Research Statement

Peng Shi

My research agenda is to develop rigorous methodologies to optimize the operations of matching markets. Examples of matching markets that I study include allocation systems for public resources, such as seats at public schools, organs from cadaveric donors, and units of subsidized housing. These are referred to as one-sided matching markets, because they match people to objects, and only the people are strategic actors. I also study two-sided marketplaces, which are platforms that help customers match with suitable service providers at suitable prices. Prominent examples include Airbnb, Angi, Care.com, Google Local Services, HomeAdvisor, TaskRabbit, Thumbtack, Upwork, and Yelp. I am motivated to pursue this research agenda because these markets are important to society, and because I believe there is a tremendous opportunity to improve these markets by developing new mathematical models and methods.

My approach to conducting research is to first learn everything I can about the context by talking to practitioners and by reading papers, news articles, forums, and anything else I can find. I seek to identify underlying issues that are not specific to one market but are relevant to many markets, and build a rigorous model to capture the key trade-offs as elegantly as I can. Below, I describe my representative research projects, and highlight three thematic contributions. Each of the three contributions is presented as a separate section. In Section 4, I describe my personal contribution in each of the co-authored papers. Section 5 outlines directions for future work.

1 Designing Choice Sets and Priorities in School Choice

In many places, students are assigned to public schools not only based on their home address, but can indicate their preferences within a certain *choice set* of eligible schools. The question is how to design an allocation mechanism that gives students their top choices as much as possible while guaranteeing a certain notion of fairness. A common way to model fairness is whether the allocation mechanism respects certain *priorities* that schools give to students, so that whenever a school admits a student with a certain priority, it also grants access to all students with higher priorities. A mechanism that guarantees this notion of fairness is the student-proposing deferred acceptance (DA) mechanism, which is currently implemented for public school choice in Boston, New York City, Chicago, Denver, New Orleans, Washington DC, among other cities. While there is a rich and extensive market design literature on school choice, all previous works treat the choice sets and priorities as exogenously determined, whereas in practice they are chosen by policy makers and therefore can be optimized. I contribute to the literature by developing a tractable methodology for optimizing these policy levers in large matching markets.

In **"Guiding School-Choice Reform through Novel Application of Operations Research"** (*Interfaces*, 2015), I showcase how choice sets and priorities can be optimized in practice. Since 1988, the choice sets of elementary school applicants in Boston have been based on partitioning the city into three geographic zones. In 2012-2013, policy makers in Boston sought to modify these choice sets so as to improve equity of access to quality, reduce school busing costs, and increase community cohesion. After attending dozens of community meetings and talking to parents and policy makers, I learned about what the various stakeholders are looking for, and gained sufficient trust for the city to give me data to analyze the various proposed plans. Using past preference data, I estimated a multinomial logic (MNL) discrete choice model of how families ranked schools, and used it to simulate outcomes of interest under the newly proposed choice sets. This simulation model was used extensively by the city in their deliberations. Moreover, one of the plans that I proposed, later referred to as the "Home-Based Plan," was chosen by the city and implemented in 2014 for assigning all public elementary schools in Boston.

A concern with the above simulation model is that if choices are affected by behavioral issues such as framing more than by underlying preferences, then a discrete choice model that fits past data might not be able to predict outcomes after a major policy reform that also changes the framing of choices. In **"How Well Do Structural Demand Models Work? Counterfactual Predictions in School Choice"** (*Journal of Econometrics*, 2021), my co-author Parag Pathak and I validate the simulation model by comparing its predictions with the actual outcomes of the 2014 reform. One of the findings is that after controlling for changes in the demographics of applicants, the MNL model based on pre-reform data performed almost as well as the MNL model refitted on post-reform data. This provides empirical support for the use of such models to guide policy reforms, as we show that preferences are stable, can be estimated accurately, and can be used for extrapolating to different environments.

While the above simulation model can be used to evaluate a new assignment plan, one cannot use it to optimize over all possible choice sets and priorities, as the number of possibilities is enormous. I develop a practical methodology for solving this optimization problem in a series of two papers. In **"Optimal Allocation without Money: an Engineering Approach"** (*Management Science*, 2015), co-authored with Itai Ashlagi, we prove that if the market is large, then the outcome of any assignment mechanism that satisfies certain natural axioms can be replicated using the DA mechanism with suitably chosen choice sets and priorities. In other words, one cannot gain much by looking into assignment mechanisms beyond DA. Moreover, we show that if student preferences are drawn according to a MNL model, the optimal choice sets and priority distributions can be computed by solving a convex optimization problem, which for the real data set from Boston can be numerically solved within a few minutes.

However, the above methodology could not handle a constraint important to policy makers during the Boston reform, which is to limit the areas schools need to cover to pick up children. In **"Optimal Priority-Based Allocation Mechanisms"** (*Management Science*, 2021), I show that the general problem of optimizing choice sets and priority distributions can be reduced to solving a certain socially-optimal assortment planning problem. This problem is closely connected to the well-studied revenue-maximizing assortment planning problem but with a different objective: instead of finding a set of items to offer so as to yield the maximum expected revenue, the goal is to find a set of schools to offer so as the maximize students' expected utilities, while accounting for the negative externality students impose on others from occupying limited resources. This connection allows me to derive efficient algorithms to solve the school choice problem under a much richer set of utility models and constraints. In particular, the new methodology allows me to account for the average area schools need to cover to pick up children, and it yields an optimized plan for Boston that simultaneously improves upon the Home-Based Plan in match quality, equity, predictability, community cohesion and transportation savings.

Beyond optimization choice sets and priorities, I show that community cohesion can be further improved. In **"Improving Community Cohesion in School Choice via Correlated-Lottery Implementation"** (*Operations Research*, 2014), co-authored with Itai Ashlagi, we show that one can significantly increase neighbors' chances of going to school together without affecting anyone's probability of being assigned to any school. We do this by rewriting the original assignment probabilities as a convex-combination of deterministic assignments, each of which seeks to assign students from the same neighborhood to the same school. For the Boston data set, we find that correlated-lottery alone improves community cohesion more than any of the reforms considered by policy makers during the 2012-2013 reform.

2 Analyzing Dynamic Allocation Systems while Accounting for Rich Preference Heterogeneity

Two other important applications of one-sided matching are the allocation of subsidized housing and of cadaveric organs. These allocation systems are difficult to analyze because they are dynamic, as both agents and objects are arriving over time and each agent faces a trade-off between seeking to be matched now and waiting for a better match in the future. As a result of these complexities, previous theoretical papers on the design of such systems require restrictive assumptions on the preferences of agents and on the scope of possible mechanisms. My contribution is in developing tractable theoretical frameworks that can handle rich preference heterogeneity and a large space of possible mechanisms.

In **"Design of Lotteries and Wait-lists for Affordable Housing Allocation"** (*Management Science*, 2020), my co-author Nick Arnosti and I analyze a model in which apartment units and applicants arrive each period, and applicants are heterogeneous in their preferences for apartments and in their outside options. Based on the assignment mechanism and the aggregate behavior of other applicants, each applicant solves a dynamic optimization problem of whether to go for the apartment which arrived this period, to wait for a better unit in the future, or to leave the system altogether to avoid incurring participation costs. We find that diverse looking mechanisms can be equivalent in their equilibrium outcomes. In particular, the independent lotteries used in New York City (NYC) for allocating affordable housing is equivalent to the wait-list without choice used in Providence RI for public housing. The wait-list with choice used in Amsterdam is equivalent to an allocation system based on virtual currency. The paper also argues that in NYC, the simple change of restricting lottery entries would lead to a more efficient allocation of scarce housing resources.

In **"Eliminating Waste in Cadaveric Organ Allocation"** (Working paper), my student Junxiong Yin and I explore how to reduce the wastage of cadaver kidneys. In the US, more than 100,000 patients are currently on the wait-list for kidneys from deceased donors, while only 25,000 transplants are happening each year; about 5,000 patients die each year while waiting. Despite the scarcity of kidneys, around 20% of successfully procured kidneys are being discarded. Much of this apparent paradox is due to patients turning down offers of older kidneys to wait for younger ones, even though the older kidneys are medically tenable and superior to dialysis. We show that much of this wastage can be eliminated if patients upon entering the system receive a randomized boost to their priority on the wait-list. This randomization acts as a coordination device, so that many patients would only wait for younger kidneys if they were to receive the priority boost, otherwise they would settle for the older kidneys.

3 Optimal Matchmaking in Two-Sided Marketplaces

Two-sided marketplaces benefit society by helping customers more easily match with suitable service providers at suitable prices. A major obstacle to unleashing this benefit is market congestion, as when there are many potential partners and competitors, market participants may need to expend much time and effort to form suitable matches. The literature on platform design has shown using field experiments and game-theoretic analyses that certain changes to platform design can significantly improve the efficiency of matchmaking. However, it is unclear how much additional gains are possible beyond the few design changes that have been explored. My contribution is in developing a rigorous theoretical framework for optimal matchmaking, in which the space of platform designs being considered is much larger than those in previous studies, and there is a well-defined notion of optimality that is tractable to analyze and can be achieved by realistic designs.

In **"Clearing Matching Markets Efficiently: Informative Signals and Match Recommendations"** (*Management Science*, 2020), co-authored with Itai Ashlagi, Mark Braverman and Yash Kaonoia, we consider an incomplete-information variant of the stable marriage model of Gale and Shapley and analyze the minimum amount of information that agents need to communicate in order for the market to reach a stable matching. (In such markets, a matching is said to be stable if no pair of agents have the incentive to leave the platform and match on their own; this represents the equilibrium market outcome when there is no congestion.) If a platform design can facilitate a stable matching using close to the minimum amount of communication, then it is considered to be near-optimal. We find that a near-optimal platform design exhibits the following features: 1) help agents estimate their chances of obtaining a particular match; 2) allow both sides of the market to search for partners and signal interest; 3) utilize a match recommendation engine that steers agents toward partners who have a significant chance of reciprocating interest.

A limitation of the above paper is that it does not allow service providers to set their own prices endogenously, which is an important feature of many markets of interest. Another limitation is that an exactly stable matching may be an overly stringent representation of the ideal market outcome, since in reality market participants stop searching when they have found a good enough match. In **"Optimal Matchmaking Strategy in Two-Sided Marketplaces"** (minor revision at *Management Science*), I overcome these limitations by modifying the above framework to consider an approximate notion of stability as well as a model of matching with endogenous prices. I find that the key determinant of optimal platform design is the degree that agent preferences are easy to describe or satisfy: When provider preferences are easy to describe or satisfy, having only customers search and initiate contact is sufficient for near-optimal matchmaking. Conversely, when customer preferences are easy to describe or satisfy, it is near-optimal to have providers search and initiate contact. When neither sides' preferences are easy to describe or satisfy, a near-optimal design requires both sides to be able to search for partners and initiate contact. When both sides' preferences are easy to describe or satisfy, it is unnecessary to allow agents to search over all possible partners, and a near-optimal design is have the platform recommend a limited number of service providers to each customer.

In **"Optimal Match Recommendations in Two-Sided Marketplaces with Endogenous Prices"** (Working Paper), I develop a prescriptive model of how platforms should make match recommendations to maximize market surplus, assuming that there is a constraint on how many recommendations can be given to each customer. Such constraints ensure that the platform does not overwhelm customers with too many options to evaluate and compare. I find that if the platform adopts a naive policy of recommending to each customer the best matches based on observable characteristics, then it would create an inefficient bottleneck in which a disproportionately large number of recommendations are channeled to relatively few providers, who would profit by either raising prices or only serving customers whose jobs require low costs. This situation is bad for market as a whole. In contrast, an optimal match recommendation policy exhibits two features: 1) pre-emptively account for the potential behavioral responses of providers in terms of raising prices or increasing selectivity; 2) utilize randomization to distribute match recommendations among a broad set of providers.

4 Future Directions

After completing the working papers outlined above, I plan to pursue the following two research directions in the next few years:

1. I plan to explore other aspects of the wastage problem of cadaver organs. For example, based on conversations with doctors, my student Junxiong and I learned that another cause of organ wastage is the miscoordination between who is offered a transplant and whom hospitals make ready for transplantation. To be ready to receive a transplant, a patient needs to undergo frequent health check ups, but due to limited capacities, hospitals can only make ready a subset of patients. We hypothesize that periodically raising and lowering the allocation priorities of patients can reduce this miscoordination, as it helps hospitals to make ready patients who are more likely to be offered a transplant, while maintaining fairness.

2. I plan to collaborate with various online platforms to operationalize the insights from my papers on optimal matchmaking. The theory requires being able to estimate the distribution of preferences of various agent types, and it is an open question how to obtain these estimates from available data. Furthermore, real-world platforms have business constraints and objectives that are not captured in the theoretical models, and it will be useful to develop an optimization model that the platforms can directly use to inform day-to-day operations.

In the long term, I aspire to become known for developing rigorous theoretical models for matching markets that yield novel insights, as well as for making an impact on the operations of real markets.