Sönmez, T., 2013. School admissions reform in chicago and england: Comparing mechanisms by their vulnerability to manipulation. American Economic Review 103 (1), 80–106.

Pope, D. G., Schweitzer, M. E., 2011. Is tiger woods loss averse? persistent bias in the face of experience, competition, and high stakes. American Economic Review 101 (1), 129–57.

Rabin, M., Weizsäcker, G., 2009. Narrow bracketing and dominated choices. American Economic Review 99 (4), 1508–43.

Rees-Jones, A., 2018. Suboptimal behavior in strategy-proof mechanisms: Evidence from the residency match. Games and Economic Behavior 108, 317–330.

Rees-Jones, A., Skowronek, S., 2018. An experimental investigation of preference misrepresentation in the residency match. Proceedings of the National Academy of Sciences 115 (45), 11471–11476.

Rosato, A., 2014. Loss aversion in sequential auctions: Endogenous interdependence, informational externalities and the" afternoon effect". Mimeo, University of Technology Sydney Business School.

Rosato, A., 2017. Sequential negotiations with loss-averse buyers. European Economic

Review 91, 290–304.

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

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

Rubinstein, A., 1991. Comments on the interpretation of game theory. Econometrica: Journal of the Econometric Society, 909–924.

Rydval, O., Ortmann, A., Prokosheva, S., Hertwig, R., 2009. How certain is the uncertainty effect? Experimental Economics 12 (4), 473–487.

Shorrer, R., Sóvágó, S., 2017. Obvious mistakes in a strategically simple college admissions environment.

Sprenger, C., 2015. An endowment effect for risk: Experimental tests of stochastic reference points. Journal of Political Economy 123 (6), 1456–1499.

Tversky, A., Kahneman, D., 1981. The framing of decisions and the psychology of choice. Science 211 (4481), 453–458.

Tversky, A., Kahneman, D., 1992. Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and uncertainty 5 (4), 297–323.

von Wangenheim, J., 2017. English versus vickrey auctions with loss averse bidders. Tech. rep., CRC TRR 190 Rationality and Competition.