Essays on Microeconomics under Non-Transferable Utility

Inauguraldissertation

zur Erlangung des Grades eines Doktors
der Wirtschaftswissenschaften

durch

die Rechts- und Staatswissenschaftliche Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Günnur Ege Bilgin

aus Izmir, Türkei

2024

Dekan:	Prof. Dr. Jürgen von Hagen
Erstreferent:	Prof. Dr. Stephan Lauermann
Zweitreferent:	Prof. Dr. Daniel Krähmer
Tag der mündlichen Prüfung:	17. July 2024

Acknowledgements

I am deeply indebted to my supervisors, Stephan Lauermann and Daniel Krähmer, whose unwavering support and guidance have been instrumental throughout my doctoral journey. Stephan Lauermann's intuitive ability to discern my intentions even before they crystallize in my own mind, unwavering support, and commitment to excellence have been a constant source of inspiration. Likewise, Daniel Krähmer's steadfast encouragement and expertise in microeconomics have broadened my horizons and enriched my academic experience. Their patience, dedication, and insistence on scientific rigor have shaped not only my research but also my approach to academia.

I would also like to express my deep gratitude to the exceptional team of researchers at the University of Bonn Economics Department, particularly the Institute for Microeconomics, who have played a pivotal role in my academic development. Benny Moldovanu's insightful perspectives, which seamlessly connect various strands of literature, have been particularly enlightening. Sven Rady's exacting standards, kindness, and exceptional teaching during the first year have left an indelible imprint on my academic path. Furthermore, the accessibility and encouragement of Francesc Dilmé, Jonas von Wangenheim, and Sarah Auster have been invaluable. Their feedback and guidance have played a pivotal role in refining my ideas and directing my research trajectory.

Special thanks are due to my former supervisor and current coauthor, Orhan Aygün, whose mentorship and collaboration have been integral to my academic journey since my master's studies at Bogazici University. He served as the first stepping stone, introducing me to the world of matching and market design, making my academic journey feasible. Working with him has consistently been an enriching adventure. His guidance and the opportunities he provided have laid the foundation for my research endeavors.

I am immensely grateful to Mehmet Ekmekci, my host at Boston College, for facilitating my integration into another esteemed academic community and his invaluable insights, which have profoundly enriched my academic journey. My gratitude to Tayfun Sönmez and Utku Ünver is immeasurable for the stimulating discussions that inspired and challenged me. I also extend my heartfelt appreciation to Bumin Yenmez for his dedicated assistance in advancing my research progress. Pietro Dall'Ara's camaraderie, accessibility, and profound understanding of the literature have been invaluable assets that have greatly enhanced my academic experience.

My heartfelt appreciation also goes out to my fellow Ph.D. students and Stephan Lauermann's lab meeting group members. Particularly, Helene Mass, Markus Möller, Carl-Christian Groh, Justus Preusser, Simon Block, Miguel Risco, and Alex Winter have not only been colleagues but also friends and mentors, enriching my academic journey in countless ways. Their contributions have significantly enhanced my papers and presentations, and I am grateful for their support and camaraderie.

I am grateful for the administrative support and assistance from the Bonn Graduate School of Economics, the Institute for Microeconomics, and the Collaborative Research Center Transregio 224. Special thanks to Stefanie Stolz and Silke Kinzig for their unwavering support. Financial support from the DAAD, the Bonn Graduate School of Economics, and Collaborative Research Center Transregio 224 has enabled me to focus on my research without financial concerns, for which I am profoundly thankful.

Last but not least, I would like to thank my family and friends for their unwavering support. To my parents, Servet Bilgin and Ahmet Mehmet Bilgin, for their unconditional love, which has been my anchor and allowed me to figure out what I want to do with my life. I would like to thank Mustafa Kaba for moving to Bonn at just the right moment, completing my family, and making research discussions even more fun.

Most importantly, I am forever grateful to my best friend, husband, colleague, and coauthor, Cavit Görkem Destan, for his unwavering presence during every twist and turn of this journey and illuminating my path with his support whenever the road felt dark. May this acknowledgment mark the beginning of a new chapter filled with endless possibilities.

Günnur Ege Bilgin
Bonn, 15. April 2024

Contents

List of Tables

Introduction

In the microeconomics world, we often think of value as money and goods that can be bought, sold, divided, or traded. For example, in the most indispensable applications of microeconomics, we find price vectors that clear the market, optimal bidding strategies in auctions as in Myerson (1981), or optimal transfers for the provision of a public good as in Vickrey (1961), Clarke (1971), Groves (1973), and Groves and Ledyard (1977). In all these examples, an equilibrium price or equilibrium transfers act as a baby pacifier in the economy so that all agents involved are content with the outcome.

Nevertheless, in many economic environments, using money as a medium is either impossible or legally prohibited. In other words, intrinsic utility is not exchangeable for money. For instance, marriage usually happens due to love rather than a monetary interaction. Similarly, bribing your way into a college is considered unethical and mostly illegal.[1] These applications are examples of matching theory, one of the prime applications of nontransferable utility. As Alvin E Roth (2015) puts forward: "*Matching* is economist-speak for how we get the many things we choose in life that also must choose us". When we consider political economy, the flavor of nontransferable utility is similar even though the environment differs from matching. Voters cannot be monetarily compensated in elections if the candidate they favor loses. Neither can they be incentivized by money to vote for someone in a secret ballot.

Since my bachelor's education in economics, I have been captivated by the intricate workings of markets, especially their ability to operate even without monetary mediums. This fascination led me to explore whether a social planner or the agents themselves can achieve an allocation that satisfies everyone involved, particularly in nontransferable utility environments. I delve into this intriguing inquiry in my thesis across three distinct chapters. Each chapter presents models within nontransferable utility frameworks and analyzes equilibria where all parties are content and unincentivized to deviate, as well as the strategic interactions between these agents.

1. For the most canonical and earliest examples, see Nash (1950), Gale and Shapley (1962), Aumann (1985), and Alvin E. Roth (1985).

In Chapter 1, I analyze a canonical two-sided many-to-one matching market such that there are two disjoint sides in an economy (firms-workers). I present my analysis within a decentralized search model with frictions, where a finite number of firms and workers meet randomly until the market clears. I compare the stable matchings of the underlying market and equilibrium outcomes of the decentralized model when time is nearly costless. This chapter focuses on whether agents can find plausible allocations themselves and how this changes when we allow firms to employ more than one worker.

In contrast to the case where each firm has just a single vacancy, as in Wu (2015), I show that stable matchings are not obtained as easily. In particular, there may be no Markovian equilibrium that uniformly implements either the worker- or the firm-optimal stable matching in every subgame. The challenge results from the firms' ability to withhold capacity strategically. Yet, this is not the case for markets with vertical preferences on one side, and I construct the equilibrium strategy profile that leads to the unique stable matching almost surely. Moreover, multiple vacancies enable firms to implicitly collude and achieve unstable but firm-preferred matchings, even under Markovian equilibria. Finally, I identify one sufficient condition on preferences to rule out such opportunities.

In Chapter 2, which is joint work with Orhan Aygün, an extension to the many-to-one placement problem is analyzed, where some doctors are exogenously guaranteed a seat at a program, which defines a lower bound on their assignment. These lower bounds are called assignment guarantees, and a placement mechanism can only place doctors in places they like at least as much as their assignment guarantees. Respecting assignment guarantees, combined with the limited capacities of programs, often violates fairness and leaves more preferred doctors unemployed. Pursuing fairness, a designer often has to deviate from the target capacities of programs. We introduce two notions tailored to the environment to prevent excessive deviations: q-fairness and avoiding unnecessary slots. We present the Assignment-Guarantees-Adjusted Mechanism (AGAM) and show that it is the unique strategy-proof mechanism that satisfies q-fairness and avoids unnecessary slots while respecting assignment guarantees. Furthermore, among the q-fair and respect guarantees mechanisms, AGAM minimizes the deviation from the target capacities.

Finally, Chapter 3, joint work with Cavit Görkem Destan, presents a model demonstrating politicians strategically adopt extreme positions even when the voters are homogeneous and moderate. We examine the behavior of voters and electoral candidates under the assumption that the salience of political issues affects voting decisions through voter preferences. Voters have limited attention which is unintentionally captured by distinctive policies, which is referred to as *salience bias* in behavioral theory.[2] We demonstrate that candidates who differ in their budget constraints, along with voters who have such limited attention, can account for extremist policies, even though voters are identical in their preferences.[3] Subsequently, we examine the elections with *decoy* candidates unlikely to win. Even though these candidates do not attract the voters, they might still influence the election outcome by altering salience. Moreover, we provide experimental evidence that salience affects consumer preferences and election outcomes using a representative sample of Turkey's vote base.

References

Aumann, Robert J. 1985. "An Axiomatization of the Non-Transferable Utility Value." *Econometrica* 53 (3). https://doi.org/10.2307/1911657. [1]

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2012. "Salience theory of choice under risk." *Quarterly Journal of Economics* 127 (3). https://doi.org/10.1093/qje/qjs018. [3]

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2013. "Salience and consumer choice." *Journal of Political Economy* 121 (5). https://doi.org/10.1086/673885. [3]

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2016. "Competition for Attention." *Review of Economic Studies* 83 (2). https://doi.org/10.1093/restud/rdv048. [3]

Clarke, Edward H. 1971. "Multipart pricing of public goods." *Public Choice* 11 (1). https://doi.org/10.1007/BF01726210. [1]

Gale, D., and L. S. Shapley. 1962. "College Admissions and the Stability of Marriage." *American Mathematical Monthly* 69 (1). https://doi.org/10.1080/00029890.1962.11989827. [1]

Groves, Theodore. 1973. "Incentives in Teams." *Econometrica* 41 (4). https://doi.org/10.2307/1914085. [1]

Groves, Theodore, and John Ledyard. 1977. "Optimal Allocation of Public Goods: A Solution to the "Free Rider" Problem." *Econometrica* 45 (4). https://doi.org/10.2307/1912672. [1]

Myerson, Roger B. 1981. "Optimal Auction Design." *Mathematics of Operations Research* 6 (1). https://doi.org/10.1287/moor.6.1.58. [1]

Nash, John F. 1950. "The Bargaining Problem." *Econometrica* 18 (2). https://doi.org/10.2307/1907266. [1]

Nunnari, Salvatore, and Jan Zápal. 2017. "Dynamic elections and ideological polarization." *Political Analysis* 25 (4). https://doi.org/10.1017/pan.2017.24. [3]

Roth, Alvin E. 2015. *Who gets what–and why: the new economics of matchmaking and market design.* Houghton Mifflin Harcourt. [1]

Roth, Alvin E. 1985. "The college admissions problem is not equivalent to the marriage problem." *Journal of Economic Theory* 36 (2). https://doi.org/10.1016/0022-0531(85)90106-1. [1]

Vickrey, William. 1961. "Counterspeculation, auctions, and competitive sealed tenders." *Journal of Finance* 16 (1). https://doi.org/10.1111/j.1540-6261.1961.tb02789.x. [1]

2. Salience bias is usually analyzed within an industrial organization framework, see Bordalo, Gennaioli, and Shleifer (2012), Bordalo, Gennaioli, and Shleifer (2013), Bordalo, Gennaioli, and Shleifer (2016).

3. Another political economy application with candidates who face the same budget constraint can be found in Nunnari and Zápal (2017).

Wu, Qinggong. 2015. "A finite decentralized marriage market with bilateral search." *Journal of Economic Theory* 160. https://doi.org/10.1016/j.jet.2015.09.005. [2]

Chapter 1

Decentralized Many-to-One Matching with Random Search*

1.1 Introduction

In the absence of a central planner, many applications of many-to-one matchings evolve over time. For instance, consider the labor market for junior positions, where a finite number of job candidates and job vacancies interact over time. The labor market initially comprises a pool of job seekers and available positions. However, with the gradual execution of contracts between some agents, the market shrinks, with the number of vacancies and available workers decreasing as positions are filled.

This unfolding job search process highlights a fundamental trade-off for job candidates and employers. Job seekers must decide whether to accept the current job offer whenever they receive one or continue searching for potentially more favorable alternatives. On the other side of the market, employers face a similar trade-off. They must decide whether to accept the available candidates or wait for potential candidates who might be a better fit for their organizations. Moreover, since firms usually have multiple vacancies, they must consider the whole team they will have at the end of the employment procedure. Of course, it is not for sure for either side that they will encounter better alternatives in the future. One

* I am grateful to my advisors Stephan Lauermann and Daniel Krähmer, and thankful for the significant comments and suggestions of Mehmet Ekmekci, Orhan Aygün, Francesc Dilmé, Jonas von Wangenheim, Alex Westkamp, Qinggong Wu, Bumin Yenmez, Cavit Görkem Destan, Carl-Christian Groh, Markus Möller, Axel Niemeyer, Justus Preusser, Simon Block, Jacopo Gambato, Miguel Risco, and Jasmina Simon. I thank workshop participants at Bonn and the 12th Conference of Economic Design in Padova and the 2023 Econometric Society European Meeting in Barcelona attendees for their helpful feedback. I am sincerely thankful for the warm hospitality of Boston College, where I had the opportunity to write parts of this paper. Support by the German Research Foundation (DFG) through CRC TR 224 (Project B03) is gratefully acknowledged.

can observe similar dynamics at graduate school admissions with exploding offers or when families are looking for schools for their children in competitive markets.[1]

Unlike decentralized search, when many-to-one matchings are centralized, a central planner collects preferences from both sides and imposes an allocation for the entire economy. A minimal requirement that we would expect from those assignments is *stability*, which ensures agents have no incentives to deviate from the proposed allocation. In a many-to-one matching context, this would mean agents do not prefer the outside option of not participating in the allocation, and no firm and worker pairs prefer each other over their assignments. When there are no complementarities between the workers, we know stable matchings exist by Roth (1985). Nevertheless, it is unclear whether agents can find stable matchings of the underlying market without the intervention of a central planner.

The contribution of this paper is to study whether equilibrium behavior by the agents in a search game leads to stable (or unstable) matchings. Doing so, I restrict attention to simple strategies, in which agents condition their behavior on the current state of the market only *(Markovian strategies)*, and consider the game outcomes that are obtained almost surely *(enforced matchings)*. I show that the worker- or firm-optimal stable matchings of the underlying market may not be enforceable in any subgame under equilibrium behavior. On the other hand, for some markets, there are equilibria that enforce unstable matchings. These results differ from the one-to-one counterpart of the model analyzed by Wu (2015), where firms hire only one worker. The difference shows that, unlike centralized markets, there is a significant disparity between finite one-to-one and many-to-one markets when analyzed within a search model with random meetings. A substantial portion of this discrepancy is attributed to the nature of the firms to employ multiple workers over time and stay in the game until they fill their capacities.

This paper studies a finite many-to-one market with general preferences à la Gale and Shapley (1962) where the matching evolves without a central planner, and there are search frictions as in Smith (2006). Firms hire multiple workers up to their capacity, whereas each worker can work for at most one firm. I analyze the finite market within a search and matching framework, in which bilateral meetings between the agents are random and time is nearly costless.[2] The market is common knowledge to the agents. Upon meeting, agents decide whether to accept or reject each other. Mutual acceptance results in leaving the market, which means irreversibly leaving the market for the worker and losing one vacancy for the firm. Therefore, as many-to-one matching evolves within the search game, the

1. For an analysis of childcare assignments in Germany, see Reischmann, Klein, and Giegerich (2021).

2. When time is too costly, the agents become so impatient that they accept all individually rational partners instead of waiting for potentially better matches.

initial market of available agents shrinks over time. In the analysis, I consider the matchings that are attained almost surely in equilibrium.[3]

First, I show that firms might find it favorable to unilaterally deviate from a strategy profile that would enforce the worker- or firm-optimal stable matching for any subgame. The one-step deviation occurs in the sense that firms hire less preferred workers than their stable partners in early periods and continue the search game with fewer vacancies. By doing so, they might ensure more favorable sets of workers, making the one-step deviation profitable. Interestingly, this result about the decentralized search game follows a similar pattern from the centralized many-to-one matching literature, i.e., when firms of the centralized market play a capacity revelation game. Konishi and Ünver (2006) show that firms might have incentives to misreport their capacities when this information is unknown to the central planner. For instance, firms may trigger rejection cycles and increase their overall welfare by underreporting their capacities. Here, by employing some workers in early periods, firms mimic capacity underreporting (hence, the rejection cycles) for subsequent periods, even though every component is common knowledge to the agents.

Nevertheless, no rejection cycle can be triggered when preferences on at least one side of the market are perfectly aligned (a consensus among workers about which firm is better or vice versa), as shown by Balinski and Sönmez (1999). Therefore, firms can only be worse off by such one-step deviations. Hence, under vertical preferences, there is an equilibrium strategy profile that enforces the unique stable matching of the underlying market. I show this by constructing the strategy profile, in which workers accept the firms they prefer at least as much as those they are matched with under the stable matching and reject all others. With a slight difference, firms accept the workers they prefer at least as much as their least preferred worker under the stable matching. The strategy profile is simple in the sense that agents “follow” the unique stable matching for any initial or remaining market, which applies to both on- and off-the-equilibrium path.

Second, I demonstrate that unstable matchings may be enforceable as equilibrium outcomes when waiting is costless. In contrast to the unilateral deviation incentives in the previous part preventing stable matchings from being enforceable in every subgame, firms form collusions to enforce unstable matchings. By collectively and credibly committing to avoid forming blocking pairs, they can switch their stable partners and achieve more favorable outcomes. However, for the firms to engage in such a strategic action, at least some firms should be willing to exchange some workers of the stable matching.[4] Naturally, some alignment in

3. Wu (2015) shows how unstable matchings can arise with positive probability for one-to-one and every one-to-one matching is also a many-to-one matching.

4. In many-to-one markets, there might be individually rational matchings that all firms prefer to the firm-optimal stable matching as shown by Roth (1985).

preferences prevents such scenarios. One such preference structure is exemplified by the Sequential Preference Condition, which not only ensures the existence of a unique stable matching in the market but also prevents unstable matchings from being enforced in equilibria.

Crucially, both of the main results are different if we assume firms have a capacity of one. This environment, traditionally referred to as a *one-to-one* or *marriage* market, is analyzed by Wu (2015) within a search and matching framework with similar features. In his paper, Wu shows that for any stable matching of the underlying market, there is a Markovian strategy profile that enforces the stable matching. Similarly, he shows that there is no equilibrium Markovian strategy profile under which an unstable matching is enforced. It turns out these results are restricted to one-to-one environments where every firm has at most one vacancy and are not robust to multiple vacancies.

These theoretical results suggest some policy implications as well. For instance, in college admission problems, the primary objective is often to ensure that better students are assigned to more preferred colleges and make it undesirable for agents, particularly colleges, to engage in strategic manipulation that could hinder *fair* outcomes. Such manipulative behavior would be counterproductive, given that colleges typically serve as public goods, and their pursuit of self-interest through strategies undermines the collective welfare they are meant to provide. However, the results of this paper suggest that, in equilibrium, colleges could collude to secure better outcomes for themselves, which comes at the expense of leaving students in a disadvantaged position. From a policy standpoint, these findings emphasize the importance of having a central clearinghouse to counteract and prevent such behavior from emerging and safeguard the interests of all parties involved, especially when agents on one side of the market have multiple capacities.

Furthermore, I analyze how the equilibrium outcomes of the decentralized many-to-one search model differentiate from the dynamic stability notion, studied by Doval (2022) for one-to-one markets and by Altinok (2022) for the many-to-one case. In dynamic stability, workers arrive over time, and firms consider the endgame implemented by evolving matchings. Despite the similarity, workers arriving over time translates into binary meeting probabilities, either 0 or 1. Thus, dynamically stable matchings may not be enforceable via the equilibria of the decentralized many-to-one search model with random search.

The remainder of the paper is structured as follows: After discussing the related literature, I introduce the model in Section 2. In Section 3, I compare the outcomes that are enforced in equilibrium with stable and unstable matchings of the underlying market. I show a sufficient condition on preferences that prevents the enforceability of unstable matchings in equilibria in Section 4. In Section 5, I compare the decentralized many-to-one search model with random search to the notion of dynamic stability by Altinok (2022). I conclude in Section 6.

1.1.1 Literature Review

The paper connects to multiple strands in both centralized matching and search literature. Naturally, the first strand is the traditional finite many-to-one market literature that determines the market I am working with (as in Gale and Shapley (1962), Roth (1985), Roth and Sotomayor (1989), Roth and Sotomayor (1992)). Many results of these papers are significant when the market is considered in a search and matching framework. For instance, Roth (1985) shows that, unlike one-to-one markets, there might be individually rational matchings that are preferred to the firm-optimal stable matching by all firms. In this paper, this fact opens up collusion opportunities for the firms in equilibrium. Moreover, Konishi and Ünver (2006) considers the many-to-one market within the framework of a capacity revelation game. They show that firms might have incentives to misreport their capacities. Therefore, neither the worker- nor the firm-optimal stable matching is necessarily implementable. In this paper, even though the market is common knowledge for agents, firms can mimic capacity underreporting by hiring in the early periods of the search game. Hence, stable matchings may not be enforceable as search equilibria.

Before discussing other strands, current work is the most closely related to Wu (2015), in the sense that Wu considers the one-to-one counterpart of the model analyzed in this paper. Similar to the title of Roth (1985), *the decentralized finite many-to-one matching with random search is not equivalent to its one-to-one counterpart*. In the many-to-one case, stable matchings may not be enforceable in any subgame in equilibrium, and unstable matchings can be enforced more easily. These results are attributed to firms employing multiple workers over time, differentiating the current paper from Wu (2015).

The second strand of related literature analyzes finite matching markets from a dynamic perspective with directed offers. Among those, Pais (2008) considers where acceptance is deferred in line with the Gale and Shapley (1962) algorithm. However, the current paper models a different trade-off. First, an agent can still improve within the game when acceptance is deferred. However, when acceptance is permanent, it eliminates the chance of meeting someone better. Hence, this paper considers the potential regret upon accepting someone. Furthermore, unlike the deferred acceptance algorithm, the meeting technology in this paper still allows the separated agents to meet again in future periods as long as they are still on the market. In addition to Wu (2015), Haeringer and Wooders (2011) and Niederle and Yariv (2009) analyze a more similar trade-off to this paper in which they consider irreversible market exit upon match with one-to-one markets. Similarly, Alcalde, Pérez-Castrillo, and Romero-Medina (1998) and Alcalde and Romero-Medina (2000) analyze a game of only two stages for many-to-one markets, and Alcalde and Romero-Medina (2005) consider the sequential proposal version of the latter paper. For a comparison between deferred and immediate acceptance for one-to-one markets, see Alcalde (1996).

Another approach in the dynamic finite markets strand is the *dynamic stability* notion, studied by Doval (2022) for one-to-one markets and by Altinok (2022) for many-to-one markets. In both papers, agents arrive, and matchings evolve over time, with the condition that there should not be blocking pairs in *intertemporal* matchings as well as in the final matching. In the last section of the paper, I show that dynamically stable matchings are not necessarily enforceable. This is mainly due to the difference in meeting technology of both models. In fact, I consider scenarios where every pair has a positive probability of meeting, but the dynamically stable matching of a market may heavily depend on how agents arrive over time - which requires some meeting probabilities being exactly 0. Similar to dynamic stability, Blum, Roth, and Rothblum (1997) analyze markets that are initially stable but destabilize through new entrants to the job market or through people who leave.

The third strand that this paper connects to is the study of steady states of a search and matching game with a continuum of agents under nontransferable utility. When comparing the steady states with the stable matchings of the underlying market, Burdett and Coles (1997), Eeckhout (1999), and Smith (2006) assume vertical preferences that result in a unique stable matching, whereas Adachi (2003) and Lauermann and Nöldeke (2014) consider general preferences. In Adachi (2003), the stock of searching agents is exogenously given, whereas Lauermann and Nöldeke (2014) consider an endogenous steady-state, endogeneity being naturally the case for a finite market. Even though mutual acceptance has the same consequence of leaving the market, rejecting a potential partner and they marrying someone else has different implications in a continuum because there are still replicas of the same agent in the market that would be available in the future. For search models under transferable utility and vertical agents, see Becker (1973) and Shimer and Smith (2000), and Chade, Eeckhout, and Smith (2017) for a detailed survey of this strand.

Importantly, all these papers are in a one-to-one setting. The steady-state analysis for a many-to-one market would allow the firms to meet the same agent repeatedly. Clearly, having multiples of the same worker is not defined in the preferences of any underlying matching market and would require technical manipulation in preferences similar to Hatfield and Kominers (2015), which is not a realistic approach within a search and matching framework.

Search problems are widely studied not only on micro levels but also on macro levels to answer bigger-scale questions and better understand the economies as a whole. Most famous examples of the search literature include Mortensen (1982) and Pissarides (1985) search models. These early papers have ex-ante homogeneous agents and focus on wage bargaining and unemployment dynamics in the economy.

1.2 Model

1.2.1 Environment

The search environment is the same as a standard many-to-one matching market, with a finite set of firms and a finite set of workers that are bilaterally and randomly searching for each other. The set of firms is denoted by $F = \{f_1, \ldots, f_n\}$ and the set of workers is $W = \{w_1, \ldots w_m\}$, where $(f, w) \in F \times W$ denotes a generic firm-worker pair.

Workers have complete and strict preferences over firms and remaining unemployed. The utility that worker w receives is denoted by $v(i, w) \in \mathbb{R}$, where $i \in F \cup \{w\}$. This means workers are indifferent between the vacancies of the same firm and their coworkers. The many-to-one structure implies that firms can employ many workers, whereas a worker can be matched to one firm at most. The capacity vector $q = (q_{f_1}, \ldots, q_{f_n})$ specifies the maximum number of workers a firm can employ. Consequently, firms have complete and strict preferences over sets of workers, which is denoted by $u(f, \Omega) \in \mathbb{R}$, where $\Omega \subset W$. The utility of not being in a match is normalized for both parties: $u(f, \emptyset) = v(w, w) = 0$. A pair $(f, w) \in F \times W$ is an acceptable pair at u, v if w is *acceptable* for f, and f is acceptable for w, that is, $u(f, \{w\}) > 0$ and $v(f, w) > 0$.[5]

An underlying preference relation over individual workers induces firms' preferences over sets of workers. In other words, if two sets differ in only one worker, the firm prefers the set with the more preferred worker. This condition is referred to as *q-responsive preferences* in the existing literature. Formally:

Definition 1.1. The preferences of firms over 2^W are q-responsive if they satisfy the following conditions:

(1) For all $\Omega \subset W$ such that $|\Omega| > q_f$, we have $u(f, \Omega) < 0$.

(2) For all $\Omega \subset W$ such that $|\Omega| < q_f$ and $w \notin \Omega$, $u(f, \Omega \cup \{w\}) > u(f, \Omega)$ if and only if $u(f, \{w\}) > u(f, \emptyset) = 0$.

(3) For all $\Omega \subset W$ such that $|\Omega| < q_f$ and $w, w' \notin \Omega$, $u(f, \Omega \cup \{w\}) > u(f, \Omega \cup \{w'\})$ if and only if $u(f, \{w\}) > u(f, \{w'\})$

A many-to-one search market is represented by $M = (F, W, q, u, v)$ and all components are common knowledge to all agents. The summary of the assumptions on the market for a search game to start is the following:

(1) Both parties have strict preferences.

(2) The utility of being single is normalized to zero for both parties.

5. For notational convenience, $\Omega \succ_f \Omega'$ is often used instead of $u(f, \Omega) > u(f, \Omega')$, and same for $f \succ_w f'$ instead of $v(f, w) > v(f', w)$.

(3) Firms have q-responsive preferences over sets of workers.
(4) The market is finite: $|F| < \infty$, $|W| < \infty$, and $q_f < \infty \quad \forall f$.
(5) The initial market is nontrivial in the sense that there are some acceptable pairs.

1.2.2 The Search Game

The game starts at $t = 0$ with the initial market M and continues for an indefinite amount of time. Each day, a random pair (f, w) meets randomly. I describe the meeting process in detail in the following subsection 1.2.3. Upon meeting, w first decides whether to apply to f or not. Then, if w applies, f decides whether to accept w. If w does not apply or f rejects w, they separate and return to the market to keep searching. If w applies and f accepts, w leaves the market, and f loses one of its vacancies. If f has more vacancies, it stays in the market but leaves if q_f is full after hiring w.

Upon hiring, w receives a one-time payoff of $v(f, w)$. The firm also receives a one-time payoff. However, this depends on the already hired workers. Namely, suppose f has already hired $\Omega \subset W$ before meeting w. In that case, the one-time payoff it gets after hiring w is $u(f, \Omega \cup \{w\}) - u(f, \Omega)$, i.e., it enjoys the additional utility it derives from hiring w. The instant utility captures the immediate utility gain of extending the labor force.[6] The meeting and the decisions take place on the same day t, and if the meeting concludes with hiring, the agents' utilities are discounted by δ^t. The common discount factor refers to the cost of time and is the first source of friction in the model.

Leaving the market upon mutual acceptance is permanent. That is, workers cannot quit, and the firms cannot fire workers. As a result, the market weakly shrinks over time. Any *submarket* or *remaining market* is then denoted by $M' = (F', W', q', u, v)$, where $F' \subset F$ denotes the remaining firms in the market, $W' \subset W$ the remaining workers, and $q' \subset Q$ the remaining capacities of the firms F'. The capacities of the firms will decrease over time, therefore $q' \leq q$ necessarily.

1.2.3 The Contact Function

The second source of friction in the decentralized many-to-one search game is the random meeting process defined by the contact function. For any day, the contact function $C(f, w, M')$ defines the probability that the pair (f, w) will meet given that the remaining market at that day is M'. The assumptions on the contact function are the following:

(1) The pair (f, w) meets with positive probability only if f has not filled its capacity and w is not matched to any firm: $C(f, w, M') = 0$ if $f \notin M'$ or $w \notin M'$.

6. I will revisit this assumption in Section 1.3.2.

(2) A meeting occurs each day: $\sum_{(f,w)\in M'} C(f,w,M') = 1$.
(3) Every meeting between the pairs in the remaining market is somewhat possible:
$\exists \epsilon > 0$ s.t. $C(f,w,M') \geq \epsilon$ if $(f,w) \in M'$.
(4) By definition, the contact function does not depend on the history itself. The pair (f,w) has the same meeting probability whenever the remaining market is the same along different histories.

The game ends if either one side of the market is fully matched or there are no mutually acceptable pairs remaining in the market, that is for any given submarket M', $\nexists(f,w)$ such that $u(f,\Omega\cup\{w\}) \geq u(f,\Omega)$ for $f \in M'$ holding Ω and $v(f,w) \geq 0$ for $w \in M'$.[7] Workers search until they are matched to a firm, or the game ends. Similarly, firms search until they fill their capacity or the game ends.

The search game is represented by the tuple $\Gamma = (F,W,q,u,v,C,\delta)$, and components of the search game are common knowledge.

1.2.4 Equilibria

Due to complete information and the finite and dynamic structure of the many-to-one bilateral search game, the appropriate equilibrium concept for analysis is subgame perfect equilibrium. The set of all histories is denoted by $\mathscr{H}$, where $h \in \mathscr{H}$ and $\hat{\mathscr{H}}$ is the set of all non-terminal histories, after which the game continues with a new pair meeting. Additional to the initial search game Γ, any subgame that follows after history h is $\Gamma(h) = (F',W',q',u,v,C,\delta)$, and $M'(h)$ is the remaining market. A strategy profile is denoted by σ, and $\sigma|\Gamma(h)$ is its restriction to the subgame $\Gamma(h)$. Furthermore, let μ_h denote the instantaneous matching at history h, i.e., all the meetings that have ended with employment.

Along with the regular subgame perfect equilibrium analysis, the paper considers some refinements to the information structure. The baseline scenario, called *full-awareness*, assumes that agents possess the capability to observe and base their decisions on every aspect of the game's history. The *private-dinner condition* refers to environments where meetings between the agents take place in private settings. Agents can observe what meetings realize and their conclusions, but in case of separation, they do not observe the reason for the split, i.e., who rejected whom. However, when a meeting concludes with employment, it indirectly signals mutual acceptance to all agents.

Taking the refinement one step further, we reach the most stringent, and the most commonly used refinement in the literature is the *Markov condition*, under which agents condition their behavior only on the current state of the market but not on any other component of the history. The current state of the market

7. Due to q-responsive preferences, $u(f,\Omega\cup\{w\}) \geq u(f,\Omega)$ is equivalent to $u(f,w) \geq 0$.

consists of an instantaneous matching for every history, which also implies a remaining market.[8] This progression of information structure refinements enriches the analysis and sheds light on the intricate dynamics of the economic model under scrutiny.

Independent of the information structure, the game dictates the following for optimal pure strategy behavior upon meeting:

(1) A worker w applies to f if the expected utility upon application exceeds the expected utility of continuing the search.
(2) Any firm f accepts a w for one of its vacancies if the instant utility gain and the expected utility gain for the remaining vacancies exceeds the expected utility gain for continuing search without accepting w.

A strategy profile σ constitutes a subgame perfect equilibrium of the search game $(F, W, q, u, v, C, \delta)$ if σ is optimal for every agent in every subgame.[9]

1.3 Connection to Centralized Many-to-One Markets à la Roth

In this section, I analyze the equilibrium outcomes attained almost surely in equilibrium. Specifically, I compare them to stable and unstable matchings of the underlying many-to-one market. Comparing search outcomes and static allocations differs from the one-to-one case. This divergence is primarily attributed to the dynamic evolution of firms' engagement with multiple workers as the game unfolds.

1.3.1 Evolving Many-to-One Matching

At each history h, the instantaneous matching μ_h collects all the meetings that have ended with recruitment so far. In compliance with the centralized matching literature, the rules of the search game ensure that any instantaneous matching is naturally a many-to-one matching and has the following properties:

(1) $|\mu_h(w)| = 1$ and if $\mu_h(w) \neq w$, then $\mu_h(w) \subset F$.
(2) $\mu_h(f) \in 2^W$ and $|\mu_h(f)| \leq q_f$.
(3) $\mu_h(w) = f$ if and only if $w \in \mu_h(f)$.

Verbally, workers are matched to only one firm at most, and if they are not single, they are matched to a firm. Firms are matched to subsets of workers that do not exceed their capacity. Lastly, a worker is matched to a firm if and only if that firm employs the worker, so the matching process is bilateral.

8. The Markov condition in Wu (2015) considers conditioning on the remaining market only, which does not change any of the results presented in this paper.

9. The conditions get stronger as we move from full awareness to Markov condition. Therefore, the set of equilibrium strategies weakly shrinks in the same direction.

Observe that any instantaneous matching μ_h implies a remaining market such that already matched agents and seats of the firms are removed from the initial market. Similarly, any contact function $C(f, w, M')$ together with a strategy profile σ induces a probability mass function on outcome matchings, details of which are to be found in the appendix. A matching μ *arises* if it appears as an outcome matching with positive probability under σ. If μ obtains almost surely under σ, I say that σ *enforces* μ.[10]

The prevailing solution concept in centralized matching theory is stability, which assesses the sustainability of a matching. In essence, it ensures that agents lack incentives to disrupt the proposed matching, whether through individual or bilateral actions, thus making it both individually rational and unblocked. Formally, a matching μ is individually rational if all the matched pairs are acceptable. A matching μ is blocked by the pair (f, w) if they are not matched under μ but prefer each other over their matches under μ. That is, μ is blocked by the pair (f, w) if at least one of the following conditions hold:

(1) If $|\mu(f)| \leq q_f$ and $\mu(w) \neq f$,
$v(f, w) > v(\mu(w), w)$ and $u(f, w) > u(f, w')$ for some $w' \in \mu(f)$.

(2) If $|\mu(f)| < q_f$ and $\mu(w) \neq f$,
$v(f, w) > v(\mu(w), w)$ and $u(f, w) > 0$

If μ is individually rational and unblocked, it is a stable matching. Since the firms' preferences are q-responsive, the set of stable matchings is non-empty for any initial market. Furthermore, by the famous Rural Hospital Theorem by Roth (1984), the set of matched agents is the same under every stable matching, and each firm that does not fill its quota has the same set of agents matched under every stable matching.

1.3.2 Enforcing Stable Matchings

In the first analysis section, the main question is the relation between the subgame perfect equilibrium outcomes of the search game and the stable outcomes of the underlying many-to-one market.

Firstly, small values of δ reflect high costs of time. If waiting is sufficiently costly, the workers would apply to acceptable firms, and firms would accept every acceptable worker (partners that give positive utility). Therefore, the rather interesting question is for larger δ, in fact, for $\delta \to 1$, that is called as *limit equilibria*.

Definition 1.2. A strategy profile σ is a **limit equilibrium** of the many-to-one search environment (F, W, q, u, v, C) if there exists some $d < 1$ such that σ is an SPE of the many-to-one search game $(F, W, q, u, v, C, \delta)$ for all $\delta > d$.

10. If σ enforces μ, the set of matchings that arise under σ may contain other elements than μ, all of which have zero probability of arising.

In the one-to-one component of the search model, Wu (2015) demonstrates that for any stable matching μ^* of any underlying market, there is a strategy profile σ^* that satisfies the Markov condition, is a limit equilibrium, and enforces μ^*. In addition, he constructs this strategy profile. As one might expect similarities with the many-to-one model (at least with responsive preferences), in the following, I show that such a Markovian strategy profile may not exist in the many-to-one search game - even with additively separable utility over the workers.[11] The crucial distinction is that firms have a collective structure and stay on the market until they fill their capacities.

In this paper, I consider limit equilibria where waiting is almost costless and enforced matchings while assuming responsive preferences. Therefore, for all results presented in this paper, the assumption about how firms derive utility from expanding their employment set can be replaced with another, simpler one: that firms only care about the end-outcome of the search game.[12] This specific focus allows us to abstract from other strategic problems that can arise to prolong or expedite the game on the firm's side and target the differences caused by cumulative employment only.

Enforcing the Worker-Optimal Stable Matching via Markov Equilibria

According to the strategy profile defined in Wu, for any given stable matching, agents accept partners they prefer at least as much as their stable partners, which is a limit equilibrium and eventually leads to the stable matching of interest. Off-the-equilibrium path, agents play the same strategy according to the firm-optimal stable matching, enforcing FOSM for off-the-equilibrium path remaining markets. In the following, I test this in many-to-one markets in the following way: Is there a many-to-one counterpart of the strategy profile by Wu that would enforce stable matchings for any subgame of the initial game? A natural starting point here is to start with extremal matchings, i.e., the worker-optimal and the firm-optimal stable matchings.

For any initial many-to-one market (with responsive preferences), there is a worker-optimal stable matching μ^W (WOSM), corresponding to the outcome of the worker-proposing deferred acceptance algorithm. In the following proposition, I show that if there is a Markovian strategy profile that enforces the WOSM for any subgame, it is not a limit equilibrium for some initial markets.

Proposition 1.3. *For some many-to-one bilateral search games, there is no Markovian strategy profile that is a limit equilibrium and enforces the worker-optimal stable matching in every subgame.*

11. The utility function satisfies: $u(f, \Omega \cup \{w\}) = u(f, \Omega) + u(f, \{w\})$ $\forall \Omega \subset W$ such that $w \notin \Omega$.

12. Details can be found in the Appendix for Proposition 1.

The proposition can be proved by constructing a market for which a Markovian strategy profile enforcing the WOSM in every subgame of any search game derived from the market is not a limit equilibrium.

Proof. Suppose otherwise, i.e., there is a Markovian strategy profile σ^* that is a limit equilibrium and that enforces the worker-optimal stable matching in any subgame of any many-to-one search game $(F, W, q, u, v.C, \delta)$. Note that σ^* restricted to any subgame $\Gamma(h)$ would enforce the WOSM of the remaining market $M'(h)$.

Now, consider the following example with $F = \{f_1, f_2, f_3\}$, $q = \{2, 1, 1\}$ and $W = \{w_1, w_2, w_3, w_4, w_5\}$. In addition to the following preference profile, suppose $\{w_3, w_5\} \succ_{f_1} \{w_1, w_4\}$, which does not violate responsive preferences. Other than that, any q-responsive completion of firm preferences is admissible.

$$f_1 : w_3 \succ w_4 \succ w_1 \succ w_2 \succ w_5$$
$$f_2 : w_1 \succ w_2 \succ w_3 \succ w_4 \succ w_5$$
$$f_3 : w_2 \succ w_3 \succ w_1 \succ w_4 \succ w_5$$

$$w_1 : f_1 \succ f_2$$
$$w_2 : f_2 \succ f_3$$
$$w_3 : f_3 \succ f_1$$
$$w_4 : f_1 \succ f_2 \succ f_3$$
$$w_5 : f_1 \succ f_2 \succ f_3$$

In this example, the WOSM is $\mu^W = \{(f_1; w_1, w_4), (f_2; w_2), (f_3; w_3)\}$. First, observe that for μ^W to be enforceable, f_1 should reject w_5 upon meeting because knowing that f_1 will accept, w_5 will apply to f_1 in any SPE. Second, since by assumption σ^* enforces the WOSM for any many-to-one search game, it enforces the worker-optimal stable matching of any submarket of the initial market. Third, in another many-to-one market where all agents and preferences are the same, but f_1 has a capacity of 1, $\mu^W = \{(f_1; w_3), (f_2; w_1), (f_3; w_2)\}$.[13]

Consider the subgame where the meeting between (f_1, w_5) results in employment. By assumption, the worker-optimal stable matching is enforced in the remaining subgame by σ^* restricted to that subgame. In the remaining market, all firms have a capacity of 1 and $\mu^W = \{(f_1; w_3), (f_2; w_1), (f_3; w_2)\}$ as noted above. Therefore, a one-step deviation to accept w_5 leads f_1 to $\{w_3, w_5\}$, instead of $\{w_1, w_4\}$. Since $\{w_3, w_5\} \succ_{f_1} \{w_1, w_4\}$, f_1 finds it plausible to employ w_5 upon application. Since every meeting between the pairs has a positive probability, the profitable one-shot deviation by f_1 to accept w_5 conflicts with σ^* enforcing the WOSM.

13. Firm f_1 rejecting w_1 in the first step of DA results in a rejection chain, resulting f_1 ending up with w_3.

In fact, $\mu^W = \{(f_1; w_3), (f_2; w_1), (f_3; w_2)\}$ is the unique stable matching of the remaining market. Therefore, Wu's off-the-path conjecture of playing according to the firm-optimal stable matching when an off-path acceptance occurs would also not work.

□

The preceding example illustrates how firms are willing to accept unstable partners initially to secure a more advantageous group of workers ultimately, which is not possible in a one-to-one search game. This one-step deviation by f_1, involving filling one capacity with a less desirable candidate in the early stages of the game, aligns with the *capacity manipulation game* analyzed in Konishi and Ünver (2006). In their research, they demonstrate that firms have incentives to misrepresent their capacities as a means to enhance their overall welfare under both the worker-proposing and firm-proposing deferred acceptance algorithm. In the aforementioned example, f_1 accepting w_5 would mimic capacity underreporting in subsequent rounds, ultimately benefiting the firm with a more favorable set of workers.

Enforcing the Firm-Optimal Stable Matching via Markov Equilibria

Given that it is a firm that engages in a one-step deviation by misreporting its capacities in the previous part, and Wu uses the firm-optimal stable matching for off-path, one might naturally contemplate that this impossibility result could be overcome if we consider the firm-optimal stable matching μ^F (FOSM). However, again, as demonstrated by Konishi and Ünver (2006), similar incentives for misreporting capacities may persist even when employing the firm-proposing deferred acceptance algorithm—a scenario reflected in the many-to-one search game by the hiring unstable workers in early periods.

Below, I present the FOSM counterpart of the previous proposition, along with an illustrative example that highlights an initial market configuration for which no Markovian strategy profile can enforce the FOSM in every subgame.

Proposition 1.4. *For some many-to-one bilateral search games, there is no Markovian strategy profile that is a limit equilibrium and enforces the firm-optimal stable matching in every subgame.*

Proof. Suppose otherwise, i.e., for any given market, there is a Markovian strategy profile that is a limit equilibrium and that enforces firm-optimal stable matching in any subgame of any many-to-one search game $(F, W, q, u, v.C, \delta)$ derived from this market. Now, consider the following example with $F = \{f_1, f_2\}$, $q = \{3, 3\}$ and $W = \{w_1, w_2, w_3, w_4, w_5, w_6\}$. In addition to the following preference profile, suppose $\{w_2, w_4, w_6\} \succ_{f_1} \{w_4, w_5\}$ and any q-responsive completion of firm preferences.

$$f_1 : w_1 \succ w_2 \succ w_4 \succ w_3 \succ w_5 \succ \emptyset \succ w_6 \qquad w_1 : f_2 \succ f_1$$
$$f_2 : w_4 \succ w_5 \succ w_1 \succ w_3 \succ w_2 \succ \emptyset \succ w_6 \qquad w_2 : f_2 \succ f_1$$
$$w_3 : f_2 \succ f_1$$
$$w_4 : f_1 \succ f_2$$
$$w_5 : f_1 \succ f_2$$
$$w_6 : f_1 \succ f_2$$

In this example, $\mu^F = \{(f_1; w_4, w_5), (f_2; w_1, w_2, w_3)\}$. Similar to above, by employing w_6 in early periods (who is even an unacceptable worker for the firm), f_1 can mimic capacity underreporting. In the following subgame, the firm-optimal stable matching is $\mu^F = \{(f_1; w_2, w_4), (f_2; w_1, w_3, w_5)\}$ and f_1 ensures a more favorable outcome overall. □

A Remedy to Enforce Stable Matchings: Vertical Preferences

Both propositions above show how firms can profitably deviate from a strategy profile that would enforce a stable matching. Even though one might expect different incentives under the worker-optimal or firm-optimal matching, they share a common characteristic of involving rejection chains in the deferred acceptance outcome, which firms initiate by underreporting their capacities. This cascading effect is mimicked by employing less favorable workers and ultimately yields a benefit for the deviating firm.

It's well-established in the literature that such profitable chains do not occur when one side of the market possesses identical preferences for the other side — in other words, when preferences are vertical on at least one side of the market. In this scenario, a unique stable matching μ^* emerges as the outcome of both worker-proposing and firm-proposing deferred acceptance algorithms.

In this part, given a matching μ for the market (F, W, q, u, v) let $\mathscr{M}_\mu$ denote the set of all markets (F', W', q', u, v) such that $\cup_{F \setminus F'} \mu(f) = W \setminus W'$ and $q'_f = q_f - |\mu(f)| \quad \forall f \in F$. In other words, $\mathscr{M}_\mu$ denotes all the markets that remain after some pairs of μ are matched.

In the following, I construct the Markovian strategy profile that is a limit equilibrium and enforces the unique stable matching μ^* for any initial many-to-one market with vertical preferences. The strategy profile enforces the unique stable matching for any history, both off- and on the equilibrium path. On the equilibrium path, all the remaining markets will be elements of $\mathscr{M}_{\mu^*}$, leading to μ^* eventually. Observe that if $M \in \mathscr{M}_{\mu^*}$, μ^* restricted to M is also the unique stable matching for M.

Definition 1.5. For any given market (F, W, q, u, v) with vertical preferences on at least one side (either firms share the same preferences up to their capacity, or workers have the same preferences over firms), σ_{μ^*} denotes the strategy profile in which the agents behave in the following way upon meeting:

(1) Workers apply only to firms such that $v(f, w) \geq v(\mu^*(w), w)$ in any submarket.
(2) Firms accept workers such that $u(f, w) \geq \min(u(f, w'))$ for $w' \in \mu^*(f)$, reject others.

In other words, σ_{μ^*} prescribes that workers apply to firms that they weakly prefer to their allocation under μ^*, and firms accept workers whom they prefer to their least favorite worker under μ^*.

Theorem 1.6. *For any given market (F, W, q, u, v) with vertical preferences on at least one side, σ_{μ^*} is a limit equilibrium for any search game $(F, W, q, u, v, C, \delta)$ and it enforces the unique stable matching μ^*.*

The idea of the proof is first to show σ^* being a limit equilibrium and then to show it indeed enforces μ^*. The second step follows from the fact that a pair accepts each other if and only if they are stable partners. For the first step, I show that any one-step deviation yields a worse payoff for the agents. First of all, since the strategy profile is Markovian, any deviation to not accepting a partner does not change the remaining market, hence, we are still on the equilibrium path. The only profitable deviation could arise in accepting a partner that is not accepted under σ_{μ^*}, which was the case for the propositions above. However, since preferences are vertical on the one side, such rejection cycles do not occur by misreporting capacities, and the outcome of the search game only changes by one worker. Since preferences are responsive, this difference is reflected in the overall preference, which concludes that no profitable deviation is possible. The detailed proof can be found in the appendix.

Intuitively, if agents are patient enough in a many-to-one market with vertical preferences, it is possible to obtain no other outcome but the unique stable matching in equilibrium. Even though the equilibrium strategy profile depends on the stable matching of the underlying market, the strategy profile is feasible to sustain with complete information. [14]

14. Furthermore, if we instead consider a many-to-many environment where workers can also work for multiple firms at the same time and have responsive preferences over firms (and one side has vertical preferences), replacing their strategy profile with "Workers apply only to firms such that $v(f, w) \geq \min(f', w)$ such that $f' \in \mu(w)$ in any submarket", the strategy profile would yield the stable many-to-many matching μ^*.

1.3.3 Enforcing Unstable Matchings

Knowing that the stable matchings of the underlying many-to-one market can be sustained as search equilibria, the natural follow-up question would be about the enforceability of unstable matchings. As Wu (2015) shows, unstable matchings may also be enforced in full-awareness equilibria and may arise with positive probability as outcome matchings. Those possibility results apply to our framework since every one-to-one matching is essentially a many-to-one matching, with the specification that $q_f = 1 \quad \forall f$.

On the other hand, Wu (2015) also shows that the only way to enforce unstable matchings under limit equilibria is by "Reward and Punishment Schemes", akin to those in Wolinsky (1990), which is the main discussion object of this section. In Wu's one-to-one model, the reward and punishment scheme works as follows for any blocking pair (f, w): On the equilibrium path, the w is punished for initiating the block by applying, and f is credibly rewarded for not obliging. Only that way, the blocking pair does not realize, and an unstable matching can be enforced as an equilibrium outcome. In the following, I show that in a many-to-one search model, there exists an increased potential for implementing such schemes by utilizing the remaining capacities of firms.

Crucially, in a one-to-one world, implementing such schemes requires knowledge about who rejected whom in past meetings. Consequently, when agents cannot condition their behavior on who rejected whom in past failed events, such schemes are not implementable. That information restriction is in line with what was defined as "private-dinner equilibria", where meetings take place in private environments, and others can only observe (and condition their behavior on) the outcome of meetings. Under the private-dinner condition, the successful meetings indirectly convey the information of mutual acceptance. However, in case of an unsuccessful meeting, it remains unknown to the agents who rejected whom. Since this restriction deactivates the reward and punishment schemes, no unstable matching can be enforced in a private-dinner equilibrium.

The following result shows that this result is not robust to transitioning to a many-to-one model, i.e., allowing for some firms to have $q_f > 1$. Introducing multiple capacities creates opportunities for firms to employ alternative forms of strategic manipulation and ensure more favorable outcomes for themselves. Inevitably, this has adverse implications for workers since responsive preferences preserve the lattice structure of the matchings Roth (1985). This aspect will become even more apparent when examining the example presented following the proposition.

Proposition 1.7. *In a many-to-one finite decentralized search model where waiting is costless, unstable matchings can be enforced by equilibria that satisfy the Markov condition.*

Proof. I prove this proposition by an example, which will also help illustrate the intuition behind the essential difference between many-to-one and one-to-one models.

Example 1.8. Suppose there are two firms $F = \{f_1, f_2\}$ such that $q_1 = 3$ and $q_2 = 2$ and four workers $W = \{w_1, w_2, w_3, w_4\}$ and the ordinal preference relation derived from the preferences of the agents is as follows:[15]

$$w_1 : f_1 \succ f_2 \qquad f_1 : w_4 \succ w_1 \succ w_2 \succ w_3$$
$$w_2 : f_1 \succ f_2 \qquad f_2 : w_1 \succ w_2 \succ w_3 \succ w_4$$
$$w_3 : f_1 \succ f_2$$
$$w_4 : f_2 \succ f_1$$

Now, consider the following strategy profile σ^U such that, at every history, w_1, w_2, w_3 only apply to f_1 as long as it is in the market (after f_1 leaves, apply to f_2), w_4 only applies to f_2 if is ranked within the remaining capacity in the remaining market for the f_2 or apply to both firms upon meeting. On the firm side, f_1 accepts the top q'_1 of the remaining workers, and f_2 waits until f_1 leaves the market (only accepts w_1, w_2, w_3 -who do not accept- until f_1 fulfills its capacity), then only accepts the top q'_2 of the remaining workers. In the off-path subgames, agents only accept candidates they like at least as much as their stable partners.

The strategy-profile σ^U is an equilibrium when $\delta = 1$ and depends only on the remaining market for every history, which is even more restricted than the private-dinner condition. Additionally, σ^U enforces $\mu^U = \{(f_1; w_1, w_2, w_4), (f_2; w_3)\}$ which is unstable due to the blocking pair (f_2, w_4).

□

It is easy to show that σ^U enforces μ^U since only pairs that end with pairing with positive probability are as prescripted in μ^U. Moreover, under σ^U, f_1 achieves its favorite employment set (due to responsive preferences). Similarly, w_1 and w_2 are employed by their favorite firm. Therefore, those agents have no incentive to deviate from σ^U at any point in the game. Similarly, w_3 not applying to f_2 or f_2 not accepting does not change the remaining market and, therefore, is no candidate for a profitable one-step deviation. When f_1 is still in the game, w_3 is indifferent between applying to f_2 or not when waiting is costless. Let's take the only candidate pair for a profitable deviation, the blocking pair $(f_2; w_4)$, and suppose they meet on the equilibrium path. The worker w_4 is already applying to f_2, and not applying will not be a profitable deviation either.

It is particularly noteworthy how f_2 rejects w_4 upon his application. Since having a successful meeting with w_1 and w_2 is impossible, f_2 would like to hire

15. Any completion for the firm side as long as responsiveness is ensured.

both w_3 and w_4. However, if f_2 deviates to accepting w_4 in any subgame following that history, w_3 will apply only to f_1, and f_1 will hire him in return. Therefore, deviating to accepting w_4 before f_1 fulfills its capacity makes f_2 lose w_3 in return, which is less favorable. In other words, since f_2 is forward-looking, it understands the downside of engaging in a block with w_4.

Intuitively, f_2 accepting w_4 starts a rejection cycle similar to Kojima and Pathak (2009). With f_1 starting accepting w_3, w_3 starts rejecting f_2. Both firms prefer μ^U over μ^*, and they can credibly switch w_3 and w_4 in the limit equilibrium by using their remaining capacities as a commitment device. The remaining capacities enable the reward and punishment schemes to be implemented even though the information is restricted to the remaining market.

This difference between one-to-one and many-to-one search models relies on the difference in the underlying static markets. In one-to-one markets, there is no individually rational matching, which is preferred to the firm-optimal stable matching by all firms. In the many-to-one case, the enforced matching μ^U is preferred to μ^* by both firms, and the firms can implement that as long as they can credibly signal each other that blocking pairs will not realize. In return, workers w_3 and w_4 are worse off from being switched by the firms.

Last, but not the least, observe that there is a unique stable matching in the underlying many-to-one market, that is $\mu^* = \{(f_1; w_1, w_2, w_3), (f_2; w_4)\}$. As previously elucidated in the literature review, the uniqueness of the centralized stable outcome usually averts unstable limit equilibrium outcomes, yet, in this particular instance, such prevention does not apply.

To summarize the findings above, in a many-to-one search model, firms can collude strategically, allowing them to implement outcomes more favorable to themselves than the stable matching, even though it may be unique. Unfortunately, such strategic behavior comes at the detriment of workers, who end up worse off as a result. From a policy proposal perspective, the normative question then arises as to whether providing firms with this room for strategic maneuvering is desirable or not. This concern is particularly evident in a setting with colleges and students, where colleges are viewed as public goods, and it becomes imperative to prevent them from developing strategies that benefit themselves at the expense of rendering students worse off. The same applies to the school choice and residency matching problems.

Comparing one-to-one and many-to-one models, we see that enforcing stable matchings with equilibrium strategies is more challenging in a many-to-one search model, whereas unstable matchings are more easily enforceable. These observations emphasize that the role of a central planner becomes essential when it comes to many-to-one matching. The central planner's intervention might be crucial as she can impose a stable allocation that not only ensures fairness but also eliminates any blocking pairs that could lead to undesirable outcomes. By doing so, the central planner might help maintain the integrity of the matching process and protect the interests of all parties involved, promoting a more equitable allocation in many-to-one markets.

1.4 A Related One-to-One Search Model à la Wu

The fact that stable matchings are more challenging to enforce in many-to-one matchings, as well as unstable matchings are more easily enforceable, is particularly striking when we consider the centralized matching literature. We know that when preferences are responsive, many-to-one and one-to-one matching markets can be seamlessly mapped onto each other as demonstrated by Roth and Sotomayor (1992).

The mapping is done via a "related marriage market", which is obtained by replicating each firm by their capacities and treating this new replica market as a one-to-one market. When firms' preferences are responsive in the original market, we know that a many-to-one matching is stable if and only if the corresponding one-to-one matching in the related marriage market is stable. This compelling finding significantly simplifies the analysis of centralized many-to-one matchings.

Propositions 1.3, 1.4, and 1.7 already hint at the many-to-one search model differing fundamentally from its one-to-one counterpart. Furthermore, this difference cannot be eliminated by imposing strict regularity conditions on within-firm preferences (such as additively separable utility). Instead, the fact that the end matching evolves through time allows for strategic behavior among firms based on their remaining capacities.

To gain more concrete insights into the distinctive search behavior in a many-to-one market, it becomes imperative to delineate a corresponding one-to-one search model for comparative analysis. By examining the differences between the many-to-one and one-to-one search models, we can comprehensively understand the unique characteristics and dynamics inherent in each setting. This analytical approach will shed light on the complexities of the many-to-one search models, paving the way for valuable insights into how search behavior evolves in such contexts, which is the objective of this section. To streamline the discussion and prevent redundancy, any aspects left unspecified in this context can be assumed to remain analogous to the original many-to-one search model.

1.4.1 The Related One-to-One Search Game

The *related* one-to-one search model will be a translation of our many-to-one search model onto a one-to-one environment. The related one-to-one market is obtained where each firm is replicated as many times as its capacity. In other words, each seat of the firms is individually present in the search market. Furthermore, the seats are individually searching for workers and hence become competitors.

When we replicate firms, we obtain q_f identical seats. Namely, the set of the seats in the related one-to-one market is $S = \{s_{11}, ..., s_{1q_1}, ..., s_{n1}, ..., s_{nq_n}\}$, where s_{ij} is the jth seat of firm i. Recall that firms' preferences in the original many-to-one market are complete and responsive. This means that we can deduct the firms' preferences over individuals. Each seat has the same preference over individuals as the firm. Regarding the environment, the replication is the same as Roth and Sotomayor (1989). The search model requires an additional adjustment with the contact function.

On the workers' side, there is no replication. Each worker from $W = \{w_1, ...w_m\}$ is still searching for himself. Workers are indifferent between the seats of the same firm, and each worker i prefers a seat in firm j over a seat in firm k if and only if he prefers firm j over firm k in the many-to-one market. Formally, $v(s_{jn}, w_i) < v(s_{km}, w_i)$ whenever $v(f_j, w_i) < v(f_k, w_i)$. For simplicity, from now on, I break the indifferences in workers' preferences such that they prefer the seat with a smaller index within the same firm: $v(s_{j1}, w_i) > ... > v(s_{jq_j}, w_i)$.[16] A related one-to-one market is then the tuple $M_R = (S, W, u, v)$. Similarly, any submarket (remaining market) is $M'_R = (S', W', u, v)$ with $S' \subset S$ and $W' \subset W$.[17]

The contact function is mildly adjusted such that for each submarket, the sum over the probabilities of w and s meeting in M'_R and s is replicated from f of M' is equal to $C(f, w, M')$ in the original market. All other assumptions on the contact function remain the same. Most importantly, all pairs of seats and workers have a positive probability of meeting in line with Wu (2015).

The related one-to-one game is denoted by the tuple $\Gamma_R = (S, W, u, v, C, \delta)$.

Recall that Wu (2015) shows that no unstable matching can be enforced in a private-dinner equilibrium, whereas Proposition 1.7 shows otherwise, even under a more restrictive information criterion. Therefore, the equilibria of one-to-one and many-to-one search models are generically not equivalent. In the following, I will provide a sufficient condition for the underlying many-to-one market that will restore the projectability between many-to-one and one-to-one models and ensure that no unstable matching will be enforced in a private-dinner equilibrium.

Definition 1.9. The pair (s, w) is called a top-pair for any related (sub)market if: Among the seats that find w acceptable, s is the best for w, and among the workers that find s acceptable, w is the best for s.

Following Wu, a related marriage market satisfies the Sequential Preference Condition (SPC) if there is an ordering of the seats and workers and a positive integer k such that:

16. This tie breaking rule reduces the number of stable matchings in the related one-to-one market, yet does not affect the analysis, see Appendix.

17. Any instantaneous matching μ translates onto the one-to-one replica such that $\mu_R(w) = s_{in} \Rightarrow \mu(w) = f_i$.

(1) For any $i \leq k$, (s_i, w_i) is a top-pair in the (sub)market.

(2) Discarding the top-pairs results in a trivial market with non-acceptable pairs.

In a one-to-one market, the top-pairs always mutually accept each other upon meeting. This is almost trivial since the top-pairs cannot hope for a better alternative to be matched in future periods. Once the top-pair leaves the market, the same applies to another because the market satisfies SPC. Since every pair has a positive probability of meeting, there is an equilibrium path that matches the top-pairs, which happens with positive probability, which ensures no unstable matching can be enforced in a limit equilibrium. In the following, I adjust this condition to many-to-one markets.

Definition 1.10. A many-to-one market satisfies the Sequential Preference Condition if the following two conditions hold:

(1) The related one-to-one market satisfies SPC.

(2) Firm preferences are lexicographic for top pairs: If (s_{ij}, w_i) is a top pair in the remaining related market, $u(f_i, \Omega_i) > u(f_i, \Omega_{-i})$ for all (Ω_i, Ω_{-i}) such that $w_i \in \Omega_i$ but $w_i \notin \Omega_{-i}$.

Proposition 1.11. *If the initial many-to-one market satisfies the Sequential Preference Condition, no unstable matching can be enforced in a limit equilibrium.*

Proof. In a market where preferences satisfy SPC, there is a unique matching that allocates top pairs to each other. In the original many-to-one search game, the top pairs would always accept each other (which, on the firm side, is ensured via lexicographic preferences). Since every pair has a positive probability of meeting, there is an equilibrium path in which top pairs meet each other in the SPC ordering, and this happens with positive probability. Therefore, enforced matching cannot be unstable when the preferences of the initial market satisfy SPC. □

Corollary 1.12. *Suppose the initial many-to-one market satisfies the Sequential Preference Condition. In that case, the only enforceable matching of the many-to-one market in a limit equilibrium corresponds to the only enforceable matching of the related one-to-one market in a limit equilibrium.*

Even though we need a modification for the Sequential Preference Condition for many-to-one markets, the result and the proof method apply from Wu (2015). The unique stable set of workers will apply to the firm, and the firm will accept those. Thus, they cannot credibly threaten each other with rejection, resulting in employment upon first meeting. Furthermore, no unstable matching can be enforced in the related market, and a many-to-one matching is stable if and only if the related one-to-one matching is stable in the related market. Consequently, the equilibrium outcomes of both models are equivalent when the contact function is realized according to the SPC ordering, as well as the remaining markets are related. Thus, SPC rebuilds the connection from the many-to-one search model to the one-to-one static market regarding enforced matchings.

Note that by preventing unstable matchings from being enforced, SPC provides a solution to the discussion at the end of Section 1.3.3 about the potential necessity of a central planner for more general preferences to prevent firms from manipulating the search outcome. However, unstable matchings can still arise with positive probability even under SPC, as shown in Wu (2015).

1.5 Connection to Dynamic Stability à la Doval

On the way from centralized matching with clearinghouses that impose an allocation on an economy to a decentralized search model, one natural stepping stone to consider would be the *dynamic stability* concept, introduced by Doval (2022) for one-to-one environments and then later incorporated into many-to-one by Altinok (2022). In this section, I compare the search outcomes to dynamically stable matchings. Since we are in a many-to-one environment, the definitions and examples below are based on Altinok (2022).

Intuitively, the dynamic stability concept also incorporates the time component, or at least the sequential acceptance of agents, even though waiting is almost costless. Formally, what differs from the current model is that all employers are around in all periods, whereas candidates arrive over time and W_t denotes the workers that arrive at period t, and the agents form matches over exogenously given T periods, $t = 1, 2, ..., T$.

For the rest of this section, I am restricting attention to $T = 2$, that is, workers arrive over two periods. Then, a history is the empty set for the first period and the 1st-period matching in period 2, and the strategies map histories into the matchings of the same period.

Definition 1.13. A t-period matching μ_t is a mapping from the set of candidates that have arrived until t to the set of employers; that is, for each t,

$$\mu_t : \bigcup_{\tau=1}^{t} W_t \to F \cup \{\emptyset\}$$

and satisfies the following properties:

(1) Capacity constraints are always respected: $|\mu_t^{-1}(f)| \leq q_f$ for each f for $t = 1, 2$.
(2) First-period matchings are irreversible: $\mu_2(w) = \mu_1(w)$ if $\mu_1(w) \neq \emptyset$,

Similarly, M_t is the set of t—period matchings for $t = 1, 2$, $h_t := (h_\tau)_1^t$ a history of matchings at t where H_t denotes all possible period-t histories, where $H_1 = \emptyset$ and $H_2 = M_1$. A strategy profile is then (s_1, s_2), where $s_t : H_t \to M_t$ for $t = 1, 2$. The second period is identical to a static market, and a first-period block refers to a blocking coalition that exists in the first period and that can implement a better second-period matching for them by forming a coalition in the first period.

Consider the following example by Altınok to illustrate the difference between the static stability vs the dynamic one:

Example 1.14. Suppose there are two firms $F = \{f_1, f_2\}$ such that $q_1 = 3, q_2 = 2$ and six workers $W = \{w_1, ..., w_6\}$ and the preferences are as below:

$$\begin{aligned} w_1 &: f_1 \succ f_2 \\ w_2 &: f_1 \succ f_2 \\ w_3 &: f_1 \succ f_2 \\ w_4 &: f_1 \succ f_2 \\ w_5 &: f_1 \succ f_2 \\ w_6 &: f_2 \succ f_1 \end{aligned} \qquad \begin{aligned} f_1 &: w_6 \succ w_1 \succ w_2 \succ w_3 \succ w_4 \succ w_5 \\ f_2 &: w_1 \succ w_2 \succ w_3 \succ w_6 \end{aligned}$$

In addition, suppose $\{w_4, w_5, w_6\} \succ \{w_1, w_2, w_3\}$ for f_1, which is still in line with responsiveness but depicts that f_1 has *extreme preferences*, in the sense that it prefers combining extreme workers rather than the average ones.

In this market, the unique stable matching is $\mu^* = \{(f_1; w_1, w_2, w_3), (f_2; w_6), (\emptyset, w_4), (\emptyset, w_5)\}$, which in fact is not dynamically stable in this particular market: Suppose instead of everybody being in the market at once, workers arrive in 2 periods, such that workers w_4 and w_5 arrive in period 1. Then, f_1 would form a period-1 matching with w_4 and w_5 (1st period block), enters the second-period with $q_1' = 1$. The unique stable μ in period 2: f_1 matches with w_6, f_2 is matched with $\{w_1, w_2\}$, so the dynamically stable matching is $\mu^* \neq \mu^D = \{(f_1; w_4, w_5, w_6), (f_2; w_1, w_2), (\emptyset, w_3)\}$

Both firms prefer the outcome μ^D to μ^*, which is not stable because of the blocking pair (f_1, s_1). What happens with dynamic stability is that the firms in a sense exchange w_6 with w_1, w_2. To make this exchange credible, f_1 fills its capacity with w_4, w_5 in the first period.

The straightforward intuition places dynamic stability between centralized and completely decentralized search models, with the number of periods given and the central planner still imposing the matching on the economy somehow. Furthermore, the dynamic stability concept and the decentralized search model share a similar flavor in the strategic commitment and manipulation room they provide to firms with multiple capacities. Nevertheless, the following proposition will show that (contrary to the connection to completely centralized markets) dynamically stable matchings may not be enforced as search equilibria.

Proposition 1.15. *Dynamically stable matchings may not be enforceable by limit equilibria.*

Proof. Take the example by Altinok (2022) one more time, and suppose the dynamically stable $\mu^D = \{(f_1; w_4, w_5, w_6), (f_2; w_1, w_2), (\emptyset, w_3)\}$ is enforced in a limit equilibrium. Since every meeting has a positive probability in the decentralized search model, suppose f_1 and w_6 meet the first day. If w_6 applies to f_1: If f_1 accepts, firms have the same preferences over remaining workers, f_1 employs $\{w_1, w_2, w_6\}$. The best f_1 can hope for: f_1 accepts w_6. Therefore, w_6 applies to f_1 even with small waiting costs upon meeting. Since every pair has a positive probability of meeting in the initial market, $\mu^D = \{(f_1; w_4, w_5, w_6), (f_2; w_1, w_2), (\emptyset, w_3)\}$ cannot be enforced by limit equilibria. □

In fact, as the proof of this proposition already suggests, the dynamic stability concept depends heavily on which workers arrive in which period, which resembles the meeting probability being exactly 0 and exactly 1 for some of the candidates. The concept is not robust to more general probabilities such as that are defined by the contact function, as well as the workers becoming strategic agents as well.

1.6 Conclusion

In this paper, I describe and analyze a finite decentralized many-to-one bilateral search model. The finite search model is the many-to-one counterpart of Wu (2015) that allows firms to employ more than one worker. The outcomes of subgame perfect equilibria are compared to the stable matchings of the underlying many-to-one market. Unlike the one-to-one finite search model, the stable matchings of the underlying market may not be enforceable by simple strategies. On the other hand, when one side of the market has vertical preferences, there is a Markovian strategy profile that enforces the unique stable matching.

Unlike the centralized matching models, the many-to-one search model presents another fundamental distinction in its projection onto one-to-one environments. In one-to-one settings, enforcing unstable matchings necessitates the application of reward and punishment schemes, which is only feasible once agents can observe the details of failed meetings. On the contrary, in the many-to-one search model, firms hold the capacity to implement reward and punishment strategies through their remaining capacities, enabling them to commit to credible strategies that yield a more advantageous matching outcome for themselves compared to the stable matching. This capability stems solely from the fact that firms possess multiple capacities, making it impractical to attempt resolving the manipulation incentives on the firm side by imposing stringent assumptions on within-firm preferences, such as additively separable utility.

The transition from one-to-one to many-to-one search models presents new challenges in enforcing stable matchings and the relative ease of enforcing unstable ones. This underscores the significance of employing a centralized clearinghouse in many-to-one markets such as student admissions and labor markets compared to their one-to-one counterparts.

Dynamically stable matchings may not be enforceable in limit equilibria either. Indeed, the concept hinges on the specific sequence of worker arrivals over time, which is not necessarily robust to non-binary meeting probabilities.

Nevertheless, when preferences of all firms satisfy the sequential preference condition, i.e., there are top pairs of firms and workers that mutually prefer each other over all other available options, the unique stable matching is the only enforceable matching in a limit equilibrium. Furthermore, the related matching in the related one-to-one market is the only enforceable outcome in any limit equilibrium in a one-to-one search model à la Wu (2015), and the instantaneous matchings of many-to-one and one-to-one models are equivalent.

The paper investigates the intricacies of finite decentralized many-to-one matching, offering a comprehensive understanding of the underlying dynamics that govern agents' behaviors. Notably, showing how many-to-one matching markets differ from their one-to-one counterparts when considered in a search model bridges the gap between centralized many-to-one and decentralized one-to-one models. Discrepancies are present with both models, even though firm preferences are responsive and agents are sufficiently patient. In general, the presence of multiple capacities within this framework introduces opportunities for strategic manipulation by firms, either by sacrificing some vacancies or through collusion. Nonetheless, it is worth noting that the characterization of matchings that can be enforced through limit equilibria remains a topic that necessitates further exploration and investigation in future research.

Appendix 1.A Omitted Proofs

Theorem 1.6:

Proof.

- σ^* enforces the stable μ^*: μ^* obtains almost surely on the equilibrium path. For any meeting function realization, another outcome than μ^* arising from σ^* has a probability of 0. Suppose f and w are not matched under μ^* but they end up together under σ^* for some meeting function realization. Mutual acceptance requires $u(f,w) \geq \min(u(f,w'))$ such that $w' \in \mu(f)$ and $v(f,w) \geq v(\mu^*(w),w)$ which contradicts with μ^* being stable under responsive preferences. This part concludes all pairs in the outcome are consistent with μ^*.
  Also, note that all existing agents meet with some positive probability. Since the pairs under μ^* accept each other, and the probability of a rejecting pair (or a rejection pair cycle) occurring has a probability of 0, μ^* obtains almost surely on the equilibrium path.
- σ^* constitutes a limit equilibrium of any search game. For any firm f, the expected utility of σ^* is $u(f,\mu^*(f))$ and for any worker w the expected utility of σ^* is $v(\mu^*(w),w)$. Note that a one-step deviation to reject a partner that is accepted under μ^* does not change the submarket, therefore we are still on-the-equilibrium-path. The only deviation by an agent that would yield a switch to off-path is accepting a partner from the other side who would not be accepted μ^*.

  Now consider a one-step deviation by w such that w accepts f instead of rejecting as under σ^*. Rejection under σ^* implies $v(f,w) < v(\mu^*(w),w)$. If f rejects w, the subgame does not change and the expected utility of w does not change. If f accepts, w receives a lower utility than $v(\mu^*(w),w)$. Therefore, it is not beneficial for w to accept f.

  A similar logic applies to f and a one-step deviation towards rejecting, even though f has multiple capacities. We know that when one side has vertical preferences, the deferred-acceptance outcome is the same as the serial dictatorship outcome. Suppose firms have the same preferences over individual workers, i.e. we are in a college admissions model. Then, there is a unique stable matching mu^* that can be achieved both by firm-proposing DA and the worker-proposing DA, as well as the serial dictatorship where the workers are ranked according to the vertical preferences of firms.

  Furthermore, this applies to any submarket of the initial market. In any subgame of the initial game that is on-the-equilibrium-path, the expected utility under μ^* is $u(f,\mu^*(f))$. Consider a one-shot deviation where f has already hired $\Omega \subset W$ and accepts a worker w that has applied to f and that it prefers less than its least preferred stable partner: i.e: $u(f,w) < \min(u(f,w'))$ such that

$w' \in \mu^*(f)$. Denote the subgame where f rejects w as Γ and the one with the deviation Γ'. In Γ under μ^*, $q'_f = q_f - |\Omega|$, and the firm receives $u(f, \mu^*(f))$, where $\mu^*(f) = \Omega \cup \Omega'$.

If $\Omega' = \emptyset$, either f has filled its capacity or $u(f, w) < 0$, ensuring $u(f, \mu^*(f) \cup \{w\}) < u(f, \mu^*(f))$ in either case. If $\Omega' \neq \emptyset$, the first $q' - 1$ workers that f employs under Γ and Γ' are the same by the serial dictatorship representation, and the end outcome differs only by one worker. Denote the worker that is employed under Γ but not under Γ' as w^*. The expected utility with one-shot deviation is $u(f, \Omega \cup \Omega' \setminus \{w^*\} \cup \{w\})$. We know that $u(f, w) < \min(u(f, w'))$ such that $w' \in \mu^*(f)$, therefore $u(f, w) < u(f, w^*)$. With responsive preferences, $u(f, \Omega \cup \Omega' \setminus \{w^*\} \cup \{w\}) < u(f, \mu^*(f))$, ensuring the one-shot deviation being not profitable.

Verbally, any deviation to accepting a worker that is not acceptable under μ^* only changes that worker with one of the stable workers. With responsive preferences, it is ensured that the difference between the specific workers is reflected in the overall preference, preventing deviation.

□

Proposition 1.3: How can we restrict attention to end-outcomes? 12

Take the same example from the Proposition, and once again, suppose there is a Markovian strategy profile σ^* that is a limit equilibrium and enforces the WOSM in any many-to-one search game. Under σ^*, f_1 either meets with w_1 first and then w_4, depending on the contact function realization. Since σ^* enforces WOSM, the expected utility gain when $\mu_h(f_1) = \emptyset$ is $u(f_1, \{w_1, w_4\})$. When f_1 deviates to accepting w_5, it first gets $u(f_1, w_5)$. Then, since σ^* is an equilibrium in every subgame by assumption, the expected utility gain is the additional utility from adding w_3. Therefore, the overall utility from employing w_5 is $U = u(f_1, w_5) + \delta^{tau}(u(f_1, \{w_3, w_5\}) - u(f_1, w_5))$ (τ referring to the expected day the match will conclude), which converges to $u(f_1, \{w_3, w_5\})$ as δ converges to 1. The same logic applies to other propositions.

More on Equivalent Equilibria of Many-to-One and One-to-One

Proposition 1.11 can be analyzed with a different perspective for related markets with indifferences, in which ties between seats are not broken. In that case, SPC ensures that the equilibria of the original many-to-one and the related one-to-one market are *equivalent*. Intuitively, the equilibria of the many-to-one search model and its related one-to-one search model are equivalent if the workers adapt the same acceptance strategy for the seats of a firm as the strategy they use for the firm, and the firms' seats use the same acceptance strategy for each worker as the firm itself.

Definition 1.16. For any many-to-one market and its related market, the equilibria of the many-to-one search game and its related one-to-one search game are equivalent if under the same information restriction and for every related subgame:

(1) Workers of the related market accept the seats that belong to the firms they apply to in the original market and reject others.
(2) Seats of the related market accept the same workers as their mother firm.

The following lemma additionally shows that we can track equilibrium equivalence from remaining markets, even though we cannot observe agents' strategy profiles.

Lemma 1.17. *For each realization of the contact function C, the equilibria of the many-to-one search model and its related one-to-one search model are equivalent if and only if the remaining markets are related for each history.*

Proof. Let C be any realization of the contact function. I will prove the lemma by proving the if statements from both directions.

(1) The equilibria of many-to-one search and the related one-to-one search are equivalent $\Rightarrow$ remaining markets are related for each history.
Easily proven by induction. Start with the initial market M. Equivalent acceptance strategies imply:

$$f \text{ accepts } w \iff s \text{ accepts } w$$
$$w \text{ accepts } f \iff w \text{ accepts } s$$

This means, for the same realized related contact function, M' after the first day is the same. Apply this to every step, the first part of the lemma concludes.

(2) Remaining markets are related for each history $\Rightarrow$ The equilibria of both search models are equivalent.
Suppose s is a seat of f. If when s, w and f, w meet after h at the related remaining markets, and the remaining market after this is also the same s,f use the same acceptance strategies.
If this holds for each remaining market and history, the equilibria are equivalent.

□

The above lemma establishes the equivalence between the equilibrium strategies and the remaining markets. Recall that Wu (2015) shows that no unstable matching can be enforced in a private-dinner equilibrium, whereas Proposition 1.7 shows otherwise, even under a more restrictive information criterion. Therefore, the equilibria of one-to-one and many-to-one search models are generically not equivalent, which, on the other hand, is ensured if the initial market satisfies the sequential preference condition.

Appendix 1.B Matchings of the Centralized Market à la Search Game

This subsection takes a quick detour into the existing literature of many-to-one matchings and describes the adaptations that will translate the existing results into a many-to-one search environment. In order to find the set of stable matchings, a linear programming approach is developed. As Vate (1989) and Rothblum (1992) characterize stable matchings in a marriage market as extreme points of a convex polytope, Baïou and Balinski (2000) extend the results to a many-to-one matching market. They show that simply replacing the parameters of a marriage market with their many-to-one counterparts does not extend the results of Rothblum, but needs a slight differentiation. Neme and Oviedo (2021) adapt their approach as well, so am I:

Given a matching μ, an assignment matrix $x^B \in \mathbb{R}^{|F|\times|W|}$ (B for Baïou and/or Balinski) is defined where all its elements are denoted by $x^B(f,w)$ where $x^B(f,w) \in \{0,1\}$ and $x^B(f,w) = 1$ if and only if $\mu(w) = f$.

Following Baïou and Balinski, let *CP* denote the convex polytope generated by the following linear inequalities:

$$\sum_{j\in W} x^B_{f,j} \leq q_f \qquad \forall f \in F \tag{1.B.1}$$

$$\sum_{i\in F} x^B_{i,w} \leq 1 \qquad \forall w \in W \tag{1.B.2}$$

$$x^B_{f,w} \geq 0 \qquad \forall (f,w) \in F \times W \tag{1.B.3}$$

$$x^B_{f,w} = 0 \qquad \text{for unacceptable pairs } (f,w) \tag{1.B.4}$$

The integer solutions to (1.B.1)-(1.B.3) are assignment matrices of simple many-to-one matchings. A matching, where some entries $x^B(f,w)$ are non-integers in the interval $(0,1)$ is called a fractional matching. We can interpret the fractional matchings as probabilities that the agents are matched to one another as well as the timeshares the respective agents spend with each other.

An example of a many-to-one assignment matrix with 2 firms $\{f_1, f_2\}$ and 2 workers $\{w_1, w_2\}$ would look like as follows:

	f_1	f_2
w_1	$x^B(f_1,w_1)$	$x^B(f_2,w_1)$
w_2	$x^B(f_1,w_2)$	$x^B(f_2,w_2)$

where all the entries are nonnegative and:

$$\sum_F x^B(f,w) \leq 1 \quad \text{for both workers}$$
$$\sum_W x^B(f,w) \leq q_f \quad \text{for both firms}$$

Adding (1.B.4) imposes the individual rationality constraint, that the match is at least as good as the outside option. As Baïou and Balinski show, adding another linear inequality to the *CP* system:

$$\sum_{u(f,j)>u(f,w)} x^B_{f,j} + q_f \sum_{v(i,w)>v(f,w)} x^B_{i,w} + q_f x^B_{f,w} \geq q_f \qquad \forall (f,w) \in A \tag{1.B.5}$$

defines the stable convex polytope *SCP* and the integer solutions to *SCP* define stable simple matchings, which are individually rational and pairwise stable.

In Kojima and Manea (2010) and Kesten and Ünver (2015), "The School-Choice Birkhoff-von Neumann Theorem states that any fractional matching can be represented as a lottery (not necessarily unique) over simple matchings", which allows us to interpret a fractional matching in a third way. Nevertheless, the intuition about the stable matchings turns out to be incorrect and the non-integer solutions of inequalities (1.B.1)-(1.B.5) do not immediately give us stability when it comes to fractional matchings.

In fact, as shown by Baïou and Balinski Baïou and Balinski (2000) and elaborated further in Neme and Oviedo Neme and Oviedo (2021), the non-integer solutions to the *SCP* might be blocked in a *fractional way*, by a firm and worker, who want to increase their timeshare together, at the expense of those they like less at a non-integer solution to *SCP*. Formally:

Definition 1.18. A matching is blocked by the firm-worker pair (f,w) in a fractional way, when $x^B(f,w) < 1$, $v(f,w) > v(f',w)$ for some f' such that $x^B(f',w) > 0$ and $u(f,w) > u(f,w')$ for some w' such that $x^B(f,w') > 0$.

Example: An example from Baïou and Balinski (2000) and Neme and Oviedo (2021) which shows an assignment matrix, which is a solution to *SCP* and blocked in a fractional way by is as follows:

	f_1	f_2
w_1	1	0
w_2	0.5	0.5
w_3	0.5	0.5
w_4	0	1

where the preferences (over individuals, derived from the preferences over sets) are such that:

$$\begin{aligned}
f_1 &: u(f_1,w_1) > u(f_1,w_2) > u(f_1,w_3) > u(f_1,w_4) \quad \text{and} \quad q_{f_1} = 2\\
f_2 &: u(f_2,w_4) > u(f_2,w_3) > u(f_2,w_2) > u(f_2,w_1) \quad \text{and} \quad q_{f_2} = 2\\
w_1 &: v(f_2,w_1) > v(f_1,w_1)\\
w_2 &: v(f_2,w_2) > v(f_1,w_2)\\
w_3 &: v(f_2,w_3) > v(f_1,w_3)\\
w_4 &: v(f_1,w_4) > v(f_2,w_4)
\end{aligned}$$

In the example above, it can easily be checked that the numbers solve the linear problem *SCP*. However, w_3 likes f_2 better than f_1, but his time is shared equally between the firms. In addition, f_2 likes w_3 better than w_2 but one seat is shared equally between those workers. In such a case, the pair (f_2, w_3) blocks the assignment above in a fractional way so that they can increase their time spent together in exchange for their other partners in the matching, f_1, and w_2.

As it is not mentioned in either of the papers, the lottery interpretation of the fractional matchings helps us understand the underlying misfunction in this example. Although the lottery over simple matchings which represents a fractional matching is generically not unique, in this example it actually is unique. The fractional matching is a lottery over the following two simple matchings with equal probability 0.5:

	f_1	f_2
w_1	1	0
w_2	1	0
w_3	0	1
w_4	0	1

	f_1	f_2
w_1	1	0
w_2	0	1
w_3	1	0
w_4	0	1

The reason why there is a blocking pair in a fractional way can be observed in the lottery as well. Although the first simple matching of the lottery is stable, the second one is not. Furthermore, the fact that it is not stable pins down the fractional blocking pair: The second simple matching is blocked by the pair (f_2, w_3).

Neme and Oviedo (2021) refer to matchings in which there are no incentives to block (neither as in the usual way nor in the fractional interpretation) as *strong stable matchings* and prove that they can be found adding the additional constraint to *SCP*:

Definition 1.19. Let (F, W, q, u, v) be a many-to-one matching market. A fractional matching is strongly stable if for each acceptable pair (f, w), x satisfies the strong stability condition:

$$\left[q_f - \sum_{u(f,j)\geq u(f,w)} x^B_{f,j}\right] \cdot \left[1 - \sum_{v(i,w)\geq v(f,w)} x^B_{i,w}\right] = 0 \tag{1.B.6}$$

Observe that the assignment matrix of a simple stable matching fulfills (1.B.6). This follows from the simple fact that if f and w are matched, the second multiplier is 0. If they are not matched with each other, at least one of them is consuming its own capacity.

For fractional matchings, when (1.B.6) does not hold for some (f,w), $q_f > \sum_{u(f,j)\geq u(f,w)} x^B_{f,j}$ and $1 > \sum_{v(i,w)\geq v(f,w)} x^B_{i,w}$, it means that there are f' and w' such that $u(f,w) > u(f,w')$, $v(f,w) > v(f',w)$ and $x(f,w') > 0$, $x(f',w) > 0$ and $x(f,w) < 1$. In such a scenario, the (f,w) would block the assignment and increase their time shared together.

Insightful Theorem 1 from Neme and Oviedo (2021) concludes: "If x^B is a strongly stable fractional matching, it can be represented as a convex combination of stable matchings. Furthermore, a lottery over simple stable matchings is strongly stable as well." This establishes the lottery interpretation as in Lauermann and Nöldeke (2014).

Both Baïou and Balinski (2000) and Neme and Oviedo (2021) constructed the matching as an assignment matrix with two sides in rows and columns, respectively. This approach is quite useful for visualizing and pointing out simple stable matchings. However, the structure in Lauermann and Nöldeke (2014) requires a different approach. The main difference is that the many-to-one papers of Baïou and Balinski and Neme and Oviedo use an ordinal utility approach, whereas Lauermann and Nöldeke employ a cardinal utility in their model, which allows them to calculate the expected utilities for the agents as well. As discussed before, cardinal utility adaptation is vital for a search structure.

If $u(f,\Omega)$ is a linear function of individual values, i.e. additively separable such that $u(f,\Omega) = \sum_{i\in\Omega} u(f,w_i)$, we could still use the assignment matrix with the agents in rows and columns to calculate expected utilities. In order to employ such a structure in a many-to-one framework and be able to calculate expected utilities at the same time, we would need a different assignment matrix $x \in \mathbb{R}^{|F|\times|2^W|}$, satisfying:

$$\sum_{F}\sum_{\substack{\Omega\subset 2^W\\ w\in\Omega}} x(f,\Omega) \leq 1 \qquad \forall w \tag{1.B.7}$$

$$\sum_{\Omega\subset 2^W} x(f,\Omega) \leq 1 \qquad \forall f \tag{1.B.8}$$

$$x(f,\Omega) = 0 \qquad \forall |\Omega| > q_f \tag{1.B.9}$$

$$x(f,\Omega) \geq 0 \qquad \forall (f,\Omega) \in F\times 2^W \tag{1.B.10}$$

The above-described assignment matrix has firms in the columns and all possible subsets of the workers in the rows. In a many-to-one matching, a specific worker cannot be matched to two firms. Hence, considering a worker would now require considering each subset that this worker appears in.

The expected utilities from a matching x then can be calculated as follows:

$$U(f;x) = \sum_{\Omega} x(f,\Omega)u(f,\Omega)$$
$$V(w;x) = \sum_{F}\sum_{\substack{\Omega\subset 2^W\\ w\in\Omega}} x(f,\Omega)v(f,w)$$

Example: An assignment matrix of a many-to-one matching market with 2 workers $\{w_1,w_2\}$ and 2 firms $\{f_1,f_2\}$ would be as follows according to the latter description:

	f_1	f_2
$\{w_1,w_2\}$	$x(f_1,\{w_1,w_2\})$	$x(f_2,\{w_1,w_2\})$
w_1	$x(f_1,w_1)$	$x(f_2,w_1)$
w_2	$x(f_1,w_2)$	$x(f_2,w_2)$

where all the entries are nonnegative and:

$$x(f_1,\{w_1,w_2\}) + x(f_1,w_1) + x(f_2,\{w_1,w_2\}) + x(f_2,w_1) \leq 1$$
$$x(f_1,\{w_1,w_2\}) + x(f_1,w_2) + x(f_2,\{w_1,w_2\}) + x(f_2,w_2) \leq 1$$
$$x(f,\{w_1,w_2\}) + x(f,w_1) + x(f,w_2) \leq 1 \quad \text{for both firms}$$
$$x(f,\Omega) = 0 \quad \text{if} \quad |\Omega| > q_f \quad \text{for all subsets}$$

With expected utilities:

$$U(f;x) = x(f,\{w_1,w_2\})u(f,\{w_1,w_2\}) + x(f,w_1)u(f,w_1) + x(f,w_2)u(f,w_2)$$
$$V(w_1;x) = \sum_{f\in F}[x(f,\{w_1,w_2\}) + x(f,w_1)]v(f,w_1)$$
$$V(w_2;x) = \sum_{f\in F}[x(f,\{w_1,w_2\}) + x(f,w_2)]v(f,w_2)$$

Proposition 1.20. *The assignment matrix described before by Baïou-Balinski and Neme-Oviedo, which has the workers instead of sets of workers in the rows can easily be calculated from the matrix described above, by setting* $x^B(f,w) = \sum_{\substack{\Omega\subset 2^W \\ w\in\Omega}} x(f,\Omega)$.

Similarly, the entries of the assignment matrix x *can be calculated if all the entries of* x^B *are integers, i.e.* $x^B(f,w) = \{0,1\}$ $\forall (f,w)$*. This does not necessarily hold if some entries of* $x^B \in (0,1)$.

Proof. The first part of the lemma is trivial with the given equality. For the second part, the example below illustrates the calculation of the assignment matrix x from x^B for simple matchings. Furthermore, the second part of the example serves as a proof that x calculated from x^B is not necessarily unique for fractional matchings.

Example: Consider again the example of a many-to-one matching market with 2 workers $\{w_1, w_2\}$ and 2 firms $\{f_1, f_2\}$. In the first scenario, let us take a simple matching with all entries are either 0 or 1. In that case, if $u(f,H)$ and $v(h,w)$ are known, the expected utilities can easily be calculated because the assignment matrix x^B implies a unique x.

x^B	f_1	f_2
w_1	1	0
w_2	1	0

x	f_1	f_2
$\{w_1, w_2\}$	1	0
w_1	0	0
w_2	0	0

However, if x^B is a fractional matching, the corresponding x is not necessarily unique. Under different x representations of x^B, the expected utility of the workers will be equal, whereas the expected utilities of the firms might differ. There is an example of a fractional matching x^B with multiple x representations below, where both x_1 and x_2 imply x^B, which serves as a proof to the Proposition 1.20 above.

x^B	f_1	f_2
w_1	0.5	0.5
w_2	0.3	0.4

x_1	f_1	f_2
$\{w_1, w_2\}$	0.3	0
w_1	0.2	0.5
w_2	0	0.4

x_2	f_1	f_2
$\{w_1, w_2\}$	0.1	0
w_1	0.4	0.5
w_2	0.2	0.4

□

When we consider the finite decentralized many-to-one search game, the acceptance decisions of the agents represent a stopping agreement. Therefore, we need to be able to calculate the expected utilities from continuing the search and compare them to the gains from an immediate acceptance decision. As the next subsection shows, this new definition of a matching matrix will enable us to do such comparisons.

From Equilibria to Assignment Matrices

In pursuance of the analysis of how the subgame perfect equilibria of the many-to-one search model relate to the stable matchings of the centralized many-to-one market, I will describe the assignment matrix that can be obtained from the search equilibrium. In fact, any equilibrium of the search model implies an assignment matrix. Intuitively, the assignment matrix will show the probability of a firm and a subset of workers being matched in equilibrium. In the following revisions, this methodology will be implemented to show whether equilibrium matchings are stable.

In order to calculate the assignment matrix from an equilibrium of the search game, we look at the terminal histories. For any finite terminal history $h \in \mathscr{H} \setminus \hat{\mathscr{H}}$ that takes T periods, let $C_h^t(f, w, M_h^t) \in \{0, 1\}$ denote the meeting function realization for any $t \in \{0, ..., T\}$, where M_h^t is the remaining market implied by μ_h^t, the instantaneous matching in the beginning of period t. [18] The meeting function realization is such that $C_h^t(f, w, M_h^t) = 1$ for firm f and worker w who meet at period t along h and $C_h^t(f, w, M_h^t) = 0$ for all other pairs. The pair (f, w) such that $C_h^t(f, w, M_h^t) = 1$ will be referred to as i_h^t, since they are the agents of the stage game.

Similarly, $a_h^t(i_h^t, \mu_h^t) \in \{A, R\}^2$ denotes the action realization for i_h^t.[19] After the action profile of t realizes, the instantaneous matching is updated to μ_h^{t+1} and remaining market is M_h^{t+1}. Since the market does not change if any of the parties reject, $M_h^{t+1} = M_h^t$ unless $a_h^t(i_h^t, \mu_h^t) = (A, A)$. If both parties accept, the worker leaves the market and the firm leaves the market if the capacity is full.

With this formulation, we can now calculate the probability of a terminal history h occurring on the equilibrium path of the search game. The probability of meeting is simply determined by the choice function, $\mathbb{P}(C_h^t(f, w, M_h^t) = 1) = C(f, w, M_h^t)$.

In order to reach day 1 on the equilibrium path of h, the agents who meet on day 0 should be aligned with h and they should decide accordingly as well. The probability of reaching day 1 under history h, denoted by $\mathbb{P}(\mu_h^1)$ and satisfies $\mathbb{P}(\mu_h^1) = C(i_h^0, M_h^0)\mathbb{P}(a_h^0(i_h^0, \mu_h^0))$. By induction, the probability of reaching any day $t \le T$ along history h can be calculated by multiplying the probability of agents meeting and behaving according to the history along the equilibrium path of h:

18. Clearly, $M_h^0 = M$ for any history.
19. Recall that we restrict attention to pure strategies.

$$\mathbb{P}(\mu_h^t) = \prod_{k=1}^{t} C(i_h^{k-1}, M_h^{k-1})\mathbb{P}(a_h^{k-1}(i_h^{k-1}, \mu_h^{k-1}))$$

Subsequently, for any given many-to-one search game $(F, W, q, u, v, C, \delta)$ once the equilibrium acceptance strategies of the agents are calculated, we can restrict attention to terminal histories and easily calculate the probability of any f, Ω being matched at the end of the game by simply adding up the probabilities of different terminal histories at the end of which f and Ω are together. By simply taking this probability equal to $x(f, \Omega)$, a matching matrix can be constructed.[20]

20. The game ends almost surely, and it can easily be shown that this matrix satisfies the properties of a many-to-one matching.

References

Adachi, Hiroyuki. 2003. "A search model of two-sided matching under nontransferable utility." *Journal of Economic Theory* 113 (2). https://doi.org/10.1016/S0022-0531(03)00085-1. [10]

Alcalde, José. 1996. "Implementation of stable solutions to marriage problems." *Journal of Economic Theory* 69 (1). https://doi.org/10.1006/jeth.1996.0050. [10]

Alcalde, José, David Pérez-Castrillo, and Antonio Romero-Medina. 1998. "Hiring Procedures to Implement Stable Allocations." *Journal of Economic Theory* 82 (2). https://doi.org/10.1006/jeth.1997.2447. [9]

Alcalde, José, and Antonio Romero-Medina. 2000. "Simple Mechanisms to Implement the Core of College Admissions Problems." *Games and Economic Behavior* 31 (2). https://doi.org/10.1006/game.1999.0743. [9]

Alcalde, José, and Antonio Romero-Medina. 2005. "Sequential decisions in the college admissions problem." *Economics Letters* 86 (2). https://doi.org/10.1016/j.econlet.2004.06.013. [9]

Altinok, Ahmet. 2022. "Dynamic Matching: Shall we commit to not competing?" *SSRN Electronic Journal,* https://doi.org/10.2139/ssrn.4029179. [8, 10, 27, 29]

Baïou, Mourad, and Michel Balinski. 2000. "The stable admissions polytope." *Mathematical Programming, Series B* 87 (3). https://doi.org/10.1007/s101070050004. [34, 35, 37]

Balinski, Michel, and Tayfun Sönmez. 1999. "A Tale of Two Mechanisms: Student Placement." *Journal of Economic Theory* 84 (1). https://doi.org/10.1006/jeth.1998.2469. [7]

Becker, Gary S. 1973. "A Theory of Marriage: Part I." *Journal of Political Economy* 81 (4). https://doi.org/10.1086/260084. [10]

Blum, Yosef, Alvin E. Roth, and Uriel G. Rothblum. 1997. "Vacancy Chains and Equilibration in Senior-Level Labor Markets." *Journal of Economic Theory* 76 (2). https://doi.org/10.1006/jeth.1997.2307. [10]

Burdett, Ken, and Melvyn G. Coles. 1997. "Marriage and class." *Quarterly Journal of Economics* 112 (1). https://doi.org/10.1162/003355397555154. [10]

Chade, Hector, Jan Eeckhout, and Lones Smith. 2017. "Sorting through Search and matching models in economics." *Journal of Economic Literature* 55 (2). https://doi.org/10.1257/jel.20150777. [10]

Doval, Laura. 2022. "Dynamically stable matching." *Theoretical Economics* 17 (2). https://doi.org/10.3982/te4187. [8, 10, 27]

Eeckhout, Jan. 1999. "Bilateral search and vertical heterogeneity." *International Economic Review* 40 (4). https://doi.org/10.1111/1468-2354.00045. [10]

Gale, D., and L. S. Shapley. 1962. "College Admissions and the Stability of Marriage." *American Mathematical Monthly* 69 (1). https://doi.org/10.1080/00029890.1962.11989827. [6, 9]

Haeringer, Guillaume, and Myrna Wooders. 2011. "Decentralized job matching." *International Journal of Game Theory* 40 (1). https://doi.org/10.1007/s00182-009-0218-x. [9]

Hatfield, John William, and Scott Duke Kominers. 2015. "Hidden Substitutes," https://doi.org/10.1145/2764468.2764469. [10]

Kesten, Onur, and M. Utku Ünver. 2015. "A theory of school-choice lotteries." *Theoretical Economics* 10 (2). https://doi.org/10.3982/te1558. [35]

Kojima, Fuhito, and Mihai Manea. 2010. "Incentives in the probabilistic serial mechanism." *Journal of Economic Theory* 145 (1). https://doi.org/10.1016/j.jet.2009.09.002. [35]

Kojima, Fuhito, and Parag A. Pathak. 2009. "Incentives and stability in large two-sided matching markets." *American Economic Review* 99 (3). https://doi.org/10.1257/aer.99.3.608. [23]

Konishi, Hideo, and M. Utku Ünver. 2006. "Games of capacity manipulation in hospital-intern markets." *Social Choice and Welfare* 27 (1). https://doi.org/10.1007/s00355-006-0097-z. [7, 9, 18]

Lauermann, Stephan, and Georg Nöldeke. 2014. "Stable marriages and search frictions." *Journal of Economic Theory* 151 (1). https://doi.org/10.1016/j.jet.2013.11.001. [10, 37]

Mortensen, Dale T. 1982. "Property Rights and Efficiency in Mating, Racing, and Related Games." *American Economic Review* 72 (5): 968–79. Accessed April 14, 2024. http://www.jstor.org/stable/1812016. [10]

Neme, Pablo, and Jorge Oviedo. 2021. "On the set of many-to-one strongly stable fractional matchings." *Mathematical Social Sciences* 110. https://doi.org/10.1016/j.mathsocsci.2020.12.002. [34–37]

Niederle, Muriel, and Leeat Yariv. 2009. "Decentralized matching with aligned preferences." *NBER Working paper series* 14840. https://doi.org/10.3386/w14840. [9]

Pais, Joana. 2008. "Incentives in decentralized random matching markets." *Games and Economic Behavior* 64 (2). https://doi.org/10.1016/j.geb.2007.12.005. [9]

Pissarides, Christopher A. 1985. "Short-Run Equilibrium Dynamics of Unemployment, Vacancies, and Real Wages." *American Economic Review* 75 (4): 676–90. Accessed April 14, 2024. http://www.jstor.org/stable/1821347. [10]

Reischmann, Tobias, Thilo Klein, and Sven Giegerich. 2021. "An Iterative Deferred Acceptance Mechanism for Decentralized, Fast and Fair Childcare Assignment." *SSRN Electronic Journal,* https://doi.org/10.2139/ssrn.3989989. [6]

Roth, Alvin E. 1984. "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory." *Journal of Political Economy* 92 (6). https://doi.org/10.1086/261272. [15]

Roth, Alvin E. 1985. "The college admissions problem is not equivalent to the marriage problem." *Journal of Economic Theory* 36 (2). https://doi.org/10.1016/0022-0531(85)90106-1. [6, 7, 9, 21]

Roth, Alvin E., and Marilda Sotomayor. 1989. "The College Admissions Problem Revisited." *Econometrica* 57 (3). https://doi.org/10.2307/1911052. [9, 25]

Roth, Alvin E., and Marilda Sotomayor. 1992. "Chapter 16 Two-sided matching." *Handbook of Game Theory with Economic Applications* 1. https://doi.org/10.1016/S1574-0005(05)80019-0. [9, 24]

Rothblum, Uriel G. 1992. "Characterization of stable matchings as extreme points of a polytope." *Mathematical Programming* 54 (1-3). https://doi.org/10.1007/BF01586041. [34]

Shimer, Robert, and Lones Smith. 2000. "Assortative matching and search." *Econometrica* 68 (2). https://doi.org/10.1111/1468-0262.00112. [10]

Smith, Lones. 2006. "The marriage model with search frictions." *Journal of Political Economy* 114 (6). https://doi.org/10.1086/510440. [6, 10]

Vate, John H. Vande. 1989. "Linear programming brings marital bliss." *Operations Research Letters* 8 (3). https://doi.org/10.1016/0167-6377(89)90041-2. [34]

Wolinsky, Asher. 1990. "Decentralized trading, strategic behaviour and the walrasian outcome." *Review of Economic Studies* 57 (1). https://doi.org/10.2307/2297543. [21]

Wu, Qinggong. 2015. "A finite decentralized marriage market with bilateral search." *Journal of Economic Theory* 160. https://doi.org/10.1016/j.jet.2015.09.005. [6–9, 14, 16, 21, 25–27, 29, 30, 33]

Chapter 2

Placement with Assignment Guarantees and Semi-Flexible Capacities*

Joint with Orhan Aygün

2.1 Introduction

In most of the standard matching theory and its applications, the analysis often includes the preferences of both sides as the main component. In a residency matching environment, this component would consist of doctors' preferences over residency programs and programs' preferences over sets of doctors. In addition to baseline preferences, many real-life matching applications include ex-ante entitlements. The entitlements constitute assignment guarantees and define a lower bound on the allocation for their owners.

One such case is when there are second-round placements for empty seats in schools or colleges. In that case, the assignment in the first round defines a lower bound for candidates who participate in the second round. The placement procedure can only send the students placed somewhere in the first round to places they prefer more. Similarly, in the case of overbooking by airline companies, all ticket owners are entitled to fly to their destination. Only occasionally is the demand to check-in higher than the plane's capacity. Companies then compensate the ineligible passengers with highly attractive offers so they voluntarily back down from their claim on the flight.

* We thank Jonas von Wangenheim, Stephan Lauermann, Francesc Dilmé, Markus Möller, and other workshop participants at Bonn, as well as the 19th Middle East Economic Association Conference, 13th Conference on Economic Design, and SAET Annual Conference 2023 participants for helpful comments and valuable feedback. Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 (Project B03) is gratefully acknowledged by Bilgin.

Nevertheless, it is important to acknowledge that the possibility of employing monetary transfers to resolve allocation issues in various contexts may be limited or completely infeasible. Yet, as an alternative approach, relaxing the capacities to some extent might be feasible. For instance, firms consist of different professionals with different skills. The firms frequently analyze their status and restructure the firm if needed. From time to time, efficient restructuring might involve replacing existing workers with new ones with different skills. However, it is often too costly for the firm to lay off existing workers due to regulations. If hiring a new worker is essential for the firm, it might deviate from the target capacity and hire the new worker anyway, even though the firm did not intend to expand in the first place.

In this paper, we take an axiomatic approach to many-to-one matching environments with assignment guarantees and *semi*-flexible (deviations from the target capacities are undesirable) capacities. Framing this within the context of the resident matching problem, doctors have preferences over residency programs, and programs have preferences over doctors. Furthermore, some doctors have assignment guarantees. A mechanism respects assignment guarantees if it assigns doctors to programs they like at least as much as their assignment guarantee.

We instantly observe that the hybrid placement problem with assignment guarantees reveals a many-to-one matching trilemma: *assignment guarantees* and *fixed capacities* are generically not compatible with *fairness*. This is because whenever the less preferred candidates have guarantees, the mechanism can not assign seats to more preferred candidates; hence is not *fair* to them. If a designer wants to respect the exogenously determined assignment guarantees and remain fair to doctors, she must relax residency programs' capacity constraints. Therefore, with such assignment guarantees, we often observe deviations from the initially determined capacities.

However, the capacities of residency programs are exogenously given in placement problems, and they reflect limited resources. Therefore, relaxing them as much as needed would probably be undesirable or infeasible for the designer. For instance, relaxing school capacities might be possible, but the number of vaccines available is a restriction that cannot be relaxed by any means. To balance fairness and capacity concerns, we present two novel axioms tailored for environments with assignment guarantees and semi-flexible capacities.

One of the axioms defines the eligible doctors for a seat in a program. This axiom is called *avoiding unnecessary slots* (shortly AUS) and ensures that a doctor can only earn a seat through her assignment guarantee or merit ranking for a program. By AUS, if a doctor is assigned a seat at a program without guarantee, we can conclude that she is ranked within the target capacity among the doctors placed there. Similarly, if a doctor is not ranked within the target capacity of a program, the placement must be due to her placement guarantee.

As mentioned above, deviations from the target capacities are inevitable if the designer aims to be fair to more preferred candidates. However, only slightly relaxing the capacities might not ensure fairness. On the contrary, the fairness of any mechanism is endangered unless the capacity limits of programs are abolished completely. When the worst doctors have assignment guarantees, imposing the traditional fairness notion requires creating additional capacities, even for some doctors, who would not have received a seat without the assignment guarantees.

Observing that the traditional notion of fairness is too strict for environments with guarantees and requires major deviations from the target capacities, the other axiom we introduce is a relaxed notion of fairness, which we name as *capacity respecting fairness* (shortly q-fairness). Given an assignment, if a doctor does not receive a seat at a program she likes better, she envies the candidates placed there. With the canonical notion of fairness, her envy is justified as soon as a candidate in that program is less preferred than her. With q-fairness, on the other hand, there is another requirement to justify the envy: Among the doctor pool of the program, the doctor has to be ranked within the target capacity. Without guarantees, q-fairness is equivalent to the traditional notion of fairness.

Crucially, q-fairness and AUS have different implications, such that q-fairness puts constraints on doctors who do not get into programs, whereas AUS constrains those who do get in. For instance, assigning every doctor to their favorite program would trivially satisfy q-fairness but fail AUS, thus heavily damaging the capacities. On the other hand, assigning everybody to their guaranteed seat would raise fairness concerns. Thus, while respecting the guarantees, a suitable mechanism must balance capacity constraints and fairness.

After introducing the axioms, we present the Assignment-Guarantees-Adjusted Mechanism (AGAM), the deferred acceptance algorithm induced by the Assignment-Guarantees-Adjusted choice function. This special choice function is tailored for the student placement environments with assignment guarantees, aiming to create the least excessive capacities possible. At each step of the algorithm, programs first admit the best candidates in their application pool, as many as their target capacity. If there are remaining doctors in the pool who have assignment guarantees at that program, they are admitted additionally. We show that AGAM admits many favorable features. First, it satisfies q-fairness and AUS, respects the assignment guarantees, and is non-wasteful. Second, it produces a stable matching and is strategy-proof on the doctor's side. In fact, it is the unique strategy-proof mechanism that is q-fair, avoids unnecessary slots, and respects assignment guarantees. Furthermore, among the mechanisms that are q-fair and respect the assignment guarantees, it minimizes the deviations from the target capacities with its AUS property.

The environment described in this paper has many applications. The main example, which also shapes the terminology of this paper, is the Re-placement of Residents Matching Problem in Turkey (see Section 2.5.1 for more detail). In

Turkey, medical students who want to continue their education with a specialty take an exam, and their placements in residency programs are determined based on their rankings in this exam. However, over the last decade, errors were discovered in calculating these scores after candidates had already been assigned to residency programs on three occasions.

In response to these errors, the centralized clearinghouse initiated a re-placement procedure. The primary objective of this re-placement process was to rectify the situation and ensure fairness for candidates whose "true" rankings were higher than initially calculated. The initial assignment granted the candidates an *acquired right*, establishing a minimum allocation guarantee for them in the second round of placements. In practical terms, the re-placement mechanism is legally obliged to assign them to a program they prefer at least as much as their initial assignment, thereby preserving their acquired rights. The conflict between the re-calculated rankings and the acquired rights resulted in some programs' relaxation of target capacities.

Likewise, in countries such as Mexico and Chile, centralized exams play a crucial role in the placement of students (Section 2.5.2). However, an intriguing feature of these systems is that they permit ties among students. This becomes particularly noteworthy when tied students compete for the last available seat at a school. In these specific applications, it is not uncommon for deviations from the target capacities of schools to arise due to the admission of extra students. This deviation is necessitated by ensuring fairness among students who are considered *equal* in terms of their exam scores and qualifications.

Moreover, in various countries, particularly within the public sector, policies ensure that married couples are granted workplaces in close geographic proximity to one another. This commitment extends to situations with no nearby vacancies, resulting in excess employment (Section 2.5.3). Similarly, within the European Union, legal provisions guarantee that new parents can return to work after parental leave. Even if the employer has hired a re-placement during their absence, the law mandates offering the returning parent a job opportunity that is at least as favorable as their previous position, which can lead to an excess of employment within specific departments (Section 2.5.4).

To better understand all such environments with entitlements, this paper provides suitable axioms that would match the policymaker's expectations. Moreover, it presents a plausible way to relax the capacities if allowed.

This paper is organized as follows: After this introduction, we discuss the related literature in the next subsection. In Section 2, we present our model. In Section 3, we show the preliminary shortcomings of the existing notions and define the axioms tailored for the specific environment. We present the Assignment-Guarantees-Adjusted Mechanism and discuss its properties in Section 4. The real-life applications introduced above are discussed in more detail in Section 5. We conclude in Section 6.

Related Literature

The paper connects to both many-to-one matching and many-to-one matching with contracts literature. On the one hand, programs have capacities and preferences over doctors, which reflects the key elements of a many-to-one environment such as in the seminal papers by Roth (1984), Balinski and Sönmez (1999), Abdulkadiroğlu and Sönmez (2003). As in those papers, the only strategic agents are the doctors, the mechanism chosen by the designer, and the residential programs' preferences are commonly observed. In addition to these essential components, the existence of assignment guarantees complexifies the many-to-one environment towards a many-to-one matching with contracts framework described by Hatfield and Milgrom (2005), Hatfield and Kojima (2008) Westkamp (2013), and many others.

Intuitively, the placement guarantees reflect yet another form of affirmative action policy. Similar to the existing literature, some candidates are exogenously prioritized at some programs, such as in Abdulkadiroğlu and Sönmez (2003), Kojima (2012), Hafalir, Yenmez, and Yildirim (2013), and Doğan (2016). Similar to them, there might be various underlying reasons for such a claim at a seat, for example, the location of the spouse or acquired rights. However, because a doctor cannot be assigned to a least preferred alternative than their guarantee, the assignment guarantees are more strict than the priorities in those papers. A similar lower bound constraint can be found in Combe et al. (2022). In their setting, the existing teachers of schools cannot be sent anywhere that they like less than their current assignment.

Similar to Westkamp (2013), Kominers and Sönmez (2016), Aygün and Sönmez (2013), the mechanism we propose is essentially a Gale-Shapley deferred acceptance algorithm Gale and Shapley (1962) along with a choice rule to be implemented at each step. Furthermore, the mechanism involves a dynamic flavor as in Aygün and Turhan (2020), in the sense that the balance between the guaranteed seats and regular seats evolves as the mechanism moves forward. Unlike those papers, doctors do not have a preference about the type of seat they acquire. Therefore, even if they were asked to reveal their guarantees to the mechanism, no candidate would have incentives to hide their guarantee status strategically.

The environment resembles the benchmark housing market as individuals' guarantees define a lower bound for them in a Pareto sense; candidates with guarantees can only be better off Shapley and Scarf (1974). However, the complex preferences of programs prevent us from implementing Gale's TTC. Furthermore, our mechanism creates additional seats. In that sense, the problem resembles the House Allocation with Existing Tenants in Abdulkadiroğlu and Sönmez (1999), with a substantial twist that creating new rooms is possible. Nevertheless, the additional rooms, if created, can only be used by the candidates with a guarantee at the specific program. Increasing efficiency further by exchanging guarantees is restricted due to stability concerns.

To the best of our knowledge, this paper is the first to study how to relax capacities when the need arises in strict preference environments while respecting the exogenous lower bounds on the assignments and upholding some fairness. Pursuing an axiomatic approach to characterize a suitable mechanism for such environments, we translate the semi-flexible capacities into the choice functions of programs, and the realized capacities then depend on the application pool. Similar to Westkamp (2013), Sönmez and Switzer (2013), and Dimakopoulos and Heller (2019), our analysis has a direct application in a current market that concerns tens of thousands of people every year.

2.2 Model

Same as in other many-to-one matching environments, and in line with the main application, we have a finite set of doctors and a finite set of residency programs denoted by $D = \{d_1, d_2, ..., d_n\}$ and $H = \{h_1, h_2, ..., h_m\}$, respectively. The generic doctor d has strict preferences, P_d (occasionally $\succ_d$ for convenience) over programs H along with an outside option $\emptyset$, where $P_{D'}$ is the collection of the preferences of doctors in set $D' \subset D$.

Similarly, the generic residency program is denoted by h and has an exogenously determined capacity q_h, where the collection of the target capacities of all programs is denoted by $q_H = \{q_{h_1}, ..., q_{h_m}\}$.

While considering a student placement problem with assignment guarantees, a convenient approach is adapted from the matching with contracts framework: We define choice functions for both sides of the matching platform. From any set of residency programs $H' \subset H$, d chooses according to the choice function that is driven from her strict preference P_d over $H \cup \{\emptyset\}$, $C_d : 2^H \to H \cup \{\emptyset\}$, such that $C_d(H') = max_{P_d}(H' \cup \{\emptyset\})$. This implies that the doctors have unit demand. When $C_d(H') = \emptyset$, doctor d prefers the outside option among the choices, i.e., to remain unemployed.

Analogously, C_h is the choice function of program h. Similar to C_d, C_h allows the programs to choose no doctor from any application pool. However, C_h is different than C_d in two aspects: First, programs choose sets of doctors from application pools, thus $C_h : 2^D \to 2^D$, and for any $D' \subset D$, $C_h(D') \subset D'$. Second, it's worth noting that choice functions within programs do not necessarily have to be determined by a preference relation, nor do the choice functions of distinct programs need to be interconnected or correlated. In fact, any choice process will involve two components:

First, each program h has an exogenously given strict preference over individual doctors that is denoted by P_h ($\succ_h$, and the collection of preferences

P_H).[1] For notational convenience, we occasionally use the rankings of doctors in an application pool D' instead of the preferences. Reasonably, for program h, the ranking of doctor d in any application pool D' is defined as a function $z_h(d|P_h, D') : D' \to \mathbb{N}^+$ and $z_h(d)$ decreasing with P_h, such that the more preferred d is for h, the higher ranking she has in an application pool.[2] Without assignment guarantees, the components described so far constitute a student placement problem: (D, H, P_D, P_H, q_H).

Nevertheless, the contribution of this paper is to implement assignment guarantees into such a many-to-one placement problem, which is the second component of a program's choice process. Doctors can have assignment guarantees at different programs. In line with this purpose, let E_h denote the set of doctors who are guaranteed a seat at program h. With a similar notation logic as above, E_H is the collection of doctors with assignment guarantees at each program. Therefore, the student placement problem with assignment guarantees consists of the tuple: $(D, H, P_D, P_H, q_H, E_H)$.

Once doctors and programs are assigned to each other, a matching μ is a set of doctor-program (d, h) pairs such that each doctor d appears in at most one pair and $\mu(d) = h$ if and only if $d \in \mu(h)$, where $\mu(d)$ and $\mu(h)$ denote the match of the doctor d and program h under matching μ, respectively.

A direct mechanism is then a function ϕ and selects a matching for each preference profile, capacity vector, and guarantee scheme of the doctors. In this paper, we denote the matching, which is the outcome of the direct mechanism ϕ as μ^ϕ.

2.3 Placement Problem with Assignment Guarantees

In this section, we formally define the concepts specific to our environment. Moreover, we discuss the peculiarities and challenges of the student placement problem with assignment guarantees that arise due to the characteristics of the problem.

2.3.1 Placement Problem with Assignment Guarantees when Capacities are Fixed

This subsection studies the incompatibility between assignment guarantees and fixed capacities in the placement problem. Namely, we show and discuss that respecting assignment guarantees and satisfying fairness is impossible while keeping the programs' capacities constant. Assignment guarantees of doctors put constraints on the outcome, such that a doctor may never be assigned to a program

1. The existence of a centralized exam score is a special case of this framework such that the programs have the same preference over doctors.

2. Formally, $z_h(d|P_h, D') = |d' \in D' : d' \succ_h d| + 1$.

that she prefers less than the program at which she has an assignment guarantee. Formally:

Definition 2.1. Mechanism ϕ **respects assignment guarantees** if for any fixed problem $(D, H, P_D, P_H, q_H, E_H)$, $\nexists d \in D$ such that $d \in E_h$ for some h and $h \succ_d \mu^\phi(d)$.[3]

Next, we introduce the *fairness* criterion, which is adapted from the student placement literature. Fairness requires that the more preferred doctors are assigned to better alternatives. Formally:

Definition 2.2. Mechanism ϕ satisfies **fairness** (or is fair) if for any fixed problem $(D, H, P_D, P_H, q_H, E_H)$, $\nexists\{d, d', h\}$ such that $d \succ_h d'$ and $h \succ_d \mu^\phi(d)$ whereas $d' \in \mu^\phi(h)$.

The outcome of a mechanism is not fair if there is an unmatched doctor-program pair (d, h), where doctor d prefers program h to her own assignment, and the program h prefers her to another doctor d' who is assigned a seat there. Observe that both criteria are quite intuitive. The difference between the two is that fairness only considers the preferences of both sides. However, programs should also consider the assignment guarantees of doctors. Therefore, assignment guarantees might diversify programs' *preferences*. The following simple example shows how fairness conflicts with assignment guarantees when programs' capacities are not relaxed.

Example 2.3. There are two doctors, $D = \{a, b\}$, and one program $H = \{h\}$ with a single capacity $q_h = 1$. Suppose a is guaranteed a seat at hospital h, meaning that if a mechanism is to respect assignment guarantees, she cannot be assigned to a worse alternative than h. Doctor b has no assignment guarantee; therefore, $E_h = \{a\}$. If $a \succ_h b$, it is quite easy for any mechanism ϕ to satisfy fairness and respect the guarantees, that is, to match a with h and leave b unemployed. However, it is more complex once $b \succ_h a$. In that case, fairness requires $\mu^\phi(b) = h$ and assignment guarantees require $\mu^\phi(a) = h$. Thereupon, it is only possible to satisfy both by creating an additional capacity at program h.

Admitting the impossibility of respecting assignment guarantees and satisfying fairness without creating additional capacities, deviation from the *target* capacities q_H is still undesired. The unintended acceptance of additional doctors results in inefficiency in many ways. First, it is harmful to the government budget to employ two doctors instead of one. Second, from program h's point of view, if the program was optimally designed for one resident only, then the additional resident may reduce the overall quality of the education.

3. Observe that this definition does not prevent a doctor from having assignment guarantees at multiple programs. If a doctor is guaranteed a seat at multiple programs, we can WLOG restrict attention to her most preferred alternative amongst her assignment guarantees.

Additionally, recall the Rural Hospitals Theorem by Roth (1986). A quite intuitive fact that names the theorem is that young residential candidates prefer the programs in the urban areas rather than the rural ones. Thus, this unintended creation of additional capacities will likely disturb the balance between hospitals regarding the number of residents employed.

On the one hand, additional capacities must be created if the designer aims to implement a fair mechanism. However, how many additional capacities will be required remains an unanswered question. We illustrate this extent with the following example:

Example 2.4. Suppose there are three doctors, $D = \{a, b, c\}$ and one program $H = \{h\}$ with a single capacity $q_h = 1$. As the previous example, only a is guaranteed a seat at hospital h. Furthermore, h ranks the candidates as $b \succ c \succ a$.

Any mechanism that respects assignment guarantees has to satisfy $\mu^{\phi}(a) = h$.

Furthermore, once a is assigned to program h, fairness would require all three doctors to be assigned to h since b and c are preferred over a. However, without assignment guarantees, c would not receive a seat in a fair mechanism. In other words, a's assignment guarantee at program h indirectly creates another assignment guarantee for c.

We could take this scenario to an even more extreme point: What if there are many doctors as c, whose ranking satisfies $b \succ c \succ \ldots \succ z \succ a$? Would all the candidates receive a seat at h only because of a's assignment guarantee? In fact, the first proposition of the paper answers this question and shows the extent of the demandingness of fairness in environments with placement guarantees:

Proposition 2.5. *There is no mechanism that satisfies fairness and respects assignment guarantees without creating capacities equal to the number of candidates for each program.*

If the designer wants to ensure fairness for all candidates, she has to abolish the capacities of all programs. The discussion above is only strengthened for no capacity regulations and, therefore, is undesirable. Apart from not being desirable, disregarding the capacity constraints is infeasible in most cases since the target capacities are initially designed to meet specific criteria. Admitting that imposing the traditional fairness axiom is too strict in such environments, we propose an alternative, relaxed notion of fairness in the next section, which is more suitable to the placement environments with assignment guarantees.

2.3.2 Capacity Respecting Fairness (q-fairness)

The main reason for proposing a new notion of fairness is to avoid creating additional capacities for those who would not acquire a seat without the assignment guarantees of other candidates. The canonical notion of fairness is too strict in that additional capacities are created unintendedly, even for doctors such as c, who are ranked outside the target capacity and do not have assignment guarantees at programs. In other words, if it were not for the guarantees, doctors such as c would not have a claim on the seats based on fairness because they are ranked outside the capacity of program h. For this very reason, we present *capacity respecting fairness*, which is a relaxed version of the traditional fairness axiom:

Definition 2.6. Mechanism ϕ satisfies **capacity respecting fairness** (or is q-fair) if for any fixed problem (D,H,P_D,P_H,q_H,E_H), $\nexists(d,h)$ such that $h \succ_d \mu^\phi(d)$ and $z(d|P_h,\mu^\phi(h)\cup\{d\}) \leq q_h$.

Intuitively, q-fairness suggests that a mechanism ϕ is *capacity respectingly fair* to doctors, as long as it doesn't assign a doctor d to a worse alternative $\mu^\phi(d)$, who would be ranked within the target capacity of a program h along with the to-h-matched doctors $\mu^\phi(h)$. Observe that q-fairness is clearly a weaker condition than fairness. Specifically, q-fairness allows fairness violation for those ranked outside the target capacity for an application pool (such as doctor c in Example 2.4 above, but not doctor b).

The problem of creating additional capacities is still not beside the mark with q-fairness. A mechanism still does not exist that satisfies q-fairness and respects assignment guarantees without creating additional capacities for some programs. This could again be observed in Example 2.4 with doctors a, b, and c. As a has a right at the only seat of h, b also has the right due to q-fairness. However, q-fairness allows us to leave c out of the program. In that sense, q-fairness is a minimal *fairness* requirement to decrease the number of additional capacities. It is the least we could expect from a mechanism with some fairness concerns when assignment guarantees are present.

If a mechanism is q-fair, a side benefit is that no seat of a program is left empty as long as some doctor prefers the seat to her current alternative. This notion is formally called *non-wastefulness*, meaning no seat is wasted throughout the procedure.[4] Furthermore, in an environment where agents can earn seats by no means but simple preferences on the program side (such as their exam scores), q-fairness is the same as the canonical fairness axiom and non-wastefulness combined.

2.3.3 Avoiding Unnecessary Slots

Capacity respecting fairness is still insufficient to prevent the creation of unnecessary capacities to the full extent. To see this, consider the following example:

Example 2.7. Suppose there are four doctors, $D=\{a,b,c,d\}$ and two programs $H=\{x,y\}$ such that $q_x=q_y=1$. The program's preferences satisfy $a \succ b \succ c \succ$

4. The formal definition, as well as the proof for q-fairness implying non-wastefulness, can be found in the appendix.

d, and the least preferred candidates have assignment guarantees, $E_x = \{c\}$ and $E_y = \{d\}$. Suppose x is preferred over y by all doctors but c (c prefers y to x). If mechanism ϕ places candidates such that $\mu^{\phi}(x) = \{a,d\}$, $\mu^{\phi}(y) = \{b,c\}$, q-fairness is still not violated. However, it is neither clear nor natural that c and d receive an additional seat at a program with no assignment guarantee.

Capacity respecting fairness constraints the envy of those who did not get into the residential programs. However, as seen above, another notion is also needed, which restricts the ones who actually receive a slot. Only that way can we ensure that the excess capacities are only created for those with assignment guarantees at programs. For this reason, we introduce the following notion:

Definition 2.8. Mechanism ϕ **avoids unnecessary slots** (or satisfies AUS) if for any fixed problem (D,H,P_D,P_H,q_H,E_H) and any pair (d,h),

$\mu^{\phi}(d) = h$ and $d \notin E_h \Rightarrow z(d|P_h,\mu^{\phi}(h)) \leq q_h$

Verbally, a mechanism avoids unnecessary slots if a doctor is placed in a program only because of her assignment guarantee or if she is ranked within the target capacity among the assigned set of doctors to the program.

The two notions of q-fairness and AUS are introduced to find a balance between doctors' preferences, programs' preferences, and assignment guarantees. Even though they might sound similar, their implications are quite different. For example, assigning every doctor to their favorite program would satisfy assignment guarantees and q-fairness trivially (because there is no envy) but would possibly create many unnecessary slots, failing AUS. On the other hand, allocating seats only to doctors with assignment guarantees would not create unnecessary seats but raise fairness concerns.

When there are assignment guarantees in addition to program preferences, the requirements we expect from a mechanism have to be relaxed relative to the benchmark student placement problem. For the two new notions of q-fairness and avoiding unnecessary slots, we can say that q-fairness considers doctors' preferences in a relaxed way compatible with assignment guarantees. In contrast, AUS considers programs' preferences in the same relaxed way. Without assignment guarantees, the Student Proposing Deferred Acceptance Algorithm is naturally q-fair (also fair) and avoids unnecessary slots.

2.4 The Assignment-Guarantees-Adjusted Mechanism

Having presented the appropriate notions, we now define a new choice function, with the intent of building towards a mechanism that creates excess additional capacities to the least while respecting assignment guarantees. Because of the rather complex nature of the student placement problem with assignment guarantees, the choice function will resemble the choice functions in matching with contracts framework, especially the choice functions in the Cadet-Branch Matching problem by Sönmez and Switzer (2013). The Assignment-Guarantees-Adjusted Choice Function (shortly AGA Choice Function, denoted by C_h^A) defines a selection rule of program h from any application pool $D' \subset D$ and proceeds as follows:

(1) Rank all the doctors in D' according to preferences of h.

(2) Based on their rankings, add doctors one-by-one to $C_h^A(D')$ until q_h is full or all doctors are considered.

(3) Add all the remaining doctors such that $d \in (E_h \cap D')$ to $C_h^A(D')$.

(4) Terminate the procedure, reject all other doctors.

For any given application pool, the choice function first considers its own preferences and adds doctors one by one according to their ranking. In this step, there may or may not be doctors who have assignment guarantees among the chosen doctors. After the target capacity is complete with the merit candidates, the program does not immediately reject all the remaining candidates. Instead, if some doctors are left in the application pool with an assignment guarantee, it adds them to the chosen set and expands its capacity.

Almost trivially, one can show that the choice function satisfies specific well-behaving properties that a mechanism designer would expect from a choice function, such as substitutes, law of aggregate demand (LAD), and irrelevance of rejected contracts (IRC), proofs of which can be found in the Appendix 2.B. Intuitively, the substitutes condition ensures no complementarities between the doctors for the programs. LAD guarantees the expansion of the rejection set as the choice set expands, and with IRC, removing the rejected alternatives does not affect the choice set.

After introducing the AGA Choice Function, we now focus on mechanism design. The mechanism we introduce to minimize the number of additional seats is the doctor proposing deferred acceptance algorithm induced by the AGA Choice Function. Formally:

Step 1: Each doctor proposes to their first choice. Each program tentatively assigns its seats to the doctors in its application pool according to the AGA Choice Function.

⋮

Step k: Each doctor rejected by any program in the previous step proposes their next choice. Each program considers the doctors it has been holding, together with the new applicants, and tentatively assigns its seats to the doctors in its new application pool according to the AGA Choice Function.

Because this mechanism uses the AGA Choice Function in every step of the deferred acceptance algorithm, we call this special mechanism *Assignment-Guarantees-Adjusted Mechanism*, shortly AGAM.

In an environment where doctors have assignment guarantees at some programs in addition to the programs' preferences, creating additional capacities is inevitable. However, the adjusted choice function helps implement assignment guarantees into the benchmark placement problem such that assignment guarantees are respected and the deviation from the target capacities is only due to the guarantees.

It is almost trivial that for any application pool of any program, the choice function itself exhibits the plausible features of the environment with assignment guarantees. If there is a single program, the choice function admits candidates in a single step such that it is q-fair, avoids unnecessary slots, and respects assignment guarantees by construction.[5] Furthermore, the candidates would not be incentivized to misreport their preferences. However, when it comes to the mechanism, which includes more than one program and takes several steps to conclude, it is not straightforward that the plausible properties are still satisfied. In the following section, we rigorously discuss the properties of AGAM. All omitted proofs can be found in Appendix 2.C.

2.4.1 Fairness, q-fairness, AUS, and Assignment Guarantees

At this point, it is already clear that AGAM violates traditional fairness.[6] It was the first acknowledgment of the paper that fairness is too strict in placement environments with assignment guarantees. On the other hand, the mechanism satisfies other properties defined in the previous chapters.

Proposition 2.9. *AGAM satisfies capacity respecting fairness, avoids unnecessary slots, and respects assignment guarantees.*

The fact that AGAM satisfies q-fairness, AUS, and respects assignment guarantees proves that the mechanism is a suitable candidate for a placement environment with assignment guarantees. The mechanism acknowledges the merit rankings of doctors and respects the exogenously given assignment guarantees. When some fairness concerns are present, recall that any mechanism has to deviate from the target capacities, but AGAM does that in a minimally harmful way. This is because AGAM ensures that the extra capacities belong to exogenously guaranteed candidates at each program. Furthermore, as mentioned before, it is non-wasteful, implied by q-fairness.

5. When there is a single program, a mechanism being q-fair is equivalent to the choice function being q-responsive as in Aygün and Turhan (2020).

6. Recall Example 2.4, the outcome of AGAM would be $\mu^A(h) = \{a, b\}$, leaving c unemployed, thus violating fairness.

2.4.2 Strategy-Proofness

While designing a mechanism for a placement problem with or without assignment guarantees, most of the properties we look for and expect from our mechanism depend on the preferences of the doctors (q-fairness of the mechanism, stability of the outcome, etc.) Therefore, if doctors are incentivized to misreport their preferences, it would be pointless to analyze those properties. Strategy-proofness is, therefore, an essential property of the mechanisms to eliminate such doctors' incentives.

Definition 2.10. A mechanism ϕ is **strategy-proof** (for doctors) if for any doctor d and preference profile (P_d, P_{-d}), where P_{-d} is the collection of the preference profiles of all doctors but d, there is no preference $P'_d \in \mathscr{P}_d$ such that $\mu^{\phi(P'_d, P_{-d})}(d) \succ_d \mu^{\phi(P_d, P_{-d})}(d)$.

Proposition 2.11. *AGAM is strategy-proof.*

Strategy-proofness ensures that the doctors reveal their true preferences to the mechanism, without which all the other properties would trivially fail according to true preferences. Moreover, note that the assignment guarantees of doctors are automatically revealed to the mechanism. However, the guarantees can only improve the placement of a doctor, and the doctors do not differentiate between different types of seats. Therefore, even if we allowed the doctors to report their assignment guarantees alongside their preferences, they would not be incentivized to hide their guarantee status either.

The fact that AGAM is also strategy-proof strengthens our claim that the mechanism is suitable for placement environments with assignment guarantees. Furthermore, the following theorem concludes that AGAM is, in fact, the unique mechanism if the designer has merit concerns, is constrained by the guarantees as well, and aims to create as least additional seats as possible while eliciting the actual preferences of doctors.

Theorem 2.12. *AGAM is the unique strategy-proof mechanism that respects assignment guarantees, is q-fair, and avoids unnecessary slots.*

Proof. The mechanism is essentially a deferred acceptance algorithm that uses the AGA choice function at each iterative step. In Appendix 2.D, we show that the axioms of respecting assignment guarantees, q-fairness, and AUS are together equivalent to stability with respect to the AGA Choice Function. The AGA Choice Function satisfies substitutes and the law of aggregate demand conditions, the deferred acceptance induced by this function produces a stable outcome and is strategy-proof. By Hirata and Kasuya (2017), the doctor proposing deferred acceptance algorithm is the unique candidate for a strategy-proof mechanism that produces a stable outcome. □

2.4.3 Deviation

Until this point, we characterized and showed many favorable properties of AGAM. In this section, we analyze how the mechanism deviates from the target capacities carefully designed by planners and deviations from which are undesired.

Admitting that the additional capacities will have to be created in a student placement problem with assignment guarantees, we calculate the **deviation** of an outcome from the original target capacity of a program by taking the difference between the realized capacity and the target capacity only if the realized capacity exceeds the target capacity. Formally, the deviation of an outcome from the original target capacity of a program is $max\{\mu(h)-q_h,0\}$. The main reasoning behind this absolute value approach is that the unintended and inevitable *excess* placements cause complications in the first place. By now, it is clear that avoiding unnecessary slots is required to control the complication dimension. The theorem below shows that, indeed, AGAM is one of the mechanisms that minimize the deviation from the target capacities while satisfying q-fairness and respecting the assignment guarantees of doctors.

Theorem 2.13. *Among the mechanisms that are q-fair and respect assignment guarantees of doctors, AGAM minimizes the deviation from the target capacities.*
Formally, for all mechanisms ϕ that are q-fair and respect assignment guarantees, and problems (D,H,P_D,P_H,q_H,E_H), we have $\forall h \in H$, $max\{\mu^A(h)-q_h,0\} \leq max\{\mu^\phi(h)-q_h,0\}$.

A very short and verbal intuition for the proof would be: To obtain an outcome that deviates less than μ^A, a chain must be constructed over μ^A, which starts with a doctor who is placed to that program under μ^A and ends at a vacant capacity. We show in the appendix rigorously that such a chain conflicts with either q-fairness or assignment guarantees or both. In fact, the deviation result relies on the AUS property of AGAM, which is one way to prevent excessive deviations. After proving this theorem, we can now conclude that in a student placement problem with assignment guarantees, if the deviation from the target capacities is undesirable, AGAM is one of the best mechanisms that can be implemented. The formal proof, as well as an illustrative example, is to be found in the appendix.

There might be other mechanisms that result in the same deviation for each program—for example, implementing TTC after AGAM is another mechanism that minimizes deviation. This mechanism fails AUS. An example of the outcome can be found here 2.7. In other words, Theorem 2.13 also informs us that within the realm of mechanisms that are q-fair and respect the guarantees, imposing avoiding unnecessary slots exhibits one viable approach to minimizing deviation from the target capacities.

2.5 Applications

In this section, we present four applications of the theoretical analysis in this paper, all of which have the feature that program preferences and assignment guarantees conflict. We start with the Re-placement of Residents Matching Problem in Turkey, where some candidates have guaranteed seats at a program during the resident placement procedure. This application is the main example, which also determines the terminology of the paper.

2.5.1 Re-placement of Residents Matching Problem in Turkey

Medical students in Turkey who want to continue their education with specialization take an exam called the Examination of Specialty in Medicine (ESM). The state agency Measuring, Selection, and Placement Center (MSPC) is responsible for conducting the ESM twice a year and the assignment procedure in the aftermath of the exam. In the ESM, residential candidates receive a score and are ranked according to those scores. After that, they are placed at residency programs according to the doctor proposing deferred acceptance algorithm (similar to many exams conducted by the MSPC). However, one of the two exams in 2010, 2013, 2014, and 2016 were exceptional, and the placement process was not straightforward. After the exam, authorities found some questions flawed and were officially canceled. Scores of the doctors were calculated according to the remaining accurate questions. As usual, placements were done by the doctor proposing deferred acceptance algorithm.

Nonetheless, after the placements, the State Council revoked the cancellation, which re-established the accuracy of the canceled questions. This led to a change in the scores and, hence, the rankings of the doctors. However, the placements had already been done and the residential candidates had started working at their assigned programs. Hence, the original placement was obviously not *fair* to some residential candidates, especially those whose rankings have increased after the score re-calculation.

As compensation, it was announced that there would be a re-placement procedure to provide fairness to the doctors with increased rankings. The re-placement procedure would preserve the acquired rights (which correspond to the assignment guarantees in our setting) of the existing doctors in the programs, i.e., who were assigned a seat during the original placement procedure. [7] In other words, any compensation mechanism should make the candidates placed in the initial placement at least as happy as before.

7. The acquired rights are defined by law and prevent the existing residents of programs being assigned to other programs which they prefer less than their initial assignments (also called as vested interests in the literature).

All in all, restoring fairness and respecting the acquired rights can only occur unlesifprograms' target capacities are relaxed. For some scores and preferences of doctors, some residential programs end up with more doctors than their target capacity. The unintended acceptance of these doctors results in inefficiency in many ways. First, from the governmental point of view, if they were not placed at the original placements, they are harming the government's budget. Second, from a program's point of view, if the program was optimally designed for one resident, then the additional resident may reduce the quality of the education for each resident. Third, if a doctor has already started with another residency program, her being accepted by another program means a loss for her original assignment. Depending on the presence and quality of the other doctors in its application pool, that program might face other complications and this situation will keep snowballing towards the less preferred programs. Additionally, young residential candidates usually prefer programs in urban areas rather than rural ones. Thus, this unintended creation of additional capacities will likely disturb the balance between rural and urban hospitals regarding the number of residents employed, who are an essential chain ring in the middle of the health industry. Last, but not least, the more preferred programs announce fewer vacancies in the subsequent years because of over-employment during the re-placement. This resembles exchanging better future students for worse doctors in the years of re-placement for them, harming both the quality of those programs and future residential candidates.

In another work in progress (Aygün and Bilgin (n.d.)), we analyze the mechanism used by MSPC and compare it rigorously to the mechanism described in this paper. Under the dominant strategy of doctors, the mechanism used by MSPC exhibits a notable feature—it generates additional seats for doctors, even in cases where they lack top-q merit or an assignment guarantee. In essence, MSPC effectively doubles the target capacities of residency programs.

For this case with exam scores, the problem is a special case of our model in this paper. Consequently, our characterization and other results directly apply to this context. Within this framework, our Assignment-Guarantees-Adjusted Mechanism creates weakly less deviation than the mechanism used by MSPC. This difference in deviation becomes significantly pronounced when we assume a positive correlation between the initial and re-calculated scores. The positive correlation implies that candidate rankings undergo only slight changes during re-calculation, aligning with the nature of the application itself since the re-calculation affects only some questions.

2.5.2 Equal Treatment of Equals: Student Assignment in Hungary, Chile, and Mexico

One of the most basic fairness axioms in matching theory is *equal treatment of equals*, which is more trivial when agents have strict preferences. On the other hand, when ties and capacity constraints enter the equation, this principle becomes considerably more complex. In Hungary, where schools maintain strict target capacities, the feasibility of increasing these capacities is not an option. Consequently, a stark reality emerges in student placement: if admitting all tied students results in a capacity increase, none of them gain admission as discussed in Biró and Kiselgof (2015).

In Chile, allocating students to colleges showcases a different approach to handling capacity limitations. Here, colleges are allowed to exceed their designated capacity, but only under the crucial condition that this over-allocation can occur solely when the last admitted students, who push the college beyond its capacity, share an identical ranking or score. In essence, this approach ensures that the final students admitted are considered equal in terms of their qualifications, thereby upholding the equal treatment of equals as in Rios et al. (2021), which may cause excessive deviations from the target capacities.

A combination of the above approaches is implemented in Mexico City, which is analyzed in Ortega Hesles (2015). When students with identical scores compete for the last available seat in a school, the centralized clearinghouse admits all or none of the tied students. This results in a capacity increase for some. particular schools.

These applications across Hungary, Chile, and Mexico City exhibit different approaches in case some students share the same qualifications for admission, underscoring one more time the importance of the balance between fairness concerns and capacity regulations. Since q-fairness implies non-wastefulness (Appendix: 2.A), we can conclude that only the Chilean approach satisfies q-fairness among these three applications.

2.5.3 Spousal Matching

Matching markets with couples is an interesting theoretical problem. The problem of matching with couples relies on the fact that spouses have a preference to be appointed to geographically similar locations, and there is already a vast literature about whether and under what conditions stability can be achieved (Roth (1984), Kojima, Pathak, and Roth (2013)).

To protect family integrity, different central planners adopt different solutions. The famous National Resident Matching Program (NRMP) states, ”When applicants participate in a match as a couple, their rank order lists form pairs of program choices that the matching algorithm considers. A couple will match to the most preferred pair of programs on their rank order lists where each partner has been offered a position”.[8] In Turkey, civil servants are guaranteed to be appointed to the same location as their spouse, provided that either their spouse is also a civil servant or the spouse has been working for the same private firm for a sufficiently long time.

For instance, after medical education, doctors must complete a mandatory civil service at a place determined by a lottery to validate their diplomas. Married doctors who satisfy the above criteria can apply for a spouse-related appointment to be separated from the general lottery and are guaranteed to be appointed to a hospital in the same city as their spouse.[9]

In that case, the departments might have to create excess capacities even though they are not looking for additional workers. The excess capacity creation is only due to the marital status of doctors, which is usually exogenous to the appointment problem.

2.5.4 Return to Work After Parental Leave

Having and raising offspring is a basic instinct for human beings. Furthermore, for a functioning social security system and a balanced society, every country needs a sufficient amount of young population to join the labor force. Nevertheless, the fertile time window usually conflicts with the early career plans of young individuals. Hence, it is not always an easy decision to take a break from their career. Therefore, many countries work on regulations that will give young individuals incentives to childbearing to have a balanced population.

In the European Union, the law ensures that “working men and women are entitled to return to their jobs or to equivalent posts on terms and conditions which are no less favorable to them”.[10] With that regulation, the potential fear of losing their job is eliminated, so young individuals are incentivized towards childbearing.

The firms might have nondeferrable needs, however fair and reasonable that protection is. Suppose the new parent’s temporarily vacant position is crucial for the firm’s structure. In that case, the firm might consider hiring an additional worker, even though it is aware that the parent has the right and will return to the same position after their parental leave. In that case, the firm will have hired an additional worker even though it does not have a prospect of expanding the company.

8. Source: The official website of NRMP https://www.nrmp.org/.
9. Source: The official website of Ministry of Health in Turkey https://yhgm.saglik.gov.tr/.
10. Source: Directive 2006/54/EC of the European Parliament and of the Council of 5 July 2006 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation https://commission.europa.eu.

2.6 Conclusion

In this paper, we analyze student placement problems with assignment guarantees, where the designer aims to preserve assignment guarantees as well as has some fairness concerns. We show that conflict between fairness and assignment guarantees is unavoidable when the programs' capacities are fixed.

Nevertheless, since the programs' capacities are already optimized, any deviation from the target capacities is costly and undesirable. Imposing the traditional notion of fairness, however, results in excessive deviations from the target capacities in such an environment. Thus, the traditional fairness notion is unsuitable for student placement environments with assignment guarantees. To reduce excessive deviations and still redeem some form of fairness, we define the notions of capacity respecting fairness and avoiding unnecessary slots. Capacity respecting fairness relaxes fairness to the extent that a mechanism can be unfair to candidates who are ranked outside the target capacity but can still be q-fair. Avoiding unnecessary slots ensures that additional seats are used either by merit candidates or guaranteed candidates.

Moreover, we define a new selection rule for the programs, the assignment-Guarantees-Adjusted Choice Function, and propose a new mechanism, the Assignment-Guarantees-Adjusted Mechanism, to be used in student placement procedures with assignment guarantees.

The Assignment-Guarantees-Adjusted Mechanism is the deferred acceptance algorithm induced by the Assignment-Guarantees-Adjusted Choice Function. It is the only strategy-proof mechanism that satisfies the q-fairness and avoids unnecessary slots while respecting assignment guarantees. Moreover, the Assignment-Guarantees-Adjusted Mechanism minimizes the deviation from the target capacities while respecting the assignment guarantees of doctors and satisfying q-fairness.

For future research, it might be helpful to consider programs' preferences more elaborately. For instance, the central planner might commit to smaller target capacities if many candidates have assignment guarantees, and the deviation is costly. On the other hand, if the emphasis is on fairness, we would observe larger target capacities. Quantifying the analysis might help us better understand environments with assignment guarantees.

Appendix 2.A Non-wastefulness

Definition 2.14. Mechanism ϕ is **non-wasteful** if for any fixed problem $(D, H, P_D, q_H, E_H, s_D)$, $\forall(d, h) \in (D \times H)$, $h \succ_d \mu^\phi(d) \implies |\mu^\phi(h)| \geq q_h$.

We can easily show that if a mechanism satisfies q-fairness, it is also non-wasteful. Suppose ϕ violates non-wastefulness. Then $\exists(d, h)$ such that $h \succ_d \mu^\phi(d)$ and $|\mu^\phi(h)| < q_h$. Then, $z(d|P_h, \mu^\phi(h) \cup \{d\}) \leq q_h$ which means ϕ also violates q-fairness.

Appendix 2.B Properties of the AGA Choice Function (C_h^A)

(1) Substitutes

Definition 2.15. Elements of Y are **substitutes** for program h if for all subsets $Y' \subset Y'' \subset D$ we have $Y' \setminus C_h(Y') \subset Y'' \setminus C_h(Y'')$. (Hatfield & Milgrom, 2005)

Substitutes condition requires that the rejection set expands (weakly) as the application pool expands. Intuitively, it implies that there are no complementarities between doctors.

Lemma 2.16. *C_h^A satisfies substitutes.*

Proof. Any violation of substitutes would require the existence of doctor d such that, $d \notin C_h^A(Y')$ but $d \in C_h^A(Y'')$ for some Y'' such that $Y' \subset Y''$. All by-h-prioritized doctors are chosen by the choice function, so our violation of substitutes, if any, must stem from the non-prioritized candidates in the application pool. Suppose d it not prioritized at h and $d \notin C_h^A(Y')$. Then d has not a high enough ranking in Y'. Clearly, doctor d's ranking in the set Y' weakly decreases while the set expands by the addition new doctors. As a consequence, d will still not be chosen from any set Y'' such that $Y' \subset Y''$ either. Thus, C_h^A satisfies substitutes. □

(2) Law of Aggregate Demand

Definition 2.17. The preferences of hospital $h \subset H$ satisfy the law of aggregate demand if for all $X' \subset X'' \subset D$, $C_h(X') \leq C_h(X'')$. (Hatfield & Milgrom, 2005)

The Law of Aggregate Demand is an intuitive condition, which implies that the chosen set does not get smaller as the application pool expands.

Lemma 2.18. *C_h^A satisfies LAD.*

Proof. Observe that for any $Y' \subset Y'' \subset D$, if $|C_h^A(Y')| \leq q_h$, at least $|C_h^A(Y')|$ candidates will be chosen from Y''. If $|C_h^A(Y')| > q_h$, it means there are prioritized candidates in (Y'), who are ranked outside the target capacity. As the set expands, those prioritized doctors will still be outside the capacity, thus again at least $|C_h^A(Y')|$ candidates will be chosen from Y'' as well. Thus, for all $Y' \subset Y'' \subset D$, $|C_h^A(Y')| \leq |C_h^A(Y'')|$. □

(3) IRC

Definition 2.19. Given a set of doctors D, a choice function satisfies the **irrelevance of rejected contracts (IRC)** if and only if:
$\forall Y \subset D, \quad \forall z \in D \setminus Y \qquad z \notin C(Y \cup \{z\}) \implies C(Y) = C(Y \cup \{z\})$. (Aygün & Sönmez, 2013)

The IRC condition requires that the removal of not chosen (rejected) contracts from the application pool do not affect the chosen set.

Lemma 2.20. *C_h^A satisfies IRC.*

Proof. Suppose the choice function chooses $C_h^A(Y)$ from the application pool Y, where $d \notin C_h^A(Y)$. As above, $d \in (Y \setminus E_h)$ and d's ranking in Y is lower than the target capacity. Then, removing d from Y would have no effect on the top q_h candidates of the prioritized candidates, and thus on the chosen set, namely $\forall d \in Y$ such that $d \notin C_h^A(Y)$, $C_h^A(Y) = C_h^A(Y \setminus \{d\})$. Thus, C_h^A satisfies IRC. □

Appendix 2.C Properties of AGAM

Proposition 2.9:

Proof. Below we show that AGAM is q-fair, AUS, and respects assignment guarantees.

(1) q-fairness:
AGAM is the mechanism that uses the AGA Choice Function for programs at each iterative step of the deferred acceptance algorithm. Therefore, if $\exists(d,h)$ such that $h \succ_d \mu^A(d)$, d must have proposed to h at earlier steps before the algorithm concluded, and h can only reject d for the doctors it prefers over d to fill its seats, which implies $z(P_h, \mu^A(h) \cup \{d\}) > q_h$.

(2) AUS:
The AUS property of AGAM straightforwardly follows from the AGA Choice Function and the deferred acceptance procedure. At each step of the DA, the choice function selects applicants such that they are either in the top-q for the program, or they have a guaranteed seat. The top-q candidates admitted in the first step can only be replaced with better-ranked candidates in the following rounds, whereas the guarantees candidates never lose their additional seats. Therefore, the end allocation ensures the same for the selected doctors for all programs.

(3) Assignment Guarantees:
Similar to AUS, respecting assignment guarantees follows from the fact that the AGA Choice Function respects guarantees at each step. A candidate who has a guaranteed seat at h proposes to h only when she is rejected from all other programs that she prefers to h and she once she proposes to h she receives either a top-q seat or an additional seat created because of her guarantee. The status of the seat can only change from top-q to an additional seat but she is never rejected by h.

□

Proposition 2.11:

Proof. AGAM is strategy-proof because the choice function that induces the DA in this mechanism satisfies the substitutes and the Law of Aggregate Demand (LAD) conditions, which are sufficient properties for a deferred acceptance algorithm to be strategy-proof (Hatfield and Milgrom (2005)). □

Appendix 2.D Stability

The minimal requirement one would expect from a mechanism is that it produces a stable matching, which is the most common equilibrium concept in matching theory. Among several different approaches to stability in the literature, we use pairwise stability, intuitively meaning that no parties can individually or mutually be better off by opting out of the mechanism.

Formally, a matching μ is stable if it is:

(1) individually rational, $C_i(\mu(i)) = \mu(i)$ for all $i \in (D \cup H)$.
(2) not blocked, $\nexists(d,h)$ pair such that $\mu(d) \neq h$, $C_d(\mu(d) \cup h) = h$ and $d \in C_h(\mu(h) \cup d)$.

Below, we show that AGAM always creates a stable outcome with respect to the AGA Choice Function. In fact, we prove this not only by relying on the properties of the choice function but also by rigorously showing that the outcome is always individually rational and there are no blocking pairs. Furthermore, we show the properties of q-fairness, AUS, and respecting guarantees together are equivalent to the stability with respect to the AGA choice function in our framework, which is yet another way of proving the stability of the outcome of AGAM.

Proposition 2.21. *AGAM produces a stable outcome with respect to the AGA Choice Function.*

Proof. As mentioned above, one proof would be relying on the properties of the choice function. Since C_h^A satisfies substitutes, LAD, and IRC conditions, the existence of a stable outcome is guaranteed. Furthermore, the deferred acceptance algorithm induced by this choice function creates a stable outcome.

Alternatively, we show that the mechanism is individually rational and there are no blocking pairs.

(1) Individual Rationality: Along the DA, doctors only propose to acceptable programs and programs only accept doctors that either belong to top q or have an assignment guarantee, which corresponds to AUS for programs.

(2) Blocking pairs: Suppose (d, h) constitute a blocking pair under μ^A. The nature of DA requires d having proposed to h. Since they are not matched under μ^A, h rejects d after the proposal because h has employed at least q_h candidates that are preferred over d and d has no guarantee at h.

□

Theorem 2.22. *Any mechanism satisfies q-fairness, avoids unnecessary slots, respects assignment guarantees, and is individually rational for doctors if and only if it is stable with respect to the AGA Choice Function.*

Proof. $\Rightarrow$ Suppose a mechanism ϕ is individually rational for the doctors, satisfies q-fairness, avoids unnecessary slots, respects assignment guarantees, and however, is not stable. Since we already assumed it is individually rational for the doctors, it can only fail IR for the programs. A doctor d is not acceptable for program h under μ^ϕ if $d \in \mu^\phi(h)$ but $d \notin C_h(\mu^\phi(h))$. Doctor d not being chosen implies d not being one of the top q_h candidates in $\mu^\phi(h)$ and $d \notin E_h$, which conflicts with ϕ avoiding unnecessary slots. Therefore, ϕ has to be individually rational.

The only possibility of ϕ not being stable is then blocking pairs. Suppose (d, h) such that $d \in C_h(\mu^\phi(h) \cup \{d\})$ and $h = C_d(\mu^\phi(d) \cup h)$. Since ϕ respects assignment guarantees, $d \notin E_h$. Then $d \in C_h(\mu^\phi(h) \cup \{d\})$ implies $z(d|P_h, \mu^\phi(h)) \leq q_h$, which conflicts with ϕ satisfying q-fairness.

$\Leftarrow$ Suppose a mechanism ϕ is stable with respect to the AGA Choice Function. It is individually rational for the doctors since it is stable. Suppose it does not respect assignment guarantees. This means $\exists(d, h)$ such that $\mu^\phi(d) = h'$, $h \succ_d h'$, and $d \in E_h$. In that case, (d, h) would constitute a blocking pair with respect to the AGA Choice Function (because $C_d(\mu^\phi(d)\ cup\{h\}) = h$ and $C_h(\mu^\phi(h) \cup \{d\})$ is either $\mu^\phi(h) \cup \{d\}$ or $\mu^\phi(h) \setminus \{d'\} \cup \{d\}$ for some d'). With the same logic, ϕ has to satisfy q-fairness (or the candidate who is ranked within capacity would form a blocking pair with the respective program). Lastly, suppose ϕ does not avoid unnecessary slots, i.e. $\exists(d, h)$ such that $\mu^\phi(d) = h$ but $z(d|p_h, \mu^\phi(h)) > q_h$ and $d \notin E_h$, which means d received a seat at h despite her ranking and having no guarantee. The AGA Choice Function would then reject at least d, $d \in \mu^\phi(h) \setminus C_h(\mu^\phi(h))$, which would imply ϕ not being individually rational for the programs. □

Appendix 2.E Deviation

Theorem 2.13:

Proof. The theorem is equivalent to the following: $\nexists$ a mechanism ϕ which satisfies q-fairness and respects guarantees, along with a problem $(D, H, P_D, P_H, q_H, E_H)$ such that $\exists h$ for which $max\{\mu^\phi(h) - q'_h, 0\} < max\{\mu^A(h) - q'_h, 0\}$.

Suppose there exists such ϕ, which satisfies q-fairness, respects guarantees, and creates less deviation at program h from the target capacities than AGAM for a fixed problem $(D, H, P_D, P_H, q_H, E_H)$. The existence of a mechanism ϕ with less deviation implies that there is excess employment under AGAM. The only way there is excess employment under AGAM is that for some h the least preferred doctor $d \in \mu^A(h)$ is a doctor with a guarantee (because of AUS). Since ϕ results in less deviation, the excess capacity creation at h must be strictly reduced.

Step 0: Furthermore, we also know that if ϕ creates less deviation, it must place some doctor, who was placed elsewhere under μ^A to a vacant capacity. This can happen via cycles and chains, but there has to be at least 1 chain. Note that this vacant capacity can either be at another hospital, or it can be the unemployment scenario. Call this doctor, who is placed in a vacant capacity under ϕ as d_n. Recall that AGAM is individually rational and non-wasteful. Hence, we know that d_n prefers $h_n = \mu^A(d_n)$ to this vacant capacity. Because ϕ respects guarantees, it also follows that $d_n \notin E_{h_n}$. Hence, $d_n \neq d$.

Step 1: Since ϕ also satisfies q-fairness, there must be at least q_{h_n} doctors under μ^ϕ at h_n that h_n prefers to d_n. In words, this means all the seats in the target capacity of the program must have been filled with better candidates so that d_n can not reclaim her seat at h_n. This also means that there is at least one doctor who is preferred to d_n, **who was matched to somewhere else under AGAM**, but is assigned to h_n under ϕ. Call the by-h_n-least-preferred doctor among those as d_{n-1}. Let d_n hypothetically point to d_{n-1} and let $h_{n-1} = \mu^A(d_{n-1})$. If d_{n-1} prefers h_n over h_{n-1}, AGAM would fail q-fairness, which by definition is impossible. Hence d_{n-1} must prefer h_{n-1} over h_n. Because ϕ respects guarantees, it follows that $d_{n-1} \notin E_{h_{n-1}}$.

⋮

Step k: Since ϕ also satisfies q-fairness, there must be at least $q_{h_{n-(k-1)}}$ doctors under μ^ϕ at $h_{n-(k-1)}$ that $h_{n-(k-1)}$ prefers to $d_{n-(k-1)}$. This also means that there is at least one doctor who is preferred to $d_{n-(k-1)}$, **who was matched to somewhere else under AGAM**, but is assigned to $h_{n-(k-1)}$ under ϕ and **who has not been pointed until this step**. Call the least preferred doctor among those as $d_{n-(k-1)-1}$. Let $d_{n-(k-1)}$ hypothetically point to $d_{n-(k-1)-1}$ and let $h_{n-(k-1)-1} = \mu^A(d_{n-(k-1)-1})$. If $d_{n-(k-1)-1}$ prefers $h_{n-(k-1)}$ to $h_{n-(k-1)-1}$, AGAM would fail q-fairness, which by definition is impossible. Hence $d_{n-(k-1)-1}$ must prefer $h_{n-(k-1)-1}$ to $h_{n-(k-1)}$. Because ϕ preserves assignment guarantees, it follows that $d_{n-(k-1)-1} \notin E_{h_{n-(k-1)-1}}$.

Observe that this induction can be traced back with finitely many steps until all the relocated candidates are pointed. Furthermore, since ϕ strictly reduces the deviation at h, the chain's last step is a doctor that was employed at h under AGAM. Then, there are two cases to consider:

(1) If we encounter d in one of the steps:
Recall that d was the least preferred doctor along the doctors in $\mu^A(h)$. We cleared before that $d \in E_h$. By construction of the pointing, $\mu^{\phi}(d) = h' \neq h$. Now, if d prefers h over h', ϕ does not respect her guarantee at h. If d prefers h' over h then AGAM fails q-fairness, contradiction.

(2) If we don't encounter d in one of the steps:
This means that $d \in \mu^{\phi}(h)$ as well. To reduce the deviation, some other doctor $d' \in \mu^A(h)$, such that $d' \succ_h d$ must have been relocated to some other program, hence constitutes one end of the chain, $d' = d_1$. For notational convenience, rename $\mu^A(d_i) = h_i$ and $\mu^{\phi}(d_i) = h_{i+1}$ for all $i = 1, ..., n$. The chain takes d_1 from h_1 to h_2, takes d_2 from h_2 to h_3,...., until the under AGAM vacant capacity under h_{n+1} is reached. From the construction of the chain, no candidate has a guarantee at their assignment under μ^A.
Since AGAM is AUS and d_1 has no priority at h_1, d_1 is ranked within top q_{h_1} for h_1. Therefore, it must be that $h_2 \succ_{d_1} h_1$ so that d_1 and h_1 do not block mu^{ϕ} (q-fairness). If $d_1 \succ_{h_1} d_2$ contradicts with q-fairness of AGAM, so $d_2 \succ_{h_1} d_1$. Again, $h_3 \succ_{d_2} h_2$ for q-fairness of ϕ. Following these steps, we have $h_{i+1} \succ_{d_i} h_i$ and $d_{i+1} \succ_{h_{i+1}} d_i$. However, for h_n, $d_n \succ_{h_n} d_{n-1}$ contradicts with ϕ's q-fairness (d_n is at a vacant seat) and $d_{n-1} \succ_{h_n} d_n$ contradicts with AGAM's q-fairness (d_{n-1} and h_n would block mu^A).

Another possibility is that the chain does not include the least preferred doctor in $\mu^A(h)$ but another one among the least preferred doctors who are ranked outside the target capacity. Such a doctor has also a guaranteed seat at h and the proof goes through. □

Let us illustrate the arguments of the proof with the example below. Similar to the proof, we try to construct an alternative outcome with less deviation from the target capacities.

In order to consider less deviation from the target capacities, the outcome of AGAM must have placed excessive residents in at least one program. There must be an existing candidate of this program, who wanted to use her guarantee to be placed in the same program. In the example, let the set of programs and doctors be $H = \{h_1, h_2, h_3\}$ and $D = \{d_1, d_2, d_3\}$ respectively, with each program having a target capacity of 1. Suppose the outcome of AGAM is such that $\mu^A(h_1) = \{d_3, d_1\}$, $\mu^A(h_2) = \{d_2\}$ and $\mu^A(h_3) = \emptyset$. $h_3 = \emptyset$ is analogous to some vacant seat at a program or the unemployment case. Let us without loss of generality assume

$d_3 \succ_{h_1} d_1$ and impose no further score relations on the candidates. For the other case, the flow will be analogous.

From the structure above, we already have some information:

- h_1 likes d_3 more than d_1 and d_1 creates excess capacity at $h_1 \implies d_1 \in E_{h_1}$.
- $\mu^A(h_3) = \emptyset \implies$ each candidate prefers own allocation under μ^A to h_3.
- d_1 used her guarantee to be placed at $h_1 \implies$
  she either prefers h_1 to h_2 or she prefers h_2 to h_1 but h_2 prefers d_2 over d_1.

Now consider the following alternative allocations μ^ϕ which deviate less than μ^A. In any alternative, we start the chain from the residential candidate placed at the vacant seat at h_3. This doctor will (hypothetically) point to the lowest-scored doctor who claimed her seat at her previous assignment by outscoring her. If the chain includes d_1, ϕ violates guarantees or AGAM violates q-fairness. If the chain does not include d_1, ϕ violates q-fairness. In either case, we find a contradiction that ϕ creates less deviation while respecting guarantees and satisfying q-fairness.

- $\mu(h_1) = d_1$, $\mu(h_2) = d_2$, $\mu(h_3) = d_3$
  d_3 cannot point to anyone, the chain ends without starting
  Since $d_3 \succ_{h_1} d_1$, ϕ fails q-fairness.
- $\mu(h_1) = d_1$, $\mu(h_2) = d_3$, $\mu(h_3) = d_2$
  d_2 points to d_3, d_3 cannot point to anyone, the chain ends. $h_2 \succ_{d_3} h_1$ (or AGAM is not q-fair). If $d_3 \succ_{h_2} d_2$, AGAM fails q-fairness. If $d_2 \succ_{h_2} d_3$, ϕ fails q-fairness.
- $\mu(h_1) = d_2$, $\mu(h_2) = d_1$, $\mu(h_3) = d_3$
  d_3 points to d_2, d_2 points to d_1.
  We encounter d_1, a guaranteed candidate. $d_2 \succ_{h_1} d_3$ (or ϕ is not q-fair), $h_2 \succ_{d_2} h_1$ (or AGAM is not q-fair), $d_1 \succ_{h_2} d_2$ (or ϕ is not q-fair). If $h_2 \succ_{d_1} h_1$ AGAM is not q-fair, if $h_1 \succ_{d_1} h_2$ ϕ fails guarantees.
- $\mu(h_1) = d_2$, $\mu(h_2) = d_3$, $\mu(h_3) = d_1$
  d_1's guarantee at h_1 is not respected.
- $\mu(h_1) = d_3$, $\mu(h_2) = d_1$, $\mu(h_3) = d_2$
  d_2 points to d_1, the chain ends. $d_1 \succ_{h_2} d_2$ (or ϕ is not q-fair). If $h_2 \succ_{d_1} h_1$ AGAM is not q-fair, if $h_1 \succ_{d_1} h_2$ ϕ fails guarantees.
- $\mu(h_1) = d_3$, $\mu(h_2) = d_2$, $\mu(h_3) = d_1$
  d_1's guarantee h_1 is not respected.

Observe that leaving $\mu(h_3) = \emptyset$ as it was in the μ^A and sending the candidates to unemployment will not be possible due to similar arguments as above. So, we can conclude that it is not possible to create an allocation with less deviation than AGAM outcome, whilst preserving assignment guarantees and satisfying q-fairness.

References

Abdulkadiroğlu, Atila, and Tayfun Sönmez. 1999. "House Allocation with Existing Tenants." *Journal of Economic Theory* 88 (2). https://doi.org/10.1006/jeth.1999.2553. [49]

Abdulkadiroğlu, Atila, and Tayfun Sönmez. 2003. "School choice: A mechanism design approach." *American Economic Review* 93 (3). https://doi.org/10.1257/000282803322157061. [49]

Aygün, Orhan, and Günnur Ege Bilgin. n.d. "Preserving Priorities in the Case of the Re-Placement of Medical Residents in Turkey," Work in progress. [61]

Aygün, Orhan, and Tayfun Sönmez. 2013. "Matching with contracts: Comment." *American Economic Review* 103 (5). https://doi.org/10.1257/aer.103.5.2050. [49]

Aygün, Orhan, and Bertan Turhan. 2020. "Dynamic reserves in matching markets." *Journal of Economic Theory* 188. https://doi.org/10.1016/j.jet.2020.105069. [49, 57]

Balinski, Michel, and Tayfun Sönmez. 1999. "A Tale of Two Mechanisms: Student Placement." *Journal of Economic Theory* 84 (1). https://doi.org/10.1006/jeth.1998.2469. [49]

Biró, Péter, and Sofya Kiselgof. 2015. "College admissions with stable score-limits." *Central European Journal of Operations Research* 23 (4). https://doi.org/10.1007/s10100-013-0320-9. [62]

Combe, Julien, Umut Mert Dur, Olivier Tercieux, Camille Terrier, and M Utku Ünver. 2022. *Market design for distributional objectives in (re) assignment: An application to improve the distribution of teachers in schools.* MIT Department of Economics. http://fmwww.bc.edu/EC-P/wp1050.pdf. [49]

Dimakopoulos, Philipp D., and C. Philipp Heller. 2019. "Matching with waiting times: The German entry-level labor market for lawyers." *Games and Economic Behavior* 115. https://doi.org/10.1016/j.geb.2019.03.009. [50]

Doğan, Battal. 2016. "Responsive affirmative action in school choice." *Journal of Economic Theory* 165. https://doi.org/10.1016/j.jet.2016.04.007. [49]

Gale, D., and L. S. Shapley. 1962. "College Admissions and the Stability of Marriage." *American Mathematical Monthly* 69 (1). https://doi.org/10.1080/00029890.1962.11989827. [49]

Hafalir, Isa E., M. Bumin Yenmez, and Muhammed A. Yildirim. 2013. "Effective affirmative action in school choice." *Theoretical Economics* 8 (2). https://doi.org/10.3982/te1135. [49]

Hatfield, John William, and Fuhito Kojima. 2008. "Matching with Contracts : Comment." *American Economic Review* 98 (3). https://doi.org/10.1257/aer.98.3.1189. [49]

Hatfield, John William, and Paul R. Milgrom. 2005. "Matching with Contracts." *American Economic Review* 95 (4): 913–35. https://doi.org/10.1257/0002828054825466. [49, 67]

Hirata, Daisuke, and Yusuke Kasuya. 2017. "On stable and strategy-proof rules in matching markets with contracts." *Journal of Economic Theory* 168. https://doi.org/10.1016/j.jet.2016.12.005. [58]

Kojima, Fuhito. 2012. "School choice: Impossibilities for affirmative action." *Games and Economic Behavior* 75 (2). https://doi.org/10.1016/j.geb.2012.03.003. [49]

Kojima, Fuhito, Parag A. Pathak, and Alvin E. Roth. 2013. "Matching with couples: Stability and incentives in large markets." *Quarterly Journal of Economics* 128 (4). https://doi.org/10.1093/qje/qjt019. [62]

Kominers, Scott Duke, and Tayfun Sönmez. 2016. "Matching with slot-specific priorities: Theory." *Theoretical Economics* 11 (2). https://doi.org/10.3982/te1839. [49]

Ortega Hesles, Maria Elena. 2015. "School Choice and Educational Opportunities: The Upper-Secondary Student-Assignment Process in Mexico City," https://dash.harvard.edu/handle/1/16461054. [62]

Rios, Ignacio, Tomás Larroucau, Giorgiogiulio Parra, and Roberto Cominetti. 2021. "Improving the Chilean college admissions system." *Operations Research* 69 (4). https://doi.org/10.1287/opre.2021.2116. [62]

Roth, Alvin E. 1984. "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory." *Journal of Political Economy* 92 (6). https://doi.org/10.1086/261272. [49, 62]

Roth, Alvin E. 1986. "On the Allocation of Residents to Rural Hospitals: A General Property of Two-Sided Matching Markets." *Econometrica* 54 (2). https://doi.org/10.2307/1913160. [53]

Shapley, Lloyd, and Herbert Scarf. 1974. "On cores and indivisibility." *Journal of Mathematical Economics* 1 (1). https://doi.org/10.1016/0304-4068(74)90033-0. [49]

Sönmez, Tayfun, and Tobias B. Switzer. 2013. "Matching With (Branch-of-Choice) Contracts at the United States Military Academy." *Econometrica* 81 (2). https://doi.org/10.3982/ecta10570. [50, 56]

Westkamp, Alexander. 2013. "An analysis of the German university admissions system." *Economic Theory* 53 (3). https://doi.org/10.1007/s00199-012-0704-4. [49, 50]

Chapter 3

Voting Under Salience Bias and Strategic Extremism*

Joint with Cavit Görkem Destan

3.1 Introduction

The political economy literature classically presumes that office-oriented candidates observe the electorate and take positions matching majority voter preferences. However, there are political candidates worldwide who take extreme positions on some issues and propose radical policies. Yet, some of these politicians get elected and implement their pledged policies.[1]

Besides, we observe another phenomenon. The policy positions of candidates influence the preferences of voters. Consequently, politicians may employ political positioning as a strategic tool to shape preferences into a more favorable distribution. This study presents a model and experimental evidence demonstrating that politicians can optimally choose extreme policies even when no voters have extreme preferences. We examine how policy proposals affect preferences and lead to extreme policies, thereby explaining the extreme policy choices of candidates.

We show that when voters exhibit salience bias (i.e., overemphasize the importance of salient issues), candidates can manipulate this bias by adopting radical stances in a policy where they have an advantage. This way, they draw attention

* We are thankful to Stephan Lauermann, Daniel Krähmer, Florian Zimmermann, and Thomas Dohmen for many discussions, and to Lucas Coffman, Frank Schilbach, and Mehmet Ekmekci for helpful comments. Funding by the Reinhard Selten Institute (RSI) through Selten Grant and support by the German Research Foundation (DFG) through CRC TR 224 (Project B03) are gratefully acknowledged. This study is preregistered under Open Science Framework: https://osf.io/cfgdy . The ethical approval by German Association for Experimental Economic Research e.V.: No. J2S1c2TN

1. For instance Donald Trump builds a multi-billion dollar wall (BBC News (2017)). See Carothers and O'Donohue (2019) for an overview of different countries.

to that issue and create demand by shifting voters' preferences. When we extend the same argument to two candidates, the electoral competition may become an arms-race scenario where each candidate aims to take an extreme position on a different issue and tries to persuade voters that their issue is the most relevant one.

We implement the probabilistic voting model by Persson and Tabellini (2002) with two politicians and a two-dimensional policy platform. As a special case of their model, our voters are not divided into groups and are quite similar to each other in taste, apart from some noise factors. The voters consider the utility they would get from each candidate, which the policy choices of candidates would drive. In addition, the utility of voters is affected by salience. A policy dimension becomes more salient as candidates become more diverse, and more salient issues are overemphasized by the voters, similar to Bordalo, Gennaioli, and Shleifer (2012). The voter then votes for the candidate whose policy choices bring higher utility.

Politicians are aware of the salience bias, and by choosing and committing to the two-dimensional policy proposal, they maximize their probability of winning, i.e., their vote share. They are constrained by the same governmental budget. However, they differ in their marginal costs to implement each policy, which reflects both pecuniary and non-pecuniary costs. Pecuniary costs reflect the candidate's resources, such as tools, factories, and workforce, which enables the candidate to provide policies at lower costs. On the other hand, non-pecuniary costs reflect the candidate's connections. For instance, if the candidate's main supporting lobby favors a policy, then shifting resources to the other is more costly for her.

In any equilibrium of the model with two candidates, always the same issue is salient for both candidates. Both candidates invest more in the salient issue, and extremism is enhanced with the salience bias. The relatively more advantageous candidate in the salient issue can increase her advantage by choosing even higher levels. Which issue will be salient in the equilibrium will be dictated by the parameters of the model, namely the relative cost advantages.

There is a new but sizable literature on extremism in politics. Unlike our approach, most studies assume that there are existing divides in each society, and politicians use that polarization to gain power. However, we show that a radical vote base is unnecessary for extreme policies. Politicians can promote an issue as the most crucial aspect of the election by taking an extreme position. This way, they can manufacture radicalization. We also show that extremism is exacerbated if an issue is already a hot topic. Hence, existing societal polarization would have a multiplicative effect on extremism.

Our model also explains increased mobilization through extremism. When candidates take disparate positions, the welfare difference between candidates gets larger. Thus, voters have a greater incentive to vote. Additionally, our model can analyze run-off elections and the effect of existing polarization. Furthermore, the tractable form of the model can be used in most of the more complex models to investigate various phenomena.

Additionally, we test the model's predictions through an experiment with a representative sample of Turkey. We ask subjects to vote on a hypothetical election where hypothetical candidates differ in their positions on climate and defense policy proposals. The experimental findings support the model and confirm that politicians can increase their vote shares by promising extreme policies. We also show that the salience of an issue is the primary driver of the voting decision, as assumed in the model.

The paper is organized as follows: In Section 3.2, we provide an overview of the related literature. In Section 3.3, the model is described, and the equilibrium analysis is provided in Section 3.4. We analyze comparative statics about the optimal choices in Section 3.5. The possible implications of the model regarding mobilization and second-term elections are explained in Section 3.6. Section 3.7 provides the experimental design and the main results. Section 3.8 concludes.

3.2 Literature

This paper lies at the intersection of two strands of literature: extremism and salience. In the extremism literature, most studies try to explain radical politicians as a response to radical voters. This bottom-up argument mainly states that the political preferences of (at least some) people in society shift toward extreme attitudes, and politicians take extreme stances to match the demands of their voters.

For instance, Matějka and Tabellini (2021) argues that small groups with stark preferences can alter the political outcomes in their favor. They advocate that electoral candidates give those groups a disproportionately large weight in their policy choices since they are more responsive than moderate voters. Similarly, Jones, Sirianni, and Fu (2022) argues that if voters with moderate preferences are less likely to vote, politicians take extreme positions to attract more eager radical voters.

Furthermore, there are studies analyzing extremism as a result of identity politics (Kuziemko and Washington (2018), Grossman and Helpman (2021)), communalism (Enke (2020), Enke, Rodríguez-Padilla, and Zimmermann (2022)), globalization (Rodrik (2021)), and polarization (Nunnari and Zápal (2017), Bursztyn, Egorov, and Fiorin (2020), Enke, Polborn, and Wu (2022)).

On the other hand, many studies show that the salience of an issue is a critical factor in voters' decisions. Colussi, Isphording, and Pestel (2021) clearly shows that anti-Muslim parties gain votes if the elections are held right after Ramadan. They also demonstrate the effect of the salience of Muslim minorities as the primary mechanism. Likewise, Aragonès and Ponsatí (2022) depicts a similar phenomenon using the data from the UK and Catalonia. They show that political parties adjust their positions when an exogenous shock makes an issue more salient.

On top of that, some studies show the effect of salience is exacerbated when combined with existing stereotypes. Bordalo, Gennaioli, and Shleifer (2020) and Bonomi, Gennaioli, and Tabellini (2021) show that a salient divide in society creates radical preferences via negative stereotypes. Furthermore, Spirig (2023) shows the strength of salience using Swiss data. When immigration becomes more salient, not only the voter preferences but also the decisions of judges become less favorable for minorities.

The over- or underweighting of specific policies or dimensions that shape the agent's utility differently can also be analyzed by rational inattention of the agents. For instance, Matějka and McKay (2015) and Caplin, Dean, and Leahy (2019) analyze agents who are aware of attention costs and avoid evaluating some options on purpose to avoid that cost. In salience bias, on the other hand, the voters' attention is involuntarily attracted towards extreme values.

Furthermore, some studies show that politicians strategically manipulate the salience of some issues to gain an advantage. For instance, Lewandowsky, Jetter, and Ecker (2020) provides evidence for Donald Trump using Twitter to manipulate the salience of some issues. Similarly, Glaeser (2005) shows that politicians can supply hate stories to shape the preferences of individuals. Balart, Casas, and Troumpounis (2022) also shows that politicians can exploit social media platforms to push radical opinions.

However, none of the papers in the literature examines the positioning of candidates as a potential manipulation of the salience of different issues. Yet, the idea of politicians positioning themselves in different attributes is very similar to firms choosing different price, quality, and quantity levels to compete with other firms. Although there are differences between firms and politicians, the closest resemblance to our model can be found in IO literature. Several papers show that firms design their menus to influence the salience of some aspects of products. The canonical paper by Bordalo, Gennaioli, and Shleifer (2016) (together with Bordalo, Gennaioli, and Shleifer (2013)) provides a model that explains the product choices of firms to exploit the salience bias of consumers.[2]

3.3 Model

Two purely office-oriented candidates are running for the election, $i = \{A, B\}$. Both candidates announce and commit to two policy choices $q = (x_i, y_i) \in \mathbb{R}^2_+$, representing the government spending they will allocate to the two subjects.

2. See the book chapter by Herweg, Müller, and Weinschenk (2018) for an analysis of these models and their implications.

There is a continuum of voters. Voters do not have the option to abstain. Following the Probabilistic Voting Model by Persson and Tabellini (2002), they simply vote for the candidate whose policy proposal is more favorable. Observing the policy choices, a single voter's utility from candidates is as follows:

$$v(i) = \ln x_i + m \ln y_i$$

However, we assume that the agents have bounded rationality and their attention is limited a la BGS. To be more specific, as the policies in one spectrum are wider spread from each other, this drives the voters' attention to that aspect, increasing the relative utility weight of that issue in their utility function. In particular, the policy choices of the politicians affect voter preferences such that for $\delta > 1$:

$$v(i) = \begin{cases} \delta \ln x_i + m \ln y_i & \text{if } \frac{|x_i - \bar{x}|}{\bar{x}} > \frac{|y_i - \bar{y}|}{\bar{y}} \\ \ln x_i + m \ln y_i & \text{if } \frac{|x_i - \bar{x}|}{\bar{x}} = \frac{|y_i - \bar{y}|}{\bar{y}} \\ \ln x_i + \delta m \ln y_i & \text{if } \frac{|x_i - \bar{x}|}{\bar{x}} < \frac{|y_i - \bar{y}|}{\bar{y}} \end{cases}$$

BGS uses a more general salience function. However, in this version of the paper, we are restricting our attention to a more specific one, which indicates that a policy attribute is more salient for a candidate whenever he deviates from the average spending relative to the other policy. Other than the partiality due to salience, the utility function is the sum of two logarithmic utility functions, with a slight adjustment by m representing the relative importance of issue y for the voters. Voters receive strictly positive utility from both policies; therefore, $m > 0$. If $m < 1$, voters care more about policy x without the interference of the salience bias.

Policy choices are not the only factors that affect voter preferences. Additionally, ideological bias towards candidate B denoted by $\beta \sim U\left[\frac{-1}{2\phi}, \frac{1}{2\phi}\right]$ and relative popularity of B denoted $\epsilon \sim U\left[\frac{-1}{2\varphi}, \frac{1}{2\varphi}\right]$ represent the noise in the elections. Once the candidates select their positions, salience is revealed, and voters calculate the utility they would get from each candidate. Furthermore, the noise factors β and ϵ realize and a voter votes for A if $v(A) > v(B) + \beta + \epsilon$.[3]

Both politicians are trying to maximize their probability of winning, which, with the logic explained above, is equal to $[v(i) - v(j)]\,\varphi + \frac{1}{2}$ for candidate i. Furthermore, they are bounded by a budget constraint $c_x^i x_i + c_y^i y_i = G$. This budget constraint represents each policy's pecuniary and non-pecuniary costs for both candidates. For example, if a candidate possesses tools to ease implementing a policy, he has a lower marginal cost. These tools might be materials such as factories, skilled teams, and other apparatus. However, they could also represent other

3. β realizes for each individual, whereas ϵ realizes as a common variable for the whole electorate.

structures, such as networks and lobbies. If the main lobby that supports a candidate favors policy x, then implementing policy y would be more costly for him. Since voters get positive utility from both policies, we impose $c_x^A < c_x^B$ and $c_y^B < c_y^A$ for a non-trivial analysis of equilibrium policy choices.

Simple intuition would hint at the fact that both candidates want to highlight the dimension in which they have a comparative advantage. At this point, a bridging fact that is shown by BGS simplifies our analysis a lot:

Lemma 3.1. *x is salient by $A \iff x$ is salient by B. (BGS 2012)*

3.4 Equilibrium Analysis

As a result of the features discussed above, a voter with $\tilde{\beta} = v(A) - v(B) - \epsilon$ indifferent between the two candidates and the vote share of A can be calculated as $\Pi_A = \mathbb{P}(\beta < \tilde{\beta}) = \left[\tilde{\beta} + \frac{1}{2\phi}\right]\phi$ and the probability of candidate A winning the election is $\mathbb{P}(\Pi_A > \frac{1}{2}) = \mathbb{P}\left(v(A) - v(B) - \epsilon + \frac{1}{2\phi} > \frac{1}{2\phi}\right) = [v(A) - v(B)]\,\varphi + \frac{1}{2}$

Furthermore, as discussed in the previous section, candidates try to maximize their probability of winning. They only have control over their policy choices and take other candidates' positions as given. Therefore, candidate A's problem is:

$$\max_{\{x_A, y_A\}} [v(A) - v(B)]\,\varphi + \frac{1}{2} \tag{3.1}$$

$$\text{s.t. } c_x^A x_A + c_y^A y_A = G \tag{3.2}$$

A critical analysis requires embranchment after this point. This is because both $v(A)$ and $v(B)$ depend on the salient issue in the election. From the lemma, we know that the same issue will be salient for both candidates. Therefore, we call it the salience issue of the election. As the first branch, suppose there exists an x—salient equilibrium. Then, the maximization problem of candidate A is relatively straightforward:

$$\max_{\{x_A, y_A\}} [\delta \ln x_A + m \ln y_A - \delta \ln x_B - m \ln y_B]\,\varphi + \frac{1}{2} \tag{3.3}$$

$$\text{s.t. } c_x^A x_A + c_y^A y_A = G \tag{3.4}$$

Since the candidates can only affect their own positions, the problem resembles a fundamental utility maximization problem of a consumer with a budget constraint. As usual, optimality of the interior solution requires the following:

$$\frac{\delta y_A}{m x_A} = \frac{c_x^A}{c_y^A}$$

Proposition 3.2. *In an x-salient equilibrium, the optimally chosen policy profiles of both candidates are as in the following table, and the equilibrium indeed is x-salient iff $\frac{c_x^B}{c_x^A} > \frac{c_y^A}{c_y^B}$.*

$x_A^* = \frac{G\delta}{(\delta+m)c_x^A}$	$x_B^* = \frac{G\delta}{(\delta+m)c_x^B}$
$y_A^* = \frac{G\delta}{(\delta+m)c_y^A}$	$y_B^* = \frac{G\delta}{(\delta+m)c_y^B}$

Observe that in such an equilibrium, $x_A^* > x_B^*$ and $y_B^* > y_A^*$. Furthermore, this equilibrium can be sustained if and only if $c_x^B/c_x^A > c_y^A/c_y^B$, meaning that the relative cost advantage of candidate A in policy x should be higher than the relative cost advantage of candidate B in policy y. Furthermore, candidate A wins if and only if $\delta \ln \frac{c_x^B}{c_x^A} - m \ln \frac{c_y^A}{c_y^B} > \epsilon$. The equilibrium policy choices and the necessary condition of a y-salient equilibrium can be found in the appendix.

3.5 Comparative Statics

In this section, we provide comparative statics of the equilibrium and provide explanations. First of all, in both x-salient and y-salient equilibria, $x_A^* > x_B^*$ and $y_B^* > y_A^*$. This is not related to salience but is solely due to the different cost functions of the candidates. Each candidate prefers higher amounts in the policy that is less costly for him.

Moreover, in x-salient equilibrium, x_i^* increases with δ and in y-salient equilibrium, y_i^* increases with δ. This explains that politicians respond to salience in the sense that they provide more on the salient issues. Thus, the salience has an overshooting effect such that voters' utility from the salient issue increases even more.

The probability of candidate A winning the election in an x-salient equilibrium increases with the salience of x and the cost advantage of A in policy x and decreases with the relative importance of issue y and the cost advantage of B in policy y as expected.

Observe that A prefers an x-salient equilibrium since he has the absolute advantage and will provide more than B in any case. However, which equilibrium is to be sustained will be determined by exogenous variables and the candidates have no means of choosing the equilibrium. With two candidates, they respond to salience only by choosing their own policies, not by the salience structure of the equilibrium.

However, even with this simple strategic behavior, in x-salient equilibrium, $x_A^* - x_B^*$ increases with δ and $y_B^* - y_A^*$ decreases with δ. This sustains the salience bias in policy x.

In the following section, we consider an extension to the model where another candidate is introduced into the environment.

3.6 Extensions and Implications

3.6.1 Introduction of a Decoy Candidate

Similar to the industrial organization literature, an interesting implication of this model occurs when a decoy candidate appears on the election platform. In marketing, the *decoy effect* is the phenomenon whereby consumers tend to have a specific change in preference between two options when also presented with a dominant third option. In social choice, it is known as *indepence of irrelevant alternatives* (Cane and Luce (1960)), and in matching theory, the notion corresponds to *irrelevance of rejected contracts* (Aygün and Sönmez (2013)). The flavor is similar in any of the fields: An alternative that the decision-makers will not choose should not affect the choice process at all.

In this paper, a candidate is a *decoy* if he is unlikely to be chosen but affects the election outcomes by interfering with salience. We show that, for a given policy choice, an initially disadvantageous candidate might benefit from the existence of a decoy candidate.

Consider an initial setup where candidates A and B choose relatively moderate locations in policy y. In contrast, their policy choices are spread wider in policy x, so policy x is the salient issue for both candidates. Additionally, suppose B chooses a higher level of x and A chooses a higher level in y for non-triviality. In such a scenario, candidate B has a relatively upper hand by choosing more in the salient issue.

Next, we introduce a third candidate, C, in the election. Candidate C is a far-extremist in policy y and will not allocate any budget to policy x. This simple assumption ensures that candidate C will not be chosen in any equilibrium due to the utility function of the voters. The following proposition shows that, even though any voter will not vote for C, his existence can affect the election outcome by interfering with salience and salience only.

Proposition 3.3. *Suppose the alignment of the candidates is as in the table below, and* $h > \varepsilon > 0$, $\frac{h}{\bar{x}} > \frac{\varepsilon}{\bar{y}}$ *and* $\bar{x} > h > \frac{\bar{x}}{3}$.

	A	B	C
x	$\bar{x} - h$	$\bar{x} + h$	0
y	$\bar{y} + \varepsilon$	$\bar{y} - \varepsilon$	ω

Then, introduction of an extremist candidate C *where* ω *is large enough* ($\omega > \frac{4\bar{x}\bar{y}+6\bar{y}h+6\bar{x}\varepsilon}{3h-\bar{x}}$ *and* $\omega > \frac{2\bar{y}h-2\bar{x}\varepsilon}{\bar{x}-h}$) *increases the vote share of candidate* A *if* $m\ln(\bar{y}+\varepsilon) > \ln(\bar{x}-h)$.

First, observe that candidate C's choice of 0 in policy x indeed ensures he is not elected. Candidate A would prefer making y salient in the initial positions. With the far extremist candidate C, policy x is still salient for candidate B. However, with three candidates, it is now possible that different issues are salient for different candidates. If C is extremist enough, policy y becomes salient for candidate A. If the utility A creates with policy y exceeds the utility A creates with x, policy y becoming salient for A increases the probability of him winning the election.

The proposition shows that if voters' rationality is bounded by salience bias, introducing a third candidate can interfere with the election outcome, even though the third candidate is *irrelevant*, so he does not attract any votes. This candidate only serves as an agenda setter and attracts voters' attention to the policy, in which the initially disadvantageous candidate has a comparative advantage.

3.6.2 Polarization in the Electorate

For this extension, suppose there is an existing polarization in the electorate. Namely, apart from their ideological bias towards candidate B, the voters also differ in the importance they attribute to policy y. Recall that in the benchmark model, m reflected the relative importance of policy y from the voters' perspective. Now, a voter either belongs to the group that intrinsically cares less about policy y with m_L (with probability p) or more with m_H (with probability $1-p$), where $m_L < m < m_H$.

Solving the model for such parameters shows that the optimal policy choices of the candidates depend only on the average relative importance of policy y in the society, namely $pm_L + (1-p)m_H$. How the optimal policy choices and winning probabilities change is the same question as the comparative statics concerning m. Interestingly, the candidates' positions are not affected as long as the weighted average of relative importance remains the same in the electorate.

3.6.3 Mobilization

In line with the probabilistic voting model, our agents vote for the candidate they like better. However, we could also consider a scenario where voters do not simply go to the ballot box. Instead, similar to Coate, Conlin, and Moro (2008), they might require the election to be sufficiently important. The utility difference between the two candidates reflects the importance of the election. The following proposition suggests that as the salience bias strengthens, no abstention is ensured, and all voters vote.

Proposition 3.4. *Suppose voting is costly, and voters vote if and only if the utility difference they get from both candidates exceeds the cost of voting. If the cost of voting is bounded from above, i.e., $c_v < \infty$, $\exists \underline{\delta} < \infty$ such that for all $\delta > \underline{\delta}$ everybody in the electorate votes.*

The above proposition suggests that, apart from affecting candidate positioning, salience bias can also be a factor that incentivizes people to vote. Therefore, increasing the salience of an issue can be used as a tool to increase voter turnout.

3.7 Experiment

As our theoretical framework suggests plausible dynamics, we also conduct a supplementary experiment to test whether the implications are applicable in real life. Namely, in the experiment, we test whether the prediction of the model about the positive effect of extremism on the vote share.

The main goal of the experiment is to investigate two conjectures of the model. First, we check if a candidate can gain more votes by choosing an extreme policy. Secondly, we assess attention as the primary driver of policy preferences and voting decisions.

3.7.1 Experimental Design

The experiment is in a survey format. Each participant answers simple questions using the online platform. Our main goal is to test the model's predictions in a stylized context. Specifically, participants are presented with a hypothetical election scenario and asked to vote for one of the two candidates. The positions of the hypothetical candidates regarding climate and defense policies are either extreme or moderate (2x2 design). The experiment is in a between-subject design; hence, subjects are only aware of a single scenario. The timeline of the experiment is as follows:

(1) Demographics: In this part, we ask simple demographic questions about age, gender, education, employment, city of residence, and per-person income in the household.

(2) Political Engagement: We use agreement with four statements to measure general interest in politics. The statements are about following the news, attachment to an ideology, being influenced by the election polls, and regular voting. We also ask participants whether they have ever voted and are registered political party members.

(3) Issue Ranking: We ask them to rank political issues such as health services, economic stability, and freedom of speech according to subjective importance. We mainly focus on the ranking of climate and defense policies.

(4) Voting: We present hypothetical candidates (A and B) and ask participants to vote for one. They see the information about the verbal proposals of candidates on climate and defense policies, in addition to their age, gender, education, and family status. Climate and defense policy can be extreme or moderate for both candidates. Treatment manipulation is implemented here.

(5) Key Factors: We ask participants to state the factors crucial for their voting choice in the previous question. We use this question to detect the salient issues.

(6) Donation: Participants are asked to divide 10.000 Turkish Liras among two charities. One participant will be randomly selected, and her choice of donations will be implemented. One charity (TEMA) is one of the most significant associations in Turkey that focuses on the environment, whereas the other charity supports the war veterans and families of martyrs. The donations reflect the importance of climate and defense policies, respectively.

Participants will be randomly allocated to one of the four (2×2) treatments differing only in the voting question:

- Moderate-Moderate (MM) Treatment: There are two candidates with moderate proposals on both climate and defense policies.
- Extreme-Moderate (EM) Treatment: Two candidates have extreme and opposing views on climate policies. One promises urgent solutions to the climate crisis, and the other does not find it necessary to take any action. Defense proposals are moderate.
- Extreme-Moderate (ME) Treatment: Two candidates have extreme and opposing views on defense policies. One considers border security a top priority, and the other needs to attach more importance to it. Climate proposals are moderate.
- Extreme-Extreme (EE) Treatment: There are two candidates with extreme and opposing views on climate policies. Defense proposals are moderate.

3.7.2 Experimental Results

The experiment was conducted with 604 participants in September 2022 in Turkey with a representative sample of the country's adult population in terms of geographical region, age, gender, and socioeconomic status. A third-party company collected the data to reach a representative subject pool. We conducted the experiment specifically in Turkey because the political conjuncture is similar to our model environment, where two opposing candidates run presidential elections. The experiment takes around 10 minutes, and the participation fee is 4 Euros.

The main result of the experiment is in line with the model prediction, such that a candidate's vote share increases as she takes more extreme positions in her firm policy. As shown in Table 3.1, participants are likelier to vote for the climate-oriented candidate (Candidate B) when climate policy proposals are extremely different, and vice-versa.

The second result of the experiment is about the underlying channel of this effect. As shown in Table 3.2, people who report that they considered climate

Table 3.1. OLS regression of voting for candidate B on treatment variations. The baseline is the MM treatment in the first two regressions.

	Vote for climate-oriented candidate			
	votes B	votes B	votes B	votes B
EM	0.185***	0.165***		
	(0.0544)	(0.0531)		
ME	-0.119**	-0.128**		
	(0.0544)	(0.0531)		
EE	0.0199	-0.00344		
	(0.0544)	(0.0530)		
extreme climate			0.162***	0.145***
			(0.0384)	(0.0374)
extreme defense			-0.142***	-0.149***
			(0.0384)	(0.0376)
constant	0.351***	0.029***	0.363***	0.033***
	(0.0385)	(0.241)	(0.0333)	(0.241)
N	604	604	604	604
Control vars.		✓		✓

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

proposals while voting have a higher probability of voting for the climate-oriented candidate, and the opposite is true for defense proposals. Crucially, the coefficients are similar when we control for the importance of those policies before the voting decision. Hence, paying more attention to a policy increases the likelihood of voting for the stronger candidate in that policy.

Table 3.2. OLS regression of voting for candidate B on indicators of (self-reported) considered policies and donation for the environmental charity.

	Vote for climate-oriented candidate			
	votes B	votes B	votes B	votes B
considered climate	0.210***	0.185***		
	(0.0398)	(0.0401)		
considered defense	-0.268***	-0.235***		
	(0.0388)	(0.0399)		
donation for climate			0.0331***	0.0277***
			(0.0107)	(0.0106)
constant	1.471***	1.077***	1.213***	0.881***
	(0.0351)	(0.233)	(0.0553)	(0.246)
N	604	604	604	604
Control vars.		✓		✓

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Both findings support the implications of our model. Politicians can attract voters by choosing extreme positions in a policy, and they achieve that by drawing the voters' attention to that specific policy. These results suggest that extremist policies can arise to stand out in the competition and catch voters' attention.

3.8 Conclusion

We provided a model that explains the mechanism behind the extreme policy proposals by electoral candidates. We assume voters involuntarily pay more attention to issues where candidates take extreme positions and overstate the importance of those salient issues. As a result, intrinsically differentiated politicians exploit this bias by strategically positioning themselves in extreme positions and trying to attract attention to their vital issues.

This model shows the top-to-bottom process of extremism and polarization. Different from the existing studies, an already polarized vote base is not necessary, and all the results hold for homogeneous voters. We also show that the supply-driven extremism we propose gets exacerbated by existing societal polarization. Hence, the results of this paper can also be a multiplier of previous findings on extremism.

Additionally, our model clearly shows the effect of extremism on the mobilization of voters. When the candidates take extreme positions to exploit salience bias, the utility difference between them for the voter gets larger. This creates an extra incentive for individuals to vote, which leads to higher turnout.

The model can also be used to analyze the second-round elections. If the opposition party chooses a moderate candidate, extremist politicians in the opposition can help her gain votes by manipulating the salient issues. For instance, in the 2020 US presidential elections, more radical politicians such as Bernie Sanders and Elizabeth Warren may have had a positive impact on Joe Biden by attracting attention to some issues different than Donald Trump's campaign.

We also conducted an experiment with a representative sample to test the predictions of the theory. The results of the experiment provide supportive evidence for our model. The vote share of candidates increases when they take extreme positions and the salience of their strong issues is the main channel of this increase.

The natural next step in this research line would be investigating how to combat this supply-driven extremism. Raising awareness about those strategies and more informative media consumption are promising channels, but their analysis is beyond the scope of this paper.

Appendix 3.A Omitted Proofs

Proof of Lemma 3.1

Proof.

$$\frac{|x_A-\bar{x}|}{\bar{x}} > \frac{|y_A-\bar{y}|}{\bar{y}} \iff \frac{|x_A-\frac{x_A+x_B}{2}|}{\bar{x}} > \frac{|y_A-\frac{y_A+y_B}{2}|}{\bar{y}} \tag{3.A.1}$$

$$\iff \frac{|\frac{x_A-x_B}{2}|}{\bar{x}} > \frac{|\frac{y_A-y_B}{2}|}{\bar{y}} \iff \frac{|\frac{x_B-x_A}{2}|}{\bar{x}} > \frac{|\frac{y_B-y_A}{2}|}{\bar{y}} \tag{3.A.2}$$

$$\iff \frac{|x_B-\frac{x_A+x_B}{2}|}{\bar{x}} > \frac{|y_B-\frac{y_A+y_B}{2}|}{\bar{y}} \iff \frac{|x_B-\bar{x}|}{\bar{x}} > \frac{|y_B-\bar{y}|}{\bar{y}} \tag{3.A.3}$$

□

Values in y-salient equilibrium: In a y-salient equilibrium, the optimally chosen policy profiles of both candidates are as in the following table, and the equilibrium indeed is x-salient iff $\frac{c_y^A}{c_y^B} > \frac{c_x^B}{c_x^A}$.

$x_A^* = \frac{G}{(1+\delta m)c_x^A}$	$x_B^* = \frac{G}{(1+\delta m)c_x^B}$
$y_A^* = \frac{G\delta m}{(1+\delta m)c_y^A}$	$y_B^* = \frac{G\delta m}{(1+\delta m)c_y^B}$

Proof of Proposition 3.3:
At the initial positioning without candidate C, policy x is salient for both candidates. However, with the introduction of candidate C, different policies may become salient for both candidates. The assumptions $h > \varepsilon > 0$ and $\frac{h}{\bar{x}} > \frac{\varepsilon}{\bar{y}}$ ensure that both policies are positive values initially. Furthermore, $\bar{x} > h > \frac{\bar{x}}{3}$ ensures x is salient for candidate B even after C comes on stage.

For ω is large enough ($\omega > \frac{4\bar{x}\bar{y}+6\bar{y}h+6\bar{x}\varepsilon}{3h-\bar{x}}$ and $\omega > \frac{2\bar{y}h-2\bar{x}\varepsilon}{\bar{x}-h}$), policy y becomes salient for candidate A, in which A proposes a higher budget than B. Since candidate C offers 0 in policy x, this candidate does not attract any votes. Then, candidate A benefits from the introduction of C if the utility it creates with policy y is larger than the utility created by the proposal for x.

Polarization in the Electorate:
Suppose that a voter either has m_L with probability p or m_H with probability $(1-p)$. Note that the salience is not affected by m values. Therefore, the valuation for both types is as follows:

$$v_L(i) = \begin{cases} \delta \ln x_i + m_L \ln y_i & \text{if } \frac{|x_i-\bar{x}|}{\bar{x}} > \frac{|y_i-\bar{y}|}{\bar{y}} \\ \ln x_i + m_L \ln y_i & \text{if } \frac{|x_i-\bar{x}|}{\bar{x}} = \frac{|y_i-\bar{y}|}{\bar{y}} \\ \ln x_i + \delta m_L \ln y_i & \text{if } \frac{|x_i-\bar{x}|}{\bar{x}} < \frac{|y_i-\bar{y}|}{\bar{y}} \end{cases}$$

$$v_H(i) = \begin{cases} \delta \ln x_i + m_H \ln y_i & \text{if } \frac{|x_i - \bar{x}|}{\bar{x}} > \frac{|y_i - \bar{y}|}{\bar{y}} \\ \ln x_i + m_H \ln y_i & \text{if } \frac{|x_i - \bar{x}|}{\bar{x}} = \frac{|y_i - \bar{y}|}{\bar{y}} \\ \ln x_i + \delta m_H \ln y_i & \text{if } \frac{|x_i - \bar{x}|}{\bar{x}} < \frac{|y_i - \bar{y}|}{\bar{y}} \end{cases}$$

Among the voters with m_L, voters with $\tilde{\beta_L} = v_L(A) - v_L(B) + \beta + \epsilon$ vote for A and among the voters with m_H, voters with $\tilde{\beta_H} = v_H(A) - v_H(B) + \beta + \epsilon$ vote for A.

Hence, vote share of A boils down to $\phi[p\tilde{\beta_L} + (1-p)\tilde{\beta_H}] + \frac{1}{2}$, which turns A's winning probability into:

$$[p(v_L(A) - v_L(B)) + (1-p)(v_H(A) - v_H(B))]\,\varphi + \frac{1}{2}$$

Therefore, A's problem becomes a weighted average:

$$\max_{\{x_A, y_A\}} [pv_L(A) + (1-p)v_H(A)] \tag{3.A.4}$$

$$\text{s.t. } c_x^A x_A + c_y^A y_A = G \tag{3.A.5}$$

In return, this leads to a replacement of m in the original problem by $pm_L + (1-p)m_H$ in the optimality conditions. Nothing else changes.

Proof of Proposition 3.4:

Suppose we are in an x-salient equilibrium. The utility difference that a voter gets from both candidates is formulated as follows:

$$|\delta lnx_A + mlny_A - \delta lnx_B - mlny_B|$$

Plugging in the equilibrium policy choices of both candidates yield

$$|\delta ln \frac{c_x^B}{c_x^A} + m \frac{c_y^B}{c_y^A}|$$

We know that $\delta \geq 1$ and $m > 0$. Because $c_x^B > c_x^A$ and $c_y^B < c_y^A$, the first term is positive and the latter is negative. If $\delta ln \frac{c_x^B}{>} c_x^A > mln \frac{c_y^A}{<} c_y^B$, the whole term in absolute value is positive and therefore increases with δ.

References

Aragonès, Enriqueta, and Clara Ponsatí. 2022. “Shocks to issue salience and electoral competition.” *Economics of Governance* 23 (1). https://doi.org/10.1007/s10101-022-00267-0. [77]

Aygün, Orhan, and Tayfun Sönmez. 2013. “Matching with contracts: Comment.” *American Economic Review* 103 (5). https://doi.org/10.1257/aer.103.5.2050. [82]

Balart, Pau, Agustin Casas, and Orestis Troumpounis. 2022. “Technological change, campaign spending and polarization.” *Journal of Public Economics* 211: 104666. https://doi.org/10.1016/J.JPUBECO.2022.104666. [78]

BBC News. 2017. “Trump orders wall to be built on Mexico border.” *BBC News* (26, 2017). Accessed January 30, 2022. https://www.bbc.com/news/world-us-canada-38740717. [75]

Bonomi, Giampaolo, Nicola Gennaioli, and Guido Tabellini. 2021. “Identity, Beliefs, and Political Conflict.” *Quarterly Journal of Economics* 136 (4). https://doi.org/10.1093/qje/qjab034. [78]

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2012. “Salience theory of choice under risk.” *Quarterly Journal of Economics* 127 (3). https://doi.org/10.1093/qje/qjs018. [76]

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2013. “Salience and consumer choice.” *Journal of Political Economy* 121 (5). https://doi.org/10.1086/673885. [78]

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2016. “Competition for Attention.” *Review of Economic Studies* 83 (2). https://doi.org/10.1093/restud/rdv048. [78]

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2020. “Memory, attention, and choice.” *Quarterly Journal of Economics* 135 (3). https://doi.org/10.1093/qje/qjaa007. [78]

Bursztyn, Leonardo, Georgy Egorov, and Stefano Fiorin. 2020. “From Extreme to Mainstream: The Erosion of Social Norms.” *American Economic Review* 110 (11): 3522–48. https://doi.org/10.1257/AER.20171175. [77]

Cane, Violet, and R. Duncan Luce. 1960. “Individual choice behavior: A theoretical analysis.” *Journal of the Royal Statistical Society. Series A (General)* 123 (4). https://doi.org/10.2307/2343282. [82]

Caplin, Andrew, Mark Dean, and John Leahy. 2019. “Rational inattention, optimal consideration sets, and stochastic choice.” *Review of Economic Studies* 86 (3). https://doi.org/10.1093/restud/rdy037. [78]

Carothers, Thomas, and Andrew O’Donohue. 2019. *Democracies Divided: The Global Challenge of Political Polarization.* Brookings Institution Press. [75]

Coate, Stephen, Michael Conlin, and Andrea Moro. 2008. “The performance of pivotal-voter models in small-scale elections: Evidence from Texas liquor referenda.” *Journal of Public Economics* 92 (3-4). https://doi.org/10.1016/j.jpubeco.2007.08.007. [83]

Colussi, Tommaso, Ingo E. Isphording, and Nico Pestel. 2021. “Minority salience and political extremism.” *American Economic Journal: Applied Economics* 13 (3). https://doi.org/10.1257/app.20190703. [77]

Enke, Benjamin. 2020. “Moral values and voting.” *Journal of Political Economy* 128 (10). https://doi.org/10.1086/708857. [77]

Enke, Benjamin, Mattias Polborn, and Alex Wu. 2022. “Values as luxury goods and political polarization.” *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.4098323. [77]

Enke, Benjamin, Ricardo Rodríguez-Padilla, and Florian Zimmermann. 2022. “Moral universalism and the structure of ideology.” *Review of Economic Studies*, https://doi.org/10.1093/restud/rdac066. [77]

Glaeser, Edward L. 2005. “The political economy of hatred.” *Quarterly Journal of Economics* 120 (1). https://doi.org/10.1162/0033553053327434. [78]

Grossman, Gene M., and Elhanan Helpman. 2021. “Identity Politics and Trade Policy.” *Review of Economic Studies* 88 (3): 1101–26. https://doi.org/10.1093/RESTUD/RDAA031. [77]

Herweg, Fabian, Daniel Müller, and Philipp Weinschenk. 2018. “Salience in markets.” *Handbook of Behavioral Industrial Organization,* 75–113. https://doi.org/10.4337/9781784718985.00010. [78]

Jones, Matthew I., Antonio D. Sirianni, and Feng Fu. 2022. “Polarization, abstention, and the median voter theorem.” *Humanities and Social Sciences Communications* 9 (1). https://doi.org/10.1057/s41599-022-01056-0. [77]

Kuziemko, Ilyana, and Ebonya Washington. 2018. “Why did the democrats lose the south? Bringing new data to an old debate.” *American Economic Review* 108 (10). https://doi.org/10.1257/aer.20161413. [77]

Lewandowsky, Stephan, Michael Jetter, and Ullrich K.H. Ecker. 2020. “Using the president’s tweets to understand political diversion in the age of social media.” *Nature Communications* 11 (1). https://doi.org/10.1038/s41467-020-19644-6. [78]

Matějka, Filip, and Alisdair McKay. 2015. “Rational inattention to discrete choices: A new foundation for the multinomial logit model.” *American Economic Review* 105 (1). https://doi.org/10.1257/aer.20130047. [78]

Matějka, Filip, and Guido Tabellini. 2021. “Electoral competition with rationally inattentive voters.” *Journal of the European Economic Association* 19 (3). https://doi.org/10.1093/jeea/jvaa042. [77]

Nunnari, Salvatore, and Jan Zápal. 2017. “Dynamic elections and ideological polarization.” *Political Analysis* 25 (4). https://doi.org/10.1017/pan.2017.24. [77]

Persson, Torsten, and Guido Tabellini. 2002. *Political Economics: Explaining Economic Policy.* MIT Press. [76, 79]

Rodrik, Dani. 2021. “Why does globalization fuel populism? Economics, culture, and the rise of right-wing populism.” *Annual Review of Economics* 13. https://doi.org/10.1146/annurev-economics-070220-032416. [77]

Spirig, Judith. 2023. “When issue salience affects adjudication: Evidence from Swiss asylum appeal decisions.” *American Journal of Political Science* 67 (1). https://doi.org/10.1111/ajps.12612. [78]