Can Ego Utility and Expectations-based Loss Aversion Explain Pessimism and Reverse Endowment in Strategy Proof Mechanisms?

Research Proposal

Hunter Guru, Pablo Noyola, Sikata Sengupta, Anthony Weng

Abstract

Market designers often celebrate strategy-proof models, as they entail truthfulness as a dominant strategy; despite this phenomenon, when faced with allocation problems, agents often misreport their preferences. Understanding these behaviors is crucial, as it can help mechanism design minimize their harm. In both lab and field settings, agents have been found to overvalue particular choices (e.g. ranking less preferred schools higher, as they are ”less selective”), omit some choices (e.g. leaving off their highest ranked choices, for they are ”too selective”), and undervalue other choices (e.g. applying to schools less desirable than an admission in hand). In isolation, these actions seem to be psychologically explainable by loosely related behavioral theories (that is loss aversion, over-pessimism, and reverse-endowment, respectively), however, we show through a modification of expectations-based loss aversion (Kőszegi and Rabin (2006), Dreyfuss, Heffetz, and Rabin (2019)), these behaviors can be synthesized within one model. We integrate a multistage experiment with the Ghana school assignment system with each stage measuring a parameter within our model, and compare the fit of the observed data to our modified model, the standard economic model, as well as the model introduced in Dreyfuss, Heffetz, and Rabin (2019). Given our observations, we conclude by offering interpretations of our expected results, including calibrated parameters and predicted behaviors. [1]

Preliminaries [2]

We note that our paper assumes a basic understanding of game theory for our discussion of strategy-proofness and dominant strategies. A background in market design is helpful, although unnecessary: one need only have read Gale and Shapley (1962) as background on deferred-acceptance. Lastly, we reference certain behavioral economics concepts: reading Kahneman and Tversky (1979) for background on Prospect Theory is recommended as well.

1 Introduction

The deferred acceptance algorithm (Gale and Shapley (1962)), is a cornerstone of market design, celebrated as a theoretically desirable implementation of the matching market. The algorithm has since been applied to a diverse range of matching domains; Roth (1984) discusses it in terms of the medical school residency match algorithm, while Abdulkadiroğlu and Sönmez (2003) extend it to the school choice problem. In addition to stability, a key property of the algorithm, first raised in Dubins and Freedman (1981), is strategy-proofness on the side of the "proposer". That is, the "proposing" side of the algorithm—men in the marriage market, students applying to schools or universities, or medical students seeking a residency—cannot improve their results by deviating from stating their true preferences. Despite these useful theoretical properties, a growing literature notes that in practice people categorically misrepresent their preferences in deferred-acceptance settings. Common deviations from true preferences include "overvaluation" of less desirable, though less risky, options; "pessimistic" truncation of preferences, wherein certain desirable options are not considered; and "undervaluation" of the status quo [3].

Psychologists and economists have applied a patchwork of behavioral theories to attempt to explain such behavior, which individually explain isolated deviations but fail to provide an overarching model that can explain such behavior. We offer a singular model that explains these behaviors, building on the work of Köszegi (2006) and Dreyfuss, Heffetz, and Rabin (2019) by modifying their expectations-based loss aversion model. In the end, we synthesize various behavioral theories, including loss aversion, over-pessimism, and the reverse-endowment effect, into our unifying framework [4].

We offer further motivation of the problem and survey the rich vein of existing research, along both market design and behavioral economics dimensions. Then, we continue by providing the theoretical foundations for our model. We propose a multistage experiment to test and calibrate our framework. Our multistage experiment compares the performance of our modified model with standard economic theory and baseline expectations-based loss aversion from Dreyfuss, Heffetz, and Rabin (2019) to predict the submitted preferences of participants within the Ghana secondary school assignment mechanism. We conclude with a discussion of our anticipated findings.

2 Prior Work and Motivation

Strategy-proofness Isn't Enough

Despite the strategy-proofness of the Gale-Shapley algorithm, there is a significant body of literature and experimentation documenting the misrepresentation of preferences in deferred-acceptance situations. Chen and Sönmez (2006) study the school choice problem—–wherein students' preferences are used to match them to schools—–and find that, though deferred-acceptance performs better than alternative mechanisms, 36% of participants misrepresent preferences. Calsamiglia, Haeringer, and Klijn (2010) build on this work, noting that in real-world contexts, students are often constrained by a cap on the number of schools they can apply to, exacerbating misrepresentation as the risk of remaining unassigned grows. Examining the medical school residency match, Rees-Jones and Skowronek (2018) find that 23% of medical students in their simulated residency match deviate from the truth, with factors such as student test scores, cognitive ability, and overconfidence correlating with misrepresentation [5].

Klijn, Pais, and Vorsatz (2010) also find that risk aversion leads students to "play it safer" than their true preferences, over-ranking lower preferred schools with the belief that it increases admission probability – an effect we will refer to as *overvaluation*. Hassidim, Romm, and R. I. Shorrer (2016) found that people who view themselves as less attractive applicants, may end up ranking non-funded positions over funded ones, even when there are no stakes attached to the financial designation.

From the studies of Calsamiglia, Haeringer, and Klijn (2010), Klijn, Pais, and Vorsatz (2010), we also observe that students omit schools which seem "impossible" from their lists altogether – an effect called *pessimism*. Pessimism is particularly relevant when there is no marginal cost for additional application; some students may forgo submitting a common application to schools they express interest in, even with a fee waiver (Edelman (2014)) [6].

Some agents participate in mechanisms with no intent to ultimately take the offer (e.g. a job-applicant undervalues their current job, a student undervalues their "safety school", etc.); these behaviors categorize *undervaluation* in economic mechanisms [7]. In comparison to the aforementioned strategic deviations (i.e. pessimism, overvaluation), undervaluation is the least commonly observed, in part because it implies a risk-seeking attitude over gains: if undervaluation is the pursuit of a prospect with less expected utility over an alternate prospect granting positive utility, this is definitionally risk-seeking behavior. The popular prospect theory posits risk-averse attitudes over gains, a tenet which has become widely accepted as both descriptive and predictive of agent actions.

Despite conventional wisdom, examples of undervaluation exist. Blavatskyy and Pogrebna (2008) analyze categorical undervaluation in the popular television gameshow Deal or No Deal. They consider two groups of contestants: in one group, contestants possess a prize box with a 20% chance of containing a large prize; in the other, the prize box has a 80% chance of the large prize. At various points in the game, contestants in both groups are given qualitatively similar price offers to sell the contents of their boxes (i.e. a certainty offer). Surprisingly, contestants in both groups accept/reject the offers at similar rates, suggesting agents possess identical risk attitudes over sufficiently large gains of low or high probability. [1]

Prospect theory adherents may explain this phenomena via diminishing sensitivity. But, this explanation may not be sufficient–as the game progresses and certainty offers are rejected, the certainty offers often converge to (and occasionally) exceed the expected value of the prize box. If diminishing sensitivity holds and agents indeed possess identical risk attitudes over sufficiently large gains of low or high probability, their utility should be more responsive to an increase in the certainty offer as opposed to an increase in the expected value of the box, leading to increased offer acceptance rates as the game progresses. However, Blavatskyy and Pogrebna find no increased propensity to accept the certainty offer, suggesting that as the game progresses, contestants are in fact undervaluing the certainty offer.

In the labor market, we observe individuals applying to job opportunities they have no intention on accepting. Liu (2010) finds that many college applicants refuse job offers, for reasons likely unattributed to the job application process whatsoever. These behaviors indicate a reverse-endowment phenomena towards "safety options" adjacent to the mechanism—and away from the unwanted, yet "endowed," offers on the table [2].

More specifically, the reverse-endowment effect refers to a reversal of the usual endowment effect (i.e., a decreased willingness to retain the endowment as compared to the higher willingness to pay to obtain said endowment) should the endowment be undesirable (Brenner et al. (2007)). When students are admitted to their safety schol, they *lose* utility, and are more willing to explore something new. These biases lead them to give up the endowment—thereby rejecting the offers—even though they might have no reason to do so. Within the existing market design literature, little work has been done to verify the existence of reverse-endowment in allocation mechanisms. As an auxiliary objective of our analysis, we seek to confirm and isolate the existence of this phenomenon by accounting for potential changes in valuation resulting from temporal discounting. [8]

Partially in response to these deviations, Li (2017) constructs a bounded rationality property of mechanisms, which he calls obviously strategy-proofness (OSP). He also shows that agents play the dominant strategy (i.e. truthfulness) at significantly higher rates under OSP mechanisms as compared to SP mechanisms. If this is the case, one may rather spend time constructing a matching mechanism that satisfies OSP. However, Ashlagi and Gonczarowski (2018) show that "no mechanism that implements a stable matching is obviously strategy-proof for any side of the market" in two-sided matching markets. As such, the need for an alternate unifying theory and recommendation persists.

[1] Note that Deal or No Deal is theoretically a single-player strategy-proof game, where the agent is incentivized to indicate when the prize passes their certainty equivalent.

[2] A caveat: the labor market is not necessarily a strategy-proof setting; our analysis considers undervaluation in a strategy-proof context.

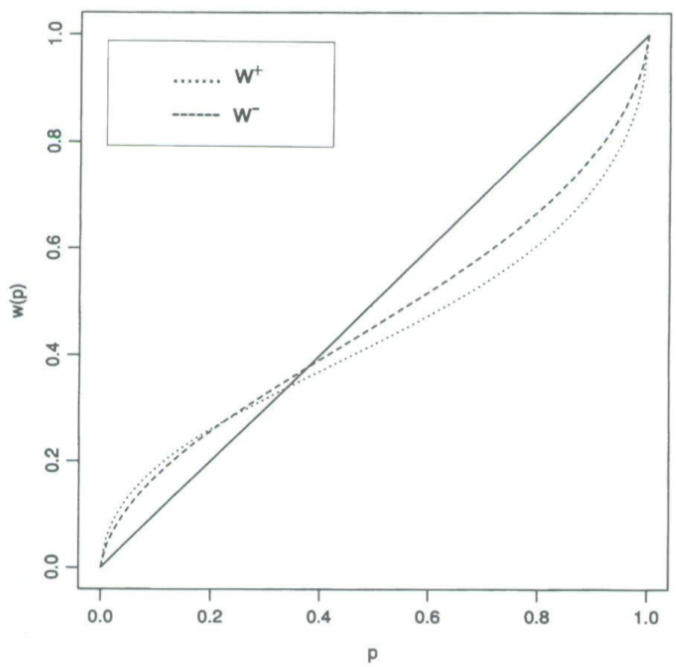

Figure 1: Prospect Theory Probability Weight Function from Tversky and Kahneman (1992).

Alternate Prevailing Behavioral Explanations

Alternative construction of an agent's utility function may individually sufficiently explain pessimism, overvaluation, or undervaluation in strategy-proof mechanisms. We review other utility models motivated by commonly accepted psychological theories and models.

Prospect Theory and Subjective Probabilities

The original prospect theory model in Kahneman and Tversky (1979) integrates loss-aversion, which influences overvaluation behavior, and introduces a probability weight function (PWF) that reflects how people understand uncertainty by relating true probabilities to an associated "decision weight", or the perceived probability people use to guide decisions (i.e., a function $\pi(p)$). The PWF recognized an overweighing of low-probability events and underweighting of high-probability events. Over time, the pair refined their estimate of the PWF, within Tversky and Kahneman (1992) estimating a now-characteristic "s-curve" by fitting experimental data using

$$\pi(p;\gamma) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{1/\gamma}}, \tag{1}$$

where they note this functional form maintains several desirable features, including: needing only one parameter, encompassing weighting functions with both concave and convex regions, and providing a reasonably good approximation to both individual and aggregate-level data. These weighting functions have been further employed and refined by numerous others; in their study of violations of the betweenness axiom and adjustments to expected utility theory, Camerer and Ho (1994) fit data from nine studies to both original and cumulative prospect theory to the functional form in Figure 1 with appreciable accuracy. Wu and Gonzalez (1996) affirm the S-shaped weighting function by proposing two preference conditions which: (1) are necessary and sufficient for the concavity and convexity of the weighting function; and (2) can be tested independently of the form of the value function. Their tests of these preference conditions validate the S-shaped weighting function, and they observe substantial fit when modeling experimental data using the Figure 1 functional form.

We anticipate that the probability weighting function will be crucial in modeling pessimism, as an agent who omits a zero-probability choice from their preference list is technically playing a weakly dominant strategy; in other words, if they perceive some probabilities $p > 0$ as $\hat{p} = 0$, then this behavior would be perfectly rational in conventional game theoretic modeling. Unfortunately, these studies and numerous others all consider weighting functions in the context of exogenously given prospects. In fact, conventional behavioral economic theory, predicts the opposite behavior to occur: that is, lower probabilities are actually over-weighted [9].

There has been little to no systematic consideration of how agents initiating or pursuing prospects may signal shifts in their weighting functions; i.e., analogous to the winner's curse phenomenon whereby a high bid indicates a high personal valuation of an item (and commonly results in overpaying), initiation of a prospect with a low-probability of success may, in part, indicate overconfidence about one's belief of success (especially if the prospect carries significant costs of initiation). Such phenomena could substantially alter the widely-accepted S-shape of the weighting function, thus demanding further study considering many important prospects are self-initiated (e.g. school/job applications).

Ego Utility

Utility function constructions which include considerations beyond pecuniary payoffs have been well-studied; i.e. the Etzioni (1986) critique of monoutility introduces "two irreducible sources of value or 'utility', pleasure and morality." Pleasure utility is commonly associated with pecuniary outcomes, but can more broadly be conceived to capture any hedonistic payoff.

Köszegi (2006) analyzes one such conception—"ego utility," or the utility an agent derives from positive views about the self. In his construction of the ego utility framework, an agent plays a multi-period game wherein she chooses to participate in one of two financial activities, one of which is better suited for skillful agents. Between periods, she may choose to collect information regarding her ability to perform well in the more ambitious of the two activities and can stop her information collection at any time. Generally, he finds that "agents become overconfident in their ability to perform the more ambitious task"; however, his findings with regard to information collection are more salient. Köszegi argues that agents are subject to two motives: a self-image protection motive and a self-image enhancement motive. The former arises when an agent is satisfied with her present beliefs, and she "distort[s] her choices to avoid receiving information about herself"; the latter occurs when she is dissatisfied with her current perception of herself and "go[es] out of her way to try to improve her beliefs".

These motives coordinate well with many of the behaviors we seek to model; for instance, agents may undervalue choices in favor of less desirable ones owing to their perceived higher probability of success (and the associated higher probability of deriving ego utility from satisfying a self-image enhancement motive), even if these agents have no intent of accepting the positive outcome of these less desirable choices. Similarly, the self-image protection motive may explain why agents sometimes omit highly desirable choices despite no significant costs of selection or overvalue less desirable choices, especially as their confidence rises: when an agent believes herself to be good enough to pursue the ambitious task—but is not quite sure—"she might avoid the ambitious option for fear of learning bad news about her ability" (Köszegi (2006)).

The conception of ego utility thus initially offers a plausible alternative to expectations-based loss aversion in our current context. However, empirically validating the existence and influence of ego utility may be difficult; doing so would require subjects be able to accurately assess how "good" they felt after given decisions and later outcomes, at which point, one could argue that they would choose to submit their true preferences, but conceal them from others.

Expectations-based Loss Aversion

To theoretically explain this seemingly irrational behavior, Dreyfuss, Heffetz, and Rabin (2019) create a model using the expectations-based reference dependent framework (EBRD), introduced in Kőszegi and Rabin (2006). The EBRD model extends the standard prospect theory model to a stochastic domain, and incorporates losses relevant to a fixed or stochastic reference point. Consider an individual's utility for the consumption-level $c \in \mathbb{R}$, given a reference level of consumption $r \in \mathbb{R}$, which we denote by $u(c \mid r)$. When c is drawn according to a probability distribution F, Kőszegi and Rabin (2006) define the utility to be

$$U(F \mid r) = \int u(c) + u(c \mid r) dF(c).$$

Furthermore, they achieve flexibility in the model by allowing the reference point r to also be stochastic; when $r \sim G$, then

$$U(F \mid G) = \int \int u(c) + u(c \mid r) dG(r) dF(c). \tag{2}$$

To gain intuition for $U(F \mid G)$, one can assume that U represents the utility of consuming a lottery F, when an agent has a reference point or expectations of that lottery corresponding to G. Via $u(c \mid r)$, the EBRD model integrates the loss-aversion property from Kahneman and Tversky (1979) to a stochastic setting; when F is a $0-100, 50/50$ lottery, $U(F \mid F)$ corresponds to the probabilistic mixture of sensation from all the possible current and future selves (e.g. distress accompanying loss when expecting a win, along with the surprise of winning while expecting loss, etc.).

In an allocation mechanism, $U(F_r \mid F_r)$ would denote the expected utility of submitting ranked-order-list (ROL) of preferences, r, where F_r corresponds to the student's belief of their match probabilities induced by their ROL (Meisner and Wangenheim (2021)). Dreyfuss, Heffetz, and Rabin (2019) then analyze an experiment running random serial dictatorship (RSD), a one-sided strategy proof mechanism, and find that high degrees of loss aversion can produce misrepresentation behavior.

Meisner and Wangenheim (2021) conduct a similar study on the same data, but their conclusion is that non-truthful preferences are optimal if and only if they satisfy top-choice monotonicity – a property which requires a preference ordering of $k \succ k-1 \succ k-2 \succ \cdots \succ 1 \succ k+1 \succ k+2 \cdots \succ s$, for any k, where the $1 \succ 2 \succ \ldots \succ s$ denotes the "true" (1-most-preferred) ranking.

Both Dreyfuss, Heffetz, and Rabin (2019) and Meisner and Wangenheim (2021), fail to consider agent pessimism and safe allocations outside the market. Neither incorporate a PWF, and results in Dreyfuss, Heffetz, and Rabin (2019) indicated biases contrary to the standard models in Figure 1; in fact, they find that individuals seem to underweigh small, but non-zero probabilities of getting their top preference.

It can be shown that when considering sufficiently risk averse individuals, the model in Dreyfuss, Heffetz, and Rabin (2019) leads to a contradiction. Given that many individuals have a safe allocation already (e.g. an existing job, admission to a local state school, etc.), if individuals were risk-averse, why would they enter the mechanism in the first place?

3 Theoretical Preliminaries

We begin by extending the behavioral models of expectations-based loss aversion and ego utility into a single model that reconciles pessimism, overvaluation, and undervaluation. Our model is in the context of the student-school deferred acceptance (DA) mechanism (student-proposing) in accordance with Dreyfuss, Heffetz, and Rabin (2019) and Meisner and Wangenheim (2021), but we note that the model should generalize to other DA contexts.

In this setting, a student is tasked with submitting a rank ordered list (ROL), $\succ$, over a set of schools, $\mathcal{S} = \{s_1, s_2, \ldots, s_n\}$. It is also given that the student is guaranteed admission to some school, $s_0 \in \mathcal{S}$ (often due to districting/neighborhood allocation). That is, the student is to submit a single $\succ$ from $\boldsymbol{\succ}$ [10], where

$$\boldsymbol{\succ} := \bigcup_{X \in \mathcal{P}(\mathcal{S})} \sigma(X),$$

$\sigma(X)$ is the set of all permutations of X, and $\mathcal{P}(X)$ is the power-set of X (this does mean the student is free to omit schools as they please).

We follow the convention of modeling the submission of $\succ$ as a lottery, as done by Dreyfuss, Heffetz, and Rabin (2019). For every school, $s_i \in \mathcal{S}$, each student has an intrinsic attractiveness x_i and a random term ε_i. x_i represents internal characteristics of the applicant observed to both the student and school (e.g. test scores, accomplishments, etc.) and ε_i corresponds to applicant characteristics that are only known by the school (e.g. idiosyncratic "fit" with the school, the other candidates' strength, etc.). The school has a cutoff, y_i, so the student is admitted to s_i if and only if $x_i + \varepsilon_i \geq y_i$. Hence, as noted in Dreyfuss, Heffetz, and Rabin (2019), if a student's rank ordered list is $s_{i_1} \succ \cdots \succ s_{i_k}$, then

$$\Pr[\text{admitted to } s_{i_n} \mid \text{rejected by } s_{i_1}, \ldots, s_{i_{n-1}}] = \Pr[\varepsilon_{i_n} \geq y_{i_n} - x_{i_n}] \equiv q_{i_n}.$$

Hence, the unconditional probability can be given as

$$p_{\succ}(s_{i_n}) = \Pr[\text{admitted to } s_{i_n}] = q_{i_n} \prod_{j=1}^{n-1} (1 - q_{i_j}),$$

which corresponds to being rejected by each school prior to s_{i_n}, and then being admitted to s_{i_n}. Hence, $p_{\succ}(s_{i_n})$ corresponds to the admissions PMF induced by the ROL, $\succ$.

Let $\pi : [0,1] \to [0,1]$ be the probability weighting function (PWF) that maps some actual probability p to some perceived probability $\hat{p} = \pi(p)$. Then, the belief "distribution" of $\succ$ corresponds to $\hat{p}_{\succ}(s_{i_n}) = \pi(p_{\succ}(s_{i_n}))$. π is crucial in our model for accounting for pessimism; in contrast with standard PWFs, we anticipate that in allocation mechanisms, $\pi(p) < p$ for smaller p – that is, peoples' anticipated probabilities of admission into a selective schools tend to be less than they truly are.

Next, we define $\mu(\cdot)$ to be the traditional linear prospect theory gain-loss utility function,

$$\mu(x) = \begin{cases} x & x \geq 0 \\ \lambda x & x < 0 \end{cases} \tag{3}$$

so λ is a parameter corresponding to the strength of the student's loss-aversion. We let $m : \mathcal{S} \to \mathbb{R}$ represent the consumption utility of receiving admission to a school. Then, the reference dependent news utility of gaining admission to a new school s', given expectations of admission to s is [11]

$$u(s' \mid s) = \mu(m(s') - m(s) - \alpha \mathbf{1}\{s' = s_0\}). \tag{4}$$

In this function, an additional disutility term is included with strength α when the mechanism grants admission to s_0 – intuitively, this corresponds to the amount that a student would pay to get in anywhere that isn't their guaranteed school. With α, we capture potential undervaluation due attributed to ego; if the student is admitted by the mechanism to their safety school, they experience disutility weighted by α [12].

Hence, from Equation 2, we can say that the utility that the student experiences when they submit the rank ordered list $\succ$ is

$$U(\succ; \mathcal{S}, \lambda, \alpha, m, s_0, \pi, q_1, \cdots, q_n) = \sum_{c \in \succ} \sum_{r \in \succ} (m(c) + u(c \mid r)) \hat{p}_{\succ}(r) \hat{p}_{\succ}(c). \quad (5)$$

As U is based on Dreyfuss, Heffetz, and Rabin (2019), given sufficiently large λ, we capture overvaluation.

We posit that the student submits $\succ^*$ such that

$$\succ^* := \arg\max_{\succ \in \succ} U(\succ). \quad (6)$$

4 Experimental Design

We propose a multistage experiment, consisting of incentive-compatible procedures that calibrate agent-specific exogenous parameters of U, including λ, α, m, and π. Our concluding stage will integrate with existing stakeholders in a particular instance of school choice DA, and we hope to contrast the similarity of ROLs selected from Equation 6 with the expected ROL from Dreyfuss, Heffetz, and Rabin (2019) and Gale and Shapley (1962). Additionally, we design our experiment in an effort to learn more about diversions in traditional parameters such as λ and α, as their calibrations can give us a social interpretation of underlying behavioral phenomena, which we anticipate may vary from standard literature. An overview of the study methodology is provided in Figure 2 [13].

4.1 Participants

Our experiment will run partnered with a school system that currently uses deferred-acceptance to match students and schools. There are several such districts in the United States, including New York City and Boston's public schools (Roth (2008)).

However, the mechanisms used by these districts use a variety of parameters in calculating student priority that can be difficult to account for (e.g. neighborhood proximity, fit, sibling accommodation, equity tie-breaking etc.). To work around these complications, we study school choice in Ghana [3].

Ghana allocates 350,000 elementary school students to 700 senior high schools (SHS) every year, using a constrained deferred acceptance algorithm in which students may apply to at most 6 schools and must specify their preferences prior to receiving their test scores (Ajayi (2014)). Unlike Boston and New York, the test score is the *only* relevant variable in determining a student's match.

[3] Though the Ghanian school system would need to agree to working with us in a research setting, we believe this may be feasible, as prior collaborations between school choice mechanism designers and school districts is common. These collaborations often flourish, as many families and students experience the benefits of improved mechanisms [14].

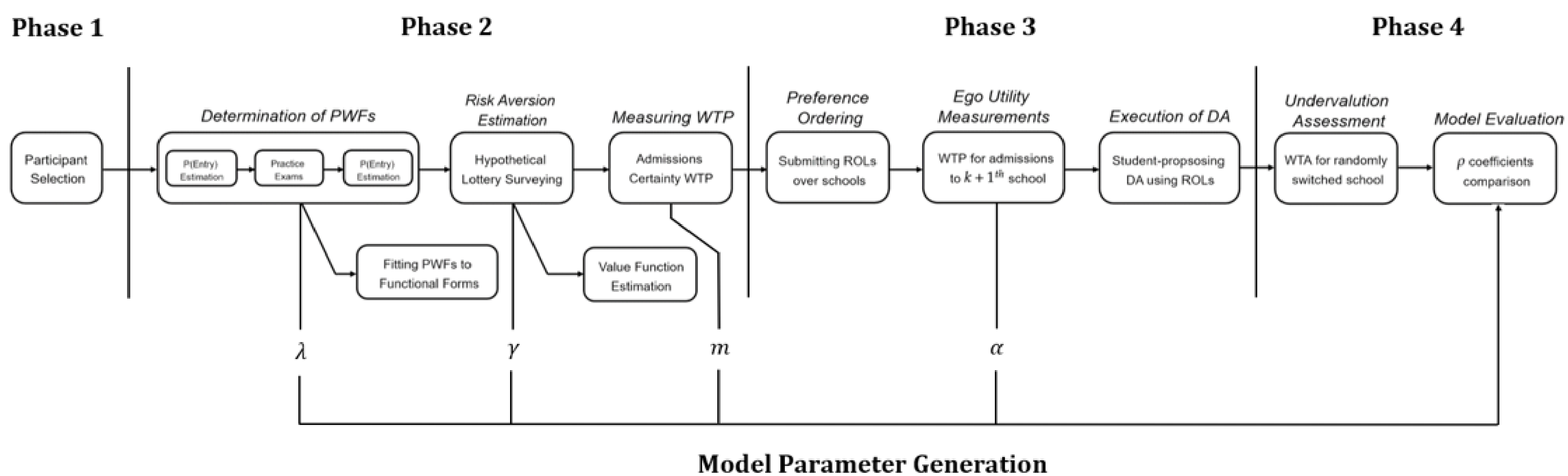

Figure 2: Graphical Depiction of Study Phases [15]

The constrained nature of this school choice is initially problematic to our study design: constrained school choice is not strategy-proof. Indeed, Ajayi (2014) recognizes that students with the same exam scores from lower-performing primary schools gain admission to less-selective secondary schools than students from higher-performing primary schools, partly because ”they are less likely to use sophisticated application strategies”; similar concerns on inequality resulting from strategizing are raised in Calsamiglia, Haeringer, and Klijn (2010).

That being said, the constrained problem is ubiquitous in real world applica-tions—New York and Boston have caps as well. So, we modify our approach to limit our participation to find subjects whose behavior will not change in response to the cap. Given that students often choose to stay within a proximal geographical radius when applying to schools, we choose to conduct our study with students interested in applying to schools only within Southern Ahafo, a diverse region in Ghana with approximately half a dozen vocational schools. For these students, the cap will most likely not induce strategic deviations. We will randomly select a sample of $N = 100$ students applying to SHS schools, and ensure that they are only interested in applying to schools geographically proximal to the Southern Ahafo region.

Of course, in limiting the sample, we need to verify that such a subsample is representative of the overall population of students. We would examine potentially confounding variables such as historic performance of the primary school (as cited as a source of variation in Ajayi (2014)), household income, and selectivity of secondary schools and determine whether the subsample is consistent with the overall population distribution. It is quite possible we find inconsistency here—a region with less schools may likely be rural, underfunded, less affluent, and featuring worse schools than Ghana as a whole. If this were to be the case, one should be cautious of the limitations of our analysis; nevertheless this procedure is the best we can do, given the reality that real-world school choice faces practical constraints.

4.2 Surveying

Variable	Description	Fitted in Section
γ	γ corresponds to a tuning parameter for the PWF functional form in 1. Though, it can also be used as an interpretation of risk aversion.	4.2.1
λ	Coefficient of Strength of Loss Aversion as used in Equation 3.	4.2.2
m	Reference-independent (individual) consumption utility function of receiving admission into each school. See Equation 4.	4.2.3
α	Strength of disutility from admission into safety option. Loosely corresponds to amount an individual would pay to get in anywhere other than safety option. See Equation 4.	4.3.2

Table 1: Each individual survey was designed to isolate and calibrate a parameter in Equation 5. [16]

4.2.1 Probability Weighting

In Ghana's school allocation mechanism, standardized tests are a great source of uncertainty for the student – the student has not even yet taken their standardized exams at the time of submitting their ROL. As a result, understanding how this uncertainty is modeled is imperative for fitting π. We focus on trying to estimate students' probability weighting functions relative to their true prospect probabilities.

First, we will ask all participants to estimate their perceived probabilities of getting into all schools in their region, $\mathcal{S}$. Then, we will ask students to take three practice exams of similar level to their entrance exam, prior to applying. After revealing them their practice exam scores, students will once more estimate their perceived probabilities of getting into all schools.

To incentivize participants to try their best on each exam, we will randomly select one participant and one practice test score out of his or her three scores. Re-scaling each test score t to be from 0 to 100, we will pay that selected participant, $t/4$. In this way, we create an incentive-compatible structure.

Given their practice exam scores, we can produce an individual-based distribution of their performance on the actual entrance exam (based on the standard error of measurement of the entrance exam, a property typically measured by exam writers). Since in Ghana's school match, the test score cutoffs for admission from previous years are public information, we can compute a relatively confident "actual" probability that an individual gets accepted to each school in $\mathcal{S}$ (especially given that historical cutoffs have minimal variance). We measure the decision weights of the perceived against actual prospect probability to characterize each subject's PWF. [4] [17]

Pless and Miller (1979) found that telephone and mail methods of survey tend to elicit more 'truthful' sets of responses on questions (particularly that pertain to socially undesirable responses) than in-person inquiry. Thus, when querying individuals on their perceived probabilities post hearing about their test scores, we opt to survey them with a follow up phone-call.

We select 10 surveyors by advertising online (i.e. through campus social media platforms and short term gig posting websites) and by sending out emails/posting flyers around public spaces, particularly within a university campus. In line with Kapor, Neilson, and Zimmerman (2016), we require each trainee to take an online training module on data confidentiality and privacy. Then, we simulate the phone-call interaction to enable them to practice their task. We train each worker to survey the subjects by phone, and pay them the standard rate of 20 dollars per hour. Given that there are at most 9 schools in the district and that each interview takes 15 minutes, each interviewer spends about one hour on 4 students, we expect to pay for 25 hours for one round of interviewing all 100 subjects. Since we perform this procedure twice, we expect to pay 1000 dollars in total to get the

[4] By querying belief probabilities before and after taking practice exams, we can study the distinction of PWFs at different levels of information. These distinctions may, in future, account for the fact that certain low-income populations are unable to take practice tests and get feedback/extra information in advance.

necessary PWF information.

After eliciting each individuals' PWF, we will fit the function both to: (1) the decision-weighting functional form proposed by Kahneman and Tversky in their 1992 work on revising prospect theory (commonly used by other studies); and (2) the inverse of (1) which will permit this functional form to capture pessimism in low-probability ranges and overconfidence/optimism in high-probability ranges (coordinating with the pessimism and overvaluation phenomenon we which to observe. The specific functional form is given in 1.

4.2.2 Risk Aversion Estimation

In line with the existing body of literature, we will conduct lottery-based surveys to estimate individual-level risk aversion to generate a value function for both gains and losses.

To do so, we present participants with lotteries of the following form:

> Consider the following lottery: You have a X% chance to earn/lose [payoff i] and an (1-X)% chance to earn/lose [payoff j]. How much would you be willing to pay to participate/avoid this lottery?

Analogous to the methodology employed by Tversky and Kahneman (1992), each lottery will be presented alongside a descending series of eleven certain outcomes logarithmically spaced between the extreme outcomes of the lottery, as well as the expected value of the lottery. Participants will indicate a preference between each of the certain outcomes and the lottery. Certainty equivalents will be estimated by the midpoint between the lowest accepted value and the highest rejected value in the set of choices.

Participants will be presented with a large, fixed number of such lotteries in the survey. To incentivize truthfulness, one of the lotteries (with purely positive outcomes) and accompanying certainty equivalents will be selected at random. Depending on their indicated preference, participants will receive either a payout scaled from the certainty equivalent or a payout scaled from the lottery (after executing a random [0,1] draw to simulate the lottery).

As noted in Tversky and Kahneman (1992) once more, assumption of preference homogeneity is both necessary and sufficient to represent the value function v as a two-part power function of the form:

$$v(x) = \begin{cases} x^{\alpha} & x \geq 0 \\ -\lambda(-x)^{\beta} & x < 0 \end{cases}$$

with behavioral studies commonly finding preference homogeneity as a reasonable approximation for lottery-based surveying. As noted in the theoretical preliminaries, we choose $\alpha, \beta = 1$.

For each participant-lottery-certainty equivalent triple, the parameters for the two-part power function will be solved for via:

$$v(E) = v(i)\pi(p_i) + v(j)\pi(p_j)$$

where i, j are the payouts of the lottery, and p_i, p_j are the associated probabilities for the respective payouts.

Here, $\pi(p)$ is defined to be the decision-weighting function in Equation 1, with γ calculated according to the following procedure (analogous to Tversky and Kahneman (1992)):

1. For each prospect of the form $(x, p; 0, 1-p)$, let c/x be the ratio of the certainty equivalent of that prospect to the nonzero outcome x.
2. For each individual, plot the median value of c/x as a function of p, for both positive and negative prospects.
3. Fit these $(p, c/x)$ points to the functional form defined in Equation 1, to estimate γ and the decision-weighting function.

Parameter estimations will be averaged over all such triples to obtain aggregate estimates of the relevant parameters.

4.2.3 Surveys of Willingness to Pay (WTP)

In the second stage of the study, we will survey (households/students) to obtain their self-reported willingness to pay (WTP) to be guaranteed admission to select schools. Their WTP for a school will be used as a proxy for their pecuniary utility of it.

Students will be presented with a questionnaire of the following form:

> Imagine you are applying for schools. You can be guaranteed admission to [school] for a fixed, one-time payment. Please note that guaranteed admission does not include guaranteed funding for costs of attendance.
> For the payment, you may spend up to [median/mean monthly household income in Ghana × some fixed constant k] from an external, independent funding source.
> Any money you do not expend for the payment you may keep.
> Please indicate the amount of money you would be willing to spend from the fund to guarantee your admission to [same school as above]:
>
> ____________________________
>
> **Amount**

Individual survey respondents will be presented with this question in the context of every school in the school set.

To elicit truthfulness, a reserve price $p \in [0 : M]$, will be drawn ex-post, where M is the maximum, as stated in the form. Then, $N/2$ survey respondents will be randomly selected to receive: (1) guaranteed admission to one of the schools among those presented to them in their questionnaires (students will not be mandated to attend this school) only if the amount reported exceeds p; and (2) their associated residual payout for that school (i.e. the amount of money not expended on guaranteeing admission to a select school, $M - p$) if they choose to enroll there, otherwise they will receive M if they did not receive admission in (1).

Payouts are structured around the median monthly household income such that the prior award acts as a sufficiently large incentive (Kahneman and Tversky's seminal work on prospect theory (Kahneman and Tversky (1979)) and others employ a similar strategy for approximating a reasonable and incentivizing payout). Given appropriate instruction, this incentive-compatible structure is well studied and utilized throughout experimental economics to measure willingness to pay (Burfurd and Wilkening (2018)) and is an adaptation of the Becker–DeGroot–Marschak method, with anchor point M.

From this procedure, we will have a function m, related to the admission utility of each school in $\mathcal{S}$.

4.3 Matching

4.3.1 Submitting ROL

All students submit their final ROL of the schools they actually apply to to the centralized Ghana clearinghouse. The algorithm will be run the following week.

Note from Section 4.2.3 we can determine the truthful ROL over schools for each student, where preference is dictated by increasing willingness to pay to guarantee admission; i.e. for any student, we will have a preference list of the following form:

$$s_1 \succ s_2 \succ ...s_k \succ s_{k+1} \succ ... \succ s_n.$$

with $m(s_1) \geq m(s_2) \geq \cdots \geq m(s_n)$. Here s_k refers to the school in which the student received guaranteed admission (if they did) in 4.2.3.

4.3.2 Ego Utility and Estimating α

Consider the set of students who were granted guaranteed admissions to some school in 4.2.3 or who are guaranteed admission to some school. For each of these students, we will ask the following, where s_{k+1} represents the first school after which $m(s_{k+1}) < m(s_k)$.

> Tomorrow is the SHSS process will take place. Please consider the following proposition:
> You can be guaranteed admission to [s_{k+1}] for an additional fixed, pre-application screening fee. Please note that guaranteed admission does not include guaranteed funding for costs of attendance.
> For the payment, you may spend up to [median/mean monthly household income in Ghana × some fixed constant k] from an external, independent funding source.

> Any money you do not expend for the payment you may keep.
> Please indicate the amount of money you would be willing to spend from the fund to guarantee your admission to [same school as above]:
>
> ______________________________
>
> **Amount**

As in 4.2.3, we will use the adapted Becker-DeGroot-Marschak method to use a reserve price to elicit truthful answers, but this time we will only select one person to receive the reserve price offer.

If the amount reported in the form, F is any value above 0, we observe ego utility. Since $m(s_{k+1}) < m(s_k)$, that means that there must be some ego term $m(s_{k+1}) + \alpha = m(s_k)$, where α corresponds to F.

4.3.3 Execution of Deferred Acceptance

At this point, student-proposing deferred acceptance is run, using each student's ROL to make matches by the Ghana clearinghouse.[18] The final matches are saved. [19]

4.4 Post-Match Analysis

We ask a survey question to verify the existence of reverse-endowment in allocation mechanisms, due to a lack of studies that do so. At this stage, we have a fit for λ, α, m, π and $q_1, \cdots, q_n$, so we also compare the predictions of our model with the models of Dreyfuss, Heffetz, and Rabin (2019) and Gale and Shapley (1962).

4.4.1 Assessment of Undervaluation and the Reverse-Endowment Effect

After conducting the initial DA matching, we will further sample the randomly selected respondents from Section 4.2.3 who were granted guaranteed admissions to one of the schools, s_k, and were admitted to a school in DA (or none at all) with $m(s) < m(s_k)$. We do so in order to evaluate the existence of the undervaluation phenomenon. We will adapt the methodology presented by Kahneman, Knetsch, and Thaler (1990), in their seminal work on the endowment effect.

That is, students from this sub-sample will be presented with a questionnaire of the following form:

> Previously, you were randomly selected to be granted guaranteed admission at [student-specific school]. Please consider the following proposition:
> You can receive a one-time, fixed sum payment in exchange for relinquishing your guaranteed admission to this specific school and instead being granted guaranteed admission to some other randomly selected school (which you will pay for as specified in the previous form).

> Please indicate the minimum payment necessary for you to accept the previously outlined proposition:
>
> ________________________________
>
> **Amount**

Denote F as the value indicated in the form. To elicit truthfulness, similar to 4.2.3, a reserve price, p, will be drawn and the transaction will go through at F if $p > F$. [5]

Some fixed number of survey respondents will be randomly selected for which the procedure will be carried through (i.e. their guaranteed admission at school s_i will be replaced with guaranteed admission at school s_j, and they will receive their indicated fixed, one-time payment). To limit anchoring effects on the WTA/WTP figures reported, we will conduct this portion of the study sufficiently after 4.2.3.

We structure the proposition alternative as such so that differences between WTP and WTA reflect only differences in valuation of guaranteed admission to a specific school, rather than including the potential valuation difference of guaranteed admission to any school.

Participants will have the option to select any value from 0 to M times some constant factor exceeding one (to accommodate the possibility of a positive endowment effect). Participants will also have the option to indicate "no amount is sufficient" if they would not prefer to relinquish their previous guaranteed admission.

After conducting these surveys, we will possess a student's pre-admission WTP to attend a specific institution and post-admission WTA to attend a potentially different institution. By comparing these figures, we aim to detect and measure the degree of undervaluation, $\mathcal{U}$, of the now-reserve option school:

$$\mathcal{U} = \beta * \delta * WTA_{\text{post}} - WTP_{\text{pre}}.$$

Over a limited time frame (e.g. one annual admissions cycle) and without the introduction of preference-altering information, standard economic theory would attribute differences between WTA_{post} and WTP_{pre} to time discounting. We capture these potential time-related differences through the inclusion of a δ and a β factor (defining a quasihyperbolic discounting function); for our estimates of δ and β, we refer to meta-analyses which report discounting factors of $\delta = 0.33$ and $\beta = 0.82$ respectively (Cheung, Tymula, and Wang (2021), Matousek, Havranek, and Irsova (2022)).

The differences (or lack thereof) between WTA_{post} and WTP_{pre} which persist despite the time-discounting corrections will then be analyzed as evidence for undervaluation.

[5] Note that the above questionnaire structures the alternative to committing to one's previous guaranteed admission to be guaranteed admission to some other randomly selected school rather than repeated participation in the mechanism described in 4.2.3

4.4.2 Assessing Model Performance

Now, we are able to assess the performance of our model by comparing its ability to predict the actual ROL submitted by the students, relative to other potential models. Here, we have three predictive ROLs for each student:

1. $\succ_1$, as predicted by our model (according to Equation 6)
2. $\succ_2$, the standard economic theory prediction that students submit true preferences, as stated in 4.3.1
3. $\succ_3$, as predicted by the model in Dreyfuss, Heffetz, and Rabin (2019), where $\lambda, q_1, \cdots, q_n, m, \pi$ are inputs.

In addition, each student has their true ROL collected in 4.3.3, notated as $\succ$.

We quantify the predictive ability of each of the three models by computing the rank correlation coefficient for each $\succ_i$ with $\succ$. We accomplish this using Spearman's ρ, which extends Pearson's correlation coefficient to ordinal, rank variables. Thus, for each student, s, we calculate a row-vector $\rho_s = [\rho_{s,1}, \rho_{s,2}, \rho_{s,3}]$, where $\rho_{s,i}$ is defined as Spearman's ρ between lists $(\succ, \succ_i)$ for student s.

Our next task is to compare the three coefficients in ρ_s to determine what model has the strongest correlation with the true ROL. As a first check, we can test whether each of these models is predictive. This amounts to testing the null hypothesis that the average correlation, over all students, for each model is zero, where we reject the null if there is a significantly nonzero correlation. As for notation, we refer to $\overline{\rho_{S,i}}$ for all $i \in \{1,2,3\}$, where $\overline{\rho_{S,i}}$ is the sample mean of $\rho_{s,i}$ over all students, $s \in S$. Then, we have the simple t-test for each of our $i = 1, 2, 3$ models:

$$H_0 : \overline{\rho_{S,i}} = 0, H_A : \overline{\rho_{S,i}} \neq 0$$

$$t_i = \frac{\overline{\rho_{S,i}} - 0}{\frac{\sigma}{\sqrt{n}}}$$

Where t has $n-1$ degrees of freedom and σ is the sample standard deviation of our individual student correlation coefficients. [6]

Next, we want to see which model performs "best," which we check by seeing whether a correlation coefficient is stronger than the other two. Since ρ is standardized on $-1 \leq \rho \leq 1$, we can also make ad-hoc comparisons from the magnitude alone and then statistically test the significance of any difference we observed.

Alternatively, we can perform an asymptotic z-test, using Fisher's r to z transform, as described in Lee and Preacher (2013) and Steiger (1980). Thus, we can test the statistical significance of the difference between each of our three correlation coefficients. This gives us our conditions to conclude that one model is the "best" predictor of $\succ$:

[6] This procedure may not be intuitive: essentially, we are able to find the true correlation between a student's ROL and the ROL produced for each model. Then, we take the average correlation coefficient, for each model, over the distribution of student correlations. Therefore, we are simply testing a sample average, and by the CLT, we can apply this relatively simple statistical procedure.

1. The average correlation coefficient is significantly nonzero
2. The average correlation coefficient is significantly stronger than the two other models.

In other words, if our model is best, we would expect to see $\overline{\rho_{S,1}} \neq 0$, $\overline{\rho_{S,1}} > \overline{\rho_{S,2}}$, and $\overline{\rho_{S,1}} > \overline{\rho_{S,3}}$.

4.5 Analysis of Budget and Payouts

- We allocated \$1025 for costs related to training and eliciting values of the PWF.
- Given that we call a ”significant payoff” \$150, we pay 100/2 students with at most \$150 each in order to reveal their preferences, which is a total of \$7500.
- Only one student is selected to receive their ego utility payment, which is at most \$150.
- In accordance with prior years, we expect no more than 20% of participants to be unmatched after deferred acceptance, so we expect no more than 10 students who need reverse-endowment payouts of at most \$200 each, which is a payout of \$2000.

In total, our worst case payoff is \$10,675, though, we anticipate the true cost to fall a ways below this threshold, recognizing that we will not need to payout \$150 to each student every time they are selected by the Becker–DeGroot– Marschak mechanism.

5 Expected Results

5.1 Model Predictions

First we share some general behaviors predicted by our model.

Result 5.1. *Consider a student, with $\alpha = 0$. Given a school list of only one, $\mathcal{S} = \{s_1\}$, where $q_1 > 0$ and $m(s_1) > 0$, under U, the student only applies if they have $\lambda < \frac{2-\pi(q_1)}{1-\pi(q_1)}$.*

Proof. The utility of the empty preference list is 0, as observed from Equation 5. The utility of the singleton ROL consisting of just s_1, is as follows:

$$\begin{aligned}
U &= \sum_{c\in[s_1,\emptyset]} \sum_{r\in[s_1,\emptyset]} (m(c) + u(c \mid r))\hat{p}_{>}(r)\hat{p}_{>}(c) \\
&= (m(\emptyset) + u(\emptyset \mid s_1))\hat{p}_{>}(s_1)(1 - \hat{p}_{>}(s_1)) + (m(s_1) + m(s_1))(1 - \hat{p}_{>}(s_1))\hat{p}_{>}(s_1) + m(s_1)\hat{p}_{>}(s_1)^2 \\
&= (-\lambda m(s_1) + 2m(s_1))\pi(q_1)(1 - \pi(q_1)) + m(s_1)\pi(q_1)^2 \\
&= (-\lambda m(s_1) + 2m(s_1))(1 - \pi(q_1)) + m(s_1)\pi(q_1) \\
&= (-\lambda + 2)(1 - \pi(q_1)) + \pi(q_1)
\end{aligned}$$

The student only applies if and only if $U > 0$, hence, they apply if and only if

$$\begin{aligned}(-\lambda+2)(1-\pi(q_1))+\pi(q_1)>0 &\iff (-\lambda+2)(1-\pi(q_1))>-\pi(q_1)\\ &\iff (\lambda-2)(1-\pi(q_1))<\pi(q_1)\\ &\iff \lambda(1-\pi(q_1))-2+2\pi(q_1)<\pi(q_1)\\ &\iff \lambda<\frac{2-\pi(q_1)}{1-\pi(q_1)}\end{aligned}$$

□

From this result, if there exists a school for which the student believes they have a 50-50 shot and they are not significantly risk averse ($\lambda < 3$), then they will certainly partake in the mechanism.

Result 5.2. *Consider a student who is faced with applying to a school list of size n, that is, $\mathcal{S} = \{s_1, \cdots, s_n\}$. Assume the student has positive consumption utility for each school. If the student is not risk averse ($\lambda \leq 1$), the student will apply to schools, as long as there is a school with non-zero perceived probability. In fact, it can be shown that any preference ordering over all schools is more preferable than submitting nothing.*

Proof. Recall that

$$U(\succ) = \sum_{c\in\succ}\sum_{r\in\succ}(m(c)+u(c\mid r))\hat{p}_{\succ}(r)\hat{p}_{\succ}(c).$$

Consider any ordering, $\succ$, of the full set of schools $\mathcal{S}$. For terms in the summation where $c = r = s$, positive utility is added, since $m(c) + u(c \mid r) = m(c) > 0$. For terms in the summation where $c \neq r$ (w.l.o.g assume that $c \succ r$), then

$$(m(c)+u(c\mid r)+m(r)+u(r\mid c))\hat{p}_{\succ}(r)\hat{p}_{\succ}(c)\geq 0 \qquad (\star),$$

is added to the sum. We show that for all possible combinations of $(c, r) \in \mathcal{S} \times \mathcal{S}\}$, where $c \neq r$, $(\star)$ is non-negative, hence, $U > 0$, eliminating the appeal of the empty list.

Since probabilities can only be positive, we instead show that

$$(m(c)+u(c\mid r)+m(r)+u(r\mid c))\geq 0 \qquad (\star\star),$$

Notice that

$$\begin{aligned}m(c)+u(c\mid r)+m(r)+u(r\mid c) &\geq m(c)+m(c)-m(r)+m(r)+\lambda(m(r)-m(c)-\alpha)\\ &= 2m(c)+\lambda(m(r)-m(c)-\alpha)\\ &= 2m(c)+\lambda m(r)-\lambda m(c)-\lambda\alpha\\ &\geq 2m(c)-\lambda m(c)-\lambda\alpha\\ &\geq 2m(c)-m(c)-\alpha\\ &= m(c)-\alpha\end{aligned}$$

Since we assumed that the consumption utility of every school was positive, that means that $\alpha \leq \min_{c \in \mathcal{S}}\{m(c)\}$, indicating that $m(c) - \alpha \geq 0$ for all $c \in \mathcal{S}$. Hence, we have shown that ($\star\star$) is satisfied. $\square$

This result indicates exactly what we wished to capture when constructing the model – for students who are not risk averse, empty preference orderings aren't very attractive.

Interactive Applet

In order to play with the model with different values and parameters, we have written an HTML/JavaScript applet, which outputs the optimal ROL for any set of legal parameters. You can access this applet by visiting:

https://hunterguru.com/econ178-final-project/applet.html

In the applet, the user can define the WTP and true probability of some schools they plan to apply to. Then, one can adjust the loss-aversion and risk perception of the individual, and see how the optimal ROL changes, as these parameters are adjusted.

5.2 Parameter Expectations and Interpretations

We estimate that λ will fall on the lower end of 1.8 and 2.1, as noted in Brown et al. (2021), since we calibrate λ by running standard prospect theory lotteries. Dreyfuss, Heffetz, and Rabin (2019) fit $\lambda = 2$ with a standard error or 0.2. We are unsure of how large α will be, though we believe that it may fall closely to $\min_{s \in \mathcal{S}} m(s)$, as the true consumption utility of the least appealing school could likely fall to 0. For, γ, given that we believe that lower probabilities will be under-weighted and higher probabilities will be over-weighted, we expect $\gamma > 1$.

We expect that the model will outperform Dreyfuss, Heffetz, and Rabin (2019), given that values of $\alpha = 0$ and $\pi(x) = x$ gives the same model; hence, our model can provably perform no worse. For students who attempt to ”game the system”, we expect that our model will out-perform the standard economic prediction; however, for students who submit true preferences, our model may be incorrect. Given that statistical models exist that predict *when* students will deviate from true preferences (based on education level, wealth level, or other demographic factors), by combining the models together, we believe that this hybrid model may significantly outperform the standard model as well.

5.3 Takeaways for the Market Designer

If our model out-predicts the standard economic model and the model in Dreyfuss, Heffetz, and Rabin (2019), one may be curious how this information could better help a market designer. When designing a mechanism to be strategy-proof, the designer could attempt to calculate a normalized proxy for m and then pick the outcome distribution that maximizes the welfare of the participants; that is, they could attempt to create a mechanism that

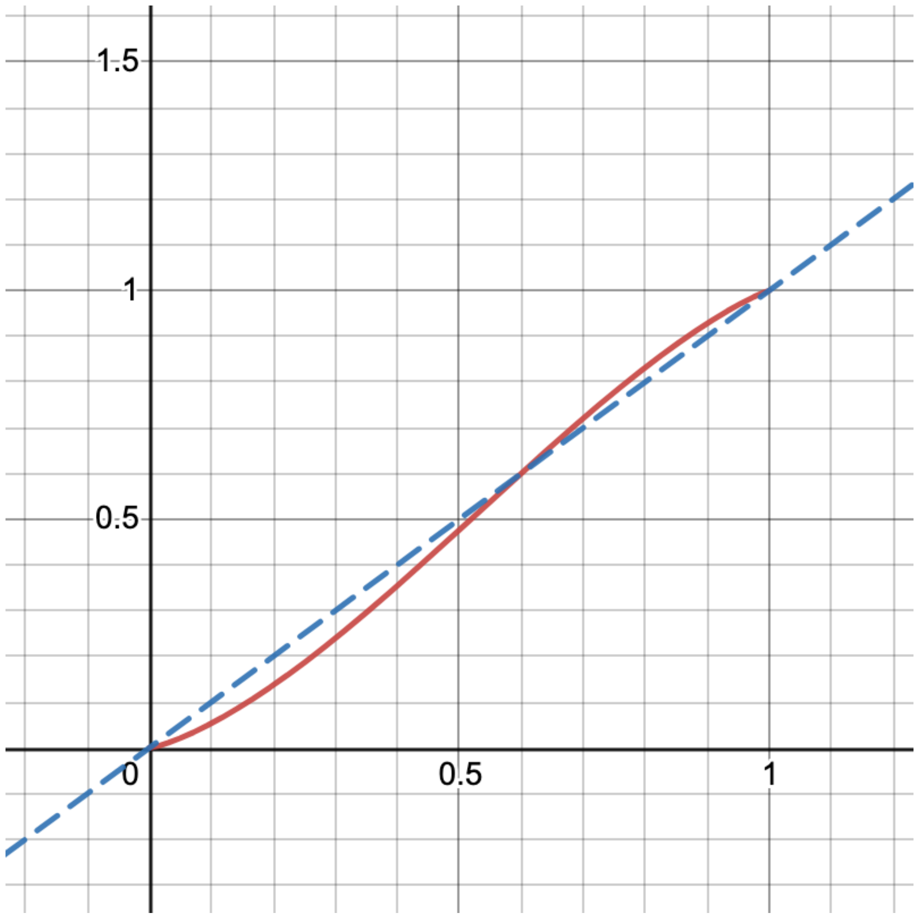

Figure 3: A sample PWF functional form with $\gamma = 1.3$

optimizes $Q_a = (q_1, \cdots, q_n)$ for each student, for the objective function,

$$\sum_{a \in \mathcal{A}} U(\succ_a^*),$$

where $\mathcal{A}$ is the set of all applicants, and $\succ_a^*$ is the ROL submitted by applicant a, as noted in Equation 6. A similar objective function can be constructed for the offer recipients, and the market designer can construct an equilibrium by finding the distributions $Q = \{Q_1, \cdots, Q_s\}$ maximizing the total utility for both recipients and proposers.

The model could also serve as a metric that predicts the deviation from submitting true preferences, which may encourage the designer to use a different mechanism in light of the circumstances. We anticipate that a more obvious mechanism, such as random serial dictatorship with participation, in the case of Ghana, may reduce the effects of λ and π.

Perhaps more interestingly, if a spectator observes $\lambda, \mathcal{S}, \pi, \alpha$, and $\succ^*$, and a learning model can accurately predict the true preference relation $\succ$ with high accuracy, then the designer could augment the preferences of the student (given that such augmentations in expectation increase welfare).

Interpretations of the parameters can also help a market designer adjust features of the mechanism. For instance, given α, and a good estimation of how far down the reference school, s_0, is ranked, the mechanism designer can increase the marginal cost of each application following a cutoff, by α.

6 Future Steps and Flaws

We recognize that there are several limitations to our approach and other ways to extend our analysis. Here, we specify a few such cases [20].

First, we note the possibility of other confounding factors leading to misrepresentation of preferences. For example, we tend to estimate γ by surveying preferences over different lotteries. It is possible for familial wealth level to be a factor that influences preferences over these lotteries. In addition, wealth level may also directly affect the ultimate ROL submitted over schools. For instance, students from families with lower wealth may tend to submit ROLs based on valuing closer distances to home as a result of other family responsibilities, even if it is a misrepresentation of their true preferences in a vacuum. Ultimately though, we think that confounding variables are slightly less of a concern for our study, compared to the possibility of alternative explanatory variables for misrepresentation like social/influence networks, reputation, family, etc.[21]

Many real-world systems (like that of Ghana) tend to put limits on the number of schools that one can apply to. As indicated in Coles and R. Shorrer (2014), the truncation of preference-lists may break the strategy-proofness of DA for the proposing side. Extending analysis to a setting in which we do not impose truncation/self select for individuals whose list at most contains the max allowable number of applications would be beneficial. Moreover, it would be interesting to broaden the model and analysis to cases in which there is additional marginal cost of applying to each extra school. This is a more realistic setting and would presumably alter both the length and choice to misrepresent by students.

Some of the surveys by which we calibrated parameters (like for γ) were based on lotteries with monetary payouts. However, we think it would be useful to consider lotteries based upon school assignment itself, given that different risk aversion coefficients may govern these sets of choices. Also, the imposition of an upper bound in our WTP surveys may lead to inaccurate elicitation of true WTP; subjects' answers may be prone to both the income effect and anchoring bias.

It would be interesting to run further experiments in which we study how individuals' perceptions of the probabilities across various prospects change before and after taking a practice test. Along these lines, we recommend that districts give the option for all students to take practice tests for free before the actual entrance exam, as information on one's own performance may have significant effects on the possible ROLs they ultimately submit.

Acknowledgements

We would like to thank Marcelo Clerici-Arias, Lily Liu, and Matt Brown for their help over the course of this project, guidance in developing a bibliography, and feedback on the design of our experiment. We are also grateful to our anonymous peer reviewers, who provided invaluable commentary in the revision process.

Peer Feedback

1. We revised our existing abstract by adding references to our study design and anticipated results to address peer comments. Other comments asking for a “slower start” to the abstract were ignored, owing to the need for brevity. For the same reason, we disregard comments asking for ”more detailed explanations about the possible result[s].”

2. Without a mechanism design background, some peer reviewers understandably found our paper technical; we thought it would be useful for our paper to begin with some useful background, though the material presented in this paper is mostly self contained.

3. Moved the problem statement to the beginning paragraph to serve as a hook.

4. Here, we provide layperson definitions of the phenomenon we intend to study in response to multiple peer reviews requesting terminology definition and noting that ”narrowing down to only certain main ideas” would help prevent readers getting lost in concepts.

5. Good catch Matt – we moved this evidence away from overvaluation, as it could be misleading

6. Added a source which further contextualizes pessimism in the context of school admissions, with emphasis on zero marginal cost; the source in fact directly states that many students may not apply out of fear of not getting in.

7. Moving definition of undervaluation before stating its properties.

8. Paragraph added to the literature review to more explicitly introduce the reverse-endowment effect and our objectives regarding it earlier in the paper.

9. Directly stating that this flies in the face of prospect theory.

10. We agree this bolding is difficult to read, but we will be keeping the notation, as it is the notation utilized in other similar papers.

11. Reworded explanation of reference dependent utility in terms of news utility

12. Included an explicit explanation of the α term

13. Peer reviewers found ”some parts [of the paper]...a little bit hard to understand, especially since the paper is so technical,” so we created a more easily-digested, graphical representation of some of the most technical aspects of the paper (i.e. the study design).

14. Explicitly addressing why this collaboration may work.

15. Resized figure sideways to improve readability

16. This table addresses the feedback we received about providing clearer definitions and/or specification of certain key concepts that are pivotal to our later results. We try to provide concise, but more intuitive descriptions of the variables in addition to referencing where they are fit later on. In conjunction, this table and the earlier diagram depicting study design stages should also address peer feedback asking

for summaries and justifications of evidence types, as well as linking various stages of the study to the overarching purpose of the paper.

17. We believe that informing subjects about their scores on their three practice exams and on historical cutoffs, gives them enough information to specify an 'actual prospect probability'. Of course, we could modify our current experiment design to inform them of an explicit probability value, but we would be better served by setting up explicit lotteries that they make rankings over in that circumstance to measure the effect of probability weighing rather than directly surveying their perceived probabilities afterwards.

18. For background and the steps to the algorithm itself, refer to Gale and Shapley (1962).

19. We've revised elements of the post-match analysis for content. We also received feedback that a "detailed explanation of how actually the post-match analysis is done would be useful," which we've largely ignored, believing our two subsections to be sufficiently thorough.

20. Identification and analysis of confounding factors and other study limitations is provided in response to peer feedback noting their potential influence on expected results

21. This paragraph addresses the feedback we received about addressing the role of confounding factors

References

Abdulkadiroğlu, Atila and Tayfun Sönmez (2003). "School choice: A mechanism design approach". In: *American economic review* 93.3, pp. 729–747.

Ajayi, Kehinde F (2014). *School choice and educational mobility: Lessons from secondary school applications in Ghana*. Citeseer.

Ashlagi, Itai and Yannai A Gonczarowski (2018). "Stable matching mechanisms are not obviously strategy-proof". In: *Journal of Economic Theory* 177, pp. 405–425.

Blavatskyy, Pavlo and Ganna Pogrebna (2008). "Risk aversion when gains are likely and unlikely: Evidence from a natural experiment with large stakes". In: *Theory and Decision* 64.2, pp. 395–420.

Brenner, Lyle et al. (2007). "On the psychology of loss aversion: Possession, valence, and reversals of the endowment effect". In: *Journal of Consumer Research* 34.3, pp. 369–376.

Brown, Alexander L et al. (2021). "Meta-analysis of empirical estimates of loss-aversion". In: *Available at SSRN 3772089*.

Burfurd, Ingrid and Tom Wilkening (2018). "Experimental guidance for eliciting beliefs with the Stochastic Becker–DeGroot–Marschak mechanism". In: *Journal of the Economic Science Association* 4.1, pp. 15–28.

Calsamiglia, Caterina, Guillaume Haeringer, and Flip Klijn (2010). "Constrained school choice: An experimental study". In: *American Economic Review* 100.4, pp. 1860–74.

Camerer, Colin F and Teck-Hua Ho (1994). "Violations of the betweenness axiom and nonlinearity in probability". In: *Journal of risk and uncertainty* 8.2, pp. 167–196.

Chen, Yan and Tayfun Sönmez (2006). "School choice: an experimental study". In: *Journal of Economic theory* 127.1, pp. 202–231.

Cheung, Stephen L, Agnieszka Tymula, and Xueting Wang (2021). "Quasi-Hyperbolic Present Bias: A Meta-Analysis". In: *Life Course Centre Working Paper* 2021-15.

Coles, Peter and Ran Shorrer (2014). "Optimal truncation in matching markets". In: *Games and Economic Behavior* 87, pp. 591–615.

Dreyfuss, Bnaya, Ori Heffetz, and Matthew Rabin (2019). *Expectations-based loss aversion may help explain seemingly dominated choices in strategy-proof mechanisms*. Tech. rep. National Bureau of Economic Research.

Dubins, Lester E and David A Freedman (1981). "Machiavelli and the Gale-Shapley algorithm". In: *The American Mathematical Monthly* 88.7, pp. 485–494.

Edelman, Jonah (Dec. 2014). "Russian Billionaire's Superyacht Mysteriously Vanishes". en. In: *The Daily Beast*. URL: https://www.thedailybeast.com/russian-billionaire-vagit-alekperovs-superyacht-galactica-super-nova-mysteriously-vanishes (visited on 03/10/2022).

Etzioni, Amitai (1986). "The case for a multiple-utility conception". In: *Economics & Philosophy* 2.2, pp. 159–184.

Gale, David and Lloyd S Shapley (1962). "College admissions and the stability of marriage". In: *The American Mathematical Monthly* 69.1. Publisher: Taylor & Francis, pp. 9–15. ISSN: 0002-9890.

Hassidim, Avinatan, Assaf Romm, and Ran I Shorrer (2016). "" Strategic" Behavior in a Strategy-Proof Environment". In: *Proceedings of the 2016 ACM Conference on Economics and Computation*, pp. 763–764.

Kahneman, Daniel, Jack L Knetsch, and Richard H Thaler (1990). "Experimental tests of the endowment effect and the Coase theorem". In: *Journal of political Economy* 98.6, pp. 1325–1348.

Kahneman, Daniel and Amos Tversky (1979). "Prospect Theory: An Analysis of Decision under Risk". In: *Econometrica* 47.2. Publisher: [Wiley, Econometric Society], pp. 263–291. ISSN: 0012-9682. DOI: 10.2307/1914185. URL: https://www.jstor.org/stable/1914185 (visited on 02/09/2022).

Kapor, Adam, Christopher Neilson, and Seth Zimmerman (2016). "Heterogeneous beliefs and school choice". In: *American Economic Review*.

Klijn, Flip, Joana Pais, and Marc Vorsatz (2010). "Preference Intensities and Risk Aversion in School Choice: An Experimental Study". In.

Köszegi, Botond (2006). "Ego Utility, Overconfidence, and Task Choice". In: *Journal of the European Economic Association* 4.4. Publisher: Oxford University Press, pp. 673–707. ISSN: 1542-4766. URL: https://www.jstor.org/stable/40004814 (visited on 02/10/2022).

Kőszegi, Botond and Matthew Rabin (Nov. 2006). "A Model of Reference-Dependent Preferences". In: *The Quarterly Journal of Economics* 121.4, pp. 1133–1165. ISSN: 0033-5533. DOI: 10.1093/qje/121.4.1133. URL: https://doi.org/10.1093/qje/121.4.1133.

Lee, Ihno A and Kristopher J Preacher (2013). *Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software].* URL: http://quantpsy.org/corrtest/corrtest2.htm.

Li, Shengwu (2017). "Obviously strategy-proof mechanisms". In: *American Economic Review* 107.11, pp. 3257–87.

Liu, Juan (2010). "A study on college job applicant's intention to refuse job offer". PhD thesis. University of Missouri–Columbia.

Matousek, Jindrich, Tomas Havranek, and Zuzana Irsova (2022). "Individual discount rates: a meta-analysis of experimental evidence". In: *Experimental Economics* 25.1, pp. 318–358.

Meisner, Vincent and Jonas von Wangenheim (2021). "School choice and loss aversion". In: *Available at SSRN 3985777*.

Pless, Ivan Barry and JR Miller (1979). "Apparent validity of alternative survey methods". In: *Journal of Community Health* 5.1, pp. 22–27.

Rees-Jones, Alex and Samuel Skowronek (2018). "An experimental investigation of preference misrepresentation in the residency match". In: *Proceedings of the National Academy of Sciences* 115.45, pp. 11471–11476.

Roth, Alvin E (1984). "The evolution of the labor market for medical interns and residents: a case study in game theory". In: *Journal of political Economy* 92.6, pp. 991–1016.

— (2008). "Deferred acceptance algorithms: History, theory, practice, and open questions". In: *international Journal of game Theory* 36.3, pp. 537–569.

Steiger, James H (1980). "Tests for comparing elements of a correlation matrix." In: *Psychological bulletin* 87.2, p. 245.

Tversky, Amos and Daniel Kahneman (1992). "Advances in prospect theory: Cumulative representation of uncertainty". In: *Journal of Risk and uncertainty* 5.4, pp. 297–323.

Wu, George and Richard Gonzalez (1996). “Curvature of the probability weighting function”. In: *Management science* 42.12, pp. 1676–1690.