NBER WORKING PAPER SERIES

INATTENTION AND INEQUITY IN SCHOOL MATCHING

Stefan F. Bucher
Andrew Caplin

Working Paper 29586
http://www.nber.org/papers/w29586

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2021

The authors wish to thank Adam Brandenburger, Ambuj Dewan, Octavia Ghelfi, Sam Kapon, Dmitry Sorokin, and seminar participants at NYU, the Virtual Behavioral Economics Seminar (VIBES), Universidad de Alicante, and the 2020 Workshop on the Cognitive Foundations of Economic Behavior. We would like to thank the NOMIS Foundation and the Alfred P. Sloan Foundation for their support. Portions of this paper appeared in Stefan Bucher's PhD dissertation at NYU. All remaining errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Inattention and Inequity in School Matching
Stefan F. Bucher and Andrew Caplin
NBER Working Paper No. 29586
December 2021
JEL No. C78,D47,D82,D83

ABSTRACT

The attractive properties of the Deferred Acceptance (DA) algorithm rest on the assumption of perfect information. Yet field studies of school matching show that information is imperfect, particularly for disadvantaged students. We model costly strategic learning when schools are ex ante symmetric, agree on their ranking of students, and learning is rationally inattentive. Our analytic solution quantifies how each student's rank, learning costs and prior beliefs interact to determine their gross and net welfare as well as the extent and form of mistakes they make. In line with the evidence, we find that lower-ranked students are affected disproportionately more by information costs, generally suffering a larger welfare loss than higher-ranked students. Interactions between mechanism design, inattention and inequity are thus of first order importance.

Stefan F. Bucher
NYU Department of Economics
19 W4th St, 6th Fl
New York, NY 10012
United States
sfb311@nyu.edu

Andrew Caplin
Department of Economics
New York University
19 W. 4th Street, 6th Floor
New York, NY 10012
and NBER
andrew.caplin@nyu.edu

Inattention and Inequity in School Matching

Stefan Bucher and Andrew Caplin*†

December 13, 2021

Abstract

The attractive properties of the Deferred Acceptance (DA) algorithm rest on the assumption of perfect information. Yet field studies of school matching show that information is imperfect, particularly for disadvantaged students. We model costly strategic learning when schools are ex ante symmetric, agree on their ranking of students, and learning is rationally inattentive. Our analytic solution quantifies how each student's rank, learning costs and prior beliefs interact to determine their gross and net welfare as well as the extent and form of mistakes they make. In line with the evidence, we find that lower-ranked students are affected disproportionately more by information costs, generally suffering a larger welfare loss than higher-ranked students. Interactions between mechanism design, inattention and inequity are thus of first order importance.

JEL classification: C78, D47, D82, D83

Keywords: Information Acquisition, School Choice, Matching, Deferred Acceptance, Rational Inattention

1 Introduction

Matching mechanisms are in widespread use, prominently in centralized school allocation systems worldwide (Neilson, 2019). The Deferred Acceptance (DA) algorithm of Gale and Shapley (1962) is particularly widely used, due in part to its theoretically attractive properties. Yet these properties rest on the assumption that agents know everything about all schools they are asked to rank, which is at odds with empirical evidence. Of particular concern is the observation that not all students are equally informed about their options, so that incomplete information can amplify inequity. Students from disadvantaged backgrounds are often unaware of their options (Hoxby and Avery, 2013) and targeted information interventions can significantly improve

*Department of Economics, New York University, New York, NY 10012, U.S.A. E-Mail: stefan.bucher@nyu.edu

†The authors wish to thank Adam Brandenburger, Ambuj Dewan, Octavia Ghelfi, Sam Kapon, Dmitry Sorokin, and seminar participants at NYU, the Virtual Behavioral Economics Seminar (VIBES), Universidad de Alicante, and the 2020 Workshop on the Cognitive Foundations of Economic Behavior. We would like to thank the NOMIS Foundation and the Alfred P. Sloan Foundation for their support. Portions of this paper appeared in Stefan Bucher's PhD dissertation at NYU. All remaining errors are our own.

the quality of school they attend (Hastings and Weinstein, 2008; Hoxby and Turner, 2015). Presciently, Roth and Sotomayor (1990) early on identified modeling of information acquisition as an important next step in the analysis of matching mechanisms. Recent years have seen increased interest in and gradual progress towards this goal (Bade, 2015; Immorlica et al., 2020; Artemov, 2021; Chen and He, 2021a,b).

The challenge faced by matching models with endogenous information is that students face three sources of uncertainty: signal-based, deriving from uncertainty about what information their learning strategy will produce; strategic, deriving from uncertainty about others' submissions and thus the resulting matching outcome; and value-based, referring to the remaining uncertainty about the student's valuation of their match.

We introduce a tractable model of strategically rational inattention in a matching market that parsimoniously captures this complexity. To focus on the interplay with inequity we assume that schools agree on their ranking of students. For analytic tractability we assume that schools are ex ante symmetric (exchangeable) and that learning is rationally inattentive (Sims, 2003; Caplin and Dean, 2015; Matejka and McKay, 2015). While our symmetry assumption implies that schools are ex ante identical, it does not require that students' valuations are independent across schools so that information on a school can update beliefs about others.

We solve the model analytically for any number of schools and students, for any exchangeable priors, and for heterogeneous marginal costs of learning. While not entirely simple, the solution is remarkably elementary given the subtle interplay of forces. It permits us to quantify how each student's rank, learning costs and prior beliefs interact to determine their gross and net welfare as well as the extent and form of mistakes they make.

A central finding is that DA exacerbates inequity. Lower-ranked students attain a lower fraction of their net welfare surplus under full information than do higher-ranked students, even if they have the same costs of learning. This is because lower-ranked students face greater uncertainty about the outcome resulting from any submission, which disperses and often dilutes their incentive to acquire information. This is in addition to, and exacerbates, any effect stemming from lower-ranked students having less ready access to information, as may be realistic. While students' net utilities respect their ranking, this need not be the case for their outcomes. It may be optimal for a lower-ranked student to invest more in learning than a higher-ranked student, in order to hedge for the higher strategic risk they face.

Three key aspects of our model allow us to solve it analytically. First, given their exchangeable priors students endogenously acquire symmetric information (cf. Roth and Rothblum, 1999), so that in equilibrium each student's ex ante probability of being matched with any particular school depends only on the school's position in the student's submitted list ("list-position sufficiency"). Each student is most likely (at least weakly) to be matched with their top-listed school, then with their second-listed school, and so on. Students' incentive to learn their true preference is fully characterized by their probability distribution of being matched with the school in each position of their list. List-position sufficiency results from the exchangeability of prior beliefs, their independence across students, and anonymity of DA, and hence holds more generally.

The second aspect of the solution is specific to the mechanism. In our setting with schools' ranking of students being unanimous, DA amounts to serial dictatorship (Kojima and Manea, 2010; Morrill, 2013). The first-ranked student is always matched with their top-listed school, with the second and each subsequent student then able to understand precisely the probability with which each school will still be available to them, and hence the likelihood that they will receive the school at each position in their list. The fact that lower-ranking students are more likely to be matched with a school further down their list results in very unequal learning incentives that policy can do little to address, since it is based on properties of the mechanism. Other mechanisms may incentivize information acquisition more equally, as we illustrate in an example for the Boston mechanism. The equalization of learning incentives in the Boston mechanism relative to DA reduces inequity.

The third feature of our model that makes the solution simple is that students' learning costs are linear in the expected reduction in Shannon entropy between their prior and posterior beliefs, as in standard rational inattention theory. Our exchangeable case allows us to apply the methods of Matejka and McKay (2015) in their simplest form, producing state-dependent choice probabilities of the undistorted multinomial logit form. Our approach is particularly amenable to rational inattention in strategic settings (Martin, 2017; Denti, 2020; Ravid, 2020; Bloedel and Segal, 2021). While the entropy-based cost function simplifies formulae, we conjecture that many of the key properties of the solution would generalize to a broader class of cost functions. By way of illustration, we revisit our leading example for the case of optimal sequential search. While the sequential search solution qualitatively confirms our findings, it is harder to generalize.

This paper contributes to a growing literature on endogenous information acquisition in matching markets. In a setting with partition-based learning, Bade (2015) shows that among all non-bossy strategy-proof mechanisms, sequential serial dictatorship is uniquely optimal in incentivizing efficient information acquisition. Artemov (2021) provides a mild condition under which fewer students acquire information in equilibrium than socially optimal. Immorlica et al. (2020) shows that when learning is costly, no matching mechanism can generally guarantee a stable outcome while inducing efficient information acquisition. Chen and He (2021a,b) illustrate differential learning incentives under DA and the Boston mechanism.

Section 2 introduces our model. Section 3 illustrates the key features of our solution and the structure of the equilibrium in a small-scale example. Section 4 provides the analytic solution to the general model and characterizes strategic uncertainty. Section 5 characterizes patterns of mistaken decisions and provides formulae for each student's gross and net welfare in equilibrium. Section 6 analyzes the potential impact of cost-reducing policies in reducing the inequity resulting from DA, illustrates how other mechanisms may be more effective in reducing inequity, and revisits the example of Section 3 with sequential search. Section 7 concludes. The full solution of the sequential search model is in Appendix B, all proofs are in Appendix C.

2 Model and Symmetry

A finite number $N = c \cdot M$ of students $i \in I = \{1, \ldots, N\}$ is matched with schools $x \in X$ in a finite set of cardinality $|X| = M$, each of capacity $c \in \mathbb{N}$, using the Deferred Acceptance (DA) algorithm.[1] Each student i submits a rank-order list $l^i \in A^i$ to the matching mechanism, a permutation of schools with $l^i(m) \in X$ indicating the school listed at position m. We assume that it is common knowledge that all schools submit the same list to the matching mechanism, according to which students are ranked in strict order given by their index i (so that student 1 is top-ranked and thus has the highest priority). This unanimous ranking of students could reflect their scores on a standardized admissions test, for example, or the random outcome of a single tie-breaking rule. When submitting list l^i, student i is matched with school $\boldsymbol{x}^i_{DA}(l^i, \boldsymbol{l}^{-i}) \in X$ which is the deterministic matching resulting from DA when the lists submitted by all other students are $\boldsymbol{l}^{-i} \in \boldsymbol{A}^{-i} \equiv \times_{j \neq i} A^j$. The utility student i derives from being matched with a school $x \in X$ is determined by their type $\theta^i \in \Theta^i \equiv V^M$, a vector of length M whose elements – denoted by θ^i_x – are the type's valuations of each school $x \in X$, which are in the finite set $V \subset \mathbb{R}$ of possible utility levels.[2] Students' types θ^i are drawn independently of each other from their priors $\mu^i \in \Delta(\Theta^i)$. Let $\boldsymbol{\Theta}^{-i} \equiv \times_{j \neq i} \Theta^j$. Students do not know their types ex ante; rather, they can resolve some of the uncertainty about their own (but not others') type (i.e. about their valuation of schools) by acquiring a costly signal before submitting their list. They choose an arbitrary (finite) information structure before observing a signal realization and submitting a rank-order list. The cost of an information structure is linear in its expected reduction of Shannon entropy.

Each student's equilibrium strategy will reflect the three-stage lottery they face: First, they face *signal uncertainty* about the outcome of their chosen information structure. Second, the matching outcome resulting from a submitted list is stochastic, due to the uncertainty about other students' submissions. We will refer to this as *strategic uncertainty*. Note that while DA's strategy-proofness implies that strategic uncertainty does not affect a student's list submission strategy, it can affect their information acquisition strategy. What is worth learning for a given student depends on how likely they are to be matched with each school on their submitted list. Characterizing this distribution is combinatorically challenging even in the simplest of cases, which would be compounded when relaxing independence across students, as observing a signal would permit inference about what others likely learn and thus schools' availability. Assuming that students' types are drawn independently of each other allows us to abstract from this complication as students have no way of learning – directly or indirectly – about others' types. Third, students' endogenously imperfect information results in remaining *value uncertainty* about the utility they obtain from the school with which they are matched. In equilibrium, each student bears these types of uncertainty in mind when choosing the extent of the effort they put into understanding schools. Solving the model thus requires us to identify each student's probability, in equilibrium, of being matched with each school on their list. This will further allow us to specify each student's probability distribution, as a function of their rank and their prior valuation of

[1] We consider the one-shot implementation of the student-proposing DA.

[2] The notation θ^i_x implies an ordering of the elements in X, which is arbitrary and assumed where convenient for brevity.

schools, over the set of feasible rank-order list submissions, as well as their posterior belief when submitting each possible list.

2.1 Exchangeable Prior Beliefs

As we focus on differences between students, we impose strong symmetry on each student's prior belief about their valuation of schools, assuming that it is *exchangeable* in the sense of de Finetti (cf. Kreps, 1988, pp. 145), i.e. for any permutation $\alpha : X \to X$ of schools

$$\mu^i\left(\theta_1^i, \ldots, \theta_M^i\right) = \mu^i\left(\theta_{\alpha(1)}^i, \ldots, \theta_{\alpha(M)}^i\right). \tag{1}$$

Exchangeability requires that all permutations of the elements of θ^i are equiprobable. This condition is more general than assuming that utilities are independently and identically distributed across schools. It will also be preserved in the remaining capacities of schools, while independence is not. Note also that priors can be heterogeneous, so long as they are all exchangeable.

The assumption of exchangeability allows for essentially arbitrary Bayesian updating so that a signal about a student's valuation of one school can carry information about their valuation of other schools. In models of costly sequential search this type of updating becomes intractable quickly. For example, it can imply that there is no reservation strategy: observing a school of high utility can imply that utilities are generally high and justify further search, while observing a school of low utility can imply the converse. Rational inattention and exchangeability, on the other hand, are highly complementary as we now demonstrate.

2.2 Symmetric Decision Problems

Exchangeable priors give rise to a strong form of symmetry in students' decision problems that we now introduce for general decision problems (μ, A, u) consisting of a prior $\mu \in \Delta(\Omega)$ over a finite state space Ω, a choice set A, and a utility function $u : A \times \Omega \to \mathbb{R}$.

Definition 1. Define the partition $\{\Omega_k\}_{1\le k\le K}$ of the state space, with $\omega, \omega' \in \Omega_k \subseteq \Omega$ for some k if and only if there exists a bijection $\alpha_{\omega\omega'} : A \to A$ such that

$$u(a, \omega) = u(\alpha_{\omega\omega'}(a), \omega') \qquad \forall a \in A. \tag{2}$$

A decision problem (μ, A, u) is **symmetric** if, for any k,

$$\mu(\omega) = \mu(\omega') \qquad \forall \omega, \omega' \in \Omega_k \tag{3}$$

and, for any $a, b \in A$ and k, there exists a bijection $\pi_{ab} : \Omega_k \to \Omega_k$ such that

$$u(a, \omega) = u(b, \pi_{ab}(\omega)) \qquad \forall \omega \in \Omega_k. \tag{4}$$

This definition states that, within each subset Ω_k of the partition, each state is equally likely and the utilities of each pair of actions and each pair of states are permutations of each other. The critical simplification is that symmetry of a decision problem ensures that the corresponding rational inattention problem has a particularly simple solution. In stating this result it is convenient to solve for optimal information acquisition strategies via choice of Bayes-consistent stochastic choice data

$$P \in \mathcal{P}(\mu, A) \equiv \left\{ P \in \Delta(A \times \Omega) | \sum_{a \in A} P(a, \omega) = \mu(\omega) \, \forall \omega \in \Omega \right\} \tag{5}$$

as is standard in the literature (Matejka and McKay, 2015; Denti, 2020). The general rational inattention problem with Shannon mutual information costs is given by

$$\max_{P \in \mathcal{P}(\mu, A)} \sum_{a \in A} \sum_{\omega \in \Omega} P(a, \omega) u(a, \omega) - K(\kappa, P) \tag{6}$$

where

$$K(\kappa, P) \equiv \kappa \sum_{a \in A} [H(\mu) - H(\gamma^a)] \sum_{\omega \in \Omega} P(a, \omega) \tag{7}$$

with the Shannon entropy $H(p) = -\sum_{\omega \in \Omega} p(\omega) \ln(p(\omega))$ and the posterior belief

$$\gamma^a(\omega) \equiv \frac{P(a, \omega)}{\sum_{\nu \in \Omega} P(a, \nu)}. \tag{8}$$

We will focus on strategies reflecting the decision problem's symmetry.

Definition 2. A strategy $P \in \mathcal{P}(\mu, A)$ is symmetric if for all $a, b \in A$

$$\sum_{\omega \in \Omega} P(a, \omega) = \sum_{\omega \in \Omega} P(b, \omega). \tag{9}$$

The symmetric solution to symmetric rational inattention problems reduces to conditional choice probabilities of the (undistorted) multinomial logit form of Matejka and McKay (2015).

Proposition 1. *If a decision problem (μ, A, u) is symmetric, then the unique symmetric solution of the rational inattention problem (eq. 6) satisfies*

$$P(a|\omega) \equiv \frac{P(a, \omega)}{\mu(\omega)} = \frac{z(a, \omega)}{\sum_{c \in A} z(c, \omega)} \qquad \forall a \in A, \omega \in \Omega \tag{10}$$

where $z(a, \omega) \equiv \exp(u(a, \omega)/\kappa)$. If the $z(a)$ are affine independent, this is the unique solution (Matejka and McKay, 2015; Caplin et al., 2019).

3 An Example

We will show in Section 4 that in equilibrium, others' symmetric equilibrium strategies cause each student's problem to be symmetric, so that they have a unique symmetric best response. In this section we illustrate the resulting solution in an example, which demonstrates how each student's symmetric solution results in lower-ranked students' problems being symmetric, so that symmetry is inherited. This implies that in equilibrium each student is unconditionally equally likely to match with any of the schools, which will permit a combinatoric answer to the question of how likely each school is to be available for ensuing students.

The example of this section involves three students $i \in \{1, 2, 3\}$ who are to be matched by DA with three schools $x \in \{a, b, c\}$ of unit capacity. Each student values schools as either good or bad, receiving utility $u_G > 0$ from matching with a good school and $u_B = 0$ from a bad one. For simplicity, we assume that all students face the same marginal cost κ of information. Since the schools unanimously rank the students in order of their index, the matching resulting from DA is characterized by serial dictatorship (see Proposition 4 below), so that each student is matched with the first school on their submitted rank-order list that has not been assigned to a higher-ranking student. We first solve for the Nash equilibrium of the resulting game between students whose strategies consist of type-dependent stochastic choice data (Matejka and McKay, 2015). The solution permits us to analyze key features of the equilibrium: what each student learns and what type of mistakes they make; inequities in the gross and net welfare they attain; and the potential for cost-reducing policies to reduce inequity. Section 4 will generalize these analyses to arbitrary numbers of schools and students, general utilities, and heterogeneous information costs.

3.1 Solving for Students' Optimal Submissions

Note first that there are eight states of the world defined by the quality in sequential order of schools a, b, and c. Since each student i has an exchangeable prior μ^i, prior probabilities are a function only of the cardinality of the set of good schools,

$$\mu^i(GGB) = \mu^i(GBG) = \mu^i(BGG) =: \mu^i(2) \geq 0 \tag{11}$$

$$\mu^i(GBB) = \mu^i(BGB) = \mu^i(BBG) =: \mu^i(1) \geq 0 \tag{12}$$

We define $\mu^i(GGG) =: \mu^i(3)$ and $\mu^i(BBB) =: \mu^i(0)$ accordingly. The only additional restriction that exchangeability imposes apart from non-negativity is that $\mu^i(3) + 3\mu^i(2) + 3\mu^i(1) + \mu^i(0) = 1$. A student's expected utility of submitting a rank-order list depends on their type θ^i, but also on their rank i. Since, by serial dictatorship, student 1 receives their first-listed school with certainty, their utility of submitting list l^1 is given by

$$u^1(l^1, \theta^1) = \theta^1_{l^1(1)} \tag{13}$$

when their type is θ^1. Ex ante, student 1 thus only faces the *signal uncertainty* associated with the random realization from their chosen signal structure, and the *value uncertainty* referring to the remaining uncertainty

$l^1 \backslash \theta^1$	GGG	GGB	GBG	BGG	GBB	BGB	BBG	BBB
$(a,*,*)$	$\frac{z}{6z}$	$\frac{z}{4z+2}$	$\frac{z}{4z+2}$	$\frac{1}{4z+2}$	$\frac{z}{2z+4}$	$\frac{1}{2z+4}$	$\frac{1}{2z+4}$	$\frac{1}{6}$
$(b,*,*)$	$\frac{z}{6z}$	$\frac{z}{4z+2}$	$\frac{1}{4z+2}$	$\frac{z}{4z+2}$	$\frac{1}{2z+4}$	$\frac{z}{2z+4}$	$\frac{1}{2z+4}$	$\frac{1}{6}$
$(c,*,*)$	$\frac{z}{6z}$	$\frac{1}{4z+2}$	$\frac{z}{4z+2}$	$\frac{z}{4z+2}$	$\frac{1}{2z+4}$	$\frac{1}{2z+4}$	$\frac{z}{2z+4}$	$\frac{1}{6}$

Table 1: Type-conditional choice probabilities $P^{1*}(l^1|\theta^1)$ of student 1, where $P^{1*}(a,*,*|\theta^1)$ is the probability of submitting *each* list that ranks school a first, i.e. (a,b,c) and (a,c,b).

in the resulting posterior about their valuation $\theta^1_{l^1(1)}$ of their top-listed school. Their learning problem is therefore a standard rational inattention problem which is symmetric as specified in Definition 1. The optimal symmetric strategy follows directly from the standard rationally inattentive solution of Matejka and McKay (2015) and is given in full in Table 1, where we define $z \equiv \exp(u_G/\kappa)$ for convenience.[3] Given the strong symmetry properties of the solution, we can effectively summarize all choice probabilities focusing on states in which the good schools are earlier in the alphabet than bad schools: if there is one good school it is school a, if there are two they are a, b. For each state we show the solution for a subset of rank-order lists which are "representative" for other lists that give rise to the same sequence of good and bad schools. The choice probabilities of Table 1 can thus be summarized as

$$P^{1*}(a,b,c|GGG) = \frac{z}{6z} \tag{14}$$

$$P^{1*}(a,b,c|GGB) = P^{1*}(a,c,b|GGB) = \frac{z}{4z+2} \tag{15}$$

$$P^{1*}(c,a,b|GGB) = \frac{1}{4z+2} \tag{16}$$

$$P^{1*}(a,b,c|GBB) = \frac{z}{2z+4} \tag{17}$$

$$P^{1*}(b,a,c|GBB) = P^{1*}(b,c,a|GBB) = \frac{1}{2z+4} \tag{18}$$

$$P^{1*}(a,b,c|BBB) = \frac{1}{6} \tag{19}$$

The solution has many noteworthy features. First, note that for student 1 only the quality of the top-ranked school matters for the probability with which a list is submitted. Second, $z \equiv \exp(u_G/\kappa)$ is a sufficient statistic for the extent to which the utility of picking a good school and costs of identifying such a school impact choice probabilities, as has been understood since Matejka and McKay (2015). Third, the simple way of understanding these probabilities in all cases is to note that the numerator reflects the transformed utility of the particular list submitted, while the denominator reflects the corresponding value added up across all lists. What this means is that the likelihood of identifying a particular school as good is higher the fewer such schools there are: With $z = 2$, for example, there is a 50% probability of listing the good school first if exactly one school is good. If exactly two schools are good, each of these is 40% likely to be top-ranked.

[3] The symmetric equilibrium restricts attention to strategies under which submissions with equivalent expected utility are equally likely.

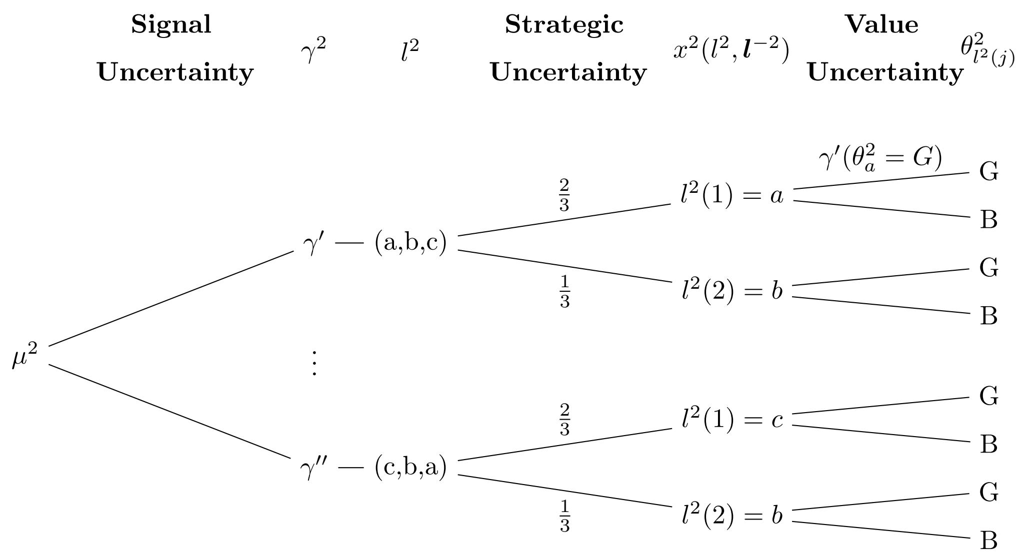

Figure 1: Posterior-based illustration of the three-stage lottery faced by student 2. Having chosen an information structure given their prior belief μ^2, student 2 faces uncertainty about the realization of the signal and the posterior belief γ^2 it induces. Given their posterior belief, they submit a utility-maximizing rank-order list l^2 to the mechanism, which results in an outcome that is stochastic as it depends on the submission of student 1. There is remaining uncertainty about their valuation of their matched school.

Fourth, and most strikingly, these conditional probabilities apply regardless of the structure of the prior, provided it is exchangeable. The same formulae thus characterize conditional choice probabilities regardless of what learning about one school might imply about the probability of the other schools being good. This contrasts profoundly with models of sequential search, in which the realized utilities of early schools can create all kinds of intricacies in the subsequent search patterns, making conditional choice probabilities in the optimal strategy strongly path dependent and far from invariant to the structure of the prior. Finally, note that the *un*conditional choice probabilities of student 1 are uniform, so ex ante they are equally likely to be matched with any of the three schools. This is critical for student 2, since whichever school student 1 is matched with will not be available to them.

In addition to value and signal uncertainty, all students other than student 1 also face *strategic uncertainty* about the outcome resulting from the list they submit. Figure 1 illustrates this for student 2 using the (equivalent) posterior-based approach for interpretability: First, nature draws a signal from the information structure chosen by the student, which results in a posterior belief $\gamma^2 \in \Delta(\Theta^2)$ about their type θ^2 under which they will submit a utility-maximizing rank-order list l^2. Second, the student faces strategic uncertainty about the outcome resulting from their submitted list. Third, there is value uncertainty remaining about the utility of their realized outcome. By choosing an information structure, the student determines how much value uncertainty will be remaining. Note that the choice of information structure is implicit in our formulation, as is now standard in the literature. Since their types are independent, student 2 cannot infer anything about the type of student 1 from what they learn about their own type. From the perspective of student 2 each of the three schools is thus equally likely to still be available, hence they will be matched with their top-ranked school with probability 2/3, and their second-ranked school with probability 1/3. For student 2 the resulting expected utility of submitting a list l^2 when their true type is θ^2 is thus

$$u^2(l^2, \theta^2) = \frac{2}{3}\theta^2_{l^2(1)} + \frac{1}{3}\theta^2_{l^2(2)}. \tag{20}$$

The strategic uncertainty about which of the schools on their list student 2 is matched with affects their learning incentives, which is reflected in their type-conditional choice probabilities. Due to the strong symmetry properties of the solution, the choice probabilities (given in full in Table 4 in the appendix) can again be summarized by focusing on states in which the good schools are earlier in the alphabet than the bad schools. The choice probabilities in such cases are:

$$P^{2*}(a, b, c|GGG) = \frac{z}{6z} \tag{21}$$

$$P^{2*}(a, b, c|GGB) = \frac{z}{2z + 2z^{\frac{2}{3}} + 2z^{\frac{1}{3}}} \tag{22}$$

$$P^{2*}(a, c, b|GGB) = \frac{z^{\frac{2}{3}}}{2z + 2z^{\frac{2}{3}} + 2z^{\frac{1}{3}}} \tag{23}$$

$$P^{2*}(c, a, b|GGB) = \frac{z^{\frac{1}{3}}}{2z + 2z^{\frac{2}{3}} + 2z^{\frac{1}{3}}} \tag{24}$$

$$P^{2*}(a, b, c|GBB) = \frac{z^{\frac{2}{3}}}{2z^{\frac{2}{3}} + 2z^{\frac{1}{3}} + 2} \tag{25}$$

$$P^{2*}(b, a, c|GBB) = \frac{z^{\frac{1}{3}}}{2z^{\frac{2}{3}} + 2z^{\frac{1}{3}} + 2} \tag{26}$$

$$P^{2*}(b, c, a|GBB) = \frac{1}{2z^{\frac{2}{3}} + 2z^{\frac{1}{3}} + 2} \tag{27}$$

$$P^{2*}(a, b, c|BBB) = \frac{1}{6} \tag{28}$$

Student 2's optimal strategy has the same qualitative features and essentially the same simplicity and separability properties as that of student 1. Again $z \equiv \exp(u_G/\kappa)$ is a sufficient statistic for the impact of learning costs as well as the utility differential between schools on patterns of mistakes. As was the case for student 1, student 2 is more likely to submit better lists for higher values of z. The conditional probabilities again apply regardless of the structure of the prior and the unconditional choice probabilities of student 2 are also uniform, so ex ante they are equally likely to be matched with any of the three schools. The solution has one more striking feature related to strategic uncertainty. Student 2 is less likely than student 1 to submit a good list because their learning incentive is diluted by the strategic uncertainty, which is reflected in the exponents of z. The formulae reveal in a precise and simple manner how the strategic uncertainty impacts what is learned. If only the first school on a list is good, z is taken to the power of 2/3 to reflect the fact that the first school is received with this probability. If only the second school on a list is good, the exponent 1/3 reflects the fact that the second school is received with this probability. If both are good, z has exponent 1 since student 2 will find themselves at a good school regardless of how the strategic uncertainty resolves.

Finally, the exponent is zero when the first two are both bad. As for student 1, the numerator can be seen as the value of the list in question, while the denominator is the total value of all lists in the given state, which depends only on the number of good schools in that state. As we will see, these findings apply equally in the general case. Finally, note that there is an even stronger form of symmetry here, as identifying the one good school gives rise to the same conditional probabilities (after cancelling) as identifying the one bad school.

Since whichever schools students 1 and 2 match with will, by serial dictatorship, not be available to student 3, their submission is inconsequential, so they optimally acquire no information and are indifferent between lists.

3.2 Value Uncertainty in Students' Posterior Beliefs

Beyond specifying students' type-conditional choices, solving the model analytically allows us to analyze the form and likelihood of the mistakes each student makes. This is revealed by their posterior beliefs $P^{i*}(\theta^i|l^i) \equiv P^{i*}(l^i,\theta^i)/\sum_{\theta\in\Theta^i} P^{i*}(l^i,\theta)$ when submitting any list, which are shown in Table 2. Students' posterior beliefs exhibit a strong form of symmetry: They depend only on the order of good and bad schools in the submitted list, but not on the schools' identities. Due to this symmetry, the nature of students' posteriors is best summarized through the value uncertainty, which is captured by the probability

$$p^i(m) \equiv \sum_{\theta^i\in\Theta^i} P^{i*}(\theta^i|l^i)\mathbf{1}\left\{\theta^i_{l^i(m)} = u_G\right\} \tag{29}$$

that the m-th listed school on any list l^i submitted by student i is a school whose valuation is good under their true type. This probability is independent of l^i due to the posteriors' symmetry.

Student 1 does not face any strategic uncertainty and their value uncertainty is standard as in any individual rational inattention model. The probability that the top-listed school (the only possible outcome) is good for student 1 is

$$p^1(1) = \mu^1(3) + 3\mu^1(2)\frac{2z}{2z+1} + 3\mu^1(1)\frac{z}{z+2} \tag{30}$$

which is increasing in z. For student 3 the strategic uncertainty is maximal, as they face a 1/3 probability of matching with any of the schools regardless of their submission. Student 3 thus will not learn anything, so their value uncertainty will be defined by the prior irrespective of information costs,

$$p^3(1) = p^3(2) = p^3(3) = \mu^3(3) + 2\mu^3(2) + \mu^3(1) =: p^3 \tag{31}$$

More interesting are the mistakes made by student 2 who faces non-trivial strategic uncertainty (unlike student 1) and non-trivial signal uncertainty (unlike student 3) and thus resembles the generic student in our general model. The nature of the mistakes depends on how the strategic uncertainty resolves: When being matched with the first school on their list, the probability of that school being good is

$$p^2(1) = \mu^2(3) + 3\mu^2(2)\frac{z + z^{\frac{2}{3}}}{z + z^{\frac{2}{3}} + z^{\frac{1}{3}}} + 3\mu^2(1)\frac{z^{\frac{2}{3}}}{z^{\frac{2}{3}} + z^{\frac{1}{3}} + 1} \tag{32}$$

$l^1 \backslash \theta^1$	GGG	GGB	GBG	BGG	GBB	BGB	BBG	BBB
$(a,*,*)$	$\mu^1(3)$	$\frac{3z}{2z+1}\mu^1(2)$	$\frac{3z}{2z+1}\mu^1(2)$	$\frac{3}{2z+1}\mu^1(2)$	$\frac{3z}{z+2}\mu^1(1)$	$\frac{3}{z+2}\mu^1(1)$	$\frac{3}{z+2}\mu^1(1)$	$\mu^1(0)$
$(b,*,*)$	$\mu^1(3)$	$\frac{3z}{2z+1}\mu^1(2)$	$\frac{3}{2z+1}\mu^1(2)$	$\frac{3z}{2z+1}\mu^1(2)$	$\frac{3}{z+2}\mu^1(1)$	$\frac{3z}{z+2}\mu^1(1)$	$\frac{3}{z+2}\mu^1(1)$	$\mu^1(0)$
$(c,*,*)$	$\mu^1(3)$	$\frac{3}{2z+1}\mu^1(2)$	$\frac{3z}{2z+1}\mu^1(2)$	$\frac{3z}{2z+1}\mu^1(2)$	$\frac{3}{z+2}\mu^1(1)$	$\frac{3}{z+2}\mu^1(1)$	$\frac{3z}{z+2}\mu^1(1)$	$\mu^1(0)$

$l^2 \backslash \theta^2$	GGG	GGB	GBG	BGG	GBB	BGB	BBG	BBB
(a,b,c)	$\mu^2(3)$	$\frac{6z\mu^2(2)}{Z^2(2)}$	$\frac{6z^{\frac{2}{3}}\mu^2(2)}{Z^2(2)}$	$\frac{6z^{\frac{1}{3}}\mu^2(2)}{Z^2(2)}$	$\frac{6z^{\frac{2}{3}}\mu^2(1)}{Z^2(1)}$	$\frac{6z^{\frac{1}{3}}\mu^2(1)}{Z^2(1)}$	$\frac{6\mu^2(1)}{Z^2(1)}$	$\mu^2(0)$
(a,c,b)	$\mu^2(3)$	$\frac{6z^{\frac{2}{3}}\mu^2(2)}{Z^2(2)}$	$\frac{6z\mu^2(2)}{Z^2(2)}$	$\frac{6z^{\frac{1}{3}}\mu^2(2)}{Z^2(2)}$	$\frac{6z^{\frac{2}{3}}\mu^2(1)}{Z^2(1)}$	$\frac{6\mu^2(1)}{Z^2(1)}$	$\frac{6z^{\frac{1}{3}}\mu^2(1)}{Z^2(1)}$	$\mu^2(0)$
(b,a,c)	$\mu^2(3)$	$\frac{6z\mu^2(2)}{Z^2(2)}$	$\frac{6z^{\frac{1}{3}}\mu^2(2)}{Z^2(2)}$	$\frac{6z^{\frac{2}{3}}\mu^2(2)}{Z^2(2)}$	$\frac{6z^{\frac{1}{3}}\mu^2(1)}{Z^2(1)}$	$\frac{6z^{\frac{2}{3}}\mu^2(1)}{Z^2(1)}$	$\frac{6\mu^2(1)}{Z^2(1)}$	$\mu^2(0)$
(b,c,a)	$\mu^2(3)$	$\frac{6z^{\frac{2}{3}}\mu^2(2)}{Z^2(2)}$	$\frac{6z^{\frac{1}{3}}\mu^2(2)}{Z^2(2)}$	$\frac{6z\mu^2(2)}{Z^2(2)}$	$\frac{6\mu^2(1)}{Z^2(1)}$	$\frac{6z^{\frac{2}{3}}\mu^2(1)}{Z^2(1)}$	$\frac{6z^{\frac{1}{3}}\mu^2(1)}{Z^2(1)}$	$\mu^2(0)$
(c,a,b)	$\mu^2(3)$	$\frac{6z^{\frac{1}{3}}\mu^2(2)}{Z^2(2)}$	$\frac{6z\mu^2(2)}{Z^2(2)}$	$\frac{6z^{\frac{2}{3}}\mu^2(2)}{Z^2(2)}$	$\frac{6z^{\frac{1}{3}}\mu^2(1)}{Z^2(1)}$	$\frac{6\mu^2(1)}{Z^2(1)}$	$\frac{6z^{\frac{2}{3}}\mu^2(1)}{Z^2(1)}$	$\mu^2(0)$
(c,b,a)	$\mu^2(3)$	$\frac{6z^{\frac{1}{3}}\mu^2(2)}{Z^2(2)}$	$\frac{6z^{\frac{2}{3}}\mu^2(2)}{Z^2(2)}$	$\frac{6z\mu^2(2)}{Z^2(2)}$	$\frac{6\mu^2(1)}{Z^2(1)}$	$\frac{6z^{\frac{1}{3}}\mu^2(1)}{Z^2(1)}$	$\frac{6z^{\frac{2}{3}}\mu^2(1)}{Z^2(1)}$	$\mu^2(0)$

Table 2: Posterior beliefs $P^{i*}(\theta^i|l^i)$ of student 1 (top) and 2 (bottom), with $Z^2(2) = 2z + 2z^{\frac{2}{3}} + 2z^{\frac{1}{3}}$ and $Z^2(1) = 2z^{\frac{2}{3}} + 2z^{\frac{1}{3}} + 2$. $P^{1*}(\theta^1|a,*,*)$ denotes the posterior belief of student 1 when submitting *any* list that lists school a first.

which is increasing in z, as was the case for student 1 above. The pattern of mistakes is more interesting if the strategic uncertainty results in student 2 being matched with their second-listed school, whose probability of being good is

$$p^2(2) = \mu^2(3) + 3\mu^2(2)\frac{z + z^{\frac{1}{3}}}{z + z^{\frac{2}{3}} + z^{\frac{1}{3}}} + 3\mu^2(1)\frac{z^{\frac{1}{3}}}{z^{\frac{2}{3}} + z^{\frac{1}{3}} + 1}. \tag{33}$$

Note that this is increasing in z if and only if $\mu^2(2) > \mu^2(1)$, so that the pattern of mistakes depends on the structure of the prior. In a tracking problem in which exactly one of the schools is good (so that $\mu^2(1) = 1/3 > 0 = \mu^2(2)$), for example, this probability decreases when information costs are reduced, as the student is increasingly successful in identifying the one good school and ranking it first ("cherry picking"). Conversely, if $\mu^2(2) = 1/3$ so that there is exactly one bad school ("lemon dropping"), this probability increases as information costs are reduced. This is because the student will be more successful in identifying the one bad school and ranking it last.

3.3 Gross Welfare and the Value of Information

The strategic and value uncertainty jointly determine students' ex ante probability of being matched with a good school.[4] For student 1 this probability is simply given by $p^1(1)$, whereas for student 2 it is

$$\frac{2}{3}p^2(1) + \frac{1}{3}p^2(2) = \mu^2(3) + \mu^2(2) + (\mu^2(2) + \mu^2(1))\frac{2z + z^{\frac{2}{3}}}{z + z^{\frac{2}{3}} + z^{\frac{1}{3}}}. \tag{34}$$

Note that both students' ex ante probability of a good outcome is always increasing in z regardless of the prior (provided that not all schools yield the same utility with certainty).

We can quantify how students' ex ante welfare and the value of information in equilibrium depend on their rank (which determines how much strategic uncertainty they face) and on their information costs. First, note that under their prior belief, a student always matches with a good school when all schools are good, matches with a good school with probability 2/3 when two schools are good, and with probability 1/3 when only one school is good. In the case of no information, all students thus attain expected utility

$$\underline{U}(\mu^i) \equiv \left(\mu^i(3) + 2\mu^i(2) + \mu^i(1)\right) u_G \tag{35}$$

regardless of their list submission. By acquiring information, students raise their ex ante gross welfare $U^{i*}(z, \mu^i)$ above this level, to an extent that reflects the value of their equilibrium information structure which depends on their rank i. The resulting value of information (Frankel and Kamenica, 2019) to each student in equilibrium is summarized in the following Proposition.

Proposition 2. *The value of students' information structure in equilibrium is*

$$U^{1*}(z, \mu^1) - \underline{U}(\mu^1) = \left(\frac{2z-2}{2z+1}\mu^1(2) + \frac{2z-2}{z+2}\mu^1(1)\right) u_G \tag{36}$$

$$U^{2*}(z, \mu^2) - \underline{U}(\mu^2) = \frac{z - z^{\frac{1}{3}}}{z + z^{\frac{2}{3}} + z^{\frac{1}{3}}}\left(\mu^2(2) + \mu^2(1)\right) u_G \tag{37}$$

and 0 *for student 3.*

[4] The signal uncertainty does not impact the probability of a good outcome, due to the symmetry of the posteriors.

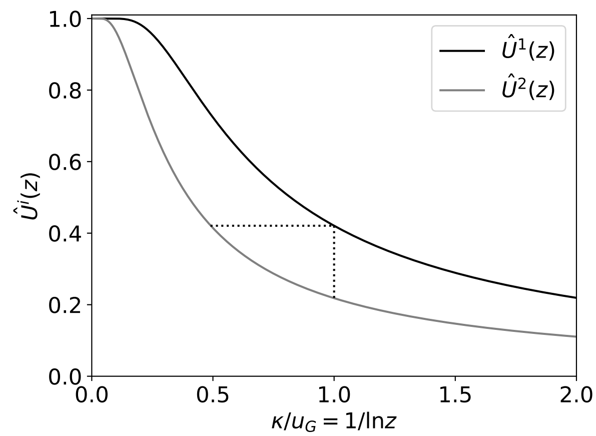

Figure 2: Relative gross welfare of students 1 and 2 as a function of $\kappa/u_G = 1/\ln(z)$, for student 1 priors satisfying $\mu^1(2) = \mu^1(1)$ and any student 2 priors.

The utility achieved under full information, given by

$$\overline{U}^1(\mu^1) \equiv \left(\mu^1(3) + 3\mu^1(2) + 3\mu^1(1)\right) u_G \tag{38}$$

$$\overline{U}^2(\mu^2) \equiv \left(\mu^2(3) + 3\mu^2(2) + 2\mu^2(1)\right) u_G, \tag{39}$$

constitutes an upper limit on each student's gross welfare. While student 1 will be matched with a good school whenever there is one, student 2 only has a 2/3 chance of being matched with the only good school. The fact that student 2 attains lower gross utility even under perfect information solely reflects the fact that schools prefer student 1 to student 2. Students' full information surplus

$$\overline{U}^1(\mu^1) - \underline{U}(\mu^1) = \left(\mu^1(2) + 2\mu^1(1)\right) u_G \tag{40}$$

$$\overline{U}^2(\mu^2) - \underline{U}(\mu^2) = \left(\mu^2(2) + \mu^2(1)\right) u_G \tag{41}$$

thus presents a natural benchmark with which to compare their value of information (which is itself a surplus). We thus define the relative gross welfare

$$\hat{U}^i(z, \mu^i) \equiv \frac{U^{i*}(z, \mu^i) - \underline{U}(\mu^i)}{\overline{U}^i(\mu^i) - \underline{U}(\mu^i)}, \tag{42}$$

which is plotted in Figure 2.

3.4 Net Welfare and the Cost of Information

Rather than analyzing students' utility without taking into account their information costs, a policy maker or planner may want to consider students' utility net of equilibrium information costs $K^{i*}(\kappa, \mu^i) \equiv K(\kappa, P^{i*})$,

$$N^{i*}(z, \mu^i) \equiv U^{i*}(z, \mu^i) - K^{i*}\left(\frac{u_G}{\ln(z)}, \mu^i\right). \tag{43}$$

Due to the symmetry in students' posteriors they all have the same entropy, so that computing students' information costs is straightforward. The resulting net utility is summarized in the following Proposition.

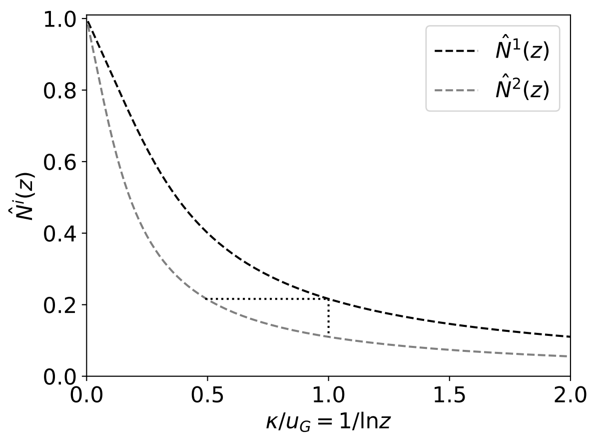

Figure 3: Relative net welfare of students 1 and 2 as a function of $\kappa/u_G = 1/\ln(z)$, for student 1 priors satisfying $\mu^1(2) = \mu^1(1)$ and any student 2 priors.

Proposition 3. *Students' net utility in equilibrium is*

$$N^{1*}(z, \mu^1) = \left[\mu^1(3)\ln z + 3\mu^1(2)\ln\left(\frac{2z+1}{3}\right) + 3\mu^1(1)\ln\left(\frac{z+2}{3}\right)\right]\frac{u_G}{\ln z} \tag{44}$$

$$N^{2*}(z, \mu^2) = \left[\mu^2(3)\ln z + 3\mu^2(2)\ln\left(\frac{z+z^{2/3}+z^{1/3}}{3}\right) + 3\mu^2(1)\ln\left(\frac{z^{2/3}+z^{1/3}+1}{3}\right)\right]\frac{u_G}{\ln z} \tag{45}$$

$$N^{3*}(z, \mu^3) = \left(\mu^3(3) + 2\mu^3(2) + \mu^3(1)\right)u_G \tag{46}$$

The expressions for the net utility exhibit the same kind of separability by prior as observed for the gross utilities of Proposition 2. As before, a meaningful comparison is again possible by computing students' relative net welfare, which is defined as

$$\hat{N}^i(z, \mu^i) \equiv \frac{N^{i*}(z, \mu^i) - \underline{N}(\mu^i)}{\overline{N}^i(\mu^i) - \underline{N}(\mu^i)} \tag{47}$$

where $\overline{N}^i(\mu^i) = \overline{U}^i(\mu^i)$ is the net utility when information is costless and $\underline{N}(\mu^i) = \underline{U}(\mu^i)$ is the net utility when information is infinitely costly (so that no information is acquired). The resulting relative net welfare is shown in Figure 3.

3.5 Inequity

Figure 2 illustrates how, at any marginal cost κ of information, student 2 attains a lower fraction than student 1 of their gross welfare surplus under full information, which is already smaller than that of student 1. At $\kappa = u_G$, for example, student 1 achieves $\hat{U}^1 \approx 42\%$ of their full information surplus. At the same cost of information, student 2 achieves only $\hat{U}^2 \approx 22\%$ of their full information surplus – a sizeable difference indicated by the vertical dotted line in Figure 2. An analyst merely examining students' behavior might be tempted to conclude that lower-ranked students face higher information costs, but here their behavior and outcome are solely due to their diluted incentives to acquire information under the mechanism.

The relative gross welfare of the two students facing the same information costs obeys the ranking

$$\frac{2z-2}{2z+1} \geq \hat{U}^1(z,\mu^1) \geq \frac{z-1}{z+2} > \hat{U}^2(z,\mu^2) > 0 \tag{48}$$

at any cost of information and regardless of their priors. As discussed above, the gross welfare surplus of student 3 is always zero. Lower-ranked students in this section's example are thus affected disproportionately more by information costs than higher-ranked students: The former attain a lower fraction $\hat{U}^i(z,\mu^i)$ of their already smaller gross welfare surplus under full information than the latter, which also implies that the value $U^{i*}(z,\mu^i) - \underline{U}(\mu^i)$ of their equilibrium information structure is lower, as is their gross welfare $U^{i*}(z,\mu^i)$ in absolute terms.

The fact that student 2 exerts less effort on information acquisition than student 1 does not change the ranking of their utilities: The relative net welfare is still lower for student 2 than for student 1, obeying the ordering

$$3\frac{\ln\left(\frac{2z+1}{3}\right)}{\ln(z)} - 2 \geq \hat{N}^1(z,\mu^1) \geq \frac{3\ln\left(\frac{z+2}{3}\right)}{2\ln(z)} - \frac{1}{2} > \hat{N}^2(z,\mu^2) > 0 \tag{49}$$

for any information costs and priors. At $\kappa = u_G$, for example, student 1 achieves $\hat{N}^1 \approx 22\%$ of their full information surplus while student 2 achieves only $\hat{N}^2 \approx 11\%$, as seen in Figure 3.

3.6 Information Policies

We have seen above that student 2 is disproportionately affected by costly information. This is because their incentive to acquire information is lower under DA than that of student 1. A policy maker to whom this type of inequity is of concern may thus want to address this informational disadvantage by leveling the informational playing field: Using the above solution, it is straightforward to determine the level $\kappa^2(\kappa^1)$ to which student 2's information costs have to be lowered in order for them to achieve the same fraction of their full information surplus as student 1 does at cost κ^1. The policy maker's goal would then be to find, for any z, the $\tilde{z}(z)$ such that

$$\hat{U}^2(\tilde{z}(z),\mu^2) = \hat{U}^1(z,\mu^1). \tag{50}$$

Returning to the example of Figure 2 where $\mu^1(2) = \mu^1(1)$, it is easy to see that for student 2 to achieve the fraction student 1 achieves at $\kappa^1 = u_G$,

$$\hat{U}^1(\exp(1),\mu^1) \approx 0.42 \approx \hat{U}^2(\exp(2.04),\mu^2) \tag{51}$$

their information cost would have to be reduced by more than half to $\kappa^2 = 0.49u_G$ as indicated by the horizontal dotted line in Figure 2. In order for student 2 to double their fraction of net welfare surplus from 11% to the 22% that student 1 achieves at $\kappa = u_G$,

$$\hat{N}^1(\exp(1),\mu^1) \approx 0.22 \approx \hat{N}^2(\exp(2.02),\mu^2) \tag{52}$$

their information cost would have to be reduced to $\kappa^2 = 0.495u_G$ as indicated by the horizontal dotted line in Figure 3.

4 Equilibrium

In this section we characterize the symmetric equilibrium of the general model. The game between students who acquire costly information structures before submitting a rank-order list can be expressed equivalently as a game in stochastic choice strategies (cf. Denti, 2020). This game consists of each student i choosing a stochastic choice strategy $P^i \in \mathcal{P}(\mu^i, A^i)$ so as to maximize, given others' strategies $\boldsymbol{P}^{-i} \equiv (P^1, \dots, P^{i-1}, P^{i+1}, \dots, P^N)$ where $P^j \in \mathcal{P}(\mu^j, A^j)$ for all $j \neq i$, expected utility

$$U^i(P^i, \boldsymbol{P}^{-i}) = \sum_{(l^i, \boldsymbol{l}^{-i}) \in A^i \times \boldsymbol{A}^{-i}} \sum_{(\theta^i, \boldsymbol{\theta}^{-i}) \in \Theta^i \times \boldsymbol{\Theta}^{-i}} P^i(l^i, \theta^i) \boldsymbol{P}^{-i}(\boldsymbol{l}^{-i}, \boldsymbol{\theta}^{-i}) \theta^i_{\boldsymbol{x}^i(l^i, \boldsymbol{l}^{-i})} \tag{53}$$

net of information costs $K(\kappa^i, P^i)$ as defined in equation 7. Note that $\boldsymbol{P}^{-i}(\boldsymbol{l}^{-i}, \boldsymbol{\theta}^{-i}) = \prod_{j \neq i} P^j(l^j, \theta^j)$ since students' types are independent, as are their simultaneous list submissions. This formulation of student i's rational inattention problem (cf. Matejka and McKay, 2015, Corollary 1) gives rise to the following equilibrium: $\boldsymbol{P}^* = (P^{1*}, \dots, P^{N*}) \in \times_{i \in I} \mathcal{P}(\mu^i, A^i)$ is a Nash equilibrium if for all $i \in I$

$$U^i(P^{i*}, \boldsymbol{P}^{-i*}) - K(\kappa^i, P^{i*}) \geq U^i(P^i, \boldsymbol{P}^{-i*}) - K(\kappa^i, P^i) \text{ for all } P^i \in \mathcal{P}(\mu^i, A^i). \tag{54}$$

It is symmetric if every strategy P^{i*} is symmetric according to Definition 2.

We focus on equilibria in symmetric strategies that reflect the fact that schools are ex ante identical under each student's prior belief. The central challenge in identifying the equilibrium is to characterize the strategic uncertainty that each student faces. This is rendered feasible by the observation that if others' strategies are symmetric, then a student's problem is symmetric in the sense of Definition 1 and thus has a unique symmetric solution itself (Proposition 1). In order to characterize the resulting symmetric equilibrium, we first show in Section 4.1 that if others' strategies are symmetric, then the probability that a student matches with a given school only depends on its position on the student's submitted list (a property we refer to as list-position sufficiency; see Lemma 1). Next, we establish that list-position sufficiency implies that each student's problem is symmetric (Lemma 2), which implies (by Proposition 1) that it has a unique symmetric solution. Given students' symmetric equilibrium strategies, we then find combinatoric expressions for each student's probability of being matched by DA with each school on their list, which fully characterizes the strategic uncertainty they face (Section 4.2). This permits us to describe the unique symmetric equilibrium analytically (Section 4.3). It is only at this last stage that the rationally inattentive cost function comes into play – most of the intricacy relates to the inescapable strategic uncertainty.

4.1 Symmetry and List-Position Sufficiency

The first key observation is that symmetry of others' strategies renders all lists equivalent in that a student's probability of being matched with a given school depends only on its position m in their submitted list, not its identity nor that of the other schools in the list. This holds true for a class of matching mechanisms beyond DA. In our setting, suppressing notation for the school side of the market, a matching mechanism $\boldsymbol{x}$

is a vector-valued function $\boldsymbol{x} : \times_{i \in I} A^i \to X^N$ such that, for all $x \in X$ and $l^1, \ldots, l^N \in \times_{i \in I} A^i$,

$$\sum_{i \in I} \mathbf{1}\{\boldsymbol{x}^i(l^1, \ldots, l^N) = x\} \leq c \tag{55}$$

where $\boldsymbol{x}^i(l^1, \ldots, l^N) \in X$ denotes the school $x \in X$ with which student $i \in I$ is matched when the submitted lists are $l^1, \ldots, l^N$. We define a class of mechanisms that are invariant to renaming of schools (see also Ehlers, 2008).

Definition 3. Given any $l^i \in A^i$ and any permutation $\alpha : X \to X$ of schools, let l^i_α be the list such that

$$l^i_\alpha(m) = \alpha(l^i(m)) \qquad m = 1, \ldots, M. \tag{56}$$

A matching mechanism $\boldsymbol{x}$ is **anonymous** if for any permutation $\alpha : X \to X$

$$\boldsymbol{x}^i(l^i_\alpha, \boldsymbol{l}^{-i}_\alpha) = \alpha(\boldsymbol{x}^i(l^i, \boldsymbol{l}^{-i})) \tag{57}$$

for all $i \in I$, $l^i \in A^i$, and $\boldsymbol{l}^{-i} \in \boldsymbol{A}^{-i}$.

To state the result formally, we define student i's probability, given $\boldsymbol{P}^{-i}$, of being matched with school $x \in X$ when submitting list $l^i \in A^i$ as

$$\pi^i_{\boldsymbol{x}}(x | l^i, \boldsymbol{P}^{-i}) \equiv \sum_{\boldsymbol{l}^{-i} \in \boldsymbol{A}^{-i}} \sum_{\boldsymbol{\theta}^{-i} \in \boldsymbol{\Theta}^{-i}} \boldsymbol{P}^{-i}(\boldsymbol{l}^{-i}, \boldsymbol{\theta}^{-i}) \mathbf{1}\left\{\boldsymbol{x}^i(l^i, \boldsymbol{l}^{-i}) = x\right\}. \tag{58}$$

Lemma 1 (List-Position Sufficiency)**.** *If, for any $i \in I$, P^j is symmetric for all $j \neq i$ in the sense of Definition 2, then for any anonymous matching mechanism $\boldsymbol{x}$ there exists $h^i \in \Delta(\{1, \ldots, M\})$ such that*

$$\pi^i_{\boldsymbol{x}}(l^i(m) | l^i, \boldsymbol{P}^{-i}) = h^i(m) \tag{59}$$

for all $l^i \in A^i$ and all $m \in \{1, \ldots, M\}$.

List-position sufficiency rests on students' symmetric information (cf. Roth and Rothblum, 1999), which they acquire endogenously in light of their exchangeable priors.

Given others' strategies $\boldsymbol{P}^{-i}$, student i's expected utility of submitting rank-order list $l^i \in A^i$ when their true type is $\theta^i \in \Theta^i$ is

$$u^i(l^i, \theta^i) = \sum_{x \in X} \pi^i_{\boldsymbol{x}}(x | l^i, \boldsymbol{P}^{-i}) \theta^i_x = \sum_{m=1}^{M} \pi^i_{\boldsymbol{x}}(l^i(m) | l^i, \boldsymbol{P}^{-i}) \theta^i_{l^i(m)}. \tag{60}$$

The distribution h^i of random variable m thus captures the entire strategic uncertainty under the matching mechanism. We next note that list-position sufficiency implies that the student's problem is symmetric, provided that their prior is exchangeable.

Lemma 2. *The problem of student i with choice set $A = A^i$, an exchangeable prior $\mu^i \in \Delta(\Theta^i)$ over state space $\Omega = \Theta^i$, and utility, for any probability mass function h^i,*

$$u^i(l^i, \theta^i) = \sum_{m=1}^{M} h^i(m) \theta^i_{l^i(m)} \tag{61}$$

is symmetric in the sense of Definition 1.

The unique symmetric equilibrium then rests on Proposition 1, by which each student has a unique symmetric best response to others' symmetric equilibrium strategies.

4.2 Characterizing Strategic Uncertainty under DA

This section fully characterizes the strategic uncertainty under DA. Due to schools' unanimous ranking of students, the matching produced by the Deferred Acceptance Algorithm is characterized by serial dictatorship (Kojima and Manea, 2010; Morrill, 2013): the matching is as if every student i was matched in order, starting with student 1, with the first school in their submitted rank-order list among those with positive remaining capacity after students $1, \dots, i-1$ have been assigned in the same manner.

Proposition 4. *Given schools' unanimous ranking of students, the DA matching is characterized by serial dictatorship.*

This permits us to find a combinatoric expression for the probability mass function h^i_{DA}. The key observation is that given that others' equilibrium strategies are symmetric, each school is equally likely to be available to any student (see Lemma 3 below). The number of schools available to a student is thus a sufficient statistic for their probability of each outcome, which is specified combinatorically in Lemma 4 below. When schools' capacities are greater than 1, then the number of available schools is stochastic, however, so Lemma 5 specifies the corresponding distribution of the number of available schools each student will face given others' symmetric strategies. In order to state these results, we define the random vector $\boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n \equiv \left\{\boldsymbol{c}^n \in \mathbb{N}_0^M : \sum_{x\in X} c^n_x = c \cdot M - n\right\}$ whose element c^n_x is the remaining capacity of school x after students $1, \dots, n$ have been assigned by serial dictatorship and whose probability distribution is induced by these students' strategies,

$$\mathbb{P}(\boldsymbol{c}^n | P^1, \dots, P^n) \equiv \sum_{l^1 \in A^1} \cdots \sum_{l^n \in A^n} \mathbf{1}\left\{ c^n_x = c - \sum_{i=1}^{n} \mathbf{1}\{\boldsymbol{x}^i_{DA}(l^i, \boldsymbol{l}^{-i}) = x\} \; \forall x \in X \right\} \prod_{j=1}^{n} \sum_{\theta^j \in \Theta^j} P^j(l^j, \theta^j). \quad (62)$$

Note that this definition, resting on serial dictatorship, would not be meaningful for more general mechanisms. The following Lemma establishes that if the strategies of higher-ranked students are symmetric, then each school is equally likely to have a given remaining capacity, and thus also the same probability of having positive remaining capacity.

Lemma 3. *If $P^1, \dots, P^n$ are symmetric in the sense of Definition 2, then the conditional distribution $\mathbb{P}(\boldsymbol{c}^n | P^1, \dots, P^n)$ is exchangeable, i.e. for all permutations $\alpha : X \to X$ of schools*

$$\mathbb{P}(c^n_{\alpha(1)}, \dots, c^n_{\alpha(M)} | P^1, \dots, P^n) = \mathbb{P}(c^n_1, \dots, c^n_M | P^1, \dots, P^n) \quad (63)$$

for all $\boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n$.

Since the distribution of $\boldsymbol{c}^n$ is exchangeable, a sufficient statistic is given by the random vector $\boldsymbol{k}^n = (k^n_0, \dots, k^n_c)$ whose elements $k^n_r \equiv \sum_{x \in X} \mathbf{1}\{c^n_x = r\}$, for $r = 0, 1, \dots, c$, capture the number of schools with remaining capacity r after n students have been assigned. All vectors $\boldsymbol{c}^n$ in the set

$$\boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n) \equiv \left\{ \boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n : \sum_{x \in X} \mathbf{1}\{c_x^n = r\} = k_r^n \quad \forall r = 0, \ldots, c \right\} \tag{64}$$

of vectors consistent with a given $\boldsymbol{k}^n$ are permutations of each other and hence – by exchangeability – equally likely. Conditional on a vector $\boldsymbol{k}^n$, the remaining capacities $\boldsymbol{c}^n$ induced by the symmetric equilibrium strategies are thus distributed uniformly across schools.

Corollary 1. *If $P^1, \ldots, P^n$ are symmetric, then the distribution of $\boldsymbol{c}^n$ conditional on $\boldsymbol{k}^n$ is uniform over $\boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)$.*

This observation permits us to find the combinatoric expression for h_{DA}^i, which we now illustrate in two steps. In the next subsection, we derive an expression for a student's probability of matching with the m-th school on any list they submit, *conditional* on a vector $\boldsymbol{k}^n$ of remaining capacities. The subsection thereafter derives a recursive expression for the distribution of remaining capacities faced by any student. Together, these two results characterize each student's probability mass function h_{DA}^i.

4.2.1 Outcome Conditional on the Number of Schools with no Remaining Capacity

By serial dictatorship, each student is matched with the first school on their list having positive remaining capacity after all higher-ranked students have been matched in the same manner. Since each school is equally likely to be available to a student, the number k_0^n of schools with no remaining capacity is a sufficient statistic for the probability that the m-th school on a student's list l^i is the first one with positive remaining capacity (and thus their matching outcome). This probability is given in the following Lemma.

Lemma 4. *Given any $\boldsymbol{k}^n$ and $l^i \in A^i$*

$$\frac{1}{|\boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)|} \sum_{\boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)} \mathbf{1}\left\{ m = \min_{\eta}\{\eta : c_{l^i(\eta)}^n > 0\} \right\} = \binom{M-m}{k_0^n + 1 - m} \Big/ \binom{M}{k_0^n} \qquad \forall m \in \{1, \ldots, M\}. \tag{65}$$

Conditional on k_0^n, this result provides the probability distribution over the rank m on their list of the school with which a student will be matched. The combinatoric expression with binomial coefficients arises because each of the $\binom{M}{k_0^n}$ configurations in which exactly k_0^n schools have no remaining capacity is equally likely. Of these configurations, $\binom{M-m}{k_0^n+1-m}$ are such that the first $m-1$ positions on a given list have a school with no remaining capacity and the m-th position features a school with positive remaining capacity, which are exactly the cases in which a student is matched with the m-th school on their list. Their probability, conditional on k_0^n, of being matched with the m-th school on their list is thus

$$\binom{M-m}{k_0^n + 1 - m} \Big/ \binom{M}{k_0^n} \tag{66}$$

for *any* list. Note that this probability is zero for all $m > k_0^n + 1$, which reflects the fact that a student can do no worse than list all unavailable schools before available ones.

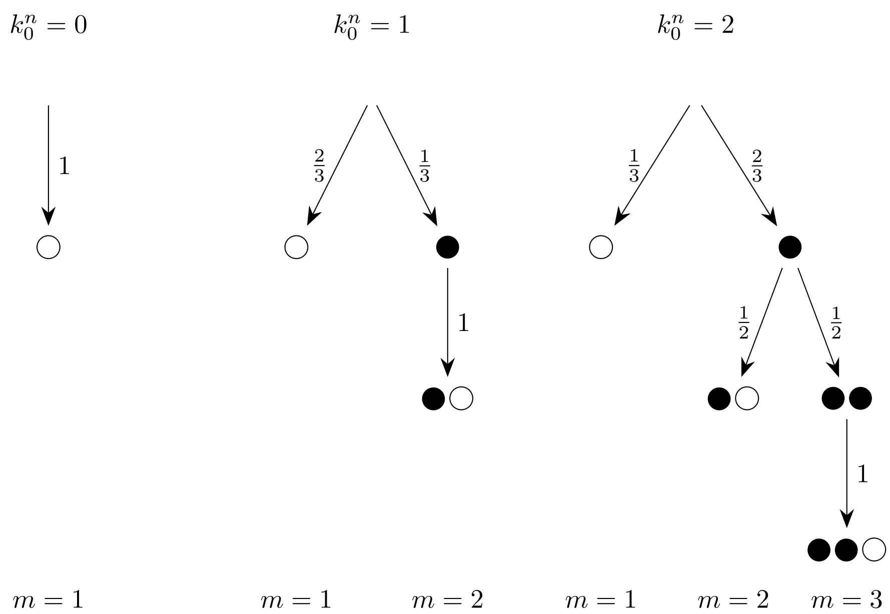

Figure 4: Illustration of Lemma 4 for $M = 3$ schools. White balls symbolize schools with positive remaining capacity, black balls schools that have no remaining capacity and are thus unavailable to a student. Each column corresponds to a number k_0^n of schools that are unavailable to a student, and each of the three schools is equally likely to be an available one. Determining the first available school in a given list thus amounts to sequential draws without replacement from an urn with $M = 3$ balls, k_0^n of which are black, until the first white draw.

This result is illustrated in Figure 4 which revisits the example of Section 3 featuring three schools of unit capacity. If $k_0^n = 0$ schools are unavailable to a student, as is the case for student 1, then the school in position $m = 1$ on their list is available with probability

$$\binom{3-1}{0+1-1} / \binom{3}{0} = 1. \tag{67}$$

If, as is the case for student 2, exactly $k_0^n = 1$ of the three schools is unavailable (specifically, any of the three schools with equal probability 1/3), then the school in position $m = 1$ on their list is available with probability

$$\binom{3-1}{1+1-1} / \binom{3}{1} = 2/3, \tag{68}$$

and otherwise their second-listed school is available. If $k_0^n = 2$ schools are unavailable, as is the case for student 3, then by analogous logic there is a 1/3 chance that their first-listed school is available, a 1/3 chance that the first-listed school is unavailable but the second one is available, and a 1/3 chance that only the third-listed school is available. Equation 65 thus parsimoniously characterizes the strategic uncertainty faced by students in the example of Section 3.

4.2.2 Characterizing the Distribution of Remaining Capacities

In the example of Section 3 there was no uncertainty about the number of schools with no remaining capacity, only about their identity. This is not true when schools have capacity greater than 1, however: The number of unavailable schools then is stochastic; it depends on the probability mass function in equilibrium of the random vector $\boldsymbol{k}^n$ after n students have been assigned, which we will denote by $f^n(\boldsymbol{k}^n)$. This probability mass function can be obtained recursively. The recursion rests on serial dictatorship and the fact that for a student with a symmetric strategy P^i, each school with positive remaining capacity is an equally likely outcome (see Lemma C.2 in the Appendix).

The resulting recursion yielding the distribution over remaining capacities is illustrated in Figure 5 for an example with three schools (as before), now of capacity 2. In the top left corner, the distribution f^0 before any student has been assigned is degenerate at $\boldsymbol{k}^0 = (0, 0, 3)$, as all three schools are certain to have a remaining capacity of two. Student 1 will thus with certainty be matched with a school of remaining capacity $r = 2$. After $n = 1$ students have been assigned, one school will thus have one slot remaining and two schools will have both slots remaining, so that $f^1(\boldsymbol{k}^1 = (0, 1, 2)) = \frac{3}{3} f^0(\boldsymbol{k}^0 = (0, 0, 3)) = 1$ is degenerate.

The distribution f^2 after two students have been assigned is non-degenerate: If students 1 and 2 did *not* match with the same school, then two schools will have one slot remaining and one school will have two slots remaining, $\boldsymbol{k}^2 = (0, 2, 1)$ (left branch). Since student 2 has an equal probability of matching with any school, this occurs with probability $f^2(\boldsymbol{k}^2 = (0, 2, 1)) = \frac{2}{3} f^1(\boldsymbol{k}^1 = (0, 1, 2)) = 2/3$. If students 1 and 2 *did* match with the same school (right branch), then $\boldsymbol{k}^2 = (1, 0, 2)$ which occurs with probability $f^2(\boldsymbol{k}^2 = (1, 0, 2)) = \frac{1}{3} f^1(\boldsymbol{k}^1 = (0, 1, 2)) = 1/3$.

How many schools are available after $n = 3$ students depends on whether students 1, 2, and 3 all matched with different schools or whether two of them matched with the same school. If they all matched with different schools then $\boldsymbol{k}^3 = (0, 3, 0)$. This occurs with probability

$$f^3(\boldsymbol{k}^3 = (0, 3, 0)) = \frac{1}{3} f^2(\boldsymbol{k}^2 = (0, 2, 1)) = 2/9 \tag{69}$$

since, conditional on $\boldsymbol{k}^2 = (0, 2, 1)$, there is a 1/3 chance that student 3 matched with the one school of remaining capacity $r = 2$ (left branch). Otherwise two of the students matched with the same school so that $\boldsymbol{k}^3 = (1, 1, 1)$, which as illustrated in the fourth row of Figure 5 occurs with probability

$$f^3(\boldsymbol{k}^3 = (1, 1, 1)) = \frac{2}{3} f^2(\boldsymbol{k}^2 = (0, 2, 1)) + \frac{2}{2} f^2(\boldsymbol{k}^2 = (1, 0, 2)) = 7/9, \tag{70}$$

because $\boldsymbol{k}^3 = (1, 1, 1)$ arises if either $\boldsymbol{k}^2 = (0, 2, 1)$ and student 3 matched (with probability 2/3) with one of the two schools of remaining capacity $r = 1$, or if $\boldsymbol{k}^2 = (1, 0, 2)$ and student 3 matched (with certainty) with one of the two schools of remaining capacity $r = 2$. Analogous logic yields the distributions in the remaining rows of Figure 5. The general characterization of remaining capacities following the logic of the preceding example is given by the following Lemma.

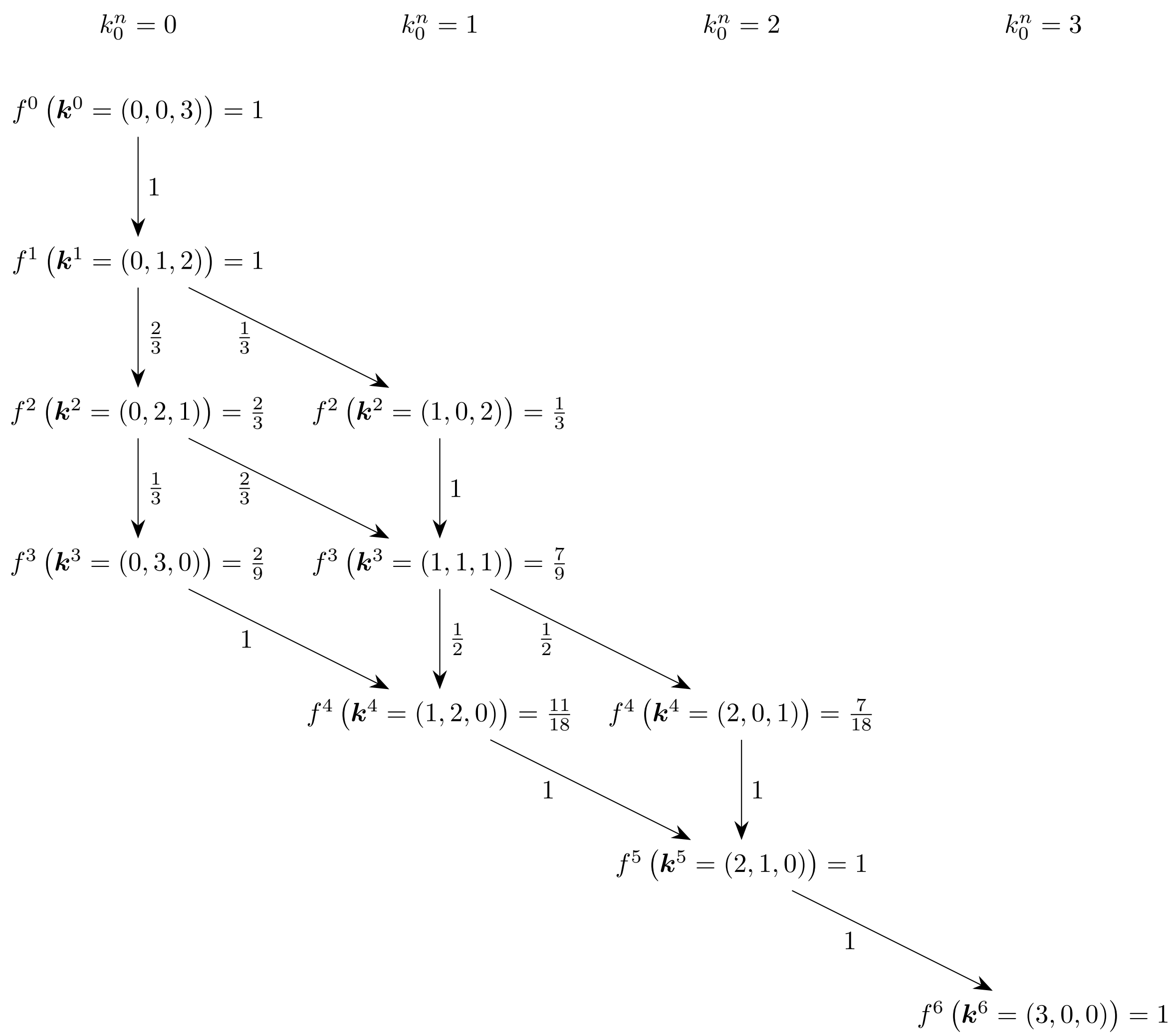

Figure 5: Illustration of the recursion of Lemma 5 for a simple example with 3 schools of capacity 2. Each row illustrates the distribution $f^n(\boldsymbol{k}^n)$ for a different number n of students that have been assigned. Each column corresponds to a given number k_0^n of schools with no remaining capacity.

Lemma 5. *For any $n = 0, \ldots, N$, if $P^1, \ldots, P^n$ are symmetric in the sense of Definition 2, then*

$$\sum_{\boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)} \mathbb{P}(\boldsymbol{c}^n | P^1, \ldots, P^n) = f^n(\boldsymbol{k}^n) \tag{71}$$

for all $\boldsymbol{k}^n$, where the probability mass functions f^n over a random vector $\boldsymbol{k} = (k_0, k_1, \ldots, k_c)$ are defined recursively as

$$\begin{aligned} f^0(\boldsymbol{k}) &= \mathbf{1}\{k_c = M, k_r = 0\,\forall r \neq c\} \\ f^{n+1}(\boldsymbol{k}) &= \sum_{r=1}^{c} \frac{k_r + 1}{M - k_0 + \mathbf{1}\{r = 1\}} f^n(\boldsymbol{k} + \boldsymbol{e}_r - \boldsymbol{e}_{r-1}) \end{aligned} \tag{72}$$

with $\boldsymbol{e}_r$ being the unit vector of length $c+1$ such that $(\boldsymbol{e}_r)_\rho = \mathbf{1}\{\rho = r\}$ for $\rho = 0, \ldots, c$, and where $f^n(\boldsymbol{k}) = 0$ unless $k_r \in \{0, \ldots, M\}$ for all r and $\sum_{r=0}^{c} k_r = M$.

Given the resulting distributions of k_0^{i-1}, each student i's probability $h^i_{DA}(m)$ of matching with the school in each position on their list is then readily obtained from Lemma 4. In the previous example, students 1 and 2 will always be matched with their first-listed school, since the schools have sufficient capacity to accommodate two students so that $k_0^0 = k_0^1 = 0$ with certainty. The probability that the first-listed school of student 3 is still available to them depends on k_0^2. With probability 2/3 it is the case that $k_0^2 = 0$, in which case student 3 gets their first choice with probability 1 (by Lemma 4). Otherwise $k_0^2 = 1$, in which case student 3 gets their first choice with probability 2/3 and their second choice with probability 1/3. Student 3 is thus matched with their first choice with probability 8/9, and with their second choice otherwise. In contrast, student 4 is matched with their first choice with probability 20/27 and their second choice otherwise, since $k_0^3 = 0$ only with probability 2/9.

4.3 Equilibrium Characterization

Proposition 1, Lemma 1, and Lemma 2 establish the unique symmetric Nash equilibrium, which is summarized by the following Theorem. If others' strategies are symmetric, then a student i's probability of matching with a school under DA only depends on its position m on their list and the distribution of remaining capacities, as characterized by Lemmas 3, 4, and 5.

Theorem 1 (Nash Equilibrium)**.** *Given an anonymous mechanism, $(P^{1*}, \dots, P^{N*})$ is a symmetric Nash equilibrium if and only if for all $i \in I$*

$$P^{i*}(l^i|\theta^i) \equiv \frac{P^{i*}(l^i, \theta^i)}{\mu^i(\theta^i)} = \frac{z^i(l^i, \theta^i)}{Z^i(\theta^i)} \qquad \forall l^i \in A^i, \theta^i \in \Theta^i \tag{73}$$

where

$$z^i\left(l^i, \theta^i\right) = \exp\left(\frac{1}{\kappa^i}\sum_{m=1}^{M} h^i(m)\theta^i_{l^i(m)}\right) \tag{74}$$

and

$$Z^i(\theta^i) = \sum_{l' \in A^i} z^i(l', \theta^i). \tag{75}$$

Under DA

$$h^i_{DA}(m) = \sum_{\boldsymbol{k} \in \mathbb{N}^{c+1}} \binom{M-m}{k_0+1-m} \Big/ \binom{M}{k_0} \cdot f^{i-1}(\boldsymbol{k}). \tag{76}$$

The key observation in establishing the result is that given others' symmetric equilibrium strategies, the problem of student i is symmetric in the sense of Definition 1, so that equation 73 is the unique symmetric best response, giving rise to choice probabilities that are *unconditionally* uniform. Regardless of how the strategic uncertainty in h^i resolves, they thus have an unconditionally equal probability of matching with any school that still has positive remaining capacity.

It is worth noting how simple these results are. For example, in a general model, changing any student's cost of learning would be expected to change the strategic uncertainty facing all subsequent students. In our symmetric framework, students have symmetric information à la Roth and Rothblum (1999). This implies that the effect instead is purely local to the student in question, so that changing one student's marginal cost of information only affects their outcome, but does not otherwise affect the equilibrium. It is this feature that allows us to conduct separable policy experiments for each student. Note also that the probability mass function h^i is monotone decreasing in m, which reflects the strategy-proofness of DA: Under any belief it is best to rank the schools in decreasing order of their expected utility. While students may submit a list inconsistent with their true type, they have no incentive to deviate from truthfully revealing their belief.

5 Students' Choices, Mistakes, and Welfare

5.1 Students' List Submissions

Theorem 1 shows that the powerful separability properties of the solution found in the example of Section 3 are fully general. Each student's optimal strategy has the same qualitative features and essentially the same simplicity as in that special case. The conditional probability of any list in any state can be computed, according to equation 73, as a ratio between a transformed payoff to that list in that state and the sum of the correspondingly transformed payoffs in that state across all lists. The properties of this solution notably depend on the precise nature of the payoff transformation in equation 74. The novel feature of this formula relates to the strategic uncertainty it captures in a student's probability $h^i(m)$ of receiving the m-th listed school on any list, which determines the expected utility of the *stochastic* outcome resulting from an action in a given state. Note that this is different from standard rational inattention models in that the learnable state does not resolve all of the uncertainty: The decision maker can only reduce value uncertainty, but not strategic uncertainty. The probability of choosing any list is then obtained by the usual transformation of dividing the expected utility of that list by the cost parameter κ^i and exponentiating, divided by the sum across all lists of that expression. It is hard to imagine a simpler formula that takes into account in a flexible manner all strategic considerations.

Note finally that the prior is separable just as in the three-school case. Variations in the prior change the pattern of behavior without in any way changing the conditional choice probabilities in each state. Of course the prior does impact posterior beliefs and hence welfare, which we will now discuss.

5.2 Value Uncertainty in Students' Posterior Beliefs

Note that the denominator in the conditional probability of equation 73, $Z^i(\theta^i)$ as specified in equation 75, depends only on the set of payoffs, but not their precise assignment to schools. The assumption of exchangeable priors implies that all states with the same such set of payoffs are equiprobable. This means that the posterior beliefs have corresponding invariance properties. From students' solution of equation 73 it is immediate to find their posterior belief

$$P^{i*}(\theta^i|l^i) = \mu^i(\theta^i)\frac{z^i(l^i,\theta^i)}{Z^i(\theta^i)/(M!)} \tag{77}$$

when submitting any list $l^i \in A^i$. The posterior distorts the prior in a systematic and continuous manner: The prior probability $\mu^i(\theta^i)$ of any state is divided by this state's average value across all lists, $Z^i(\theta^i)/(M!)$, and then multiplied by the state's value $z^i(l^i,\theta^i)$ when submitting a given list l^i. This implies that states giving higher-than-average utility for a list are more likely under a student's posterior belief when submitting that list than they are under the student's prior. Exchangeability again implies that $\mu^i(\theta^i)$ and $Z^i(\theta^i)$ are the same for all states featuring the same distribution of utility levels.

This explicit solution allows us to analyze the form and likelihood of the mistakes each student makes. As in the example of Section 3, the structure of learning is best understood by analyzing students' expected utility after strategic uncertainty has resolved. Student i's expected utility in equilibrium of being matched with their m-th listed school is given by

$$U^{i*}\left(\kappa^i,\mu^i|m\right) \equiv \sum_{l^i\in A^i}\sum_{\theta^i\in\Theta^i} P^{i*}(l^i,\theta^i)\theta^i_{l^i(m)} = \sum_{\theta^i\in\Theta^i}\mu^i(\theta^i)\frac{\sum_{l^i\in A^i} z^i(l^i,\theta^i)\theta^i_{l^i(m)}}{Z^i(\theta^i)}. \tag{78}$$

This quantity's additive separability by type θ^i, while compelling and intuitive, is one of the attractive consequences of the exchangeability assumption that could not have been expected ex ante. This position-conditional expected utility is a response to, and reflects features of, the strategic uncertainty in h^i. This function, like h^i, will be monotonically decreasing in m, as students list schools in order of decreasing expected utility.

5.3 Gross Welfare and the Value of Information

From equation 78 it is immediate that in equilibrium, student i attains gross utility $U^{i*}(\kappa^i,\mu^i) \equiv U^i(P^{i*},\boldsymbol{P}^{-i*})$ given by

$$U^{i*}(\kappa^i,\mu^i) = \sum_{m=1}^{M} h^i(m)U^{i*}(\kappa^i,\mu^i|m) = \kappa^i\sum_{\theta^i\in\Theta^i}\mu^i(\theta^i)\frac{\sum_{l^i\in A^i} z^i(l^i,\theta^i)\ln z^i(l^i,\theta^i)}{Z^i(\theta^i)}. \tag{79}$$

A student who does not acquire any information and instead relies on their prior attains expected utility

$$\underline{U}(\mu^i) = \sum_{\theta^i\in\Theta^i}\mu^i(\theta^i)\frac{\sum_{x\in X}\theta^i_x}{M}, \tag{80}$$

as they are equally likely to be matched with any of the M schools. The following Proposition quantifies the resulting value of information (Frankel and Kamenica, 2019) to each student.

Proposition 5. *The value of student i's information structure in equilibrium is*

$$U^{i*}(\kappa^i,\mu^i) - \underline{U}(\mu^i) \quad = \quad \sum_{m=1}^{M}\left[h^i(m) - \frac{1}{M}\right]U^{i*}\left(\kappa^i,\mu^i|m\right). \tag{81}$$

The value of information is again separable by the prior, a property that is inherited from $U^i\left(\kappa^i, \mu^i | m\right)$ in equation 78. Proposition 5 highlights the value of information's dependence on strategic uncertainty. Strategic uncertainty affects $U^i\left(\kappa^i, \mu^i | m\right)$ as discussed above, but it also enters the expression for the value of information directly. For example, the value of information is highest for student 1 for whom $h^1(1) = 1$, and zero for student N for whom $h^N(m) = 1/M$ for all m. The expression above also opens the door for comparative statics of gross welfare, as exemplified in Section 3.

5.4 Net Welfare and the Cost of Information

Analogous expressions can be obtained for the information costs incurred by any student and their resulting net utility. In equilibrium, student i incurs information costs $K^{i*}(\kappa, \mu^i) \equiv K(\kappa, P^{i*})$ of

$$K^{i*}(\kappa^i, \mu^i) = \kappa^i \sum_{\theta^i \in \Theta^i} \mu^i(\theta^i) \sum_{l^i \in A^i} \frac{z^i(l^i, \theta^i)}{Z^i(\theta^i)} \ln\left(\frac{M! z^i(l^i, \theta^i)}{Z^i(\theta^i)}\right). \tag{82}$$

Their resulting net utility is given in the following Proposition.

Proposition 6. *In equilibrium, student i attains net utility $N^{i*}(\kappa^i, \mu^i) \equiv U^{i*}(\kappa^i, \mu^i) - K^{i*}(\kappa^i, \mu^i)$ given by*

$$N^{i*}(\kappa^i, \mu^i) = \kappa^i \sum_{\theta^i \in \Theta^i} \mu^i(\theta^i) \ln\left(\frac{Z^i(\theta^i)}{M!}\right). \tag{83}$$

The functional form capturing students' net utility is remarkably simple and depends on the average value of a state across all feasible lists, $Z^i(\theta^i)/(M!)$. This expression facilitates a direct and intuitive comparison between students' net utilities in equilibrium. The net utility surplus is given accordingly by

$$N^{i*}(\kappa^i, \mu^i) - \underline{N}(\mu^i) = \sum_{\theta^i \in \Theta^i} \mu^i(\theta^i) \left[\kappa^i \ln\left(\frac{Z^i(\theta^i)}{M!}\right) - \frac{1}{M} \sum_{x \in X} \theta^i_x\right]. \tag{84}$$

6 Inequity, Policy, and Sequential Search

This section analyzes the implications of costly information for students' ex ante welfare, and discusses the potential for cost-reducing policies to alleviate the inequity resulting from costly information.

6.1 Inequity

Figure 6 illustrates, using an example with eight students and four schools of capacity 2, that lower-ranked students attain a lower fraction of their (already smaller) net welfare surplus and generally lower gross welfare, as well.

However, information costs can also result in lower-ranked students attaining *greater* expected gross utility than higher-ranked students, even though higher-ranked students are preferred by all schools. This counter-intuitive flipping of the utility order is due to the fact that lower-ranked students may be incentivized to learn more, as illustrated in the following example. Consider six students and three schools of capacity 2

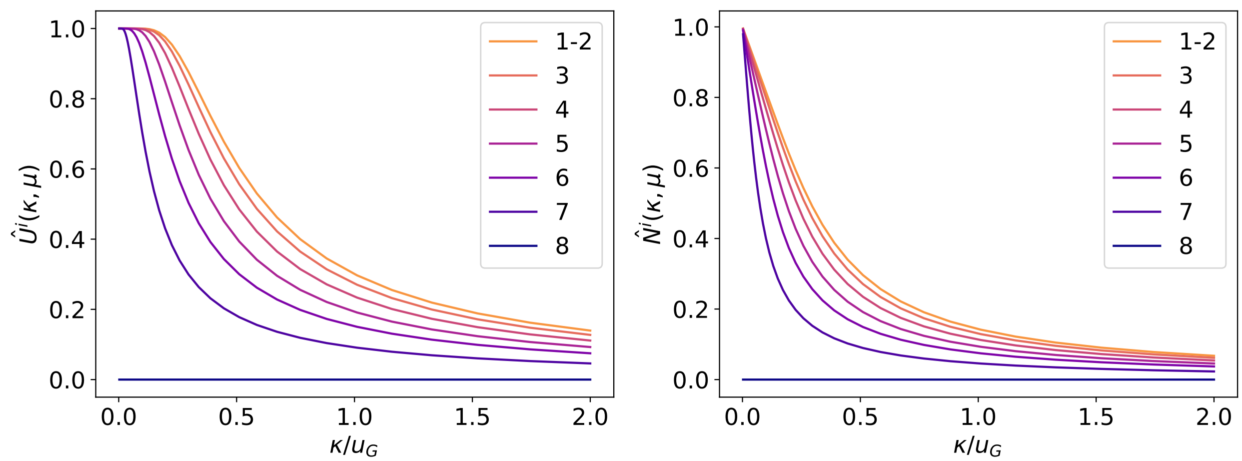

Figure 6: Relative gross (left) and net (right) welfare in an example with eight students and four schools of capacity 2, with priors under which each student considers exactly one of the schools to be good ($u_G = 1$) and three bad ($u_B = 0$).

as in the example of Figure 5 in Section 4, and assume that exactly two schools are good while one is bad ("lemon dropping"). Students' position-conditional expected utilities are shown for $\kappa = 0.05$ and $u_G = 1$ in Table 3. The expected utility of the first-listed school of students 1 through 5 is 1, meaning that they list a good school in the first position with certainty. Students 1 and 2 list the second good school in second or third position with equal probability 1/2, but always match with a good school since they are always matched with their first-listed school. Student 3, who faces (as derived in Section 4) a 1/9 probability of matching with their second-listed school, lists the second good school second with probability 90.2% and third otherwise. They thus attain an ex ante expected utility of 0.989, meaning that they match with a good school with probability 98.9%. Student 4 faces a 7/27 chance of matching with their second-listed school (see Section 4 above). This outcome is more likely than it is for student 3, and incentivizes student 4 to exert more effort in identifying the second good school. They thus list the second good school second

$i\backslash m$	1	2	3
1	1	1/2	1/2
2	1	1/2	1/2
3	1	0.902	0.098
4	1	0.994	0.006
5	1	0.983	0.017
6	2/3	2/3	2/3

Table 3: Position-conditional expected utility $U^i(\kappa^i, \mu^i|m)$ for $\kappa^i = 0.05$ and μ^i under which exactly two schools yield $u_G = 1$ and one has utility 0.

with probability 99.4% and third otherwise, and therefore match with a good school with probability 99.8%, which is a better outcome than that of student 3. While student 3 could do better by following the learning strategy of student 4, its higher cost is not justified for them. Student 5 lists the second good school second with probability 98.3% and third otherwise; student 6 has a 2/3 chance in each slot of having a good school, thus matching with a good school with probability 2/3.

6.2 Information Policies

Our model allows us to quantify the relationship between students' rank and welfare loss, as well as the effect of cost-lowering policies. A particularly attractive feature of the equilibrium is that students' solutions do not depend on others', so that changing one student's information costs does not have any equilibrium effects. Even after a student's learning costs are reduced to offset the informational disadvantage as in Section 3, note that higher-ranked students will still achieve higher utility, as the policy maker offsets the informational disadvantage resulting from a student's lower rank but *not* the lower priority itself. Note also that cost-lowering policies can offset lower-ranked students' welfare loss by reducing their value uncertainty, but they cannot reduce their strategic uncertainty itself, which weakens their impact. Student N is a particularly stark example, as they have no incentive to acquire any information regardless of how much a policy may reduce information costs. This stresses the need for mechanism design that takes into account information acquisition incentives – future research for which our model opens the door and for which the symmetric setting seems to be particularly promising. This will likely require reducing strategic uncertainty.

6.3 Inequity in the Boston Mechanism

Artemov (2021) and Chen and He (2021b) have observed that the Immediate Acceptance (IA) algorithm or Boston mechanism, while challenged for its lack of strategy-proofness (cf. Abdulkadiroğlu and Sönmez, 2003; Abdulkadiroğlu et al., 2005, 2009), may incentivize better information acquisition than DA. An interesting observation is that IA also incentivizes information acquisition more equally: In the three student example of Section 3, for example, it would give rise to strategic uncertainty captured by the probability mass function

$$h_B^1 = (1, 0, 0) \tag{85}$$

$$h_B^2 = \left(\frac{6}{9}, \frac{2}{9}, \frac{1}{9}\right) \tag{86}$$

$$h_B^3 = \left(\frac{4}{9}, \frac{5}{18}, \frac{5}{18}\right) \tag{87}$$

Student 2 still has a 2/3 chance of being matched with their first pick, which is the case unless that school is also the first pick of student 1. If student 2 does not receive their first choice, they are no longer guaranteed to be matched with their second choice, however: They will miss out on their second choice if it is the first choice of student 3, which is the case with probability 1/3. Student 2 thus has a lower incentive to resolve their preference under IA than they did under DA. Student 3, on the other hand, now has positive value of information. They will be matched with their first choice if it is neither the first choice of student 1 nor of student 2, which is the case with probability 4/9. In this example IA is thus more equitable in incentivizing information acquisition.

6.4 Sequential Search

Modeling students as rationally inattentive, with learning costs that are linear in the expected reduction in Shannon entropy between their prior and posterior beliefs, gives rise to equilibrium strategies that reflect the strategic learning incentives in a remarkably compact and simple form. Yet Section 4 makes clear that the central analytic challenge, irrespective of learning costs, is solving for the strategic uncertainty. It is thus natural to conjecture that many of the qualitative features of the equilibrium are robust to changes in the cost function. Most significant perhaps is to consider qualitatively different costs, such as those that characterize standard models of sequential search. While the Shannon cost function is uniformly posterior separable (Caplin et al., 2021) and as such can be micro-founded based on continuous incremental learning and optimal stopping (Hébert and Woodford, 2021; Bloedel and Zhong, 2021), sequential search is a more appropriate description for situations in which agents incur fixed costs to perfectly learn their valuation of a school (e.g. through a campus visit). The most realistic cases may be of hybrid form, with both fixed and flow costs. Appendix B revisits the example of Section 3 and shows that its key conclusions are indeed robust to replacing the Shannon cost function with optimal sequential search. Recall that the example involves three schools of unit capacity and three students whose valuation of schools is either good or bad, yielding utility $u_G > 0$ or $u_B = 0$, respectively.

To reduce notational burden, we assume that all students incur the same fixed search cost $\kappa \in (0, 1)$ for fully revealing their valuation of a school, and that they have the same exchangeable prior μ with full support $\mu(g) > 0$ (to avoid trivial cases) over the number $g = 0, 1, 2, 3$ of good schools, so that

$$\mu(3) + 3\mu(2) + 3\mu(1) + \mu(0) = 1. \tag{88}$$

To reflect the focus on symmetry, we assume that search order among schools is uniform and independent across students. One of the central findings of Section 3 is that lower-ranked students are affected disproportionately by information costs. Figure 7, which plots students' relative gross and net welfare analogous to Figures 2 and 3, demonstrates that this finding is robust. While the precise shape depends on parameters, the broad qualitative features of relative welfare remain intact in this example. First, the relative welfare falls faster for student 2 as search costs rise above zero. Second, it is always (weakly) higher for student 1 than for student 2. It is this second property which implies that no search is optimal for student 2 at a lower cutoff than for student 1. The figure also illustrates that with sequential search costs, gross welfare is a discontinuous function of information costs.

While Figure 7 is qualitatively similar to Figures 2 and 3, note that sequential search gives rise to piece-wise linear functions, while the Shannon case is smooth. This reflects the different nature of the two

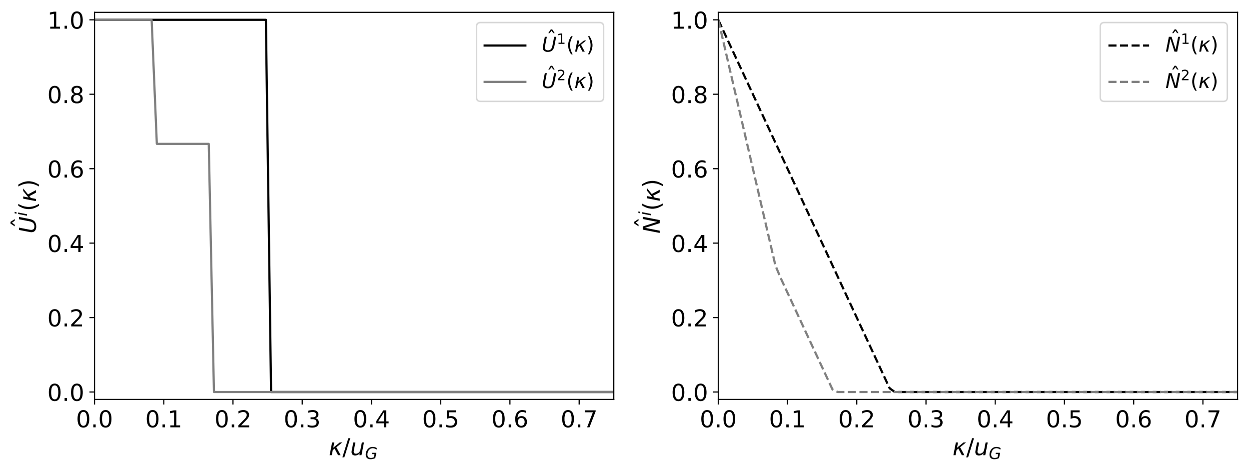

Figure 7: Relative gross (left) and net (right) welfare of student 1 and 2 under sequential search, as a function of κ, for $\mu(3) = \mu(2) = \mu(1) = \mu(0) = 1/8$.

information cost functions. In sequential search, students pay a fixed cost to fully reveal their valuation of a school. This results in strategies that vary discontinuously with search costs, which would make the equilibrium analysis beyond this example challenging. Shannon information costs, in contrast, make it easier to analyze the resulting equilibrium even for relatively general cases.

7 Conclusion

The attractive properties of DA rest on the assumption that applicants are fully informed about their options, yet in practice this is often not the case as learning is costly and information is incomplete. We introduce a tractable model of costly strategic learning and characterize the welfare consequences of endogenously imperfect information for DA. We assume that schools agree on their ranking of students, that schools are ex ante symmetric, and that learning is rationally inattentive. We find that lower-ranked students are affected disproportionately more by information costs, generally suffering a larger welfare loss than higher-ranked students. This raises questions about the mechanism's consequences for equity and underlines the importance of exploring the trade-off between how a mechanism incentivizes information acquisition and other possible desiderata such as strategy-proofness.

References

Atila Abdulkadiroğlu and Tayfun Sönmez. School Choice: A Mechanism Design Approach. *American Economic Review*, 93(3):729—747, 2003.

Atila Abdulkadiroğlu, Parag A Pathak, Alvin E Roth, and Tayfun Sönmez. The Boston Public School Match. *American Economic Review*, 95(2):368–371, 2005. ISSN 0002-8282. doi: 10.1257/000282805774669637.

Atila Abdulkadiroğlu, Parag A. Pathak, and Alvin E. Roth. Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match. *American Economic Review*, 99(5):1954—1978, 2009.

Georgy Artemov. Assignment mechanisms: Common preferences and information acquisition. *Journal of Economic Theory*, 198:105370, 2021. ISSN 0022-0531. doi: 10.1016/j.jet.2021.105370.

Sophie Bade. Serial dictatorship: The unique optimal allocation rule when information is endogenous. *Theoretical Economics*, 10(2):385–410, 2015. ISSN 1555-7561. doi: 10.3982/te1335.

Alexander W. Bloedel and Ilya Segal. Persuading a Rationally Inattentive Agent. *Working Paper*, 2021.

Alexander W. Bloedel and Weijie Zhong. The Cost of Optimally Acquired Information. *Working Paper*, 2021.

Andrew Caplin and Mark Dean. Revealed Preference, Rational Inattention, and Costly Information Acquisition. *American Economic Review*, 105(7):2183—2203, 2015.

Andrew Caplin, Mark Dean, and John Leahy. Rational Inattention, Optimal Consideration Sets, and Stochastic Choice. *Review of Economic Studies*, 86:1061—1094, 2019. doi: 10.1093/restud/rdy037/5060717.

Andrew Caplin, Mark Dean, and John Leahy. Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy. *Working Paper*, 2021.

Yan Chen and YingHua He. Information acquisition and provision in school choice: An experimental study. *Journal of Economic Theory*, 197:105345, 2021a. ISSN 0022-0531. doi: 10.1016/j.jet.2021.105345.

Yan Chen and YingHua He. Information acquisition and provision in school choice: a theoretical investigation. *Economic Theory*, 2021b. ISSN 0938-2259. doi: 10.1007/s00199-021-01376-3.

Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley-Interscience, Hoboken, NJ, 2006. ISBN 9780471241959.

Tommaso Denti. Unrestricted Information Acquisition. *Working Paper*, 2020.

Lars Ehlers. Truncation Strategies in Matching Markets. *Mathematics of Operations Research*, 33(2):327–335, 2008. ISSN 0364-765X. doi: 10.1287/moor.1070.0294.

Alexander Frankel and Emir Kamenica. Quantifying Information and Uncertainty. *American Economic Review*, 109(10):3650–3680, 2019. ISSN 0002-8282. doi: 10.1257/aer.20181897.

David Gale and Lloyd Shapley. College Admissions and the Stability of Marriage. *The American Mathematical Monthly*, 69(1):9—15, 1962.

Justine S Hastings and Jeffrey M Weinstein. Information, School Choice, and Academic Achievement: Evidence from Two Experiments. *Quarterly Journal of Economics*, 123(4):1373–1414, 2008. ISSN 0033-5533. doi: 10.1162/qjec.2008.123.4.1373.

Caroline Hoxby and Christopher Avery. The Missing "One-Offs": The Hidden Supply of High-Achieving, Low-Income Students. *Brookings Papers on Economic Activity*, 2013(1):1–65, 2013. ISSN 1533-4465. doi: 10.1353/eca.2013.0000.

Caroline M. Hoxby and Sarah Turner. What High-Achieving Low-Income Students Know About College. *American Economic Review: Papers & Proceedings*, 105(5):514—517, 2015.

Benjamin Hébert and Michael Woodford. Rational Inattention when Decisions Take Time. *Working Paper*, 2021.

Nicole Immorlica, Jacob Leshno, Irene Lo, and Brendan Lucier. Information Acquisition in Matching Markets: The Role of Price Discovery. *SSRN Electronic Journal*, 2020. doi: 10.2139/ssrn.3705049.

Fuhito Kojima and Mihai Manea. Axioms for Deferred Acceptance. *Econometrica*, 78(2):633–653, 2010. ISSN 1468-0262. doi: 10.3982/ecta7443.

David M. Kreps. *Notes on the theory of choice*. Westview Press, Boulder, 1988.

Daniel Martin. Strategic pricing with rational inattention to quality. *Games and Economic Behavior*, 104: 131–145, 2017. ISSN 0899-8256. doi: 10.1016/j.geb.2017.03.007. URL http://dx.doi.org/10.1016/j.geb.2017.03.007.

Filip Matejka and Alisdair McKay. Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model. *American Economic Review*, 105(1):272—298, 2015.

Thayer Morrill. An alternative characterization of the deferred acceptance algorithm. *International Journal of Game Theory*, 42(1):19–28, 2013. ISSN 0020-7276. doi: 10.1007/s00182-011-0311-9.

Christopher A. Neilson. The Rise of Centralized Choice and Assignment Mechanisms in Education Markets around the World, 2019. URL www.ccas-project.org.

Doron Ravid. Ultimatum Bargaining with Rational Inattention. *American Economic Review*, 110(9):2948–2963, 2020. ISSN 0002-8282. doi: 10.1257/aer.20170620.

Alvin E. Roth and Uriel G. Rothblum. Truncation Strategies in Matching Markets—in Search of Advice for Participants. *Econometrica*, 67(1):21–43, 1999. ISSN 1468-0262. doi: 10.1111/1468-0262.00002.

Alvin E. Roth and M. A. O. Sotomayor. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge University Press, New York, 1990.

Christopher A. Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690, 2003. ISSN 0304-3932. doi: 10.1016/s0304-3932(03)00029-1.

A Tables

$l^2 \backslash \theta^2$	GGG	GGB	GBG	BGG	GBB	BGB	BBG	BBB
(a,b,c)	$\frac{z}{Z^2(3)}$	$\frac{z}{Z^2(2)}$	$\frac{z^{\frac{2}{3}}}{Z^2(2)}$	$\frac{z^{\frac{1}{3}}}{Z^2(2)}$	$\frac{z^{\frac{2}{3}}}{Z^2(1)}$	$\frac{z^{\frac{1}{3}}}{Z^2(1)}$	$\frac{1}{Z^2(1)}$	$\frac{1}{Z^2(0)}$
(a,c,b)	$\frac{z}{Z^2(3)}$	$\frac{z^{\frac{2}{3}}}{Z^2(2)}$	$\frac{z}{Z^2(2)}$	$\frac{z^{\frac{1}{3}}}{Z^2(2)}$	$\frac{z^{\frac{2}{3}}}{Z^2(1)}$	$\frac{1}{Z^2(1)}$	$\frac{z^{\frac{1}{3}}}{Z^2(1)}$	$\frac{1}{Z^2(0)}$
(b,a,c)	$\frac{z}{Z^2(3)}$	$\frac{z}{Z^2(2)}$	$\frac{z^{\frac{1}{3}}}{Z^2(2)}$	$\frac{z^{\frac{2}{3}}}{Z^2(2)}$	$\frac{z^{\frac{1}{3}}}{Z^2(1)}$	$\frac{z^{\frac{2}{3}}}{Z^2(1)}$	$\frac{1}{Z^2(1)}$	$\frac{1}{Z^2(0)}$
(b,c,a)	$\frac{z}{Z^2(3)}$	$\frac{z^{\frac{2}{3}}}{Z^2(2)}$	$\frac{z^{\frac{1}{3}}}{Z^2(2)}$	$\frac{z}{Z^2(2)}$	$\frac{1}{Z^2(1)}$	$\frac{z^{\frac{2}{3}}}{Z^2(1)}$	$\frac{z^{\frac{1}{3}}}{Z^2(1)}$	$\frac{1}{Z^2(0)}$
(c,a,b)	$\frac{z}{Z^2(3)}$	$\frac{z^{\frac{1}{3}}}{Z^2(2)}$	$\frac{z}{Z^2(2)}$	$\frac{z^{\frac{2}{3}}}{Z^2(2)}$	$\frac{z^{\frac{1}{3}}}{Z^2(1)}$	$\frac{1}{Z^2(1)}$	$\frac{z^{\frac{2}{3}}}{Z^2(1)}$	$\frac{1}{Z^2(0)}$
(c,b,a)	$\frac{z}{Z^2(3)}$	$\frac{z^{\frac{1}{3}}}{Z^2(2)}$	$\frac{z^{\frac{2}{3}}}{Z^2(2)}$	$\frac{z}{Z^2(2)}$	$\frac{1}{Z^2(1)}$	$\frac{z^{\frac{1}{3}}}{Z^2(1)}$	$\frac{z^{\frac{2}{3}}}{Z^2(1)}$	$\frac{1}{Z^2(0)}$

Table 4: Type-conditional choice probabilities $P^{2*}(l^2|\theta^2)$ of student 2 in the example of Section 3, with $Z^2(3) = 6z$, $Z^2(2) = 2z + 2z^{\frac{2}{3}} + 2z^{\frac{1}{3}}$, $Z^2(1) = 2z^{\frac{2}{3}} + 2z^{\frac{1}{3}} + 2$, and $Z^2(0) = 6$.

B Sequential Search

This section derives the solution of the sequential search model outlined in Section 6.4. The solution is simplified by the fact that searching all three schools is a strictly dominated strategy under any mechanism. Rather than search the third school, the optimal strategy is to list the unsearched school last if the first two are both good, first if they are both bad, and in the middle if one is good and one bad. There are thus five strategies (denoted as indicated in parentheses) whose expected net utilities we consider: search twice regardless of what the first search reveals ("2"); search twice if the first search reveals a bad school and once if it reveals a good school ("2B"); search once if the first is bad, twice if it is good ("2G"); search once regardless ("1"); and do not search at all ("0"). The strategy of not searching at all yields an expected utility of

$$V(0;\mu) = [\mu(3) + 2\mu(2) + \mu(1)]\, u_G \tag{B.1}$$

where the right-hand side is simply the unconditional probability that a randomly selected school is good. To specify the value $V^i(1;\kappa,\mu)$ of searching exactly once, we take account, conditional on each search outcome, of the outcome of the optimal listing strategy, which is to list the one searched school first if it is revealed to be good and last if it is bad. How likely a student using this strategy is matched with a good school depends on the number of good schools, so that

$$V^i(1;\kappa,\mu) = \left[\mu(3) + \mu(2)\left(3h^i(1) + 2h^i(2) + h^i(3)\right) + \mu(1)\left(2h^i(1) + h^i(2)\right)\right] u_G - \kappa. \tag{B.2}$$

If there are 3 good schools, which occurs with probability $\mu(3)$, then the student is sure to match with a good school. With probability $3\mu(2)$ there are two good schools, in which case there is a $2/3$ chance that the first searched will be good and a $1/3$ chance that it is bad. If it is good and thus top-listed, then the second good school will be listed second or third with probability 1/2 each. If it is bad and thus listed last, then the two good schools will be in the first two positions for sure. Hence the first school on the list is certain to be good, the second on the list is good with probability 2/3, and the third on the list is good with probability 1/3. The remaining state in which a good match is possible is the state with 1 good school, which has probability $3\mu(1)$. In this case the one search reveals a good school with probability 1/3 and a bad one otherwise. If the revealed school is the good one then it will be listed first, otherwise the good school has an equal 1/2 chance of being listed first or second. Combining these two cases there is an overall 2/3 chance of the good school being listed first, and a 1/3 chance of it being listed second. Lastly, we subtract the search cost which is incurred once. Analogous logic rationalizes the remaining three equations as expanded on below.

$$\begin{aligned} V^i(2G;\kappa,\mu) &= \left[\mu(3)+3\mu(2)\left(h^i(1)+h^i(2)\right)+\mu(1)\left(2h^i(1)+h^i(2)\right)\right]u_G-\left[1+\mu(3)+2\mu(2)+\mu(1)\right]\kappa \\ V^i(2B;\kappa,\mu) &= \left[\mu(3)+\mu(2)\left(3h^i(1)+2h^i(2)+h^i(3)\right)+3\mu(1)h^i(1)\right]u_G-\left[1+\mu(2)+2\mu(1)+\mu(0)\right]\kappa \\ V^i(2;\kappa,\mu) &= \left[\mu(3)+3\mu(2)\left(h^i(1)+h^i(2)\right)+3\mu(1)h^i(1)\right]u_G-2\kappa \end{aligned}$$

To understand $V^i(2G;\kappa,\mu)$ note that it differs from $V^i(1;\kappa,\mu)$ in only two respects: It is more likely to identify a good school when there are two good ones, and an additional search cost is paid if the first one is good. The key is that when the first-searched school is good, the second search is only beneficial when in fact two schools are good, and in this case the search means that the second good school is certainly listed second, either as the second identified good one or as the unsearched one. Hence the 1/3 probability that the last-listed school is one of the good ones is no longer applicable, as the first- and second-listed school are surely the two good ones. Note that in the state in which only one school is good, the second search after finding the good one has no value.

To understand $V^i(2B;\kappa,\mu)$ note likewise that it differs from $V^i(1;\kappa,\mu)$ in only two respects. This time it is more likely to identify a good school when there is only one good school, and an additional search cost is paid if the first one is bad. The key is that when the first-searched school is bad, the second search is only beneficial when there is only one good school, and in this case the second search ensures that it is certainly listed first. Finally, to understand $V^i(2;\kappa,\mu)$ note that it has both of the conditional benefits associated with searching a second time only following one of the outcomes, at the cost of incurring a second search cost for sure.

Which search strategy attains the highest ex ante net value under DA depends on a student's information costs, prior, as well as their rank, since $h^1_{DA}=(1,0,0)$, $h^2_{DA}=(2/3,1/3,0)$ and $h^3_{DA}=(1/3,1/3,1/3)$. Student 3, as residual claimant, never searches. From the above equations it follows that student 1 stops searching for sure if the first school is revealed to be good, so they are optimizing among the three strategies whose net value is illustrated in the left panel of Figure 8 for the case in which schools have an independent 1/2 probability of being good. For sufficiently high costs, not searching at all is the dominant strategy, while for low enough costs it is optimal to search twice only after the first-searched school is revealed to be bad. Whether or not it is ever optimal to search exactly once regardless of the outcome depends on the

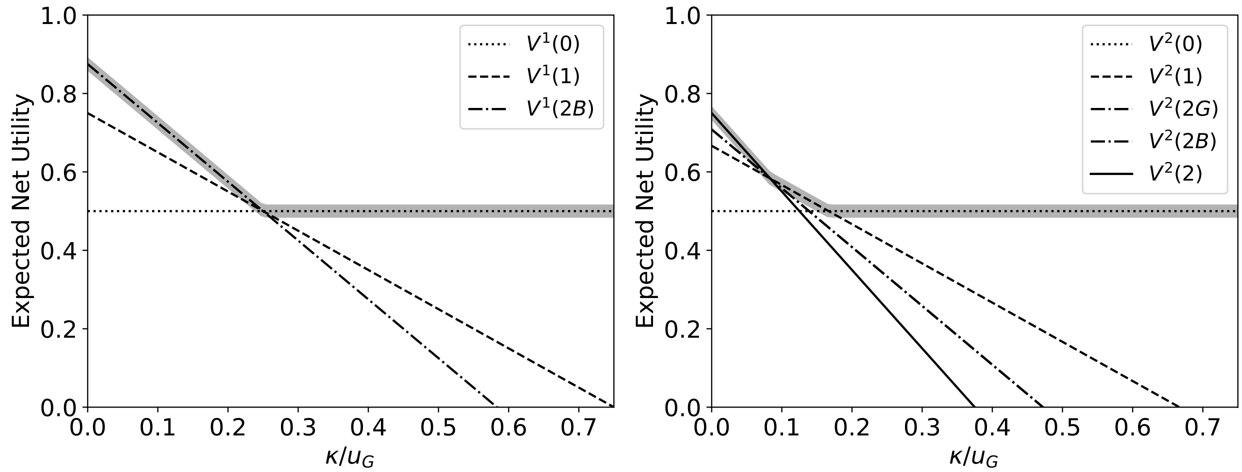

Figure 8: Expected net utility for student 1 (left) and 2 (right) under different sequential search strategies, for $\mu(3) = \mu(2) = \mu(1) = \mu(0) = 1/8$. The optimal strategy at each information cost is highlighted in gray.

parameters.

The set of possibly optimal strategies for student 2 is larger than for student 1. In particular, neither the strategy of searching twice only if the first-searched is good, nor that of always searching twice, is dominated. This most search intensive strategy indeed is optimal as search costs approach zero, with no search still being optimal for sufficiently high search costs. Which, if any, of the other strategies can be optimal depends on the parameters. The right panel of Figure 8 reveals that for the previously discussed independent uniform prior, student 2's optimal search strategy for intermediate information costs is to search exactly once, and thus less than does student 1. In both figures we have marked the upper envelope defining the optimal search strategy, which is convex as the maximum of linear functions. Note that student 2 stops searching altogether at a lower cost of information than student 1. At the lowest information costs, however, student 2 searches more than does student 1, which is reminiscent of the example of Table 3.

The resulting welfare is shown in Figure 7 in the main text, whose properties can be rationalized by a few simple observations. First, in accordance with equations 40 and 41, the full information surplus of student 1 and 2 is

$$V^1(2; \kappa = 0, \mu) - V(0; \mu) \quad = \quad [\mu(2) + 2\mu(1)]\, u_G \tag{B.3}$$

$$V^2(2; \kappa = 0, \mu) - V(0; \mu) \quad = \quad [\mu(2) + \mu(1)]\, u_G. \tag{B.4}$$

As search costs rise initially from zero, it is optimal for student 2 to search twice while student 1 searches less, stopping after one search if the school is good. Relative net welfare in this leftmost region of the figure is therefore

$$\frac{V^1(2B;\kappa,\mu)-V(0;\mu)}{V^1(2;\kappa=0,\mu)-V(0;\mu)} = 1-\frac{1+\mu(2)+2\mu(1)+\mu(0)}{\mu(2)+2\mu(1)}\frac{\kappa}{u_G} \tag{B.5}$$

$$\frac{V^2(2;\kappa,\mu)-V(0;\mu)}{V^2(2;\kappa=0,\mu)-V(0;\mu)} = 1-\frac{2}{\mu(2)+\mu(1)}\frac{\kappa}{u_G}. \tag{B.6}$$

which explains the steeper negative slope of the relative net welfare as search costs rise above zero for student 2 compared to student 1 in Figure 7.

A second qualitative observation concerns the cutoff cost level at which each student stops search altogether. The surplus of each search strategy relative to the no-search benchmark for student 1 is

$$V^1(1;\kappa,\mu)-V(0;\mu) = [\mu(2)+\mu(1)]\,u_G-\kappa \tag{B.7}$$

$$V^1(2B;\kappa,\mu)-V(0;\mu) = [\mu(2)+2\mu(1)]\,u_G-[1+\mu(2)+2\mu(1)+\mu(0)]\,\kappa. \tag{B.8}$$

Hence for student 1 the cutoff $\bar{\kappa}^1$ above which no search is optimal is given by

$$\bar{\kappa}^1=\max\left\{\mu(2)+\mu(1),\frac{\mu(2)+2\mu(1)}{1+\mu(2)+2\mu(1)+\mu(0)}\right\}u_G. \tag{B.9}$$

For student 2

$$V^2(1;\kappa,\mu)-V(0;\mu) = \left[\frac{2}{3}\mu(2)+\frac{2}{3}\mu(1)\right]u_G-\kappa \tag{B.10}$$

$$V^2(2G;\kappa,\mu)-V(0;\mu) = \left[\mu(2)+\frac{2}{3}\mu(1)\right]u_G-[1+\mu(3)+2\mu(2)+\mu(1)]\,\kappa \tag{B.11}$$

$$V^2(2B;\kappa,\mu)-V(0;\mu) = \left[\frac{2}{3}\mu(2)+\mu(1)\right]u_G-[1+\mu(2)+2\mu(1)+\mu(0)]\,\kappa \tag{B.12}$$

$$V^2(2;\kappa,\mu)-V(0;\mu) = [\mu(2)+\mu(1)]\,u_G-2\kappa \tag{B.13}$$

so the cutoff $\bar{\kappa}^2$ above which no search is optimal for student 2 is given by

$$\bar{\kappa}^2=\max\left\{\frac{2}{3}\left(\mu(2)+\mu(1)\right),\frac{\mu(2)+\frac{2}{3}\mu(1)}{1+\mu(3)+2\mu(2)+\mu(1)},\frac{\frac{2}{3}\mu(2)+\mu(1)}{1+\mu(2)+2\mu(1)+\mu(0)}\right\}u_G. \tag{B.14}$$

Qualitatively, the key observation is that student 1 continues to earn surplus at cost levels at which student 2 has stopped searching and effectively picks at random,

$$\bar{\kappa}^1\geq\mu(2)+\mu(1)>\bar{\kappa}^2. \tag{B.15}$$

Hence at high as well as at low levels of the search cost, the reduction in net welfare surplus is greater for student 2 than for student 1, just as in the case with rational inattention.

C Proofs

C.1 Proof of Proposition 1

First note that by Definition 1 there exists, given $\omega, \omega' \in \Omega_k$ for some $1 \leq k \leq K$, a bijection $\alpha_{\omega\omega'} : A \to A$ satisfying equation 2 so that $u(c, \omega) = u(\alpha_{\omega\omega'}(c), \omega')$ for all $c \in A$. It follows that

$$Z(\omega) \equiv \sum_{c \in A} z(c, \omega) = \sum_{c \in A} z(\alpha_{\omega\omega'}(c), \omega') = Z(\omega') \qquad \forall \omega, \omega' \in \Omega_k \tag{C.1}$$

where the last equality follows from the fact that $\alpha_{\omega\omega'}$ is a permutation. Note moreover that by Definition 1 there exists, given any $a, b \in A$ and $1 \leq k \leq K$, a bijection $\pi_{ab} : \Omega_k \to \Omega_k$ satisfying equation 4 so that $u(a, \omega) = u(b, \pi_{ab}(\omega))$ for all $\omega \in \Omega_k$, from which it follows that

$$\sum_{\omega \in \Omega_k} z(a, \omega) = \sum_{\omega \in \Omega_k} z(b, \pi_{ab}(\omega)) = \sum_{\omega \in \Omega_k} z(b, \omega) \qquad \forall a, b \in A \tag{C.2}$$

where the last equality follows from the fact that π_{ab} is a permutation. Definition 1 therefore implies that for any symmetric decision problem it is the case, for all $\omega, \omega' \in \Omega_k$ given some $1 \leq k \leq K$, that $\mu(\omega) = \mu(\omega') =: \mu_k$ by equation 3 and $Z(\omega) = Z(\omega') =: Z_k$ by equation C.1, so that

$$\sum_{\omega \in \Omega_k} \mu(\omega) \frac{z(a, \omega)}{Z(\omega)} = \frac{\mu_k}{Z_k} \sum_{\omega \in \Omega_k} z(a, \omega) = \frac{\mu_k}{Z_k} \sum_{\omega \in \Omega_k} z(b, \omega) = \sum_{\omega \in \Omega_k} \mu(\omega) \frac{z(b, \omega)}{Z(\omega)} \qquad \forall a, b \in A \tag{C.3}$$

where the second equality follows from equation C.2. This implies, for any $a, b \in A$, that

$$\sum_{\omega \in \Omega} \frac{z(a, \omega)}{Z(\omega)} \mu(\omega) = \sum_{k=1}^{K} \sum_{\omega \in \Omega_k} \frac{z(a, \omega)}{Z(\omega)} \mu(\omega) = \sum_{k=1}^{K} \sum_{\omega \in \Omega_k} \frac{z(b, \omega)}{Z(\omega)} \mu(\omega) = \sum_{\omega \in \Omega} \frac{z(b, \omega)}{Z(\omega)} \mu(\omega) \tag{C.4}$$

so that in a symmetric decision problem

$$\sum_{\omega \in \Omega} \frac{z(a, \omega)}{Z(\omega)} \mu(\omega) = 1/|A| \qquad \forall a \in A. \tag{C.5}$$

In a symmetric decision problem, any symmetric strategy $P \in \mathcal{P}(\mu, A)$ therefore satisfies

$$\sum_{\omega \in \Omega} \frac{z(a, \omega)\mu(\omega)}{\sum_{c \in A} z(c, \omega) \sum_{\omega \in \Omega} P(c, \omega)} = |A| \sum_{\omega \in \Omega} \frac{z(a, \omega)}{Z(\omega)} \mu(\omega) = 1 \qquad \forall a \in A \tag{C.6}$$

where the first equality follows from the fact that for a symmetric strategy $\sum_{\omega \in \Omega} P(c, \omega) = 1/|A|$ for all $c \in A$ by definition. The necessary and sufficient conditions for optimality (Caplin et al., 2019, Proposition 1) then imply that a symmetric strategy is a solution if and only if the state-dependent choice probabilities satisfy the necessary conditions of Matejka and McKay (2015), which in this case reduce to equation 10.

C.2 Derivation and Analysis of Equations 32, 33, and 34

The probability of the first-listed school of student 2 being good is obtained immediately by summing, in Table 2, the posterior probabilities of all states in which this is the case, so that (cf. equation 78)

$$p^2(1) = \mu^2(3) + 3\mu^2(2) \frac{z + z^{\frac{2}{3}}}{z + z^{\frac{2}{3}} + z^{\frac{1}{3}}} + 3\mu^2(1) \frac{z^{\frac{2}{3}}}{z^{\frac{2}{3}} + z^{\frac{1}{3}} + 1} \tag{32}$$

The probability that the second-listed school of student 2 is good is given analogously by (cf. equation 78)

$$p^2(2) = \mu^2(3) + 3\mu^2(2)\frac{z + z^{\frac{1}{3}}}{z + z^{\frac{2}{3}} + z^{\frac{1}{3}}} + 3\mu^2(1)\frac{z^{\frac{1}{3}}}{z^{\frac{2}{3}} + z^{\frac{1}{3}} + 1} = \mu^2(3) + 3\mu^2(2) + 3\frac{\left[\mu^2(1) - \mu^2(2)\right] z^{\frac{2}{3}}}{z + z^{\frac{2}{3}} + z^{\frac{1}{3}}}. \quad (33)$$

The sign of its derivative with respect to z only depends on the sign of the derivative's numerator, so that

$$\begin{aligned} sgn\left(\frac{\partial p^2(2)}{\partial z}\right) &= sgn\left(\left[\mu^2(1) - \mu^2(2)\right] \cdot \left[2z^{-\frac{1}{3}}(z + z^{\frac{2}{3}} + z^{\frac{1}{3}}) - z^{\frac{2}{3}}(3 + 2z^{-\frac{1}{3}} + z^{-\frac{2}{3}})\right]\right) \\ &= sgn\left(\left[\mu^2(2) - \mu^2(1)\right] \cdot \left[z^{\frac{2}{3}} - 1\right]\right) \end{aligned}$$

so $p^2(2)$ is increasing in z if and only if $\mu^2(2) > \mu^2(1)$, since $z > 1$. For a tracking problem in which exactly one of the schools is good (so that $\mu^2(1) = 1/3 > 0 = \mu^2(2)$), this reduces to

$$p^2(2) = \frac{z^{\frac{1}{3}}}{z^{\frac{2}{3}} + z^{\frac{1}{3}} + 1}$$

which, due to $z > 1$, is decreasing in z. Conversely, if $\mu^2(2) = 1/3$ then the probability is

$$p^2(2) = 1 - \frac{z^{\frac{1}{3}}}{z^{\frac{2}{3}} + z^{\frac{1}{3}} + 1}$$

which is increasing in z. Student 2's ex ante probability of being matched with a good school is

$$\begin{aligned} \frac{2}{3}p^2(1) + \frac{1}{3}p^2(2) &= \mu^2(3) + \frac{2}{3}\left(3\frac{(\mu^2(2) + \mu^2(1))z + \mu^2(2)z^{\frac{2}{3}}}{z + z^{\frac{2}{3}} + z^{\frac{1}{3}}}\right) + \frac{1}{3}\left(3\mu^2(2) + 3\frac{\left[\mu^2(1) - \mu^2(2)\right] z^{\frac{2}{3}}}{z + z^{\frac{2}{3}} + z^{\frac{1}{3}}}\right) \\ &= \mu^2(3) + \mu^2(2) + (\mu^2(2) + \mu^2(1))\frac{2z + z^{\frac{2}{3}}}{z + z^{\frac{2}{3}} + z^{\frac{1}{3}}} \quad (34) \end{aligned}$$

which is non-decreasing in z.

C.3 Proof of Proposition 2

The gross utility of student 1 follows as a special case of equations 78 and 79 as

$$\begin{aligned} U^{1*}(z, \mu^1) = U^{1*}(z, \mu^1|1) &= \kappa\left[\mu^1(3)\frac{3z\ln z}{3z} + 3\mu^1(2)\frac{2z\ln z}{2z + 1} + 3\mu^1(1)\frac{z\ln z}{z + 2}\right] \\ &= \left[\mu^1(3) + 3\mu^1(2)\frac{2z}{2z + 1} + 3\mu^1(1)\frac{z}{z + 2}\right] u_G \end{aligned}$$

in accordance with equation 30. The gross utility of student 2, conditional on matching with the first or second school on their list, respectively, is obtained from equation 78 as

$$\begin{aligned} U^{2*}(z, \mu^2|1) &= \left[\mu^2(3) + 3\mu^2(2)\frac{z + z^{2/3}}{z + z^{2/3} + z^{1/3}} + 3\mu^2(1)\frac{z^{2/3}}{z^{2/3} + z^{1/3} + 1}\right] u_G \\ U^{2*}(z, \mu^2|2) &= \left[\mu^2(3) + 3\mu^2(2)\frac{z + z^{1/3}}{z + z^{2/3} + z^{1/3}} + 3\mu^2(1)\frac{z^{1/3}}{z^{2/3} + z^{1/3} + 1}\right] u_G \end{aligned}$$

in accordance with equations 32 and 33. The expected gross utility of student 2 is thus given by

$$\begin{aligned} U^{2*}(z, \mu^2) &= \frac{2}{3}U^{2*}(z, \mu^2|1) + \frac{1}{3}U^{2*}(z, \mu^2|2) \\ &= \left[\mu^2(3) + 3\mu^2(2)\frac{z + \frac{2}{3}z^{2/3} + \frac{1}{3}z^{1/3}}{z + z^{2/3} + z^{1/3}} + 3\mu^2(1)\frac{\frac{2}{3}z^{2/3} + \frac{1}{3}z^{1/3}}{z^{2/3} + z^{1/3} + 1}\right] u_G \\ &= \left[\mu^2(3) + \mu^2(2)\frac{3z + 2z^{2/3} + z^{1/3}}{z + z^{2/3} + z^{1/3}} + \mu^2(1)\frac{2z + z^{\frac{2}{3}}}{z + z^{\frac{2}{3}} + z^{\frac{1}{3}}}\right] u_G \end{aligned}$$

or equivalently, using equation 79,

$$\begin{aligned}
U^{2*}(z,\mu^2) &= \left[\mu^2(3)\ln z + 3\mu^2(2)\frac{2z\ln z + 2z^{\frac{2}{3}}\ln z^{\frac{2}{3}} + 2z^{\frac{1}{3}}\ln z^{\frac{1}{3}}}{2z+2z^{\frac{2}{3}}+2z^{\frac{1}{3}}} + 3\mu^2(1)\frac{2z^{\frac{2}{3}}\ln z^{\frac{2}{3}} + 2z^{\frac{1}{3}}\ln z^{\frac{1}{3}}}{2z^{\frac{2}{3}}+2z^{\frac{1}{3}}+2}\right]\kappa \\
&= \left[\mu^2(3)\ln z + 3\mu^2(2)\frac{z\ln z + \frac{2}{3}z^{\frac{2}{3}}\ln z + \frac{1}{3}z^{\frac{1}{3}}\ln z}{z+z^{\frac{2}{3}}+z^{\frac{1}{3}}} + 3\mu^2(1)\frac{\frac{2}{3}z\ln z + \frac{1}{3}z^{\frac{2}{3}}\ln z}{z+z^{\frac{2}{3}}+z^{\frac{1}{3}}}\right]\kappa \\
&= \left[\mu^2(3) + \mu^2(2)\frac{3z+2z^{\frac{2}{3}}+z^{\frac{1}{3}}}{z+z^{\frac{2}{3}}+z^{\frac{1}{3}}} + \mu^2(1)\frac{2z+z^{\frac{2}{3}}}{z+z^{\frac{2}{3}}+z^{\frac{1}{3}}}\right]u_G \\
&= \left[\mu^2(3) + \mu^2(2) + \frac{2z+z^{\frac{2}{3}}}{z+z^{\frac{2}{3}}+z^{\frac{1}{3}}}(\mu^2(2)+\mu^2(1))\right]u_G
\end{aligned}$$

Using equation 35 the gross utility surplus is then obtained as

$$\begin{aligned}
U^{1*}(z,\mu^1) - \underline{U}(\mu^1) &= \left[\mu^1(3) + 3\mu^1(2)\frac{2z}{2z+1} + 3\mu^1(1)\frac{z}{z+2}\right]u_G - \left[\mu^1(3)+2\mu^1(2)+\mu^1(1)\right]u_G \\
&= \left[\left(3\frac{2z}{2z+1}-2\right)\mu^1(2) + \left(3\frac{z}{z+2}-1\right)\mu^1(1)\right]u_G \\
&= \left[\frac{2z-2}{2z+1}\mu^1(2) + \frac{2z-2}{z+2}\mu^1(1)\right]u_G \qquad (36)
\end{aligned}$$

and

$$\begin{aligned}
U^{2*}(z,\mu^2) - \underline{U}(\mu^2) &= \left[\mu^2(3) + \mu^2(2) + \frac{2z+z^{\frac{2}{3}}}{z+z^{\frac{2}{3}}+z^{\frac{1}{3}}}(\mu^2(2)+\mu^2(1))\right]u_G - \left[\mu^2(3)+2\mu^2(2)+\mu^2(1)\right]u_G \\
&= \left[\frac{2z+z^{\frac{2}{3}}}{z+z^{\frac{2}{3}}+z^{\frac{1}{3}}} - 1\right](\mu^2(2)+\mu^2(1))u_G \\
&= \frac{z-z^{\frac{1}{3}}}{z+z^{\frac{2}{3}}+z^{\frac{1}{3}}}\left(\mu^2(2)+\mu^2(1)\right)u_G \qquad (37)
\end{aligned}$$

C.4 Proof of Proposition 3

The proposition is a corollary of Proposition 6: From equation 83 it follows that

$$\begin{aligned}
N^{1*}(z,\mu^1) &= \kappa\left[\mu^1(3)\ln\left(\frac{6z}{6}\right) + 3\mu^1(2)\ln\left(\frac{4z+2}{6}\right) + 3\mu^1(1)\ln\left(\frac{2z+4}{6}\right)\right] \\
&= \left[\mu^1(3)\ln z + 3\mu^1(2)\ln\left(\frac{2z+1}{3}\right) + 3\mu^1(1)\ln\left(\frac{z+2}{3}\right)\right]\frac{u_G}{\ln z} \qquad (44)
\end{aligned}$$

and

$$\begin{aligned}
N^{2*}(z,\mu^2) &= \kappa\left[\mu^2(3)\ln\left(\frac{6z}{6}\right) + 3\mu^2(2)\ln\left(\frac{2z+2z^{\frac{2}{3}}+2z^{\frac{1}{3}}}{6}\right) + 3\mu^2(1)\ln\left(\frac{2z^{\frac{2}{3}}+2z^{\frac{1}{3}}+2}{6}\right)\right] \\
&= \left[\mu^2(3)\ln z + 3\mu^2(2)\ln\left(\frac{z+z^{\frac{2}{3}}+z^{\frac{1}{3}}}{3}\right) + 3\mu^2(1)\ln\left(\frac{z^{\frac{2}{3}}+z^{\frac{1}{3}}+1}{3}\right)\right]\frac{u_G}{\ln z} \qquad (45)
\end{aligned}$$

and

$$N^{3*}(z,\mu^3) = \left(\mu^3(3)+2\mu^3(2)+\mu^3(1)\right)u_G \qquad (46)$$

C.5 Derivations of Section 3.5

Relative Gross Welfare (Equation 48 and Figure 2)

From Proposition 2 and equations 40 and 41, it follows by the definition of equation 42 that students' relative gross welfare is

$$\hat{U}^1(z,\mu^1) = \frac{2z-2}{2z+1}\frac{\mu^1(2)}{\mu^1(2)+2\mu^1(1)} + \frac{z-1}{z+2}\frac{2\mu^1(1)}{\mu^1(2)+2\mu^1(1)} \tag{C.7}$$

$$\hat{U}^2(z,\mu^2) = \frac{z-z^{\frac{1}{3}}}{z+z^{\frac{2}{3}}+z^{\frac{1}{3}}} \tag{C.8}$$

so that, for all μ^1 and μ^2 and all $z>1$,

$$1 \geq \frac{2z-2}{2z+1} \geq \hat{U}^1(z,\mu^1) \geq \frac{z-1}{z+2} > \hat{U}^2(z,\mu^2) > 0. \tag{48}$$

Relative Net Welfare (Equation 49 and Figure 3)

Recalling that

$$\underline{N}(\mu^i) = \underline{U}(\mu^i) = \left(\mu^i(3)+2\mu^i(2)+\mu^i(1)\right)u_G$$

it follows from Proposition 3 that

$$\begin{aligned} N^{1*}(z,\mu^1) - \underline{N}(\mu^1) &= \left[\mu^1(3)\ln z + 3\mu^1(2)\ln\left(\frac{2z+1}{3}\right) + 3\mu^1(1)\ln\left(\frac{z+2}{3}\right)\right]\frac{u_G}{\ln z} - \left(\mu^1(3)+2\mu^1(2)+\mu^1(1)\right)u_G \\ &= \left[\left(3\ln\left(\frac{2z+1}{3}\right) - 2\ln(z)\right)\mu^1(2) + \left(3\ln\left(\frac{z+2}{3}\right)-\ln(z)\right)\mu^1(1)\right]\frac{u_G}{\ln z} \end{aligned}$$

and

$$\begin{aligned} N^{2*}(z,\mu^2) - \underline{N}(\mu^2) &= \left[\mu^2(3)\ln z + 3\mu^2(2)\ln\left(\frac{z+z^{2/3}+z^{1/3}}{3}\right) + 3\mu^2(1)\ln\left(\frac{z^{2/3}+z^{1/3}+1}{3}\right)\right]\frac{u_G}{\ln z} \\ &\quad - \left(\mu^2(3)+2\mu^2(2)+\mu^2(1)\right)u_G \\ &= \left[\left(3\ln\left(\frac{z+z^{\frac{2}{3}}+z^{\frac{1}{3}}}{3}\right) - 2\ln(z)\right)\mu^2(2) + \left(3\ln\left(\frac{z^{\frac{2}{3}}+z^{\frac{1}{3}}+1}{3}\right)-\ln(z)\right)\mu^2(1)\right]\frac{u_G}{\ln z}. \end{aligned}$$

Since

$$\overline{N}^i(\mu^i) - \underline{N}(\mu^i) = \overline{U}^i(\mu^i) - \underline{U}(\mu^i)$$

students' relative net welfare as defined in equation 47 is thus

$$\hat{N}^1(z,\mu^1) = \left(3\ln\left(\frac{2z+1}{3}\right) - 2\ln(z)\right)\frac{1}{\ln(z)}\frac{\mu^1(2)}{\mu^1(2)+2\mu^1(1)} + \left(3\ln\left(\frac{z+2}{3}\right)-\ln(z)\right)\frac{1}{\ln(z)}\frac{\mu^1(1)}{\mu^1(2)+2\mu^1(1)} \tag{C.9}$$

and

$$\begin{aligned}
\hat{N}^2(z,\mu^2) &= \left(3\ln\left(\frac{z+z^{\frac{2}{3}}+z^{\frac{1}{3}}}{3}\right)-2\ln(z)\right)\frac{1}{\ln(z)}\frac{\mu^2(2)}{\mu^2(2)+\mu^2(1)} \\
&\quad+\left(3\ln\left(\frac{z^{\frac{2}{3}}+z^{\frac{1}{3}}+1}{3}\right)-\ln(z)\right)\frac{1}{\ln(z)}\frac{\mu^2(1)}{\mu^2(2)+\mu^2(1)} \\
&= \left(3\left[\ln\left(\frac{z^{\frac{2}{3}}+z^{\frac{1}{3}}+1}{3}\right)+\frac{1}{3}\ln(z)\right]-2\ln(z)\right)\frac{1}{\ln(z)}\frac{\mu^2(2)}{\mu^2(2)+\mu^2(1)} \\
&\quad+\left(3\ln\left(\frac{z^{\frac{2}{3}}+z^{\frac{1}{3}}+1}{3}\right)-\ln(z)\right)\frac{1}{\ln(z)}\frac{\mu^2(1)}{\mu^2(2)+\mu^2(1)} \\
&= 3\ln\left(\frac{z^{\frac{2}{3}}+z^{\frac{1}{3}}+1}{3}\right)/\ln(z)-1. \qquad \text{(C.10)}
\end{aligned}$$

Since $z>1$, note that for any μ^1

$$3\ln\left(\frac{2z+1}{3}\right)/\ln(z)-2 \geq \hat{N}^1(z,\mu^1) \geq \frac{1}{2}\left(3\ln\left(\frac{z+2}{3}\right)/\ln(z)-1\right)$$

so that

$$3\frac{\ln\left(\frac{2z+1}{3}\right)}{\ln(z)}-2 \geq \hat{N}^1(z,\mu^1) \geq \frac{3\ln\left(\frac{z+2}{3}\right)}{2\ln(z)}-\frac{1}{2} > 3\frac{\ln\left(\frac{z^{\frac{2}{3}}+z^{\frac{1}{3}}+1}{3}\right)}{\ln(z)}-1=\hat{N}^2(z,\mu^2)>0. \qquad (49)$$

C.6 Proof of Lemma 1

Given $\boldsymbol{P}^{-i}$ and fixing the "numeraire" list $l^I=(1,\ldots,M)$, note that for any permutation $\alpha: X\to X$

$$\begin{aligned}
\pi^i_{\boldsymbol{x}}(\alpha(x)|l^I_\alpha,\boldsymbol{P}^{-i}) &= \sum_{\boldsymbol{l}^{-i}\in\boldsymbol{A}^{-i}}\sum_{\boldsymbol{\theta}^{-i}\in\boldsymbol{\Theta}^{-i}}\boldsymbol{P}^{-i}(\boldsymbol{l}^{-i},\boldsymbol{\theta}^{-i})\mathbf{1}\left\{\boldsymbol{x}^i(l^I_\alpha,\boldsymbol{l}^{-i})=\alpha(x)\right\} \\
&= \sum_{\boldsymbol{l}^{-i}\in\boldsymbol{A}^{-i}}\sum_{\boldsymbol{\theta}^{-i}\in\boldsymbol{\Theta}^{-i}}\boldsymbol{P}^{-i}(\boldsymbol{l}^{-i}_\alpha,\boldsymbol{\theta}^{-i})\mathbf{1}\left\{\boldsymbol{x}^i(l^I_\alpha,\boldsymbol{l}^{-i}_\alpha)=\alpha(x)\right\} \\
&= \sum_{\boldsymbol{l}^{-i}\in\boldsymbol{A}^{-i}}\mathbf{1}\left\{\boldsymbol{x}^i(l^I_\alpha,\boldsymbol{l}^{-i}_\alpha)=\alpha(x)\right\}\prod_{j\neq i}\sum_{\theta^j\in\Theta^j}P^j(l^j_\alpha,\theta^j) \\
&= \sum_{\boldsymbol{l}^{-i}\in\boldsymbol{A}^{-i}}\mathbf{1}\left\{\alpha(\boldsymbol{x}^i(l^I,\boldsymbol{l}^{-i}))=\alpha(x)\right\}\prod_{j\neq i}\sum_{\theta^j\in\Theta^j}P^j(l^j_\alpha,\theta^j) \\
&= \sum_{\boldsymbol{l}^{-i}\in\boldsymbol{A}^{-i}}\mathbf{1}\left\{\boldsymbol{x}^i(l^I,\boldsymbol{l}^{-i})=x\right\}\prod_{j\neq i}\sum_{\theta^j\in\Theta^j}P^j(l^j_\alpha,\theta^j) \\
&= \sum_{\boldsymbol{l}^{-i}\in\boldsymbol{A}^{-i}}\mathbf{1}\left\{\boldsymbol{x}^i(l^I,\boldsymbol{l}^{-i})=x\right\}\prod_{j\neq i}\sum_{\theta^j\in\Theta^j}P^j(l^j,\theta^j)
\end{aligned}$$

where the first equality is by definition of equation 58, the second equality follows from observing that summing over all $\boldsymbol{l}^{-i}_\alpha$ amounts to summing over all $\boldsymbol{l}^{-i}$ because α is a permutation, the fourth equality follows by equation 57 from the fact that $\boldsymbol{x}^i$ is an anonymous mechanism (Definition 3), the fifth equality follows from α being a permutation, and the last equality follows by equation 9 from the symmetry of all P^j. Student i's probability of being matched with the m-th school on any list $l^i \in A^i$ is then obtained, by setting $\alpha = l^i$ and noting that $l^I_{l^i} = l^i$, as

$$\pi^i_{\boldsymbol{x}}(l^i(m)|l^i, \boldsymbol{P}^{-i}) = \sum_{\boldsymbol{l}^{-i} \in \boldsymbol{A}^{-i}} \mathbf{1}\left\{\boldsymbol{x}^i(l^I, \boldsymbol{l}^{-i}) = m\right\} \prod_{j \neq i} \sum_{\theta^j \in \Theta^j} P^j(l^j, \theta^j) =: h^i(m) \tag{C.11}$$

for all $m \in \{1, \dots, M\}$. We have shown that if others' strategies are symmetric, then there exists h^i satisfying equation 59 which is independent of l^i. To complete the proof, note that this is indeed a probability mass function, as

$$\begin{aligned}
\sum_{m=1}^{M} h^i(m) &= \sum_{m=1}^{M} \sum_{\boldsymbol{l}^{-i} \in \boldsymbol{A}^{-i}} \mathbf{1}\left\{\boldsymbol{x}^i(l^I, \boldsymbol{l}^{-i}) = m\right\} \prod_{j \neq i} \sum_{\theta^j \in \Theta^j} P^j(l^j, \theta^j) \\
&= \sum_{\boldsymbol{l}^{-i} \in \boldsymbol{A}^{-i}} \prod_{j \neq i} \sum_{\theta^j \in \Theta^j} P^j(l^j, \theta^j) \sum_{m=1}^{M} \mathbf{1}\left\{\boldsymbol{x}^i(l^I, \boldsymbol{l}^{-i}) = m\right\} \\
&= \sum_{\boldsymbol{l}^{-i} \in \boldsymbol{A}^{-i}} \prod_{j \neq i} \sum_{\theta^j \in \Theta^j} P^j(l^j, \theta^j) \\
&= 1.
\end{aligned}$$

C.7 Proof of Lemma 2

Define a partition $\{\Theta^i_k\}_{1 \leq k \leq K}$ of Θ^i such that $\theta, \theta' \in \Theta^i_k$ for some k if and only if there exists a permutation $\rho_{\theta\theta'} : X \to X$ of schools such that

$$\theta'_{\rho_{\theta\theta'}(x)} = \theta_x \tag{C.12}$$

for all $x \in X$.

We first show that this partition conforms to the requirement that $\theta, \theta' \in \Theta^i_k$ for some k if and only if there exists a bijection $\alpha_{\theta\theta'} : A^i \to A^i$ satisfying equation 2. To see that $\theta, \theta' \in \Theta^i_k$ only if there exists such an $\alpha_{\theta\theta'}$, define, for any $\theta, \theta' \in \Theta^i_k$, the mapping $\alpha_{\theta\theta'}$ as

$$\left(\alpha_{\theta\theta'}(l)\right)(m) = \rho_{\theta\theta'}(l(m)) \qquad \forall l \in A^i,\, m \in \{1, \dots, M\} \tag{C.13}$$

and note that it is a bijection since it is constructed from a concatenation of two permutations of schools ($\rho_{\theta\theta'}$ and l). It follows, for all $l \in A^i$, that

$$u^i(\alpha_{\theta\theta'}(l), \theta') = \sum_{m=1}^{M} h^i(m)\theta'_{(\alpha_{\theta\theta'}(l))(m)} = \sum_{m=1}^{M} h^i(m)\theta'_{\rho_{\theta\theta'}(l(m))} = \sum_{m=1}^{M} h^i(m)\theta_{l(m)} = u^i(l, \theta) \tag{C.14}$$

where the second equality follows from equation C.13 and the third equality from equation C.12. To see the converse direction, namely that existence of *some* permutation $\alpha_{\theta\theta'}$ satisfying equation C.14 implies $\theta, \theta' \in \Theta^i_k$ for some k, note that for the preceding equation to hold for arbitrary h^i, it must be the case that

$$\theta'_{(\alpha_{\theta\theta'}(l))(m)} = \theta_{l(m)}. \tag{C.15}$$

Since $(\alpha_{\theta\theta'}(l)) \circ l^{-1}$ is a permutation of schools (because $\alpha_{\theta\theta'}(l) \in A^i$ and $l \in A^i$ are themselves permutations of schools), this implies that there exists a permutation of schools which, by equation C.15, satisfies equation C.12, so that $\theta, \theta' \in \Theta^i_k$ by definition. The partition defined via the permutation $\rho_{\theta\theta'}$ satisfying equation C.12 thus conforms to the requirements of Definition 1.

Given this partition, we can now show that the problem of student i is symmetric. First note that, given some k, equation 3 is satisfied for all $\theta, \theta' \in \Theta^i_k$, since

$$\mu^i(\theta'_1, \ldots, \theta'_M) = \mu^i(\theta'_{\rho_{\theta\theta'}(1)}, \ldots, \theta'_{\rho_{\theta\theta'}(M)}) = \mu^i(\theta_1, \ldots, \theta_M)$$

where the first equality follows from μ^i being exchangeable and the second equality from equation C.12. Next, define, for any $l, l' \in A^i$ and k, the mapping $\pi_{ll'} : \Theta^i_k \to \Theta^i_k$ as

$$(\pi_{ll'}(\theta))_{l'(m)} = \theta_{l(m)} \qquad \forall \theta \in \Theta^i_k,\ m \in \{1, \ldots, M\}. \tag{C.16}$$

Note that $\pi_{ll'}(\theta) \in \Theta^i_k$ for all $\theta \in \Theta^i_k$, since $l' \circ l^{-1}$ is a permutation of schools which, by equation C.16, satisfies equation C.12, from which it follows that $\theta \in \Theta^i_k$ implies $\pi_{ll'}(\theta) \in \Theta^i_k$. Note moreover that $\pi_{ll'}$ is a bijection (since l and l' are permutations of schools) that satisfies equation 4 for all $\theta \in \Theta^i_k$, since

$$u^i(l', \pi_{ll'}(\theta)) = \sum_{m=1}^{M} h^i(m) \left(\pi_{ll'}(\theta)\right)_{l'(m)} = \sum_{m=1}^{M} h^i(m) \theta_{l(m)} = u^i(l, \theta)$$

where the second equality follows from equation C.16. This concludes the proof.

C.8 Proof of Proposition 4

The fact that, given the unanimous ranking of students by schools, the matching resulting from the Deferred Acceptance Algorithm is equivalent to that resulting from serial dictatorship follows from Theorem 3 of Kojima and Manea (2010) which states that deferred acceptance is the unique allocation rule that is stable and satisfies weak Maskin monotonicity.[5] To see this, first note that the (exogenously specified) priority structure C (the same at all schools) induced by schools' unanimous ranking of students is substitutable[6] and, since all students are acceptable, acceptant.

Let φ be the allocation rule given by serial dictatorship. The following arguments apply to students' rank-order lists which amount to strict "preferences" in the classical full-information sense, even though actual preferences in our setting need not be. Note that φ is a stable allocation rule at the priority structure C. Towards a contradiction, assume that there is a blocking pair. Since $N = c \cdot M$ so that there are no empty seats, this must consist of two students, $i < i'$ w.l.o.g., where student i listed school $\tilde{x}$ that student i' is matched with above their current match, and vice versa. But this is not possible under serial dictatorship, for student i would have been matched with school $\tilde{x}$, as they are ranked above student i'. Note also that φ

[5] Also see Morrill (2013, p. 20).

[6] E.g. by virtue of inducing a responsive (Roth and Sotomayor, 1990, Definition 4 of Chapter 16) preference relation over sets of students (Roth and Sotomayor, 1990, p. 499).

satisifes weak Maskin monotonicity (Morrill, 2013, p. 20). Then by Theorem 3 of Kojima and Manea (2010) it must be the case that φ is the deferred acceptance rule for C.

C.9 Auxiliary Lemma C.1

Lemma C.1 (List-Position Sufficiency under DA). *If, for any $i \in I$, P^j is symmetric for all $j \neq i$ in the sense of Definition 2, then under DA*

$$\pi^i_{\boldsymbol{x}_{DA}}(l^i(m)|l^i, \boldsymbol{P}^{-i}) = h^i_{DA}(m) = \sum_{\boldsymbol{c}^{i-1} \in \boldsymbol{\mathcal{C}}^{i-1}} \mathbf{1}\left\{m = \min_{\eta}\{\eta : c^{i-1}_{\eta} > 0\}\right\} \mathbb{P}(\boldsymbol{c}^{i-1}|P^1, \ldots, P^{i-1}) \tag{C.17}$$

for all $l^i \in A^i$ and all $m \in \{1, \ldots, M\}$, where $\mathbb{P}(\boldsymbol{c}^{i-1}|P^1, \ldots, P^{i-1})$ is as defined in equation 62.

Proof. Define, for any $n = 0, \ldots, N-1$, the vector-valued function

$$\boldsymbol{C}^n : \times_{j \leq n} A^j \to \mathbb{N}^M$$

whose element

$$C^n_x(l^1, \ldots, l^n) = c - \sum_{i=1}^{n} \mathbf{1}\{\boldsymbol{x}^i_{DA}(l^i, \boldsymbol{l}^{-i}) = x\} \tag{C.18}$$

is the remaining capacity of school x after students $1, \ldots, n$ submitting lists $l^1, \ldots, l^n$ have been assigned. Note, for any $i \in I$, that by serial dictatorship (Proposition 4)

$$\boldsymbol{x}^i_{DA}(l^i, \boldsymbol{l}^{-i}) = l^i \left(\min_{\eta}\{\eta : C^{i-1}_{l^i(\eta)}(l^1, \ldots, l^{i-1}) > 0\}\right). \tag{C.19}$$

Since DA is an anonymous mechanism (Lemma 3), it follows from Lemma 1 that, given a symmetric $\boldsymbol{P}^{-i}$, student i's probability of being matched with the m-th school on their rank-order list $l^i \in A^i$ is, for any $m = 1, \ldots, M$ and $l^i \in A^i$, given by

$$\begin{aligned}
&\pi^i_{\boldsymbol{x}_{DA}}(l^i(m)|l^i, \boldsymbol{P}^{-i}) \\
&= \sum_{\boldsymbol{l}^{-i} \in \boldsymbol{A}^{-i}} \mathbf{1}\left\{\boldsymbol{x}^i_{DA}(l^I, \boldsymbol{l}^{-i}) = m\right\} \prod_{j \neq i} \sum_{\theta^j \in \Theta^j} P^j(l^j, \theta^j) \\
&= \sum_{\boldsymbol{l}^{-i} \in \boldsymbol{A}^{-i}} \mathbf{1}\left\{\min_{\eta}\{\eta : C^{i-1}_{\eta}(l^1, \ldots, l^{i-1}) > 0\} = m\right\} \prod_{j \neq i} \sum_{\theta^j \in \Theta^j} P^j(l^j, \theta^j) \\
&= \sum_{\boldsymbol{c}^{i-1} \in \boldsymbol{\mathcal{C}}^{i-1}} \mathbf{1}\left\{\min_{\eta}\{\eta : c^{i-1}_{\eta} > 0\} = m\right\} \sum_{\boldsymbol{l}^{-i} \in \boldsymbol{A}^{-i}} \mathbf{1}\left\{\boldsymbol{C}^{i-1}(l^1, \ldots, l^{i-1}) = \boldsymbol{c}^{i-1}\right\} \prod_{j \neq i} \sum_{\theta^j \in \Theta^j} P^j(l^j, \theta^j) \\
&= \sum_{\boldsymbol{c}^{i-1} \in \boldsymbol{\mathcal{C}}^{i-1}} \mathbf{1}\left\{\min_{\eta}\{\eta : c^{i-1}_{\eta} > 0\} = m\right\} \sum_{l^1 \in A^1} \cdots \sum_{l^{i-1} \in A^{i-1}} \mathbf{1}\left\{\boldsymbol{C}^{i-1}(l^1, \ldots, l^{i-1}) = \boldsymbol{c}^{i-1}\right\} \prod_{j < i} \sum_{\theta^j \in \Theta^j} P^j(l^j, \theta^j) \\
&= \sum_{\boldsymbol{c}^{i-1} \in \boldsymbol{\mathcal{C}}^{i-1}} \mathbf{1}\left\{\min_{\eta}\{\eta : c^{i-1}_{\eta} > 0\} = m\right\} \mathbb{P}(\boldsymbol{c}^{i-1}|P^1, \ldots, P^{i-1}) \qquad \text{(C.17)} \\
&=: h^i_{DA}(m)
\end{aligned}$$

where the first equality follows from equation C.11, the second equality from equation C.19 applied to the "numeraire" list $l^I = (1, \ldots, M)$, and the fifth equality from the definition of equation 62. □

C.10 Proof of Lemma 3

Since $\boldsymbol{x}_{DA}$ is an anonymous mechanism, the remaining capacities function of equation C.18 satisfies, for any permutation $\alpha : X \to X$ of schools and l^i_α as in Definition 3,

$$\begin{aligned} C^n_{\alpha(x)}(l^1_\alpha, \dots, l^n_\alpha) &= c - \sum_{i=1}^n \mathbf{1}\{\boldsymbol{x}^i_{DA}(l^i_\alpha, \boldsymbol{l}^{-i}_\alpha) = \alpha(x)\} \\ &= c - \sum_{i=1}^n \mathbf{1}\{\alpha(\boldsymbol{x}^i_{DA}(l^i, \boldsymbol{l}^{-i})) = \alpha(x)\} \\ &= c - \sum_{i=1}^n \mathbf{1}\{\boldsymbol{x}^i_{DA}(l^i, \boldsymbol{l}^{-i}) = x\} \\ &= C^n_x(l^1, \dots, l^n) \end{aligned} \tag{C.20}$$

where the second equality follows from equation 57 and the third equality from α being a permutation. It follows that if $P^1, \dots, P^n$ are symmetric, then equation 63 is satisfied for all $\boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n$, as

$$\begin{aligned} &\mathbb{P}(c^n_{\alpha(1)}, \dots, c^n_{\alpha(M)} | P^1, \dots, P^n) \\ &= \sum_{l^1 \in A^1} \cdots \sum_{l^n \in A^n} \mathbf{1}\left\{\boldsymbol{C}^n(l^1, \dots, l^n) = \boldsymbol{c}^n_\alpha\right\} \prod_{j=1}^n \sum_{\theta^j \in \Theta^j} P^j(l^j, \theta^j) \\ &= \sum_{l^1 \in A^1} \cdots \sum_{l^n \in A^n} \prod_{x \in X} \mathbf{1}\left\{C^n_x(l^1, \dots, l^n) = c^n_{\alpha(x)}\right\} \prod_{j=1}^n \sum_{\theta^j \in \Theta^j} P^j(l^j, \theta^j) \\ &= \sum_{l^1 \in A^1} \cdots \sum_{l^n \in A^n} \prod_{x \in X} \mathbf{1}\left\{C^n_{\alpha(x)}(l^1_\alpha, \dots, l^n_\alpha) = c^n_{\alpha(x)}\right\} \prod_{j=1}^n \sum_{\theta^j \in \Theta^j} P^j(l^j, \theta^j) \\ &= \sum_{l^1 \in A^1} \cdots \sum_{l^n \in A^n} \prod_{x \in X} \mathbf{1}\left\{C^n_x(l^1_\alpha, \dots, l^n_\alpha) = c^n_x\right\} \prod_{j=1}^n \sum_{\theta^j \in \Theta^j} P^j(l^j_\alpha, \theta^j) \\ &= \sum_{l^1 \in A^1} \cdots \sum_{l^n \in A^n} \mathbf{1}\left\{\boldsymbol{C}^n(l^1, \dots, l^n) = \boldsymbol{c}^n\right\} \prod_{j=1}^n \sum_{\theta^j \in \Theta^j} P^j(l^j, \theta^j) \\ &= \mathbb{P}(c^n_1, \dots, c^n_M | P^1, \dots, P^n) \end{aligned}$$

where the first equality follows from the definition of equation 62, the third equality from equation C.20, the fourth equality from α being a permutation and by equation 9 from the symmetry of all P^j, and the fifth equality from the fact that α is a permutation and that we are summing over all $l^1, \dots, l^n$.

C.11 Proof of Corollary 1

Define, given any $\boldsymbol{k}^n$ such that $\exists\, \boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)$ with $\mathbb{P}(\boldsymbol{c}^n | P^1, \dots, P^n) > 0$, the conditional probability

$$\mathbb{P}(\boldsymbol{c}^n | \boldsymbol{k}^n, P^1, \dots, P^n) \equiv \frac{\mathbb{P}(\boldsymbol{c}^n | P^1, \dots, P^n)}{\mathbb{P}(\boldsymbol{k}^n | P^1, \dots, P^n)} \mathbf{1}\{\boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)\} \tag{C.21}$$

where

$$\mathbb{P}(\boldsymbol{k}^n | P^1, \dots, P^n) \equiv \sum_{\boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)} \mathbb{P}(\boldsymbol{c}^n | P^1, \dots, P^n). \tag{C.22}$$

Corollary. *If $P^1, \ldots, P^n$ are symmetric, then, given any $\boldsymbol{k}^n$ s.t. $\exists\, \boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)$ with $\mathbb{P}(\boldsymbol{c}^n | P^1, \ldots, P^n) > 0$,*

$$\mathbb{P}(\boldsymbol{c}^n | \boldsymbol{k}^n, P^1, \ldots, P^n) = \frac{\mathbf{1}\{\boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)\}}{|\boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)|}. \tag{C.23}$$

Proof. If $\boldsymbol{c}^n, \tilde{\boldsymbol{c}}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)$, then by the definition of $\boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)$ in equation 64

$$\sum_{x \in X} \mathbf{1}\{c_x^n = r\} = k_r^n = \sum_{x \in X} \mathbf{1}\{\tilde{c}_x^n = r\} \qquad \forall r = 0, \ldots, c$$

This implies that there exists a permutation $\alpha : X \to X$ of schools such that

$$\tilde{c}_x^n = c_{\alpha(x)}^n \qquad \forall x \in X$$

If $P^1, \ldots, P^n$ are symmetric then, by Lemma 3, $\mathbb{P}(\boldsymbol{c}^n | P^1, \ldots, P^n)$ is exchangeable, so it follows that

$$\mathbb{P}(\boldsymbol{c}^n | P^1, \ldots, P^n) = \mathbb{P}(\tilde{\boldsymbol{c}}^n | P^1, \ldots, P^n) \qquad \forall \boldsymbol{c}^n, \tilde{\boldsymbol{c}}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)$$

and therefore

$$\mathbb{P}(\boldsymbol{c}^n | \boldsymbol{k}^n, P^1, \ldots, P^n) = \mathbb{P}(\tilde{\boldsymbol{c}}^n | \boldsymbol{k}^n, P^1, \ldots, P^n) \qquad \forall \boldsymbol{c}^n, \tilde{\boldsymbol{c}}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)$$

which implies the result. □

C.12 Proof of Lemma 4

Note that

$$|\boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)| = M! / \prod_{r=0}^{c} k_r^n!$$

which is the number of ways in which the remaining capacities of $\boldsymbol{k}^n$ can be assigned to the M schools. It follows that

$$\begin{aligned} \frac{1}{|\boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)|} \sum_{\boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)} \mathbf{1}\left\{\bigcap_{\eta=1}^{m} \left(c_{l^i(\eta)}^n = 0\right)\right\} &= \frac{\prod_{r=0}^{c} k_r^n!}{M!} \sum_{\boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)} \mathbf{1}\left\{\bigcap_{\eta=1}^{m} \left(c_{l^i(\eta)}^n = 0\right)\right\} \\ &= \frac{\prod_{r=0}^{c} k_r^n!}{M!} \frac{(M-m)!}{(k_0^n - m)! \prod_{r=1}^{c} k_r^n!} \\ &= \frac{(M-m)! k_0^n!}{M!(k_0^n - m)!} \end{aligned} \tag{C.24}$$

where the second equality follows from counting the number of vectors $\boldsymbol{c}^n$ such that the first m schools on list l^i have no remaining capacity. The result then follows from observing that, for all $m \in \{1, \ldots, M\}$,

$$
\begin{aligned}
&\frac{1}{|\boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)|} \sum_{\boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)} \mathbf{1}\left\{m = \min_{\eta}\{\eta : c^n_{l^i(\eta)} > 0\}\right\} \\
&= \frac{1}{|\boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)|} \sum_{\boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)} \mathbf{1}\left\{\left(c^n_{l^i(m)} > 0\right) \bigcap_{\eta=1}^{m-1} \left(c^n_{l^i(\eta)} = 0\right)\right\} \\
&= \frac{1}{|\boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)|} \sum_{\boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)} \left[\mathbf{1}\left\{\bigcap_{\eta=1}^{m-1} \left(c^n_{l^i(\eta)} = 0\right)\right\} - \mathbf{1}\left\{\bigcap_{\eta=1}^{m} \left(c^n_{l^i(\eta)} = 0\right)\right\}\right] \\
&= \frac{(M-m+1)!k_0^n!}{M!(k_0^n - m + 1)!} - \frac{(M-m)!k_0^n!}{M!(k_0^n - m)!} \\
&= \left(\frac{M-m+1}{k_0^n - m + 1} - 1\right) \frac{(M-m)!k_0^n!}{M!(k_0^n - m)!} \\
&= \left(\frac{M - k_0^n}{k_0^n - m + 1}\right) \frac{(M-m)!k_0^n!}{M!(k_0^n - m)!} \\
&= \frac{(M-m)!}{(k_0^n - m + 1)!(M - k_0^n - 1)!} \frac{(M-k_0^n)!k_0^n!}{M!} \\
&= \binom{M-m}{k_0^n + 1 - m} \Big/ \binom{M}{k_0^n}
\end{aligned}
$$

where the third equality follows from equation C.24.

C.13 Auxiliary Lemma C.2

Lemma C.2. *Given $\boldsymbol{c}^{i-1}$ (and thus $\boldsymbol{k}^{i-1}$), if P^i is symmetric then*

$$
\tilde{\pi}^i(x|P^i, \boldsymbol{c}^{i-1}) \equiv \sum_{l^i \in A^i} \mathbf{1}\left\{l^i\left(\min_{\eta}\{\eta : c^{i-1}_{l^i(\eta)} > 0\}\right) = x\right\} \sum_{\theta^i \in \Theta^i} P^i(l^i, \theta^i) = \frac{\mathbf{1}\{c_x^{i-1} > 0\}}{M - k_0^{i-1}} \tag{C.25}
$$

for all $x \in X$.

First note that the number of lists whose first school with positive remaining capacity is a given $x \in X$ is

$$
\begin{aligned}
&\sum_{l^i \in A^i} \mathbf{1}\left\{l^i\left(\min_{\eta}\{\eta : c^{i-1}_{l^i(\eta)} > 0\}\right) = x\right\} \\
&= \sum_{m=1}^{M} \sum_{l^i \in A^i} \mathbf{1}\{l^i(m) = x\} \mathbf{1}\left\{m = \min_{\eta}\{\eta : c^{i-1}_{l^i(\eta)} > 0\}\right\} \\
&= \sum_{m=1}^{M} \sum_{l^i \in A^i} \mathbf{1}\{l^i(m) = x\} \mathbf{1}\left\{c^{i-1}_{l^i(m)} > 0\right\} \mathbf{1}\left\{\bigcap_{\eta=1}^{m-1} \left(c^{i-1}_{l^i(\eta)} = 0\right)\right\} \\
&= \mathbf{1}\left\{c_x^{i-1} > 0\right\} \sum_{m=1}^{M} \sum_{l^i \in A^i} \mathbf{1}\{l^i(m) = x\} \mathbf{1}\left\{\bigcap_{\eta=1}^{m-1} \left(c^{i-1}_{l^i(\eta)} = 0\right)\right\}.
\end{aligned} \tag{C.26}
$$

It remains to count the number of ways in which the $M-1$ schools *other than* $l^i(m) = x$ can be arranged in the $M-1$ other positions of l^i in such a manner that the first $m-1$ positions feature an empty school.

Note that, since there are $k_0^{i-1} = \sum_{x \in X} \mathbf{1}\{c_x^{i-1} = 0\}$ empty schools, there are

$$\frac{k_0^{i-1}!}{(k_0^{i-1} - (m-1))!}$$

ways to select $m-1$ empty schools for positions $l^i(1), \dots, l^i(m-1)$, and $(M-m)!$ ways of arranging the remaining $M-m$ schools in positions $l^i(m+1), \dots, l^i(M)$. It follows, for $m - 1 \leq k_0^{i-1}$, that

$$\begin{aligned} \sum_{l^i \in A^i} \mathbf{1}\{l^i(m) = x\} \mathbf{1}\left\{ \bigcap_{\eta=1}^{m-1} \left(c_{l^i(\eta)}^{i-1} = 0 \right) \right\} &= \frac{k_0^{i-1}!}{(k_0^{i-1} - (m-1))!}(M-m)! \\ &= k_0^{i-1}!(M - k_0^{i-1} - 1)! \binom{M-m}{M - k_0^{i-1} - 1} \end{aligned} \tag{C.27}$$

and therefore

$$\begin{aligned} &\sum_{l^i \in A^i} \mathbf{1}\left\{ l^i \left(\min_{\eta} \{\eta : c_{l^i(\eta)}^{i-1} > 0\} \right) = x \right\} \\ &= \mathbf{1}\{c_x^{i-1} > 0\} k_0^{i-1}!(M - k_0^{i-1} - 1)! \sum_{m=1}^{k_0^{i-1}+1} \binom{M-m}{M - k_0^{i-1} - 1} \\ &= \mathbf{1}\{c_x^{i-1} > 0\} k_0^{i-1}!(M - k_0^{i-1} - 1)! \sum_{m'=M-k_0^{i-1}-1}^{M-1} \binom{m'}{M - k_0^{i-1} - 1} \\ &= \mathbf{1}\{c_x^{i-1} > 0\} k_0^{i-1}!(M - k_0^{i-1} - 1)! \binom{M}{M - k_0^{i-1}} \\ &= \mathbf{1}\{c_x^{i-1} > 0\} \frac{M!}{M - k_0^{i-1}} \end{aligned} \tag{C.28}$$

where the first equality follows from plugging equation C.27 into equation C.26, the second equality from the substitution $m' = M - m$, and the third equality from the hockey-stick identity

$$\sum_{m'=r}^{n} \binom{m'}{r} = \binom{n+1}{r+1}.$$

Given that P^i is symmetric, student i's probability of matching with school x is thus, conditional on $\boldsymbol{c}^{i-1}$,

$$\tilde{\pi}^i(x|P^i, \boldsymbol{c}^{i-1}) = \frac{1}{M!} \sum_{l^i \in A^i} \mathbf{1}\left\{ l^i \left(\min_{\eta} \{\eta : c_{l^i(\eta)}^{i-1} > 0\} \right) = x \right\} = \frac{\mathbf{1}\{c_x^{i-1} > 0\}}{M - k_0^{i-1}}$$

where the first equality follows by $\sum_{\theta^i \in \Theta^i} P^i(l^i, \theta^i) = 1/(M!)$ for all $l^i \in A^i$ from the definition of equation C.25, and the second equality follows from equation C.28. Note also that since, by Corollary 1, each school is equally likely to have positive remaining capacity, student i has an equal probability of matching with any school $x \in X$.

C.14 Proof of Lemma 5

The proof is by induction over n.

Base Case: $n = 0$

Before any student has been assigned, each school $x \in X$ has capacity $c_x^0 = c$ with certainty, so

$$\mathbb{P}(\boldsymbol{c}^0) = \mathbf{1}\{\boldsymbol{c}^0 = (c, \dots, c)\}.$$

It follows that $k_r^0 = M \cdot \mathbf{1}\{r = c\}$ with certainty so equation 71 is satisfied for $n = 0$, because

$$\sum_{\boldsymbol{c}^0 \in \boldsymbol{\mathcal{C}}^0(\boldsymbol{k}^0)} \mathbb{P}(\boldsymbol{c}^0) = \mathbf{1}\left\{k_c = M, k_r = 0 \, \forall r \neq c\right\} = f^0\left(\boldsymbol{k}^0\right).$$

Induction Step: $n - 1 \to n$

Assuming that equation 71 holds for $n-1$, we show that it then holds for n. Define a random vector $\boldsymbol{i} \in \mathcal{I} \equiv \left\{\boldsymbol{i} \in \{0,1\}^M : \sum_{x \in X} i_x = 1\right\}$ with the interpretation that $i_x = 1$ if and only if student n is matched with school x. Its distribution conditional on $\boldsymbol{c}^{n-1}$ is, by Proposition 4 and equation C.19, given by

$$\mathbb{P}_{\boldsymbol{i}|\boldsymbol{c}^{n-1}}\left(\boldsymbol{i}|P^n, \boldsymbol{c}^{n-1}\right) = \mathbf{1}\{\boldsymbol{i} \in \mathcal{I}\} \sum_{x \in X} \mathbf{1}\{i_x = 1\} \sum_{l^n \in A^n} \mathbf{1}\left\{l^n\left(\min_{\eta}\{\eta : c_{l^n(\eta)}^{n-1} > 0\}\right) = x\right\} \sum_{\theta^n \in \Theta^n} P^n(l^n, \theta^n).$$

After student n is matched with a school, the remaining capacity of this school will decrease from $c_x^{n-1} = r$ to $c_x^n = r - 1$. It follows, given any $\boldsymbol{k}^n$ and $\boldsymbol{c}^{n-1} \in \boldsymbol{\mathcal{C}}^{n-1}(\boldsymbol{k}^{n-1})$ and where $\boldsymbol{e}_r$ is the unit vector defined in the Lemma, that

$$\begin{aligned}
&\sum_{\boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)} \mathbb{P}_{\boldsymbol{i}|\boldsymbol{c}^{n-1}}\left(\boldsymbol{c}^{n-1} - \boldsymbol{c}^n|P^n, \boldsymbol{c}^{n-1}\right) \\
&= \sum_{\boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)} \mathbf{1}\{\boldsymbol{c}^{n-1} - \boldsymbol{c}^n \in \mathcal{I}\} \sum_{x \in X} \mathbf{1}\{c_x^{n-1} - c_x^n = 1\} \frac{\mathbf{1}\{c_x^{n-1} > 0\}}{M - k_0^{n-1}} \\
&= \frac{1}{M - k_0^{n-1}} \sum_{\boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)} \mathbf{1}\{\boldsymbol{c}^{n-1} - \boldsymbol{c}^n \in \mathcal{I}\} \sum_{x \in X} \mathbf{1}\{c_x^{n-1} - c_x^n = 1\} \sum_{r=1}^{c} \mathbf{1}\{c_x^{n-1} = r\} \\
&= \frac{1}{M - k_0^{n-1}} \sum_{\boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)} \mathbf{1}\{\boldsymbol{c}^{n-1} - \boldsymbol{c}^n \in \mathcal{I}\} \sum_{r=1}^{c} \sum_{x \in X} \mathbf{1}\{c_x^{n-1} = r\} \mathbf{1}\{c_x^n = r - 1\} \\
&= \frac{1}{M - k_0^{n-1}} \sum_{r=1}^{c} \sum_{x \in X} \mathbf{1}\{c_x^{n-1} = r\} \sum_{\boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)} \mathbf{1}\{\boldsymbol{c}^{n-1} - \boldsymbol{c}^n \in \mathcal{I}\} \mathbf{1}\{c_x^n = r - 1\} \\
&= \frac{1}{M - k_0^{n-1}} \sum_{r=1}^{c} \sum_{x \in X} \mathbf{1}\{c_x^{n-1} = r\} \mathbf{1}\{\boldsymbol{k}^n = \boldsymbol{k}^{n-1} - \boldsymbol{e}_r + \boldsymbol{e}_{r-1}\} \\
&= \sum_{r=1}^{c} \frac{k_r^{n-1}}{M - k_0^{n-1}} \mathbf{1}\{\boldsymbol{k}^{n-1} = \boldsymbol{k}^n + \boldsymbol{e}_r - \boldsymbol{e}_{r-1}\} \qquad \text{(C.29)}
\end{aligned}$$

where the first equality follows by equation C.25 of Lemma C.2 from the symmetry of P^n, the third equality from $c_x^{n-1} - c_x^n = 1$, the fourth from rearranging, the fifth equality from the fact that given $\boldsymbol{c}^{n-1}$ there is exactly one vector $\boldsymbol{c}^n$ consistent with a decrease of the capacity of school x from r to $r-1$, and the last equality from the definition of k_r^{n-1}. Since

$$\mathbb{P}\left(\boldsymbol{c}^n | P^1, \ldots, P^n\right) = \sum_{\boldsymbol{c}^{n-1} \in \boldsymbol{\mathcal{C}}^{n-1}} \mathbb{P}_{\boldsymbol{i}|\boldsymbol{c}^{n-1}}\left(\boldsymbol{c}^{n-1} - \boldsymbol{c}^n | P^n, \boldsymbol{c}^{n-1}\right) \mathbb{P}\left(\boldsymbol{c}^{n-1} | P^1, \ldots, P^{n-1}\right) \tag{C.30}$$

this implies, given $\boldsymbol{k}^n$, that

$$\begin{aligned}
&\sum_{\boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)} \mathbb{P}(\boldsymbol{c}^n | P^1, \ldots, P^n) \\
&= \sum_{\boldsymbol{c}^{n-1} \in \boldsymbol{\mathcal{C}}^{n-1}} \mathbb{P}\left(\boldsymbol{c}^{n-1} | P^1, \ldots, P^{n-1}\right) \sum_{\boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)} \mathbb{P}_{\boldsymbol{i}|\boldsymbol{c}^{n-1}}\left(\boldsymbol{c}^{n-1} - \boldsymbol{c}^n | P^n, \boldsymbol{c}^{n-1}\right) \\
&= \sum_{\boldsymbol{k}^{n-1}} \sum_{\boldsymbol{c}^{n-1} \in \boldsymbol{\mathcal{C}}^{n-1}(\boldsymbol{k}^{n-1})} \mathbb{P}\left(\boldsymbol{c}^{n-1} | P^1, \ldots, P^{n-1}\right) \sum_{r=1}^{c} \frac{k_r^{n-1}}{M - k_0^{n-1}} \mathbf{1}\{\boldsymbol{k}^{n-1} = \boldsymbol{k}^n + \boldsymbol{e}_r - \boldsymbol{e}_{r-1}\} \\
&= \sum_{r=1}^{c} \sum_{\boldsymbol{k}^{n-1}} \frac{k_r^{n-1}}{M - k_0^{n-1}} \mathbf{1}\{\boldsymbol{k}^{n-1} = \boldsymbol{k}^n + \boldsymbol{e}_r - \boldsymbol{e}_{r-1}\} \sum_{\boldsymbol{c}^{n-1} \in \boldsymbol{\mathcal{C}}^{n-1}(\boldsymbol{k}^{n-1})} \mathbb{P}\left(\boldsymbol{c}^{n-1} | P^1, \ldots, P^{n-1}\right) \\
&= \sum_{r=1}^{c} \sum_{\boldsymbol{k}^{n-1}} \frac{k_r^{n-1}}{M - k_0^{n-1}} \mathbf{1}\{\boldsymbol{k}^{n-1} = \boldsymbol{k}^n + \boldsymbol{e}_r - \boldsymbol{e}_{r-1}\} f^{n-1}(\boldsymbol{k}^{n-1})
\end{aligned} \tag{C.31}$$

where the first equality follows from equation C.30, the second equality from equation C.29, the third from rearranging, and the fourth equality from the inductive assumption of equation 71 for $n-1$. It then follows that equation 71 also holds for n, as for any $\boldsymbol{k}^n$

$$\begin{aligned}
&\sum_{\boldsymbol{c}^n \in \boldsymbol{\mathcal{C}}^n(\boldsymbol{k}^n)} \mathbb{P}(\boldsymbol{c}^n | P^1, \ldots, P^n) \\
&= \sum_{r=1}^{c} \sum_{\boldsymbol{k}^{n-1}} \frac{k_r^n + 1}{M - (k_0^n - \mathbf{1}\{r = 1\})} \mathbf{1}\{\boldsymbol{k}^{n-1} = \boldsymbol{k}^n + \boldsymbol{e}_r - \boldsymbol{e}_{r-1}\} f^{n-1}(\boldsymbol{k}^n + \boldsymbol{e}_r - \boldsymbol{e}_{r-1}) \\
&= \sum_{r=1}^{c} \frac{k_r^n + 1}{M - (k_0^n - \mathbf{1}\{r = 1\})} f^{n-1}(\boldsymbol{k}^n + \boldsymbol{e}_r - \boldsymbol{e}_{r-1}) \sum_{\boldsymbol{k}^{n-1}} \mathbf{1}\{\boldsymbol{k}^{n-1} = \boldsymbol{k}^n + \boldsymbol{e}_r - \boldsymbol{e}_{r-1}\} \\
&= \sum_{r=1}^{c} \frac{k_r^n + 1}{M - k_0^n + \mathbf{1}\{r = 1\}} f^{n-1}(\boldsymbol{k}^n + \boldsymbol{e}_r - \boldsymbol{e}_{r-1}) \qquad (72) \\
&= f^n(\boldsymbol{k}^n)
\end{aligned}$$

where the first equality follows from equation C.31 noting that since the capacity of one school decreases from r to $r-1$ it is the case that $k_r^{n-1} = k_r^n + 1$ and $k_0^{n-1} = k_0^n - \mathbf{1}\{r = 1\}$. Note also that unless $k_{r-1}^n > 0$ and $k_r^n < M$, the requirement that $k_{r-1}^{n-1} = k_{r-1}^n - 1 \geq 0$ and $k_r^{n-1} = k_r^n + 1 \leq M$ is violated, in which case $f^{n-1}(\boldsymbol{k}^n + \boldsymbol{e}_r - \boldsymbol{e}_{r-1}) = 0$.

C.15 Proof of Theorem 1

We use Proposition 1, Lemma 1, and Lemma 2 to show that for each student $i \in I$, P^{i*} satisfying equation 73 is the unique symmetric best response to others' symmetric strategies $\boldsymbol{P}^{-i*}$. From equation 53 it follows by Lemma 1 for symmetric $\boldsymbol{P}^{-i*}$ that

$$\begin{aligned}
U^i(P^i, \boldsymbol{P}^{-i*}) &= \sum_{(l^i,\boldsymbol{l}^{-i})\in A^i\times\boldsymbol{A}^{-i}} \sum_{(\theta^i,\boldsymbol{\theta}^{-i})\in\Theta^i\times\boldsymbol{\Theta}^{-i}} P^i(l^i,\theta^i)\boldsymbol{P}^{-i*}(\boldsymbol{l}^{-i},\boldsymbol{\theta}^{-i})\theta^i_{\boldsymbol{x}^i(l^i,\boldsymbol{l}^{-i})} \\
&= \sum_{l^i\in A^i}\sum_{\theta^i\in\Theta^i} P^i(l^i,\theta^i) \sum_{\boldsymbol{l}^{-i}\in\boldsymbol{A}^{-i}}\sum_{\boldsymbol{\theta}^{-i}\in\boldsymbol{\Theta}^{-i}} \boldsymbol{P}^{-i*}(\boldsymbol{l}^{-i},\boldsymbol{\theta}^{-i})\theta^i_{\boldsymbol{x}^i(l^i,\boldsymbol{l}^{-i})} \\
&= \sum_{l^i\in A^i}\sum_{\theta^i\in\Theta^i} P^i(l^i,\theta^i) \sum_{x\in X}\theta^i_x \sum_{\boldsymbol{l}^{-i}\in\boldsymbol{A}^{-i}}\sum_{\boldsymbol{\theta}^{-i}\in\boldsymbol{\Theta}^{-i}} \boldsymbol{P}^{-i*}(\boldsymbol{l}^{-i},\boldsymbol{\theta}^{-i})\mathbf{1}\left\{\boldsymbol{x}^i(l^i,\boldsymbol{l}^{-i}) = x\right\} \\
&= \sum_{l^i\in A^i}\sum_{\theta^i\in\Theta^i} P^i(l^i,\theta^i) \sum_{x\in X}\theta^i_x \pi^i_{\boldsymbol{x}}(x|l^i,\boldsymbol{P}^{-i*}) \\
&= \sum_{l^i\in A^i}\sum_{\theta^i\in\Theta^i} P^i(l^i,\theta^i) \sum_{m=1}^{M} \pi^i_{\boldsymbol{x}}(l^i(m)|l^i,\boldsymbol{P}^{-i*})\theta^i_{l^i(m)} \\
&= \sum_{l^i\in A^i}\sum_{\theta^i\in\Theta^i} P^i(l^i,\theta^i) \sum_{m=1}^{M} h^i(m)\theta^i_{l^i(m)}
\end{aligned}$$

where the fourth equality follows from equation 58 and the last equality from Lemma 1. The problem of student i according to equation 54 is therefore, given others' symmetric $\boldsymbol{P}^{-i*}$,

$$\max_{P^i\in\mathcal{P}(\mu^i,A^i)} U^i(P^i,\boldsymbol{P}^{-i*}) - K(\kappa^i,P^i) = \max_{P^i\in\mathcal{P}(\mu^i,A^i)} \sum_{l^i\in A^i}\sum_{\theta^i\in\Theta^i} P^i(l^i,\theta^i)u^i(l^i,\theta^i) - K(\kappa^i,P^i)$$

with

$$u^i(l^i,\theta^i) = \sum_{m=1}^{M} h^i(m)\theta^i_{l^i(m)}. \tag{61}$$

By Lemma 2 this problem is symmetric in the sense of Definition 1, so by Proposition 1 its unique symmetric solution is indeed given by

$$P^{i*}\left(l^i|\theta^i\right) \equiv \frac{P^{i*}\left(l^i,\theta^i\right)}{\mu^i(\theta^i)} = \frac{z^i(l^i,\theta^i)}{\sum_{l'\in A^i} z^i(l',\theta^i)} \qquad \forall l^i\in A^i, \theta^i\in\Theta^i \tag{73}$$

with

$$z^i\left(l^i,\theta^i\right) = \exp\left(\frac{1}{\kappa^i}\sum_{m=1}^{M} h^i(m)\theta^i_{l^i(m)}\right) \qquad \forall l^i\in A^i, \theta^i\in\Theta^i \tag{74}$$

which is indeed symmetric in the sense of Definition 2, as

$$\sum_{\theta^i\in\Theta^i} P^{i*}(l^i|\theta^i)\mu^i(\theta^i) = 1/|A^i| \qquad \forall l^i\in A^i.$$

This concludes the equilibrium proof. Under DA we further know by Lemma C.1 that

$$h^i_{DA}(m) = \sum_{\boldsymbol{c}^{i-1}\in\boldsymbol{\mathcal{C}}^{i-1}} \mathbf{1}\left\{m = \min_{\eta}\{\eta : c^{i-1}_\eta > 0\}\right\}\mathbb{P}(\boldsymbol{c}^{i-1}|P^{1*},\ldots,P^{(i-1)*}). \tag{C.17}$$

Recalling the definition

$$\mathbb{P}(\boldsymbol{k}^{i-1}|P^{1*},\ldots,P^{(i-1)*}) \equiv \sum_{\boldsymbol{c}^{i-1}\in\boldsymbol{\mathcal{C}}^{i-1}(\boldsymbol{k}^{i-1})} \mathbb{P}(\boldsymbol{c}^{i-1}|P^{1*},\ldots,P^{(i-1)*}) \tag{C.22}$$

it therefore follows from Lemma 3, Lemma 4, and Lemma 5 that

$$\begin{aligned}
h^i_{DA}(m) &= \sum_{\boldsymbol{k}^{i-1}} \mathbb{P}(\boldsymbol{k}^{i-1}|P^{1*},\ldots,P^{(i-1)*}) \sum_{\boldsymbol{c}^{i-1}\in\boldsymbol{\mathcal{C}}^{i-1}(\boldsymbol{k}^{i-1})} \frac{\mathbb{P}(\boldsymbol{c}^{i-1}|P^{1*},\ldots,P^{(i-1)*})}{\mathbb{P}(\boldsymbol{k}^{i-1}|P^{1*},\ldots,P^{(i-1)*})} \mathbf{1}\left\{m = \min_{\eta}\{\eta : c^{i-1}_{l^I(\eta)} > 0\}\right\} \\
&= \sum_{\boldsymbol{k}^{i-1}} \mathbb{P}(\boldsymbol{k}^{i-1}|P^{1*},\ldots,P^{(i-1)*}) \frac{1}{|\boldsymbol{\mathcal{C}}^{i-1}(\boldsymbol{k}^{i-1})|} \sum_{\boldsymbol{c}^{i-1}\in\boldsymbol{\mathcal{C}}^{i-1}(\boldsymbol{k}^{i-1})} \mathbf{1}\left\{m = \min_{\eta}\{\eta : c^{i-1}_{l^I(\eta)} > 0\}\right\} \\
&= \sum_{\boldsymbol{k}^{i-1}} \mathbb{P}(\boldsymbol{k}^{i-1}|P^{1*},\ldots,P^{(i-1)*}) \binom{M-m}{k_0^{i-1}+1-m} \Big/ \binom{M}{k_0^{i-1}} \\
&= \sum_{\boldsymbol{k}^{i-1}} \binom{M-m}{k_0^{i-1}+1-m} \Big/ \binom{M}{k_0^{i-1}} \sum_{\boldsymbol{c}^{i-1}\in\boldsymbol{\mathcal{C}}^{i-1}(\boldsymbol{k}^{i-1})} \mathbb{P}(\boldsymbol{c}^{i-1}|P^{1*},\ldots,P^{(i-1)*}) \\
&= \sum_{\boldsymbol{k}^{i-1}} \binom{M-m}{k_0^{i-1}+1-m} \Big/ \binom{M}{k_0^{i-1}} f^{i-1}(\boldsymbol{k}^{i-1})
\end{aligned} \tag{76}$$

where the first equality follows from equation C.17, the second equality by equation C.23 from Corollary 1 of Lemma 3, the third equality by equation 65 from Lemma 4, the fourth equality from the definition of equation C.22, and the last equality follows by equation 71 from Lemma 5.

C.16 Proof of Proposition 5

From equation 78 it follows that

$$\begin{aligned}
U^{i*}(\kappa^i,\mu^i) &= \sum_{m=1}^{M} h^i(m) U^{i*}(\kappa,\mu^i|m) \\
&= \sum_{\theta^i\in\Theta^i} \mu^i(\theta^i) \sum_{l^i\in A^i} \frac{z^i(l^i,\theta^i)}{\sum_{l'\in A^i} z^i(l',\theta^i)} \sum_{m=1}^{M} h^i(m)\theta^i_{l^i(m)} \\
&= \sum_{\theta^i\in\Theta^i} \mu^i(\theta^i) \sum_{l^i\in A^i} \frac{z^i(l^i,\theta^i)}{\sum_{l'\in A^i} z^i(l',\theta^i)} u^i(l^i,\theta^i) \\
&= \kappa^i \sum_{\theta^i\in\Theta^i} \mu^i(\theta^i) \frac{\sum_{l^i\in A^i} z^i(l^i,\theta^i)\ln z^i(l^i,\theta^i)}{\sum_{l'\in A^i} z^i(l',\theta^i)}
\end{aligned} \tag{79}$$

where the last equality follows from the fact that $u^i(l^i,\theta^i) = \kappa^i \ln z^i(l^i,\theta^i)$ by definition. Note that

$$\begin{aligned}
\underline{U}(\mu^i) &= \frac{1}{M}\sum_{m=1}^{M} U^{i*}(\kappa^i,\mu^i|m) \\
&= \sum_{\theta^i\in\Theta^i} \mu^i(\theta^i) \frac{1}{M} \sum_{m=1}^{M} \sum_{x\in X} \frac{\sum_{l^i\in A^i} z^i(l^i,\theta^i)\theta^i_x \mathbf{1}\{x = l^i(m)\}}{\sum_{l'\in A^i} z^i(l',\theta^i)} \\
&= \sum_{\theta^i\in\Theta^i} \mu^i(\theta^i) \frac{1}{M} \sum_{x\in X} \theta^i_x \frac{\sum_{l^i\in A^i} z^i(l^i,\theta^i) \sum_{m=1}^{M} \mathbf{1}\{x = l^i(m)\}}{\sum_{l'\in A^i} z^i(l',\theta^i)} \\
&= \sum_{\theta^i\in\Theta^i} \mu^i(\theta^i) \frac{1}{M} \sum_{x\in X} \theta^i_x
\end{aligned} \tag{80}$$

so that

$$
\begin{aligned}
U^{i*}(\kappa^i,\mu^i) - \underline{U}(\mu^i) &= \sum_{\theta^i\in\Theta^i}\mu^i(\theta^i)\left[\frac{\sum_{l^i\in A^i} z^i(l^i,\theta^i)\kappa^i \ln z^i(l^i,\theta^i)}{\sum_{l'\in A^i} z^i(l',\theta^i)} - \frac{\sum_{k=1}^M \theta^i_k}{M}\right] \\
&= \sum_{\theta^i\in\Theta^i}\mu^i(\theta^i)\sum_{l^i\in A^i}\frac{z^i(l^i,\theta^i)}{\sum_{l'\in A^i} z^i(l',\theta^i)}\left[u^i(l^i,\theta^i) - \sum_{k=1}^M \frac{1}{M}\theta^i_k\right] \\
&= \sum_{\theta^i\in\Theta^i}\mu^i(\theta^i)\sum_{l^i\in A^i}\left[\sum_{m=1}^M h^i(m)\theta^i_{l^i(m)} - \sum_{m=1}^M \frac{1}{M}\theta^i_{l^i(m)}\right]\frac{z^i(l^i,\theta^i)}{\sum_{l'\in A^i} z^i(l',\theta^i)} \\
&= \sum_{m=1}^M\left[h^i(m) - \frac{1}{M}\right]\sum_{\theta^i\in\Theta^i}\mu^i(\theta^i)\sum_{l^i\in A^i}\frac{z^i(l^i,\theta^i)}{\sum_{l'\in A^i} z^i(l',\theta^i)}\theta^i_{l^i(m)} \\
&= \sum_{m=1}^M\left[h^i(m) - \frac{1}{M}\right]U^{i*}\left(\kappa^i,\mu^i|m\right). \qquad (81)
\end{aligned}
$$

C.17 Proof of Proposition 6

The information costs in equilibrium are, writing $P^{i*}(l^i) = \sum_{\theta^i\in\Theta^i} P^{i*}(l^i,\theta^i)$, given by

$$
\begin{aligned}
K^{i*}(\kappa^i,\mu^i) &= K(\kappa^i, P^{i*}) \\
&= \kappa^i \sum_{l^i\in A^i} P^{i*}(l^i)\left[H(\mu^i) - H(\gamma^{l^i})\right] \\
&= \kappa^i \sum_{l^i\in A^i} P^{i*}(l^i)\sum_{\theta^i\in\Theta^i}\left[-\mu^i(\theta^i)\ln\mu^i(\theta^i) + \frac{P^{i*}(l^i|\theta^i)\mu^i(\theta^i)}{P^{i*}(l^i)}\ln\left(\frac{P^{i*}(l^i|\theta^i)\mu^i(\theta^i)}{P^{i*}(l^i)}\right)\right] \\
&= \kappa^i \sum_{\theta^i\in\Theta^i}\mu^i(\theta^i)\left[\sum_{l^i\in A^i} P^{i*}(l^i|\theta^i)\left(\ln\left(\frac{P^{i*}(l^i|\theta^i)}{P^{i*}(l^i)}\right) + \ln\left(\mu^i(\theta^i)\right)\right) - \ln\mu^i(\theta^i)\right] \\
&= \kappa^i \sum_{\theta^i\in\Theta^i}\mu^i(\theta^i)\left[\sum_{l^i\in A^i} P^{i*}(l^i|\theta^i)\ln\left(\frac{P^{i*}(l^i|\theta^i)}{P^{i*}(l^i)}\right) + \left(\sum_{l^i\in A^i} P^{i*}(l^i|\theta^i) - 1\right)\ln\mu^i(\theta^i)\right] \\
&= \kappa^i \sum_{\theta^i\in\Theta^i}\mu^i(\theta^i)\sum_{l^i\in A^i} P^{i*}(l^i|\theta^i)\ln\left(\frac{P^{i*}(l^i|\theta^i)}{P^{i*}(l^i)}\right) \\
&= \kappa^i \sum_{\theta^i\in\Theta^i}\mu^i(\theta^i)\sum_{l^i\in A^i}\frac{z^i(l^i,\theta^i)}{\sum_{l'\in A^i} z^i(l',\theta^i)}\ln\left(\frac{M!z^i(l^i,\theta^i)}{\sum_{l'\in A^i} z^i(l',\theta^i)}\right) \qquad (82)
\end{aligned}
$$

where the second equality follows from equation 7, the sixth equality corresponds to equation 2.36 of Cover and Thomas (2006), and the last equality follows from equation 73 which also implies $P^{i*}(l^i) = 1/M!$.

The net utility is then given by

$$
\begin{aligned}
N^{i*}(\kappa^i, \mu^i) &= U^{i*}(\kappa^i, \mu^i) - K^{i*}(\kappa^i, \mu^i) \\
&= \kappa^i \sum_{\theta^i \in \Theta^i} \mu^i(\theta^i) \left[\frac{\sum_{l^i \in A^i} z^i(l^i, \theta^i) \ln z^i(l^i, \theta^i)}{\sum_{l' \in A^i} z^i(l', \theta^i)} - \sum_{l^i \in A^i} \frac{z^i(l^i, \theta^i)}{\sum_{l' \in A^i} z^i(l', \theta^i)} \ln \left(\frac{M! z^i(l^i, \theta^i)}{\sum_{l' \in A^i} z^i(l', \theta^i)} \right) \right] \\
&= \kappa^i \sum_{\theta^i \in \Theta^i} \mu^i(\theta^i) \sum_{l^i \in A^i} \frac{z^i(l^i, \theta^i)}{\sum_{l' \in A^i} z^i(l', \theta^i)} \left[\ln z^i(l^i, \theta^i) - \ln z^i(l^i, \theta^i) - \ln \left(\frac{M!}{\sum_{l' \in A^i} z^i(l', \theta^i)} \right) \right] \\
&= \kappa^i \sum_{\theta^i \in \Theta^i} \mu^i(\theta^i) \sum_{l^i \in A^i} \frac{z^i(l^i, \theta^i)}{\sum_{l' \in A^i} z^i(l', \theta^i)} \ln \left(\frac{1}{M!} \sum_{l' \in A^i} z^i(l', \theta^i) \right) \\
&= \kappa^i \sum_{\theta^i \in \Theta^i} \mu^i(\theta^i) \ln \left(\frac{1}{M!} \sum_{l' \in A^i} z^i(l', \theta^i) \right) \qquad (83)
\end{aligned}
$$

where the second equality follows from equations 79 and 82. Note that without information acquisition, the net utility is given by

$$
\underline{N}(\mu^i) = \underline{U}(\mu^i) = \sum_{\theta^i \in \Theta^i} \mu^i(\theta^i) \frac{1}{M} \sum_{x \in X} \theta^i_x.
$$

The net utility surplus is therefore

$$
N^{i*}(\kappa^i, \mu^i) - \underline{N}(\mu^i) = \sum_{\theta^i \in \Theta^i} \mu^i(\theta^i) \left[\kappa^i \ln \left(\frac{1}{M!} \sum_{l' \in A^i} z^i(l', \theta^i) \right) - \frac{1}{M} \sum_{x \in X} \theta^i_x \right]. \qquad (84)
$$