Discussion Paper Series – CRC TR 224

Discussion Paper No. 312
Project B 03

School Choice and Loss Aversion

Vincent Meisner [1]
Jonas von Wangenheim [2]

July 2021

[1] Technical University Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany, vincent.meisner@tuberlin.de
[2] University of Bonn, Institute for Microeconomics, Adenauerallee 24-42, 53113 Bonn, Germany, jwangenheim@uni-bonn.de

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

School choice and loss aversion*

Vincent Meisner[†] Jonas von Wangenheim[‡]

July 16, 2021

Abstract

Evidence suggests that participants in direct student-proposing deferred-acceptance mechanisms (DSPDA) play dominated strategies. To explain the observed data, we introduce expectation-based loss aversion into a school-choice setting and characterize choice-acclimating personal equilibria in DSPDA. We find that non-truthful preference submissions can be strictly optimal if and only if they are top-choice monotone. In equilibrium, DSPDA may implement allocations with justified envy. Specifically, it discriminates against students who are more loss averse or less confident than their peers, and amplifies already existing discrimination. To level the playing field, we propose sequential mechanisms as an alternative that is robust to these biases.

JEL-Classification: C78, D47, D78, D81, D82, D91.

Keywords: Market design, Matching, School choice, Reference-dependent preferences, Loss aversion, Deferred acceptance.

*We thank Georgy Artemov, Inácio Bó, Rustamdjan Hakimov, Fabian Herweg, Peter Katuščák, Dorothea Kübler, Takeshi Murooka, Antonio Rosato, Maybritt Schillinger, Roland Strausz, Georg Weizsäcker, and seminar participants in Aachen, Berlin, and Munich, as well as at CED'19, ESEM'19, MIP'19, VfS'19, AMETS, and CMID'20 for useful comments and suggestions. Funding by the German Research Foundation (DFG) through CRC TR 190 and 224 (Project B03), and UniCredit Foundation (Modigliani Grant) is gratefully acknowledged.

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

[‡]University of Bonn, Institute for Microeconomics, Adenauerallee 24-42, 53113 Bonn, Germany, jwangenheim@uni-bonn.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. Consequently, this mechanism is implemented in many existing school choice programs.[1] By submitting their true preferences in DSPDA, students can maximize the probability of getting 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 (especially, but not only) students with low priority tend to conceal their preferences for popular schools and mimic preferences for district schools despite the dominance of the truthful strategy. Hence, potentially, none of the desired properties may be 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 the expectations to which final match outcomes are compared. Hiding a preference for a popular school by ranking it behind a less preferred school is always costly in terms of the expected match utility, as such a rank-ordered list (ROL) shifts a part of the match probability to an inferior school. However, it also mitigates disappointment, and not even trying to get into the popular school by dropping it shields off disappointment completely. We characterize that ROLs are strictly rationalizable as a choice-acclimating equilibrium (CPE) in DSPDA if and only if they satisfy a property we call top-choice monotonicity, which is a testable prediction.[2] This theoretical foundation of commonly observed deviations is the first contribution of this paper.

As a second contribution, we show that these misrepresentations may give rise to justified envy and inefficiency in equilibrium. We analyze choice-acclimating Bayesian Nash equilibria (CBNE) when heterogeneously loss-averse students compete for scarce seats at elite schools. More specifically, loss-averse students decide to apply to their district schools over the elite schools if they are pessimistic about their admission chances. Consequently, weaker students with a lower degree of loss aversion (or higher degree of confidence), who submit true preferences, are accepted instead.

Third, we delineate how social segregation can arise if certain characteristics are correlated with demographics. 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. For example, evidence suggests that overconfidence

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

[2] An ROL is top-choice monotone if it ranks all schools preferred to the one ranked first in decreasing order of their preference and all other schools in increasing order. Such ROLs are indeed common in the data by Li (2017).

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 are perceive to be) marginalized when discrimination distorts priority scores, because EBLA incentivizes such students to shy away from ranking better schools in first place.[3] In that sense, DSPDA does not "level the playing field," voiding one of the crucial advantages prominently named by Pathak and Sönmez (2008). Our model also 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 rank certain schools does not necessarily mean that they prefer other schools.

Finally, we investigate how alternative mechanisms remedy these weaknesses. If all possible preference realizations have a unique stable matching, we suggest sequential school-proposing deferred-acceptance as an alternative to foster truthful reporting on the student side. If school preferences are homogeneous, this mechanism collapses to a simple serial dictatorship mechanism. Crucially, a remedy mechanism necessarily has to be sequential as a static mechanism can implement the student-optimal stable allocation if and only if DSPDA can do so. Letting students choose sequentially allows the mechanism (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 learn their match values for each school and their individual degree of loss aversion. Moreover, they receive a signal about their relative priorities compared to the other students at each school. Generally, given beliefs about the other students' priorities and strategies, a student's preference report corresponds to a lottery over 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 match utility alone, truthful reporting is a dominant strategy and, thus, induces a lottery that first-order stochastically dominates any lottery induced by any other ROL. Following the CPE framework by Kőszegi and Rabin (2007), the chosen outcome lottery constitutes the reference point. In addition to match utility, students receive psychological utility from comparing an outcome to the reference point. Since losses with respect to the reference point are weighted stronger than gains, any uncertainty in the match utility distribution generates a cost in expected utility.

As Kőszegi and Rabin (2007) have already proved, CPE allows for a preference for stochas-

[3]Through differences in perceived discrimination, our model can resolve apparently contradictory findings in the data. While Shorrer and Sóvágó (2017) document that students with a 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 may cause a more optimistic belief about getting into an elite school.

tically dominated lotteries, if an agent's loss aversion is sufficiently strong. Indeed, a 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. Such motifs can explain the evidence, suggesting that low- and mid-priority students are prone to misrepresentations but high-priority and optimistic students are not.

We draw on the extensive literature on matching mechanisms, but depart from the standard framework where preferences are only ordinal. In their seminal paper, Gale and Shapley (1962) introduce the deferred-acceptance mechanism as a solution to find the optimal stable matchings for the proposing side in the one-to-one matching problem. Dubins and Freedman (1981) and Roth (1982) show that it is strategyproof for the proposing side. Roth (1982) proves that it is not strategyproof for the receiving side. Balinski and Sönmez (1999) show that DSPDA is constrained efficient in the sense that no other fair mechanism Pareto-dominates it. Our model introduces a fundamentally different structure of incentives and questions all of these classical insights. Roth (1989) and Ehlers and Massó (2007) are the first to study matching with incomplete information.

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, they find that "misrepresentation rates are higher in weaker segments of markets" and increase "when applicants expect to face stronger competition," in line with the predictions of our model. In field data, misrepresentations are hard to identify since the true preferences are not observable. However, Hassidim et al. (2017b), Shorrer and Sóvágó (2017) and Artemov et al. (2020) exploit objective rankings in their data to expose "obvious misrepresentations" and find the same pattern.[4] Artemov et al. (2020) and Hassidim et al. (2017b) find that 1 - 20% and 2–8% of obvious misrepresentations are ex-post costly, respectively. Shorrer and Sóvágó (2017) estimate that the 12–19% costly obvious misrepresentations amount to \$3,000–\$3,500 on average (unconditionally \$347–\$738 per misrepresentation). That is, even when restricting attention to obvious misrepresentations, consequential deviations can be observed.

Truthfulness is easier to detect in the lab where preferences are induced by the experimental design. While the pioneers Chen and Sönmez (2006) focused on a comparison

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

of different mechanisms, more recently researchers have investigated patterns in preference manipulations. Hakimov and Kübler (2020) provide a well-structured overview of 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 which vary widely between studies. Rather than rooted in behavioral theory, most experimental studies are descriptive. 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.

In contrast to our paper, where students deliberately feed DSPDA incorrect preferences, misrepresentations have most commonly been interpreted as cognitive failures to identify the dominant strategy.[5] Li (2017) points out that DSPDA is not "obviously strategyproof" (OSP), and shows that most "mistakes" vanish when replacing DSPDA with sequential serial dictatorship. Notably, students reveal their true preferences in this OSP mechanism under EBLA as well, and our predictions are able to explain the most common deviations documented by Li (2017).[6] While both explanations may be relevant in practice, our paper raises the question of the extent to which the non-truthful play in DSPDA and the more truthful play in sequential serial dictatorship are due to non-standard preferences or cognitive mistakes.

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. Dreyfuss et al. (2019) recently and independently raised the point that EBLA can help explain misrepresentations in DSPDA. Alongside 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.I.

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 the choice-acclimating personal equilibrium

[5] See, e.g., Basteck and Mantovani (2018). However, when priorities and preferences are induced by the experimenter, we see the same individuals play a dominant or dominated strategy depending on their assigned score. Hassidim et al. (2017b) observe the same pattern in a high-ability population (compared to the general population). Controlling for cognitive ability, Shorrer and Sóvágó (2017) and Artemov et al. (2020) find a causal relationship between admission selectivity and dominated choices.

[6] We discuss the differences between the two concepts in Section A.II.

(CPE), the equilibrium concept we adopt. It essentially captures disappointment aversion similar to Bell (1985), Loomes and Sugden (1986) or Gul (1991), who model the reference point as the lottery's certainty equivalent. We choose CPE where outcomes are compared to the lottery's full distribution because it allows for "mixed feelings" and because it is unclear what the certainty equivalent of lottery over real school placements is supposed to be.

EBLA is supported by evidence from the field, such as Crawford and Meng (2011) or Pope and Schweitzer (2011). Evidence from the lab is mixed. While the conflicting evidence of Ericson and Fuster (2011) and Heffetz and List (2014) is affirmatively mended by Heffetz (2021), who introduces an extra treatment causing expectations to "sink in," the evidence on real-effort experiments with EBLA (Abeler et al., 2011; Gneezy et al., 2017) does not allow a clear verdict, yet. EBLA has been applied to a variety of economic models, 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), consumer search (Karle and Schumacher, 2020), and auctions (Lange and Ratan, 2010; **?**; von Wangenheim, 2020).

2 The model

Players: We consider finite sets of students, $\mathcal{I} := \{i_1, \ldots, i_n\}$, and schools, $\mathcal{S} := \{1, \ldots, m\}$. 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.[7] Similarly, each element of vector $\mathbf{w}_i = (w_{i,s})_{s\in\mathcal{S}}$ represents the payoff school s receives from being matched with student i. 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. For some results, we consider the special case that schools have the same preferences over students:[8]

Assumption 1 (Homogeneous school preferences)**.** $w_{i,s} = \omega_i \quad \forall s \in \mathcal{S}$ *and* ω_i *is uniformly distributed on* $[0, 1]$.[9]

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

[8] 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)

[9] 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.

The ordinal preference over schools corresponding to type θ_i is captured by a rank-ordered list (ROL). Formally, an ROL is a permutation of set $\mathcal{S}$, where ROL $(s_1, s_2, \ldots, s_m)$ is interpreted as school s_1 being most preferred, s_m least preferred, and s_k having k-th highest preference.[10] Let $\mathfrak{S}(\mathcal{S})$ be the set of all such permutations.

Mechanism: Our results refer to the direct student-proposing deferred-acceptance algorithm (DSPDA) defined below. We assume that schools always report their true preferences over students.[11] 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, which fully reveals the true ROL for all types, is optimal.

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 temporarily 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 all steps are executed mechanically based on the reported ROL. 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: Each student reports the preferences maximizing her expected utility. Students are expectation-based loss averse in the sense of Kőszegi and Rabin (2006, 2007). 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

$$u(\theta_i, s|r) = v_{i,s} + \begin{cases} \eta_i(v_{i,s} - v_{i,r}) & \text{if } v_{i,s} \geq v_{i,r}, \\ \eta_i\lambda_i(v_{i,s} - v_{i,r}) & \text{if } v_{i,s} < v_{i,r}, \end{cases} \tag{1}$$

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 degree of loss aversion, whereas

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

[11] 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).

$\eta_i \geq 0$ is the weight assigned to the gain-loss utility. As we will show in Equation (7), behavior is driven by a summarizing parameter $\Lambda_i = \lambda_i \eta_i - \eta_i$ called the loss dominance. 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 $\sigma_i(\theta_i)$ 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 $\mathcal{F}_i(\theta_i, \sigma_{-i})$ be the set of feasible lotteries. The expected utility from a lottery F_i evaluated with respect to some reference lottery $G = (g_s)_{s\in\mathcal{S}}$ is then

$$\mathcal{U}_i(\theta_i, F_i|G) = \sum_{s\in\mathcal{S}} f_{i,s} \left(\sum_{r\in\mathcal{S}} u(\theta_i, s|r) g_r \right). \tag{2}$$

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

$$U_i(\theta_i, F_i) := \mathcal{U}_i(\theta_i, F_i|F_i) \geq \mathcal{U}_i(\theta_i, F_i'|F_i') \quad \forall F_i' \in \mathcal{F}_i(\theta_i, \sigma_{-i}). \tag{3}$$

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

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. Let $\nu = (\nu_i)_{i\in\mathcal{I}}$ be the profile of true ROLs. An allocation rule is strategyproof if

$$v_{i,\alpha(\nu)[i]} \geq v_{i,\alpha(\nu_i',\nu_{-i})[i]} \quad \forall\, i \in \mathcal{I}, \forall\theta. \tag{4}$$

An allocation M is stable[12] if there is no pair i, s such that student i has justified envy,

$$v_{i,s} > v_{i,M(i)} \quad \text{and} \quad w_{s,i} > w_{s,i'} \quad \text{for some } i' \in M^{-1}(s), \tag{5}$$

[12]We use the classic definition of pairwise stability and use this word synonymously with "no justified envy," but we focus on the latter meaning, i.e., interpreting it as a fairness notion, following Abdulkadiroğlu and Sönmez (2003) in the spirit of Balinski and Sönmez (1999). To interpret stability as "no coalition can profitably deviate" in our setting, we would have to take into account that student i's approach to school s would create expectations and therefore scope for disappointment.

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

$$v_{i,M(i)} \geq v_{i,M'(i)} \quad \text{for any stable matching } M'. \tag{6}$$

3 Analysis

In Section 3.1, we consider the individual decision problem of a single student, 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 sequential mechanisms to remedy the harmful strategic behavior that arises due to loss aversion. In general, proofs are relegated to the appendix, Section A.III.

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 $v_1 > v_2 > \cdots > v_m$.

Match probabilities and attainability

By construction of DSPDA, student i is rejected by school s if at some step of the algorithm more than q_s students with a 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 before are filled by students that these schools individually prefer over student i. Given the other students' ROLs, we define school s as attainable for student i if and only if i obtains a seat at school s when ranking it first.

Lemma 1. *DSPDA assigns a student to her highest-ranked attainable school.*

Let $A_s \in \{1, 0\}$ be a binary variable determining whether school s is attainable for student i and we drop index i following the convention of this section. Whether a school is attainable for student i depends on the strategies of other students and on the schools' preferences over students, but not on the ROL submitted by the student herself. The submitted ROL does, however, determine which of the attainable schools is ranked first, and hence constitutes the student's match. Therefore, the submitted ROL determines the match outcome distribution F, and selecting an ROL effectively corresponds to choosing a lottery over match outcomes.

More formally, a student's beliefs about other students' types and strategies and school preferences lead to a joint probability distribution P on $\Pi_{s=1}^{n} A_s$ of being attainability at each school. We denote for each school s with $p_s = P(A_s = 1)$ the unconditional probability of being attainable at that school.[13] Since the student is matched with her highest-ranked attainable school, a reported ROL $(s_1, s_2, ..., s_m)$ leads to a lottery with match probabilities $F = (f_1, ..., f_m)$, where f_{s_k} is the joint probability that school s_k is attainable, while schools $s_1, ..., s_{k-1}$ are not. Importantly, the attainability probabilities are usually not independent, even when types are independent draws. Given a set of schools $\mathcal{S}$ and the individual loss parameter Λ, everything a student needs to know to find an optimal ROL is captured by the individual decision environment $\xi = (P, (v_s)_{s\in\mathcal{S}})$.

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. We can include the possibility to drop a school, i.e., not listing it in the ROL, by defining the outside option as a fictional school m with unlimited capacity and normalized $v_m = 0$. Depeding on the context the outside option may refer to remaining unmatched or being matched to some "default" option. Different ROLs which rank the same schools in arbitrary order behind the outside option are equivalent in the sense that they induce the same match probabilities. Because the outside option is always attainable, ranking a school after it corresponds to dropping this school with which a match is excluded. While it is never optimal to list schools with $v_s < 0$, we will show that it can be optimal to drop schools with $v_s > 0$.

Payoffs

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

$$
\begin{aligned}
U(\theta, F) &= \sum_{s\in\mathcal{S}} f_s \left(\sum_{r\in\mathcal{S}} u(\theta, s|r) f_r \right) \\
&= \sum_{s=1}^{m} f_s \left[\left(\sum_{r=1}^{m} f_r \right) v_s + \sum_{r=1}^{s-1} f_r \lambda \eta (v_s - v_r) + \sum_{r=s+1}^{m} f_r \eta (v_s - v_r) \right] \\
&= \underbrace{\sum_{s=1}^{m} f_s v_s}_{\text{classical utility}} \underbrace{- \Lambda \sum_{s=1}^{m} \sum_{r=s+1}^{m} f_s f_r (v_s - v_r)}_{\text{gain-loss utility}} .
\end{aligned} \tag{7}
$$

Each pairwise comparison is weighted by $f_s f_r$ and shows up twice: once as a gain and once as a loss, its total factor is $\Lambda > 0$. Since losses are weighted stronger than gains, expected gain-loss utility always enters negatively. Under our notational convention, the difference $(v_s - v_r)$ is 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, $\lambda = 1$, would

[13] 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 students' preferences or draw wrong inference on other students' used strategies.

result in $\Lambda = 0$ such that students only maximize classical utility. If $\Lambda > 1$, gain-loss utility dominates match utility, which will become central.

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 can be rationalized under EBLA 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 each, 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,*

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

Schools' preferences are determined by scores ω_i 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. Table 1 provides the distribution of attainability probabilities for $\omega = 1/4$ and $\varepsilon = 1/20$.

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

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

Evidently, both schools are attainable for A only if she has the highest score, and neither school is if she has the lowest score. Only one of the schools is attainable if she has the second highest score and the student the with highest score prefers the other school. Note that the attainability probabilities are interdependent, even though preferences and scores are drawn independently.

From the attainability 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 it is attainable, with school 2 if and only if is attainable but school 1 is not, and to no match if and only if both schools are unattainable. 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 both schools being attainable) from school 1 to 2, which decreases not only 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. ROLs listing the outside option first induce identical degenerate lotteries.*

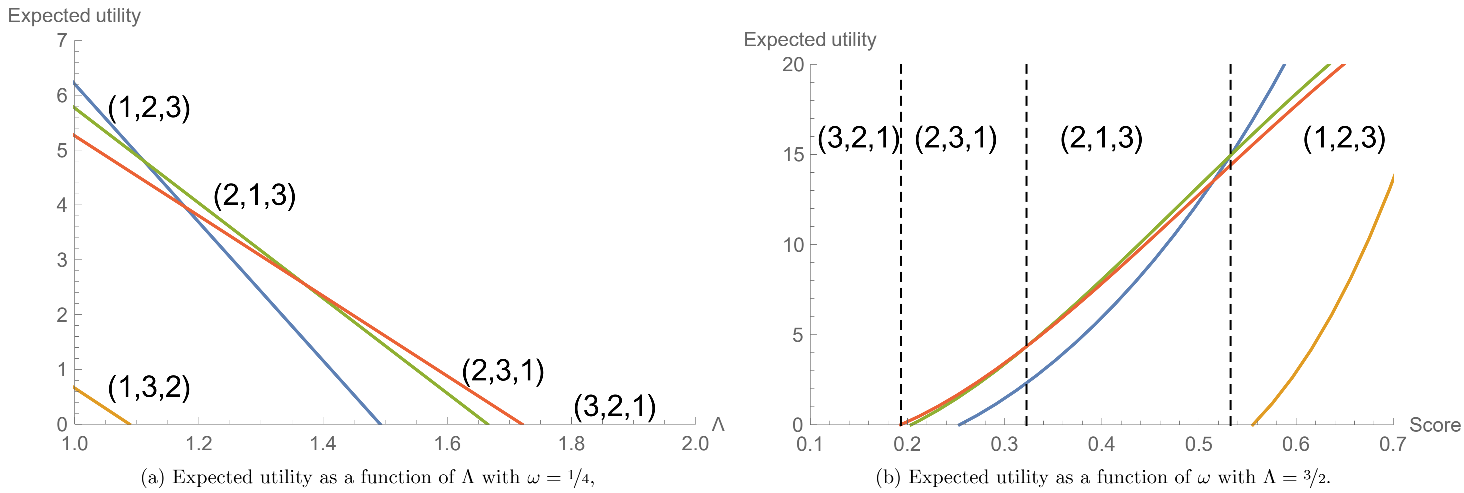

Figure 1: The expected utilities induced by every ROL as a function of (a) Λ and (b) ω, setting $v_1 = 100, v_2 = 30, v_3 = 0$ and $\varepsilon = {}^{1}/_{20}$.

Given the lotteries, we can calculate the 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 a 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 as worse, ultimately culminating in submitting an empty ROL when the perceived cost of uncertainty is sufficiently high.*[14] *Notably, any optimal manipulation involves a flipping (or dropping) of the most preferred*

[14]Abstaining from the mechanism by choosing a dominated outside option is reminiscent of the "uncertainty effect" documented by Gneezy et al. (2006).

option – ROL $(1,3,2)$ is never optimal. Figure 1b illustrates that students tend to become more truthful as their scores increase and they become more optimistic.

A large Λ by itself does not lead to profitable deviations from the true ROL, as they are inherently linked to incomplete information. If students had full information about other students' reports and schools' preferences, they could infer their match outcome from the mechanism, and would have no cost of being truthful, such that DSPDA would implement the student-optimal stable matching. Moreover, the optimal ROLs of this example cannot be explained by simple risk aversion, because the truthful lottery first-order stochastically dominates every other feasible lottery.

Characterization of optimal ROLs

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. As we will show in Proposition 2, 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 $\Lambda \leq 1$, and, by the dominance of the truthful strategy with standard preferences, the truthful lottery first-order stochastically dominates any other feasible lottery. Hence, the truthful strategy is a CPE in DSPDA for any profile and all possible beliefs if and only if $\Lambda \leq 1$.

While many applied papers restrict attention to $\Lambda \leq 1$ ("no dominance of gain-loss utility"), we explicitly allow (all or only some) students to be dominantly loss averse, in order to explain deviations from truth-telling.[15] There is substantial evidence that a large fraction of the population is indeed dominantly loss averse.[16] While the possible preference for first-order stochastically dominated lotteries that comes with this assumption may appear counterintuitive, it is not only observable in the matching context.[17]

When searching for a best-responding ROL given a type, beliefs, and others' reporting strategies, the following property turns out to be both necessary and sufficient condition for optimality for some decision environment $\xi = (P, (v_s)_{s\in\mathcal{S}})$.

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

[16]$\Lambda > 1$ matches the conventional wisdom that "losses loom about twice as large as gains." 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 a meta-analysis of 150 articles, Brown et al. (2021) find that the mean loss aversion coefficient λ with $\eta = 1$ is between 1.8 and 2.1 and about 38% out of 586 estimates find $\lambda > 2$, corresponding to $\Lambda > 1$.

[17]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.

Definition 1. *An ROL is top-choice monotone if it*

- *reverses the order of schools preferred over the top choice, i.e., it ranks school k first and all schools $1, \ldots, k-1$ in decreasing order, and*
- *preserves the order of schools not preferred over the top choice, i.e., it ranks school k first and all schools $k+1, \ldots, m$ in increasing order.*

For example, the ROL $(1, 3, 2, 4)$ is not top-choice monotone, because 2 is ranked behind 3 although $v_2 > v_3$ such that the preference order of schools considered worse than top-choice 1 is not reflected in the ranking. Similarly, ROL $(3, 1, 2, 4)$ violates the property, while ROL $(3, 2, 1, 4)$ satisfies it as the preference ranking of schools preferred over top-choice 3 is reversed. Table 3 exhibits further examples.

Proposition 1. *Take any $\mathcal{S} = \{1, \ldots, m\}$ with the implied ordinal ranking and any $\Lambda > 1$.*

a) For any decision environment ξ, a strictly optimal ROL must be top-choice monotone.

b) For any ROL L which is top-choice monotone with respect to the ordinal ranking there is a decision environment ξ such that L is strictly optimal.

If all ROLs correspond to different lotteries over match outcomes, Proposition 1 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 is top-choice monotone.[18] Only a comparably small set of ROLs is top-choice monotone.[19]

The formal proof of the proposition is the appendix, but its general idea is easily understood by example. Table 3 shows all possible ROLs for a setting with $m = 4$ as an outside option. The bold numbers are the listed schools, as schools ranked after 4 can be interpreted as "dropped from the ranking." The shaded ROLs are never strictly optimal as they violate top-choice monotonicity. For instance, $(1, 3, 2, 4)$ reverses 2 and 3 which are not preferred over top choice 1. Intuitively, if the student were willing to reduce

[18]For this reason, we render ROLs equivalent for which only the ranking after the outside option differs. In Table 3 the darkly shaded ROLs are in this sense redundant as they represent the same lottery as a unique top-choice monotone analog. 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 that are not attainable does not matter. There are no equivalent ROLs if for any subset of schools the probability of all of them being attainable is strictly between zero and one.

[19]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), just 2^{m-1} are top-choice monotone.

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 school 1 to 3, i.e., $(3,1,2,4) \succ_i (1,3,2,4)$. Hence, $(1,3,2,4)$ can never be strictly optimal. Similarly, $(3,1,2,4)$ cannot be strictly optimal as either $(1,3,2,4) \succ_i (3,1,2,4)$ or $(3,2,1,4) \succ_i (3,1,2,4)$.

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. The lightly (and darkly) shaded ROLs are not top-choice monotone and thus never strictly optimal.

As a first impression of our theory's predictive power, we briefly consider an 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 between \$0 and \$1.25. Then, participants simultaneously submit an ROL about the prizes to a mechanism which calculates the DSPDA allocation.

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%
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%
TCM	91.1%	77.4%	77.5%	88.7%	75.9%	91.9%	87.2%	98.2%	95.7%	93.8%	87.5%

Table 4: Share of most commonly submitted ROLs and total share of top-choice monotone ROLs for each priority score. Most common deviation from truth-telling in bold.

Table 4 summarizes 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 top-choice monotone. 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, higher scores are more likely to submit truthful ROLs.

Proposition 1 implies that any manipulation of the ROL will concern the most preferred schools: the true ROL is strictly optimal if and only if it is strictly 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. The attainability probability p_1 only depends on beliefs about what other students do and their priority relative to our representative student. Hence, Proposition 2 gives precise bounds on when DSPDA is incentive-compatible for loss-averse students based only on fundamentals that are exogenous in this section. 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 attainability probabilities and also the cardinal utilities.

We say a school is exclusive if attainability at that school implies non-attainability of some other school.[20]

Proposition 2. *Suppose the most preferred school is not exclusive. Let p_1 be the probability that a student's most preferred school is attainable.*

1. *If $p_1 > 1 - 1/\Lambda$, then the true ROL is strictly optimal.*
2. *If $p_1 < \frac{1-1/\Lambda}{2}$, then the true ROL is strictly suboptimal.*

This proposition implies that under Assumption 1 sufficiently high scores report truthfully whereas sufficiently low scores misrepresent whenever seats at their preferred school are scarce. This result is in line with the evidence that suggests 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 in our setting. Importantly, we have made no assumptions on whether the beliefs determining the attainability probabilities are rational. Consequently, EBLA is a channel which renders other well-documented biases distorting the beliefs as decisive. Here, 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 participants are more likely to be truthful.[21] Without our theory, this observation may appear counterintuitive as this bias usually steers behavior away from a rational unbiased benchmark.

Truncated lists are prevalent in the data. Since constraining the ROLs to a fixed number

[20] For instance, a boy school would be exclusive, if there were a girl school in the set of schools. Evidently, if a school is exclusive with other schools, then the rank of the exclusive school among these schools in an ROL is inconsequential for attainability, and hence multiple ROLs induce the same outcome lottery.

[21] 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.

of schools can destroy both the strategyproofness and the stability of DSPDA, economists often advocate against such restrictions. While prohibiting complete ROLs introduces strategic motifs into DSPDA with standard preferences, such motifs are already present in our setting. As in our models truncations formally correspond to ranking a school behind the outside option, Proposition 1 implies the following statement on truncations:

Corollary 1. *It is never strictly optimal to drop some desirable school k from the ROL, but list some preferred school $\ell < k$.*

3.2 Strategic interaction

In this section, we investigate the structure of the (essentially) unique choice-acclimating Bayesian Nash equilibrium (CBNE) with homogeneous school preferences. We rationalize the prevalent district-school bias as an equilibrium phenomenon in a setting with district and elite schools.

To establish the general existence of a CBNE for arbitrary school preferences, note that a CBNE is just a standard Bayesian Nash Equilibrium, where individual utilities over actions are given by utility function (7).[22] Equilibrium existence is then implied by Theorem 1 in Milgrom and Weber (1985).[23]

In general, strategic interaction between loss-averse agents is difficult to analyze and has to date only been sparsely studied. A key observation in the analysis of strategic interaction with homogeneous preferences of schools is that a student's match outcome is only affected by the behavior of other students with a higher score. By construction of DSPDA, a student will only be rejected by a school she proposes to if this school also has an application 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 submitted ROLs by students of higher scores. For ease of exposition, we now assume that students' type space is finite and all schools break ties between students in the same publicly known deterministic way. An equilibrium is essentially unique if it is unique after imposing a rule for how students break ties when they are indifferent between multiple ROLs.

Proposition 3. *With homogeneous school preferences and discrete types, there exists*

[22] This interpretation subtly involves a 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 is realized only *after* the player has chosen it and has formed her reference point.

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

an essentially unique CBNE in pure strategies.

Misallocations with elite schools and the district-school bias

We now employ a simplified setting to derive the district-school bias first explored by Chen and Sönmez (2006).[24] For complex preference structures, there are numerous interdependencies, each giving rise to potential risks of instability. For now, we focus on the district-school bias and neglect other sources of misrepresentations, such as differences in preferences for schools within the set of desirable schools.

Suppose that there is a set $\mathcal{E} \subset \mathcal{S}$ of elite schools. Each school from this set is unambiguously preferred by each student over some safe outside option, the district school. To simplify, we assume that all elite schools induce the same match utility $v > 0$, whereas the safe outside option induces a normalized utility of zero.

Suppose further that Assumption 1 holds: each student i independently draws a score ω_i from the uniform distribution. Let a student's loss dominance Λ_i be independently drawn from a common distribution with discrete support $\{\Lambda^0, \Lambda^1, \ldots, \Lambda^l\}$. 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 $\Lambda^1 > 1$. By the following lemma, we can, without loss of generality, focus on just one elite school with capacity $q = \sum_{s \in \mathcal{E}} q_s < n$ instead of a set $\mathcal{E}$ of elite schools.

Lemma 2. *For any belief on the attainability probabilities of elite schools, the best response is to either rank all elite schools adjacently (in any order) or no elite school before the district school.*

In CBNE, a student's decision as to whether to apply to the elite school depends on her probability of attaining it, i.e., the probability that fewer than q students of higher score apply there. Hence, the attainability probability is a function which is weakly increasing in her score ω and depends on the other students' reporting strategy σ_{-i}. Fixing σ_{-i}, payoff function (7) implies that listing the elite school before the outside option is optimal for any $\omega > 0$ if and only if

$$f(\omega)v - \Lambda f(\omega)(1 - f(\omega))v \geq 0 \iff \Lambda \leq \frac{1}{1 - f(\omega)} \iff f(\omega) \geq 1 - \frac{1}{\Lambda}. \tag{8}$$

Consequently, for any score $\omega \in (0, 1)$, there is a cutoff $\overline{\Lambda}(\omega) = {}^{1}/_{(1-f(\omega))}$ such that applying to the elite school is a best response if and only if $\Lambda \leq \overline{\Lambda}(\omega)$. 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,

[24]Hakimov and Kübler (2020) state the phenomenon that "the district school (or safe school) is ranked higher in the reported list than in the true preferences," and document its prevalence in a wide range of experiments. In our two-school setting, it is equivalent to the small-school bias.

any student with loss dominance Λ^{l-1} can infer her score cutoff $\overline{\omega}(\Lambda^{l-1})$ below which she drops the elite school, and so on.

Lemma 3. *In the elite school problem, there is an essentially unique CBNE. In this equilibrium, 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 insight that, in CBNE, pessimistic loss-averse students shy away from applying has relevant ramifications for affirmative action policy beyond the scope of this simplified model. While the cutoff attainability probability in (8) only depends on the score and the loss dominance, beliefs can be additionally skewed which leads directly to unfavorable allocations. For instance, if some students expect their scores to be lower than they are, because of (perceived) discrimination, these students may not apply to the elite school although they would be assigned a seat in the stable allocation. Consequently, DSPDA aggravates the discrimination by discouraging truthful revelation, and whether this discrimination is real or caused by underconfidence or doubts about how schools assess abilities, is irrelevant. Thus, downplaying the cost of discrimination when marginalized students do not rank discriminating schools in DSPDA is inherently flawed in models incorporating EBLA, because the submitted ROLs may not reflect true preferences in equilibrium.

3.3 Possible remedies

In keeping with Proposition 2, misreporting in our model is inherently linked to beliefs on attainability and thereby on how a student expects to compare to others in terms of priority. While we modelled the source of uncertainty about this relative priority as uncertainty about other students' priorities $\mathbf{w}_{-i}$ at schools, the same holds true when it stems from uncertainty about how the schools assess own abilities $\mathbf{w}_i$. Notably, this uncertainty is unlikely to depend on market size. Moreover, our model abstracts away from other inherent sources of uncertainty in the market than the relative priority, and other students' preferences that may aggravate the problem, such as institutional uncertainties. In general, any regulation that may help a student to better assess his own standing and capacities of schools is likely to reduce students' uncertainty and hence the misrepresentation of preferences. In this section, we study how the choice of the mechanism itself may mitigate the problem of uncertainty and help to prevent justified envy in the equilibrium allocation.

Static mechanisms

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

Lemma 4. *For any distribution of preferences, there exists a static mechanism that implements the student-optimal stable allocation as CBNE for all realizations if and only if DSPDA is truthful.*

Hence, if DSPDA fails, there is no hope for remedies in the class of static mechanisms. Our argument is akin to 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 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 eliciting preferences. 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, see, e.g., Myerson (1979). In settings with dynamic information and EBLA, however, this revelation principle does not apply. As students evaluate outcomes with respect to their beliefs, information endogenously affects their preferences over alternatives.

For a dynamic equilibrium concept in the context of EBLA, we follow **?** in his straightforward extension of CPE to dynamic settings. At each decision node of an extensive form game, a student correctly anticipates her choices at any future node. Based on the induced beliefs and using backward induction, she selects the lottery most-preferred under the static CPE at every decision node, with the reference point at each choice being her beliefs about final match outcomes conditional on the information available at that stage.

Formally, any stage of the mechanism may reveal new information that alters the beliefs about the final match outcome, which depends on the student's behavior in future stages. Let $\mathcal{F}_{i,k}(\theta_i, \sigma_{-i})$ 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 a lottery corresponding to some σ_i. Given some σ_{-i}, strategy σ_i is a sequential CPE (SCPE) if, at every decision node k, it selects a lottery $F_{i,k}$ such that

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

Accordingly, we call a strategy profile where each player's strategy is an SCPE, given other players' strategies, a sequential CBNE (SCBNE). Equipped with the tools to analyze incentives in sequential mechanisms, we now evaluate the sequential student-receiving (school-proposing) deferred-acceptance mechanism (SSRDA).

Definition 2 (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. 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 even with standard preferences DA is not truthful for the receiving side, as it implements the optimal stable allocation for the proposing side. Under full information, the receiving side could coordinate on their preferred stable match by strategically rejecting all alternatives. In general, a strategic rejection of a school s can trigger a more preferred offer s' because another student may free capacity at s' by accepting s. Such considerations play no role when the stable matching is unique. For standard ordinal preferences, Ehlers and Massó (2007) show that under incomplete information truthfulness is an ordinal Bayesian Nash equilibrium if and only if for all realizations of preferences the stable match is unique.[25]

The literature has identified numerous sufficient conditions to ensure the uniqueness of stable matches for given preferences (e.g., Eeckhout (2000), Niederle and Yariv (2009) and Karpov (2019)).

Proposition 4. *If there is a unique stable allocation for all realizations of preferences, the truthful strategy profile is an SCBNE in SSRDA, and this stable allocation is implemented for all preference realizations.*

In contrast to the student-proposing DA, EBLA tends to mitigate the incentives for misreporting in the student-receiving DA. This finding is in line with evidence that, in contrast to the proposing side, strategizing on the receiving side seems to play no major role.[26] Intuitively, a student with an offer in SSRDA can obtain a seat with certainty if she accepts. A rejection increases the uncertainty over the match outcome. Since the loss parameter Λ can be interpreted as a cost parameter for uncertainty, strategic rejections become even less appealing for a larger Λ.[27]

[25] A strategy profile is an ordinal Bayesian Nash equilibrium if it is a Bayesian Nash equilibrium for any cardinal representation of the ordinal preferences.

[26] For instance, Klijn et al. (2019) find that SSRDA outperforms the (static and sequential) student-proposing DA mechanism in terms of stability and average payoffs.

[27] This effect is related to the logic in Fernandez (2018), who identifies anticipated regret for the case where manipulations do not pay off as a possible explanation for the observed truthful behavior in

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 with two students and $q = 1$. In the unique stable matching, the student with the 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* (8) *holds. Consequently, if the score of both students is below* $1 - {}^{1}/_{\Lambda}$, *both students attend the district school, and the match outcome is neither stable nor student optimal.*

Sequential 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 sequential serial dictatorship in which homogeneous priority scores determine the order in which students pick their school. Because each student determines her match with certainty, the dominance of choosing the most preferred among the available options is obvious regardless of EBLA. Since with homogeneous school preferences the stable allocation is unique, the following corollary is an immediate consequence of Proposition 4.

Corollary 2. *Under Assumption 1, the truthful strategy is an SCBNE in SSRDA, and the unique optimal stable allocation is obtained.*

Li (2017) compares the outcome of DSPDA with the outcome of a sequential serial dictatorship mechanism 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, where the only outcomes considered are those that follow from the information sets where both strategies first diverge. Hence, dominance in an OSP mechanism may be easier to detect by agents with cognitive limitations.

SSRDA.

Ashlagi and Gonczarowski (2018) show in their Example 1 that the sequential 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 always implements stable match outcomes. In Appendix A.II, we build on their Example 2 to demonstrate that a stable OSP mechanism may fail to induce stability with EBLA. This example sets the two concepts apart and suggests how to disentangle the explanations experimentally. Our model conveys EBLA as an alternative explanation for the observed misrepresentations. Rather than a mistake because of cognitive limitations, we see them as the deliberate optimal choice of students who suffer from a behavioral bias.

4 Conclusion

We have identified a possible reason why students play dominated strategies in the strategyproof direct student-proposing deferred-acceptance mechanism (DSPDA). The truthful equilibrium in dominant strategies may not be a choice-acclimating personal equilibrium (CPE) with dominant expectation-based loss aversion (EBLA). In other contexts, evidence consistent with dominant EBLA has been found 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 want to get into competitive programs. Our theoretical predictions fit this pattern in experimental and field data, and we also provide a formalized framework for the pervasive district-school and small-school biases. Our characterization of optimal play in Proposition 1 and 2 is testable.

The extensive evidence of 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 that 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. If (as the data suggests) these characteristics are correlated with demographics, DSPDA indirectly but inherently implements an imbalanced allocation and amplifies discrimination against already marginalized groups. The misallocation problem does not vanish as markets grow large.

We have discussed remedy mechanisms and our theory suggests that sequential mechanisms outperform static mechanisms in terms of truthfulness. Under homogeneous school preferences, a sequential serial dictatorship mechanism delivers the unique stable allocation in dominant strategies and in a truthful CPE, i.e., it succeeds where the celebrated DSPDA fails. Indeed, Li (2017) documented that this mechanism outperforms its static version in his experiment. While he attributes this to obvious strategyproofness, we suggest that also reference-dependent preferences may drive this difference. That is, it is not clear whether misrepresentation originates from a behavioral bias (loss aversion) or a cognitive impairment (difficulties understanding the dominance of a strategy). We see the presence of at least some students with non-standard preferences as hard to deny. Our paper provides first steps into understanding this dimension, but more experimental work is needed to disentangle loss aversion and cognitive limitations.

Appendix

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

Similar to our paper, Dreyfuss et al. (2019) find that EBLA can explain non-truthful ROLs. 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 is resolved, which gives rise to additional gain-loss utilities. The essential intuition for how students use manipulations to shield off potential disappointment is, however, similar in both models.

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 choose our modeling approach to draw a clear comparison to the existing experimental literature, where stakes are simply money, and values are hence fully comparable between schools.

The uncertainty in Dreyfuss et al. (2019) stems from iid shocks on how individual schools assess 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 attainability probabilities are independent between schools, which is 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 light of EBLA, we delve deeper into its theoretical implications, and analyze the set of rationalizable strategies, strategic interaction, and evaluate remedy mechanisms.

A.II 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, all students will report truthfully in DSPDA if their probability p_1 that their favorite school is attainable is sufficiently large. This condition certainly does not imply that DSPDA is OSP. For example, ranking the favorite school 1 second can yield a match with 1 as a best case, whereas the true ROL can lead to a worse match.

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 with standard preferences, 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 $\epsilon > 0$. Here, DSPDA is not OSP. For instance, if student A prefers school 2 truth-telling is not obviously dominant as the true ROL $(2, 1)$ may result in a match with school 1, whereas ROL $(1, 2)$ may result in a match with 2 with positive probability.

Ashlagi and Gonczarowski (2018, Figure 2) propose the following OSP sequential mechanism. 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 an SCPE at this final decision node). If A prefers school 1, deviating yields her either $v_{A,1}$ or $v_{A,2}$ 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 $v_{A,1} < v_{A,2}$ and being truthful leads to a lottery over $v_{A,2}$ and $v_{A,2}$. Because even the worst outcome from being truthful is not worse than the best (only) outcome from deviating, the truth is an obviously dominant strategy, making the mechanism OSP. However, for any $\Lambda > 1$, truth-telling is not an SCPE for student A if $\epsilon < 1 - {}^{1}/_{\Lambda}$. Hence, because of the uncertainty effect, even OSP mechanisms can fail to have a truthful SCPE.

A.III Proofs

Proof of Lemma 1. The claim follows from the fact that DSPDA is strategyproof for every student with standard preferences. Take an arbitrary ROL and let s be the highest-ranked attainable school in it.

Suppose that under this ROL the student is matched with s' ranked before s. But then, since s' is unattainable (i.e., would not get in if ranked first), she would prefer this ROL over her true ROL if s' was her most preferred school, a contradiction to strategyproofness.

Suppose that under this ROL she is matched with s'' ranked behind s. But then, if the ROL was true, she would prefer a match with s over s'', and ranking s first would achieve this match, again a contradiction to strategyproofness. $\square$

Proof of Proposition 1. a) We start with a practical lemma which identifies when flipping two adjacently ranked schools in an ROL is profitable. Consider two otherwise identical ROLs swapping two 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 x and y being attainable but no school which is ranked before.

Lemma 5. *$U(\cdot, F) \geq U(\cdot, \underline{F})$ if and only if*

$$\frac{\varepsilon}{\Lambda} \geq \varepsilon \left(-\sum_{s=1}^{x} f_s + \varepsilon + \sum_{s=x+1}^{y-1} f_s \, \frac{v_x + v_y - 2v_s}{v_x - v_y} + \sum_{s=y}^{m} f_s \right) \tag{9}$$

with equality in (9) *only in the case of indifference.*

Proof of Lemma 5. By (7), we have

$$U(\cdot, F) - U(\cdot, \underline{F}) = \sum_{s=1}^{m} (f_s - \underline{f}_s) v_s - \Lambda \sum_{s=1}^{m} \sum_{r=s+1}^{m} (f_s f_r - \underline{f}_s \underline{f}_r)(v_s - v_r). \tag{10}$$

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 $\underline{f}_x = f_x - \varepsilon$, $\underline{f}_y = f_y + \varepsilon$, with $\varepsilon \geq 0$. Hence, if we split the second sum over s into five summands, we obtain

$$\begin{aligned}
&\sum_{s=1}^{m} \sum_{r=s+1}^{m} (f_s f_r - \underline{f}_s \underline{f}_r)(v_s - v_r) \\
&= \sum_{s=1}^{x-1} \Big[(f_s f_x - f_s \underline{f}_x)(v_s - v_x) + (f_s f_y - f_s \underline{f}_y)(v_s - v_y) \Big] + \sum_{r=x+1}^{m} (f_x f_r - \underline{f}_x \underline{f}_r)(v_x - v_r) \\
&\quad + \sum_{s=x+1}^{y-1} (f_s f_y - f_s \underline{f}_y)(v_s - v_y) + \sum_{r=y+1}^{m} (f_y f_r - \underline{f}_y f_r)(v_y - v_r) + 0 \\
&= \sum_{s=1}^{x-1} [f_s \varepsilon (v_s - v_x) - f_s \varepsilon (v_s - v_y)] + \sum_{r=x+1}^{m} \varepsilon f_r (v_x - v_r) + (\underline{f}_x f_y - \underline{f}_x \underline{f}_y)(v_x - v_y) \\
&\quad + \sum_{s=x+1}^{y-1} f_s (-\varepsilon)(v_s - v_y) + \sum_{r=y+1}^{m} (-\varepsilon) f_r (v_y - v_r)
\end{aligned}$$

$$
\begin{aligned}
&= -\sum_{s=1}^{x-1} \varepsilon f_s(v_x - v_y) + \sum_{r=y}^{m} \varepsilon f_r(v_x - v_r) + (f_x - \varepsilon)(-\varepsilon)(v_x - v_y) \\
&\quad + \sum_{s=x+1}^{y-1} \varepsilon f_s(v_x - 2v_s + v_y) + \sum_{r=y+1}^{m} (-\varepsilon) f_r(v_y - v_r) \\
&= \varepsilon(v_x - v_y) \left(-\sum_{s=1}^{x} f_s + \varepsilon + \sum_{s=y}^{m} f_s + \sum_{s=x+1}^{y-1} \frac{v_x - 2v_s + v_y}{v_x - v_y} \right)
\end{aligned}
$$

Since the difference in classical utility satisfies $\sum_{s=1}^{m}(f_s - \underline{f}_s)v_s = \varepsilon(v_x - v_y)$, we have $U(\cdot, F) - U(\cdot, \underline{F}) \geq 0$ if and only if

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

Dividing by $\Lambda(v_x - v_y) > 0$ yields the result for the inequality. For the statement about indifference, replace all inequalities with equality. $\square$

We now prove the next auxiliary lemma by contradiction.

Lemma 6. *If a strictly optimal ROL ranks school b after school c for $b < c$, it ranks the schools $1, ..., b-1, b$ in decreasing order.*

Proof of Lemma 6. Suppose that for some $1 \leq a < b < c \leq m$, the strictly optimal ROL ranks b behind c but a before b. Let c be the least preferred school, i.e., the one with the highest index, for which such a triple exists in this ROL. Given this c, select b and a such that a is the lowest-index school, i.e., the most preferred one, satisfying the requirement.

Since a is ranked before b, the optimal ROL has one of the following forms:

i) $(..., a, ..., c, ..., b, ...)$

ii) $(..., c, ..., a, ..., b, ...)$

We make first considerations for both cases.

i) Since, by assumption, a is the lowest-index school ranked before b, the list must be increasing from a, and eventually decreasing (possibly at b) to a number above a. Call $\overline{x}$ the first school where the list starting from a has decreased. Now, by choosing $\underline{x}$ appropriately in the list between a 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, c is the highest-index school for which there exists b and a with b ranked behind c but a before b, the list must be decreasing from c, but eventually increasing (possibly immediately after a) to a number below c. Call $\underline{y}$ the first school after c where the list is increasing. Now, by choosing $\overline{y}$ appropriately in the list between

c 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 rest of the proof is identical for both cases.

Since the ROL is supposed to be strictly optimal, it must be strictly preferred to an otherwise equivalent ROL that swaps $\overline{x}$ and $\overline{y}$. 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

$$\overline{f}_{\overline{x}} = f_{\overline{x}} + \overline{\varepsilon} \quad \text{and} \qquad\qquad \overline{f}_{\overline{y}} = f_{\overline{y}} - \overline{\varepsilon}, \tag{11}$$

where $\overline{\varepsilon}$ is the probability that $\overline{x}$ and $\overline{y}$ are attainable, but any school ranked before $\overline{x}$ and $\overline{y}$ in the optimal ROL is not. By the strict optimality, $\overline{\varepsilon} > 0$.

Hence, by Lemma 5

$$\begin{aligned}\frac{1}{\Lambda} &< -\sum_{s=1}^{\overline{x}} \overline{f}_s + \overline{\varepsilon} + \sum_{s=\overline{x}+1}^{\overline{y}-1} \overline{f}_s \frac{v_{\overline{x}} + v_{\overline{y}} - 2v_s}{v_{\overline{x}} - v_{\overline{y}}} + \sum_{s=\overline{y}}^{m} \overline{f}_s \\ &= -\sum_{s=1}^{\overline{x}} f_s + \sum_{s=\overline{x}+1}^{\overline{y}-1} f_s \frac{v_{\overline{x}} + v_{\overline{y}} - 2v_s}{v_{\overline{x}} - v_{\overline{y}}} + \sum_{s=\overline{y}}^{m} f_s - \overline{\varepsilon}\end{aligned} \tag{12}$$

Similarly, the ROL must be strictly preferred to an otherwise equivalent ROL that swaps $\underline{x}$ and $\underline{y}$. Hence, by Lemma 5

$$\frac{1}{\Lambda} > -\sum_{s=1}^{\underline{x}} f_s + \underline{\varepsilon} + \sum_{s=\underline{x}+1}^{\underline{y}-1} f_s \frac{v_{\underline{x}} + v_{\underline{y}} - 2v_s}{v_{\underline{x}} - v_{\underline{y}}} + \sum_{s=\underline{y}}^{m} f_s \tag{13}$$

Both inequalities can only simultaneously hold if the right-hand side of (12) is strictly larger than the right-hand side of (13), which we bring to a contradiction. Since $-\overline{\varepsilon} < 0 < \underline{\varepsilon}$ it suffices to show that for each s the respective summand in (12) is (weakly) smaller than in (13).

For $s \in [\underline{x}+1, \overline{x}]$, we have $-1 = \frac{v_{\underline{x}}+v_{\underline{y}}-2v_s}{v_{\underline{x}}-v_{\underline{y}}} + 2\frac{v_s - v_{\underline{x}}}{v_{\underline{x}}-v_{\underline{y}}} < \frac{v_{\underline{x}}+v_{\underline{y}}-2v_s}{v_{\underline{x}}-v_{\underline{y}}}$, since $v_{\underline{x}} > v_s > v_{\underline{y}}$. For $s \in [\overline{x}+1, \underline{y}-1]$, we have

$$\frac{v_{\overline{x}} + v_{\overline{y}} - 2v_s}{v_{\overline{x}} - v_{\overline{y}}} = \frac{1}{1 + \frac{2v_s - 2v_{\overline{y}}}{v_{\overline{x}}+v_{\overline{y}}-2v_s}} < \frac{1}{1 + \frac{2v_s - 2v_{\underline{y}}}{v_{\underline{x}}+v_{\underline{y}}-2v_s}} = \frac{v_{\underline{x}} + v_{\underline{y}} - 2v_s}{v_{\underline{x}} - v_{\underline{y}}},$$

where the inequality follows since $v_{\underline{x}} > v_{\overline{x}}$, $v_{\underline{y}} > v_{\overline{y}}$, and the term is increasing in both, $v_{\overline{x}}$ and $v_{\overline{y}}$. For $s \in [\underline{y}, \overline{y}-1]$, we have $\frac{v_{\overline{x}}+v_{\overline{y}}-2v_s}{v_{\overline{x}}-v_{\overline{y}}} = 1 + 2\frac{v_{\overline{y}}-v_s}{v_{\overline{x}}-v_{\overline{y}}} < 1$ since $v_{\overline{x}} > v_s > v_{\overline{y}}$. $\square$

Now, a) of the proposition follows immediately. That schools preferred over top choice b are optimally ranked in decreasing order is just Lemma 6. An ROL not ranking schools worse than a top choice a in increasing order would rank some $a < b < c$ in the form of $(a, ..., c, ..., b, ...)$. As seen by the contradiction of case (i) in the proof of Lemma 6, this cannot be optimal.

We now establish assertion b). First, by Proposition 2 the truthful ROL is always strictly optimal for a sufficiently large attainability probability p_1 at the most preferred school.[28] Hence, we can focus on non-truthful, top-choice monotone ROLs.

We say that the joint distribution of attainability probabilities has full support if for any subset of schools without the safe school m the probability of these schools being attainable and these schools only is strictly between zero and one. We now show by induction over the number of schools that for any non-truthful top-choice monotone ROL there exists an environment $\xi = (P, (v_s)_{s \in \mathcal{S}})$ with full support on the distribution of attainability probabilities such that the ROL is strictly optimal.

For the base case $\mathcal{S}_2 = \{1, 2\}$, Lemma 5 establishes that $(2, 1)$ is strictly optimal for any $v_1 > v_2$ if p_1 satisfies

$$\frac{p_1}{\Lambda} < p_1 \left(-p_1 + 1\right) \quad \Leftrightarrow \quad 0 < p_1 < 1 - \frac{1}{\Lambda}.$$

For the induction step suppose that the statement holds for any set of $m - 1$ schools. Let L be an arbitrary non-truthful top-choice monotone list for the set $\mathcal{S}_m = \{1, ..., m\}$. Since any such ROL must rank school 1 behind school 2, list L must be of the form $([a], 2, [b], 1, [c])$, where $[a]$, $[b]$, and $[c]$ stand for some (potentially empty) ordered subsets of schools $\{3, 4, \ldots, m\}$.

By induction assumption, for the school set $\mathcal{S}_{m-1} = \{\tilde{2}, 3, ..., m\}$ there is some $\widetilde{\xi} = (\widetilde{P}, (\widetilde{v}_s)_{s \in \mathcal{S}_{m-1}})$ with full support such that $\widetilde{L} = ([a], \widetilde{2}, [b], [c])$ is strictly optimal. Intuitively, we now construct the environment for set $\mathcal{S}_m$ that makes L strictly optimal by splitting school $\widetilde{2}$ into two schools, 1 and 2. Formally, we define $v_1 > v_{\widetilde{2}}$ and $v_s = v_{\tilde{s}}$ for all $s \geq 2$. Attainability probabilities are defined exactly as on the set $\mathcal{S}_{m-1}$ with the additional assumption that whenever school $\tilde{2}$ is attainable school 1 but not 2 is attainable with probability ϵ, school 2 but not 1 with $1 - 2\epsilon$, and both schools 1 and 2 with probability ϵ.[29] Note that the resulting distribution has full support. We need to show that for suitable choices of v_1 and ϵ list L becomes strictly optimal. Foreshadowing the structure of the argument, we split the remaining proof into proving the following three claims.

(i) If $p_1 = \epsilon p_{\widetilde{2}} < \frac{1-1/\Lambda}{2}$, then there is some $\overline{v} > v_2$ such that for $v_1 = \overline{v}$ any optimal

[28]Note, that this is not a circular argument, as Proposition 2 only builds on Proposition 1a.

[29]More formally, for any $E \in \Pi_{s=3}^{m} A_s$ we define $P(A_1 = A_2 = 1, E) = \epsilon \widetilde{P}(A_{\widetilde{2}} = 1, E)$, $P(A_1 = 1, A_2 = 0, E) = \epsilon \widetilde{P}(A_{\widetilde{2}} = 1, E), P(A_1 = A_2 = 0, E) = \widetilde{P}(A_{\widetilde{2}} = 0, E)$, $P(A_1 = 0, A_2 = 1, E) = (1 - 2\epsilon)\widetilde{P}(A_{\widetilde{2}} = 1, E)$.

ROL ranks school 1 *after outside option* m.

(ii) *For sufficiently small* ϵ *and* $v_1 \leq \overline{v}$, *any strictly optimal ROL of set* $\mathcal{S}_m$ *ranks schools* $\{2, \dots, m\}$ *in the order* $([a], 2, [b], [c])$.

(iii) *For sufficiently small* ϵ *there is some* $v_1 \in (v_2, \overline{v}]$ *such that* $([a], 2, [b], 1, [c])$ *is strictly optimal.*

Proof of Claim (i): For any ROL that does not rank school 1 last, consider Lemma 5 for a switch of ranks between school $x = 1$ and school y ranked directly after it. Define

$$\alpha(v_1, v_s, v_y) = \frac{v_1 + v_y - 2v_s}{v_1 - v_y} = 1 - 2\frac{v_s - v_y}{v_1 - v_y}$$

and let $\underline{\alpha}(v_1) = \min_{2 \leq s < y \leq m} \alpha(v_1, v_s, v_y)$. Note that $\underline{\alpha}(v_1)$ is strictly increasing in v_1 with $\underline{\alpha}(v_1) \in (-1, 1)$ for all v_1, and $\lim_{v_1 \to \infty} \underline{\alpha}(v_1) = 1$.

Considering the swap of 1 and y, we have

$$-f_1 + \varepsilon + \sum_{s=2}^{y-1} f_s \, \alpha(v_1, v_s, v_y) + \sum_{s=y}^{m} f_s > -f_1 + \underline{\alpha}(v_1) \sum_{s=2}^{m} f_s > 1 - 2f_1 + (\underline{\alpha}(v_1) - 1),$$

as $\sum_{s=2}^{m} f_s = (1 - f_1)$ and $\underline{\alpha}(v_1) < 1$. If $p_1 < \overline{p} \equiv \frac{1 - 1/\Lambda}{2}$, then $f_1 \leq p_1$ implies

$$1 - 2f_1 + (\overline{\alpha}(v_1) - 1) > 1 - 2\overline{p} + (\overline{\alpha}(v_1) - 1) = {}^1\!/\!_\Lambda + (\overline{\alpha}(v_1) - 1).$$

Since the inequality is strict and $\lim_{v_1 \to \infty} (\overline{\alpha}(v_1) - 1) = 0$, there is some $\overline{v}$, for which swapping 1 and y is profitable by Lemma 5 as $1 - 2\overline{p} + (\overline{\alpha}(v_1) - 1) > {}^1\!/\!_\Lambda$ and $\varepsilon > 0$ by full support. Hence, it is always profitable to switch ranks of 1 and y for any school y ranked behind it, and, by iteration, ranking 1 before m is strictly suboptimal for $v_1 = \overline{v}$.

Proof of Claim (ii): Intuitively, for small ϵ the position of school 1 is unsubstantial for expected utility such that the optimality of the order $([a], 2, [b], [c])$ follows from the strict optimality of $([a], \tilde{2}, [b], [c])$.

More formally, let $L' = ([d], 1, [e])$ be an arbitrary ROL of school set $\mathcal{S}_m$. We calculate a bound on the utility difference between L' with attainability distribution P and ROL $\tilde{L}' = ([d], [e])$ under school set $\mathcal{S}_{m-1}$ and attainability distribution $\widetilde{P}$ (indentifying school 2 with school $\tilde{2}$).

Let $p_1 = \epsilon p_{\tilde{2}}$ be the unconditional probability of school 1 being attainable. Denote with $(f'_s)_{s \in \mathcal{S}_m}$ the match probabilities of list L' and with $(\tilde{f}'_s)_{s \in \mathcal{S}_{m-1}}$ the match utilities of list $\tilde{L}'$. Let further be $\Delta U([d], [e])$ the absolute value of the utility difference between L' and $\tilde{L}'$. Employing triangle inequality and the fact that the match probabilities with schools

$2,...,m$ each differ by less then p_1 between ROLs L' and $\tilde{L}'$, we obtain by Equation 7

$$\begin{aligned}\Delta U([d],[e]) &\leq \sum_{s=1}^{m}|\tilde{f}'_s - f'_s|v_s + \Lambda\sum_{s=1}^{m}\Big(\Big|f'_s\sum_{r=s+1}^{m}f'_r(v_s-v_r) - \tilde{f}'_s\sum_{r=s+1}^{m}f'_r(v_s-v_r)\Big| \\ &\quad + \tilde{f}'_s\sum_{r=s+1}^{m}p_1(v_s-v_r)\Big) \\ &< \sum_{s=1}^{m}p_1\overline{v} + \Lambda\sum_{s=1}^{m}\left(p_1\overline{v}\sum_{r=s+1}^{m}f'_r + \tilde{f}'_s(m-s-1)p_1\overline{v}\right) \\ &< mp_1\overline{v} + \Lambda 2mp_1\overline{v} = \epsilon p_{\tilde{2}}m\overline{v}(1+2\Lambda).\end{aligned}$$

As $([a],\tilde{2},[b],[c])$ is strictly optimal for for $\mathcal{S}_{m-1}$ and $\tilde{P}$, let C be the utility difference to the second best ROL of $\mathcal{S}_{m-1}$. Let $\epsilon < \frac{C}{2p_{\tilde{2}}m\overline{v}(1+2\Lambda)}$. Hence, whenever the order of schools $\{2,...,m\}$ in $([d],[e])$ is different to the order in $([a],2,[b],[c])$ we know that the utility of $([a],2,1,[b],[c])$ exceeds the utility of $([d],1,[e])$ by more that

$$-\Delta U([a],2,[b],[c]) + C - \Delta U([d],[e]) > -C/2 + C - C/2 = 0,$$

which shows that any optimal ROL of n schools ranks schools $2,...,m$ in the order $([a],2,[b],[c])$.

Proof of Claim (iii): By top choice monotonicity, $([b],[c])$ is of form $(k,k+1,...,m)$ with $k\geq 3$.. Also by top choice monotonicity, it is suboptimal to rank 1 behind 2. We need to show that for sufficiently small ϵ any position of school 1 in $(k,k+1,...,m)$ can be achieved as strictly optimal by choosing v_1 appropriately. We employ Lemma 5 to consider a swap of school $x=1$ and school y ranked directly behind 1 (if any). For fixed match probabilities $f_1,...,f_m$, the term in brackets on the right-hand side of (9) is strictly decreasing in y for any v_1, since $\alpha(v_1,v_s,v_y)<1$ is increasing in v_y and v_y is decreasing in $y\in\{k,k+1,...,m\}$. Moreover, the amount by which this term decreases depends continuously on v_1, and hence the term attains a minimum strictly larger than zero at some $v_1\in[v_2,\overline{v}]$. Since the match probabilities f_s and ε can differ by at most $p_1=\epsilon\widetilde{p_2}$ between two different ROLs it follows that the term in the brackets of (9) is also strictly decreasing in $y\in\{k,k+1,...,m\}$ for all $v_1\in[v_2,\overline{v}]$ for sufficiently small ϵ. Hence, an order $([a],2,k,...,y-1,1,y,...,m)$ is optimal if and only if there is a v_1 for which the term in the brackets of (9) is strictly larger than $\frac{1}{\Lambda}$ for the adjacent swap of 1 and $y-1$, but strictly smaller than $\frac{1}{\Lambda}$ for the adjacent swap of 1 and y. Since for each y the term in the brackets of (9) is continuously increasing in v_1 and decreasing in y, such $v_1\in(v_2,\overline{v}]$ exists by the intermediate value theorem for all y if $([a],2,1,k,...m)$ is optimal for v_1 sufficiently close to v_2 and $([a],2,k,...,m,1)$ is optimal for $v_1=\overline{v}$. The latter has been shown in Claim (i). For v_1 sufficiently close to v_2 the optimality of $L=([a],2,1,k,k+1,...,m)$ follows from the optimality of $\widetilde{L}=([a],\widetilde{2},k,k+1,...,m)$ for schools $\mathcal{S}_{m-1}$: By Lemma 5, optimality of $\tilde{L}$ implies for this list

$$\frac{1}{\Lambda} > -f_{\tilde{2}} + \varepsilon + \sum_{s=3}^{k-1} f_s \frac{v_{\tilde{2}} + v_k - 2v_s}{v_{\tilde{2}} - v_k} + \sum_{s=k}^{m} f_s.$$

Since the match probabilities for schools 3,...,m are the same in $([a], 2, 1, k, k+1, ..., m)$ and $\tilde{L}$, the match probabilities f_1 and f_2 satisfy $f_1 + f_2 = f_{\tilde{2}}$, and for $v_2 = v_{\tilde{2}}$, we have

$$\begin{aligned} \frac{1}{\Lambda} &> -f_1 + \frac{v_2 + v_k - 2v_2}{v_2 - v_k} f_2 + \epsilon\varepsilon + \sum_{s=3}^{k-1} f_s \frac{v_2 + v_k - 2v_s}{v_2 - v_k} + \sum_{s=k}^{m} f_s \\ &= \lim_{v_1 \to v_2} \left(-f_1 + \epsilon\varepsilon + \sum_{s=2}^{k-1} \frac{v_1 + v_k - 2v_s}{v_1 - v_k} f_s + \sum_{s=k}^{m} f_s \right), \end{aligned}$$

which shows that it is unprofitable to swap 1 and k in $([a], 2, 1, k, k+1, ..., m)$ when v_1 is sufficiently close to v_2 . This concludes the proof of Claim (iii). □

Proof of Proposition 2. 1. Suppose, by way of contradiction, that truth-telling is suboptimal. By Proposition 1a, this implies that there is an optimal ROL which does not rank school 1 first. Let $\overline{F} = (\overline{f}_1, ..., \overline{f}_m)$ be the lottery induced by thatl ROL. Let $F = (f_1, ..., f_m)$ be the lottery induced by the ROL which ranks school 1 first and all other schools in the same order as the optimal ROL. Let further $\varepsilon_i = f_i - \overline{f}_i$ be the shifts in probability when choosing F instead of $\overline{F}$. Evidently, $\varepsilon_i \leq 0$ for all $i \geq 2$ and $\varepsilon_1 = -\sum_{i=2}^{m} \varepsilon_i > 0$, where the strict inequality comes from the fact that school 1 is not exclusive. To find a contradiction to the optimality of $\overline{F}$, we show that $U(\cdot, F) > U(\cdot, \overline{F})$. By (7),

$$\begin{aligned} U(\cdot, F) - U(\cdot, \overline{F}) &= \sum_{s=1}^{m} f_s \left[v_s - \Lambda \sum_{r=s+1}^{m} f_r (v_s - v_r) \right] - \sum_{s=1}^{m} \overline{f}_s \left[v_s - \Lambda \sum_{r=s+1}^{m} \overline{f}_r (v_s - v_r) \right] \\ &\geq \sum_{s=1}^{m} f_s \left[v_s - \Lambda \sum_{r=s+1}^{m} f_r (v_s - v_r) \right] - \sum_{s=1}^{m} \overline{f}_s \left[v_s - \Lambda \sum_{r=s+1}^{m} f_r (v_s - v_r) \right] \\ &= \left(-\sum_{i=2}^{m} \varepsilon_i \right) \left[v_1 - \Lambda \sum_{r=2}^{m} f_r (v_1 - v_r) \right] + \sum_{s=2}^{m} \varepsilon_s \left[v_s - \Lambda \sum_{r=s+1}^{m} f_r (v_s - v_r) \right] \\ &= \sum_{s=2}^{m} \varepsilon_s \left[v_s - v_1 + \Lambda \sum_{r=2}^{s} f_r (v_1 - v_r) + \Lambda \sum_{r=s+1}^{m} f_r (v_1 - v_s) \right] \\ &> \sum_{s=2}^{m} \varepsilon_s \left[v_s - v_1 + \Lambda \sum_{r=2}^{s} f_r (v_1 - v_s) \right] \\ &= \sum_{s=2}^{m} \varepsilon_s \left[v_s - v_1 + \Lambda (1 - p_1)(v_1 - v_s) \right] > 0, \end{aligned}$$

where the last inequality exploits that, by assumption, $p_1 > 1 - 1/\Lambda$.

2. Suppose, by way of contradiction, that truth-telling is strictly optimal. Hence,

$(1, 2, ..., m) \succ_i (2, 1, ..., m)$, and by Lemma 5

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

where ε is the probability that the student is attainable at schoo 1 and school 2. Since school 1 is not exclusive, we have $\varepsilon > 0$, and hence

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

which can be rearranged to $p_1 > 0.5\,(1 - {}^{1}/_{\Lambda})$, a contradiction. □

Proof of Proposition 3. The fact that all schools use the same deterministic tie-breaking rule for students of the same score is isomorphic to an assumption that no two students share the same score in their supports. The existence of a pure strategy equilibrium then follows iteratively by scores.

Start with the highest possible score. As this type will be accepted at any school, she submits (according to the fixed tie-breaking rule) an ROL that lists her most preferred school (which, of course, depends on $\mathbf{v}_i$) first. Next, consider the second highest possible score and the associated student with her preference profile. If the initial support of scores is asymmetric between students, this could be the same student. In that case, she knows again that no other student has a higher score and will report the true ROL. If another student holds this score, she correctly infers the probability that another student has a higher score and the probability distribution over her submitted ROLs. From that, she correctly infers the distribution over her attainability probabilities and picks her best-responding 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 2. By (7), an ROL which lists any subset of elite schools before the outside option induces an expected utility of $fv - \Lambda f(1-f)v$, where f is the probability that at least one elite school of the subset is attainable. Since the utility is a convex function in f, it is maximized by either maximizing or minimizing f. Hence, by either listing all or none of the elite schools before the outside option. □

Proof of Lemma 3. The probability that there are less than capacity q students with a score above ω among $(n-1)$ students,

$$P_{n-1:q}(\omega) := \sum_{k=0}^{q-1} \binom{n-1}{k} (1-\omega)^k \omega^{n-1-k}, \tag{14}$$

is continuously and monotonically increasing in ω from 0 to 1. Thus, there is a unique $\overline{\omega}^l$

such that $P_{n-1:q}(\overline{\omega}^l) = 1 - {}^1\!/_{\Lambda^l} \in (0,1)$. Because $f(\omega)$ is minimal when all other students choose to apply, $f(\omega) \geq P_{n-1:q}(\omega)$ for all ω and all reporting strategies of the other students. Hence, for any $\Lambda \leq \Lambda^l$ and any $\omega \geq \overline{\omega}^l$, we have $f(\omega) \geq P_{n-1:q}(\omega) \geq 1 - {}^1\!/_\Lambda$, meaning that applying to the elite school is a best response for all such types, as (8) holds.

Knowing that all students of score $\omega \geq \overline{\omega}^l$ apply, a student of type Λ^l infers that for score $\omega \geq \overline{\omega}^l$ she has attainability probability $f(\omega) = P_{n-1:q}(\omega)$, and by construction applies if and only if her score satisfies $\omega \geq \overline{\omega}^l$.

Next, because $\Lambda^{l-1} < \Lambda^l$, there are types $\omega < \overline{\omega}^l$ who prefer to apply as well. A student of score $\omega < \overline{\omega}^l$ and sufficiently close to $\overline{\omega}^l$ expects acceptance if there are less than q other students with a score either above $\overline{\omega}^l$ or a score in $[\omega, \overline{\omega}^l]$ and $\Lambda < \Lambda^l$. Again, this attainability probability is strictly and continuously increasing in ω which implies a unique cutoff $\overline{\omega}^{l-1}$ such that $f(\overline{\omega}^{l-1}) = 1 - {}^1\!/_{\Lambda^{l-1}}$. Hence, truthful reporting for type Λ^{l-1} is optimal if and only if $\omega > \overline{\omega}^{l-1}$. Proceeding with this manner iteratively, we obtain an essentially unique CBNE. $\square$

Proof of Lemma 4. If DSPDA is truthful, it is a static mechanism that 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 for all realizations. 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, the associated direct mechanism $\left(\prod \mathfrak{S}(\mathcal{S}), o \circ (\sigma_1, ..., \sigma_n)\right)$ has a truthful CBNE by construction, and implements the student-optimal stable allocation. This direct mechanism asks students for their type vector and implements the student optimal-stable stable match (hence the DSPDA match outcome) based on the ordinal preferences of their cardinal utility vector. As in this truthful equilibrium students have no incentive to misrepresent their cardinal utility vector $\mathbf{v}_i$ they have a forteriori no incentive to misrepresent in DSPDA, which only asks for ordinal preferences to implement the student-optimal stable match. Hence, DSPDA is truthful. $\square$

Proof of Proposition 4. Fix some student A, label schools according to her preference order (as in Section 3.1), and suppose that all other students behave truthfully in the mechanism. We start by showing that the decision of whether to accept or reject a school does not change the probability of receiving an offer from a more preferred school. We then show that truthfulness is an SCPE such that all students best-respond to each other.

Suppose A at some stage in the mechanism receives an offer from a school when she currently holds no other offer. If A rejects rather than accepts the school, then all students will receive weakly more proposals. Indeed, the rejected school will potentially send an additional proposal to another student B in the next step. If B rejects the proposal, the school will offer the seat to yet another student C in the subsequent round. If B accepts the proposal in favor of another school, then this other school will offer an additional seat in the next round. By iterating this argument, we can conclude that due to the initial rejection all students, including A, 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 was strict and A accepted all proposals by more preferred schools, then the rejection can implement a stable match outcome, which is weakly preferred by all students. This contradicts the uniqueness of the stable match. Hence, the decision of the student to accept or reject a school can only affect attainability at schools that she prefers less.

Next, suppose A accepts or rejects some school k while holding (or simultaneously receiving) an offer by some school ℓ. We show that this decision does not change the probability of receiving an offer from any school preferred to both, ℓ and k. 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 the offer from k occurs, since both strategies induce the same cascade of school proposals and students are truthful. According to the previous paragraph, rejecting both schools instead does not change the probability of proposals from schools preferred to k. Similarly, we obtain the same proposal probabilities for schools preferred to ℓ when accepting ℓ and rejecting k. Hence, 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 student A. Going backwards in her decision tree, 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 A at this decision node. Let $F_R = (f_1, ..., f_n)$ be the lottery over match outcomes if A rejects k, and let $\tilde{F}_A = (\tilde{f}_1, ..., \tilde{f}_n)$ be the respective lottery if she accepts. We distinguish three cases:

(i) The student currently holds an offer from a school $\ell < k$ such that rejecting would be truthful. Since, by assumption, truthfulness is optimal in all subsequent decision nodes, $f_i = \tilde{f}_i = 0$ for $i > k$. As shown above, the probability of proposals from schools $i < \ell$ does not depend on the decision about k such that $f_i = \tilde{f}_i$ for $i < \ell$ and $f_\ell = \sum_{i=\ell}^{k} \tilde{f}_i$. Hence, expected payoffs at the decision node satisfy

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

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

and truthful rejection is optimal, a contradiction.

(ii) The student currently holds no offer from a school preferred to k, and $v_k > v_m = 0$ such that school k is preferred to remaining unmatched and accepting would be truthful. Analogously to (i), $f_i = \tilde{f}_i$ for all $i < k$ and $\tilde{f}_k = \sum_{i=k+1}^{m} f_i$. Then,

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

and truthfully accepting is optimal, a contradiction.

(iii) The student currently holds no offer from a school preferred to k, and $v_k < v_m = 0$, such that rejecting would be truthful. As A is assumed to be truthful in all following decision nodes and is always free to take the outside option in later steps, the decision only makes a difference if the game ends after it, i.e., $f_m = \tilde{f}_k$. Hence,

$$
\begin{aligned}
U(\cdot, F_R) &= \sum_{i=1}^{m-1} f_i v_i + f_m v_m - \Lambda \sum_{1 \leq i \leq j \leq m-1} f_i f_j (v_i - v_j) - \Lambda \sum_{i=1}^{m} f_i f_m (v_i - v_m) \\
&\geq \sum_{i=1}^{m-1} \tilde{f}_i v_i + \tilde{f}_k v_k - \Lambda \sum_{1 \leq i \leq j \leq m-1} \tilde{f}_i \tilde{f}_j (v_i - v_j) - \Lambda \sum_{i=1}^{m} \tilde{f}_i \tilde{f}_k (v_i - v_k) \\
&= U(\cdot, \tilde{F}_A),
\end{aligned}
$$

and truthfully rejecting is optimal, a contradiction. All in all, truth-telling is an SCPE against others' truthfulness, completing the SCBNE. □

A.IV 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%
TCM	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 top-choice monotone ROLs are marked, and the frequencies of the most common misrepresentations for each priority score are in bold.

References

Abdulkadiroğlu, A., Sönmez, T., 2003. School choice: A mechanism design approach. American economic review 93, 729–747.

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

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

Artemov, G., Che, Y.K., He, Y., 2020. 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, 73–94.

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

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

Bell, D.E., 1985. Disappointment in decision making under uncertainty. Operations Research 33, 1–27.

Brown, A., Imai, T., Vieider, F., Camerer, C., 2021. Meta-analysis of empirical estimates of loss-aversion .

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

Chen, L., Pereyra, J.S., 2019. Self-selection in school choice. Games and Economic Behavior 117, 59–81.

Chen, Y., Sönmez, T., 2006. School choice: an experimental study. Journal of Economic Theory 127, 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, 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.

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, 485–494.

Eeckhout, J., 2000. On the uniqueness of stable marriage matchings. Economics Letters 69, 1–8.

Ehlers, L., Massó, J., 2007. Incomplete information and singleton cores in matching markets. Journal of Economic Theory 136, 587–600.

Ericson, K.M.M., Fuster, A., 2011. Expectations as endowments: Evidence on reference-dependent preferences from exchange and valuation experiments. The Quarterly Journal of Economics 126, 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, 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, 9–15.

Gill, D., Prowse, V., 2012. A structural analysis of disappointment aversion in a real effort competition. American Economic Review 102, 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, 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, 1283–1309.

Gul, F., 1991. A theory of disappointment aversion. Econometrica: Journal of the Econometric Society , 667–686.

Hakimov, R., Kübler, D., 2020. Experiments on centralized school choice and college admissions: a survey. Experimental Economics , 1–55.

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

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

Heffetz, O., 2021. Are reference points merely lagged beliefs over probabilities? Journal of Economic Behavior & Organization 181, 252–269.

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

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

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

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

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

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

Karle, H., Engelmann, D., Peitz, M., 2019. Student Performance and Loss Aversion. Technical Report.

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

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

Karle, H., Schumacher, H., 2020. Consumer Search and the Uncertainty Effect. Technical Report.

Karpov, A., 2019. A necessary and sufficient condition for uniqueness consistency in the stable marriage matching problem. Economics Letters 178, 63–65.

Klijn, F., Pais, J., Vorsatz, M., 2019. Static versus dynamic deferred acceptance in school choice: Theory and experiment. Games and Economic Behavior 113, 147–163.

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

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

Kőszegi, B., Rabin, M., 2009. Reference-dependent consumption plans. American Economic Review 99, 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, 634–645.

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

Loomes, G., Sugden, R., 1986. Disappointment and dynamic consistency in choice under uncertainty. The Review of Economic Studies 53, 271–282.

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

Milgrom, P.R., Weber, R.J., 1985. Distributional strategies for games with incomplete information. Mathematics of Operations Research 10, 619–632.

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

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

Niederle, M., Yariv, L., 2009. Decentralized matching with aligned preferences. Technical Report. National Bureau of Economic Research.

Pathak, P.A., Sönmez, T., 2008. Leveling the playing field: Sincere and sophisticated players in the boston mechanism. American Economic Review 98, 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, 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, 129–57.

Rabin, M., Weizsäcker, G., 2009. Narrow bracketing and dominated choices. American Economic Review 99, 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, 11471–11476.

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

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

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, 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, 1456–1499.

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

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

von Wangenheim, J., 2020. English versus Vickrey Auctions with Loss Averse Bidders. Technical Report.