



Management Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Reversing Reserves

Parag A. Pathak, Alex Rees-Jones, Tayfun Sönmez

To cite this article:

Parag A. Pathak, Alex Rees-Jones, Tayfun Sönmez (2023) Reversing Reserves. Management Science

Published online in Articles in Advance 23 Feb 2023

. <https://doi.org/10.1287/mnsc.2023.4669>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2023 The Author(s)

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Reversing Reserves

Parag A. Pathak,^{a,b} Alex Rees-Jones,^{b,c,*} Tayfun Sönmez^d

^aDepartment of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; ^bNational Bureau of Economic Research, Cambridge, Massachusetts 02138; ^cDepartment of Business Economics and Public Policy, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104; ^dDepartment of Economics, Boston College, Chestnut Hill, Massachusetts 02467

*Corresponding author

Contact: ppathak@mit.edu,  <https://orcid.org/0000-0001-8621-3864> (PAP); alre@wharton.upenn.edu,

 <https://orcid.org/0000-0001-5293-2158> (AR-J); sonmezt@bc.edu,  <https://orcid.org/0000-0002-9153-0466> (TS)

Received: November 6, 2020

Revised: December 21, 2021

Accepted: March 14, 2022

Published Online in Articles in Advance:
February 23, 2023

<https://doi.org/10.1287/mnsc.2023.4669>

Copyright: © 2023 The Author(s)

Abstract. Affirmative action policies are often implemented through reserve systems. In this study, we demonstrate that reserve systems face widespread misunderstanding by the public. This misunderstanding can lead individuals to support policies that ineffectively pursue their interests. To establish these claims, we present 1,013 participants in the Understanding America Study with choices between pairs of reserve systems. Participants are members of the group receiving affirmative action and are financially incentivized to choose the system that maximizes their chance of admission. Using this data, we apply a novel approach to identifying the rate of uptake of different decision rules used by participants. We find that participants rarely use a fully optimal decision rule. In contrast, we find that many choices—40% in our primary estimates—are rationalized by a *nearly* correct decision rule, with errors driven solely by failing to appreciate the importance of processing order. Failing to account for processing order causes individuals to fail to distinguish between two policies that achieve different degrees of affirmative action: policies that provide nonbinding minimum guarantees of the number of spaces allocated and policies that provide spaces over-and-above what would be allocated absent a reserve. Confusion about the importance of processing order helps to explain otherwise surprising decisions made in field applications of reserve systems. We discuss implications for managers and policy makers who are trying to implement reserve systems and who are accountable to the public.

History: Accepted by Axel Ockenfels, behavioral economics and decision analysis.



Open Access Statement: This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as “*Management Science*. Copyright © 2023 The Author(s). <https://doi.org/10.1287/mnsc.2023.4669>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>.”

Funding: The authors thank the National Science Foundation and the Wharton Behavioral Laboratory for financial support.

Supplemental Material: The data files and online appendices are available at <https://doi.org/10.1287/mnsc.2023.4669>.

Keywords: affirmative action • reserve systems • experimental economics • behavioral market design

1. Introduction

When setting rules for assessing applicants, managers and policy makers commonly face a balancing act. On the one hand, selections are often made based on well-justified rules for priority. As examples, jobs may be granted based on measures of merit, seats at schools may be granted according to lotteries, and visas may be granted based on the order of applications. On the other hand, both internal and external stakeholders may be concerned about the distribution of characteristics held by successful applicants. Returning to our examples, one might see particular value in hiring from historically underrepresented groups, in admitting students from a local area, or in granting visas to individuals with special qualifications.

Resolving this tension is often a fraught process. In many cases, the decision maker must strike a compromise between constituencies advocating for reliance on the standard measure of priority and constituencies advocating for greater distributional consideration. A common approach to striking this compromise has been the adoption of a *reserve system*. In a reserve system, some of the positions being allocated are reserved for the group targeted for preferential treatment. When reserve slots are processed, the members of the targeted group with the highest priority receive them. When unreserved slots are processed, only priority is considered (regardless of group membership). Because such a system nests both a respect for priority and a way to advance the targeted group, it may be viewed as a tolerable middle ground.

To reach this middle ground, the designer has access to two levers that influence the advantage given to targeted applicants. The first lever is the number of positions reserved. All else equal, a member of the targeted group is better off if more seats are reserved for people like him. The second lever is the order in which reserve positions are processed. As documented in recent papers (Dur et al. 2018, 2020; Pathak et al. 2020), changing processing order dramatically changes the degree of advantage conferred by a fixed number of reserved positions. When reserves are processed first, the number of positions reserved serves as a minimum guarantee. A member of the targeted group who would receive a position based on priority alone counts toward the reserves, and thus reserves become relevant only if fewer than the reserved number of positions would be filled by the targeted group absent reserves. By contrast, when reserves are processed last, the number of positions reserved denotes the number of additional positions to grant the targeted group over and above what would be determined by priority over unreserved positions. This confers advantage to the target group regardless of their initial allocation of unreserved positions and the benefits do not terminate once a target number of positions have been attained.

Our study was motivated by our belief that the critical importance of processing order may be misunderstood. Although the importance of the number of reserved positions is relatively obvious and salient, the importance of processing order can easily be overlooked or dismissed as a technical detail. This misunderstanding can be consequential. Constituencies that do not appreciate the importance of processing order could deploy reserve systems in a manner that blunts the degree of affirmative action achieved by a reserve of a fixed size. Additionally, they may view a proposed reserve system as a fair compromise when it does not significantly advance their goals.

Recent papers have presented suggestive evidence of this type of misunderstanding in several high-profile applications. In one focal example, Boston public schools enacted a policy of reserving 50% of seats for walk-zone applicants explicitly as a compromise between constituencies supporting local school assignment and constituencies supporting unrestricted school choice (Dur et al. 2018). Processing order was not explicitly considered during this compromise and was arbitrarily resolved with reserves processed first—a minimum guarantee. Over a decade later, market designers uncovered how the reserve system had been enacted and revealed that it did very little to advantage the applications of walk-zone students. This led the reserve system to be abandoned because of its unpalatable, but previously unnoticed, lack of compromise and its perceived lack of transparency. In another focal example, the 2004 reform of the U.S. H-1B visa system mandated a policy of reserving visas for advanced degree applicants, but left processing order

unspecified in the legislation. In the years since, theoretically important but unlegislated details have been modified on several occasions for reasons purely of logistical convenience. These reforms led to large changes in the degree of preference given to highly skilled immigrants. However, unlike typical changes in U.S. immigration policy, these changes were typically not publicly noticed or debated and indeed may not have been intended (Pathak et al. 2020). Although the knowledge and motivations of all parties in these examples are not fully documented, we believe that cases like these suggest that misunderstanding of reserve systems may be both common and consequential.¹

Motivated by field applications like these, we sought to design a means to infer the understanding of reserve systems held by the populace. We deployed a preregistered online experiment to 1,013 members of a nationwide survey panel that is approximately representative on a broad range of demographic variables. In this experiment, subjects faced simple scenarios mirroring the two applications of reserve systems discussed previously: allocation of seats at a high school or allocation of work visas. In the scenarios, subjects are members of a group that will have positions reserved. Subjects face financial incentives to maximize the chance that their admission is attained in a simulation. They then choose how they would like the reserve system to be administered, selecting from pairs of policies that differ in the both the *number* of seats reserved and in the *order* that the reserve seats are processed.

Our experiment was designed to reveal the rate at which subjects adopt several competing decision rules. In our empirical model, the population consists of individuals choosing from a rich set of potential *choice functions*. These choice functions dictate which policy the subject prefers given the number of seats assigned to both the “reserves-first” and “reserves-last” policies. If subjects choose optimally, they switch to preferring the reserves-first policy from preferring the reserves-last policy when the number of reserves-first seats surpasses a known threshold. Optimal behavior then leads to a discontinuity in choice probability at that threshold, and the size of the discontinuity identifies the fraction of decisions made using that choice function. Similarly, if subjects understand that more seats are better but treat processing order as irrelevant, they switch to preferring the reserves-first policy from preferring the reserves-last policy when the number of reserves-first seats surpasses the number of reserves-last seats. This naïve behavior then leads to a discontinuity in choice probability at that different threshold, and again the size of the discontinuity identifies the fraction of decisions made using that choice function.

Our results illustrate that the optimal decision rule is rarely applied and that subjects often miss the importance of processing order. Our primary estimates suggest that

3% (standard error (*s.e.*) = 2 percentage points (*pp*)) of decisions are made by applying the optimal decision rule; we are unable to reject the hypothesis that the optimal decision rule is never applied. We are further able to rule out that more than 7% of decisions are made using decision rules that are “close” to optimal. In contrast, we estimate that 40% (*s.e.* = 2*pp*) of decisions are made using a decision rule that responds to reserve size but treats processing order as irrelevant, reflecting widespread coordination on behavior that would be optimal if not for ignoring a single important comparative static. We find evidence consistent with some of the remaining decisions being made with decision rules that treat processing order as relevant but underestimate its effect size, although this behavior is less common than ignoring processing order entirely. The widespread misunderstanding of processing order helps explain the frequency of experimental decisions that are not payoff maximizing for subjects.

This paper builds on a long tradition of using laboratory-experimental methods to test for understanding of matching mechanisms (Chen and Sönmez 2006, Calsamiglia et al. 2010, Echenique et al. 2016, Rees-Jones and Skowronek 2018).² Within this literature, we make two primary contributions.

First, whereas most studies examine the preferences that participants express in a designed market, this study probes participants’ preferences over the design of the market itself. This alternative focus is important because the design of many markets is guided by public approval and subject to public oversight. As a result, misunderstanding harbored by the public can influence a market’s design. We return to a more detailed discussion of when and how the public’s understanding becomes relevant in the conclusion of the paper.

Second, these findings reinforce a growing body of work showing large potential for misunderstanding of matching-market incentives. Although clear and transparent explanation of a matching procedure is often thought to be sufficient for widespread understanding to arise, our results suggest that this is insufficient in reserve systems. These findings mirror similar results showing that misunderstanding of the deferred acceptance or top trading cycles algorithms persist even in settings with substantial training and feedback (Ding and Schotter 2017, Guillen and Hakimov 2018, Rees-Jones and Skowronek 2018). Several forces lead us to worry that eliminating misunderstanding of processing order will be particularly challenging. We document that subjects with higher education, subjects with higher performance on cognitive ability tests external to our survey, and subjects with a higher performance on comprehension tests within our survey all show a greater likelihood of adopting our misguided decision rule of interest. These findings suggest that misunderstanding

of processing order is not simply resolved by greater attention, numeracy, or careful thinking. Training people out of this mistake requires teaching them careful consideration of relatively subtle statistical selection problems—a class of problems that remains challenging even for the highly educated. Relatedly, the individuals who run the market may often not understand the importance of these issues, or, worse yet, may be actively incentivized to foster misunderstanding. In such cases, reliance on the internal provision of training and advice will clearly be insufficient to ensure that the final policy adopted efficiently pursues the goals of the populace adopting it.

In addition to our contributions to the experimental market design literature, we also make methodological contributions to the broader experimental literature concerned with the classification of “behavioral types.” Laboratory or survey experiments are often interested in partitioning experimental participants into groups according to features of the preferences they reveal. This is challenging, because individuals must be classified according to only the decisions available in the experiment—typically too few for individual-level estimation without restrictive assumptions on the nature of measurement error and heterogeneity. The empirical strategy that we describe in Section 3 provides a new approach to type estimation that relies on points of discontinuity in individual choice rules. This approach ultimately allows for estimation of the frequency of adoption of behavioral rules with relatively minimal structure placed on the nature of measurement error or heterogeneity. We believe that this general approach may be useful in other settings.

The paper proceeds as follows. In Section 2, we present a brief review of the theory of reserve systems. In Section 3, we formally present our candidate models of decision rules and our econometric strategy for identifying their rate of adoption. In Sections 4 and 5, we describe the design and deployment of our experiment. In Section 6, we present results. In Section 7, we conclude.

2. Review of the Theory of Reserve Systems

In this section, we briefly present existing theoretical results on the functioning of reserve systems. This summary primarily draws on the work of Kominers and Sönmez (2016) and Dur et al. (2018).

2.1. Decision Environment

Consider a setting in which some number of objects must be allocated. For concreteness, say the objects to be assigned are seats at a school. The school has q seats available. In the absence of reserve considerations, a

mechanism would assign these seats to applicants according to a linear priority order (e.g., outcomes of a standardized test or results from a lottery). However, this school wishes to provide some advantage to a particular group of applicants. Call this group the reserve applicants. Call those outside of this group the general-category applicants.

To help advantage the reserve applicants, the school labels q_r of their q seats as reserved seats (with $0 < q_r < q$). The remaining $q - q_r$ seats are open seats.

To determine the assignment of seats at the school, the school fills seats sequentially and one at a time. When processing an open seat, the school admits the student with the highest priority among all those not yet admitted. Reserve-category status is not considered. When processing a reserved seat, the school admits the reserve applicant with the highest priority among all those not yet admitted. General-category applicants are ineligible for these seats.

To fully specify the assignment procedure, the sole remaining requirement is to describe the processing order for reserved and open seats. Conceptually, any permutation is possible: One could process one reserved seat, followed by seven open seats, followed by two reserved seats, and so on. In practice, however, these systems are commonly administered in one of two configurations: processing all reserve seats either prior to all open seats or after all open seats. We will restrict attention to these two extremal policies.

2.2. Comparative Statics of Interest

In a system like that just specified, two key comparative statics govern the degree of advantage conferred to the reserve group.

Seat-number comparative static: Hold fixed the priority order and the processing order. Increasing the number of reserved seats weakly increases the number of admitted reserve students.

The seat-number comparative static captures an obvious and intuitive determinant of assignments: saving more seats for a group helps the group. A second more subtle comparative static follows from the work in Dur et al. (2018).

Processing-order comparative static: Hold fixed the priority order and the number of reserved seats. Switching from processing the reserved seats first to processing the reserved seats last weakly increases the number of admitted reserve students.

2.3. Potential for Misunderstanding of Processing Order

Existing results from behavioral economics suggest that there are significant psychological hurdles to comprehension of the processing-order comparative static. To summarize these hurdles, we direct attention to two

elements of reasoning that are essential for understanding the importance of processing order.

First, the processing-order comparative static is partially attributable to a selection effect. When reserved seats are processed last, reserve applicants are admitted in the first-stage processing of open seats at a rate determined by their distribution of priorities relative to general-category applicants. Except for differences in priorities, competition for the open seats is effectively a level playing field between the two groups. In contrast, when reserved seats are processed first, the highest-priority members of the reserve group are removed from the applicant pool before the processing of the open seats. The competition for open seats is therefore between all members of the general category and the comparatively low-priority members of the reserve group, tilting admissions in favor of the general-category applicants.

Selection effects like these are known to pose problems to many decision makers. Enke (2020) documents that experimental subjects have a tendency to treat selected samples as if they were representative, with this error leading to failures of economic decision-making in a signal-extraction task. These findings accord with literature in psychology that compares human judgments to those of “naïve intuitive statisticians.” This literature emphasizes humans’ ability to quickly and accurately forecast simple sample properties of presented data (Spencer 1961, 1963) but also points out humans’ systematic tendency to ignore the ways in which the samples they are presented are non-representative (Fiedler and Juslin 2006, Juslin et al. 2007). In short, substantial existing research suggests that decision makers may fail to attend to selection effects, and proper consideration of selection effects is essential to understanding reserve systems.

Second, the processing-order comparative static is partially attributable to a composition effect. To illustrate, when reserve seats are processed last, competition for the open seats is between all general-category applicants and all reserve-category applicants. In contrast, when reserve seats are processed first, competition for open seats is between all general-category applicants and the reserve applicants with q_r group members already removed. In the latter situation, reserve applicants make up a smaller portion of the total applicant pool. As a result, even without selection effects, admissions are again tilted in favor of the general-category applicants.

Ability to appreciate this composition effect would naturally be influenced by base-rate neglect, a foundational bias in probabilistic reasoning in the literature on judgment and decision making.³ Base-rate neglect refers to individuals’ tendency to ignore base rates when forecasting the comparative likelihood of outcomes. An individual affected by this bias would fail to appreciate that the different proportions of reserve-group applicants that enter the second round of processing would affect the likelihood of their admissions.

More broadly, misunderstanding of reserve systems can naturally be understood to arise as a consequence of bounded rationality (Selten 1990, Simon 1990) and rational inattention (Wiederholt 2016). Matching systems may not be intuitive, and full understanding of them may be costly to develop. Because most individuals face reserve systems infrequently, the incentive to invest in understanding them is comparatively low, especially if one's initial intuitive understanding suggests that they already know all that is relevant. The strong intuitive appeal of believing that "more seats are better" could naturally lead a boundedly rational agent to hold incorrect beliefs when first encountering reserve systems. And while incorrect initial beliefs are often assumed to be corrected by repeated experiences and feedback, that disciplining force may be weak in this setting for two reasons. First, if priors that processing order is irrelevant are strong enough, a rationally inattentive agent may fail to mentally catalog processing order as experiences accumulate, thus preventing him from noticing that this factor is relevant (as in Hanna et al. 2014). Second, and more simply, the infrequency with which most individuals face reserve systems may result in too little feedback to correct wrong intuitions.

This past literature guided our belief that misunderstanding of processing order may be widespread, motivating us to design an approach to testing this hypothesis.

3. Identifying Subjects' Understanding of Reserve Systems

In this section, we present our empirical model for inferring understanding of reserve systems. The experiment that we present in the remainder of the paper was tailored for utilization of this empirical model.

3.1. Model of Preferences for Reserve Systems

Consider an individual (i) facing an assignment problem like that described in Section 2. This individual is a member of the group that qualifies for reserve seats. He is presented with two potential policies that could be applied to determine admissions: a "reserves-first" (RF) policy with s^{RF} reserve seats and a "reserves-last" (RL) policy with s^{RL} reserve seats. Beyond seat numbers and processing order, all other features of decision environment are held fixed. The individual's task is to choose between these two policies.

In this environment, the primitives of our model are individuals' *choice functions*, denoted by $C: (s^{RF}, s^{RL}) \rightarrow [0, 1]$. Given an assigned number of reserve seats for both the RF and the RL policies ($(s^{RF}, s^{RL}) \in \mathbb{R}_+^2$), a choice function outputs the individual's probability of indicating a preference for the RF policy. When holding fixed all other elements of the assignment problem, such a function completely characterizes an individual's

observable preferences. At times we will consider a choice function adopted by a specific individual, in which case it will be subscripted by i .

If the choice function were observed, it would provide a direct means of testing an individual's understanding of the theory described in Section 2. For any given number of RL seats, there exists a threshold number of RF seats ($T^*(s^{RL})$) such that the RF policy will be most favorable to the individual if and only if its number of reserve seats exceeds the threshold. An individual who correctly analyzes the environment and chooses the policy in his best interest would therefore adopt the choice function

$$C^*(s^{RF}, s^{RL}) = \begin{cases} 1 & \text{if } s^{RF} > T^*(s^{RL}) \\ 0 & \text{if } s^{RF} \leq T^*(s^{RL}). \end{cases}$$

Adopting this choice function would serve as strong evidence in support of a sophisticated understanding of the decision problem.⁴

Just as observation of the choice function would allow for the identification of sophistication, it is also useful for identification of the type of misunderstanding that we have posited. Consider next the choice function that would be observed among individuals who understand the seat-number comparative static but who are unaware of the processing-order comparative static. Such individuals adopt the choice function

$$C^n(s^{RF}, s^{RL}) = \begin{cases} 1 & \text{if } s^{RF} > s^{RL} \\ 0 & \text{if } s^{RF} \leq s^{RL}. \end{cases}$$

This choice function dictates choosing the policy that offers more seats, regardless of order. The superscript n denotes the fact that this choice function reflects a degree of naïveté in his understanding of incentives.

In our framework, we allow for individuals to probabilistically apply different choice functions at different times. Consider an individual's *average choice function*:

$$\bar{C}_i(s^{RF}, s^{RL}) = p_i^* C^*(s^{RF}, s^{RL}) + p_i^n C^n(s^{RF}, s^{RL}) + \sum_k p_i^k C^k(s^{RF}, s^{RL}).$$

In this equation, the term p_i^* denotes the individual's probability of using the optimal choice function; p_i^n denotes the probability of using the naïve choice function of interest; and the p_i^k terms denote the probabilities of using a set of other arbitrary choice functions. This set of other choice functions is included in the framework for two reasons. First, these other choice functions can capture other heuristics. Second, their inclusion also provides a means of modeling mistakes. For example, an individual who always *tries* to apply the optimal choice rule but periodically fails to apply it correctly could be modeled as having, for example, $p_i^* = 0.9$ with the remaining 10% probability weight

placed on choice function that assigns a 50-50 chance to each choice regardless of the seats assigned. As another example, an individual who attempts to apply the optimal choice rule but assesses the optimal threshold $T^*(s^{RL})$ with error could be modeled with a choice function that replaces the discontinuity in choice probabilities at $T^*(s^{RL})$ with a smooth transition in choice probabilities occurring in the vicinity of $T^*(s^{RL})$. Because of the inclusion of these alternative choice functions, the interpretation of p_i^* and p_i^n is the probability that the subject applies the *exact* choice function of interest, as opposed to the choice function with standard notions of error allowed. However, in our empirical analysis, we will additionally examine the rate of adoption of choice functions with discontinuities “close” to the optimal threshold.

This framework for modeling individual decisions is extremely general. This generality comes at a cost. Estimating the parameters of average choice functions at the individual level would require having individual subjects complete a very large number of scenarios. Because attention and response quality decline precipitously as subjects are repeatedly asked minor variants of the same question, we believe that this approach is infeasible. This consideration leads us to formulate our approach to testing based on the aggregate choice function that would arise from a potentially heterogeneous population of individuals making these decisions. This modeling decision allows us to derive tests that require a large sample size achieved across subjects rather than within subject.

Denote the aggregate choice function, $\mathcal{C} : (s^{RF}, s^{RL}) \rightarrow [0, 1]$, as

$$\begin{aligned} \mathcal{C}(s^{RF}, s^{RL}) &= \mathbb{E}[\bar{C}_i(s^{RF}, s^{RL}) | s^{RF}, s^{RL}] \\ &= \mathbb{E}[p_i^*]C^*(s^{RF}, s^{RL}) + \mathbb{E}[p_i^n]C^n(s^{RF}, s^{RL}) \\ &\quad + \sum_k \mathbb{E}[p_i^k]C^k(s^{RF}, s^{RL}). \end{aligned} \quad (1)$$

In this equation, \mathbb{E} is used to denote the expectation taken over all individuals i , with individuals applying potentially heterogeneous average choice functions \bar{C}_i . In this formulation, the relative weight placed on each choice function is its average rate of use in the population.

For the interested reader, Online Appendix D presents a simple numerical example of the application of this approach and the interpretation of parameters in the aggregate choice function.

3.2. Approach to Estimating Rate of Choice-Function Adoption

The formulation of the aggregate choice function permits a regression-discontinuity based approach to measuring the rate of use of our choice functions of primary interest.

Under the additional assumption that all auxiliary choice functions are continuous in the neighborhood of the sets of (s^{RF}, s^{RL}) values satisfying $s^{RF} = s^{RL}$ and $s^{RF} = T^*(s^{RL})$, these average rates may be isolated through the following relationships:

$$\lim_{\delta \rightarrow 0} \mathcal{C}(T^*(s^{RL}) + \delta, s^{RL}) - \mathcal{C}(T^*(s^{RL}) - \delta, s^{RL}) = \mathbb{E}[p_i^*], \quad (2)$$

$$\lim_{\delta \rightarrow 0} \mathcal{C}(s^{RL} + \delta, s^{RL}) - \mathcal{C}(s^{RL} - \delta, s^{RL}) = \mathbb{E}[p_i^n]. \quad (3)$$

To help understand these equations, consider the case where we hold s^{RL} constant and vary s^{RF} . As s^{RF} crosses the threshold $T^*(s^{RL})$, the optimal choice function dictates that the probability of choosing the reserves-first policy changes discontinuously from zero to one. For the naive choice function and all auxiliary choice functions (because of the previous continuity assumption), no such discontinuity exists. Thus, any discontinuity observed at this point may be attributed to the rate of use of the optimal choice function. Furthermore, the magnitude of the discontinuity will simply be the predicted change in choice probability (known to be one) multiplied by the rate of use of the optimal choice function ($\mathbb{E}[p_i^*]$). This explains the reasoning behind Equation (2); Equation (3) holds by an analogous argument applied at the point where s^{RF} crosses the threshold s^{RL} .

These equations imply that the average rate of use of these choice functions may be estimated by standard regression-discontinuity techniques applied at the two thresholds of interest. We designed our experiment to apply this empirical strategy.

4. Experimental Design

In this section, we present the details of our experiment. Complete text of the experiment, along with details of all data collected, are available in the Understanding America Study (UAS) Experimental Codebook.⁵

4.1. Overview of Design

The primary purpose of our experiment is to present subjects with incentivized scenarios posing choices between RF and RL policies. In these scenarios, subjects are presented with either a high-school admissions problem or a work-visa allocation problem. Seats are assigned based on a randomly generated priority, but with some number of seats set aside for the reserve group. The subjects know they are members of the reserve group, and are given a series of choices between an RF and an RL policy with varying reserves. One of their choices is used to determine the final policy that is applied, and if the subject is allocated a school seat or visa as a result of this policy they are given a \$5 bonus payment.

These data allow us to examine the probabilities of choosing the RF policy across a range of (s^{RF}, s^{RL})

values, thus allowing us to deploy the empirical strategy described in Section 3.

On average, our study took eight minutes to complete. Subjects received a baseline payment of \$5 and an average bonus of \$3.91.

4.2. Walk-Through of Survey Content

To illustrate the nature of our experimental task, we present the text associated with the school-choice version of our experimental protocol. The visa version of this protocol is similar, with differences primarily comprised of replacing references to “students” with references to “workers” and references to “seats at a school” with references to “work visas.”

The study began with an overview:

In this study, we are interested in understanding how you think about school admissions policies. Your bonus payment for taking this study will be affected by a simulation of such policies. You will have the opportunity to choose some features of the policy.

Followed by a further elaboration:

To begin, we will explain the type of school admissions policies we will be considering.

Imagine you are applying for a position at an elite high school. Only 100 students will be admitted. The school considers two factors when deciding whom to admit. First, it considers a randomly generated lottery number. Second, it considers group composition.

There are two groups of people, the Blue students and the Green students. Due to their historical underrepresentation, the school particularly values admitting Blue students.

As is illustrated by this text, “Blue” and “Green” labeling dictated group membership. We chose to avoid the use of more standard racial, gender-based, or income-based group definitions to avoid inviting the subject to rely on beliefs about the desirability of affirmative action for these groups. Although the two groups are always labeled Blue and Green, we randomly assign which of these groups is chosen to be favored.⁶

This introduction was followed by an initial presentation of possible reserve policies:

In order to meet its goal of admitting Blue students, the school is considering two policies. In this example, both policies will involve reserving 30 seats for the Blue students. When applying either policy, students will be admitted one at a time.

Admissions will happen in two stages.

In one stage, seats are available to both Blue and Green students. When each seat is assigned, it will be given to the student with the highest lottery number who has not yet been admitted. Color will not be considered.

In the other stage, seats are reserved for Blue students only. When each seat is assigned, it will be given to the Blue student with the highest lottery number who has not yet been admitted.

The policies that the school is considering differ in the order of these stages.

Policy 1: Save the last 30 seats for the Blue students.

- Stage 1: The first 70 seats will be assigned to the 70 students who have the highest lottery numbers, regardless of color.

- Stage 2: The remaining 30 seats will be assigned to the 30 Blue students who have the highest lottery numbers of all Blue students not yet admitted.

Policy 2: Save the first 30 seats for the Blue students.

- Stage 1: The first 30 seats will be assigned to the 30 Blue students who have the highest lottery numbers.

- Stage 2: The remaining 70 seats will be assigned to the 70 students who have the highest lottery numbers of all students not yet admitted, regardless of color.

The assignment of the RF and RL policies to policy 1 and policy 2 was randomized at the subject level. After the initial randomization, these policy number assignments remained constant throughout the survey.

To test for understanding of the policies presented, this screen contained four comprehension-check questions following the previous text. Across these four questions, the subject was asked to consider several students and determine who among them would be selected for the first seats assigned by policy 1 and 2 and the last seats assigned by policy 1 and 2. To motivate careful thought, a \$1 reward was given if all four comprehension-check questions were answered correctly. After answers were submitted, a feedback screen reported the correct answer for each question and highlighted where mistakes were made.

At this stage, subjects were introduced to our primary experimental task:

To better understand how you think about these policies, we will now present you with a series of choices. Your choices will affect the bonus you earn in this study.

In each choice, you will face a simulated school admissions process like the one that we have been considering. You must choose between two policies describing different ways of assisting the Blue students. In the simulation, you are one of the Blue students, so you will benefit if you choose the policy that is most favorable for this group.

Across these policies, we will vary both the order in which reserve seats are processed and the number of seats that are reserved.

On the following page:

Simulation Details:

All six of the choices you face will have the same basic set-up.

Consider a setting where 200 students are applying to the school. 100 students are Blue and 100 students are Green. You are one of the Blue students.

As before, only 100 students can be admitted. Admissions decisions are still made based on lottery numbers and on diversity considerations. Lottery numbers will be simulated by assigning each student a random number between 1 and 100. All students' numbers, regardless of color, are randomly drawn from the same uniform distribution, so there are no differences across groups in lottery numbers. If two students have the same lottery number, ties will be broken randomly.

Compensation Details:

One of the six choices you make will be randomly selected to be the choice that "counts." After you answer all six questions, we will reveal the question that "counts" and simulate the admissions decision in the scenario you chose. If you are admitted based on this simulation, an additional \$5 will be added to your bonus.

Since you do not know which of the six choices will be chosen to "count," it is in your best interest to answer all six carefully.

Following these screens, subjects faced six screens presenting choices as described previously. On each screen, subjects must choose one of two policies. Because subjects do not know their lottery number, admission under either policy is probabilistic. Based on financial incentives, subjects should choose the policy most favorable to individuals of their group. Each screen took the following format:

Consider the following two ways in which the school could implement its admissions policy.

Policy 1: Save the last (s^{RL}) seats for the Blue students.

- Stage 1: The first ($100-s^{RL}$) seats will be assigned to the ($100-s^{RL}$) students who have the highest lottery numbers, regardless of color.

- Stage 2: The remaining (s^{RL}) seats will be assigned to the (s^{RL}) Blue students who have the highest lottery numbers of all Blue students not yet admitted.

Policy 2: Save the first (s^{RF}) seats for the Blue students.

- Stage 1: The first (s^{RF}) seats will be assigned to the (s^{RF}) Blue students who have the highest lottery numbers.

- Stage 2: The remaining ($100-s^{RF}$) seats will be assigned to the ($100-s^{RF}$) students who have the highest lottery numbers of all students not yet admitted, regardless of color.

As a Blue student, which policy would you prefer?

In the previous text, items in parentheses are placeholders for the seat numbers that were randomly

simulated—for example, (s^{RF}) could be replaced with 60 and ($100-s^{RF}$) could be replaced with 40.

As described in the prior section, our empirical strategy relies on observing choices between RF and RL policies for a range of (s^{RF}, s^{RL}) tuples. These values were randomly generated for each choice the subject faced. The six decisions presented six values of s^{RL} . These were assigned deterministically but in random order: 40, 44, 48, 52, 56, and 60 seats. For each of these scenarios, the required number of seats needed for the RF policy to be optimal was 70, 72, 74, 76, 78, and 80, respectively. For each s^{RL} value, s^{RF} was uniformly sampled from 13 potential values: $-5, -3, -1, +1, +3,$ or $+5$ seats relative to both the optimal and naïve thresholds, as well as an additional point approximately between the two thresholds. By sampling values in the vicinity of our two thresholds of interest, this design ensures that we are well powered to deploy our proposed regression-discontinuity approach.

Following these choices, one of the six scenarios was randomly selected for simulation as described previously. Their chosen policy was implemented, their admissions decision was simulated as specified, and the results of the simulation and the associated payoffs were announced. The study concluded with a brief elicitation of their degree of interest in the survey and an opportunity for free-response comments on the study.⁷

4.3. Preregistration

Our experiment was preregistered on aspredicted.com. For reference, the preregistration document is included in the online appendix. In this document, we specify our exact hypotheses of interest and the details of our regression discontinuity approach. We also commit to our sample size and exclusion restrictions. Although we will also present some exploratory analyses that were not preregistered, we do not deviate from this preregistration in our presentation of primary results.

5. Experimental Deployment and Sample

We deployed our experiment in the UAS.⁸ The UAS is an online panel of American households. The advantage of this panel is its established infrastructure for reaching a broad group of respondents and its substantial efforts to achieve representative sampling. Additionally, by using this panel we can merge data from many other surveys into our analyses, which gives us access to a variety of demographic variables and external measures of cognitive ability.

Our survey was deployed to the UAS population in December 2019 and January 2020. To achieve our targeted sample size of 1,000 responses, the UAS invited a random subsample of 1,500 respondents from their full panel. The survey was closed shortly after the target

sample size was attained, resulting in 1,013 complete observations and a 67% response rate.

Table 1 summarizes basic demographics of our respondents. As is seen across panels of this table, our sample is demographically diverse. However, because of the selection that occurs in the process of recruitment to online panels, our sample differs from the general U.S. population in several ways. Compared with the general adult population of the United States, members of our sample are somewhat more likely to be female, married, and U.S. citizens. Our sample also skews to be somewhat older and somewhat more likely to be white.

Although there is some evidence of selection on observables influencing the general UAS population, we find little evidence that such effects influence which UAS participants respond to our survey. In the final column of this table, we present formal tests for differences in the

demographic variable across respondents who did and did not participate. Only two of the nine tests conducted reach significance at traditional levels.⁹ First, participants are slightly less likely to be employed (59.2% versus 66.1%; $p = 0.01$), consistent with the possibility that those not working have more time to complete online studies. Second, participants who completed our study have a notably different age distribution. On average, those who completed our survey are 3.79 years older than those who did not ($s.e. = 0.90$; $p = 0.00$).

6. Experimental Results

6.1. Test of Misguided Policy Choices

In this section, we present the preregistered tests of our primary hypothesis: that a substantial fraction of respondents mistakenly believe that processing order does not matter in a reserve system.

Table 1. Demographic Information and Sample Selection

	Survey completion status			
	(1) Complete	(2) Incomplete	(3) All recruits	(4) Test of difference
Basic demographics				
Female	56.2	57.2	56.5	$p = 0.71$
Married	58.4	59.1	58.7	$p = 0.80$
Working	59.2	66.1	61.4	$p = 0.01$
U.S. citizen	97.9	97.9	97.9	$p = 0.98$
Hispanic or Latino	10.8	13.6	11.7	$p = 0.12$
Race				
White only	82.2	77.1	80.5	$p = 0.11$
Black only	9.0	10.1	9.4	
American Indian or Alaska Native only	1.3	2.3	1.6	
Asian only	2.8	2.7	2.7	
Hawaiian/Pacific Islander only	0.5	0.6	0.5	
Multiple races indicated	4.3	7.2	5.2	
Education				
<12th grade	5.4	4.5	5.1	$p = 0.30$
High school graduate	19.3	19.9	19.5	
Some college	20.0	25.1	21.7	
Associate degree	13.6	13.3	13.5	
Bachelor's degree	24.7	21.6	23.7	
Master's degree+	16.9	15.6	16.5	
Household income				
<\$10,000	6.7	7.2	6.9	$p = 0.87$
\$10,000–\$24,999	12.9	13.8	13.2	
\$25,000–\$49,999	21.5	21.9	21.6	
\$50,000–\$74,999	18.2	17.8	18.1	
\$75,000–\$99,999	14.3	12.0	13.6	
≥\$100,000	26.4	27.3	26.7	
Age (yr)				
18–29	10.5	15.7	12.2	$p = 0.00$
30–39	19.3	21.3	19.9	
40–49	17.1	22.5	18.9	
50–59	19.3	17.4	18.6	
60+	33.9	23.1	30.4	

Notes. This table presents summary statistics characterizing the demographic features of our sample. With the exception of p values, all numbers presented are the percentage of respondents with a given row's classification. The first group of rows, labeled "Basic demographics," characterizes a series of binary demographic variables. The groups of rows that follow present tabulations of individual categorical variables. The first column presents results for subjects included in our primary analyses. To help assess selection into our study, the second and third columns present results for the subjects who were contacted but did not complete the study and all contacted subjects, respectively. The final column provides p values for chi-squared tests for differences in the distribution of the categorical variable by survey completion status.

To test this hypothesis, we estimate models of the form

$$Y_{ij} = \alpha + \beta^n N_{ij} + \beta^* O_{ij} + f(s_{ij}^{RF}, s_{ij}^{RL}) + \epsilon_{ij}. \quad (4)$$

Subscripts i and j index the respondent and choice number, respectively. In this model, the dependent variable Y_{ij} is an indicator for whether the RF policy was chosen by individual i in a given binary choice j . Variables N_{ij} and O_{ij} provide the value of Y_{ij} dictated by the naïve or optimal choice function. Formally, $N_{ij} = I(s_{ij}^{RF} > s_{ij}^{RL})$ and $O_{ij} = I(s_{ij}^{RF} > T^*(s_{ij}^{RL}))$, where $I()$ denotes the indicator function taking the value of 1 when the statement in parentheses is true; $(s_{ij}^{RF}, s_{ij}^{RL})$ denotes the number of seats assigned to each policy, as before; and $f(s_{ij}^{RF}, s_{ij}^{RL})$ denotes a function meant to control for the number of each type of seats assigned. Across specifications, we will consider a variety of approaches to handling this control, including modeling f as a local polynomial, a cubic spline, or a fifth-order polynomial.¹⁰

Interpreted in light of our model from Section 3, β^n serves as an estimate of $\mathbb{E}[p_i^n]$ and β^* serves as an estimate of $\mathbb{E}[p_i^*]$. Despite this interpretation, the model in Equation (4) does not constrain the sign of β^n or β^* to be positive. In principle, this means that these estimates could yield invalid probabilities. We would interpret the detection of a (statistically significant) negative value for these parameters as a rejection of our framework for type estimation. The flexible term $f(s_{ij}^{RF}, s_{ij}^{RL})$ is interpreted as a fit of all auxiliary choice functions that are used beyond our two focal choice functions of interest (i.e., as an approximation of term $\sum_k \mathbb{E}[p_i^k] C^k(s_{ij}^{RF}, s_{ij}^{RL})$ in Equation (1)).

Table 2 presents our estimates of this model. In columns 1 and 2, we report estimates of this model with the data restricted to s_{ij}^{RF} values that are within five seats of the two thresholds. This amounts to a simple difference in means of the rate of choosing the RF policy when s_{ij}^{RF} is immediately above versus immediately below each threshold. Formally, no term controlling for $f(s_{ij}^{RF}, s_{ij}^{RL})$ is included in the regression; instead, the influence of this term is assumed to be nearly constant for a sufficiently narrow region of s_{ij}^{RF} values, and the estimation sample is correspondingly restricted to a narrow region near the threshold.

Interpreting the results from column 1, we see that on average, the RF policy is 40 percentage points ($s.e. = 2pp$) more likely to be chosen when the number of RF seats is just above (versus just below) the number of RL policy seats. This finding is consistent with respondents using the naïve choice function for 40% of decisions.

In contrast, column 2 demonstrates that on average, the RF policy is only 3 percentage points ($s.e. = 2pp$) more likely to be chosen when the number of RF seats is just above (versus just below) the threshold from the optimal decision function. This coefficient is statistically distinguishable from zero ($p = 0.03$), but quantitatively suggests that effectively no respondents apply the optimal choice function.

In the remaining columns of the table, we report estimates with different methods for controlling for $f(s_{ij}^{RF}, s_{ij}^{RL})$. All approaches provide similar results. Varying our approach to controlling for $f(s_{ij}^{RF}, s_{ij}^{RL})$ with a local polynomial, a spline, or a high-order polynomial, our estimates of

Table 2. Estimates of Choice Functions Governing Policy Preferences

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\beta^n: N_{ij}$	0.40 (0.02)		0.39 (0.04)		0.36 (0.02)	0.37 (0.04)	0.36 (0.02)	0.36 (0.04)
$\beta^*: O_{ij}$		0.03 (0.02)		0.02 (0.03)	-0.01 (0.02)	0.03 (0.03)	-0.01 (0.02)	0.03 (0.03)
Control for s^{RF} (f)	Sample restriction		Local polynomial		Cubic spline		Fifth-order polynomial	
s^{RL} fixed effects	No		No		Yes		Yes	
s^{RL} fixed effects $\times f$	No	No	No	No	No	Yes	No	Yes
Respondents	990	991	1,013	1,013	1,013	1,013	1,013	1,013
N	2,865	2,709	6,078	6,078	6,078	6,078	6,078	6,078
R^2	0.163	0.002			0.198	0.200	0.198	0.200

Notes. This table reports regressions of an indicator for choosing the RF policy on controls for the number of seats reserved. Variables N_{ij} and O_{ij} provide the value that Y should take if the respondent adopts the naïve or optimal choice function defined in Section 2. Across columns, we present a variety of approaches to estimating the model $Y_{ij} = \alpha + \beta^n N_{ij} + \beta^* O_{ij} + f(s_{ij}^{RF}, s_{ij}^{RL}) + \epsilon_{ij}$, varying the means of controlling for the number of seats assigned to the RF and RL policies through term $f(s_{ij}^{RF}, s_{ij}^{RL})$. In columns 1 and 2, we attempt to control for this term by assuming that it is approximately constant in a small enough window. Each column restricts the data to observations in which the number of seats in the RF policy is within five seats of the relevant threshold. In columns 3 and 4, we present estimates arising from local polynomial regressions, applying a rectangular kernel with a bandwidth of 3. In column 5, $f(s_{ij}^{RF}, s_{ij}^{RL})$ is approximated with a cubic spline over s_{ij}^{RF} combined with fixed effects for the six possible values of s_{ij}^{RL} . In column 6, the spline is interacted with the fixed effects, effectively allowing for s_{ij}^{RL} -value-specific splines over s_{ij}^{RF} . Columns 7 and 8 follow the same format as 5 and 6, replacing the splines with fifth-order polynomials. Standard errors, clustered by respondent, are reported in parentheses.

the rate of utilization of the naïve choice functions range from 36% to 37%. Across these specifications, the estimated rate of utilization of the optimal choice function never exceeds 3% and is generally statistically indistinguishable from zero.

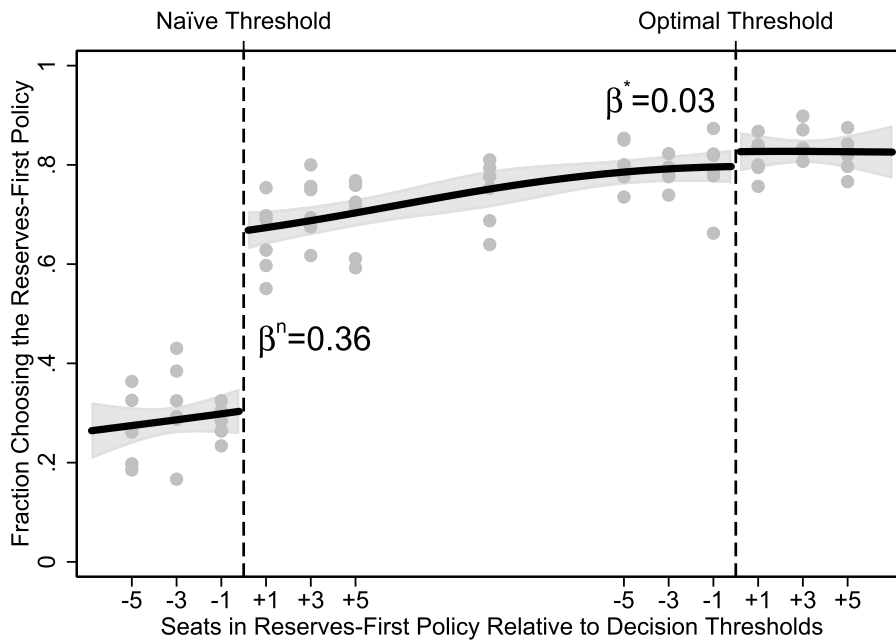
Figure 1 helps in visualizing these results. For a fixed number of RL seats, the number of RF seats takes values of $-5, -3, -1, +1, +3,$ or $+5$ seats relative to each of the thresholds of interest. One additional point was sampled between the two thresholds. In this figure, each dot illustrates the average rate of choosing the RF policy for the number RF seats illustrated on the x axis, with the six dots above each point summarizing choices under the six RL seat amounts. The solid line presents a fitted spline analogous to that in column 5 of Table 2. This figure illustrates a stark change in the rate of choosing RF at the naïve threshold of interest. In contrast, there is no apparent discontinuity at the threshold where it should occur among optimizing agents.

Figure 1 helps to assess the rate of several other choice-functions of interest. First, consider individuals who use choice functions “near” our two choice functions of interest (e.g., attempting to apply the optimal

choice function but assessing its threshold with error). Subjects using these choice functions would change their choices in the vicinity of either threshold. However, the aggregate choice function is estimated to be relatively flat near each discontinuity. Focusing specifically on the vicinity of the optimal threshold, the 95% confidence interval allows us to reject the hypothesis that the probability of choosing the RF policy increased by more than 7 percentage points in the range from five seats below the optimal threshold to five seats above the optimal threshold. This allows us to statistically reject that choice functions applying a “nearly optimal” threshold (defined to be within five seats of the optimal threshold) were applied in more than 7% of decisions.

Next consider individuals who understand that processing order matters but underestimate its quantitative effect. Subjects using these choice functions would change their choices between the two thresholds. Examining Figure 1, the rate of choosing the reserves-first policy ranges from 65% immediately over the naïve threshold to 79% immediately before the optimal threshold, a change of 14 percentage points ($s.e. = 3pp$). Of course, other choice functions could result in changing behavior

Figure 1. Regression Discontinuity Estimates



Notes. This figure illustrates the discontinuities in choice probabilities that occur at the thresholds of interest. In our experiment, subjects faced six scenarios containing choices between reserves-first and reserves-last policies. The scenarios always contained the same six reserves-last policies. In each scenario, the number of seats in the reserves-first policy was randomly drawn from 13 values spanning the x axis, defined by their position relative to two thresholds. Vertical dashed lines demarcate these thresholds: the point where the number of reserves-first seats comes to exceed the number of reserves-last seats (the naïve threshold), and the point where the number of reserves-first seats comes to exceed the amount needed to make choosing the reserves-first policy optimal (the optimal threshold). The six dots above each point on the x axis illustrate the average rate of choosing the reserves-first policy across the six reserves-last seat numbers. As seen in this figure, subjects’ average propensity to choose the reserves-first policy increases substantially when the naïve threshold is exceeded, but does not change substantially when the optimal threshold is exceeded. The plotted line is a fitted cubic spline over these points with its associated 95% confidence interval. Reported in the figure are the formal estimates of the discontinuity at these two points arising from this spline, which closely matches the results from Table 2.

in this region: this estimate does not identify the rate of choice function adoption as clearly as the discontinuity-based approaches described above. However, with the additional assumption that the rate of choosing the reserves-first policy is monotone in the seat difference for all choice functions in use, this provides an upper bound on the rate of underestimation of reserve-order's effect size. The 95% confidence interval allows us to reject that more than 20% of respondents may understand processing order while underestimating its effect size. Although potentially present among a noticeable fraction of respondents, this type of misunderstanding is substantially less prevalent than mistakenly assuming that processing order has no effect.

In summary, we estimate that a large fraction of decisions (40% in our primary regression) were made according to a choice function that reflects an understanding of the seat-number comparative static while reflecting ignorance of the processing-order comparative static. Respondents making these decisions understand that more seats are better but do not see the benefits of the reserves-last design.

6.2. Summary of Additional Results and Robustness Analyses

In Online Appendix F, we document a large battery of additional results, considerations, and robustness analyses. We summarize those findings briefly here, but direct an interested reader to the appendix for full details.

Online Appendix F.1 documents an important correlate of use of the naïve choice function: cognitive ability. Perhaps surprisingly, subjects with higher education, subjects with higher performance on cognitive ability tests external to our survey, and subjects with a higher performance on comprehension tests within our survey all show a greater likelihood of adopting our misguided decision rule of interest. This contrasts with a common finding in the behavioral market design literature that misreaction to matching-mechanisms' incentives is more prevalent among those of lower cognitive ability (Basteck and Mantovani 2018, Rees-Jones 2018, Rees-Jones and Skowronek 2018, Shorrer and Sóvágó 2018, Rees-Jones et al. 2020, Hassidim et al. 2021). In this instance, however, the finding may be rationalized by noting that adoption of this decision rule reflects a general understanding of incentives in this procedure. Our decision rule of interest is *almost* sophisticated, missing one subtle component of large ultimate importance.

Online Appendix F.2 documents two additional experiments supporting the claims presented in this paper. Prior to the deployment of our study, we ran two large-scale pilots on Amazon Mechanical Turk. Both pilots examined the "school choice" version of the study. The first pilot assessed the rate of optimal choice in a single scenario with nonrandomized seat numbers. The second pilot was nearly identical to the study deployed in the UAS,

with the exception of excluding the visa version of the scenarios. Across these two pilots, we find extremely similar qualitative and quantitative results as reported in this paper. Because of the larger incentives offered in our UAS study, along with the comparatively high quality of the UAS panel's recruitment procedures, we believe the results derived from this sample are the most credible. However, we view the fact that closely analogous results are obtained on other platforms reassuring.

The remainder of Online Appendix F provides analyses and discussions that inform robustness considerations. In particular, we document that our estimates only minimally vary by scenario (Section F.3) and only minimally change with the inclusion of survey weights that account for sample selection (Section F.4). We additionally discuss considerations related to the stake size of decisions in our experiment and analogous decisions in the field.

7. Discussion

In this paper, we examined the general understanding of reserve systems held by the U.S. populace. In experimental choices presented to participants in the Understanding America Study, we found that very few choices were guided by a choice function that reflects a fully sophisticated understanding of these systems. In contrast, a plurality of choices—40% in our leading specification—were guided by a *nearly* sophisticated choice function, demonstrating general understanding of the decision environment but misunderstanding of the critical importance of processing order.

Having established that misunderstanding is prevalent, we return to the question of why it matters. As discussed in Section 2, reserve systems are commonly deployed when achieving compromise between competing stakeholders is a first-order design consideration. Because of this use, widespread understanding of these systems has clear importance: The success of a compromise is always in jeopardy if the stakeholders who assess it do not understand it.

Moving beyond the specific setting of reserve systems, we believe that the beliefs and attitudes of the public about market design are more important than is commonly appreciated. To some, the beliefs of the general public about how to design a market may appear irrelevant. Elaborate matching procedures are typically not designed and deployed by arbitrary members of the public, but instead are ideally managed by benevolent dictators who are well versed in the market design literature. Although the understanding of the public would be comparatively unimportant in such a situation, it can become significantly more important when any of the elements of this ideal fail (as they commonly do).

The understanding of the public is relevant when the market organizer is not purely benevolent. Consider a

shrewd market organizer who wishes to convince a targeted group that he is reforming the system for their advancement but who does not sincerely value such a reform. When proposing a new reserve system, this organizer can consider two options: a reserves-last policy with a small fraction of all seats reserved (say, 5%), or a reserves-first policy with a significantly larger fraction of all seats reserved (say, 30%). In many situations, the former policy will be more effective at advancing the reserve group, and indeed the latter policy may have no effect at all. This market organizer would face strong incentives to propose and enact the less effective policy, and then subsequently enjoy the public's misguided appreciation of the proposal stemming from their incorrect belief that reserving 30% of seats always grants more affirmative action than reserving 5% of seats.

The understanding of the public is relevant when the market organizer is not a dictator. In many situations, the individual designing a market is formally tasked with enacting the wishes of his constituency and is incentivized to enact those wishes through the threat of removal. In cases where the workings of a proposed system are not apparent to the public, the public may vote or lobby for ill-designed policies. They may additionally punish a market organizer for enacting objectively desirable policy because of their misunderstandings, for example, by removing them from the decision-making position in an organization or voting them out of office.

The understanding of the public is relevant when the market organizer is not well versed in the market design literature. If, for example, the market organizer does not appreciate the importance of processing order and decides the order arbitrarily, that arbitrary decision could stop the reserve system from achieving its desired or agreed on outcome. If the system is presented transparently, an informed member of the public may notice this error and pursue its correction. If, however, the public harbors the same misunderstanding as the market organizer, correction of the error becomes substantially less likely.

The literature on market design is replete with examples of these types of considerations arising as new markets were adopted. Further research on the role the public plays, and how their potentially imperfect beliefs impact that role, may be important and productive. We believe our study provides a template for pursuing this style of research.

Returning attention specifically to reserve systems, we believe these considerations have been relevant in several of their high-profile applications, including the Boston Public School and H-1B systems that served as our motivating examples. Furthermore, as both formalized matching systems and reserve systems quickly proliferate, we expect the potential for these problems to grow.¹¹ Concretely, organizations are increasingly relying on formal

systems to reserve places for members of underrepresented groups. Our study sheds light on the misunderstandings that a manager may hold when designing and implementing such systems and the misunderstandings that may be active when other members of the organization or the general public assess the wisdom of the manager's approach. Our hope is that our work helps to foster the transparent implementation of reserve systems in such settings.

Acknowledgments

The authors thank Ori Heffetz, Ted O'Donoghue, Germán Reyes, Ran Shorrer, Jason Somerville, and seminar audiences at Cornell University, Pennsylvania State University, and the University of Bonn for helpful conversations. The project described in this paper relies on data from surveys administered by the Understanding America Study (UAS), which is maintained by the Center for Economic and Social Research at the University of Southern California (USC). The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of USC or UAS.

Endnotes

¹ For a more detailed summary of the history of these reserve systems, see Online Appendix C.

² For a recent review of experimental examinations of matching markets, see Hakimov and Kübler (2021). For a recent review of the interaction between market design and behavioral economics, see Chen et al. (2021).

³ Appreciation of base-rate neglect as a systematic phenomenon traces back at least to Kahneman and Tversky (1973). For a recent review of this literature, see Benjamin (2019).

⁴ At the point of indifference ($s^{RF} = T^*(s^{RL})$), any choice probability can be rationally supported. The choice functions written in this section resolve the indeterminacy at the point of indifference arbitrarily. In our experimental design, we intentionally do not present such cases to respondents.

⁵ Available at <https://uasdata.usc.edu/survey/UAS+210>.

⁶ Additionally, we simulated a scenario where priority (i.e., the lottery) is independent of group membership. Although this assumption holds in some applications of reserve systems, in others, the distribution of priority can differ across groups. Cross-group differences in the distribution of priority affects the quantitative benefit of a reserve system. As a result, to the extent that misunderstanding is driven by incorrectly assessing the effect size, our estimates of the rate of misunderstanding could be less externally valid when applying them to situations where such cross-group differences are present. However, when misunderstanding is driven by the incorrect belief that processing order has no effect, external validity concerns regarding the interaction between processing order and group-specific priorities are less relevant.

⁷ The inclusion of these final two questions is standard practice in the Understanding America Study; these questions were not proposed by the researchers.

⁸ For a detailed description of the UAS, see Alattar et al. (2018). In Online Appendix E, we summarize relevant details of the UAS's sample procedure and its advantages for our purposes.

⁹ Relatedly, we find no evidence of differences in the geographic distributions of participants and non-participants (see Online Appendix E.2).

¹⁰ The inclusion of these terms helps to avoid common worries about linear probability models. When they are included, the model allows for a flexible nonlinear relationship between the number of seats and the probability of choosing the RF policy. This approach is less restrictive than assuming a particular functional form of this nonlinear relationship, as in a logit or a probit model, and thus allows us to better approximate a broader class of aggregate choice functions.

¹¹ To illustrate, the need for understanding of reserve systems became significantly more urgent in the course of the COVID-19 pandemic. In light of the unequal impacts of the pandemic, the National Academies of Sciences, Engineering, and Medicine (NASEM) advised that the procedure for vaccine allocation feature a 10% reserve for disadvantaged communities. However, the October 2020 report (National Academies of Sciences, Engineering, and Medicine 2020) proposing this reserve did not describe how it should be processed. This omission is important: based on other design parameters, processing order makes difference between a system that significantly favors the disadvantaged and a system that does not.

References

- Alattar L, Messel M, Rogofsky D (2018) An introduction to the Understanding America Study internet panel. *Soc. Security Bull.* 78(2):13–28.
- Basteck C, Mantovani M (2018) Cognitive ability and games of school choice. *Games Econom. Behav.* 109:156–183.
- Benjamin DJ (2019) Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics: Applications and Foundations*, vol. 2 (Elsevier, New York), 69–186.
- Calsamiglia C, Haeringer G, Klijn F (2010) Constrained school choice: An experimental study. *Amer. Econom. Rev.* 100(4):1860–1874.
- Chen Y, Sönmez T (2006) School choice: An experimental study. *J. Econom. Theory* 127:202–231.
- Chen Y, Cramton P, List JA, Ockenfels A (2021) Market design, human behavior, and management. *Management Sci.* 67(9):5317–5348.
- Ding T, Schotter A (2017) Matching and chatting: An experimental study of the impact of network communication on school-matching mechanisms. *Games Econom. Behav.* 103(C):94–115.
- Dur U, Pathak PA, Sönmez T (2020) Explicit vs. statistical targeting in affirmative action: Theory and evidence from Chicago's Exam Schools. *J. Econom. Theory* 187:104996.
- Dur U, Kominers S, Pathak PA, Sönmez T (2018) Reserve design: Unintended consequences and the demise of Boston's walk zones. *J. Political Econom.* 126(6):2457–2479.
- Echenique F, Wilson A, Yariv L (2016) Clearinghouses for two-sided matching: An experimental study. *Quant. Econom.* 7(2):449–482.
- Enke B (2020) What you see is all there is. *Quart. J. Econom.* 135(3):1363–1398.
- Fiedler K, Juslin P (2006) Taking the interface between mind and environment seriously. *Information Sampling and Adaptive Cognition* (Cambridge University Press, Cambridge, UK), 3–29.
- Guillen P, Hakimov R (2018) The effectiveness of top-down advice in strategy-proof mechanisms: A field experiment. *Eur. Econom. Rev.* 101(C):505–511.
- Hakimov R, Kübler D (2021) Experiments on centralized school choice and college admissions: A survey. *Experiment. Econom.* 24(2):434–488.
- Hanna R, Mullainathan S, Schwartzstein J (2014) Learning through noticing: Theory and evidence from a field experiment. *Quart. J. Econom.* 129(3):1311–1353.
- Hassidim A, Romm A, Shorrer RI (2021) The limits of incentives in economic matching procedures. *Management Sci.* 67(2):951–963.
- Juslin P, Winman A, Hansson P (2007) The naïve intuitive statistician: A naïve sampling model of intuitive confidence intervals. *Psych. Rev.* 114(3):678.
- Kahneman D, Tversky A (1973) On the psychology of prediction. *Psych. Rev.* 80:237–251.
- Kominers SD, Sönmez T (2016) Matching with slot-specific priorities: Theory. *Theoretical Econom.* 11(2):683–710.
- National Academies of Sciences, Engineering, and Medicine (2020) *Framework for Equitable Allocation of COVID-19 Vaccine* (The National Academies Press, Washington, DC).
- Pathak PA, Rees-Jones A, Sönmez T (2020) Immigration lottery design: Engineered and coincidental consequences of H-1B reforms. NBER Working Paper No. 26767, National Bureau of Economic Research, Cambridge, MA.
- Rees-Jones A (2018) Suboptimal behavior in strategy-proof mechanisms: Evidence from the residency match. *Games Econom. Behav.* 108:317–330.
- Rees-Jones A, Skowronek S (2018) An experimental investigation of preference misrepresentation in the residency match. *Proc. National Acad. Sci. USA* 115(45):11471–11476.
- Rees-Jones A, Shorrer RI, Tergiman C (2020) Correlation neglect in student-to-school matching. NBER Working Paper No. 26734, National Bureau of Economic Research, Cambridge, MA.
- Selten R (1990) Bounded rationality. *J. Institute Theoretical Econom.* 146(4):649–658.
- Shorrer RI, Sóvágó S (2018) Obvious mistakes in a strategically simple college admissions environment: Causes and consequences. Preprint, submitted June 28, <https://dx.doi.org/10.2139/ssrn.2993538>.
- Simon HA (1990) Bounded rationality. *Utility and Probability* (Palgrave Macmillan UK), 15–18.
- Spencer J (1961) Estimating averages. *Ergonomics* 4(4):317–328.
- Spencer J (1963) A further study of estimating averages. *Ergonomics* 6(3):255–265.
- Wiederholt M (2016) Rational inattention. *The New Palgrave Dictionary of Economics* (Palgrave Macmillan UK), 1–5.