LOTTERY DESIGN FOR SCHOOL CHOICE

NICK ARNOSTI, UNIVERSITY OF MINNESOTA

ABSTRACT. We study outcomes of the Deferred Acceptance Algorithm in large random matching markets where priorities are generated either by a single lottery or by independent lotteries. In contrast to prior work, our model permits students to submit lists of varying lengths, and schools to vary in their popularity and number of seats. In a limiting regime where the number of students and schools grow while the length of student lists and number of seats at each school remain bounded, we provide exact expressions for the number of students who list l schools and match to one of their top k choices, for each $k \leq l$. We use these expressions to provide three main insights.

First, we identify a persistent tradeoff between using a single lottery and independent lotteries. For students who submit short lists, the rank distribution under a single lottery stochastically dominates the corresponding distribution under independent lotteries. However, the students who submit the longest lists are always more likely to match when schools use independent lotteries.

Second, we compare the total number of matches in the two lottery systems. We find that shape of the list length distribution plays a key role. If this distribution has an increasing hazard rate, then independent lotteries match more students. If it has a decreasing hazard rate, the comparison reverses. To our knowledge, this is the first analytical result comparing the size of stable matchings under different priority rules.

Finally, we study the fraction of assigned students who receive their first choice. Under independent lotteries, this fraction may be arbitrarily small, even if schools are equally popular. Under a single lottery, we provide a tight lower bound on this fraction which depends on the ratio r of the popularity of the most to least popular school. When each school has a single seat, the fraction of assigned students who receive their first choice is at least $\sqrt{r}/(1+r)$. This guarantee increases to $2/(1+\sqrt{r})$ as the number of seats at each school increases.

1. Introduction

Many cities across the world allow students to choose which public school they will attend. In the United States, 56% of major school districts have a school choice program, up from 29% in 2000 (Whitehurst, 2017). Advocates argue that school choice offers increased equity and efficiency relative to neighborhood assignment: students from neighborhoods with undesirable schools can go elsewhere, and students can select schools with curricula and clubs that match their interests. These advantages are mitigated by capacity constraints, which prevent schools from admitting every student who wishes to attend. As a result, districts must have a policy to determine who gets seats in popular schools.

The Deferred Acceptance algorithm is an increasingly popular approach, having been adopted in cities across the United States and Europe.[1] This algorithm was introduced by Gale and Shapley (1962), and is also used to match doctors to residencies at US hospitals (Roth and Peranson, 1999; Abdulkadiroglu and Sönmez, 2003). One key input to this algorithm is a priority list (ranking of students) for each school, which determines who to reject when too many students express interest. Students are often ranked using coarse criteria such as whether they live within a designated zone, or have a sibling who attends the school. This results in many ties, which must be broken in order to generate a strict ranking of students. Pathak (2016) remarks that "Questions related to tie-breaking have re-appeared in nearly every city I have interacted with using [Deferred Acceptance]."

There are two common tie-breaking procedures. One uses independent lotteries at each school: if two students are tied at multiple schools, each school flips a coin to decide which student to prioritize. The second uses a single district-wide lottery: for any two students, one will receive priority over the other at all schools where they are tied. Parents often feel intuitively that independent lotteries are "more fair," but academics have cautioned that they may produce inefficient outcomes ex post (Abdulkadiroglu and Sönmez, 2003; Abdulkadiroglu et al., 2009). To provide guidance to practitioners, Abdulkadiroglu et al. (2009) compare these tie-breaking procedures using simulations based on data from Boston and New York. They find that using a single lottery gives more students their first choice, but leaves more students unassigned. de Haan et al. (2018) observe a similar pattern in Amsterdam. Pathak (2011) writes, "these empirical results raise the need for quantitative results in matching theory that provide guidance on what features of the student preferences and school priorities are responsible for these differences."

The goal of this paper is to answer this call by comparing the distribution of match outcomes under the two tie-breaking procedures. One outcome of particular interest is the number of students who fail to match to any of the schools on their list. This can happen even when there are more seats than students, because students typically list only a limited number of schools. As a result, 3.8% of students in Boston, 7.1% of students in New York (Abdulkadiroglu et al., 2009), and 18.7% of students in New Orleans (EnrollNOLA, 2017) fail to match to any of their listed schools. In some cities, these students are administratively assigned to a school with unfilled seats; in others,

[1] American cities using Deferred Acceptance include New York City (Abdulkadiroglu et al., 2005a), Boston (Abdulkadiroglu et al., 2005b), Chicago (Pathak and Sönmez, 2013), New Orleans (Abdulkadiroglu et al., 2017), Denver, Washington DC, and Newark (Pathak, 2016). Internationally, it is used to in Amsterdam (de Haan et al., 2018), Paris (Fack et al., 2019), and London (Pathak and Sönmez, 2013), as well as for higher-education in many countries (see Table 1 by Fack et al. (2019)).

they are invited to apply in a second round.[2] Regardless of which approach is taken, these students are often disappointed with the process, and may leave for a private school, generate negative headlines, or create administrative burdens by challenging the outcome. For these reasons, school districts are motivated to match as many students as possible to one of the schools on their list.

Which lottery procedure leaves fewer students unassigned? Although recent papers by Ashlagi et al. (2017), Ashlagi et al. (2019), and Ashlagi and Nikzad (2020) study match outcomes under single and independent lotteries, none of them address this question. All three papers assume that students submit complete lists, implying that all students will match so long as there are more seats than students.[3] By contrast, we consider a model in which students submit incomplete lists of varying lengths. In addition, we relax the assumption of homogeneous schools in the aforementioned papers by allowing schools to differ in their capacity and popularity. The generality of our model allows us to (i) identify how the choice of lottery procedure affects the total number of matches, as well as the match rate for different student populations, and (ii) identify how the number of students getting their first choice when using a single lottery depends on the relative popularity of different schools.

We summarize our key findings below, while highlighting connections to a few closely related papers. Section 2 provides a more complete discussion of prior work. Sections 3 and 4 present our model and results, and Section 5 concludes by discussing policy implications of our work, as well as its limitations.

Exact Asymptotics. We consider a model with a finite number of students and schools, in which schools may have different capacities. Each student's list is drawn through a two-stage process. First, the length of the list is sampled from a distribution on $\mathbb{N}$. Next, the list is generated by sampling schools without replacement from a fixed distribution over schools. This procedure enables us to capture features of real-world markets, such as that students submit incomplete lists of varying lengths, and some schools are more popular than others. We assume that school priorities are generated either using a single lottery, or using independent lotteries at each school. Finally, a matching is determined using the student-proposing Deferred Acceptance algorithm.

We analyze our model by letting the number of students and schools grow, while the number of schools listed by each student and the capacity of each school remain bounded. In this limit, Immorlica and Mahdian (2005) and Kojima and Pathak (2009) show that there is a "nearly unique" stable matching, but do not describe this matching in detail. We do, by showing that aggregate outcome measures concentrate around a deterministic limit which we give explicitly. More specifically, for any integers $k \leq l$, Theorem 1 provides exact asymptotic expressions for the fraction of students who list l schools and match to one of their top k choices under each lottery procedure. This result requires a more precise analysis than those of Immorlica and Mahdian (2005) and Kojima and Pathak (2009), forcing us to use different proof techniques. Our proof constructs Markov Chains corresponding to each procedure, and then applies the differential equation method

[2] New York changed from the latter procedure to the former for students entering high school in 2020.

[3] Ashlagi et al. (2019) briefly consider independent lotteries with incomplete lists, and establish that when there are more seats than students, the number of unassigned students is bounded by a function which decreases exponentially in the length of student lists. However, their bound is not tight, and cannot be used to compare unassigned students across lottery procedures.

of Wormald (1999) to establish that certain match outcomes concentrate around the solution to a set of differential equations. We describe our proof technique in more detail in Section 3.

The limiting expressions from Theorem 1 provide significant insight into match outcomes. Our remaining results use these expressions to compare outcomes under single and independent lotteries.

Establishing a Tradeoff. Theorem 2 establishes that a single lottery always assigns more students to their first choice school. This is consistent with the findings of Ashlagi et al. (2019), and Ashlagi and Nikzad (2020), although our model is more general in several ways. Theorem 2 also establishes a single-crossing property that implies a tradeoff between these lottery procedures: although a single lottery gives more students their first choice, it also results in a lower probability of matching for students who submit the longest lists.

It is fairly intuitive that students with longer lists are those most likely to benefit from the use of independent lotteries, as they get the most independent draws. However, a priori it seems possible that one lottery procedure could dominate the other. The intuition for why this cannot occur is as follows. If student outcomes were uniformly better under one lottery procedure, then that procedure would result in fewer total proposals by students. However, fewer proposals also implies fewer matches, contradicting the assumption that one procedure dominates the other. This reasoning is formalized by Lemma 7 in the Appendix.

The tradeoff identified by Theorem 2 parallels a finding of Ashlagi and Nikzad (2020) when there are fewer students than seats. However, it may seem at odds with their finding that a single lottery almost dominates independent lotteries when there is a shortage of seats. The results can be reconciled by noting that in our model, if there is a shortage of seats and all students submit reasonably long lists, the crossing point occurs in the far right tail (see Figure 1). As such, the findings can be viewed as complementary: our work shows that a tradeoff always exists, and they point out circumstances in which this tradeoff almost vanishes. However, we also show that the dominance of a single lottery when there is a shortage of seats partly depends on the assumption that students submit lists of the same length. When some students submit short lists while other submit long ones, using a single lottery may harm the latter group. We elaborate upon this point when discussing policy implications of our work in Section 5.

Comparing the Total Number of Matches. Which lottery procedure matches more students? When all students submit lists of the same length, Theorem 2 implies that the answer is independent lotteries. However, there are also cases where a single lottery yields more matches. Theorem 3 and Proposition 1 state that in general, the answer depends on the distribution of student list lengths: if it has an increasing hazard rate, independent lotteries will assign more students, while the opposite holds if the distribution has a decreasing hazard rate. Interestingly, these conclusions hold for *any* joint distribution of school popularity and capacity. To our knowledge, our work is the first to analytically compare the number of assigned students under different priorities, and the first to identify the shape of the list length distribution as a parameter of interest.

We now attempt to give intuition for why the shape of the list length distribution is so crucial. Consider a scenario where student A and student B have both proposed to a school, which A lists in position k_A, and B lists in position $k_B > k_A$. Suppose that the school has room for only one of these students. We ask two questions. First, which student is most likely to be rejected? And if rejected, which student is more likely to go unmatched?

The answer to the first question does not depend on the list length distribution. With independent lotteries, the school is equally likely to accept each student. With a single lottery, the fact that B has been rejected from more schools than A suggests that B is likely to have the lower lottery number. Therefore, the school is more likely to accept A, and reject B.

The answer to the second question depends on the list length distribution:

- If both students list the same number of schools, then B faces a greater risk of going unmatched than A: because B has been rejected from more schools, she has fewer remaining opportunities to match.
- If most students are "selective" and submit lists of length k_A, while a few "unselective" students submit very long lists, then we know that B is unselective (and thus likely to find a match somewhere), but suspect that A is selective (and thus at risk of going unassigned).

When the list length distribution has an increasing hazard rate, then students proposing to schools further down their list (such as B in this example) are likely to have fewer options remaining, and therefore face the greatest risk of going unassigned. In these cases, a single lottery (which is more likely to reject B) will result in fewer matches. This reasoning reverses if the list length distribution has a decreasing hazard rate.

In practice, list length distributions from New York, Amsterdam, Hungary, and Chile (shown in Figure 4) have neither an increasing nor a decreasing hazard rate. However, all four distributions satisfy a weaker condition under which Theorem 3 predicts that independent lotteries will result in more matches. This is consistent with simulation results from Boston and New York (Abdulkadiroglu et al., 2009) and Amsterdam de Haan et al. (2018). Our work gives an explanation for these observations, and suggests that typically, independent lotteries will match more students.

We also demonstrate that the difference in the number of unassigned students can be significant. Figure 5 provides an example where moving from a single lottery to independent lotteries reduces the number of unassigned students from 6.4% to 3.4%. Under a single lottery, reducing the number of unassigned students below 3.4% would require doubling the length of student lists, or increasing the number of schools by 8%. Proposition 2 establishes the rate at which the number of unassigned students goes to zero under each lottery procedure, as the average list length ℓ increases. In a balanced market, this number decays exponentially in $\sqrt{\ell}$ with independent lotteries, but decays at a rate of $1/\ell$ with a single lottery.

Bounding the Number of Students Matched to their First Choice. Although our focus is on comparing the number of matches under each procedure, the expressions from Theorem 1 are also useful in other ways. We illustrate this fact by studying the fraction of assigned students who match to their first choice school. In markets where schools are equally popular and each have a single seat, Ashlagi et al. (2019) show that this fraction can be arbitrarily small with independent lotteries, but is always at least 1/2 with a single lottery. Our Proposition 3 generalizes this result to cases where schools each have $C \geq 1$ seats and differ in their popularity.

Increasing the number of seats C improves the lower bound for a single lottery: when schools are equally popular, the lower bound of $1/2$ for $C = 1$ increases to $5/8$ when $C = 2$, is above $3/4$ when $C = 5$, and approaches 1 as $C \to \infty$.

Meanwhile, when some schools are more popular than others, this lower bound decreases, because many students have the same first choice school. Proposition 3 quantifies this effect. When the most popular school is listed r times as often as the least popular school, it establishes that the fraction of assigned students who get their first choice is at least $\sqrt{r}/(1+r)$. Thus, the bound of 50% when $r = 1$ decays to 40% for $r = 4$ and 30% for $r = 9$. These bounds are tight when schools have a single seat. Proposition 3 also gives (larger) tight lower bounds for any $C > 1$. To our knowledge, Proposition 2 offers the first analysis of how the number of students receiving their first choice depends on the relative popularity of different schools.

2. Related Work

In their seminal paper, Gale and Shapley (1962) defined the concept of a stable matching, and proved that the Deferred Acceptance algorithm will always find such a matching. Roth (1984) identified this algorithm was being used by the NRMP to match medical residents to hospitals, and Roth and Peranson (1999) describe the redesign of this algorithm. Abdulkadiroglu and Sönmez (2003) proposed using the Deferred Acceptance algorithm to assign seats at public schools, and it has since been adopted in many cities (Pathak (2016) provides a partial list). There is a vast academic literature on stable matching: for surveys, see Roth and Sotomayor (1990), Pathak (2011), Abdulkadiroglu and Sonmez (2013) and Kojima (2015).

This paper's focus is on lotteries to break ties in priority. Although there are many ways that this could be done, Erdil and Ergin (2008) show that it is impossible to produce a student-optimal stable matching in a strategy-proof manner. Subsequent papers (and real-world implementations) have focused on the two simple procedures that we study: using independent lotteries at each school, and using a single lottery that applies to all schools. Our literature review focuses on work that studies match outcomes under these procedures.

2.1. **Top Trading Cycles with Lotteries.**

It is not a priori obvious that the method of breaking ties matters. For example, the Top Trading Cycles algorithm (attributed to David Gale, introduced by Shapley and Scarf (1974), and adapted for school choice by Abdulkadiroglu and Sönmez (2003)) uses the same input as Deferred Acceptance: preferences, priorities, and capacities. It turns out that under TTC, independent lotteries at each school result in an *identical* distribution of outcomes as using a single lottery! This was first observed by Knuth (1996) and Abdulkadiroglu and Sonmez (1998), and has since been extended to more general settings and mechanisms (Sonmez and Unver, 2005; Pathak and Sethuraman, 2011; Lee and Sethuraman, 2011; Carroll, 2014; Ekici, 2015). Most recently, Bade (2019) showed that given any deterministic, strategy-proof, Pareto efficient and non-bossy allocation rule, randomly permuting the roles of agents yields a distribution over allocations that is equivalent to a random serial dictatorship.[4]

[4]If the district uses TTC with coarse deterministic priorities and lotteries serving only as a tiebreaker, then using a single lottery and using independent lotteries are no longer equivalent.

2.2. **Deferred Acceptance with Lotteries.** This equivalence result does not apply to the Deferred Acceptance algorithm, which is bossy and not Pareto efficient. Indeed, under Deferred Acceptance, the lottery method matters. Abdulkadiroglu and Sönmez (2003) observe that using independent lotteries at each school may produce an outcome that is not Pareto efficient. Abdulkadiroglu et al. (2009) show that for any student-optimal stable matching, there exists a single lottery draw that produces it. However, Pathak (2016) notes that "Their results represent an ex post perspective, and as far as I know, there is no stronger ex ante argument for single versus multiple tie-breaking based on the distribution of matchings."

2.3. **Quantitative Results in Markets with a Large Number of Students and Schools.** Pittel (1989) considers a large one-to-one market with an equal number n of students and schools, complete preferences drawn independently and uniformly at random, and priorities drawn from independent lotteries at each school. His work shows that under the student-proposing Deferred Acceptance algorithm, students' average rank is approximately $\log(n)$. Pittel (1992) shows that under these same assumptions, with high probability every student receives a school ranked no worse than $\log^2(n)$. Ashlagi et al. (2017) relax the assumption that there are equally many students and schools. They show that adding a single student dramatically increases students' average rank, to $n/\log(n)$.

Ashlagi et al. (2017) also study outcomes when using a single lottery, and find that this results in a similar average rank to independent lotteries when there are enough seats for all students, but a much better average rank when there is a shortage of seats. Ashlagi and Nikzad (2020) extend the comparison beyond students' average rank. They show that when there is a shortage seats, the student rank distribution when using a single lottery "almost stochastically dominates" that which arises from independent lotteries. By contrast, when there are enough seats for all students, the two distributions are incomparable. Ashlagi et al. (2019) study another metric: the fraction of students receiving one of their first k choices. They show that when there is a shortage of seats, this fraction tends to zero as the market grows when using independent lotteries, whereas it remains constant under a single lottery. Collectively, these papers tell a consistent story: using a single lottery is preferable when there is a shortage of seats, but the comparison is more ambiguous when there are enough seats for all students.

As mentioned in the introduction, all of these papers assume ex-ante identical schools, and that students submit complete lists. We relax both assumptions. We adopt the preference formation model of Immorlica and Mahdian (2005) and Kojima and Pathak (2009), which allows schools to differ in their popularity. Furthermore, we allow the length of students' lists to follow an arbitrary distribution. The generality of our model allows us to study new questions: to our knowledge, Theorem 3 provides the first analytical comparison of the number of matches under different priority rules, while Proposition 2 offers the first analysis of how the number of students receiving their first choice depends on the relative popularity of different schools.

We note that Che and Tercieux (2019) also consider a large market model in which schools differ in their popularity. Their model for generating student lists sums a school quality term and an idiosyncratic term, which is more general than our approach of sampling schools without replacement from a fixed distribution. However, they assume that students submit complete lists, implying that the number of matches does not depend school priorities. Furthermore, their assumption that idiosyncratic terms are drawn iid from a bounded distribution implies that any assignment that matches each student to one of her top k choices is asymptotically optimal so long as k is $o(n)$: in their model, there is no meaningful difference between matching a student to her first choice and her fifth. By contrast, our work treats these outcomes very differently, and characterizes the exact rank distribution in the final assignment.

2.4. **Large Markets with a Finite Number of Schools.** Azevedo and Leshno (2016) take a different type of large market limit, in which the number of schools is fixed and the capacity of each school grows with the number of students. An advantage of their approach is that in the limit, outcomes at each school become deterministic, enabling tractable analysis for a richer class of preferences and priorities. For example, Shi (2021) uses a variant of this model in which students have cardinal utilities to optimize aggregate welfare over a large class of priority rules.

The advantages of this approach also come with disadvantages. Assuming that each school is large eliminates certain frictions, causing the model to make overly optimistic predictions. For example, when schools are homogeneous and there are more seats than students, this model predicts that it will be possible to assign every student to their first choice school, rendering priorities irrelevant. By contrast, in our model, the number of seats at each school is fixed and the number of students who list each school first is stochastic. As a result, even when schools are equally popular, not all students will be able to match to their first choice school, and priorities play a role in determining final outcomes.

3. Model

For expositional purposes, we begin by introducing a model with ex-ante homogeneous schools. Section 3.3 extends the model to allow schools to differ in their popularity and capacity, and all of our results hold for this more general model.

3.1. **Finite Market with Homogeneous Schools.** There are n schools and ρn students, for some $\rho \in \mathbb{R}_+$. There are $C \in \mathbb{N}$ seats at each school. Each student submits a ranked list of schools, which does not include every school. Each student also has a priority score at each school, and schools prioritize students with higher scores. Given student lists and priorities, the final allocation is determined using the student-proposing deferred acceptance algorithm. This algorithm starts with each student "proposing" to the first school on her list. From there, the following steps are repeated until convergence:

i. Schools consider students that have proposed to them, rejecting any for which the number of higher-priority interested students is at least the school's capacity.
ii. Students for whom all proposals have been rejected propose to the next school on their list (if any such school exists).

We assume that lists and priority scores are drawn iid across students. To generate student lists, each student samples a list length l from a distribution $\mathcal{L}$ on $\mathbb{N}$, and lists l uniformly random schools in a uniformly random order. For $l \in \mathbb{N}$ we let $\mathcal{L}(l)$ denote the probability of listing l schools, and define $\mathcal{L}(\geq l) = \sum_{j\geq l} \mathcal{L}(j)$ and $\mathcal{L}(>l) = \sum_{j>l} \mathcal{L}(j)$ to be the probability of listing at least l schools and more than l schools, respectively.

We consider two ways to generate priority scores. In one, schools use *independent lotteries*: students' priorities at each school are drawn independently and uniformly on $[0, 1]$. In the other, schools use a *single lottery*: each student receives a score drawn uniformly on $[0, 1]$, which is their priority at every school.

To summarize, our model with homogeneous schools has five parameters: the number of schools n, the ratio of students to schools ρ, the capacity of each school C, the list length distribution $\mathcal{L}$, and the priority rule $R \in \{I, S\}$ ($R = I$ for independent lotteries, $R = S$ for a single lottery).

Given these parameters, our goal is to understand the distribution of outcomes for students. For any fixed $\rho, C, \mathcal{L}$ and any $k, l \in \mathbb{N}$ with $k \leq l$, let $G_I^n(k, l)$ denote a random variable equal to the number of students who list l schools and match one of their top k choices when n schools use independent lotteries. Let $G_S^n(k, l)$ be the analogous quantity when schools use a single lottery.

3.2. **Large Market Limit.** Analyzing the distribution of G_R^n is challenging. Although student lists are independent, student interest is not. Whether a student proposes to her third choice depends on whether she gets into her first two choices, which in turn depends on how many other students propose to these schools. In this section, we provide a tractable approximation for the statistics G_R^n. We first offer heuristic motivation for our approximation, and then establish the accuracy of this approximation as the number of schools n grows large.

The intuition underlying our approximation is that when students list only a small number of schools, there is little difference between sampling schools with and without replacement. Hence, each time a student proposes, there is approximately a $1/n$ chance that their proposal goes to any fixed school. For this reason, we expect that if a total of λn proposals have been sent by students, the number of proposals received by each school should be approximately Binomial with parameters λn and $1/n$, which is in turn well-approximated by the Poisson distribution with mean λ. We will define values λ_I and λ_S, which are intended to approximate the expected number of proposals received by each school under independent and single lotteries, respectively.

To define these values, we make use of two functions derived from the Poisson distribution. If the number of proposals received by a school with capacity C follows a Poisson distribution with mean λ, then the probability that this school has at least one $\mathcal{V}$acancy is

$$\mathcal{V}(\lambda) = \sum_{j=0}^{C-1} \frac{e^{-\lambda}\lambda^j}{j!}, \tag{1}$$

Meanwhile, if this school ranks students according to a uniform lottery, then a student who proposes to this school will be $\mathcal{A}$ccepted with probability

$$\mathcal{A}(\lambda) = \sum_{j=0}^{\infty} \frac{e^{-\lambda}\lambda^j}{j!} \min\left(\frac{C}{j+1}, 1\right). \tag{2}$$

This is because the probability that j other students apply to the school is $\frac{e^{-\lambda}\lambda^j}{j!}$. When j other students apply, the probability of acceptance is 1 if $j < C$, and $C/(j + 1)$ otherwise (since the school selects students at random).

We now define one more important function. For a student who has list length drawn from $\mathcal{L}$ and is accepted independently with probability q at each listed school, the expected number of schools in which she is interested is given by

$$\mu(q) = \sum_{k=0}^{\infty} \mathcal{L}(>k)(1-q)^k. \tag{3}$$

This uses the fact that for a non-negative integer-valued random variable N, $\mathbb{E}[N] = \sum_{k=0}^{\infty} \mathbb{P}(N > k)$, alongside the observation that a student proposes to more than k schools if and only if she lists more than k schools and is rejected from the first k schools on her list. Note that $\mu(1)$ is the fraction of students listing at least one school, which we assume without loss of generality to be one, while $\mu(0) = \sum_{k\geq 0} \mathcal{L}(>k)$ is the average list length.

We are now prepared to define the scalar λ_I, which represents the expected number of proposals received by each school under independent lotteries. This should equal the expected number of proposals made by each student, times the ratio of students to schools ρ. Given λ_I, students should find that each proposal is accepted with probability $\mathcal{A}(\lambda_I)$, as explained above, and thus should send an average of $\mu(\mathcal{A}(\lambda_I))$ proposals. Therefore, λ_I should satisfy the following consistency condition:

$$\lambda_I = \rho \cdot \mu(\mathcal{A}(\lambda_I)). \tag{4}$$

Theorem 1 establishes that for any $\rho, C, \mathcal{L}$, there is a unique value $\lambda_I \in \mathbb{R}_+$ satisfying (4). Given λ_I, the probability that a student who lists k schools is accepted by at least one of them is

$$F_I(k) = 1 - (1 - \mathcal{A}(\lambda_I))^k. \tag{5}$$

Next, we turn to our approximation for outcomes under a single lottery. Here, rather than calculating a scalar λ_S, we will define a function $\Lambda_S : [0,1] \to \mathbb{R}_+$, where $\Lambda_S(t)$ represents the average number of proposals that each school receives from students with priority above t, when priorities are generated by a single lottery. Fixing $\rho, C, \mathcal{L}$, we define Λ_S to be the solution to the following differential equation:

$$\Lambda_S(1) = 0, \qquad \Lambda_S'(t) = -\rho \cdot \mu(\mathcal{V}(\Lambda_S(t))). \tag{6}$$

The motivation is that a market where schools have identical priorities can be analyzed in a top-down manner. A student of priority t is accepted to a school only if the number of higher-priority students who have proposed to that school is below its capacity C. If the number of such students follows a Poisson distribution with mean $\Lambda_S(t)$, then $\mathcal{V}(\Lambda_S(t))$ gives the acceptance probability for a student of priority t, and therefore $\mu(\mathcal{V}(\Lambda_S(t)))$ gives the expected number of schools to which this student will propose. The expected number of students with priority in an interval of width dt is $\rho n \cdot dt$, so the expected number of proposals from these students received by the average school is $\rho \cdot \mu(\mathcal{V}(\Lambda_S(t)))dt$.

Given Λ_S, we can calculate the fraction of students with list length k who are matched to a school on their list. Each time a student with priority t proposes to the next school on her list, her probability of being accepted by this school is $\mathcal{V}(\Lambda_S(t))$. It follows that she is rejected from her first k choices with probability $(1-\mathcal{V}(\Lambda_S(t)))^k$. Integrating across priorities, we define

$$F_S(k) = 1 - \int_0^1 (1-\mathcal{V}(\Lambda_S(t)))^k \, dt. \tag{7}$$

Our first result establishes that the heuristic reasoning above can be made rigorous, and that as the market grows large, the values $F_I(k)$ and $F_S(k)$ give the fraction of students who get a top k choice, among those who list at least k schools.

Theorem 1. *Fix $\rho \in \mathbb{R}_+, C \in \mathbb{N}$, and $\mathcal{L}$ satisfying $\mathcal{L}(>\ell) = 0$ for some $\ell \in \mathbb{N}$. There is a unique function F_I defined by (4) and (5), and a unique function F_S defined by (6) and (7). For any $1 \le k \le l \le \ell$ and $\epsilon > 0$, as $n \to \infty$,*

$$\mathbb{P}\left(\left|\frac{G_I^n(k,l)}{\rho n} - \mathcal{L}(l)F_I(k)\right| > \epsilon\right) \to 0, \qquad \mathbb{P}\left(\left|\frac{G_S^n(k,l)}{\rho n} - \mathcal{L}(l)F_S(k)\right| > \epsilon\right) \to 0.$$

This says that in large markets, the fraction of students who list exactly l schools and match to a top k choice is close to $\mathcal{L}(l)F_R(k)$ with high probability.

The proof of this result defines discrete time Markov Chains that correspond to the execution of the Deferred Acceptance algorithm on randomly drawn preferences. These chains deploy the "principle of deferred decisions" (Pittel, 1992; Immorlica and Mahdian, 2005), meaning that they reveal the next school on a student's list only when it is time for that student to propose. They also make use of the implementation of Deferred Acceptance described by McVitie and Wilson (1971), in which students are invited to propose sequentially, rather than in parallel. Each round of the algorithm captures the effect of adding a single student to the market, tracking a series of rejections until a student either fills a vacancy or reaches the end of her list.

Past work has used a similar approach, but has been limited by the challenge of tracking the evolution of a very high-dimensional Markov chain. There are two primary ways to overcome this challenge: reduce the dimensionality of the state space by imposing restrictive assumptions (such as homogeneity of students and schools), or conduct a coarse analysis that limits the types of conclusions that can be drawn.[5]

In order to get precise results in a general model, we deploy the differential equation method of Wormald (1999). The first step of this method is to show that the evolution of the Markov Chain can be well-approximated using only a limited number of summary statistics. From this, it follows that the values of these summary statistics remain close to the solution of a corresponding set of differential equations. The key to making this approach work was coming up with a set of summary statistics that is rich enough to capture the first-order behavior of the Markov Chain, but simple enough to result in a tractable set of differential equations. Our proof involves a high-dimensional set of differential equations, which we solve by using the intuition outlined above and in remarks provided in the Appendix to guess a solution (which we then verify).

The final product is a very precise description of outcomes in large markets. Our main results study and compare the values F_I and F_S. Before getting to these results, we first show how to extend our model to handle heterogeneous schools.

[5] Ashlagi et al. (2017) use the former approach, Immorlica and Mahdian (2005) and Kojima and Pathak (2009) use the latter, and Ashlagi et al. (2019) and Ashlagi and Nikzad (2020) deploy a combination of the two.

3.3. **Adding School Heterogeneity.** We now extend the model to allow schools which differ in their popularity and capacity. We adopt the preference formation model of Immorlica and Mahdian (2005) and Kojima and Pathak (2009): student list lengths are drawn iid from $\mathcal{L}$, and then lists are completed by sampling schools without replacement from a fixed distribution over the set of schools. We say that a school's "popularity" is n times its probability of being selected first. Therefore, the average school popularity is one by definition.

In this model, the relevant information about a school is captured by two numbers: its capacity and its popularity. We refer to this pair as a school's "type." We consider a limiting regime in which there is a finite set of school types $\mathcal{T} \subseteq \mathbb{N} \times \mathbb{R}_+$, and a large number of schools of each type. We let $\mathcal{D}$ be the empirical distribution of school types. In other words, each school has a type $\tau = (C_\tau, p_\tau) \in \mathcal{T}$ (where C_τ represents the capacity of the school and p_τ represents its popularity), and $\mathcal{D}_\tau \in [0, 1]$ represents the fraction of schools with type τ. We place no restriction on the distribution $\mathcal{D}$ other than that it has bounded support. Because the average school popularity has been normalized to one, the probability that the first school on a student's list is of type τ is $p_\tau \mathcal{D}_\tau$. This motivates us to define the following generalizations of $\mathcal{V}(\cdot)$ and $\mathcal{A}(\cdot)$:

$$\mathcal{V}(\lambda) = \sum_{\tau \in \mathcal{T}} p_\tau \mathcal{D}_\tau \nu(p_\tau \lambda, C_\tau), \tag{8}$$

$$\mathcal{A}(\lambda) = \sum_{\tau \in \mathcal{T}} p_\tau \mathcal{D}_\tau \alpha(p_\tau \lambda, C_\tau) \tag{9}$$

where ν and α are our new names for the expressions in (1) and (2):

$$\nu(\lambda, C) = \sum_{k=0}^{C-1} \frac{e^{-\lambda}\lambda^k}{k!} = \mathbb{P}(\mathbf{Po}(\lambda) < C), \tag{10}$$

$$\alpha(\lambda, C) = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} \min\left(\frac{C}{k+1}, 1\right). \tag{11}$$

The model with homogeneous schools corresponds to a singleton type space $\mathcal{T} = \{(C, 1)\}$, in which case the expressions in (8) and (9) coincide with (1) and (2). The intuition for these more general expressions is that if the average school receives λ proposals, and a school of type τ is p_τ times more popular than average, then the number of proposals received by a school of type τ should be Poisson with mean $p_\tau \lambda$, and therefore the probability of a vacancy at that school should be $\nu(p_\tau \lambda, C_\tau)$, with a corresponding acceptance rate of $\alpha(p_\tau \lambda, C_\tau)$. The expressions in (8) and (9) calculate a weighted average of these probabilities, accounting for the fact that $p_\tau \mathcal{D}_\tau$ gives the probability of proposing to a type-τ school. We continue to define F_I and F_S by (4), (5), (6), (7).

Small modifications to the proof establish that Theorem 1 continues to hold in markets with an arbitrary fixed type space $\mathcal{T}$ and many schools of each type. Incorporating an arbitrary distribution of school types $\mathcal{D}$ significantly generalizes the model, permitting it to capture setting where some schools are listed much more frequently than others, some are larger than others, and the popularity of large schools differs systematically from the popularity of small schools. All of our main results hold in this general model, and Proposition 3 exploits this generality to study how the number of students getting their first choice depends on the distribution of school popularity.

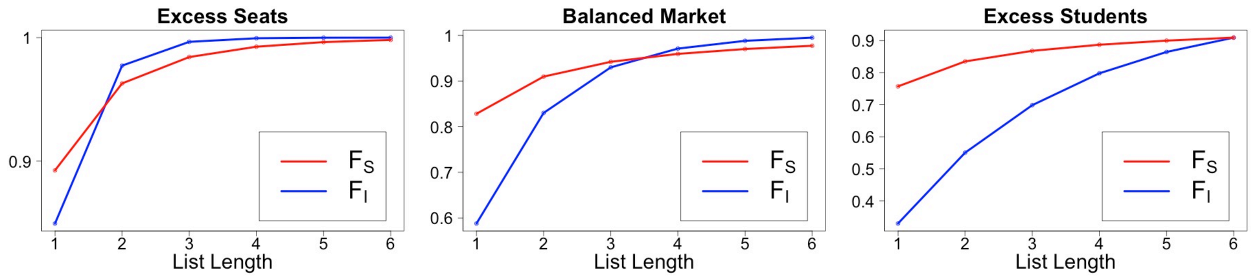

FIGURE 1. The probability of matching, as a function of list length, under single and independent lotteries. Each panel displays results for a different market. In each market, schools are equally popular and have $C = 10$ seats, and all students list 6 schools. The ratio of students to seats varies across markets: on the left, there are 10% fewer students than seats, on the right there are 10% more students than seats, and the middle panel shows a balanced market. The location of the crossing point established by Theorem 2 depends on the market imbalance. When there is a surplus of seats (left panel), students are more likely to get one of their top two choices under independent lotteries, while with a shortage of seats (right panel), a single lottery appears to dominate independent lotteries.

4. Insights from Our Model

We will study properties of the quantities $F_R(k)$. Theorem 1 establishes that $F_R(k)$ is equal to the fraction of students who get one of their top k choices, among those who list at least k schools. Alternatively, $F_R(k)$ can be thought of as the probability of matching for students who list exactly k schools.

4.1. Identifying a Tradeoff.

We begin with our first main result, which establishes a tradeoff when choosing between a single lottery and independent lotteries.

Theorem 2. *Given any $\rho, \mathcal{L}, \mathcal{D}$, the functions F_I and F_S cross exactly once: there exists $\ell \in \mathbb{N}$ with $\mathcal{L}(> \ell) > 0$ such that $F_I(k) < F_S(k)$ for $k \leq \ell$, and $F_I(k) \geq F_S(k)$ for $k > \ell$.*

Recall from Theorem 1 that $F_R(k)$ can be interpreted as the probability of matching for students who submit lists of length k. Therefore, if we call a list of length l *short* if $l \leq \ell$, and *long* if $l > \ell$, then Theorem 2 states that students with long lists are more likely to match when using independent lotteries, while those with short lists are more likely to match when using a single lottery. This establishes a tradeoff between these procedures. Theorem 2 additionally implies that a single lottery always matches more students to their first choice school, and that for students submitting short lists, the rank distribution under a single lottery stochastically dominates the rank distribution under independent lotteries.

4.1.1. *The Location of the Crossing Point.* Although Theorem 2 establishes that F_S and F_I cross at a single position, it is silent about where the crossing point falls, as well as the magnitude of the difference between F_S and F_I. This will depend on the market imbalance ρ, the distribution of student list length $\mathcal{L}$, and the size and popularity of various schools, as captured by $\mathcal{D}$.

Figure 1 shows how the crossing point depends on the market imbalance, in markets where all students list 6 schools. When there are 10% fewer students than seats, independent lotteries look

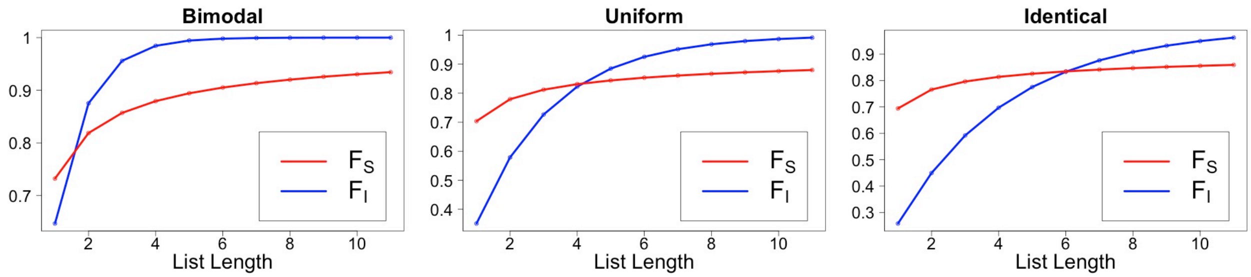

FIGURE 2. The probability of matching, as a function of list length, under single and independent lotteries. Each panel displays results for a different market. In each market, schools are equally popular and have $C = 10$ seats, there are 20% more students than seats, and students list an average of 6 schools. The list length distribution varies across markets: on the left, half of students list 1 school and the other half list 11; in the middle, list lengths are uniformly distributed on $\{1, 2, \ldots, 10, 11\}$; and on the right, all students list 6 schools. In the latter case, a single lottery seems the clearly superior choice (because all students list six schools, the part of the graph to the right of six is irrelevant). When list lengths vary, students with shorter lists favor a single lottery, while those with longer lists are more likely to match under independent lotteries.

somewhat attractive, as they result in more students getting one of their top two choices. When there are 10% more students than seats, a single lottery "almost dominates" independent lotteries: $F_S(k) > F_I(k) - \epsilon$ for all k and some small ϵ. These findings are consistent with the conclusions of Ashlagi and Nikzad (2020), who write "In a market with a surplus of seats... there are efficiency tradeoffs between the two tie-breaking rules. However, a common lottery is always preferable when there is a shortage of seats."

Figure 2 shows that this conclusion may no longer hold when student lists vary in length. It displays results from markets where each school has 10 seats and there are 20% more students than seats. Despite this imbalance, when the list length distribution is bimodal (with some very long lists and other very short ones), independent lotteries match more students with long lists, while a single lottery matches more students with short lists. Thus, even in markets with a shortage of seats, a meaningful tradeoff emerges. There may be cases where it is most important to match students with long lists, in which case independent lotteries might be preferable. We elaborate on this point in Section 5.

4.1.2. *Multiple Priority Classes.* Our model assumes that priorities are determined purely by lottery. In practice, students are often given priority if they live in the neighborhood, or have a sibling who attends the school. Lotteries are then used to break ties within each priority class.

Although Theorem 2 does not directly apply to these cases, the intuition underlying it is fairly robust. Independent lotteries give a relative advantage to students who submit longer lists, who benefit from having the most independent draws. Meanwhile, using a single lottery increases the number of students who receive their first choice, by minimizing the chance that a student who has been rejected from one school will have high priority at another.

We illustrate this point using simulation results of more complex priority structures. Figure 3 displays simulation results when there are no inherent priorities alongside corresponding results when some students have priority at their first choice school, or at a random school, and lotteries

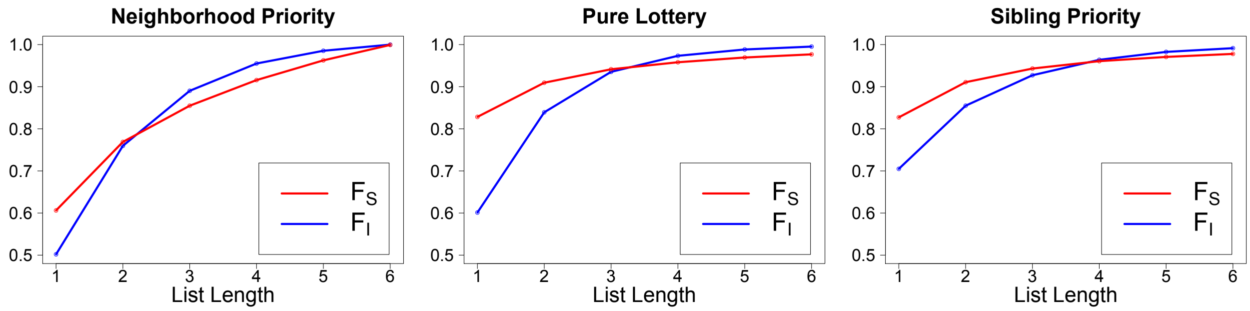

FIGURE 3. Although our model assumes that priorities are determined solely by lottery, simulation results show that our qualitative findings continue to hold when there are multiple priority classes and lotteries are used only as a tiebreaker. Each panel shows simulations from markets where there are 100 schools with 10 seats each, and 1000 students, each of whom list 6 schools. In the center, priorities are purely by lottery: these results match the predictions in the center panel of Figure 1. On the left, each student has "neighborhood priority" at one of the six schools on her list. This significantly reduces the number of students matching to their first choice (because of the difficulty of matching to a non-neighborhood school), but increases the number of matches under a single lottery (because even students with bad lottery numbers have high priority at their neighborhood school). On the right, 40% of students have "sibling priority" at the first school on their list. This notably increases the number of students who get their first choice under independent lotteries. Although introducing underlying priorities affects aggregate outcomes, the insights from Theorem 2 continue to apply: a single lottery gives more students their first choice, and F_I and F_S cross exactly once.

are used only to break ties in priority. Adding priority classes affects the distribution of outcomes under both lottery procedures, and reduces the importance of the lottery. However, the qualitative conclusions from Theorem 2 continue to apply in all three scenarios: a single lottery gives more students their first choice, independent lotteries result in more matches for students submitting long lists, and the values F_I and F_S cross exactly once.

4.2. Comparing the Number of Assigned Students.

Theorem 2 established a tradeoff: students with short lists are more likely to match when using a single lottery, while those with the longest lists are more likely to match when using independent lotteries. Which procedure matches more students overall?

The answer to this question would seem to depend on the crossing point (which determines what counts as a "short" or a "long" list), the number of students submitting lists of each length, and the magnitude of the differences between F_I and F_S for each list length k. Each of these, in turn, depend on the ratio of students to schools ρ, the distribution $\mathcal{D}$ of schools' capacity and popularity, and the distribution $\mathcal{L}$ of students' list lengths, as Figures 1 and 2 illustrate. A priori, there is no reason to expect a simple characterization.

In one special case, however, Theorem 2 provides the answer: if all students list the same number of schools, then all students "submit the longest lists", so independent lotteries assign more students. In this case, knowing the list length distribution $\mathcal{L}$ suffices to reach a conclusion that holds for *any* ρ and $\mathcal{D}$. Theorem 3 generalizes this observation by establishing conditions on $\mathcal{L}$ which ensure that one procedure matches more students than the other, regardless of the parameters ρ and $\mathcal{D}$. We define

$$F_R(\mathcal{L}) = \mathop{\mathbb{E}}_{\ell\sim\mathcal{L}}[F_R(\ell)] \tag{12}$$

to be the fraction of students who match under lottery rule R.

Theorem 3. *Fix any $\rho, \mathcal{D}$. If the list length distribution $\mathcal{L}$ is such that $f(q) = 1/\mu(q)$ is*

- *concave, then more students are assigned under a single lottery: $F_I(\mathcal{L}) \le F_S(\mathcal{L})$.*
- *convex, then more students are assigned under independent lotteries: $F_I(\mathcal{L}) \ge F_S(\mathcal{L})$.*

This establishes that either lottery procedure may result in more matches than the other, and that the comparison depends on whether the function f is concave or convex. A natural question is which case, if either, should be considered "typical"? To answer this question, we examine list length data from New York City, Amsterdam, Hungary, and Chile in Figure 4. In all four cases, the function f is convex. Therefore, Theorem 3 suggests that using a single lottery to break ties should result in fewer assigned students. This prediction is consistent with the simulation results presented by Abdulkadiroglu et al. (2009) and de Haan et al. (2018).

4.2.1. *A simple sufficient condition.* Although Theorem 3 provides a clean characterization, it is also opaque. What is the meaning of the function f? One interpretation is as follows. Suppose that there is a population of students with list lengths drawn from $\mathcal{L}$, who are accepted to each school on their list independently with probability q. Then the fraction of students receiving their first choice is q, while the fraction who match can be shown to be $q \cdot \mu(q)$. Therefore, the fraction of assigned students who attend their first choice school is $f(q)$. Even with this interpretation, however, it is difficult to think about whether a particular list length distribution will cause f to be convex or concave. Proposition 1 simplifies the process by relating the convexity or concavity of f to the hazard rate of the list length distribution.

Proposition 1. *Define the **hazard rate** of the list length distribution at k to be $\frac{\mathcal{L}(k)}{\mathcal{L}(\ge k)}$. If the hazard rate of the list length distribution weakly increases in k, then $f(q) = 1/\mu(q)$ is convex. If the hazard rate of the list length distribution weakly decreases in k, then $f(q) = 1/\mu(q)$ is concave.*

Many common distributions have an increasing hazard rate, including deterministic distributions, uniform distributions on an interval, binomial distributions, and truncated Poisson distributions. An example of distribution with a decreasing hazard rate is the Power law distribution $\mathcal{L}(\ge k) = k^{-d}$ for some $d \ge 2$. The geometric distribution has a constant hazard rate. Therefore, Theorem 3 and Proposition 1 jointly imply that if the list length distribution is geometric, the two lottery procedures match the same number of students.

It is important to note that a distribution with a decreasing hazard rate is necessarily unbounded, and therefore cannot arise in practice. There do exist distributions with bounded support for which f is concave, but Proposition 1 along with the evidence in Figure 4 suggests that the convex case is much more common. Therefore, Theorem 3 suggests that we should typically expect independent lotteries to produce more matches.

4.2.2. *How big is the difference?* Theorem 3 does not address the magnitude of the difference between the number of matches with single and independent lotteries. In some cases, these procedures

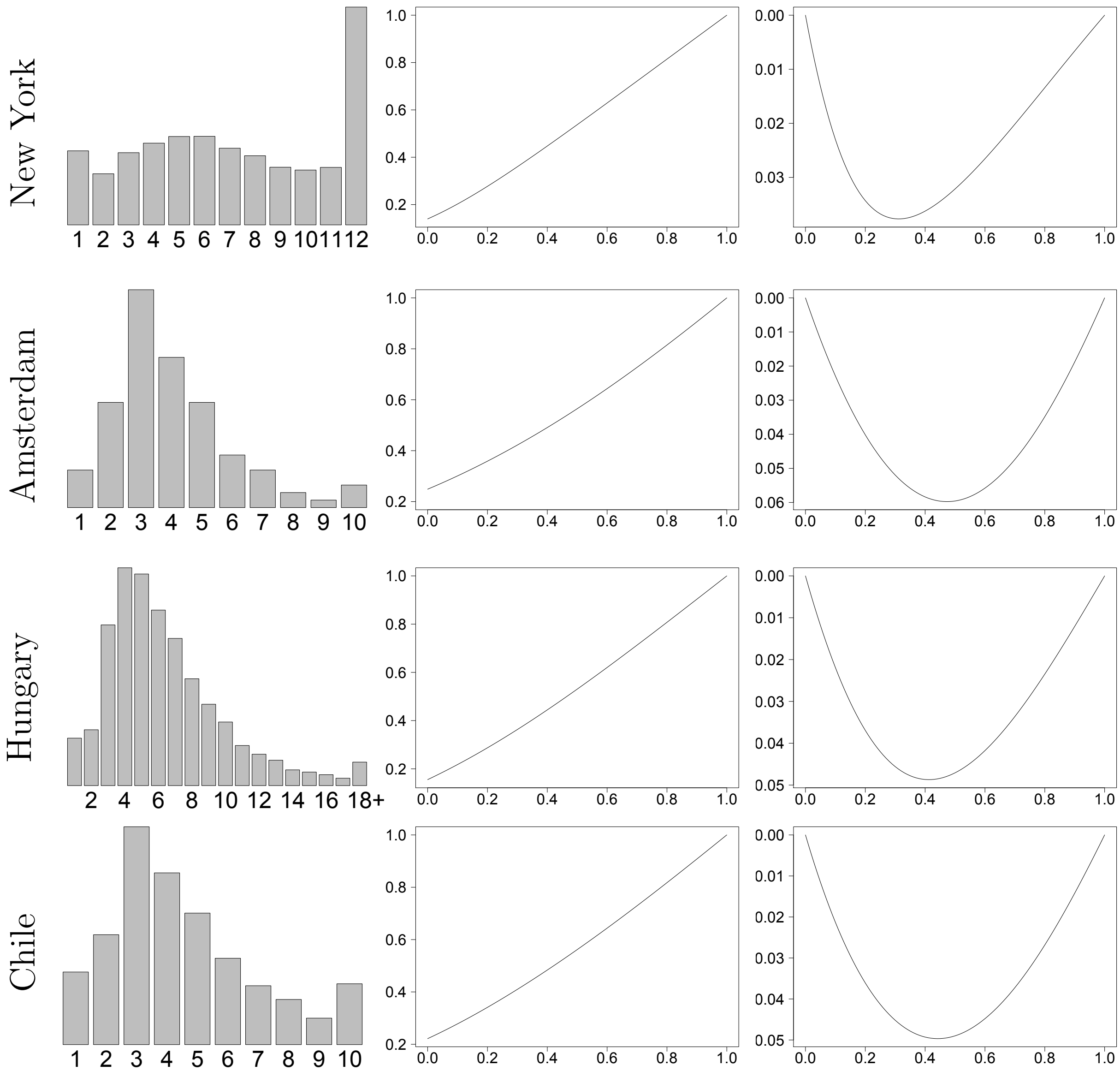

FIGURE 4. Left panel: list length distributions from four centralized admissions programs. Middle panel: the corresponding functions $f(q)$. Despite differences in the average list length and the shape of the distributions, in all cases, f is convex, as seen by removing the linear trend and plotting $f(q) - qf(1) - (1-q)f(0)$ (right panel). Theorem 3 suggests that when f is convex, breaking ties using a single lottery should match fewer students. This prediction is consistent with the empirical findings from New York (Abdulkadiroglu et al. (2009)) and Amsterdam (de Haan et al. (2018)). Hungarian list length data taken from Aue et al. (2019), and Chilean data from Larroucau and Rios (2019).

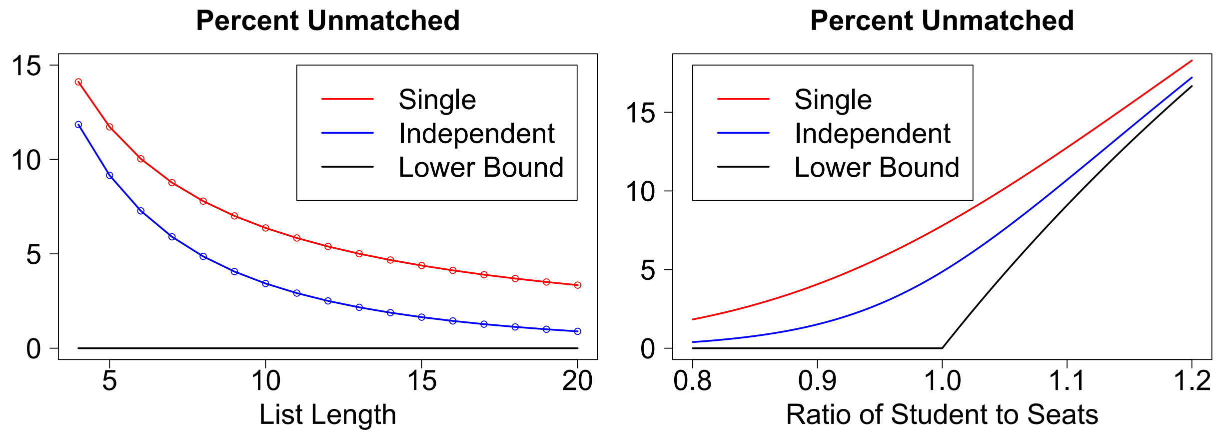

FIGURE 5. Comparison of unassigned students when using single and independent lotteries to a lower bound based on the fact that the number of assigned students is at most the total number of seats. When students list the same number of schools, Theorems 2 and 3 establish that a single lottery results in more unassigned students. This figure shows that the difference can be significant. In both panels, schools are equally popular and have capacity $C = 1$. On the left, the market is balanced ($\rho = 1$) and the list length ℓ varies. When $\ell = 10$, moving from a single lottery to independent lotteries reduces the fraction of unassigned students from 6.4% to 3.4%. In order to reduce the number of unassigned students below 3.4% while using a single lottery, lists must be lengthened to $\ell = 20$ schools. On the right, students list eight schools, and the ratio of students to seats varies. When the market is balanced, moving from a single lottery to independent lotteries causes a larger reduction in unassigned students than increasing the number of schools by 7%.

leave similar numbers of students unassigned. For example, if the market is very imbalanced or lists are very long, the number of matches will be close to the size of the short side of the market. However, in other cases, independent lotteries leave significantly fewer students unassigned. Figure 5 shows results for markets with homogeneous schools, each with a single seat. When the market is balanced and each student lists 10 schools, 6.4% of students go unassigned when using a single lottery, compared to 3.4% with independent lotteries. In order to reduce the number of unassigned students below 3.4% while using a single lottery, lists must be lengthened to $\ell = 20$ schools!

This motivates us to study the rate at which the number of unmatched students decreases as lists get longer. To obtain simple expressions, we consider the special case where $C = 1$, schools are equally popular, and student list lengths follow a Poisson distribution with mean ℓ.[6]

[6] Qualitatively similar results hold for other parameter choices, but the expressions are not as clean.

Proposition 2. *Suppose that schools are equally popular and have a single seat, and the list length distribution is Poisson with mean ℓ. If $\rho = 1$, then*

$$1 - F_I(\mathcal{L}) = e^{-\sqrt{\ell F_I(\mathcal{L})}}, \qquad 1 - F_S(\mathcal{L}) = \log(2 - e^{-\ell})/\ell.$$

If $\rho < 1$, then as $\ell \to \infty$,

$$\frac{1}{\ell}\log(1 - F_I(\mathcal{L})) \to \frac{\rho}{\log(1-\rho)}, \qquad \frac{1}{\ell}\log(1 - F_S(\mathcal{L})) \to -(1-\rho).$$

$$\frac{1}{\ell}\log(1 - F_I(\mathcal{L})) = \frac{\rho F_I(\mathcal{L})}{\log(1-\rho F_I(\mathcal{L}))} \to \frac{\rho}{\log(1-\rho)}. \tag{13}$$

$$\frac{1}{\ell}\log(1 - F_S(\mathcal{L})) = \frac{1}{\ell}\log\left(1 - \frac{1}{\rho} + \frac{1}{\rho\ell}\log(1 + e^{\ell(1-\rho)} - e^{-\ell\rho})\right) \to -(1-\rho). \tag{14}$$

Proposition 2 says that in a balanced one-to-one market, the fraction of unassigned students decreases at very different rates under the two procedures: it is approximately $e^{-\sqrt{\ell}}$ with independent lotteries, and approximately $\log(2)/\ell$ under a single lottery. When there are fewer students than seats ($\rho < 1$), then the number of unassigned students decays exponentially in ℓ under both procedures, because each additional proposal has a constant probability of being sent to a school with a vacancy. However, the rate of exponential decay is very different: $1 - F_I(\mathcal{L}) \approx e^{\frac{\rho}{\log(1-\rho)}\ell}$, while $1 - F_S(\mathcal{L}) \approx e^{-(1-\rho)\ell}$. If $\rho = 0.9$, this means that for large ℓ, $1 - F_I(\mathcal{L}) \approx e^{-0.4\ell}$, while $1 - F_S(\mathcal{L}) \approx e^{-0.1\ell}$.

4.3. **Number of First Choices.** Theorem 2 implies that more students are assigned to their first choice when using a single lottery, but how big can the difference be? If the number of students (parameterized by ρ) is very large, then the fraction of students who match (and therefore the fraction who get their first choice) is necessarily small under any allocation procedure. Accordingly, it makes sense to study the fraction of *assigned* students who receive their first choice.

Ashlagi et al. (2019) show that the difference between single and independent lotteries can be extreme: when schools have a single seat and are equally popular, the fraction of assigned students who receive their first choice can be arbitrarily small under independent lotteries, but is always at least 1/2 when using a single lottery.

Our next result generalizes the lower bound for independent lotteries to cases where schools have multiple seats and differ in their popularity. Unsurprisingly, our bound is largest when schools are equally popular, and close to zero if schools differ dramatically in their popularity. Proposition 3 quantifies the decay in the number of students receiving their first choice as discrepancies in popularity become more pronounced.

Proposition 3. *If the type distribution $\mathcal{D}$ is such that each school has C seats and the most popular school is at most r times as popular as the least popular school, then*

$$\frac{F_S(1)}{F_S(\mathcal{L})} \geq LBT(r, C) = \frac{2}{(1+\sqrt{r})^2}\int_0^\infty \sqrt{r}\cdot\nu(Cx, C)^2 + \nu(Cx, C)\nu(Cx/r, C)dx.$$

The proof in Appendix A.4 establishes that the lower bound of $LBT(r, C)$ is tight (it is attained when students submit long lists, there are as many students as seats, and schools either have

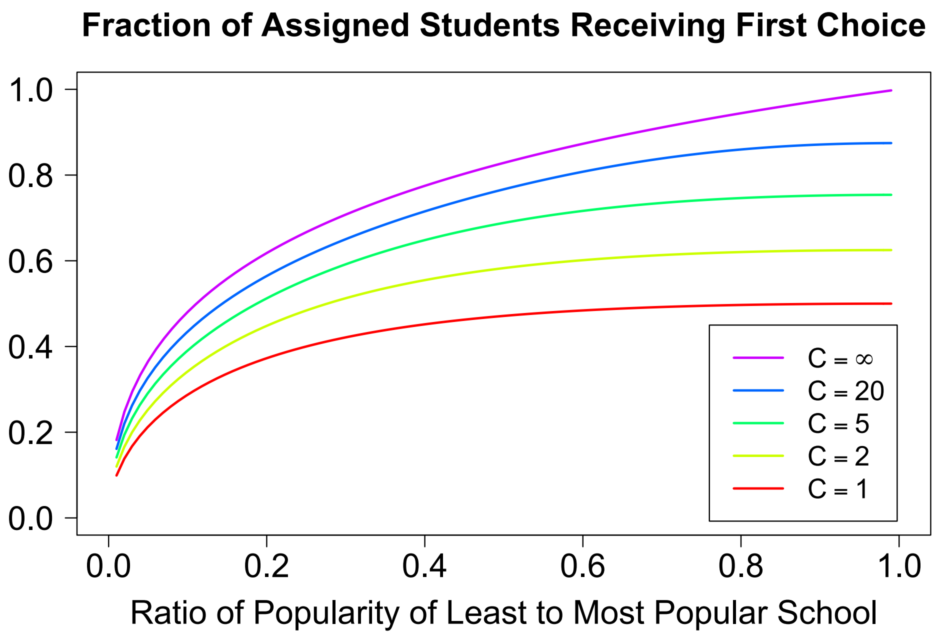

FIGURE 6. Using a single lottery ensures that a sizeable fraction of assigned students receive their first choice school. The figure shows tight lower bounds on this quantity when schools have identical capacity C. The x axis represents the ratio in popularity of the least to most popular school, which is the reciprocal of the parameter r mentioned in Proposition 3. On the far right, schools are equally popular, and at least half of assigned students receive their first choice. This guarantee increases to one as C grows. Moving to the left, the discrepancy in popularity r increases, and the number of students receiving their first choice falls. Proposition 3 states that when $C = 1$ the fraction of assigned students who receive their first choice is at least $\sqrt{r}/(1+r)$. This bound approaches $2/(1+\sqrt{r})$ as C increases.

popularity $\sqrt{r}$ or $1/\sqrt{r}$). While this bound appears complex, it is decreasing in r and increasing in C, and can be readily evaluated for each C. In particular,

$$LBT(r,1) = \frac{\sqrt{r}}{1+r}.$$

$$LBT(r,2) = \frac{2}{(1+\sqrt{r})^2}\left(1 + \frac{5}{8}\sqrt{r} - \frac{2r+1}{(r+1)^3}\right).$$

$$\lim_{C\to\infty} LBT(r,C) = \frac{2}{1+\sqrt{r}}.$$

The values $LBT(r,C)$ are displayed in Figure 6. They imply that if no school is more than four times more popular than any other, then at least 40% of assigned students get their top choice. If no school is more than nine times more popular than any other, then at least 30% of assigned students get their top choice. As $C \to \infty$, at least 2/3 of assigned students receiving their first choice when $r = 4$ and 1/2 of assigned students receiving their first choice when $r = 9$.

5. Discussion

This paper addresses an operational question facing school districts that use the Deferred Acceptance Algorithm: how should lotteries to determine school priorities be implemented? Prior work has focused on metrics such as students' average rank, or the number of students receiving their top choice, and generally concluded that a single lottery is superior to independent lotteries. Our work is the first to compare the number of matches under each procedure. It identifies the importance of a parameter – the list length distribution – which has received little attention in prior work.

5.1. **Policy Implications.** Our findings provides new information for policymakers to consider. For distributions that arise in practice, our results suggest that using a single lottery will typically match fewer students. Therefore, if maximizing the number of assigned students is sufficiently important, independent lotteries may be preferable. In fact, the logic underlying our results suggests that *negatively* correlated lotteries should match even more students, as those who are rejected from their top choices will tend to have good lottery draws at schools further down their lists.

In practice, policymakers may care not only about aggregate statistics, but also about outcomes for certain subpopulations of students. Our work takes a first step towards understanding *which* students benefit from each lottery procedure. Even if using a single lottery improves aggregate statistics, Theorem 2 implies that this may hurt students who submit long lists.

To make this idea concrete, consider the example discussed in Section 4.1.1, where the list length distribution is bimodal: some students list only one school, while others submit long lists. In that case, a single lottery matches more students to their first choice and more students overall. Based on these statistics, it might seem like a clear winner. However, for the population of students who submit long lists, a single lottery leaves 6.6% unmatched, while independent lotteries match them all. One might imagine that students who list a single school tend to have strong outside options, such as attending an elite private school. In this case, adopting a single lottery could have the unintended consequence of exacerbating inequality, by transferring positions at top public schools away from underprivileged students who submit long lists, and towards students whose parents can afford private school.

This is merely one story that could be told. An alternative narrative might conclude that students submitting long lists are those whose parents have the time and resources to research many options. Ultimately, the question of which students submit short lists is an empirical one. However, our work is the first in its line to consider outcomes for specific subpopulations of students. It identifies not only how lottery design affects the total number of matches, but also how it affects *who* matches. This information could be valuable to policymakers tasked with implementing tie-breaking procedures.

Although we focus on lottery design, the expressions for match outcomes provided in Theorem 1 are a contribution which could offer many other insights. One question that merits further study is the effect of lengthening student lists. Policymakers often encourage students to list more schools, in order to reduce their chance of going unassigned. While this is good advice at an individual level, it could lead to worse societal outcomes, as lengthening one student's list increases the competition that other students face. Our model is well-suited to provide quantitative guidance on this tradeoff.

For example, Proposition 2 highlights that when using a single lottery, lengthening student lists may not be especially effective at reducing the number of unassigned students. With independent lotteries, lengthening lists is more effective at increasing matches, but is also more likely to impose externalities on other students. Future studies might use our results to study how the length of student lists affects the number of students getting one of their top choices.

5.2. **Model Limitations.** Despite the generality of our model, it is important to note that it still imposes a lot of structure on student preferences, and does not capture certain features expected in real-world markets. Our procedure for generating student preferences allows some schools to be much more popular than others, but also ensures that learning a subset of a student's list conveys almost no information about the remaining schools on the list. In reality, students are likely to list multiple schools that are in a similar location, or have similar academic offerings. Generalizing our results to richer models of student preferences is an important direction for future work. Similarly, it would be valuable to move beyond single and independent lotteries to richer priority structures that more closely reflect reality.

A second caveat is that our comparisons implicitly assume that changing priorities doesn't affect student lists. In theory, this seems a reasonable assumption: when using student-proposing Deferred Acceptance, it is a dominant strategy for students to submit their true preferences, regardless of how priorities are determined (Dubins and Freedman, 1981; Roth, 1984).

In practice, however, there are several reasons to believe that students might respond to changes in priority. One is that most school districts limit the number of schools that each student can rank (Pathak and Sönmez, 2013), and students who cannot list all acceptable schools have an incentive to strategically list those where they have high priority. Additionally, students who are confident of admission to a particular school might neglect to list inferior options, while students who consider a school unattainable might not list it, even if there is no cost to doing so (Chen and Pereyra, 2019; Hassidim et al., 2021). In all of these cases, changes to priorities could plausibly cause students to list additional schools, or drop schools from their lists.

Despite these concerns, we think that the assumption that lists do not change is a reasonable point to start analysis. Regarding list length constraints, most students do not submit the maximum number of allowed schools. For example, only 22% do so in New York, 27% in Denver, and 5% in New Orleans (Pathak, 2016). Additionally, even for students who submit the maximum number of schools, it is not clear that their reports respond significantly to the mechanism. As evidence, consider a policy change made by Chicago's exam schools in 2009. Before the change, these schools had been assigned using a variant of the Boston mechanism. Preferences were collected from over 13,000 families in 2009, before a decision was made to dramatically change the mechanism (to Deferred Acceptance with a list length limit) and the priorities (giving affirmative action based on socioeconomic status rather than race). Students were allowed to resubmit their preferences, but only 11% made any changes (Pathak, 2016). Pathak (2016) concludes, "The magnitude of the behavioral response is swamped, for instance, by the mechanical change in how applicants are processed by the mechanism." Meanwhile, Hastings et al. (2009) study data from Charlotte, North Carolina, and conclude that there is little evidence that applicants' preferences reacted to a change in neighborhood boundaries. Thus, while changes to the assignment mechanism no doubt affect some students' lists, the effect appears to be modest even for dramatic changes, and would likely be smaller for changes such as the implementation of the lottery.[7]

For those who insist on taking seriously the idea that students generate lists strategically, we note that any analysis of this behavior requires students to form beliefs about what options are available to them. This has generally been viewed as intractable unless the students have perfect information or view all schools symmetrically. The model presented in this paper generates tractable predictions of an asymmetric market where students have incomplete information about others' preferences, and therefore could be used as a building block for the analysis of strategic listing.

References

Abdulkadiroglu, A., Che, Y.-K., Pathak, P. A., Roth, A. E., and Tercieux, O. (2017). Minimizing justified envy in school choice: The design of new orleans' oneapp.

Abdulkadiroglu, A., Pathak, P. A., and Roth, A. E. (2005a). The new york city high school match. *American Economic Review, Papers and Proceedings*, 95(2):364–367.

Abdulkadiroglu, A., Pathak, P. A., and Roth, A. E. (2009). Strategy-proofness versus efficiency in matching with indifferences: Redesigning the new york city high school match. *American Economic Review*, 99(5):1954–1978.

Abdulkadiroglu, A., Pathak, P. A., Roth, A. E., and Sonmez, T. (2005b). The boston public school match. *American Economic Review, Papers and Proceedings*, 95(2):368–371.

Abdulkadiroglu, A. and Sonmez, T. (1998). Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica*, 66(3):689–702.

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

Abdulkadiroglu, A. and Sonmez, T. (2013). *Matching Markets: Theory and Practice*, volume 1. Advances in Economics and Econometrics, Econometric Society.

Ashlagi, I., Kanoria, Y., and Leshno, J. (2017). Unbalanced random matching markets: The stark effect of competition. *Journal of Political Economy*, 125(1).

Ashlagi, I. and Nikzad, A. (2020). What matters in tie-breaking rules? how competition guides design. *Journal of Economic Theory*, 190.

Ashlagi, I., Nikzad, A., and Romm, A. I. (2019). Assigning more students to their top choices: A tiebreaking rule comparison. *Games and Economic Behavior*, 115:167–187.

Aue, R., Klein, T., and Ortega, J. (2019). What happens when separate and unequal school districts merge?

Azevedo, E. and Leshno, J. (2016). A supply and demand framework for two-sided matching markets. *Journal of Political Economy*, 124(5).

Bade, S. (2019). Random serial dictatorship: The one and only. *Mathematics of Operations Research*.

Carroll, G. (2014). A general equivalence theorem for allocation of indivisible objects. *Journal of Mathematical Economics*, 51:163–177.

Che, Y.-K. and Tercieux, O. (2019). Efficiency and stability in large matching markets. *Journal of Political Economy*, 127(5):2301–2342.

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

de Haan, M., Gautier, P., Oosterbeek, H., and van der Klaauw, B. (2018). The performance of school assignment mechanisms in practice.

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

Ekici, O. (2015). House allocation with existing tenants: Two equivalence results.

EnrollNOLA (2017). Enrollnola annual report.

[7]In fact, Abdulkadiroglu et al. (2009) report that in the first year of implementation, policymakers in New York did not commit to which lottery procedure they would use, leaving students no opportunity to respond to this choice.

Erdil, A. and Ergin, H. (2008). What's the matter with tie-breaking? improving efficiency in school choice. *American Economic Review*, 98(3):669–689.

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

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

Hassidim, A., Romm, A., and Shorrer, R. (2021). The limits of incentives in economic matching procedures. *Management Science*, 67(2).

Hastings, J. S., Kane, T. J., and Staiger, D. O. (2009). Heterogeneous preferences and the efficacy of public school choice.

Immorlica, N. and Mahdian, M. (2005). Marriage, honesty, and stability. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '05, pages 53–62, Philadelphia, PA, USA. Society for Industrial and Applied Mathematics.

Knuth, D. E. (1996). An exact analysis of stable allocation. *Journal of Algorithms*, 20(2):431–442.

Kojima, F. (2015). Recent developments in matching theory and its practical applications. *Advances in Economics and Econometrics; 11th world congress of Econometric Society* .

Kojima, F. and Pathak, P. A. (2009). Incentives and stability in large two-sided matching markets. *The American Economic Review*, pages 608–627.

Larroucau, T. and Rios, I. (2019). Do "short-list" students report truthfully? strategic behavior in the chilean college admissions problem.

Lee, T. and Sethuraman, J. (2011). Equivalence results in the allocation of indivisible objects: A unified view.

McVitie, D. and Wilson, L. (1971). The stable marriage problem. *Communications of the ACM*, 14(7):486–490.

Pathak, P. A. (2011). The Mechanism Design Approach to Student Assignment. *Annual Review of Economics*, 3(1):513–536.

Pathak, P. A. (2016). What really matters in designing school choice mechanisms. In *Advances in Economics and Econometrics, 11th World Congress of the Econometric Society*. Cambridge University Press.

Pathak, P. A. and Sethuraman, J. (2011). Lotteries in student assignment: An equivalence result. *Theoretical Economics*, 6:1–17.

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

Pittel, B. (1989). The average number of stable matchings. *SIAM Journal on Discrete Mathematics*, 2(4):530–549.

Pittel, B. (1992). On likely solutions of a stable marriage problem. *The Annals of Applied Probability*, 2(2):358–401.

Roth, A. E. (1984). The evolution of the labor market for medical interns and residents: A case study in game theory. *Journal of Political Economy*, 92(6):pp. 991–1016.

Roth, A. E. and Peranson, E. (1999). The redesign of the matching market for american physicians: Some engineering aspects of economic design. Working Paper 6963, National Bureau of Economic Research.

Roth, A. E. and Sotomayor, M. (1990). *Two-Sided Matching*. Cambridge University Press.

Shapley, L. S. and Scarf, H. (1974). On cores and indivisibility. *Journal of Mathematical Economics*, 1(1):23–37.

Shi, P. (2021). Optimal priority-based allocation mechanisms. *Management Science*.

Sonmez, T. and Unver, U. (2005). House allocation with existing tenants: an equivalence. *Games and Economic Behavior*, 52(1):153–185.

Whitehurst, G. J. (2017). Education choice and competition index 2016: Summary and commentary.

Wormald, N. C. (1995). Differential equations for random processes and random graphs. *The Annals of Applied Probability*, 5(4):pp. 1217–1235.

Wormald, N. C. (1999). The differential equation method for random graph processes and greedy algorithms. In *Lectures on approximation and randomized algorithms*, pages 73–155.

A. Proofs From Section 4

A.1. **Preliminaries.** We begin with several technical lemmas that will come in handy. When possible, we include remarks interpreting the quantities involved. Our first lemma is a well-known result about non-negative random variables.

Lemma 1. *Let N be a random variable on $\mathbb{N}$. Then $\mathbb{E}[N] = \sum_{k=0}^{\infty} \mathbb{P}(N > k)$.*

Proof.

$$\mathbb{E}[N] = \sum_{j=0}^{\infty} j\, \mathbb{P}(N = j) = \sum_{j=0}^{\infty} \sum_{k=0}^{j-1} \mathbb{P}(N = j) = \sum_{k=0}^{\infty} \sum_{j=k+1}^{\infty} \mathbb{P}(N = j) = \sum_{k=0}^{\infty} \mathbb{P}(N > k).$$

□

We now establish basic facts about the Poisson distribution. We start by defining several additional functions of interest.

$$p_j(\lambda) = \frac{e^{-\lambda}\lambda^j}{j!}, \tag{15}$$

$$\varepsilon(\lambda, C) = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} \min(k, C), \tag{16}$$

$$\mathcal{E}(\lambda) = \sum_{\tau \in \mathcal{T}} \mathcal{D}_\tau \varepsilon(p_\tau \lambda, C_\tau). \tag{17}$$

Remark 1. *A school that receives interest from k students and has capacity C will enroll $\min(k, C)$ students. Therefore, $\varepsilon(\lambda, C)$ gives expected εnrollment at a school with capacity C when the number of interested students follows a Poisson distribution with mean λ. When the average interest per school is λ, the average interest per school of type τ is $p_\tau\lambda$, so $\mathcal{E}(\lambda)$ gives the average $\mathcal{E}$nrollment across schools, given that the average interest per school is λ.*

Lemma 2. *For $C \in \mathbb{N}$,*

$$\frac{d}{d\lambda}\nu(\lambda, C) = -\frac{e^{-\lambda}\lambda^{C-1}}{(C-1)!}, \qquad \alpha(\lambda, C) = \varepsilon(\lambda, C)/\lambda, \qquad \frac{d}{d\lambda}\varepsilon(\lambda, C) = \nu(\lambda, C), \qquad \lim_{\lambda\to\infty} \varepsilon(\lambda, C) = C.$$

Therefore for any $\mathcal{T}, \mathcal{D}$,

$$\left|\mathcal{V}'(\lambda)\right| \le \sum_{\tau \in \mathcal{T}} \mathcal{D}_\tau p_\tau^2, \qquad \mathcal{A}(\lambda) = \mathcal{E}(\lambda)/\lambda \qquad \mathcal{E}'(\lambda) = \mathcal{V}(\lambda), \qquad \lim_{\lambda\to\infty} \mathcal{E}(\lambda) = \sum_{\tau \in \mathcal{T}} \mathcal{D}_\tau C_\tau.$$

Proof. By (10),

$$\frac{d}{d\lambda}\nu(\lambda, C) = -\sum_{k=0}^{C-1} \frac{e^{-\lambda}\lambda^k}{k!} + \sum_{k=0}^{C-2} \frac{e^{-\lambda}\lambda^k}{k!} = -\frac{e^{-\lambda}\lambda^{C-1}}{(C-1)!} \tag{18}$$

From (10) and (18) it follows that

$$\left|\mathcal{V}'(\lambda)\right| = \left|\frac{d}{d\lambda} \sum_{\tau \in \mathcal{T}} \mathcal{D}_\tau p_\tau \nu(p_\tau\lambda, C_\tau)\right| = \sum_{\tau \in \mathcal{T}} \mathcal{D}_\tau p_\tau^2 \frac{e^{-p_\tau\lambda}(p_\tau\lambda)^{C_\tau - 1}}{(C_\tau - 1)!} \le \sum_{\tau \in \mathcal{T}} \mathcal{D}_\tau p_\tau^2.$$

Meanwhile, we have

$$\alpha(\lambda, C) = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} \min\left(\frac{C}{k+1}, 1\right) = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{(k+1)!} \min(k+1, C) = \frac{1}{\lambda}\sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} \min(k, C) = \frac{1}{\lambda}\varepsilon(\lambda, C).$$

From this, it follows that

$$\mathcal{A}(\lambda) = \sum_{\tau\in\mathcal{T}} p_\tau \mathcal{D}_\tau \alpha(p_\tau\lambda, C_\tau) = \sum_{\tau\in\mathcal{T}} \mathcal{D}_\tau \varepsilon(p_\tau\lambda, C_\tau)/\lambda = \mathcal{E}(\lambda)/\lambda.$$

By Lemma 1 it follows that

$$\varepsilon(\lambda, C) = \mathbb{E}[\min(\mathbf{Po}(\lambda), C)] = \sum_{k=0}^{C-1} \mathbb{P}(\mathbf{Po}(\lambda) > k) = \sum_{k=0}^{C-1}(1 - \nu(\lambda, k+1))$$

Therefore, by (18),

$$\frac{d}{d\lambda}\varepsilon(\lambda, C) = \frac{d}{d\lambda}\sum_{k=0}^{C-1}(1 - \nu(\lambda, k+1)) = \sum_{k=0}^{C-1} \frac{e^{-\lambda}\lambda^k}{k!} = \nu(\lambda, C). \tag{19}$$

Similarly, (16) and (19) imply that $\mathcal{E}'(\lambda) = \mathcal{V}(\lambda)$.

Finally, as $\lambda \to \infty$, it is clear that $\varepsilon(\lambda, C) = \mathbb{E}[\min(\mathbf{Po}(\lambda), C)] \to C$, and $\mathcal{E}(\lambda) \to \sum_{\tau\in\mathcal{T}} \mathcal{D}_\tau C_\tau$. □

Remark 2. *Lemma 2 has an intuitive explanation. Recall that $\varepsilon(\lambda, C)$ represents the expected enrollment at a school with C seats which receives a number of proposals that follows a Poisson distribution with mean λ. Meanwhile, $\alpha(\lambda, C)$ represents the acceptance rate at this school, and $\nu(\lambda, C)$ represents the probability that this school has at least one vacancy. Thus, the result $\alpha(\lambda, C) = \varepsilon(\lambda, C)/\lambda$ states that the acceptance rate is equal to the expected number of accepted students divided by the expected number of interested students, while the result $\frac{d}{d\lambda}\nu(\lambda, C) = -\frac{e^{-\lambda}\lambda^{C-1}}{(C-1)!}$ states that an additional student expressing interest eliminates the last vacancy if and only if $C-1$ students had previously proposed. The result $\frac{d}{d\lambda}\varepsilon(\lambda, C) = \nu(\lambda, C)$ states that the increase in enrollment due to an additional student expressing interest is equal to the probability that the school previously had a vacancy. The result $\lim_{\lambda\to\infty}\varepsilon(\lambda, C) = C$ states that if the school has a lot of interest, it will almost always fill all seats.*

Lemma 3. *$\lambda\mathcal{A}'(\lambda) = \mathcal{V}(\lambda) - \mathcal{A}(\lambda) \leq 0$, with strict inequality for $\lambda > 0$.*

Proof. The equality follows from Lemma 2:

$$\mathcal{A}'(\lambda) = \frac{d}{d\lambda}\frac{\mathcal{E}(\lambda)}{\lambda} = \frac{\lambda\mathcal{V}(\lambda) - \mathcal{E}(\lambda)}{\lambda^2} = \frac{\mathcal{V}(\lambda) - \mathcal{A}(\lambda)}{\lambda}.$$

The inequality follows because

$$\mathcal{A}(\lambda) = \mathcal{E}(\lambda)/\lambda = \int_0^1 \mathcal{V}((1-t)\lambda)dt > \int_0^1 \mathcal{V}(\lambda)dt = \mathcal{V}(\lambda).$$

(Note that the first equality holds by definition, the second follows by the substitution $u = (1-t)\lambda$ and Lemma 2, and the inequality follows because $\mathcal{V}(\cdot)$ is decreasing.) □

Remark 3. *The quantity $\mathcal{V}(\lambda)$ represents the fraction of interest that is directed to schools with a vacancy. The quantity $\mathcal{A}(\lambda)$ represents the aggregate acceptance rate. Because schools with* *vacancies accept all interested students, it follows that $\mathcal{V}(\lambda) \le \mathcal{A}(\lambda)$, and that their difference can be interpreted as the fraction of interest that triggers rejection of another student. Therefore, Lemma 3 states that the change in acceptance rate due to additional students expressing interest is equal to the number of additional rejections divided by the total amount of interest λ.*

Lemma 4. *For any $\mathcal{L}$, $\mu(\cdot)$ is decreasing and convex, with*

$$\mu(q) \le \sum_{k=0}^{\infty} \mathcal{L}(>k) = \mathop{\mathbb{E}}_{l\sim\mathcal{L}}[l] \tag{20}$$

$$\mu'(q) \le \sum_{k=0}^{\infty} k\mathcal{L}(>k) = \frac{1}{2}\mathop{\mathbb{E}}_{l\sim\mathcal{L}}[l^2 - l]. \tag{21}$$

Furthermore, $q \cdot \mu(q)$ is increasing in q, and at most 1.

Proof. The fact that $\mu(q)$ is less than the average list length $\mu(0) = \sum_{k=0}^{\infty} \mathcal{L}(>k)$ follows by inspection of (3). Differentiate (3) to see that

$$\mu'(q) = -\sum_{k=0}^{\infty} k\mathcal{L}(>k)(1-q)^k, \tag{22}$$

from which it follows that μ is decreasing and convex. Furthermore, (22) implies (21).

Monotonicity of $q \cdot \mu(q)$ holds because

$$\begin{aligned} q \cdot \mu(q) &= \sum_{k=0}^{\infty}(1-q)^k \mathcal{L}(>k) - \sum_{k=0}^{\infty}(1-q)^{k+1}\mathcal{L}(>k) \\ &= \sum_{k=0}^{\infty}(1-q)^k \mathcal{L}(>k) - \sum_{k=1}^{\infty}(1-q)^{k}\mathcal{L}(>k-1) \\ &= 1 - \sum_{k=1}^{\infty} \mathcal{L}(k)(1-q)^k, \end{aligned}$$

which is clearly increasing in q. The claim that $q \cdot \mu(q) \le 1$ follows by taking $q = 1$. □

Remark 4. *The intuition for Lemma 4 is as follows. Consider a student for whom each application is accepted with probability q. The quantity $\mu(q)$ represents the expected number of schools to which this student will propose. This quantity increases as the acceptance probability decreases, and is bounded by the average list length. The quantity $q\,\mu(q)$ represents the probability that this student matches, which is clearly less than one, and increases with q.*

Lemma 5. *For $R \in \{I, C\}$ the following equalities hold:*

$$\mathcal{E}(\lambda_R) = \rho \mathop{\mathbb{E}}_{k\sim\mathcal{L}}[F_R(k)]. \tag{23}$$

$$\lambda_R = \rho \sum_{k=0}^{\infty} \mathcal{L}(>k)(1 - F_R(k)). \tag{24}$$

Proof. We first prove (23) for $R = I$. By definition of $\mathcal{A}$ in (9) we have

$$\begin{aligned}\mathcal{E}(\lambda_I) &= \lambda_I \mathcal{A}(\lambda_I) \\ &= \rho \mathcal{A}(\lambda_I)\mu(\mathcal{A}(\lambda_I)) \\ &= \rho(1 - \mathop{\mathbb{E}}_{\ell\sim\mathcal{L}}[(1-\mathcal{A}(\lambda_I))^\ell]) \\ &= \rho \mathop{\mathbb{E}}_{\ell\sim\mathcal{L}}[F_I(\ell)].\end{aligned}$$

The second equality follows from (4). The third follows from the definition of μ in (3) and the observation that

$$q\mu(q) = q \mathop{\mathbb{E}}_{\ell\sim\mathcal{L}}[\sum_{k=0}^{\ell-1}(1-q)^k] = \mathop{\mathbb{E}}_{\ell\sim\mathcal{L}}[1-(1-q)^\ell]. \tag{25}$$

The fourth equality follows from the definitions of F_I in (5) and (12).

Next, we prove (23) for $R = S$. By definition of F_S in (7) and (12), we have

$$\begin{aligned}\mathop{\mathbb{E}}_{\ell\sim\mathcal{L}}[F_S(\ell)] &= \int_0^1 \mathop{\mathbb{E}}_{\ell\sim\mathcal{L}}[1-(1-\mathcal{V}(\Lambda_S(t)))^\ell]dt \\ &= \int_0^1 \mathcal{V}(\Lambda_S(t))\mu(\mathcal{V}(\Lambda_S(t)))dt. \\ &= \frac{1}{\rho}\int_0^{\Lambda_S(0)} \mathcal{V}(\lambda)d\lambda \\ &= \mathcal{E}(\Lambda_S(0))/\rho.\end{aligned}$$

The second equality follows from (25), the third from the substitution $\lambda = \Lambda_S(t)$ and (6), which states that $\frac{d\lambda_S}{dt} = -\rho\mu(\mathcal{V}(\Lambda_S(t)))$, and the final equality from Lemma 2.

We now prove (24) for $R = I$. This follows immediately from the definitions of λ_I in (4), μ in (3), and F_I in (5):

$$\begin{aligned}\lambda_I &= \rho\mu(\mathcal{A}(\lambda_I)) \\ &= \rho\sum_{k=0}^{\infty}\mathcal{L}(>k)(1-\mathcal{A}(\lambda_I))^k \\ &= \rho\sum_{k=0}^{\infty}\mathcal{L}(>k)(1-F_I(k)).\end{aligned}$$

Finally, we prove (24) for $R = S$. By the definition of F_S in (7),

$$\begin{aligned}\rho\sum_{k=0}^{\infty}\mathcal{L}(>k)(1-F_S(k)) &= \rho\sum_{k=0}^{\infty}\mathcal{L}(>k)\int_0^1(1-\mathcal{V}(\Lambda_S(t)))^k dt \\ &= \rho\int_0^1 \mu(\mathcal{V}(\Lambda_S(t)))dt \\ &= \int_0^{\Lambda_S(0)} d\lambda = \Lambda_S(0).\end{aligned}$$

The final line follows from substituting $\lambda = \Lambda_S(t)$, and noting that $\frac{d\lambda}{dt} = -\rho\mu(\mathcal{V}(\Lambda_S(t)))$ by (6). Note also that $\lambda_S = \Lambda_S(0)$ by definition. $\square$

Remark 5. *Recall that $\mathcal{E}(\lambda_R)$ is interpreted as the average school enrollment. This must equal the ratio of students to schools ρ times the fraction of students who are assigned, $\mathbb{E}_{\ell\sim\mathcal{L}}[F_R(\ell)]$. Meanwhile, the quantity $\mathcal{L}(>k)(1-F_R(k))$ gives the fraction of students who propose to at least $k+1$ schools (they list more than k schools and do not get into any of their first k choices). Therefore, by Lemma 1, the right side of (24) is the average number of schools at which a student expresses interest, times the ratio of students to schools, which must equal schools' average number of interested students λ_R.*

A.2. **Proof of Theorem 2.** The proof of Theorem 2 follows from Lemmas 6 and 7. Lemma 6 says that F_I and F_S have at most one crossing point, implying that either there is one crossing point or one dominates the other. Lemma 7 states that it is not possible for F_I to dominate F_S, or vice versa.

Lemma 6. *If $F_I(k) > F_S(k)$, then $F_I(k') > F_S(k')$ for all $k' > k$.*

Proof. Recall that

$$F_I(k) = 1 - (1-\mathcal{A}(\lambda_I))^k, \qquad F_S(k) = 1 - \int_0^1 (1-\mathcal{V}(\Lambda_S(t)))^k dt.$$

It follows that for $k' \geq k$,

$$\begin{aligned} 1 - F_I(k') &= (1-F_I(k))^{k'/k} \\ &< (1-F_S(k))^{k'/k} \\ &= \left(\int_0^1 (1-\mathcal{V}(\Lambda_S(t)))^k dt\right)^{k'/k} \\ &\leq \int_0^1 (1-\mathcal{V}(\Lambda_S(t)))^{k'} dt \\ &= 1 - F_S(k'), \end{aligned}$$

where the first inequality follows because $F_I(k) > F_S(k)$, and the second by Jensen's inequality. $\square$

Lemma 7. *Let $L = \{k \in \mathbb{N} : \mathcal{L}(k) > 0\}$ be the support of the list length distribution. There exists $k \in L$ such that $F_I(k) > F_S(k)$ if and only if there exists $k' \in L$ such that $F_I(k') < F_S(k')$.*

Proof. Seeking a contradiction, suppose that one of the procedures dominates the other. That is, for $R, \tilde{R} \in \{I, S\}$,

$$F_R(k) \geq F_{\tilde{R}}(k) \text{ for all } k \in L, \text{ with strict inequality for some } k. \tag{26}$$

Note that (12) and (26) jointly imply that $\mathbb{E}_{\ell\sim\mathcal{L}}[F_R(\ell)] > \mathbb{E}_{\ell\sim\mathcal{L}}[F_{\tilde{R}}(\ell)]$. But (24) and (26) imply that $\lambda_R < \lambda_{\tilde{R}}$, and therefore that $\mathcal{E}(\lambda_R) < \mathcal{E}(\lambda_{\tilde{R}})$. By (23), this implies $\mathbb{E}_{\ell\sim\mathcal{L}}[F_R(\ell)] < \mathbb{E}_{\ell\sim\mathcal{L}}[F_{\tilde{R}}(\ell)]$, which is a contradiction. Thus, (26) cannot hold. $\square$

A.3. **Proofs from Section 4.2: Theorem 3, Proposition 1, and Proposition 2.**

Remark 6. *This remark provides interpretation for the quantities used in the proof of Theorem 3. The quantity λ'_R can be interpreted as the expected amount of additional interest triggered by the addition of a new student to the market under priority rule R. Note that this may be higher than the number of schools at which this student expresses interest, because the student may cause others to be rejected. Because schools on each student's list are sampled iid, there is a one-to-one correspondence between the total interest λ_R and enrollment $\mathcal{E}(\lambda_R)$. Thus, the strategy is to show that whenever the two methods have matched the same number of students so far, then the number of new proposals (or equivalently, the number of new matches) triggered by the addition of one student is higher under one procedure than the other.*

Proof of Theorem 3. Because $\mathcal{E}(\cdot)$ is monotonic, it suffices to compare λ_I to $\lambda_S = \Lambda_S(0)$. We do this by fixing $\mathcal{L}$ and $\mathcal{D}$, and considering the values λ_R as functions of ρ.

Because $\lambda_I(0) = \lambda_S(0) = 0$ and both λ_I and λ_S are continuous and differentiable functions of ρ, to show that $\lambda_I(\rho) \geq \lambda_S(\rho)$ for all $\rho \geq 0$, it suffices to show that if $\rho_I, \rho_S \in \mathbb{R}_+$ satisfy $\lambda_I(\rho_I) = \lambda_S(\rho_S)$, then $\lambda'_I(\rho_I) > \lambda'_S(\rho_S)$. Similarly, to show that $\lambda_I(\rho) \leq \lambda_S(\rho)$ for all $\rho \geq 0$, it suffices to show that

$$\lambda_I(\rho_I) = \lambda_S(\rho_S) \Rightarrow \lambda'_I(\rho_I) < \lambda'_S(\rho_S).$$

We now compute λ'_I and λ'_S. Differentiating (4) with respect to ρ (and dropping the implicit dependence of λ_I and λ'_I on ρ), we get

$$\lambda'_I = \mu(\mathcal{A}(\lambda_I)) + \rho\, \mu'(\mathcal{A}(\lambda_I))\, \mathcal{A}'(\lambda_I)\, \lambda'_I. \tag{27}$$

From Lemma 3 and (4), it follows that that

$$\rho \cdot \mathcal{A}'(\lambda_I) = \rho\, \frac{\mathcal{V}(\lambda_I) - \mathcal{A}(\lambda_I)}{\lambda_I} = \frac{\mathcal{V}(\lambda_I) - \mathcal{A}(\lambda_I)}{\mu(\mathcal{A}(\lambda_I))}.$$

Substituting this expression into (27) and solving for λ'_I, we see that the derivative of λ_I with respect to ρ is

$$\lambda'_I = \frac{\mu(\mathcal{A}(\lambda_I))}{1 - (\mathcal{V}(\lambda_I) - \mathcal{A}(\lambda_I))\frac{\mu'(\mathcal{A}(\lambda_I))}{\mu(\mathcal{A}(\lambda_I))}}. \tag{28}$$

Meanwhile, we claim that

$$\lambda'_S = -\frac{1}{\rho}\Lambda'_S(0) = \mu(\mathcal{V}(\lambda_S)). \tag{29}$$

This can be seen by noting that if we view Λ_S as a function of t and ρ, then for $\overline{\rho} > \rho$, $\lambda_S(\rho) = \Lambda_S(0, \rho) = \Lambda_S(1 - \rho/\overline{\rho}, \overline{\rho})$ (both quantities represent the total interest expressed by the first ρ students). Differentiating this equation with respect to ρ yields (29).

It follows from (28) and (29) that if $\lambda_I(\rho_I) = \lambda_S(\rho_S) = \lambda$, then (letting $\mathcal{V} = \mathcal{V}(\lambda), \mathcal{A} = \mathcal{A}(\lambda)$)

$$\begin{aligned}
\lambda_S'(\rho_S) < \lambda_I'(\rho_I) &\Leftrightarrow \mu(\mathcal{V}) < \frac{\mu(\mathcal{A})}{1 - (\mathcal{V} - \mathcal{A})\frac{\mu'(\mathcal{A})}{\mu(\mathcal{A})}} \\
&\Leftrightarrow \mu(\mathcal{V}) - \mu(\mathcal{A}) < (\mathcal{V} - \mathcal{A})\mu'(\mathcal{A})\frac{\mu(\mathcal{V})}{\mu(\mathcal{A})} \\
&\Leftrightarrow \frac{\mu(\mathcal{V}) - \mu(\mathcal{A})}{\mu(\mathcal{V})\mu(\mathcal{A})} < (\mathcal{A} - \mathcal{V})f'(\mathcal{A}). \\
&\Leftrightarrow \frac{f(\mathcal{A}) - f(\mathcal{V})}{\mathcal{A} - \mathcal{V}} < f'(\mathcal{A}),
\end{aligned}$$

where we have used the fact that $f'(q) = -\mu'(q)/\mu(q)^2$. Because $\mathcal{V} < \mathcal{A}$ by Lemma 3, this holds if f is convex; if f is concave, the inequality reverses.

□

Proof of Proposition 1. Differentiating the function f we obtain

$$f''(q) = \frac{2\mu'(q)^2 - \mu(q)\mu''(q)}{\mu(q)^3}.$$

Thus, to determine convexity or concavity of f, it suffices to study the sign of $2\mu'(q)^2 - \mu(q)\mu''(q)$. Straightforward algebra reveals that

$$f''(q)\mu(q)^3 = 2\mu'(q)^2 - \mu(q)\mu''(q) = \sum_{k=0}^{\infty} q_k(1-q)^k, \tag{30}$$

where

$$q_{k-2} = \frac{1}{2}\sum_{i=0}^{k}\left(4i(k-i) - i(i-1) - (k-i)(k-i-1)\right)(1-\mathcal{L}(i))(1-\mathcal{L}(k-i)).$$

For $k \in \mathbb{N}$, recall that $\mathcal{L}(>k)$ is the probability of listing more than k schools. For $i \le k$, define

$$\begin{aligned}
r_i^k &= 4i(k-i) - i(i-1) - (k-i)(k-i-1), \\
s_i^k &= \mathcal{L}(>i)\mathcal{L}(>k-i),
\end{aligned}$$

so that

$$q_{k-2} = \frac{1}{2}\sum_{i=0}^{k} r_i^k s_i^k.$$

Noting that $r_i^k = r_{k-i}^k$ and $s_i^k = s_{k-i}^k$, we can express q_{k-2} as follows:

$$q_{k-2} = \sum_{i=0}^{\lfloor k/2 \rfloor} \tilde{r}_i^k s_i^k, \tag{31}$$

where we define $\tilde{r}_i^k = r_i^k$ if $i < k/2$ and $\tilde{r}_i^k = r_i^k/2$ if $i = k/2$.

We will show that

i) The $\tilde{r}_i^k$ "cross zero" once and sum to zero.
ii) If $\mathcal{L}$ has a weakly increasing hazard rate, then s_i^k is increasing on $\{0, 1, \ldots, \lfloor k/2 \rfloor\}$.
   If $\mathcal{L}$ has a weakly decreasing hazard rate, then s_i^k is decreasing on $\{0, 1, \ldots, \lfloor k/2 \rfloor\}$.

From i), ii) and (119), Lemma 8 implies that if $\mathcal{L}$ has a weakly increasing hazard rate, then each q_k is non-negative; if $\mathcal{L}$ has a weakly increasing hazard rate, then each q_k is non-positive. By (30) the proposition immediately follows.

All that remains is to establish i) and ii). Note that

$$r_i^k = 6i(k-i) - k(k-1), \tag{32}$$

which is a quadratic in i. It is negative at $i = 0$ and obtains its maximal value at $k/2$. It follows that the sequence $\{r_i^k\}_{i=1}^{\lfloor k/2 \rfloor}$ is initially negative and later non-negative, so the same is true of $\tilde{r}_i^k$. Furthermore, the definition of $\tilde{r}_i^k$ implies that

$$\sum_{i=0}^{\lfloor k/2 \rfloor} \tilde{r}_i^k = \sum_{i=0}^{k} r_i^k = 0,$$

where the final equality follows from (32) by applying standard formulas for $\sum_{i=0}^k i$ and $\sum_{i=0}^k i^2$.

Point ii) follows because for $i \geq 1$,

$$s_{i-1}^k \leq s_i^k \Leftrightarrow \frac{\mathcal{L}(>k-i+1)}{\mathcal{L}(>k-i)} \leq \frac{\mathcal{L}(>i)}{\mathcal{L}(>i-1)} \Leftrightarrow 1 - \frac{\mathcal{L}(>i)}{\mathcal{L}(>i-1)} \leq 1 - \frac{\mathcal{L}(>k-i+1)}{\mathcal{L}(>k-i)}. \tag{33}$$

If $i \leq \lfloor k/2 \rfloor$, then $i < k - i + 1$, so if $\mathcal{L}$ has an increasing hazard rate, then all inequalities in (33) hold, and if $\mathcal{L}$ has a decreasing hazard rate, then all inequalities reverse.

□

Lemma 8. *Let $\{r_i\}_{i=0}^k$ be a sequence of real numbers with mean zero that "crosses zero" once. That is, $\sum_{i=0}^k r_i = 0$ and for some $j < k$, it holds that $r_i < 0$ for $i \leq j$ and $r_i \geq 0$ for $j < i \leq k$. Let $\{s_i\}_{i=0}^k$ be a sequence of non-negative numbers. If s_i is weakly increasing, then $\sum_{i=0}^k r_i s_i \geq 0$, and if s_i is weakly decreasing, then $\sum_{i=0}^k r_i s_i \leq 0$.*

Proof. If s_i is weakly increasing, then

$$\sum_{i=0}^{k} r_i s_i = \sum_{i=0}^{j} r_i s_i + \sum_{i=j+1}^{k} r_i s_i \geq \sum_{i=0}^{j} r_i s_j + \sum_{i=j+1}^{k} r_i s_j = s_j \sum_{i=1}^{k} r_i = 0,$$

where the inequality follows because the first sum (consisting of negative terms) has been made "less negative" and the second sum (consisting of non-negative terms) has been made "more positive." Analogous logic holds when s_i is weakly decreasing. □

Proof of Proposition 2. With a single lottery, we use Lemma (2) and (6) to note that

$$\frac{d}{dt}\mathcal{E}(\Lambda_S(t)) = \mathcal{V}(\Lambda_S(t))\Lambda_S'(t) = -\rho\mathcal{V}(\Lambda_S(t))\mu(\mathcal{V}(\Lambda_S(t))). \tag{34}$$

When the number of schools listed follows a Poisson distribution with mean ℓ,

$$q\mu(q) = 1 - e^{-q\ell}. \tag{35}$$

When schools are equally popular and have a single seat,

$$\mathcal{V}(\lambda) = e^{-\lambda} = 1 - \mathcal{E}(\lambda). \tag{36}$$

Substituting (35) and (36) into (34) yields

$$\frac{d}{dt}\mathcal{E}(\Lambda_S(t)) = -\rho(1 - e^{-\ell(1-\mathcal{E}(\Lambda_S(t)))}),$$

which (with initial condition $\mathcal{E}(\Lambda_S(1)) = 0$) has solution

$$\mathcal{E}(\Lambda_S(t)) = 1 - \frac{1}{\ell}\log(1 + (e^{\ell} - 1)e^{-\rho\ell(1-t)}).$$

The claimed expressions for the single lottery follows from $F_S(\mathcal{L}) = \mathcal{E}(\Lambda_S(0))/\rho$ (see Lemma 5).

With independent lotteries,

$$\rho F_I(\mathcal{L}) = \mathcal{E}(\lambda_I) = 1 - e^{-\lambda_I},$$

where the second equality uses (36). We can solve this for λ_I to get

$$\lambda_I = -\log(1 - \rho F_I(\mathcal{L})),$$

from which it follows that

$$\mathcal{A}(\lambda_I) = \frac{\rho F_I(\mathcal{L})}{-\log(1 - \rho F_I(\mathcal{L}))} \tag{37}$$

Furthermore, (4) implies that

$$\lambda_I = \rho\mu(\mathcal{A}(\lambda_I)) = \rho\frac{\mathcal{E}(\ell\mathcal{A}(\lambda_I))}{\mathcal{A}(\lambda_I)},$$

where the second equality uses (35) and (36). Multiplying each side by $\mathcal{A}(\lambda_I)/\rho$ and applying Lemma 5 we get

$$F_I(\mathcal{L}) = \mathcal{E}(\ell\mathcal{A}(\lambda_I)) = 1 - e^{-\ell\mathcal{A}(\lambda_I)},$$

from which it follows that

$$\mathcal{A}(\lambda_I) = -\frac{1}{\ell}\log(1 - F_I(\mathcal{L})).$$

Combining this with (37) yields

$$\frac{1}{\ell}\log(1 - F_I(\mathcal{L})) = \frac{\rho F_I(\mathcal{L})}{\log(1 - \rho F_I(\mathcal{L}))},$$

which converges to $\rho/\log(1-\rho)$ as $F_I(\mathcal{L}) \to 1$. □

A.4. **Proof of Proposition 3.** We first show that if the mass of students exceeds the total mass of seats, then under either lottery procedure the total amount of interest grows linearly in the length of student lists.

Lemma 9. *Suppose that all students list ℓ schools ($\mathcal{L}(\ell) = 1$), and $\rho > \sum_{\tau\in\mathcal{T}}\mathcal{D}_\tau C_\tau$. Then*

$$\min(\lambda_I, \lambda_S) \geq \ell(\rho - \sum_{\tau\in\mathcal{T}}\mathcal{D}_\tau C_\tau). \tag{38}$$

Proof. We start by noting that when all lists have length ℓ,

$$\mu(q) = \sum_{k=0}^{\infty}\mathcal{L}(>k)(1-q)^k = \sum_{k=0}^{\ell-1}(1-q)^k \geq \ell(1-q)^{\ell}. \tag{39}$$

It follows from (4) that

$$\lambda_I = \rho\mu(\mathcal{A}(\lambda_I)) \geq \ell\rho(1-\mathcal{A}(\lambda_I))^\ell = \ell\rho(1-F_I(\mathcal{L})) = \ell(\rho - \mathcal{E}(\lambda_I)). \tag{40}$$

The inequality follows from (39), the subsequent equality from the definition of F_I in (5) and (12), and the final equality from Lemma 5. Similarly (6) implies that

$$\begin{aligned}\lambda_S = \Lambda_S(0) &= \int_0^1 \rho\mu(\mathcal{V}(\Lambda_S(t)))dt.\\ &\geq \ell\rho\int_0^1 (1-\mathcal{V}(\Lambda_S(t)))^\ell dt\\ &= \ell\rho(1-F_S(\mathcal{L})) = \ell(\rho - \mathcal{E}(\Lambda_S(0))). \end{aligned} \tag{41}$$

The inequality uses (39), and the final line uses the definition of F_S in (7) and (12) and Lemma 5. The result follows Lemma 2, which states that $\mathcal{E}(\cdot)$ is upper-bounded by $\sum_{\tau\in\mathcal{T}}\mathcal{D}_\tau C_\tau$. $\square$

Remark 7. *The bound in Lemma 9 has a natural interpretation: capacity constraints imply that the ratio of unassigned students to schools must be at least $\rho - \sum_{\tau\in\mathcal{T}}\mathcal{D}_\tau C_\tau$, and each unassigned student proposes to all ℓ schools on her list.*

We next provide a tight lower bound on the fraction of assigned students who receive their first choice when using each procedure.

Lemma 10. *Fix $\mathcal{T},\mathcal{D}$. Under a single lottery, the fraction of assigned students who receive their first choice is at least*

$$\inf_{\rho,\mathcal{L}} \frac{F_S(1)}{\mathbb{E}_{\ell\sim\mathcal{L}}[F_S(\ell)]} = \frac{\int_0^\infty \mathcal{V}(\lambda)^2 d\lambda}{\int_0^\infty \mathcal{V}(\lambda)d\lambda},$$

whereas under independent lotteries,

$$\inf_{\rho,\mathcal{L}} \frac{F_I(1)}{\mathbb{E}_{\ell\sim\mathcal{L}}[F_I(\ell)]} = 0.$$

Proof of Lemma 10. We begin with the statement about independent lotteries. The fraction of students who receive their top choice is $F_I(1) = \mathcal{A}(\lambda_I)$, and by Lemma 5 the fraction who are matched is $F_I(\mathcal{L}) = \frac{1}{\rho}\mathcal{E}(\lambda_I)$. By (9), it follows that

$$\frac{F_I(1)}{F_I(\mathcal{L})} = \frac{\rho\cdot\mathcal{A}(\lambda_I)}{\mathcal{E}(\lambda_I)} = \frac{\rho}{\lambda_I}.$$

Lemma 9 implies that for $\rho > \sum_\tau \mathcal{D}_\tau C_\tau$, this ratio can be made arbitrarily small by letting $\mathcal{L}(\ell) = 1$ for some sufficiently large ℓ.

We now turn to the statement for the single lottery, using the shorthand $\lambda_S = \Lambda_S(0)$. By Lemma 5, the fraction of students who are assigned is

$$F_S(\mathcal{L}) = \frac{1}{\rho}\mathcal{E}(\lambda_S) = \frac{1}{\rho}\int_0^{\lambda_S}\mathcal{V}(\lambda)d\lambda,$$

where the second equality uses Lemma 2. The fraction of students who receive their top choice is

$$F_S(1) = \int_0^1 \mathcal{V}(\Lambda_S(t))dt = \frac{1}{\rho}\int_0^{\lambda_S}\frac{\mathcal{V}(\lambda)}{\mu(\mathcal{V}(\lambda))}d\lambda \geq \frac{1}{\rho}\int_0^{\lambda_S}\mathcal{V}(\lambda)^2 d\lambda,$$

where we use the fact that $1/\mu(\mathcal{V}(\lambda)) \geq \mathcal{V}(\lambda)$ by Lemma 4. It follows that

$$\frac{F_S(1)}{F_S(\mathcal{L})} \geq \frac{\int_0^{\lambda_S} \mathcal{V}(\lambda)^2 d\lambda}{\int_0^{\lambda_S} \mathcal{V}(\lambda) d\lambda} \geq \frac{\int_0^{\infty} \mathcal{V}(\lambda)^2 d\lambda}{\int_0^{\infty} \mathcal{V}(\lambda) d\lambda}. \tag{42}$$

The final inequality follows because

$$\frac{d}{d\lambda_S} \log\left(\frac{\int_0^{\lambda_S} \mathcal{V}(\lambda)^2 d\lambda}{\int_0^{\lambda_S} \mathcal{V}(\lambda) d\lambda}\right) = \frac{\mathcal{V}(\lambda_S)^2}{\int_0^{\lambda_S} \mathcal{V}(\lambda)^2 d\lambda} - \frac{\mathcal{V}(\lambda_S)}{\int_0^{\lambda_S} \mathcal{V}(\lambda) d\lambda} < 0,$$

with the inequality following from the fact that $\mathcal{V}(\cdot)$ is decreasing.

We now show that the inequalities in (42) can be made tight. For $\ell \in \mathbb{N}$, take $\mathcal{L}(k) = \mathbf{1}(k \geq \ell)$, meaning that all students list exactly ℓ schools. Then

$$\begin{aligned} F_S(1) - \frac{1}{\rho}\int_0^{\lambda_S} \mathcal{V}(\lambda)^2 d\lambda &= \frac{1}{\rho}\int_0^{\lambda_S} \mathcal{V}(\lambda)\left(1/\mu(\mathcal{V}(\lambda)) - \mathcal{V}(\lambda)\right) d\lambda \\ &\leq \frac{1}{\rho\ell}\int_0^{\lambda_S} \mathcal{V}(\lambda) d\lambda \\ &\leq \frac{1}{\rho\ell}\sum_{\tau\in\mathcal{T}} \mathcal{D}_\tau C_\tau. \end{aligned} \tag{43}$$

The second inequality follows by Lemma 2, and the first because when $\mathcal{L}(k) = \mathbf{1}(k \geq \ell)$, we have that for $q \in (0, 1]$,

$$1/\mu(q) - q = \frac{q(1-q)^\ell}{1-(1-q)^\ell},$$

which is positive, decreasing in q, and approaches $1/\ell$ as $q \to 0$. Note that (43) implies that the first inequality in (42) becomes tight as $\ell \to \infty$. Furthermore, Lemma 9 states that if $\rho > \sum_{\tau\in\mathcal{T}} \mathcal{D}_\tau C_\tau$, then $\lambda_S \to \infty$ as $\ell \to \infty$, implying that the second inequality in (42) becomes tight. □

Remark 8. *The lower bound for a single lottery given in Lemma 10 can be explained as follows. The worst case occurs when students outnumber seats, and all students submit long lists. In this case, all schools will receive a lot of interest, and all seats will fill. Recalling that $\mathcal{V}(\lambda) = \mathcal{E}'(\lambda)$ (see Lemma 2), the denominator can be re-expressed as $\mathcal{E}(\infty) = \sum_\tau \mathcal{D}_\tau C_\tau$. That is, it represents the total number of matches. Meanwhile, for students with priority t, a fraction $\mathcal{V}(\Lambda(t))$ of their proposals are accepted. Therefore, they propose to an average of $1/\mathcal{V}(\Lambda(t))$ schools for each acceptance, implying that only a fraction $\mathcal{V}(\Lambda(t))$ of their proposals go to their first choice. It follows that among these students, the fraction of all proposals that both go to a first choice school and are accepted is $\mathcal{V}(\Lambda(t))^2$. Adding this up across students reveals that the number of first-choices is lower-bounded by the numerator.*

In light of Lemma 10, given a school type space $\mathcal{T}$ and a distribution $\mathcal{D}$ on $\mathcal{T}$, define

$$LB(\mathcal{T}, \mathcal{D}) = \frac{\int_0^{\infty} \mathcal{V}(\lambda)^2 d\lambda}{\int_0^{\infty} \mathcal{V}(\lambda) d\lambda}. \tag{44}$$

Note that the dependence on $\mathcal{T}$ and $\mathcal{D}$ is implicit, through the function $\mathcal{V}(\cdot)$.

Our next step in the proof of Proposition 3 is Lemma 11, which establishes that when schools have identical capacities, the distribution $\mathcal{D}$ that minimizes the bound in (44) places all of its mass on the extreme points of $\mathcal{T}$. That is, the worst-case distribution of popularity is to have two types of schools: one popular, and the other unpopular.

Lemma 11. *Fix $\mathcal{T}$ such that $C_\tau = C \in \mathbb{N}$ for all $\tau \in \mathcal{T}$, and define*

$$\underline{p} = \min_{\tau \in \mathcal{T}} p_\tau, \qquad \overline{p} = \max_{\tau \in \mathcal{T}} p_\tau. \tag{45}$$

Let $\mathcal{D}^*(\mathcal{T})$ be the distribution that places probability β on $(\underline{p}, C)$ and probability $1-\beta$ on $(\overline{p}, C)$, where $\beta\underline{p} + (1-\beta)\overline{p} = 1$. Then $LB(\mathcal{T}, \mathcal{D}^*(\mathcal{T})) = \inf_{\mathcal{D}} LB(\mathcal{T}, \mathcal{D})$.

Proof. In this proof, it will be helpful to make the dependence of $\mathcal{V}$ and $\mathcal{E}$ on $\mathcal{D}$ explicit. Thus, we write $\mathcal{V}(\lambda, \mathcal{D})$ and $\mathcal{E}(\lambda, \mathcal{D})$ in place of $\mathcal{V}(\lambda)$ and $\mathcal{E}(\lambda)$. Because $\mathcal{T}$ is fixed, we write $\mathcal{D}^*$ for $\mathcal{D}^*(\mathcal{T})$.

By Lemma 2, for any $\mathcal{D}$ we have

$$\int_0^\infty \mathcal{V}(\lambda, \mathcal{D}) d\lambda = \lim_{\lambda \to \infty} \mathcal{E}(\lambda, \mathcal{D}) = \sum_{\tau \in \mathcal{T}} \mathcal{D}_\tau C_\tau.$$

When $C_\tau = C$ for all τ, it follows that this expression is C, so minimizing the lower bound in (44) is equivalent to minimizing the numerator $\int_0^\infty \mathcal{V}(\lambda, \mathcal{D})^2 d\lambda$.

We show that if $\mathcal{D}$ places positive mass $q > 0$ on a school type $\tau' \in \mathcal{T}$ with $p_{\tau'} = p \in (\underline{p}, \overline{p})$, then moving this mass to the extreme points $\underline{p}$ and $\overline{p}$ decreases LB. Formally, let $\tilde{\mathcal{D}}$ be the modification of $\mathcal{D}$ in which the probability of τ' under $\tilde{\mathcal{D}}$ is reduced to zero, and the probability of $(\underline{p}, C)$ and $(\overline{p}, C)$ are increased by $q\beta$ and $q(1-\beta)$, respectively. The definition of β ensures that $\tilde{\mathcal{D}}$ remains a probability distribution and that average popularity $\sum_\tau \tilde{\mathcal{D}}_\tau p_\tau = 1$. We will show that

$$\int_0^\infty \mathcal{V}(\lambda, \tilde{\mathcal{D}})^2 d\lambda \le \int_0^\infty \mathcal{V}(\lambda, \mathcal{D})^2 d\lambda. \tag{46}$$

Note that for any $\mathcal{D}$, $\mathcal{V}(\lambda, \mathcal{D})$ is one at $\lambda = 0$ and decreases to zero as $\lambda \to \infty$. Our first step is to show that the random variable X that has CDF $1 - \mathcal{V}(\cdot, \mathcal{D})$ second order stochastically dominates that the random variable $\tilde{X}$ with CDF $1 - \mathcal{V}(\cdot, \tilde{\mathcal{D}})$. That is, for any $\lambda \in \mathbb{R}_+$,

$$\int_0^\lambda \mathcal{V}(x, \mathcal{D}) - \mathcal{V}(x, \tilde{\mathcal{D}}) dx = \mathcal{E}(\lambda, \mathcal{D}) - \mathcal{E}(\lambda, \tilde{\mathcal{D}}) \ge 0, \tag{47}$$

where the equality above uses Lemma 2. By (17) we have $\mathcal{E}(\lambda, \mathcal{D}) = \mathbb{E}_{\tau \sim \mathcal{D}}[\varepsilon(p_\tau \lambda, C)]$. Lemma 2 implies that ε is concave in its first argument, from which it follows that if we let $\tau \sim \mathcal{D}$ and $\tilde{\tau} \sim \tilde{\mathcal{D}}$, then $\mathbb{E}[p_\tau] = \mathbb{E}[p_{\tilde{\tau}}]$ and by Jensen's inequality (47) holds.

Having established second order stochastic dominance, (46) follows from a chain of inequalities:

$$\begin{aligned}
\int_0^\infty \mathcal{V}(\lambda, \tilde{\mathcal{D}})^2 d\lambda &= -\int_0^\infty \mathcal{E}(\lambda, \tilde{\mathcal{D}}) \mathcal{V}'(\lambda, \tilde{\mathcal{D}}) d\lambda \\
&\le -\int_0^\infty \mathcal{E}(\lambda, \tilde{\mathcal{D}}) \mathcal{V}'(\lambda, \mathcal{D}) d\lambda \\
&= \int_0^\infty \mathcal{V}(\lambda, \mathcal{D}) \mathcal{V}(\lambda, \tilde{\mathcal{D}}) d\lambda \\
&= -\int_0^\infty \mathcal{E}(\lambda, \mathcal{D}) \mathcal{V}'(\lambda, \tilde{\mathcal{D}}) d\lambda \\
&\le -\int_0^\infty \mathcal{E}(\lambda, \mathcal{D}) \mathcal{V}'(\lambda, \mathcal{D}) d\lambda \\
&= \int_0^\infty \mathcal{V}(\lambda, \mathcal{D})^2 d\lambda.
\end{aligned}$$

The notation $\mathcal{V}'(\lambda, \mathcal{D})$ denotes the derivative of $\mathcal{V}$ with respect to λ. All equalities above follow from integration by parts, using the fact that $\mathcal{E}'(\lambda) = \mathcal{V}(\lambda)$ (Lemma 2), $\mathcal{E}(0) = 0$, and $\mathcal{V}(\infty) = 0$. The inequalities follow from second-order stochastic dominance, which implies that for any concave increasing function u,

$$-\int_0^\infty u(\lambda)\mathcal{V}'(\lambda, \tilde{\mathcal{D}})d\lambda = \mathbb{E}[u(\tilde{X})] \leq \mathbb{E}[u(X)] = -\int_0^\infty u(\lambda)\mathcal{V}'(\lambda, \mathcal{D})d\lambda.$$

In the first inequality, we use $u(\cdot) = \mathcal{E}(\cdot, \tilde{\mathcal{D}})$, and in the second we use $u(\cdot) = \mathcal{E}(\cdot, \mathcal{D})$. Both functions are concave and increasing by Lemma 2. □

Having determined the worst-case distribution $\mathcal{D}$ for a given $\mathcal{T}$, we now determine the worst-case type space $\mathcal{T}$ satisfying the constraints that all schools have equal capacity and no school is more than r times as popular as any other. Given $r \geq 1$ and $C \in \mathbb{N}$, define

$$T(r, C) = \{\mathcal{T} : C_\tau = C \text{ for all } \tau \in \mathcal{T}, \text{ and } \underline{p} \leq 1 \leq \overline{p} \leq r\underline{p}\}. \tag{48}$$

$$LBT(r, C) = \inf_{\mathcal{T} \in T(r,C)} LB(\mathcal{T}, \mathcal{D}^*(\mathcal{T})). \tag{49}$$

The following Lemma completes the proof of Proposition 3. It shows that the worst case can be achieved by a $\mathcal{T}$ for which some schools have popularity $\sqrt{r}$ and others have popularity $1/\sqrt{r}$.

Lemma 12. *Fix $r \geq 1$ and $C \in \mathbb{N}$, and define $\mathcal{T}^* = \{(\sqrt{r}, C), (1/\sqrt{r}, C)\} \in T(r, C)$. Then*

$$LBT(r, C) = LB(\mathcal{T}^*, \mathcal{D}^*(\mathcal{T}^*)), \tag{50}$$

from which it follows that

$$LBT(r, 1) = \frac{\sqrt{r}}{1 + r}, \qquad \lim_{C \to \infty} LBT(r, C) = \frac{2}{1 + \sqrt{r}}. \tag{51}$$

Proof. For any $\mathcal{T} \in T(r, C)$ it follows from Lemma 2 and (8) that

$$LB(\mathcal{T}, \mathcal{D}^*(\mathcal{T})) = \frac{1}{C}\int_0^\infty \mathcal{V}(\lambda)^2 d\lambda = \frac{1}{C}\int \left(\beta\overline{p}\nu(\overline{p}\lambda, C) + (1-\beta)\underline{p}\nu(\underline{p}\lambda, C)\right)^2 d\lambda.$$

Expanding this expression yields

$$\frac{1}{C}\left(\beta^2\overline{p}^2\int \nu(\overline{p}\lambda, C)^2 d\lambda + 2\beta(1-\beta)\overline{p}\underline{p}\int \nu(\overline{p}\lambda, C)\nu(\underline{p}\lambda, C)d\lambda + (1-\beta)^2\underline{p}^2\int \nu(\underline{p}\lambda, C)^2 d\lambda\right).$$

Applying the u-substitution $\lambda = Cx/\overline{p}$ to the first two integrals and $\lambda = Cx/\underline{p}$ to the third, we get

$$LB(\mathcal{T}, \mathcal{D}^*(\mathcal{T})) = \left(\beta^2\overline{p} + (1-\beta)^2\underline{p}\right)\int_0^\infty \nu(Cx, C)^2 dx + 2\beta(1-\beta)\underline{p}\int_0^\infty \nu(Cx, C)\nu(Cx\underline{p}/\overline{p}, C)dx. \tag{52}$$

To find $LBT(r, C)$ we must choose $\underline{p}$ and $\overline{p}$ to minimize this expression. The argument used in Lemma 11 implies that $LB(\mathcal{T}, \mathcal{D}^*(\mathcal{T}))$ is decreasing in $\overline{p}$ for fixed $\underline{p}$, so the worst case $\mathcal{T} \in T(r, C)$ satisfies $\overline{p} = r\underline{p}$. Combining this with the equation $\beta\overline{p} + (1-\beta)\underline{p} = 1$ that defines β in Lemma 11 and solving for $\overline{p}$, we get

$$\overline{p} = \frac{r}{\beta r + 1 - \beta} \tag{53}$$

We find $LBT(r, C)$ by minimizing over $\overline{p}$, or equivalently over β. We adopt the latter formulation. Define

$$\gamma_1(\beta, r) = \beta^2\overline{p} + (1-\beta)^2\overline{p}/r = \frac{\beta^2 r + (1-\beta)^2}{\beta r + 1 - \beta}, \tag{54}$$

$$\gamma_2(\beta, r) = 2\beta(1-\beta)\overline{p}/r = \frac{2\beta(1-\beta)}{\beta r + 1 - \beta}, \tag{55}$$

$$h(\beta, r, C) = \gamma_1(\beta, r)\int_0^\infty \nu(Cx, C)^2 dx + \gamma_2(\beta, r)\int_0^\infty \nu(Cx, C)\nu(Cx/r, C)dx. \tag{56}$$

Then by (52),

$$LBT(r, C) = \min_{\beta\in[0,1]} h(\beta, r, C) \tag{57}$$

We fix r and C, and drop dependence of γ_1, γ_2, h on these parameters. Because h is differentiable, its minimum is either at $\beta \in \{0, 1\}$ or a solution to $h'(\beta) = 0$. The value at the end points is

$$h(0) = h(1) = \int_0^\infty \nu(Cx, C)^2 dx.$$

We now seek solutions to $h'(\beta) = 0$. To simplify the calculations, we note that (54) and (55) imply

$$\gamma_2(\beta) = \frac{2}{1+r}(1 - \gamma_1(\beta)). \tag{58}$$

From this and (56), it follows that

$$h'(\beta) = \gamma_1'(\beta)\left(\int_0^\infty \nu(Cx, C)^2 dx - \frac{2}{1+r}\int_0^\infty \nu(Cx, C)\nu(Cx/r, C)dx\right) \tag{59}$$

If the term in parentheses is nonzero, this implies that $h'(\beta) = 0$ if and only if $\gamma_1'(\beta) = 0$, which by (58) occurs if and only if $\gamma_2'(\beta) = 0$ (if the term in parentheses is zero, then h is constant). But

$$\gamma_2'(\beta) = -2\frac{(r-1)\beta^2 + 2\beta - 1}{((r-1)\beta + 1)^2},$$

from which it follows that the only solution to $\gamma_2'(\beta) = 0$ on $[0, 1]$ is $\beta = \frac{1}{1+\sqrt{r}}$, which corresponds to $\overline{p} = \sqrt{r}$. One can verify that indeed this is a minimum, not a maximum (that is, $h(\frac{1}{1+\sqrt{r}})$ is less than $h(0) = h(1)$). This establishes (50). From this and (52), it follows that

$$LBT(r, C) = \frac{2}{(1+\sqrt{r})^2}\int_0^\infty \sqrt{r}\cdot\nu(Cx, C)^2 + \nu(Cx, C)\nu(Cx/r, C)dx.$$

For $C = 1$, $\nu(Cx, C) = e^{-x}$, and the expression above evaluates to $\sqrt{r}/(1 + r)$.

As C grows, $\nu(Cx, C)$ converges pointwise to $g(x) = \mathbf{1}(x < 1)$. Because $\nu(Cx, C)$ is pointwise less than e^{1-x}, by the dominated convergence theorem we have

$$\lim_{C\to\infty} LBT(r, C) = \frac{2}{(1+\sqrt{r})^2}\int_0^\infty \sqrt{r}\cdot g(x)^2 + g(x)g(x/r)dx = \frac{2}{1+\sqrt{r}},$$

where the final equality follows because for $r \geq 1$, $g(x)^2 = g(x) = g(x)g(x/r)$ and therefore the integral above is equal to $1 + \sqrt{r}$. $\square$