Simultaneous Search: Beyond Independent Successes*

Ran I. Shorrer
Penn State

August 26, 2019

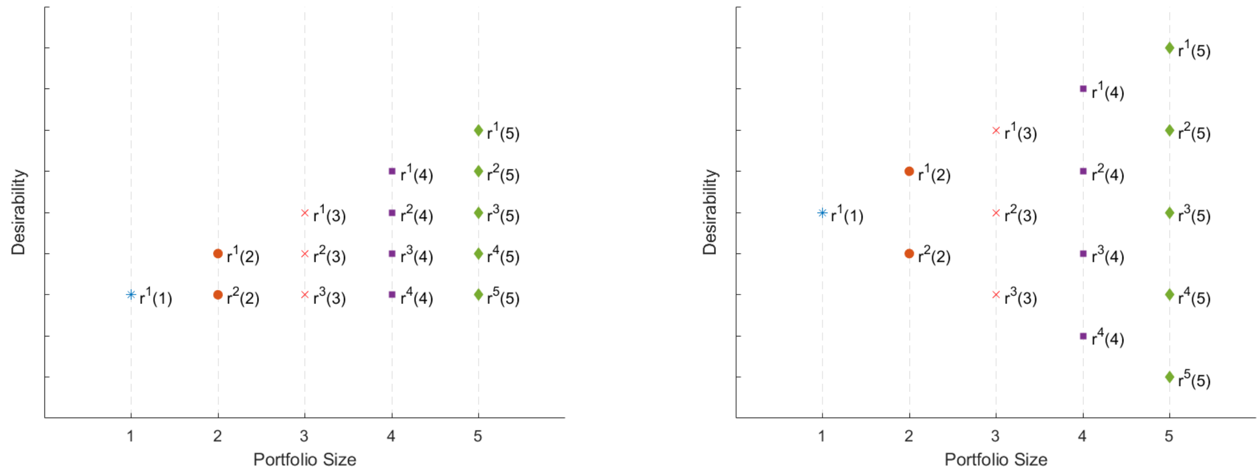
Abstract

A key decision commonly faced by students is how to optimally choose their portfolio of college applications. Students are often advised to apply to a combination of "reach," "match," and "safety" schools. Empirically, when reductions in the cost of application permit students to apply to more schools, they expand the range of selectivity of schools to which they apply both upwards and downwards. However, this ubiquitous practice of diversification is difficult to reconcile with existing theoretical analyses of search decisions, which assume that schools' admission decisions are independent conditional on the student's information. I develop a framework for simultaneous search problems that relaxes this assumption, and generates these patterns. My framework shows that popular school allocation mechanisms—arguably designed to level the playing field and promote integration—may generate segregation endogenously, as they provide students with better outside options stronger incentives to apply to selective schools.

When applying to schools and colleges, a key decision commonly faced by students is how to optimally choose their portfolio of applications. In many cases, large numbers of schools are available, but due to costs or constraints, students apply only to a few, often without perfect information about how the school will respond to their application. Determining which subset of schools to apply to—balancing the desire to attend sought-after schools with the need to hedge—forms a critical part of a student's decision problem. To achieve this balance, students are often advised to apply to a combination of "reach," "match," and "safety" schools (Avery, Howell and Page, 2014). In practice, one sees that when reductions in application costs permit students to apply to more schools, they expand the range of schools to which they apply *both* upwards and downwards, i.e., including more selective and safer schools (Ajayi, 2011; Pallais, 2015).

Although it is well-understood that applying to a diverse set of colleges is crucial for students' success and carries significant implications, this practice of diversification is difficult to reconcile with existing theoretical analyses of search decisions. In a leading analysis, Chade and Smith (2006)

*I am grateful to Kehinde Ajayi who introduced me to this problem. I also thank Mohammed Akbarpour, Nageeb Ali, Nick Arnosti, Kalyan Chatterjee, Oren Danieli, Henrique de Oliveira, Pieter Gautier, Michael Gechter, Ed Green, Nima Haghpanah, Avinatan Hassidim, Sándor Sóvágó, Kala Krishna, Vijay Krishna, Tomás Larroucau, Jacob Leshno, Dilip Mookherjee, Amanda Pallais, Alex Rees-Jones, Ignacio Rios, Assaf Romm, Karl Schurter, Shouyong Shi, Andy Skrzypacz, and Ricky Vohra. Nilufer Gok and Xiao Lin provided excellent research assistance. This research was supported by the United States–Israel Binational Science Foundation (BSF), Jerusalem, Israel (grant 2016015).



(a) Independent admission decisions

(b) Perfectly aligned admission decisions

Figure 1: Optimal portfolios under independent and perfectly aligned admission decisions

Panel 1b illustrates my result that when admission decisions are perfectly aligned, optimal portfolios of larger sizes span a wider range of schools and refine the grid. Panel 1a illustrates the findings of Chade and Smith (2006), which I slightly strengthen in Theorem 4: i) optimal portfolios of size $k+1$ nest the optimal portfolios of size k, and ii) a school that appears in the optimal portfolio of size $k+1$ but not in the optimal portfolio of size k is more aggressive than any school in the optimal portfolio of size k (when each school has close substitutes).

show that students should not apply to "safety" schools. Furthermore, they show that if application costs decrease, students should expand the range of schools only in an upward direction, by applying to the same schools and to additional schools of higher selectivity (as illustrated in Panel 1a of Figure 1). The predictions emphasized by Chade and Smith (2006) rely on the common assumption—used for both theoretical and empirical studies—that admission decisions are independent conditional on the information known to the applicant.[1] This assumption means that admission decisions at some schools convey no information about admissions decision at other schools (e.g., finding out that one is rejected at Beauxbatons is *not* bad news with respect to the probability of acceptance at Hogwarts). In Section 1, I show that this lack of consistency between theory and evidence may carry important implications for inferences that rely on such theoretical models, and can lead to misguided policies.

This paper develops a new analysis of simultaneous search in which the optimal behavior involves diversifying and expanding the range of schools one applies to when application costs decrease (as illustrated in Panel 1b of Figure 1). The important departure that I make from existing models is that schools' admission decisions are not independent. Instead, I assume that admissions decisions are perfectly aligned—conditional on the student's information, admissions are based on a common

[1]Apart from Chade and Smith (2006), examples of other studies that make the common assumption that schools' admission decisions are independent conditional on the information known to the student include Card and Krueger (2005), Chade, Lewis and Smith (2014), and Fu (2014).

index with different schools having different bars for admission, and students face uncertainty about their index value. As a result, admission decisions at some schools do convey information about the probability of acceptance at other schools (e.g., finding out that one is rejected at Beauxbatons is bad news with respect to the probability of acceptance at Hogwarts).

The assumption that admissions decisions are perfectly aligned is a stark one. However, it accurately describes many empirically relevant settings. For example, admissions decisions are perfectly aligned when a future centralized admission exam will determine admissions, which is a common practice for high-school and college admissions across the world. Similarly, admissions decisions are perfectly aligned in school districts in the U.S. and around the world where a single lottery is used to break priority ties in overdemanded schools (i.e, ones with more applicants than seats. See, e.g., De Haan et al., 2015; Abdulkadiroğlu et al., 2017). More broadly, my model captures potentially important features of environments with high degree of correlation, even if admission decisions are not perfectly aligned. Such environments are ubiquitous. For example, even in the absence of a centralized admissions exam, correlation in admission decisions is present when college applicants think that colleges seek students with certain characteristics, but applicants face uncertainty about where they stand relative to others.[2]

While I focus on students applying to colleges, *simultaneous search* problems—portfolio choice problems in which an agent chooses a portfolio of stochastic options but only consumes the best realized one—occur in many other settings (e.g., consumer search, De Los Santos, Hortaçsu and Wildenbeest, 2012; labor markets, Galenianos and Kircher, 2009; and industrial organization, Wong, 2014). In labor markets, for example, correlation arises when common factors affect all firms' hiring decisions (Lee and Wang, 2018). Similarly, a monopolist may face consumers who are vertically differentiated in their tastes for quality, where introducing new varieties is costly (Wong, 2014).

My analysis begins with the following observation: Students only attend less desirable schools if they cannot attend more desirable ones. As a result, application portfolios can be represented by a Rank-Order List (ROL), sorted according to the student's preferences, where the probability that the k-th-ranked option on the ROL is consumed depends only on higher-ranked options (specifically, the probability that they are consumed).[3]

Building upon this observation, I show that these portfolio choice problems can be solved by dynamic programming—a concept similar to backwards induction. The idea is that the optimal continuation (or suffix) of an ROL can be calculated by conditioning on the event that the agent will be rejected by all the options that are ranked higher on the ROL (prefix). Moreover, since the optimal suffix is identical for many prefixes, the optimal ROL can be found "quickly," in a running time that is polynomial in the number of options. This can be useful for empirical work: in

[2]Many studies of application patterns to American colleges use SAT test-score sending as a proxy for application (e.g., Card and Krueger, 2005). An important institutional detail is that *before* taking the exam applicants choose up to four schools that will receive their scores (they have the option to add schools to the list before and after the exam for a fee). This feature suggests that students, especially first-time test takers, may be facing substantial uncertainty about their strength as applicants when determining their application portfolio.

[3]In addition to decentralized markets, many popular centralized mechanisms have this property with respect to the ROL students are required to submit. Examples include deferred acceptance (Gale and Shapley, 1962), serial dictatorship (Abdulkadiroğlu and Sönmez, 1998), and top trading cycles (Abdulkadiroğlu and Sönmez, 2003).

the absence of this solution, existing empirical work has had to use approximations of the optimal portfolio (e.g., Ajayi and Sidibe, 2015).[4]

I show that allocation mechanisms that are commonly used in centralized school choice and that constrain the number of schools each student can apply to may generate segregation *endogenously.* This is the case even if all applicants share the same preferences, and in spite of the mechanism treating all applicants symmetrically (cf. Hastings, Kane and Staiger, 2009; Oosterbeek, Sóvágó and van der Klaauw, 2019). The reason is that these mechanisms, which include constrained deferred acceptance (Haeringer and Klijn, 2009; Calsamiglia, Haeringer and Klijn, 2010), give applicants with better outside options (e.g., access to private schools) incentives to apply more aggressively. As a result, applicants with stronger outside options are over-represented in the best schools.[5] I also show that students who are less risk averse and students with more optimistic beliefs about their admission chances apply more aggressively. This suggests that search frictions that are present in popular allocation mechanisms may introduce disparities across gender lines (Dargnies, Hakimov and Kübler, forthcoming; Pan, 2019).

The effect of better outside options explains the upward expansion of optimal portfolios. The optimal portfolio of size $k+1$ can be thought of as the optimal portfolio of size k where the student's outside option is the hope of being accepted to her $k+1$-st-ranked school (rather than nothing, in the case of choosing the optimal size-k portfolio).

The effect of higher beliefs explains the downward expansion of optimal portfolios. Recall that admission to less desirable schools only matters when a student is rejected by the most desirable school in his portfolio. Thus, one can think of the optimal portfolio of size $k+1$ as consisting of the most desirable school in the portfolio and the optimal size-k continuation portfolio conditional on being rejected from this school. When admission decisions are aligned, news of rejection lead to lower beliefs, so this continuation portfolio is less aggressive than the optimal portfolio of size k.

While the explanation for the upward expansion applies equally to the cases of perfectly aligned and independent admissions, optimal portfolios do not expand downwards when admission decisions are independent. The divergence in predictions between the two environments stems from the fact that when admissions decisions are independent, a rejection by one's first choice (or any other school) carries no information about admission chances at other schools. Thus, although more optimistic students apply more aggressively in this environment too (Chade and Smith, 2006), being rejected by one's first choice should not make one more pessimistic.

[4]In the context of centralized school choice, Calsamiglia, Fu and Güell (forthcoming) make this observation and use it to calculate optimal ROLs in their empirical study of school choice in Barcelona, where a variant of the Immediate Acceptance (Boston) mechanism is in place. In Shorrer (2019) I show that this approach is more broadly applicable as the conditions above are satisfied in many empirically relevant portfolio choice problems.

[5]Delacroix and Shi (2006) derive similar results in a model of on-the-job search, where workers differ in their current jobs, which they can hold on to. In Shorrer (2019) I show that the resutls on outside options (as well as those on risk aversion and on beliefs) generalize to many other environments, including immidiate acceptance mechanisms. Calsamiglia, Martínez-Mora and Miralles (2015) and Akbarpour and van Dijk (2018) find that when immediate acceptance is in place, students with better outside options are over represented in the best school.

1 Motivating Example

The following example illustrates how agents should reason about their portfolio of applications when admissions decisions are perfectly aligned, and then illustrates the consequences for an analyst of ignoring the resulting correlation in admission decisions.

Example 1. A school district makes admissions decisions based on a single lottery, where higher numbers have higher priority. There is a unit mass of students, half of whom reside in the East Neighborhood, and the others in the West Neighborhood. After applying to schools, students in both neighborhoods draw lottery numbers uniformly from the interval $[0, 1]$.

In each location, $x \in \{\text{East, West}\}$, there are two schools: A good school, g_x, with capacity $\frac{1}{4}$, and a bad school, b_x, with unlimited capacity. Students prefer closer and better schools. Specifically, the utility of student s from attending school m is $u_s(m) := \beta \mathbb{1}_{m \text{ is local for } s} + \gamma \mathbb{1}_{m \text{ is good}}$, where β and γ are greater than zero, and where the utility from the outside option is normalized to zero. Students can only apply to two schools. The above facts are all commonly known.

In equilibrium, each student applies to both of the schools in her neighborhood. To see this, note that under this profile of strategies a student is admitted to a good school if her lottery score is greater than one half. Thus, if she can be admitted to the good school in the other neighborhood, she will also be admitted to the good school in her own neighborhood, which she prefers. Consequentially, students prefer to apply only to the good local school and to use their second application to guarantee admission to the bad local school, in case they are not admitted to the good school.

Next, imagine an analyst who believes that schools admission decisions are independent. Except for this fact, the analyst's model is correctly specified, and he believes that students have rational expectations and that they choose their applications portfolio optimally. Observing students' applications and admission probabilities the analyst concludes that

$$\frac{1}{2} \cdot (\beta + \gamma) + \frac{1}{2} \cdot 1 \cdot \beta \geq \frac{1}{2} \cdot (\beta + \gamma) + \frac{1}{2} \cdot \frac{1}{2} \cdot \gamma,$$

where the left hand side represents his (correct) perception of students' expected utility from the portfolios they choose and the right hand side is his (mistaken) perception of the expected utility from applying to both of the good schools (see Agarwal and Somaini, 2018). This inequality implies that parameters in the identified set satisfy the inequality

$$2\beta \geq \gamma.$$

Thus, if students attribute high relative importance to quality (e.g., $\beta = 1$, $\gamma = 100$) this preference will not be reflected in the analyst's estimates. A policy maker that relies on the analyst's estimates may make misguided choices, such as focusing on reducing travel distance instead of prioritizing school quality.

2 Model

I now turn to the decision problem of interest: simultaneous search when admissions decisions are perfectly aligned and a student gets to attend the best school to which she applied whose admission cutoff she surpassed. This arises, for example, when applicants are vertically differentiated and they face uncertainty about their market position. A common case is when schools make admission decisions based on a centralized entrance exam (or central lottery) where admission to each school depends on passing a school-specific (agent-specific) score cutoff that is known to the agent,[6] and students are required to make their application decision not knowing their score on the admission exam.

There is a set, X, of stochastic *options* (also referred to as *schools* or *colleges*). Unless otherwise mentioned, the set is finite and the options are indexed by the integers 1 to N. Agents (also referred to as *students* or *applicants*) are expected-utility maximizers. They can attend at most one school, and for each school i they know the admission cutoff, c_i, and the utility they will derive from placement in that school, u_i.[7] I envision that the number of students is large, and consider the problem of a single student. I assume, without loss of generality, that lower integers (schools) are (weakly) more desired by this student. The utility from being unassigned is normalized to zero.

A Rank-Order List (ROL) is an ordered list of schools. For an ROL, r, let r^l denote the l-th-highest-ranked school on r. I assume that optimal application portfolios can be summarized by an ROL. This is the case when either: 1) a student applies to a subset of X and chooses the most preferred school that admits her, or 2) a student submits an ROL of alternatives in X to a centralized mechanism that assigns her to the highest ranked alternative whose admission cutoff she surpasses. The model, therefore, also captures centralized markets that are cleared using algorithms such as single-lottery deferred acceptance, random serial dictatorship, and top trading cycles,[8] as well as markets such as British college admissions, where students can only rank a limited number of admission offers that are contingent on obtaining a minimal score (Broecke, 2012).

The cost associated with a portfolio of size k is $C(k)$, where $C(0) = 0$ and $C(\cdot)$ is increasing. I sometimes further assume that $C(\cdot)$ is convex. Special cases include constrained choice ($C(x) = 0$ if $x \leq k$, and $C(x) = \infty$ otherwise) and constant marginal cost ($C(x) = cx$).

Without loss of generality, I make the following assumptions: 1) The student's belief about her admission score is an atomless distribution over all possible scores. Furthermore, I assume that scores are distributed uniformly between 0 and 1 (otherwise, apply the probability distribution transform to all scores and admission cutoffs). 2) Every school is preferred to the outside option whose value is normalized to 0.

Although in the environment I study there exists, generically, a unique optimal portfolio, I

[6]In the absence of aggregate uncertainty (i.e., correlation in the distribution of types in the population), when schools offer the same number of seats and the market is large, it is safe to assume that admission cutoffs are stable from one year to the next (Azevedo and Leshno, 2016).

[7]The assumption that agents know their utility from prizes is common, but there are some exceptions (e.g., Immorlica et al., 2018; Albrecht, Gautier and Vroman, 2006).

[8]Schools admissions decisions are perfectly aligned when markets are cleared by top trading cycles, regardless of the randomization method used to break priority ties (Che and Tercieux, 2017; Leshno and Lo, 2017).

choose to treat the general case and do not rule out the possibility that agents have multiple optimal portfolios. The reason is that, as Example 1 illustrates, my results hold in equilibrium in large markets, and there is no reason to think that agents have unique optimal portfolios in equilibrium. This comes at the cost that statements and proofs are slightly more cumbersome.

3 The Portfolio Choice Problem

The portfolio choice problem can be summarized by the following equation:

$$\max_{r} \sum_{1 \leq i \leq |r|} \max \left\{ 0, \min_{0 \leq j < i} \{c_{r^j}\} - c_{r^i} \right\} u\left(r^i\right) - C\left(|r|\right),$$

where $c_{r^0} \equiv 1$, and $\min\limits_{j<i} \{c_{r^j}\}$ is the i-th Most Informative Disqualification (MID, Abdulkadiroğlu et al., 2017)—the most forgiving cutoff the applicant failed to pass if the i-th-ranked alternative becomes relevant—which is a sufficient statistic for updating one's beliefs about her priority score in this event.

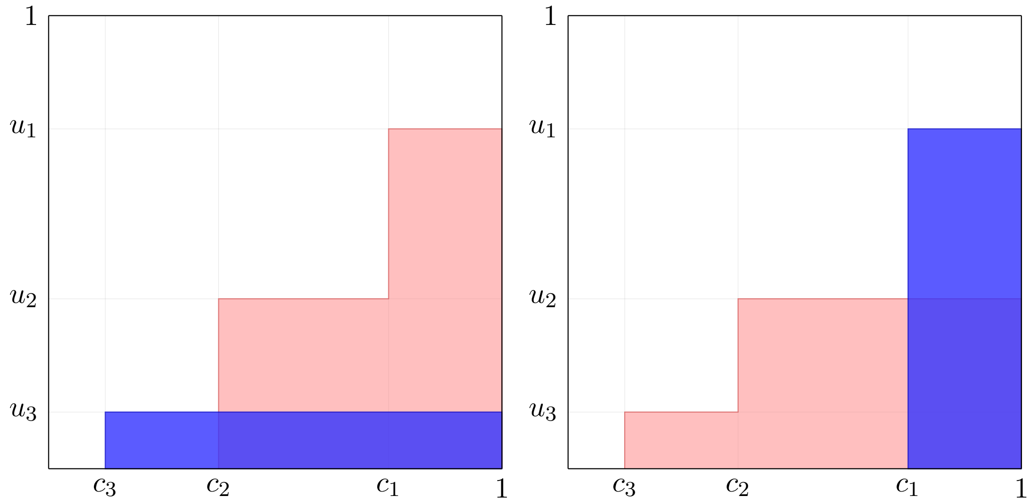
This portfolio choice problem can also be represented as an instance of the well-studied NP-hard max-coverage problem (Hochbaum, 1996). Recall that I assume, without loss of generality, that scores are distributed uniformly on the unit interval. Thus, the probability of admission to School i is equal to the distance between c_i and 1. As a result, the expected utility from applying to School i only is equal to the area of the rectangle that has vertices at $(c_i, 0)$ and $(1, u_i)$, as illustrated in Figure 2. Furthermore, the expected utility from a portfolio of k schools is equal to the area covered by the union of the k corresponding rectangles, as illustrated in Figure 3.

This formulation of the problem already gives intuition for the incentive to apply to a diverse set of colleges in terms of selectivity, as recommended by the College Board (Avery, Howell and Page, 2014): the rectangles corresponding to schools of similar selectivity (similar cutoffs) have a large overlap, so choosing several of them does not increase the covered area substantially relative to choosing just one of them (the most desirable one).[9] For example, consider a portfolio choice problem as depicted in Figure 2, where each rectangle covers approximately the same area, but the one corresponding to School 2 is slightly larger. As a result, the optimal size-1 portfolio consists of School 2, but the optimal size-2 portfolio consists of Schools 1 and 3, as the large overlap of other rectangles with the rectangle corresponding to School 2 dwarfs the size advantage of this rectangle (see Example 2 in the supplementary Appendix).

This example also shows that the "greedy" approach, used by Chade and Smith (2006) to solve for the optimal portfolio, cannot be used in the current setting.[10] To see this, note that a greedy algorithm identifies an optimal portfolio of any size only when the optimal portfolios are nested.

[9] The economic intuition resembles that of Athey and Ellison (2011) in the context of position auctions with consumer search. Athey and Ellison (2011) study the cases where the probabilities that websites meet the needs of a consumer are independent, and where similar sellers are likely to meet the same needs. They find that in the latter case the optimal auction incentivizes product variety.

[10] The same holds for the generalized environments studied in Chade and Smith (2005) and Olszewski and Vohra (2016).



(a) "safety" school

(b) "reach" school

Figure 2: The utility from different size-1 portfolios
This figure assumes that the distribution of scores is uniform. As a result, the probability of acceptance to school i is equal to $1 - c_i$, the probability of passing this school's admission cutoff. This quantity is equal to the distance between c_i and 1 on the horizontal axis. Thus, the shaded areas represent the expected utility from a portfolio consisting of a single school (School 3 in Panel 2a and School 1 in Panel 2b): the probability of admission, $1 - c_i$, times the utility from attending, u_i. School 3 is a "safety" school—it has a high probability of admission, but yields a low utility from attending. School 1 is a "reach" school—it has a low probability of admission, but yields a high utility from attending.

Therefore, an alternative approach is needed to identify optimal portfolios.

A first observation is that when looking for an optimal portfolio, there is no loss in ignoring schools such that $i > j$ and $c_j \leq c_i$ for some j. Graphically, these are schools whose rectangles in the corresponding coverage problem are covered by another rectangle. To keep the notation simple, I assume that all such schools have been removed from the menu, and there are $M \leq N$ schools such that $c_i > c_j \iff u_i > u_j$. Still, in what follows I use N, not M, to denote the number of schools.

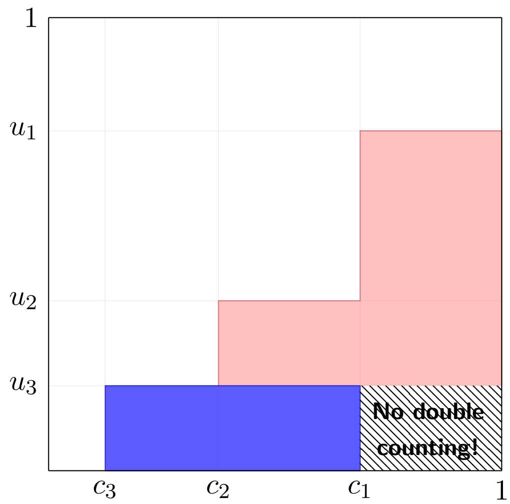
Lemma 1. *There always exists an optimal ROL that is strictly decreasing in selectivity and in desirability (i.e., lower-ranked schools have higher indices and strictly lower admission cutoffs).*

Proof. An ROL that is inconsistent with one's preferences (i.e., not monotonic in indices) is weakly dominated (e.g., Haeringer and Klijn, 2009). The result follows from this fact and the discussion in the paragraph above.

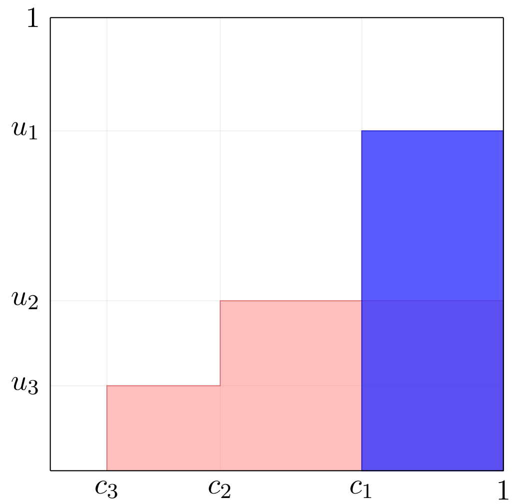
□

The portfolio choice problem can thus be simplified to

$$\max_{r | c_{r^{i-1}} > c_{r^i}} \sum_{1 \leq i \leq |r|} \left(c_{r^{i-1}} - c_{r^i}\right) u(r^i) - C(|r|).$$



(a) "safety" school



(b) "reach" school

Figure 3: The utility from a size-2 portfolio

This figure illustrates the expected utility from applying to both School 1 and School 3. It highlights the fact that a student can only attend a single school. As a result, when she attends the more selective School 1, she does not derive further utility from being also admitted to School 3, which she likes less. This is illustrated by the area missing from the "safety school" rectangle (labeled "No double counting!"). This rectangle corresponds to cases where the student is admitted to both schools, and thus attends the more preferred "reach" school.

3.1 Unconstrained Choice

I begin by considering the case of unconstrained choice ($C \equiv 0$). In this case, agents clearly have a weakly dominant strategy of applying everywhere and attending the best school that accepts them (i.e., using an ROL of all the alternatives in order of preference). The utility an agent derives from this strategy is equal to the area of the union of all rectangles.

An applicant is *effectively unconstrained* if the marginal cost of applying to an additional school at some optimal portfolio is zero. In what follows, I assume that applicants are effectively constrained.

3.2 Constrained Choice

In a constrained choice problem, the applicant may apply only to a limited number of schools. To gain intuition, I start by considering the special case of constrained choice where the constraint is extreme: the agent can choose at most one school to apply to ($C(x) = 0$ if $x \leq 1$, and $C(x) = \infty$ otherwise). In this case, the optimal strategy is clear: choose the school that maximizes $(1 - c_i)\, u(i)$. Graphically, in Figure 2, this would correspond to the largest rectangle.

The agent faces a trade-off between high ex-post utility and high ex-ante admission probability. This is illustrated by comparing the different size-1 portfolios depicted in Figure 2. The rectangle depicted in Panel 2a represents an application to a "safety school," School 3, with a high admission probability and low ex-post utility from consumption. The rectangle depicted in Panel 2b represents an application to the more selective School 1, which yields a higher ex-post utility from consumption, but entails more risk in terms of the probability of admission.

Next, we consider the optimal ROL of length k (which corresponds to the solution in the case where $C(x) = 0$ if $x \leq k$, and $C(x) = \infty$ otherwise). Graphically, the problem corresponds to the coverage problem of identifying k rectangles as above whose union covers the largest area. The challenge is that the rectangles overlap, but their intersections—representing cases where the student is admitted to more than one school—should not be double counted (Figure 3). When rectangles can be placed arbitrarily in the space, this problem is NP-hard (e.g., Hochbaum, 1996).

To address this challenge, I propose the Probabilistically Sophisticated Algorithm (PSA), which uses dynamic programming to solve for an optimal ROL in time polynomial in N. I denote by $r(k,c)$ an optimal size-k ROL when the score is uniformly distributed in $[0,c]$, and let $v(k,c)$ denote the corresponding expected utility. Furthermore, $r_F(k,c)$ and $v_F(k,c)$ are similarly defined, where the score is distributed according to F conditional on it being lower than c. I use the convention that $v(0,\cdot) = c_{r(0,\cdot)} \equiv 0$.

The success of the algorithm hinges on the following lemma.

Lemma 2. *For any integer, k, there exists optimal ROLs such that $r^1(k+1,c) \in \arg\max_{i|c_i \leq c} \frac{1}{c}\{(c-c_i)u_i + c_i v(k,c_i)\}$ and for all $k+1 \geq j > 1$, $r^j(k+1,c) = r^{j-1}(k, c_{r^1(k+1,c)})$.*[11] *Furthermore, for any integer, k, and any distribution of scores, F, there exists optimal ROLs such that $r^1_F(k+1,c) \in \arg\max_{i|c_i \leq c} \frac{1}{F(c)}\{(F(c)-F(c_i))u_i + F(c_i)v_F(k,c_i)\}$, and, for all $k+1 \geq j > 1$, $r^j_F(k+1,c) = r^{j-1}_F(k, c_{r^1_F(k+1,c)})$.*

Algorithm 1 Probabilistically Sophisticated Algorithm (PSA)

Step 1 For all possible values of r^{k-1} calculate $r^1(1, c_{r^{k-1}})$ and $v(1, c_{r^{k-1}})$.

Step j<k For all possible values of r^{k-j} calculate $r^1(j, c_{r^{k-j}})$ and $v(j, c_{r^{k-j}})$, as in Lemma 2.

Step k Calculate $r^1(k,1)$, and set $r^1 = r^1(k,1)$, $r^i = r^1(k+1-i, c_{r^{i-1}})$.

Remark 1. For ease of exposition, the algorithm does not choose the optimal continuation in cases where the prefix is not monotonic in selectivity (this would require conditioning on the score being lower than the MID, rather than the preceding cutoff). These cases are (weakly) suboptimal (Lemma 1), and strictly so under the assumption that the agent is effectively constrained.

Remark 2. For ease of exposition, the algorithm calculates a single optimal ROL. In case that there are multiple optimal ROLs, there will be ties in certain steps of the algorithm. Standard arguments show that in such situations any tie-breaking rule will lead to an optimal ROL, and any optimal ROL will result from some tie-breaking rule.

[11] The set $\{i|c_i \leq c\}$may be empty. In this case, I use the the convention that the arg max is the empty set and the ROL will have an empty enrty. This does not occur on the run of the algorithm (it will repeat the same option). In any case, such choices are inconsequential for the agent's assignment.

3.3 Costly Choice

Each step of the PSA requires no more than N^2 steps. Thus, it is possible to find an optimal portfolio of any size in less than N^3 steps. When portfolio costs differ in size, one can solve for the optimal portfolio of each size and then choose the best one by accounting for costs.

When the cost is convex, one can avoid some computation by leveraging the fact that the marginal benefit from increasing the size of the portfolio (k) is decreasing. Although intuitive, the decreasing marginal benefit property is not at all trivial due to the possibility that optimal portfolios of varying sizes are not nested.[12] I prove this property later (Theorem 2), as the proof invokes Theorem 1.

3.4 The Dual Choice Problem

In this section, I offer another perspective on the portfolio choice problem which proves useful in several of the proofs. The main insight of this section is that if one transposes a coverage problem over the diagonal connecting $(1,0)$ and $(0,1)$, one is looking at an equivalent coverage problem. For portfolio choice problems, the result is another portfolio choice problem, with the only difference being that higher indices correspond to higher rectangles (Figure 4).

Formally, consider the (primal) problem defined by $(c_i, u_i)_{i=1}^N$, the uniform distribution over the unit interval, and some cost function. First, if $\max\{u_i\}$ is greater than 1, multiply each u_i and the cost function by $\frac{1}{\max\{u_i\}}$. Note that there is no loss of generality from this normalization, which assures that the agent derives a utility of 1 from the alternative he likes best.

Next, for all i, let

$$\bar{c}_i = 1 - u_i$$

and

$$\bar{u}_i = 1 - c_i.$$

Since cutoffs lie in the unit interval, $\bar{u}_i \in [0,1]$ for all $1 \le i \le N$, and, thanks to the normalization of utilities, $\bar{c}_i \in [0,1]$ for all $1 \le i \le N$. The problem defined by $(\bar{c}_i, \bar{u}_i)_{i=1}^N$—with the same cost function and the uniform distribution—is the dual choice problem of the primal problem.

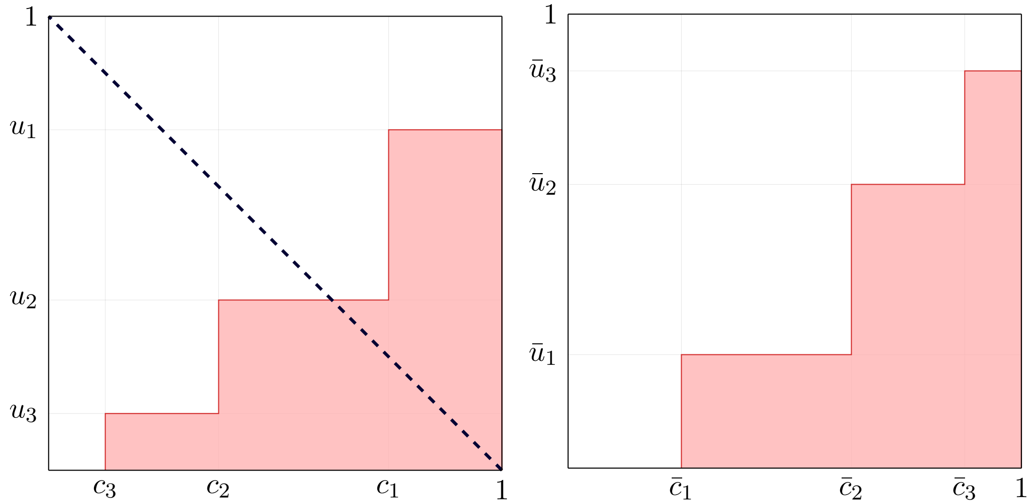
Definition. Given an ROL, r, of length k, let $\overset{\circlearrowright}{r}$ be the ROL such that $\overset{\circlearrowright}{r}_i = r_{k-i}$ for all i.

In words, $\overset{\circlearrowright}{r}$ is the ROL r turned upside down.

Proposition 1. *ROL r is a solution of a portfolio choice problem iff $\overset{\circlearrowright}{r}$ is a solution to the dual problem.*

Proof. Note first that in the dual problem, lower index schools are less desirable. Thus, since it is optimal to rank consistently with preferences, if the optimal portfolio consists of the same alternatives it is optimal to rank them in reverse order.

[12]Indeed, Example 3 in the supplementary Appendix provides an instance of the coverage problem by rectangles where this is not the case. The example does not, however, possess the structure of the problems studied in this paper, as the rectangles do not share a common bottom-right corner.



(a) primal problem

(b) dual problem

Figure 4: A problem and its dual problem

Panel 4a depicts a (primal) choice problem ($c_3 = .1$, $c_2 = .5$, $c_1 = .7$; $u_3 = .3$, $u_2 = .4$, $u_1 = .6$). Panel 4b depicts its dual choice problem ($\bar{c}_3 = .7$, $\bar{c}_2 = .6$, $\bar{c}_1 = .4$; $\bar{u}_3 = .9$, $\bar{u}_2 = .5$, $\bar{u}_1 = .3$).

To see that the same alternatives should be ranked in both problems, note that the corresponding coverage problem is isometric (the only difference is that the plane has been transposed). $\square$

Remark 3. I previously assumed that applicants are effectively constrained. Without this assumption (or the assumption that alternatives that are more selective and less desirable have been removed from the menu), the statement would not be precise, since the applicant might be able to rank undesirable alternatives in irrelevant positions on an optimal ROL. The precise statement would require that $\overset{\circlearrowright}{r}$ be sorted according to preferences.

Remark 4. The proposition suggests that solving the dual problem "backwards" is tantamount to solving the primal problem "forwards." A forward-looking version of the PSA, which is sometimes more convenient, is provided in the supplementary Appendix.

4 Monotonicity in Beliefs

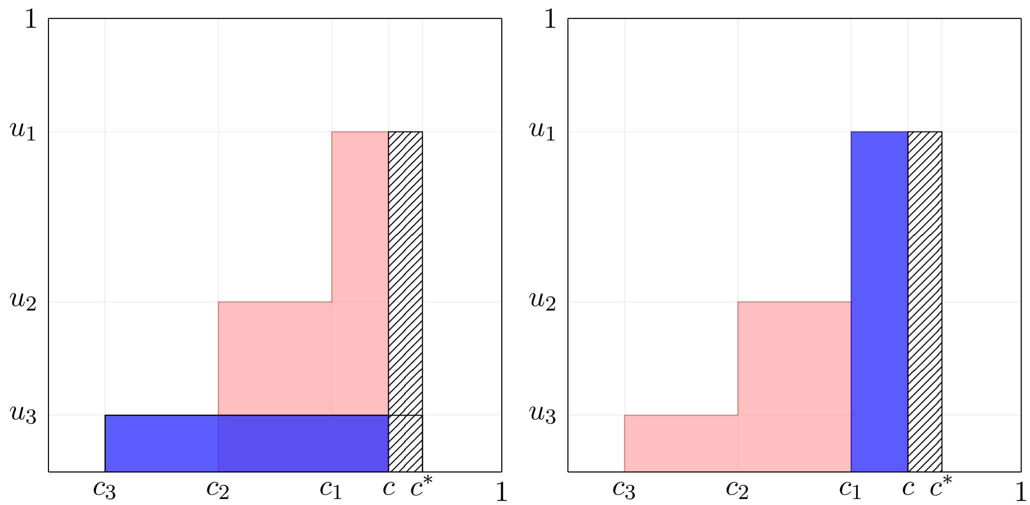
I now consider how the optimal size-k portfolio changes with the decision maker's beliefs. To this end, I dispense with the assumption that the distribution of scores is uniform.

Definition. Given two portfolios of size k, r and r', we say that r *at least as aggressive as* r' if $r^j \leq r'^j$ (r^j is at least as desirable as r'^j) for all $1 \leq j \leq k$.

Definition. Given two agents, s and b with beliefs f_s and f_b, s's beliefs about her score *MLRP-dominate* b's beliefs (or are *MLRP more optimistic*) if the ratio $\frac{f_s(x)}{f_b(x)}$ is weakly increasing in x on the union of the supports of f_s and f_b (where $\frac{c}{0} \equiv \infty$ for $c > 0$).

Proposition 2. *Let s and b be agents with identical preferences, but s is MLRP more optimistic. Then for all $k \leq N$, i) there exists an s-optimal size-k portfolio that is at least as aggressive as any b-optimal size-k portfolio, and ii) there exists a b-optimal size-k portfolio such that any s-optimal size-k portfolio is at least as aggressive.*

Figure 5 provides intuition in a special case. The figure shows how news that one's score is below a certain cutoff (which induces MLRP-lower beliefs) disproportionately decreases the expected utility from applications to more selective schools.



(a) "safety" school

(b) "reach" school

Figure 5: The effect of news of rejection

The figure illustrates the effect of negative news on one's score on the expected utility from size-1 portfolios. Specifically, it shows that being rejected by a school with cutoff c, which is less selective than c^*, affects the expected utility from (the more selective) School 1 (Panel 5b) more than it affects the expected utility from (the less selective) School 3 (Panel 5a).

Proof. I prove part i). The proof of part ii) is completely analogous. The proof is by induction on k. The case of $k = 1$ is obvious, because the MLRP shift to beliefs implies that admission chances to schools that yield higher ex-post utility are disproportionately higher. For $k > 1$, let r_s and r_b denote optimal portfolios for s and b, respectively.

Assume that the assertion is correct for all $l < k \leq N$. By Lemma 2 and the inductive hypothesis, if $r_s^1 \leq r_b^1$ we are done, since s's beliefs conditional on her being rejected from r_s^1, which is more selective than r_b^1, MLRP-dominate s's beliefs conditional on being rejected by r_b^1, and these beliefs MLRP-dominate b's beliefs conditional on b being rejected from r_b^1 by assumption.

To complete the proof, I rule out the case that $r_s^1 > r_b^1$ for all s-optimal size-k portfolios. Without loss of generality, assume that b's beliefs are distributed uniformly on the unit interval and that s's beliefs are given by the increasing probability distribution function f, and assume that

$r_s^1 > r_b^1$. Let j be the minimal index such that $r_s^j \leq r_b^j$ (if such index exists). Then $r_s^i \leq r_b^i$ for all $i \in \{j, ..., k\}$ by the inductive hypothesis and the argument in the previous paragraph.

A possible deviation for b is to drop r_b^1 from her portfolio, and to add the most desirable school on r_s not on r_b, while continuing to rank schools according to her true preferences (note that $r_s^1 = r_b^2$ is not excluded). Denote the resulting portfolio by $\bar{r}_b$. This deviation will cause her to lose $x > 0$ utils if the realized score is above $c_{r_b^1}$, and to gain $y > 0$ utils if the realized score is above the cutoff of the new addition, $c_{r_s^n}$, but lower than the cutoffs of all higher-ranked schools (i.e., lower than $c_{r_b^m} := \min\left\{c_{\bar{r}_b^i} | \bar{r}_b^i < r_s^n\right\}$). Since these higher-ranked schools are all ranked lower than r_b^1, their cutoffs are all lower than $c_{r_b^1}$. Since r_b is chosen by b over $\bar{r}_b$, $A_1 := \left(1 - c_{r_b^1}\right) \cdot x$ is weakly greater than $B_1 := \left(c_{r_b^m} - c_{r_s^n}\right) \cdot y$, where $\left(1 - c_{r_b^1}\right)$ and $\left(c_{r_b^m} - c_{r_s^n}\right)$ are the probabilities that the relevant events are realized under the assumption that beliefs are uniform.

If $\bar{r}_b^i \neq r_s^i$ for some $i < j$, the argument can be iterated. This time we get $A_1 + A_2 \geq B_1 + B_2$, where A_2 is the probability that the second change causes a utility loss times the magnitude of the loss, and B_2 is the probability of a gain times the magnitude of the gain (both are the incremental gains/losses after the first change having taken place). Clearly, scores contributing to B_2 are all lower than those contributing to A_2.

In general, for all $m < j$, we get $\sum_{i=1}^{m} A_i \geq \sum_{i=1}^{m} B_i$ and scores that contribute to B_m are all lower than those that contribute to A_m. Of note, the inequalities do not necessarily hold for each i separately. That is, B_2 may be higher than A_2, but the difference must be smaller than the difference between A_1 and B_1.

Now, consider s's deviation of changing her ROL to equal r_b^i for all $i < j$. For simplicity, assume first that $r_s^j = r_b^j$. The gain/loss score ranges are the same and so are the magnitudes of gain/loss, but the probabilities that the realized score is in these ranges change. Denote by A_i^f and B_i^f the corresponding quantities for s, and let f_i denote the value of f at the lowest score in the range associated with A_i (and A_i^f). This score is at least as high as the highest score in the range associated with B_i (and B_i^f); thus, since f is increasing, we get $A_i^f \geq f_i A_i$ and $f_i B_i \geq B_i^f$. Hence,

$$A_1^f \geq f_1 A_1 \geq f_1 B_1 \geq B_1^f,$$

and so

$$\begin{aligned} A_1^f + A_2^f = B_1^f + (A_1^f - B_1^f) + A_2^f \geq f_1 B_1 + f_1(A_1 - B_1) + f_2 A_2 \geq f_1 B_1 + f_2(A_1 - B_1) + f_2 A_2 = \\ = f_1 B_1 + f_2(A_1 - B_1 + A_2) \geq f_1 B_1 + f_2 B_2 \geq B_1^f + B_2^f, \end{aligned}$$

The inequalities use the fact that f_i is decreasing in i (since f is increasing and A_i's critical scores are decreasing by construction) and that $A_1 \geq B_1$, and later that $A_1 + A_2 \geq B_1 + B_2$.

Generally,

$$
\begin{aligned}
A_1^f + A_2^f + ... + A_m^f &= B_1^f + B_2^f + ... + B_{m-1}^f + A_m^f + \sum_{i<m}(A_i^f - B_i^f) \geq \\
&\geq B_1^f + B_2^f + ... + B_{m-1}^f + f_m A_m + \sum_{i<m} f_i(A_i - B_i) = \\
&= B_1^f + B_2^f + ... + B_{m-1}^f + f_m A_m + f_m \sum_{i<m}(A_i - B_i) + \sum_{i<m}(f_i - f_{i+1}) \cdot \sum_{j\leq i}(A_i - B_i) \geq \\
&\geq B_1^f + B_2^f + ... + B_{m-1}^f + f_m \left(A_m + \sum_{i<m}(A_i - B_i) \right) \geq \\
&\geq B_1^f + B_2^f + ... + B_{m-1}^f + f_m B_m \geq B_1^f + B_2^f + ... + B_m^f,
\end{aligned}
$$

but this means that s's deviation is weakly profitable.[13]

Finally, if $r_s^j \neq r_b^j$, it must be that $r_b^j > r_s^j$ (by the definition of j). This means that B_{j-1} is actually smaller from the perspective of s (as her outside option is more desirable). This only increases s's incentive to use the above deviation.

Thus, replacing the first $j-1$ entries on r_s with those of r_b yields an s-optimal ROL that is at leasat as aggressive than r_b. Repeating this process, enumerating over all b-optimal ROLs completes the proof (since each iteration only makes s's ROL more aggressive). □

4.1 "Falling through the Cracks"

It is clear that better students, who have higher expectations about their desirability to schools, are sometimes more likely to be assigned to some school. I now show that the opposite is also possible: in some cases better students are more likely to remain unmatched due to search frictions.

Proposition 3. *For all $k < N$, there exist pairs of agents s and b with identical preferences, but s is MLRP more optimistic, such that under their respective optimal size-k portfolios s is more likely than b to be unassigned.*

The proof is provided in the supplementary Appendix. It relies on cases where a marginal improvement to beliefs leads the an agent to choose a more selective safety school.

Next, I consider the case that applying to an additional school is costly but the size of the portfolio is not constrained. Specifically, I concentrate on the case of constant marginal cost. I show that the combination of uncertainty and application cost may lead to further failures in the assortativeness of the match.[14]

Proposition 4. *Let the marginal cost of application be $c > 0$. Then, there exist pairs of agents s and b with identical preferences, but s is MLRP more optimistic, such that:*

[13] I have used the fact that f_i's are decreasing, and so the expressions $(f_i - f_{i+1}) \cdot \sum_{j\leq i}(A_i - B_i)$ are a product of positive numbers.

[14] Nagypal (2004) derives similar comparative statics in a model with a continuum of schools whose selectivity varies continuously, where students' beliefs belong to certain families.

1) s's optimal portfolio is larger than b's,

2) s's optimal portfolio is smaller than b's,

3) s's optimal portfolio is identical to b's.

Remark 5. At points of indifference between a large and a small portfolio, the larger one spans a wider range of schools in terms of selectivity (by Theorem 1 below). Hence, in cases where a marginal improvement in beliefs leads to a reduction in the portfolio size, assortativeness is compromised.

Remark 6. I focus on the simple case of constant marginal cost to reassure the reader that the result is not driven by the cost function. The proof can easily be generalized to cover other cost functions.

The proof is provided in the supplementary Appendix. Intuitively, it shows that in the presence of application costs superstar applicants need not apply to more than one school, and that low-ability applicants should not apply at all.

5 Relaxation of Constraints

What happens to the optimal portfolio when the constraint on the number of schools is relaxed?[15] To answer this question, the following lemma, which is a corollary of Proposition 2, will be useful.

Lemma 3. *For all k, $c' < c \implies$ the most aggressive $r(k,c)$ is at least as aggressive as any $r(k,c')$.*

Theorem 1. *Optimal portfolios of larger sizes span a wider range of schools in terms of desirability (selectivity), and they refine the grid. Formally, for any size-k optimal portfolio, $r(k)$, there exists a size-$k{+}1$ optimal portfolio, $r(k{+}1)$, such that $r^1(k{+}1) \leq r^1(k) \leq r^2(k{+}1) \leq r^2(k)... \leq r^{k+1}(k{+}1)$.*

Example 2 in the supplementary Appendix shows that all the inequalities in the statement of Theorem 1 can hold strictly.

Proof. I provide a proof for the case that $r(k+1)$ is unique. The proof for the other case is similar, and is deferred to the supplementary Appendix.

When $r(k+1)$ is unique, then for all $j \leq k$, $r^{j+1}(k+1,c) = r^j(k, c_{r^1(k+1,c)})$. Since $c_{r^1(k+1,1)} \leq 1$ we have $r^j(k, c_{r^1(k+1,1)}) \geq r^j(k,1)$ for any $r(k,1)$. Thus, $r^j(k,1) \leq r^{j+1}(k+1,1)$ for all $j \leq k$. The same argument applied to the dual problem implies that $\overset{\circlearrowright}{r}{}^{j}(k,1) \geq \overset{\circlearrowright}{r}{}^{j+1}(k+1,1)$ for all $j \leq k$, and thus $r^{k-j}(k,1) \geq r^{k-j}(k+1,1)$ for all $j \leq k$, by Proposition 1. $\square$

To gain intuition, note that when a school is ranked highly, the probability that the next-ranked school will be relevant is scaled down additively (that is, the probability of assignment to the next-highest school decreases by the same amount, regardless of the identity of this school). This means

[15]There are two ways to address this question. Here, I consider the effect of relaxing the constraint for an individual agent in a large market (i.e., cutoffs are fixed). In a separate project, I consider the equilibrium implications of relaxing the constraint for all agents, which affects admission cutoffs (see Hafalir et al., 2018, for a special case).

that the attractiveness of schools with higher ex-post value (high u_i) decreases more when another school is ranked above them, or when the school above them is replaced with a lower-cutoff school (as illustrated in Figure 5). Symmetrically, when the constraint on the portfolio size is relaxed the continuation value is higher, making aggressive gambles more appealing. Intuitively, these forces give agents incentives to widen the range of selectivity of their applications.

I now prove that the marginal benefit from relaxing the constraint on the size of the portfolio is decreasing. The proof invokes Theorem 1 and uses a technique similar to that of Wong (2014).

Theorem 2. *The marginal benefit from increasing the size of the portfolio is decreasing. Formally, the function* $MB(k) := v(k,1) - v(k-1,1)$ *is decreasing in* k.

Proof. Denote by $r(k)$ an optimal portfolio of size k. Then

$$v(k+1) = U(r(k+1)) = \sum_{i=1}^{k+1} \left(c_{r^{i-1}(k+1)} - c_{r^{i}(k+1)}\right) u(r^i(k+1)),$$

where $U(\cdot)$ represent the expected utility from a portfolio, and similarly

$$v(k-1) = U(r(k-1)) = \sum_{i=1}^{k-1} \left(c_{r^{i-1}(k-1)} - c_{r^{i}(k-1)}\right) u(r^i(k-1)).$$

I will identify two (potentially suboptimal) size-k portfolios, p and p', such that

$$v(k+1) + v(k-1) \leq u(p) + u(p').$$

By the definition of $v(k)$ this will imply that

$$v(k+1) + v(k-1) \leq 2v(k)$$

and hence

$$v(k+1) - v(k) \leq v(k) - v(k-1)$$

as required.

To construct the portfolios p and p' (see Figure 6 for an illustration), list all the $2k$ entries of $r(k+1)$ and $r(k-1)$ such that entries are sorted according to weakly-decreasing order of desirability of the corresponding school and for all $i \in \{1,2,...k-1\}$ the entry $r^i(k+1)$ appears before $r^i(k-1)$ on the sorted list. It is possible to choose $r(k+1)$ and $r(k-1)$ such that this last requirement is satisfied by Theorem 1.

Since $r(k+1)$ has two more entries than $r(k-1)$, there must exist a minimal m such that $r^m(k+1)$ and $r^{m+1}(k+1)$ appear consecutively on the sorted list. Let p be the ROL that is equal to $r(k+1)$ up to the m-th entry followed by the last $k-m$ entries of $r(k-1)$, and p' be the ROL consisting of the first $m-1$ entries of $r(k-1)$ followed by the last $k+1-m$ entries of $r(k+1)$.

Then

$$v(k+1)+v(k-1)-u(p)-u(p') = \left(c_{r^m(k+1)} - c_{r^{m-1}(k-1)}\right)\left(u(r^{m+1}(k+1)) - u(r^m(k-1))\right) \leq 0,$$

where the inequality follows by the following observations. First, $c_{r^m(k+1)} \leq c_{r^{m-1}(k-1)}$ because $r^{m-1}(k-1)$ appears before $r^m(k+1)$ on the sorted list (by the minimality of m), and thus $r^{m-1}(k-1)$ is weakly more desirable. If the inequality did not hold, $r^m(k+1)$ would have been a dominated choice (by $r^{m-1}(k-1)$), in contradiction to the optimality of $r(k+1)$. Second, $u(r^{m+1}(k+1)) \geq u(r^m(k-1))$, again by the definition of m. This completes the proof. □

6 Monotonicity in Risk Aversion

How does risk aversion affect the optimal portfolio? Adopting the definition of Coles and Shorrer (2014), I say that agent b is *more risk averse* than s if there exists a concave nondecreasing function, ϕ, such that $\phi(0) = 0$ and for any school, i, the utility that b derives from attending it, u_i^b, is equal to $\phi(u_i^s)$. In other words, the two agents share the same ordinal preferences, but the relative marginal benefit from attending a preferred school is smaller for the more risk averse agent. The assumption that $\phi(0) = 0$ is just a convenient normalization, as the concavity of ϕ is unaffected by positive affine transformations. This normalization holds the value of the outside option fixed.

Proposition 5. *Let s and b be agents with identical beliefs and ordinal preferences, but b is more risk averse than s. Then for all k, i) there exists an s-optimal size-k portfolio that is* at least as aggressive as *any b-optimal size-k portfolio, and ii) there exists a b-optimal size-k portfolio such that any s-optimal size-k portfolio is at least as aggressive.*

Proof. What are the implications of b being more risk averse than s for the dual decision problems? To begin with, note that $\bar{u}_i^s = \bar{u}_i^b$ for all i, as these depend on values of c_i exclusively, and these are equal for both students.

Next, note that if b is more risk averse than s, there must exist a concave transformation, ϕ, such that $\phi(0) = 0$ and $u_i^b = \phi(u_i^s)$. Set the convention that $c_0 \equiv 1$ and $u_{N+1} \equiv 0$, and observe that since ϕ is concave and increasing, it follows that for all $N \geq i > 1$,

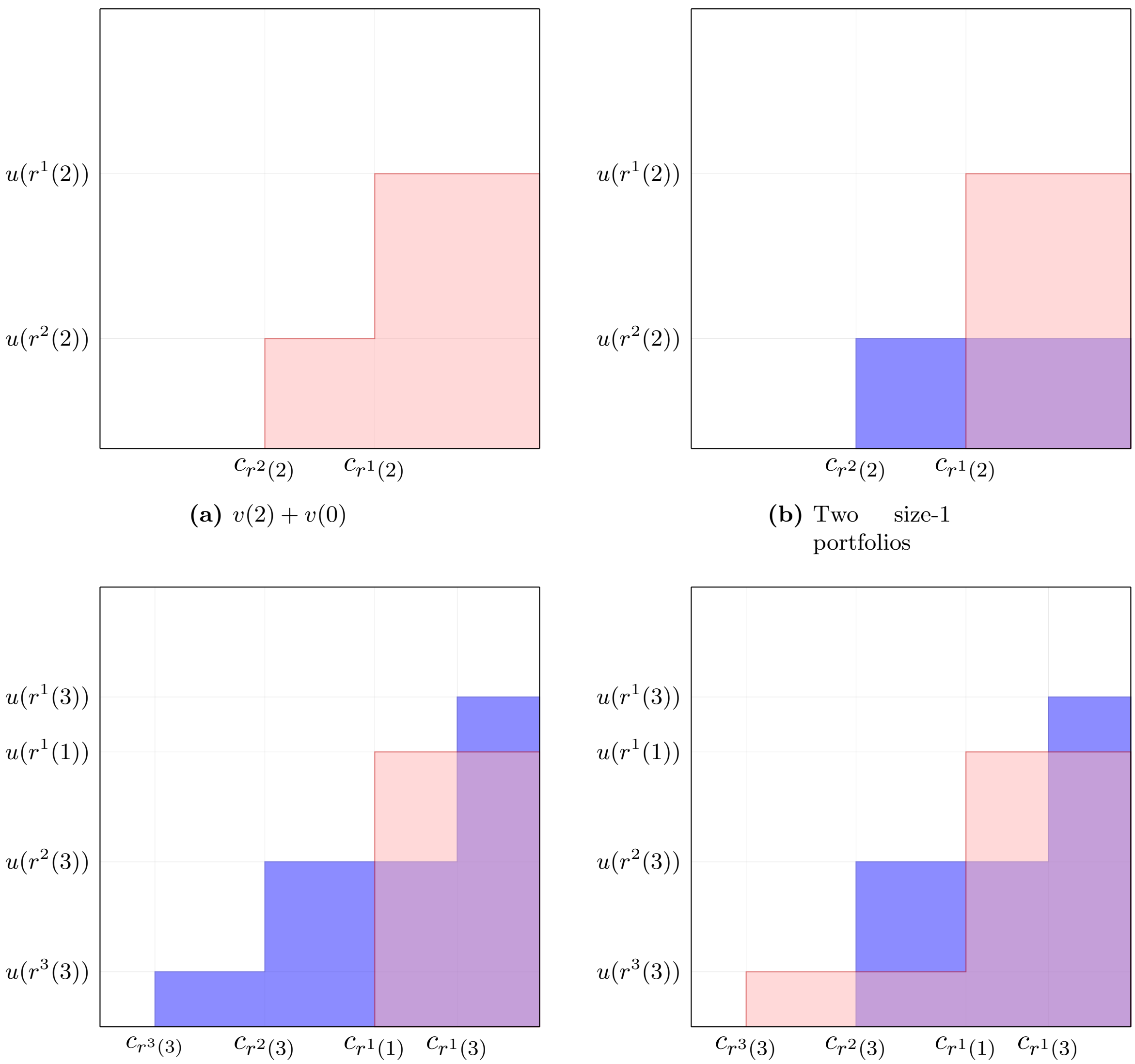
$$\frac{\phi(u_{i-1}^s) - \phi(u_i^s)}{u_{i-1}^s - u_i^s} \geq \frac{\phi(u_i^s) - \phi(u_{i+1}^s)}{u_i^s - u_{i+1}^s}.$$

Adding and reducing 1 everywhere yields

$$\frac{[1-\phi(u_i^s)] - \left[1-\phi(u_{i-1}^s)\right]}{[1-u_i^s] - \left[1-u_{i-1}^s\right]} \geq \frac{\left[1-\phi(u_{i+1}^s)\right] - [1-\phi(u_i^s)]}{\left[1-u_{i+1}^s\right] - [1-u_i^s]},$$

which can be rewritten as

$$\frac{\bar{c}_{i-1}^b - \bar{c}_i^b}{\bar{c}_{i-1}^s - \bar{c}_i^s} \geq \frac{\bar{c}_i^b - \bar{c}_{i+1}^b}{\bar{c}_i^s - \bar{c}_{i+1}^s}.$$



(a) $v(2) + v(0)$

(b) Two size-1 portfolios

(c) $v(3) + v(1)$

(d) Two size-2 portfolios

Figure 6: Illustration of the proof of Theorem 2

Panels 6a and 6b show two size-1 portfolios whose covered areas sum to more than the area covered by the optimal size-2 portfolio (plus zero, the coverage of the optimal size-0 portfolio). Panels 6c and 6d show two size-2 portfolios whose covered areas sum to more than the sum of areas covered by the optimal portfolios of size 1 and 3. In each panel, each of the portfolios is represented by a different color (pink or light blue). The purple areas indicate areas that are double counted when summing the areas covered by the two portfolios as they are covered by both. The figures highlight that the union of the two portfolios cover the same area, but in the cases of same size portfolios there is larger overlap—i.e., a larger area is double counted.

Now, note that given a set of options ordered by their selectivity, the only thing that matters for the purposes of decision making is the probabilities of the score realizations between each pair of cutoffs (and not the cardinal value of the cutoff, which is sufficient when the distribution is assumed to be uniform). Based on this observation, and on the above inequality, there are choice problems equivalent to $(\bar{c}_i^s, \bar{u}_i^s)_{i=1}^N$ and $\left(\bar{c}_i^b, \bar{u}_i^b\right)_{i=1}^N$ such that $\bar{c}_i^s = \bar{c}_i^b$ for all i, $\bar{u}_i$'s are unchanged (and thus equal), and the beliefs of b are MLRP more optimistic.[16] Thus, by Proposition 2, in the dual problem there is a b-optimal portfolio that is at least as aggressive as any s-optimal portfolio, and an s-optimal portfolio that is weakly less aggressive than any b-optimal portfolio. The Proposition follows by Proposition 1. □

Figure 7 provides graphical illustration for the relationship between risk aversion and beliefs in a special case.

7 Monotonicity in Outside Options

How do agents' outside options affect their optimal portfolios? To answer this question, I dispense with the normalization of the value of the outside option to 0. Instead, I compare the optimal behavior of individuals who are identical, except that they have different access to outside options. This may occur, for example, when a centralized school-choice mechanism allocates seats in public schools, but families differ in their ability to pay for a private school, or when students have different default assignments in case they are not placed by the mechanism (as is the case in New Orleans; Gross, DeArmond and Denice, 2015).

Definition 1. A *(stochastic) outside option* for agent i is a random variable, $\boldsymbol{o}$, with finite support, that is independent from schools' acceptance decisions and whose realization is available for i to consume regardless of her application portfolio.

Definition 2. Let $\boldsymbol{o}$ and $\boldsymbol{o'}$ be two outside options with probability mass functions f_o and $f_{o'}$ respectively. Then $\boldsymbol{o}$ MLRP-dominates $\boldsymbol{o'}$ if, restricting attention to values of x in the union of their supports, $\frac{f_o(x)}{f_{o'}(x)}$ is weakly increasing (where $\frac{c}{0} \equiv 0$ when $c > 0$).

Theorem 3. *Let s and b be agents who are identical, except that b's outside option MLRP-dominates s's outside option (and both outside options are independent of the score). Then for all k, i) there exists a b-optimal size-k portfolio that is at least as aggressive as any s-optimal size-k portfolio, and ii) there exists an s-optimal size-k portfolio such that any b-optimal size-k portfolio is at least as aggressive.*

Proof. Let $\boldsymbol{o_x}$ denote the (stochastic) outside option of agent x. Write $\hat{u}_x(z) = \mathrm{E}\left[\max\left\{o_x, u_x(z)\right\}\right]$,

[16]For example, hold $\bar{c}_i^s$'s fixed and let the distribution of scores for b be constant $\left(\frac{\bar{c}_i^b - \bar{c}_{i+1}^b}{\bar{c}_i^s - \bar{c}_{i+1}^s}\right)$ between consecutive pairs of cutoffs.

where 0 stands for no assignment, and $u_x(0) = 0$. Note that the optimal portfolio for agent x solves

$$\max_r \sum_{1 \le i \le |r|} \max\left\{0, \min_{j<i}\{c_{r^j}\} - c_{r^i}\right\} \hat{u}_x(r^i) + \max\left\{0, \min_{r^j \in r}\{c_{r^j}\}\right\} \mathrm{E}\left[o_x\right] - C(|r|),$$

where the "new" term corresponds to the outside option that was normalized to 0 previously (equivalently, deduct from each $\hat{u}_x$ the expected value to x from consuming her outside option, $\mathrm{E}\left[o_x\right]$, and use the previous formulation).

Denote the CDF of the random variable $\boldsymbol{o_x}$ by F_x, and note that

$$\hat{u}_x(z) = u_x(z) + \int_{u_x(z)}^{\infty} [1 - F_x(s)]ds,$$

thus, b is less risk averse than s if o_b MLRP-dominates o_s.[17] □

Corollary 1. *Let s and b be agents who are identical, except that b's deterministic outside option is more desirable than s's. Then for all k, i) there exists a b-optimal size-k portfolio that is at least as aggressive as any s-optimal size-k portfolio, and ii) there exists an s-optimal size-k portfolio such that any b-optimal size-k portfolio is at least as aggressive.*

Figure 7 is an illustration of Corollary 1.

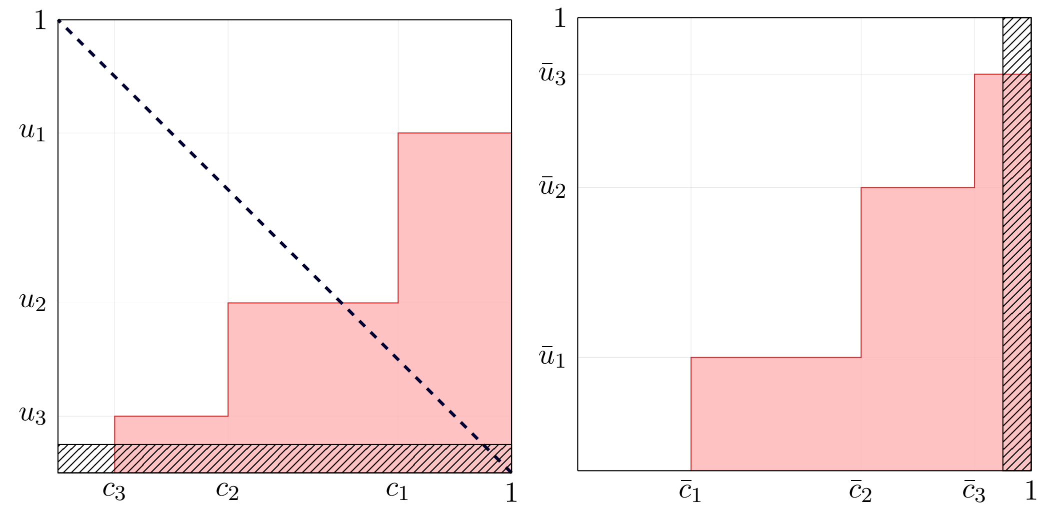
8 Discussion

There are many important simultaneous search problems that present a nontrivial correlation structure in the probabilities of success. I showed that an important case can be solved using dynamic programming. This approach, which is in fact more general, provides a practical tool for both theoretical and empirical research (Calsamiglia, Fu and Güell, forthcoming; Shorrer, 2019).

I have also shown that this decision problem can be represented as a max-coverage problem. This representation is a special case of Segal (1989; 1993). It sheds light on the relation between the concavity of the utility function and of beliefs, through a notion of duality, not unrelated to that of Yaari (1987). This relationship has been documented, in other contexts, in numerous studies of non-expected utility theory (e.g., Yaari, 1987; Hong, Karni and Safra, 1987).

I have shown that when applicants are vertically differentiated but face uncertainty about their standing relative to others, an application-cost reduction leading to an increase in the number of schools one applies to makes the optimal portfolio wider. This stands in contrast to the prediction of the model where admission decisions are independent, and is consistent with behavior in the

[17] At values of x that are not deterministically lower than s's outside option (where $\hat{u}_x$ is flat for both agents), the slope of the function ϕ that maps $\hat{u}_s(x)$ to $\hat{u}_b(x)$ is $\frac{F_b(x)}{F_s(x)}$, which is increasing in $\hat{u}_b(x)$. I avoided using the inverse notation in the previous sentence, but I Implicitly used the fact that $\hat{u}_s$ is strictly increasing in this domain, and hence invertible.



(a) primal problem

(b) dual problem

Figure 7: The effect of an improved outside option (Corollary 1)

Panel 7a illustrates the effect of an improvement to an agent's outside option. It shows that applications to less selective schools are disproportionately affected (the reduction in the area the corresponding rectangles cover is disproportionately large). Panel 7b illustrates the dual problem. It corresponds to the same coverage problem, but the graph is transposed over the line connecting $(1, 0)$ and $(0, 1)$. The effect on the dual problem is similar to the one illustrated in Figure 5, suggesting an alternative proof for Corollary 1 (by Proposition 1 and Lemma 3).

(centralized) Ghanaian high-school admissions and in the (decentralized) U.S. college admissions, suggesting that agents are facing uncertainty about their strength as applicants.

Centralized clearinghouses determine the school assignment of millions of students around the world. Pathak and Sönmez (2008) show that strategy-proof assignment mechanisms, ones that give no incentives for applicants to misrepresent their preferences, "level the playing field" by protecting strategically unsophisticated applicants. I provide an alternative argument in support of strategy-proof mechanisms: when popular manipulable mechanisms are in place, strong applicants, ones with good outside options (such as private schools, or access to a separate admissions process), will be over represented in the most-desirable schools, even when all applicants are strategically sophisticated. The same holds for applicants with differing levels of confidence or risk preferences, which may lead to disparities across gender lines.

There are many interesting research directions that are beyond the scope of this paper. First, revisiting studies that make the assumption of independence and replacing this assumption. Second, analyzing other simultaneous search environments using the dynamic programming approach. Finally, studying equilibrium behavior in labor markets where firms' preferences are correlated.

References

Abdulkadiroğlu, Atila, and Tayfun Sönmez. 1998. "Random serial dictatorship and the core from random endowments in house allocation problems." *Econometrica*, 66(3): 689–701.

Abdulkadiroğlu, Atila, and Tayfun Sönmez. 2003. "School choice: A mechanism design approach." *American Economic Review*, 93(3): 729–747.

Abdulkadiroğlu, Atila, Joshua D Angrist, Yusuke Narita, Parag A Pathak, and Roman A Zarate. 2017. "Regression discontinuity in serial dictatorship: Achievement effects at Chicago's exam schools." *American Economic Review*, 107(5): 240–45.

Agarwal, Nikhil, and Paulo Somaini. 2018. "Demand analysis using strategic reports: An application to a school choice mechanism." *Econometrica*, 86(2): 391–444.

Ajayi, Kehinde, and Modibo Sidibe. 2015. "An empirical analysis of school choice under uncertainty." Boston University Working paper, http://people. bu. edu/kajayi/research. html.

Ajayi, Kehinde F. 2011. "School choice and educational mobility: Lessons from secondary school applications in Ghana." *University of California-Berkeley Working Paper*.

Akbarpour, Mohammad, and Winnie van Dijk. 2018. "School choice with unequal outside options." Working paper.

Albrecht, James, Pieter A. Gautier, and Susan Vroman. 2006. "Equilibrium directed search with multiple applications." *The Review of Economic Studies*, 73(4): 869–891.

Athey, Susan, and Glenn Ellison. 2011. "Position auctions with consumer search." *The Quarterly Journal of Economics*, 126(3): 1213–1270.

Avery, Christopher, Jessica S Howell, and Lindsay Page. 2014. "A review of the role of college applications on students' postsecondary outcomes. Research brief." *College Board*.

Azevedo, Eduardo M., and Jacob D. Leshno. 2016. "A supply and demand framework for two-sided matching markets." *Journal of Political Economy*, 124(5): 1235–1268.

Broecke, Stijn. 2012. "University selectivity and earnings: Evidence from UK data on applications and admissions to university." *Economics of Education Review*, 31(3): 96–107.

Calsamiglia, Caterina, Chao Fu, and Maia Güell. forthcoming. "Structural estimation of a model of school choices: The Boston mechanism vs. its alternatives." *Journal of Political Economy*.

Calsamiglia, Caterina, Francisco Martínez-Mora, and Antonio Miralles. 2015. "School choice mechanisms, peer effects and sorting." *Univ. of Leicester Dept. Ecoomics WP*, , (15/01).

Calsamiglia, Caterina, Guillaume Haeringer, and Flip Klijn. 2010. "Constrained school choice: An experimental study." *American Economic Review*, 100(4): 1860–1874.

Card, David, and Alan B Krueger. 2005. “Would the elimination of affirmative action affect highly qualified minority applicants? Evidence from California and Texas.” *ILR Review*, 58(3): 416–434.

Chade, Hector, and Lones Smith. 2005. “Simultaneous search.”

Chade, Hector, and Lones Smith. 2006. “Simultaneous search.” *Econometrica*, 74(5): 1293–1307.

Chade, Hector, Gregory Lewis, and Lones Smith. 2014. “Student portfolios and the college admissions problem.” *Review of Economic Studies*, 81(3): 971–1002.

Che, Yeon-Koo, and Olivier Tercieux. 2017. “Top trading cycles in prioritized matching: An irrelevance of priorities in large markets.”

Coles, Peter, and Ran I. Shorrer. 2014. “Optimal truncation in matching markets.” *Games and Economic Behavior*, 87: 591–615.

Dargnies, Marie-Pierre, Rustamdjan Hakimov, and Dorothea Kübler. forthcoming. “Self-confidence and unraveling in matching markets.” *Management Science*.

De Haan, Monique, Pieter A. Gautier, Hessel Oosterbeek, and Bas Van der Klaauw. 2015. “The performance of school assignment mechanisms in practice.”

Delacroix, Alain, and Shouyong Shi. 2006. “Directed search on the job and the wage ladder.” *International Economic Review*, 47(2): 651–699.

De Los Santos, Babur, Ali Hortaçsu, and Matthijs R. Wildenbeest. 2012. “Testing models of consumer search using data on web browsing and purchasing behavior.” *American Economic Review*, 102(6): 2955–80.

Fu, Chao. 2014. “Equilibrium tuition, applications, admissions, and enrollment in the college market.” *Journal of Political Economy*, 122(2): 225–281.

Gale, David, and Lloyd S. Shapley. 1962. “College admissions and the stability of marriage.” *American Mathematical Monthly*, 69(1): 9–15.

Galenianos, Manolis, and Philipp Kircher. 2009. “Directed search with multiple job applications.” *Journal of Economic Theory*, 144(2): 445–471.

Gross, Betheny, Michael DeArmond, and Patrick Denice. 2015. “Common enrollment, parents, and school choice: Early evidence from Denver and New Orleans. Making School Choice Work Series.” *Center on Reinventing Public Education*.

Haeringer, Guillaume, and Flip Klijn. 2009. “Constrained school choice.” *Journal of Economic Theory*, 144(5): 1921–1947.

Hafalir, Isa E., Rustamdjan Hakimov, Dorothea Kübler, and Morimitsu Kurino. 2018. "Self-confidence and unraveling in matching markets." *Journal of Economic Theory*, 176: 886–934.

Hastings, Justine, Thomas J Kane, and Douglas O Staiger. 2009. "Heterogeneous preferences and the efficacy of public school choice." *NBER Working Paper*, 2145: 1–46.

Hochbaum, Dorit S. 1996. *Approximation algorithms for NP-hard problems.* PWS Publishing Co.

Hong, Chew Soo, Edi Karni, and Zvi Safra. 1987. "Risk aversion in the theory of expected utility with rank dependent probabilities." *Journal of Economic theory*, 42(2): 370–381.

Immorlica, Nicole S, Jacob D Leshno, Irene Y Lo, and Brendan J Lucier. 2018. "Costly information acquisition and stable matching mechanisms."

Kircher, Philipp. 2009. "Efficiency of simultaneous search." *Journal of Political Economy*, 117(5): 861–913.

Lee, Joonbae, and Hanna Wang. 2018. "Ranking and search effort in matching."

Leshno, Jacob, and Irene Lo. 2017. "The simple structure of top trading cycles in school choice."

Nagypal, Eva. 2004. "Optimal application behavior with incomplete information." *Manuscript, Econ. Dept., Northwestern Univ*, 1.

Olszewski, Wojciech, and Rakesh Vohra. 2016. "Simultaneous selection." *Discrete Applied Mathematics*, 200: 161–169.

Oosterbeek, Hessel, Sándor Sóvágó, and Bas van der Klaauw. 2019. "Why are schools segregated? Evidence from the secondary-school match in Amsterdam."

Pallais, Amanda. 2015. "Small differences that matter: Mistakes in applying to college." *Journal of Labor Economics*, 33(2): 493–520.

Pan, Siqi. 2019. "The instability of matching with overconfident agents." *Games and Economic Behavior*, 113: 396–415.

Pathak, Parag A., and Tayfun Sönmez. 2008. "Leveling the playing field: Sincere and sophisticated players in the Boston mechanism." *American Economic Review*, 98(4): 1636–1652.

Segal, Uzi. 1989. "Anticipated utility: A measure representation approach." *Annals of Operations Research*, 19(1): 359–373.

Segal, Uzi. 1993. "The measure representation: A correction." *Journal of Risk and Uncertainty*, 6(1): 99–107.

Shorrer, Ran I. 2019. "Unequal outcomes on a level playing field: The role of search frictions." Penn State.

Wong, Adam Chi Leung. 2014. "The choice of the number of varieties: Justifying simple mechanisms." *Journal of Mathematical Economics*, 54: 7–21.

Yaari, Menahem E. 1987. "The dual theory of choice under risk." *Econometrica: Journal of the Econometric Society*, 95–115.

Supplementary Material

A Relaxation of Constraints with Independent Admissions

In this appendix, I consider the environment where schools' admissions decisions are independent conditional on the information available to the student. I formalize the observation made by Chade and Smith (2006), which they term the upward diversity of optimal portfolios (see also Kircher, 2009). To make the result transparent, I assume that there are many schools of each type, where a type is defined by (u_c, p_c), i.e., the cardinal utility from attending the school c and the probability of admission. There may also be infinitely many school types (as in the labor market models of Galenianos and Kircher, 2009 and Kircher, 2009).

Theorem 4. *Optimal portfolios are upwardly diverse in the sense that the decision maker derives higher utility from consuming options that belong to larger portfolios only. Formally, if a school c is in $r(k+1)$, an optimal size-$k+1$ portfolio, but not in $r(k)$, an optimal size-k portfolio which is included in $r(k+1)$, then $u_c \geq u_{c'}$ for any other school c' in $r(k)$. Moreover, if a school c is in $r(k+1)$ but not in $r(k)$, then $p_c \leq p_{c'}$ for any other school c' in $r(k)$.*

Proof. First note that in the current setting if $u_c \geq u_{c'}$ and $p_c \geq p_{c'}$ with one inequality being strict, then schools of the same type as c' will never appear on any optimal ROL as they are dominated by schools of the same type as c, and I assumed that there are arbitrarily many schools of each type. Next, note that the beliefs of an applicant about admission probabilities to a school do not change if the applicant learns she is rejected by other schools (due to independence). Thus, the decision maker is facing the same problem of finding the optimal portfolio, with the only difference being that the length of the ROL to be chosen is shorter by one (as there are many copies of c). That is, $r(k) = \big(r^1(k), r(k-1)\big)$. Denoting by $v(r)$ the expected utility from the ROL r, I note that $r^1(k)$ is the solution to the following problem:

$$r^1(k) = \operatorname*{argmax}_c p_c u_c + (1 - p_c) v(r(k-1)).$$

Since $v(k-1) \colon = v(r(k-1))$ is an increasing function of k, and the same types of schools are available for any k (since there are plenty), it is clear that higher u (and lower p) schools become increasingly attractive as k grows large. To see this, note that the problem of choosing an optimal size-1 portfolio is identical to the problem where agents are vertically differentiated (correlation in admission chances only matters when portfolios have more than one school). Thus, the choice of

$r^1(k)$ is identical to the choice of an optimal size-1 portfolio with an outside option of $v(k-1)$. The result therefore follows from Corollary 1 and the fact that $v(k-1)$ is strictly increasing in k (hence potential previous ties are broken). □

B Additional Proofs

Proposition 3

Proof. Let b be an agent with full-support beliefs whose optimal portfolio does not consist of the top k choices. Fixing preferences, it is easy to identify such beliefs, by making high scores implausible.

Let r_b^k be the last-ranked school on b's ROL. Mix b's belief (with weight α) with her belief conditional on her score being above c_k (with weight $1-\alpha$). For low enough values of α, the optimal portfolio has a strictly more selective last choice (the k-th school). Also, for any value of α, the optimal portfolio is at least as aggressive as b's (by Proposition 2). Denote by $\hat{\alpha}$ the supremum of values of α for which the last choice is strictly more selective than r_b^k. Comparing the optimal ROLs for $\hat{\alpha}-\delta$ and $\hat{\alpha}+\delta$ for a small $\delta>0$ gives an example as required, since admission chances increase only marginally (thanks to improved beliefs) but decrease discontinuously (due to the shift to a more selective last choice). □

Proposition 4

Proof. Assume that b has a uniform belief on $[0,1]$, and the marginal cost is low enough so that b's optimal portfolio consists of more than one school. If s's belief is uniform on $[c_1,1]$, her optimal portfolio consists only of the most selective school (since she is certain she will be admitted).

Similarly, let s have a uniform belief on $[0,1]$, and let c be low enough so that s's optimal portfolio consists of more than one school. If $b's$ belief is uniform on $[0,c_{N-1}]$, her optimal portfolio consists of either the least selective school or no school at all (since she is certain she will be rejected by any other school).

The third case follows from continuity of the expected utility in beliefs. □

Theorem 1

Proof. (continued). Consider an optimal portfolio of size k, $r(k)$. If there exists an optimal portfolio of size $k+1$, $p(k+1)$ such that $r^i(k)=p^i(k+1)$ or $r^i(k)=p^{i+1}(k+1)$, then the proof holds by induction: For the first case, set $r^j(k+1)$ to equal $r^j(k)$ for all $j\le i$ and choose the ROL of size $k-i+1$ that refines the $k-i$ suffix of $r(k)$, which exists by induction. For the second case, apply the same argument to the dual market.

Otherwise, all optimal portfolios of size $k+1$ do not satisfy all inequalities. Consider the first element of the most aggressive optimal ROL of size $k+1$ and it's least aggressive continuation. By Proposition 2, it's last element is weakly less desirable than $r^k(k)$, and since we are not in the first case, strictly so. The same argument applied to the dual problem establishes that the first element on the size-$k+1$ ROL is strictly more desirable than $r^1(k)$. I will show, by way of contradiction, that these portfolios satisfy the inequalities in the statement of the theorem.

Towards contradiction, assume that they do not. Then it must be that there exist i and $j \in \{i, i+1\}$ such that $c_{r^{j-1}(k+1)} > c_{r^{i-1}(k)} > c_{r^i(k)} > c_{r^j(k+1)}$ (recall the convention that $c_{r^{j-1}(k+1)} = 1$ if $j = 1$). But this contradicts the optimality of the two portfolios, because by swapping the suffixes that start with $r^j(k+1)$ and $r^i(k)$ the sum of the expected utilities from the two portfolios will increase by

$$
\begin{aligned}
&\left[\left(c_{r^{j-1}(k+1)} - c_{r^i(k)}\right) u\left(r^i(k)\right) + \left(c_{r^{i-1}(k)} - c_{r^j(k+1)}\right) u\left(r^j(k+1)\right)\right] - \\
&\left[\left(c_{r^{j-1}(k+1)} - c_{r^j(k+1)}\right) u\left(r^j(k+1)\right) + \left(c_{r^{i-1}(k)} - c_{r^i(k)}\right) u\left(r^i(k)\right)\right] = \\
&\left(c_{r^{j-1}(k+1)} - c_{r^{i-1}(k)}\right) u\left(r^i(k)\right) + \left(c_{r^{i-1}(k)} - c_{r^{j-1}(k+1)}\right) u\left(r^j(k+1)\right) = \\
&\qquad \left(c_{r^{j-1}(k+1)} - c_{r^{i-1}(k)}\right)\left(u\left(r^i(k)\right) - u\left(r^j(k+1)\right)\right) > 0
\end{aligned}
$$

where the inequality holds since the first term is positive since $c_{r^{j-1}(k+1)} - c_{r^{i-1}(k)}$ and the second is positive as otherwise $r^i(k)$ would be dominated by $r^j(k+1)$, which is less selective. Thus, the sum of the expected utilities from the two portfolios that result from the swap is strictly greater than the sum of the expected utilities from the original portfolios. But this contradicts the optimaility of the original portfolios, because the swap leaves us with one size-k portfolio and one size-$k+1$ portfolio, and if both achieve no more than the original portfolios, the above cannot hold. □

C Forward-Looking PSA

Algorithm 2 Forward-Looking PSA

Step 1 For all possible values of r^2, find $r^1(r^2) = \arg\max_{c_{r^1} \geq c_{r^2}} (1 - c_{r^1})(u_{r^1} - u_{r^2})$. Set $v_1(r^2) := \max_{c_{r^1} \geq c_{r^2}} (1 - c_{r^1})(u_{r^1} - u_{r^2})$.

Step i<k For all possible values of r^{i+1}, find $r^i(r^{i+1}) = \arg\max_{c_{r^i} \geq c_{r^{i+1}}} (1 - c_{r^i})(u_{r^i} - u_{r^{i+1}}) + v_{i-1}(r^i)$. Set $v(r^{i+1}) = \max_{c_{r^i} \geq c_{r^{i+1}}} (1 - c_{r^i})(u_{r^i} - u_{r^{i+1}}) + v(r^i)$.

Step k Set $r^k = \arg\max_i (1 - c_{r^k}) u_{r^k} + v_{k-1}(r^k)$, and for all $j < k$, $r^j = r^j(r^{j+1}(...(r^k)))$.

D Failure of the "Greedy" Approach

Chade and Smith (2006) show that the "greedy" marginal improvement algorithm identifies the optimal portfolio when admission decisions are independent conditional on the information available to the applicant. The following example shows that this result does not extend to the case of perfectly aligned admission decisions. A greedy algorithm identifies an optimal portfolio of any size only when the optimal portfolios are nested. In the example, the optimal portfolios of size 1 and of size 2 are disjoint. This example also shows that all of the inequalities in Theorem 1 may hold strictly.

Example 2. Consider an environment with three schools such that $c_1 = \frac{3}{4}$, $c_2 = \frac{1}{2}$, and $c_3 = 0$, and $u_1 = 4$, $u_2 = 2.01$, and $u_3 = 1$. The values were selected so that singleton portfolios yield the same expected utility (i.e., $(1 - c_i) \cdot u_i = 1$), except that School 2 yields a slightly higher expected utility. It is thus clear that the best singleton portfolio consists of School 2.

Next we consider the optimal portfolio of size 2. Denote by v_{ij} the value from the two-school portfolio $\{i, j\}$. Then

$$v_{12} = \frac{1}{4} \cdot 4 + (\frac{1}{2} - \frac{1}{4}) \cdot 2.01 \approx \frac{6}{4},$$

$$v_{13} = \frac{1}{4} \cdot 4 + (1 - \frac{1}{4}) \cdot 1 = \frac{7}{4},$$

$$v_{23} = \frac{1}{2} \cdot 2.01 + (1 - \frac{1}{2}) \cdot 1 \approx \frac{6}{4}.$$

Hence, the optimal portfolio of size 2 consists of Schools 1 and 3, and does not include School 2. A greedy algorithm will only achieve a fraction of approximately $\frac{6}{7}$ of the expected utility from the optimal portfolio.

Ajayi and Sidibe (2015) use a greedy approach to approximate the optimal portfolio in their empirical study of school choice in Ghana. To get a satisfactory approximation, they enhance the simple algorithm in various ways. The following proposition provides an explanation of their success: the baseline that they chose, the greedy algorithm, cannot perform too poorly. Even before introducing the modifications of Ajayi and Sidibe (2015), a greedy algorithm is assured to achieve at least 63% of the expected utility from the optimal portfolio.

E Nondecreasing Marginal Benefit

Example 3. Consider the following six rectangles:[18]

$A = ((0, 0), (3, 0), (0, 3), (3, 3))$

$B_1 = (-0.4, 0), (-0.4, 1.5), (3, 0), (3, 1.5))$

$B_2 = ((-0.4, 1.5), (-0.4, 3), (3, 1.5), (3, 3))$

$C_1 = ((0, 0), (0, 4), (1, 0), (1, 4))$

$C_2 = ((1, 0), (1, 4), (2, 0), (2, 4))$

$C_3 = ((2, 0), (2, 4), (3, 0), (3, 4))$

Direct calculation shows that the maximal coverage by a single rectangle is by A with a covered area of 9, the maximal coverage by a pair of rectangles is by $B_1 \cup B_2$ with a covered area of 10.2, and the maximal coverage by three rectangles is by $C_1 \cup C_2 \cup C_3$ with a covered area of 12. Thus,

[18] I thank Avinatan Hassidim for providing this example.

the marginal benefit from relaxing the constraint from 1 to 2 is 1.2, but the marginal benefit from relaxing the constraint from 2 to 3 is $1.8 > 1.2$.