UNIVERSITY OF CALIFORNIA SAN DIEGO

**Essays in Microeconomic Theory**

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Economics

by

Ce Liu

Committee in charge:

Professor Christopher P. Chambers, Co-Chair
Professor Joel Sobel, Co-Chair
Professor S. Nageeb Ali
Professor Snehal Banerjee
Professor Ramamohan Paturi
Professor Joel Watson

2019

The dissertation of Ce Liu is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Co-Chair

University of California San Diego

2019

# DEDICATION

I dedicate this dissertation to the family members and friends who have provided encouragement and support throughout my life. To Gent, Jin, Marianna, Paul and Yanjun: thank you for your friendship and support throughout this journey.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

## ACKNOWLEDGEMENTS

I would like to thank my thesis committee members for their continuing guidance, support, and training. I'd also like to thank the friends and family who continue to support and encourage me.

Chapter 1, in full, is currently being prepared for submission for publication of the material. Liu, Ce. "Stability in Repeated Matching Markets". The dissertation author was the sole investigator and author of this material.

Chapter 2, in full, is currently being prepared for submission for publication of the material. Chambers, Christopher P; Liu, Ce; Rehbeck, John. "Costly Information Acquisition". The dissertation author was a primary author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in Journal of Economic Theory, 2016. Chambers, Christopher P; Liu, Ce; Martinez, Seung-Keun. "A Test for Risk-Averse Expected Utility". The dissertation author was a primary author of this paper. The copyright of this article is held by Elsevier.

## VITA

| | |
|---|---|
| 2013 | B. A. in Economics, University of Hong Kong |
| 2019 | Ph. D. in Economics, University of California, San Diego |

## PUBLICATIONS

Chambers, Christopher P., Ce Liu, and Seung-Keun Martinez. "A Test for Risk-Averse Expected Utility", *Journal of Economic Theory*, 163 (2016), 775-785.

ABSTRACT OF THE DISSERTATION

**Essays in Microeconomic Theory**

by

Ce Liu

Doctor of Philosophy in Economics

University of California San Diego, 2019

Professor Christopher P. Chambers, Co-Chair
Professor Joel Sobel, Co-Chair

The first chapter studies repeated matching markets, where in every period, a new generation of short-lived workers is matched to a fixed set of long-lived firms on the other. I characterize self-enforcing arrangements for two types of environments. When wages are rigid, as in the matching market for hospitals and medical residents, players can be partitioned into two sets: regardless of patience level, some players can be assigned only according to a static stable matching; when firms are patient, the other players can be assigned in ways that are unstable in one-shot interactions. I also discuss these results' implications for market design. When wages can be flexibly adjusted, I show that repeated interaction resolves well-known non-existence

issues: while static stable matchings may fail to exist with complementarities and/or peer effects, self-enforcing matching processes always exist if firms are sufficiently patient.

The second chapter provides revealed preference characterizations for choices made under various forms of costly information acquisition. We examine nonseparable, multiplicative, and constrained costly information acquisition. In particular, this allows the possibility of unknown time delay for acquiring information. The techniques we use parallel the duality properties in the standard consumer problem.

The third chapter provides a universal condition for rationalizability by risk-averse expected utility preference in a demand-based framework with multiple commodities. Our test can be viewed as a natural counterpart of a classical test of expected utility, due to Fishburn (1975), in a demand setting.

# Chapter 1

# Stability in Repeated Matching Markets

## 1.1 Introduction

**Summary:** College admission, hospital-resident matching, and entry-level hiring are all ongoing matching processes that take place every year. One side of these markets—the institutions—are long-lived players, whereas the other side—namely the students, residents, or workers—participate in the matching process on only a few occasions (sometimes only once). Yet, much of our theoretical analysis of matching environments treats both sides of the market as being short-lived, ignoring the possibility for dynamic incentives that could be used to motivate long-lived players. To understand the scope for such dynamic interactions, this paper develops a framework to study ongoing matching processes in *repeated* matching markets, where one side of the market is long-lived while the other is short-lived. In each period, every long-lived **institution** is matched to multiple short-lived **agents** on the other side of the market.

In this environment, I define and study a solution concept called **self-enforcing matching process**. A matching process is a complete contingent plan specifying a current matching outcome as a function of past histories. A matching process is self-enforcing if it is immune to not only unilateral deviations by institutions or agents, but also blocking coalitions that simultaneously involve an institution and groups of agents over a possibly infinite horizon. A self-enforcing matching process therefore represents self-fulfilling market expectations that are immune to both individual and coalitional deviations. A central theme of my analysis is that in repeated matching markets, long-lived coalition members—the institutions— can be disciplined through continuation play. In other words, dynamic incentives can enforce matching outcomes that are not stable in one-shot interactions. The goal of my paper is to characterize what can be supported, both with and without flexible wages between institutions and their employees.

The first contribution of this paper is to provide a framework and solution concept that takes into account the possibility of institutions' dynamic interactions. In two-sided matching markets, any appropriate stability notion must allow for joint deviations that are carried out by coalitions of agents from opposite sides of the market. Modeling these deviations through a non-cooperative game-theoretic approach, however, would inevitably anchor the analysis to specific extensive-form assumptions on the coalition formation process. Instead, I adopt a cooperative approach consistent with the analysis of static matching environments. I assume that any deviation involving coalitions of players must be profitable for all its members, and study matching processes that admit no such deviations. By meshing coalitional reasoning directly into a repeated matching environment, the resulting framework explicitly accounts for the institutions' intertemporal tradeoffs, while relying only on assumptions on the players' primitive payoff structure.

In Section 1.3, I analyze markets where wages are fixed, as in Gale and Shapley (1962). This is the canonical assumption for the labor market of medical residents. I show that there is a dichotomy of players in this environment: in every self-enforcing matching process, players in the **top coalition sequence**, identified through an algorithm, must be matched according to a *static* stable matching, regardless of institutions' patience level; by contrast, when institutions are patient, it is possible to assign players outside of the top coalition sequence in ways that are unstable in one-shot interactions—I characterize the set of self-enforcing arrangements with patient institutions. In an extreme case, when all agents share a common ordinal ranking over institutions, every player is in the top coalition sequence. As a result, no self-enforcing matching process can enforce any outcome beyond the unique static stable outcome—dynamic enforcement is completely powerless. In the motivation section, I discuss these results' implication for the matching market of medical residents.

Section 1.4 turns to the environment where institutions and their employees can redistribute match surplus through wages—this is the case considered in Kelso and Crawford (1982). I show that in contrast to the findings from the static matching literature, self-enforcing arrangements always exist when institutions are patient, despite complementarities and peer effects in preferences. In particular, treating groups of agents as objects, a Random Serial Dictatorship among the the institutions is always self-enforcing when they are sufficiently patient.

**Motivation:** The interest in repeated matching markets and self-enforcing matching processes is motivated by both normative and positive considerations.

From a normative standpoint, in the matching market for medical residents, dynamic incentives can be harnessed as an instrument for enforcing matching outcomes that favor rural hospitals. In the U.S., rural communities typically feature more elderly patients, higher rate of accidental injuries, and lower average income; the closure of rural hospitals has also been associated to significant changes in medical utilization and the health status of local population. See, for example, Bindman, Keane, and Lurie (1990), Hadley and Nair (1991) and Rosenbach and Dayhoff (1995). These have motivated the interests in policy intervention (U.S. Government Accountability Office 1991); nevertheless, the lack of complete medical facilities, excessive workloads, and cultural and geographic isolation make it difficult for rural hospitals to attract healthcare providers.

According to the Rural Hospital Theorem (Roth 1986), static market incentives are completely powerless for the purpose of allocating more residents to rural hospitals: regardless of what a (static) matching program may propose, a rural hospital failing to fill its vacancies will end up with the exact same matches across all stable market outcomes. A history-dependent matching program, however, can tap into hospitals' dynamic incentives, and propose matching processes that sustain more favorable outcomes for rural hospitals. One particular way of exploiting history dependence is to excludes hospitals that go against the program's proposed matchings from future participation, effectively recommending empty matches to the deviating hospitals in subsequent periods—this is the method used by the Japanese medical resident matching program (JRMP) (Kamada and Kojima 2015). However, in both the U.S. and Japanese markets, matchings proposals are not externally enforced: hospitals can circumvent the program and hire through direct negotiation (see also, for example, Roth 1991). The proposed matching process must then be self-enforcing in order not to create incentives for defection.

The results in Section 1.3 are valuable for understanding what can be sustained by a history-dependent matching program. In particular, Theorem 2 suggests that even in the complete absence of coercive power, a history-dependent matching program can alleviate the staffing shortages faced by rural hospitals—this should be implemented by targeting hospitals outside of the top coalition sequences. Theorem 1, on the other hand, indicates that residents in the top coalition sequence can only be allocated to rural hospitals through alternative means of intervention: existing measures to help rural hospitals, such as student loan repayment/forgiveness programs, should prioritize residents in the top coalition sequence. In fact, as indicated in Corollary 1, if medical residents have highly homogenous preferences over hospitals, no changes in the assignment of residents can be achieved through history-dependence alone.

This results in Section 1.4 make a methodological contribution. Complementarities and peer effects are important features in labor market. But the matching literature has found it difficult to establish existence once such features are incorporated. Theorem 4 fills this gap by identifying a new channel that stabilizes the market, namely the firms' collusive motives and fear of retaliation. Such incentives have previously been overlooked by static stability notions (see, for example, Gale and Shapley 1962, Kelso and Crawford 1982 and Hatfield and Milgrom 2005).

There are real-life examples for such collusive motives in matching markets. For instance, in a 2010 antitrust litigation, the Department of Justice alleged that Adobe, Apple, Google, Intel, Intuit and Pixar had reached unwritten agreements that "eliminated a significant form of competition" in order to suppress the wages paid to their matched employees. In particular, according to the DOJ, these agreements were "not ancillary to any legitimate collaboration", and "much broader than reasonably necessary for the formation or implementation of any collaborative effort", but instead aimed at "disrupting the normal price-setting mechanisms that apply in the labor setting" (Department of Justice 2010).

In the next section, I illustrate the framework and results of this paper through a stylized example based on the NRMP.

### 1.1.1 Motivating Example:

Three hospitals $I_1, I_2$ and $I_r$ each has 2 positions to fill every year. Each year, five medical students $\{a_1, a_2, a_3, a_4, a_5\}$ enter the market looking for internship. In the NRMP, wages are non-negotiable, so players' payoffs are determined by the identity of their match partners.

Students are short-lived players, so only their ordinal preferences are relevant. The left panel of Table 1.1 lists every student's preference ranking over hospitals. Observe that hospital $I_r$ is the least favorite for every student. In the matching literature, $I_r$ is usually interpreted as a hospital located in the rural area that faces difficulties in filling its vacancies.

Hospitals' Bernoulli utility from matched students are shown in the right panel of Table 1.1. Assume each hospital has additively separable utility from matched students, and derives zero utility from unfilled positions. Observe that $a_5$ is every hospital's least preferred student.

**Static Stability:** There are only two static stable matchings in this market: as is shown in Fig. 1.1, the matchings $m_I$ and $m_{\mathcal{A}}$ correspond, respectively, to the hospital-optimal and student-optimal matchings. Both these matchings leave the rural hospital $I_r$ with an unfilled position, while the other position is filled by the worst student $a_5$.

**Table 1.1**: Preferences of Students and Hospitals

| | $\succ$ | | |
|---|---|---|---|
| $a_1$ | $I_2$ | $I_1$ | $I_r$ |
| $a_2$ | $I_2$ | $I_1$ | $I_r$ |
| $a_3$ | $I_1$ | $I_2$ | $I_r$ |
| $a_4$ | $I_1$ | $I_2$ | $I_r$ |
| $a_5$ | $I_1$ | $I_2$ | $I_r$ |

| $u_I(a)$ | 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|---|
| $I_1$ | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ |
| $I_2$ | $a_3$ | $a_2$ | $a_4$ | $a_1$ | $a_5$ |
| $I_r$ | $a_2$ | $a_4$ | $a_3$ | $a_1$ | $a_5$ |

For the welfare of rural population, the matching $m_0$, shown in Fig. 1.2, is a more desirable outcome: $I_r$ still has one unfilled position, but its situation is improved compared to $m_I$ or $m_{\mathcal{A}}$, as it gets to hire the more desirable resident $a_2$. However, $m_0$ is unstable: $I_1$ and $a_2$ will form a blocking pair. This phenomenon is known more generally as the Rural Hospital Theorem:

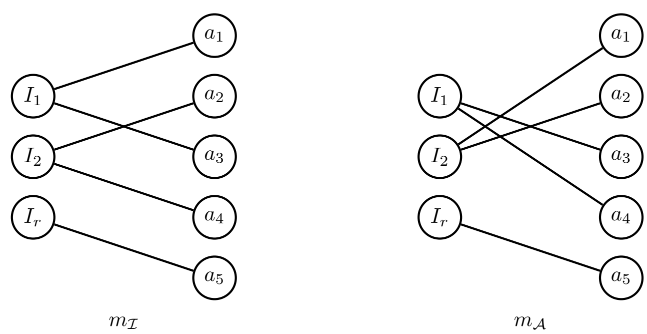


**Figure 1.1**: Static Stable Matchings

**Rural Hospital Theorem.** *(Roth 1986).* When preferences over individuals are strict, any hospital that does not fill its quota at some (static) stable matching is assigned precisely the same set of students at every (static) stable matching. According to the Rural Hospital Theorem, reallocating physicians to rural hospitals will inevitably create unstable matches.

**A Self-Enforcing Matching Process:** Now suppose the matching program can vary proposed matches based on past history. Consider the matching process $\mu^0$ depicted in Fig. 1.3:

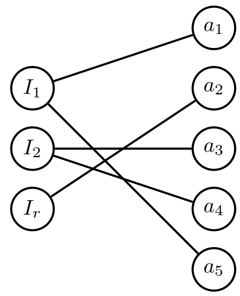


**Figure 1.2**: $m_0$: An Unstable Matching

hospitals and students start off by matching according to $m_0$. So long as no blocking involving a hospital has been observed in the past, players are repeatedly matched at $m_0$; if any blocking involving a hospital is observed, players will be matched according to $m_{\mathcal{A}}$ forever. Hospitals' per-period cardinal payoffs from $m_{\mathcal{A}}$ and $m_0$ are also listed in the diagram for easy reference.

If the market evolves according to $\mu^0$, then in every period the static matching $m_0$ is realized. It remains to check that $\mu^0$ is self-enforcing.

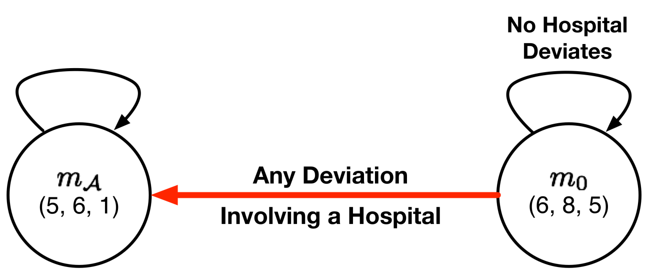


**Figure 1.3**: A Self-Enforcing Matching Process Implementing $m_0$

Students are short-lived players. The only deviation a student can do *unilaterally* is to sever the match with his hospital.

Hospitals, on the other hand, are long-lived players. Hospital $I_1$ can contemplate a **deviation plan** $d^1$, which is a complete contingent plan specifying a group of students, $d^1(h)$, that $I_1$ intends to be matched with at every possible future history $h$. Recall that the matching process $\mu^0$ is also a complete contingent plan. At history $h$, $\mu^0$ also specifies a group of students, denoted $\mu^0(I_1|h)$, for $I_1$. Whenever $d^1(h)$ and $\mu^0(I_1|h)$ are in conflict, $I_1$ can *unilaterally* fire everyone in $\mu^0(I_1|h)\backslash d_1(h)$. Students in $d_1(h)\backslash\mu^0(I_1|h)$, however, need to be poached from other hospitals. Since poaching involves *joint* deviations by both hospital and student, we say $d_1(h)$ is **feasible** when everyone in $d_1(h)\backslash\mu^0(I_1|h)$ finds $I_1$ more desirable than their matched hospital recommended by $\mu^0$.

For the matching process $\mu^0$ to be self-enforcing, it must be immune to both the unilateral and joint deviations outlined above. A one-shot deviation principle, established in the Appendix, implies that in markets with fixed wages like the NRMP, a matching process is self-enforcing if and only if at every possible history of the market:

1. no student wishes to unilaterally sever her match, and
2. no hospital finds it profitable to deviate once with a *feasible* group of student, and permanently revert to following $\mu^0$.

Suppose no deviation involving hospitals has occurred. Every student is getting a strictly positive payoff, so no students would want to unilaterally leave their match. By following $\mu^0$, $I_1$ receives a payoff of 6 every period from matching with $\{a_1, a_5\}$. If $I_1$ were to fire $a_5$ and block with $a_2$, it would enjoys a one-period payoff of 9, but subsequently receive per-period payoff of 5. As long as $I_1$'s discount factor $\delta$ satisfies

$$(1-\delta)9+\delta 5<6,$$

or $\delta > \frac{3}{4}$, such blocking will not be profitable. A similar argument rules out any profitable joint deviations involving $I_2$.

Now suppose a hospital-deviation has occurred. According to $\mu^0$, they should expect to be matched at $m_{\mathcal{A}}$ forever. Again, no students are willing to individually break off from their matches. For hospitals, current period blocking no longer carries any future consequences—the market is permanently stuck at $m_{\mathcal{A}}$ regardless of what transpires in the matching market. Since $m_{\mathcal{A}}$ is statically stable, by definition, a hospital's matched partners in $m_{\mathcal{A}}$ are more desirable than any other students that are feasible, so no hospital has any profitable one-shot blocking either.

In light of the one-shot deviation principle, $\mu^0$ is a self-enforcing matching process. Note that in a static world, pairwise blocking between $I_1$ and $a_2$ would have rendered the matching $m_0$ unstable. In the matching process $\mu^0$, however, one-shot blockings like this will permanently throw the market into a less desirable matching, $m_{\mathcal{A}}$, for $I_1$. Such blocking is no longer jointly profitable.

**The Problem with Common Ranking over Hospitals:** The analysis so far has demonstrated the possibility of using history dependence to enforce outcomes that are not stable in one-shot interactions. Now I illustrate the limitation of dynamic enforcement when all students share a common preference over hospitals.

**Table 1.2**: A Market with Aligned Preferences

| | $\succ$ | | |
|---|---|---|---|
| $a_1$ | $I_1$ | $I_2$ | $I_r$ |
| $a_2$ | $I_1$ | $I_2$ | $I_r$ |
| $a_3$ | $I_1$ | $I_2$ | $I_r$ |
| $a_4$ | $I_1$ | $I_2$ | $I_r$ |
| $a_5$ | $I_1$ | $I_2$ | $I_r$ |

| $u_I(a)$ | 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|---|
| $I_1$ | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ |
| $I_2$ | $a_3$ | $a_2$ | $a_4$ | $a_1$ | $a_5$ |
| $I_r$ | $a_2$ | $a_4$ | $a_3$ | $a_1$ | $a_5$ |

Consider the market in Table 1.2, where everything is identical to the one in Table 1.1, except now all students share a common ranking $I_1 \succ I_2 \succ I_r$ over hospitals.

There is a unique static stable matching $m^*$ in this market, as depicted in Fig. 1.4. I will show that regardless of how patient the hospitals are, no self-enforcing matching process can implement any matching other than $m^*$.

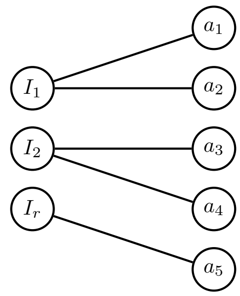


**Figure 1.4**: $m^*$: The Unique Stable Matching

To see why, first observe that as the top choice for all the students, $I_1$ always finds its most preferred students $\{a_1, a_2\}$ feasible: in any matching process $\mu$ and at any history $h$, $\{a_1, a_2\}$ are either already matched to $I_1$, or matched to a hospital that they find less desirable than $I_1$. If $I_1$ is matched with any other students at a history $h$, then $I_1$ has a feasible deviation plan: match with $\{a_1, a_2\}$ at every future history, including $h$. This is profitable because it offers $I_1$ an immediate improvement at $h$, as well as the highest possible continuation value in any matching process. To rule out profitable deviations like this, all self-enforcing matching processes must match $I_1$ with $\{a_1, a_2\}$ regardless of the history of the market.

Since $a_1$ and $a_2$ are always matched to $I_1$ in a self-enforcing matching process, there is no way to reward $I_2$ with a better match than $\{a_3, a_4\}$. On the other hand, the only way to punish $I_2$ with a match worse than $\{a_3, a_4\}$ is to match $a_3$ or $a_4$ with $I_1$—otherwise $I_2$ can simply poach them back. This is, again, impossible because the hiring quota for $I_1$ is always filled by $a_1$ and $a_2$ at every possible history of the market. It follows that in any self-enforcing matching process, $I_2$ is always matched precisely with $\{a_3, a_4\}$, on and off equilibrium path.

A similar "peeling" argument along students' shared preference list ensures $I_r$ is matched with $a_5$ in any self-enforcing matching process. The only self-enforcing matching process therefore always has players matched according to $m^*$.

In the above example, depending on players' preference configuration, history dependence may or may not expand the set of stable outcomes. It is also worth pointing out that the uniqueness of static stable outcome is *not* responsible for the collapse of dynamic enforcement: in fact, if the hospitals share a common (cardinal or ordinal) preference over residents, the market will also have a unique static stable matching; nevertheless, it is still possible to expand the set of stable outcomes through history dependence. The results in Section 1.3 characterizes how the scope of this expansion relates to players' preferences.

### 1.1.2 Related Literature

There is now a growing interest in dynamic interactions in matching markets. The existing literature can largely be divided into two strands, based on the nature of the matching environment being investigated.

The first strand focuses on matching markets where players leave the market permanently once matched: this is the natural assumption for organ transplant or child adoption. In such context, players optimally trade off the cost of waiting against the arrival of better matching opportunities. Ünver (2010), Anderson et al. (2015), Baccara, Lee, and Yariv (2015), Leshno (2017), and Akbarpour, Li, and Oveis Gharan (2017) investigate the welfare implications of various matching algorithms; Du and Livne (2016) and Doval (2018) consider the existence of self-enforcing matching arrangements.

Another strand of the literature investigates matching between long-lived players, where matching arrangements can be revised over time. Depending on the application, Corbae, Temzelides, and Wright (2003), Damiano and Lam (2005), Kurino (2009), Newton and Sawa (2015), Kotowski (2015), Kadam and Kotowski (2018a) and Kadam and Kotowski (2018b) propose and analyze different solution concepts in this environment. The most closely related paper to mine is Corbae, Temzelides, and Wright (2003), which allows full history dependence in directed matching, and considers applications in monetary theory.

My paper differs from the above papers in that I consider a different kind of matching environment: a fixed set of long-lived players on one side of the market repeatedly match with new generations of short-lived players from the other. This is the environment that arises naturally from college admission and various entry-level labor markets. Dynamic incentives also play a different role in such environments: they can be used as carrots and sticks for disciplining long-lived players.

My analysis builds on the techniques from the repeated games literature. In particular, the environment I consider bears many similarities to repeated games with perfect monitoring; I also adapt the techniques for equilibrium construction in Fudenberg and Maskin (1986) and Abreu, Dutta, and Smith (1994). The main difference of my paper is that I consider a two-sided cooperative environment, where the basic unit of analysis is not actions for individual players, but moves by coalitions. This introduces novel features that are not present in standard repeated games: regardless of patience level, the matching environment can impose severe restrictions on the payoffs of long-lived players. This is in sharp contrast to the "anything goes" result from folk theorems.

The top coalition sequence, featured prominently in the analysis in Section 1.3, is closely related to, but different from the top coalition property. The top coalition property is first introduced in Banerjee, Konishi, and Sönmez (2001) in the context of coalition formation, and subsequently used in a variety of cooperative game settings. See, for example, Niederle and Yariv (2009) and Doval (2018) for the use in matching settings. The top coalition property is an assumption which requires that every subset of players to have a top coalition; top coalition sequence, by contrast, is not an assumption about the underlying game, but is instead the collection of top coalitions that can be found through iterative eliminations.

The current paper is not the first attempt to mesh coalitional reasoning with dynamic consideration. Bernheim and Slavov (2009) considers the dynamic extension of Condorcet Winner in majoritarian voting settings; Ali and Liu (2018) consider coalition moves in a general repeated cooperative environments. Both these papers allow full history dependence—the most important difference of my approach from these papers is in the form of effective coalitions: in both these papers, the effective coalitions consist of subsets of long-lived players; in the matching environment I consider, however, the effective coalitions are those that consist of a single long-lived player and multiple generations of short-lived players. There is also an extensive literature on farsighted coalition formation: Chwe (1994), Ray and Vohra (1997), Ray and Vohra (1999), Konishi and Ray (2003), Gomes and Jehiel (2005), and Ray (2007), among others, consider the (implicit or explicit) coalition formation process among forward-looking players. My approach is different from these papers in that I allow full history dependence in a specialized matching environment, where there is no persistent state variable.

Finally, the dynamic interactions analyzed in this paper speak to two strands literature on static matching markets. The first strand is the literature on matching with complementarities or externalities. See, for example, Sasaki and Toda (1996), Dutta and Massó (1997), Echenique and Yenmez (2005), Kominers (2010), Pycia (2012), Flanagan (2015), and Pycia and Yenmez (2017). In particular, Che, Kim, and Kojima (2017) and Azevedo and Hatfield (2018) consider existence of stable matching in large markets, where institution preferences may exhibit complementarities. My results complement the existing literature by pointing out a novel channel for maintaining stability, when market size is not necessarily large. The second strand is the literature on matching with constraints. See, for example, Kamada and Kojima (2015), Fragiadakis and Troyan (2017), Goto et al. (2017), Kamada and Kojima (2017), and Kamada and Kojima (2018). This literature focuses on understanding and improving the welfare properties of a variety of real-life matchings markets, which are subject to restrictions other than those covered by standard capacity constraints. My paper complements this literature by providing a framework for understanding the enforceability of these restrictions, when the matching authority can exploit institutions' dynamic incentives.

### 1.1.3 Structure of the Paper

Section 1.2 introduce the model and discuss the solution concept, making minimal assumptions on players' preferences and allowing for flexible wages between the institutions and their matched agents. In Section 1.3, I focus on the canonical model for the matching market of medical residents: I assume that wages are held at an exogenously fixed level, and impose standard restrictions on preferences. Section 1.4 analyzes the flexible-wage model with general preferences. Finally, Section 1.5 concludes. All omitted proofs can be found in the Appendix.

## 1.2 Model and Solution Concept

### 1.2.1 Repeated Matching Market

Time is discrete. In every period $t = 0, 1, 2 \ldots$, a new generation of **agents** $\mathcal{A} \equiv \{a_1, \ldots, a_J\}$ enter a matching market to match with the **institutions** $I \equiv \{I_1, \ldots, I_K\}$. Institutions are long-lived players that persist through time. Agents are short-lived, and remain in the market for only one period.

Each institution $I_k$ has a per-period hiring quota $q_k > 0$, and Bernoulli utility function $u_k : 2^{\mathcal{A}} \to \mathbb{R}$ over the sets of agents it matches with in every period, where $2^{\mathcal{A}}$ is the set of all subsets of $\mathcal{A}$. The utility of staying unmatched, $u_k(\emptyset)$, is normalized to 0. Institutions' preferences are *not* assumed to satisfy the gross substitutes condition (see, for example, Kelso and Crawford 1982). Institutions share a common discount factor $\delta$, and evaluate a sequence of flow utilities $(u^1, u^2, \ldots)$ through discounting:

$$(1-\delta)\sum_{t=0}^{\infty} \delta^t u^t$$

Agents are allowed to care about not only the institution they are matched with, but also their colleagues. Each agent $a_j$ has utility function $v_j : \left(I \times 2^{\mathcal{A}}_{a_j}\right) \cup \{(\emptyset, \emptyset)\} \to \mathbb{R}$ over institutions and colleagues, where $2^{\mathcal{A}}_{a_j}$ is the set of subsets of $\mathcal{A}$ that contain $a_j$, while $(\emptyset, \emptyset)$ represents staying unmatched. For every agent $a_j$, $v_j(\emptyset,\emptyset)$ is normalized to 0.

Finally, I assume players have quasilinear utilities over money.

### 1.2.2 Stage-Game Matchings:

In this section, I introduce the static matching game that is played repeatedly in every period. As a point of comparison for the dynamic solution concept, I then review what it means for a matching to be stable in one-shot interactions.

In each period, a single institution can recruit multiple agents who are currently in the market—institutions and agents are matched according to a static one-to-many matching. A stage-game matching is described by a static assignment of players along with wage vectors from institutions.

**Definition 1.2.1.** A **static assignment** $\phi$ is a mapping defined on the set $I \cup \mathcal{A}$ which satisfies for all $I_k \in I$ and $a_j \in \mathcal{A}$:

1. agents are either unmatched, or matched to institution and peers: $\phi(a_j) \in (I \times 2^{\mathcal{A}}_{a_j}) \cup \{(\emptyset,\emptyset)\}$;
2. institutions are either unmatched, or matched to agents: $\phi(I_k) \in 2^{\mathcal{A}}$ and $|\phi(I_k)| \leq q_k$; and
3. consistency requirements for the mapping $\phi$: if $a_j \in \phi(I_k)$ then $\phi(a_j) = (I_k, \phi(I_k))$; if $\phi(a_j) = (I_k, A)$ then $\phi(I_k) = A$.

Let $\Phi$ denote the set of all static assignments.

A **wage vector** $\zeta_k \in \mathbb{R}^J$ from institution $I_k$ describes the wages paid by institution to each agent. Any non-zero wage payments can be made from an institution only to its employees.

**Definition 1.2.2.** A **static matching** is a pair $m = (\phi, \{\zeta_k\}_{k=1}^K)$, where $\phi$ is a static assignment, and each $\zeta_k = [\zeta_{kj}]_{j=1}^J$ is a vector of wages from institution $I_k$ to agents. Together $\phi$ and $\{\zeta_k\}_{k=1}^K$ satisfy $\zeta_{kj} \neq 0$ only if $a_j \in \phi(I_k)$.

I use $M$ to denote the set of all static matchings.

Players' preferences over static matchings are induced by their preferences over their assigned partners and the payments. For $m = (\phi, \{\zeta_k\}_{k=1}^K)$, I write

$$u_k(m) = u_k(\phi(I_k)) - \sum_{a_j \in \phi(I_k)} \zeta_{kj},$$

and

$$v_j(m) = v_j(\phi(a_j)) + \zeta_{kj}$$

for all $a_j \in \phi(I_k)$.

In canonical models of static matching markets, stability is defined in a symmetric way for institutions and agents. For example, as in Gale and Shapley (1962) and Kelso and Crawford (1982), the set of possible deviations consist of those that can be carried out by an institution, an agent, or a coalition involving players from both sides of the market.

In my model, institutions are long-lived while agents are short-lived. Due to this inherent asymmetry between two sides of the market, it is convenient to define stability in a way that emphasizes the role played by institutions, which is still consistent with the canonical stability notions.

**Definition 1.2.3.** A **feasible deviation** by institution $I_k$ from static matching $m = (\phi, \{\zeta_l\}_{l=1}^K)$ is a vector $(A'_k, \zeta'_k)$ where $A'_k$ is a set of agents that satisfies $\left|A'_k\right| \leq q_k$, and $\zeta'_k \in \mathbb{R}^J$ a vector of wages from institution $I_k$. Together $A'_k$ and $\zeta'_k$ satisfy:

1. institutions can make non-zero wage payments only to their employees: $\zeta'_{kj} = 0$ for all $a_j \notin A'_k$; and

2. a deviation is feasible only if all involved agents find it profitable: $v_j(I_k, A'_k) + \zeta'_{kj} > v_j(m)$ for all $a_j \in A'_k$.

I use $D_k(m)$ to denote the set of all feasible deviations from $m$ by institution $I_k$.

A feasible deviation thus encodes both an institution's ability to fire any subset of its employees, as well as its ability to organize a blocking coalition with agents from the other side of the market: such coalitions are available to $I_k$ as long as every member of $A'_k$ finds it profitable. A feasible deviation $(A'_k, \zeta'_k)$ is profitable for $I_k$ if

$$u_k(A'_k) - \sum_{a_j \in A'_k} \zeta'_{kj} > u_k(m).$$

A player is said to be autarkic in the stage game if it is not matched to any other player. The following definition of static stability extends that of Kelso and Crawford (1982) to the current environment.

**Definition 1.2.4.** A static matching $m = (\phi, \{\zeta_k\}_{k=1}^{K})$ is stable if

1. no agent prefers autarky: $v_j(m) \geq 0$ for every $a_j \in \mathcal{A}$; and
2. no institution has any profitable feasible deviation: if $I_k$ has a feasible deviation $(A'_k, \zeta'_k)$ from $m$, then $(A'_k, \zeta'_k)$ is not profitable for $I_k$.

Notice that this definition allows for peer effects. I use $M^*$ to denote the set of all static stable matchings.

### 1.2.3 Matching Processes

I now turn to repeated matching markets, and introduce the dynamic counterparts for the concepts reviewed in the previous section. In particular, a matching process is the dynamic extension of a static matching; self-enforcing matching process replaces stable matching outcome as the stability notion in this dynamic environment.

In a repeated matching market, a matching process specifies a *proposed or anticipated* stage-game matching today based on the history of the market.

At any time $t = 0, 1, \ldots$, a **period $t$ ex ante history** records past realizations of the public randomization device and the ensuing matching outcomes up to the beginning of period $t$. Formally, let $\overline{\mathcal{H}}_0 \equiv \emptyset$; for $t > 0$, the set of period $t$ ex ante histories is $\overline{\mathcal{H}}_t \equiv (\Omega \times M)^t$, where $\Omega$ is the state space of the arbitrary public randomization device. I use $\overline{\mathcal{H}} \equiv \cup_{t=0}^{\infty} \overline{\mathcal{H}}_t$ to denote the set of all possible ex ante histories. $\mathcal{H}_t \equiv \overline{\mathcal{H}}_t \times \Omega$ is the set of **period $t$ ex post histories**, and $\mathcal{H} \equiv \overline{\mathcal{H}} \times \Omega$ the set of all ex post histories. Let $\overline{\mathcal{H}}_\infty = (\Omega \times M)^\infty$ be the set of **outcome paths** of the game. For each $h \in \mathcal{H} \cup \overline{\mathcal{H}} \cup \overline{\mathcal{H}}_\infty$, let $\omega_t(h)$ and $m_t(h)$ denote the period $t$ realization of public randomization device and stage-game matching in $h$, respectively. For each $\widehat{\omega} \in \Omega$, I use $\mathcal{H}(\widehat{\omega}) \equiv \{h \in \mathcal{H} : \omega_0(h) = \widehat{\omega}\}$ to denote the set of ex post histories with $\widehat{\omega}$ as the initial draw from the public randomization device.

**Definition 1.2.5.** A **matching process** $\mu$ is a mapping $\mu : \overline{\mathcal{H}} \to \Delta(M)$.

Equivalently, making explicit the realization from the public randomization device, a matching process is also a mapping $\mu : \mathcal{H} \to M$.

In what follows, I will often write $\mu = (\psi, \{\xi_k\}_{k=1}^K)$. $\psi : \mathcal{H} \to \Phi$ is an **assignment process**, which maps an ex post history $h$ to the static assignment in the stage-game matching $\mu(h)$; for each institution $I_k \in I$, $\xi_k$ is a **wage plan** $\xi_k : \mathcal{H} \to \mathbb{R}^J$ such that $\xi_k(h)$ is the wage vector from institution $I_k$ in $\mu(h)$.

For an assignment process $\psi$, ex post history $h$, and institution $I_k$, I use $\psi(I_k|h)$ to denote the (possibly empty) set of agents matched to $I_k$ in the static assignment $\psi(h)$, and $\psi(a_j|h)$ for the institution and peers matched to agent $a_j$.

A matching process can capture various ways in which players coordinate their expectations. In markets like the NRMP, one can interpret a matching process as a matching protocol that proposes a (lottery over) stage-game matchings in every period based on the market's history. In labor markets without an explicit coordination mechanism, a matching process represents a shared understanding among players on how hiring outcomes in the past impact the market's future employment decisions.

A matching process $\mu$, together with the public randomization device, induces a probability measure over $\overline{\mathcal{H}}_\infty$. Institutions' preferences over matching processes are induced from the distribution over the sequences of stage-game matchings. That is, institutions evaluate matching processes by

$$u_k(\mu) \equiv \mathbb{E}^{\mu}_{\overline{\mathcal{H}}_\infty}\Big[(1-\delta)\sum_{t=0}^{\infty}\delta^t u_k(m_t(\overline{h}_\infty))\Big].$$

For every $t \geq 0$, a matching process $\mu$ also induce a probability measure over the period $t$ ex post histories $\mathcal{H}_t$. Let $u_k^t(\mu) \equiv \mathbb{E}^{\mu}_{\mathcal{H}_t}\big[u_k\big(\mu(h)\big)\big]$ denote institution $I_k$'s expected payoff in period $t$ from the matching process $\mu$. We have

$$u_k(\mu) = \sum_{t=0}^{\infty}\delta^t u_k^t(\mu).$$

As long-lived players, in a repeated matching market, institutions can conceive a feasible deviation plan from a matching process. A feasible deviation plan for an institution is a complete contingent plan, which specifies a static feasible deviation for the same institution after every possible history of the market.

**Definition 1.2.6.** A **feasible deviation plan** from the matching process $\mu = (\psi, \{\xi_l\}_{l=1}^K)$ by institution $I_k$ is a vector $(d_k', \xi_k')$. $d_k' : \mathcal{H} \to 2^{\mathcal{A}}$ is a mapping from ex post histories to sets of agents, and $\xi_k' : \mathcal{H} \to \mathbb{R}^J$ is a wage plan. Together $d_k'$ and $\xi_k'$ satisfy that $(d_k'(h), \xi_k'(h))$ is a feasible deviation from $\mu(h)$ for $I_k$, at every ex post history $h$.

A feasible deviation plan $(d_k', \xi_k')$ from matching process $\mu = (\psi, \{\xi_l\}_{l=1}^K)$ is a **one-shot deviation** if there is a unique ex post history $\widetilde{h}$ in the domain of $(d_k', \xi_k')$, such that

$$d_k'(h) = \psi(I_k|h) \text{ and } \xi_k'(h) = \xi_k(h) \quad \text{for all } h \in \mathcal{H}\backslash\{\widetilde{h}\}.$$

I will refer to a one-shot deviation $(d_k', \xi_k')$ by $(d_k'(\widetilde{h}), \xi_k'(\widetilde{h}))$.

Similar to Definition 1.2.3, the *feasibility* of a deviation plan encodes the constraint that participation in a coalitional deviation must be voluntary: for a coalitional deviation to be available to institution $I_k$ at history $h$, every participating agent must find the deviation profitable.

A feasible deviation plan can be thought of as an institution's plan for how to manipulate a matching process, in every possible contingency of the market. If the plan is carried out, it generates, together with the matching process, a path of realized stage-game matchings. Institutions evaluate a feasible deviation plan through the discounted sum of utilities collected from this path of manipulated stage-game matchings.

A conceptual question needs to be resolved in order to determine an institution's payoff from a feasible deviation plan: if institution $I_k$ were to commit a stage-game deviation $(A'_k, \zeta'_k)$ when players anticipate the stage-game outcome to be matching $m$, what is the resulting stage-game matching? Since the matching process adjusts future expectations based on past realized matchings, without an answer to the question above, it will be impossible to determine how a matching process responds to a deviation as history unfolds.

To this end, I assume that in a stage-game, if institution $I_k$ were to commit a deviation, other players do not further deviate, and remain matched according to $m$ whenever possible. Formally, given a static matching $m = (\phi, \{\zeta_l\}_{l=1}^K)$ and a feasible deviation $(A'_k, \zeta'_k)$ from $m$ by institution $I_k$, let $[m, (I_k, A'_k, \zeta'_k)]$ denote the **manipulated matching**. I make the following assumption.

**Assumption 1.** *Suppose* $m = (\phi, \{\zeta_l\}_{l=1}^K)$. *The static matching* $[m, (I_k, A'_k, \zeta'_k)] = (\overline{\phi}, \{\overline{\zeta}_l\}_{l=1}^K)$ *satisfies*

1. $\overline{\phi}(I_k) = A'_k$; $\overline{\phi}(I_l) = \phi(I_l) \backslash A'_k$ *for all* $l \neq k$; *and* $\overline{\phi}(a) = (\emptyset, \emptyset)$ *for all* $a \in \phi(I_k) \backslash A'_k$.
2. $\overline{\zeta}_k = \zeta'_k$; $\overline{\zeta}_{lj} = \zeta_{lj}$ *for* $l \neq k$, $a_j \in \overline{\phi}(I_l)$; *and* $\overline{\zeta}_{lj} = 0$ *for* $l \neq k$, $a_j \notin \overline{\phi}(I_l)$.

Most static matching models are silent about what happens in the wake of a deviation because in one-shot interactions, the profitability of a deviation does not depend on how other players react.[1]

The only role Assumption 1 plays in my model is in establishing identification: any deviation involving a single institution can be unequivocally attributed to the said institution. One can adopt alternative assumptions on how other players may further deviate in the stage-game upon seeing an initial deviation. As long as players understand which institution was involved in the *initial* deviation, all results in this paper remain unaffected.

I now define the manipulated matching process resulting from a feasible deviation plan, as well as the profitability of a feasible deviation plan.

**Definition 1.2.7.** Given a matching process $\mu$ and a feasible deviation plan $(d'_k, \xi'_k)$ by institution $I_k$, the **manipulated matching process** $[\mu, (I_k, d'_k, \xi'_k)] : \mathcal{H} \to M$ is a matching process defined by

$$[\mu, (I_k, d'_k, \xi'_k)](h) = [\mu(h), (I_k, d'_k(h), \xi'_k(h))] \quad \forall h \in \mathcal{H}$$

A feasible deviation plan $(d'_k, \xi'_k)$ is **profitable** for institution $I_k$ if $u_k([\mu, (I_k, d'_k, \xi'_k)]) > u_k(\mu)$.

To state what it means for a matching process to be self-enforcing, it is convenient to define the continuation of a matching process and a deviation plan, at both *ex ante* and *ex post* histories.

**Definition 1.2.8.** Fix a matching process $\mu$. At ex ante history $\widehat{h} \in \overline{\mathcal{H}}$, the **continuation matching process** $\mu|_{\widehat{h}} : \mathcal{H} \to M$ is defined by

$$\mu|_{\widehat{h}}(h) = \mu(\widehat{h}, h) \quad \text{for all } h \in \mathcal{H};$$

At ex post history $\widetilde{h} = \widehat{h} \times \widehat{\omega} \in \mathcal{H}$, the continuation matching process $\mu|_{\widetilde{h}} : \mathcal{H}(\widehat{\omega}) \to M$ is defined

[1] There are a few exceptions in the literature on matching with externailities, where the utility from a deviation depends on how other players are matched. In these static models, the reactions of other player are often built implicitly into the players' choice functions. See, for example, Sasaki and Toda (1996), Bando (2012), Pycia and Yenmez (2017).

by

$$\mu|_{\widetilde{h}}(h) = \mu(\widehat{h}, h) \quad \text{for all } h \in \mathcal{H}(\widehat{\omega});$$

The continuation of a feasible deviation plan is defined analogously.

**Definition 1.2.9.** Fix a feasible deviation plan $(d'_k, \xi'_k)$. At ex ante history $\widehat{h} \in \overline{\mathcal{H}}$, the **continuation deviation plan** $(d'_k|_{\overline{h}}, \xi'_k|_{\overline{h}}) : \mathcal{H} \to 2^{\mathcal{A}} \times \mathbb{R}^J$ is defined by

$$d'_k|_{\overline{h}}(h) = d'_k(\overline{h}, h) \;\text{ and }\; \xi'_k|_{\overline{h}}(h) = \xi'_k(\overline{h}, h) \quad \text{for all } h \in \mathcal{H}.$$

At ex post history $\widetilde{h} = \widehat{h} \times \widehat{\omega} \in \mathcal{H}$, the continuation deviation plan $(d'_k|_{\widetilde{h}}, \xi'_k|_{\widetilde{h}}) : \mathcal{H}(\widehat{\omega}) \to 2^{\mathcal{A}} \times \mathbb{R}^J$ is defined by

$$d'_k|_{\widetilde{h}}(h) = d'_k(\widehat{h}, h) \;\text{ and }\; \xi'_k|_{\widetilde{h}}(h) = \xi'_k(\widehat{h}, h) \quad \text{for all } h \in \mathcal{H}(\widehat{\omega}).$$

At an ex post history, the continuations of a matching process and a deviation plan have to follow the most recent realization from the public randomization device, which imposes the restriction on their domains. To reduce the notational burden, however, I will suppress this dependence on the realization of $\omega$ whenever it causes no confusion.

As the dynamic counterpart to static stable matching, a self-enforcing matching process represents players' self-fulfilling expectations over how the repeated matching market evolves. Such expectations must be robust against both individuals and coalitional deviations.

**Definition 1.2.10.** A matching process $\mu : \mathcal{H} \to M$ is **self-enforcing** if at every ex post history $h$,

1. no agent prefers autarky: $v_j(\mu(h)) \geq 0$ for every $a_j \in \mathcal{A}$; and
2. no institution has any profitable feasible deviation plan from the continuation matching process $\mu|_h$.

The lack of profitable deviation is imposed on the continuation matching processes after every possible ex post history: this essentially imposes the same sequential rationality requirement that subgame perfection does in a non-cooperative environment.

## 1.3 Markets with Fixed Wages

In this section, I first specialize the general environment introduced in Section 1.2 to markets where wages are fixed. This case is considered in Gale and Shapley (1962) and Roth and Sotomayor (1992), and is the canonical environment to study the NRMP. I characterize the set of self-enforcing assignment processes in this environment, and discuss how these results provide insights and caveats on using dynamic enforcement to redistribute medical residents.

### 1.3.1 Model

In the fixed-wage model, players' utilities are determined by the identity of their match: there are no flexible transfers that adjust matching payoffs. Institutions's preferences over agents satisfy the gross substitutes condition. In addition, agents are indifferent about which other agents are matched to the same institution. Finally, all players' preferences in the stage-game are strict.

**Market:** Formally, in a fixed-wage environment, a stage-game matching $m = (\phi, \{\zeta_k\}_{k=1}^{K})$ must satisfy $\zeta_k = \mathbf{0}$ for all $1 \leq k \leq K$, where $\mathbf{0} \in \mathbb{R}^J$ is the zero wage vector. This is without loss of generality, as the fixed wages's impacts can be incorporated into the functions $u_k(.)$ and $v_j(.)$. I will suppress the redundant wage vectors—instead of stage-game matchings, I will focus on stage-game *assignments* in this section.

Instead of matching processes, it is without loss to focus on assignment processes $\psi : \overline{\mathcal{H}}^F \to \Delta(\Phi)$, where $\overline{\mathcal{H}}^F \equiv \cup_{t=0}^{\infty}(\Omega \times \Phi)^t$ is the set of ex ante histories under fixed-wage environment. The set of ex post histories, $\mathcal{H}^F$, and outcome paths, $\overline{\mathcal{H}}_{\infty}^F$, are also modified accordingly to reflect the restriction on wages.

In the matching market for medical residents, an assignment process represents a matching program that can suggest different future assignments based on past history. For example, the matching program can exclude certain hospitals from the matching program in response to a past violation, which is equivalent to suggesting an assignment that leaves these hospitals unmatched.

The wage offer in any feasible deviation $(A'_l, \zeta'_l)$ must also be held at $\zeta'_l = \mathbf{0}$. I will also suppress the wage offer here and use $D_k^{\mathcal{A}}(\phi)$ to denote the subsets of agents that, combined with a zero wage offer, constitute a feasible deviation from the assignment $\phi$:

$$D_k^{\mathcal{A}}(\phi) \equiv \{B \subseteq \mathcal{A} : (B, \mathbf{0}_k) \in D_k(\phi, \{\mathbf{0}_l\}_{l=1}^K)\}$$

where $\mathbf{0}_l \in \mathbb{R}^J$ is the zero wage vector for all $1 \leq l \leq K$. A feasible deviation plan $(d'_k, \zeta'_k)$ reduces to the mapping $d'_k : \mathcal{H}^F \to 2^{\mathcal{A}}$.

**Preferences:** For each institution $I_k$, the utility from matched agents, $u_k : 2^{\mathcal{A}} \to \mathbb{R}$ is strict: $u_k(.)$ satisfies $u_k(A) \neq u_k(A')$ for different sets of agents $A \neq A'$. The gross substitutes condition, when specialized to the fixed-wage environment, requires that for all $A \subseteq \mathcal{A}$ and all $a_j, a_{j'} \in \mathcal{A}$,

$$a_j \in \underset{B \subseteq A \cup \{a_j, a_{j'}\}, |B| \leq q_k}{\arg\max} u_k(B) \;\Rightarrow\; a_j \in \underset{B \subseteq A \cup \{a_j\}, |B| \leq q_k}{\arg\max} u_k(B)$$

For each agent $a_j$, the utility function $v_j : I \to \mathbb{R}$ is only a function of his matched institution, and satisfies $v_j(I) \neq v_j(I')$ for all $I \neq I' \in I$. Without concerns for peer effects, there is no need to keep track of agents' peers when describing an assignment: instead of writing $\phi(a_j) \in I \times 2^{\mathcal{A}}_{a_j}$, I use $\phi(a_j) \in I$ to denote agent $a_j$'s matched institution.

### 1.3.2 The Scope of Dynamic Enforcement

In this section, I explore the possibility of using assignment processes to redistribute medical residents to rural hospitals. Before stating the main results, it is well known that static stable assignment exists when institutions' preferences satisfy the gross substitutes condition and agents' preferences are absent of peer effects (see, for example, Roth and Sotomayor (1992), Theorem 6.5). The following result guarantees the existence of self-enforcing assignment process for any patience level, since the infinite repetition of a static stable matching is a self-enforcing.

**Lemma 1.** *When institutions' preferences satisfy the gross substitutes condition and agents do not care about peer effects, self-enforcing assignment processes exist for every discount factor* $0 \leq \delta < 1$.

I first define the "top coalition sequence" in the matching market. Fix an arbitrary subset of institutions and agents $I' \cup \mathcal{A}' \subseteq I \cup \mathcal{A}$. An institution $I_k \in I'$ and a set of agents $\widehat{A}_k \subseteq \mathcal{A}'$ form a **top coalition** of $I' \cup \mathcal{A}'$ if

$$u_k(\widehat{A}_k) \geq u_k(B) \text{ for all } B \subseteq \mathcal{A}' \quad \text{and} \quad v_j(I_k) \geq v_j(I) \text{ for all } I \in I'.$$

Players in a top coalition find each other to be the most desirable match, given that $I' \cup \mathcal{A}'$ are the potential pool of players to be matched with. A **top coalition sequence** generalizes this notion by iteratively finding and eliminating top coalitions in the remaining players, until no new top coalition can be found.

**Definition 1.3.1.** [2] The top coalition sequence is the ordered set $\mathcal{T} = \{(I_{k_g}, \widehat{A}_{k_g})\}_{g=1}^{G}$ produced by the following algorithm:

1. Set $g = 0$ and $\mathcal{T}_0 = \emptyset$;

2. For $g \geq 1$:

   - If $(I \cup \mathcal{A}) \backslash \mathcal{T}_{g-1}$ has a top coalition $(I_{k_g}, \widehat{A}_{k_g})$: set $\mathcal{T}_g = \mathcal{T}_{g-1} \cup \{(I_{k_g}, \widehat{A}_{k_g})\}$ and $g = g+1$. If there are multiple top coalitions, break ties in favor of the institution with the smallest index;

[2] Whenever it causes no confusion, I will use $\mathcal{T}$ and $\mathcal{T}_g$ to denote the $(I, A)$ pairs or the set of players that are contained therein without distinction.

- If $(I \cup \mathcal{A}) \backslash \mathcal{T}_{g-1}$ has no top coalition: set $\mathcal{T} = \mathcal{T}_{g-1}$ and stop;

For each top coalition $(I, A)$ in $\mathcal{T}$, I will refer to $I$ and $A$ as each other's **top coalition partners**.

I use $\Phi_{\mathcal{T}}$ to denote the set of static assignments that assign players in $\mathcal{T}$ to their top coalition partners, and are individually rational for all the agents:

$$\Phi_{\mathcal{T}} = \{\phi \in \Phi : \phi(I_k) = \widehat{A}_k \ \forall I_k \in \mathcal{T}, \text{ and } v_j(\phi) \geq 0 \ \forall a_j \in \mathcal{A}\}$$

All static stable assignments must be contained in $\Phi_{\mathcal{T}}$. In particular, in a static stable assignment, players in the top coalition sequence must be matched to their top coalition partners.

**Lemma 2.** *If $\phi$ is a static stable assignment, then $\phi \in \Phi_{\mathcal{T}}$.*

Let $\kappa(I \backslash \mathcal{T}) \equiv \{k : I_k \in I \backslash \mathcal{T}\}$ denote the indices for the institutions not in the top coalition sequence. For every $k \in \kappa(I \backslash \mathcal{T})$, the **fixed-wage minmax payoff**, $\underline{u}_k^{\mathcal{T}}$, captures the institution's participation constraint, and is defined as

$$\underline{u}_k^{\mathcal{T}} = \min_{\phi \in \Phi_{\mathcal{T}}} \max_{B \in D_k^{\mathcal{A}}(\phi)} u_k(B).$$

**Theorem 1.** *For every $0 \leq \delta < 1$ and in every self-enforcing assignment process $\psi$:*

1. *regardless of history, players in the top coalition sequence must be matched with their top coalition partners; all agents must obtain at least their autarkic payoffs:*

$$\psi(h) \in \Phi_{\mathcal{T}} \ \ \forall h \in \mathcal{H}^F;$$

2. *all institutions must obtain at least their fixed-wage minmax payoffs:*

$$u_k(\psi) \geq \underline{u}_k^{\mathcal{T}} \ \ \forall I_k \in I \backslash \mathcal{T}.$$

Together with Lemma 2, Theorem 1 implies that in the matching market for medical residents, a history-dependent matching program can never alter the assignments for $\mathcal{T}$ beyond what is static stable. In addition, the worst possible continuation value for each hospital cannot be lower than $\underline{u}_k^{\mathcal{T}}$.

The intuition behind Theorem 1 is that in a top coalition sequence $\mathcal{T} = \{(I_{k_g}, \widehat{A}_{k_g})\}_{g=1}^{G}$, $I_{k_1}$'s favorite agents $\widehat{A}_{k_1}$ is always a feasible deviation, so there is no credible way to vary $I_{k_1}$'s continuation value. As the result, $I_{k_1}$ must always behave in the myopically optimal way— being matched to $\widehat{A}_{k_1}$. By induction, the only credible way to vary $I_{k_g}$'s continuation value is to match agents in $\widehat{A}_{k_g}$ to institutions in $\{I_{k_i}\}_{i \leq g-1}$. This is impossible because $\{I_{k_i}\}_{i \leq g-1}$ must always be matched to their own top coalition partners. Accordingly, $I_{k_g}$ must always be matched to $\widehat{A}_{k_g}$. For the second half of the theorem, since players have to be matched according to some assignment in $\Phi_{\mathcal{T}}$, every $I_k \in I \backslash \mathcal{T}$ can ensure a long-run payoff of $\underline{u}_k^{\mathcal{T}}$ by blocking with a coalition of feasible agents in every period.

When all agents share a common ranking over institutions, every player is in the top coalition sequence. In particular, if medical residents' preferences over hospitals are highly homogenous, it is impossible to enforce any assignment than the (unique) static stable assignment— history dependent matching programs offer no room for improvement compared to static ones.

**Corollary 1.** *When agents share a common ranking over institutions, there is a unique self-enforcing assignment process, where players match according to the unique static stable assignment at every ex post history of the market.*

Theorem 2 complements Theorem 1, and shows that with patient hospitals, if a randomization over assignments in $\Phi_{\mathcal{T}}$ gives hospitals in $I \backslash \mathcal{T}$ payoffs higher than $\underline{u}_k^{\mathcal{T}}$, then it can be supported as the time-invariant outcome of a self-enforcing assignment process.

Define $\Lambda^* \equiv \{\lambda \in \Delta(\Phi_{\mathcal{T}}) : \mathbb{E}_\lambda[u_k(\phi)] > \underline{u}_k^{\mathcal{T}} \; \forall k \in \kappa(I \backslash \mathcal{T})\}$.

**Theorem 2.** *For every $\lambda \in \Lambda^*$, there is a $\underline{\delta}$ such that for every $\delta \in (\underline{\delta}, 1)$, there exists a self-enforcing assignment process that randomizes over stage-game assignments according to $\lambda$ in* *every period.*

The proof for Theorem 2 adapts the "Folk Theorem" equilibrium construction from Fudenberg and Maskin (1986) and Abreu, Dutta, and Smith (1994). To see the connection, note that the condition $\mathbb{E}_\lambda[u_k(\phi)] > \underline{u}_k^{\mathcal{T}}$ $\forall k \in \kappa(I \backslash \mathcal{T})$ in the definition of $\Lambda^*$ ensures that the equilibrium payoffs for all players in $I \backslash \mathcal{T}$ are above their respective minmax levels. I then construct a matching process by varying continuation play within $\Phi_{\mathcal{T}}$, and using the analogues for "player-specific punishments" and "minmax punishment" in the matching environment.

To wrap up, the findings in this section suggest that in repeated matching markets, history dependence can substantially expand the set of feasible allocations. However, the scope of dynamic enforcement is tempered by the presence of top coalition sequence. As a result, caution should be exercised when a dynamic matching program attempts to enforce outcomes beyond those that are static stable—this is especially relevant when there are reasons to suspect homogeneity in agents preferences.

## 1.4 Markets with Flexible Wages

In this section, I return to the model introduced in Section 1.2: wages can be flexibly adjusted; in addition, I allow institutions to have complementary preferences over agents, and agents to have non-trivial preferences over colleagues.

Let $\chi(I_k, A)$ denote the total surplus $I_k$ can extract from matching with agents $A$. That is,

$$\chi(I_k, A) \equiv u_k(A) + \sum_{a_j \in A} v_j(I_k, A).$$

In particular, this leaves every agent in $A$ with 0 payoff. Use $Q \equiv \sum_{I_l \in I} q_l$ to denote all institutions' combined hiring capacity. For $I_k$,

$$\underline{\pi}_k \equiv \min_{\{A \subseteq \mathcal{A}: |A| \leq Q\}} \max_{B \subseteq \mathcal{A} \setminus A} \chi(I_k, B)$$

is the most severe punishment that can be imposed on $I_k$, in a bidding war for employees, by $I \setminus \{I_k\}$ and $I_k$'s employees.

To understand the definition of $\underline{\pi}_k$, notice that in a bidding war for employees, $I \setminus \{I_k\}$ can conspire to bid away a total of $(Q - q_k)$ agents from $I_k$; meanwhile, $I_k$ itself is matched to $q_k$ agents, all of whom are employed at premium wages. If $I_k$ were to rejects such an arrangement, its best alternative is to abandon the current employees, and hire from outside of the combined $Q$ agents. The minmax operator in $\underline{\pi}_k$ captures $I_k$'s best response to the most effective punishment.

Define

$$M^\circ \equiv \{m \in M : v_j(m) \geq v_j(\emptyset, \emptyset)\ \forall a_j \in \mathcal{A}\}.$$

Matchings in $M^\circ$ ensure that no players in $\mathcal{A}$ have static incentive to deviate. Define

$$\Lambda^\circ \equiv \{\lambda \in \Delta(M^\circ) : \mathbb{E}_\lambda[u_k(m)] > \underline{\pi}_k\ \forall I_k,\ \text{supp}(\lambda) \text{ is bounded}\}.$$

$\Lambda^\circ$ is the set of lotteries over $M^\circ$ with bounded support, which gives each institution an expected payoff higher than its punishment payoff $\underline{\pi}_k$.

As Theorem 3 demonstrates, the sets $M^\circ$ and $\Lambda^\circ$ provide a tight characterization for the set of self-enforcing matching processes: every element in $\Lambda^\circ$ can be sustained as the stationary outcome in a self-enforcing matching process with patient institutions; on the other hand, matchings in $M^\circ$ are the only ones that can be played in any self-enforcing matching process for any patience level.

**Theorem 3.** *For every $\lambda \in \Lambda^\circ$, there is a $\underline{\delta}$ such that for every $\delta \in (\underline{\delta}, 1)$, there exists a self-enforcing matching process that randomizes over stage-game matchings according to $\lambda$ in every* *period.*

*For every* $0 \leq \delta < 1$ *and in every self-enforcing matching process* $\mu$*, institutions are no worse off than when targeted in a bidding war:*

$$u_k(\mu) \geq \underline{\pi}_k \ \ \forall I_k.$$

The proof for Theorem 3 adapts the techniques in Fudenberg and Maskin (1986) and Abreu, Dutta, and Smith (1994): note that the condition $\mathbb{E}_\lambda[u_k(m)] > \underline{\pi}_k \ \forall I_k$ in the definition of $\Lambda^\circ$ ensures that the equilibrium payoffs for all institutions are above their respective minmax levels.

**Theorem 4.** *The sets* $M^\circ$ *and* $\Lambda^\circ$ *are non-empty.*

Theorem 4 is proved by considering a Random Serial Dictatorship among institutions: let institutions randomize over priorities, then take turns to extract full surplus from their favorite agents according to the realized priority. The remainder of the proof is to show that in expectation, the Random Serial Dictatorship gives every $I_k$ a payoff higher than $\underline{\pi}_k$.

Together with Theorem 3, Theorem 4 delivers the existence result.

## 1.5 Conclusion

This paper considers stability in two-sided matching markets, where one side is long-lived and the other is short-lived. The analysis builds on elements from both the matching literature and the repeated games literature. In the fixed-wage environment, some players can be motivated through intertemporal tradeoffs, while others are immune to dynamic enforcement. These results have normative implications for the design of matching programs. In the flexible-wage environment, the results show that dynamic incentives constitute a new channel for sustaining stability even if static stable matchings fails to exist.

The current paper provides a framework for understanding history dependence in repeated matching markets, but much remains unanswered about these environments. For future work, more study is merited in order to better understand matching processes under fixed patience levels, or when players may have a restricted class of preferences that reflect the specific matching environment being analyzed. Another important avenue for future research is the communication of private information in matching processes. In static environments, the deferred acceptance algorithm is strategy-proof for one side of the market. In repeated matching markets, if a matching algorithm goes beyond proposing static stable outcomes, it remains an open question whether such algorithms can still be made strategy-proof, or have to rely on weaker incentive-compatibility conditions. I hope to address these questions in future research.

## 1.6 Acknowledgement

Chapter 1, in full, is currently being prepared for submission for publication of the material. Liu, Ce. “Stability in Repeated Matching Markets”. The dissertation author was the sole investigator and author of this material.

# Chapter 2

# Costly Information Acquisition

## 2.1 Introduction

Acquiring information is an integral part of decisions under uncertainty. Most existing research on costly information acquisition studies costs that are additively separable from the expected payoff. This assumes the cost incurred acquiring information is independent of expected payoff. One can interpret these preferences as an individual having a fixed production technology to acquire information. However, the cost structures of information acquisition can be more complicated.

For example, there may be significant costs incurred from the time delay waiting for information to arrive. Consider when an oil company is deciding between locations to drill for oil. To acquire information, in addition to the monetary expenses to finance geological surveys, there are also significant costs incurred from delayed realization of profits. Suppose the payoff from drilling at a site for each state of the world is given by a net-present-value from the time drilling *begins*. If the oil sites have a higher net-present-value, then this translates into higher costs incurred through discounting. Importantly, costs from time delay now interact with the expected payoff.

In this paper, we study a general model of costly information acquisition that allows for interactions between the information cost and the expected payoff from the decision problem. Apart from the standard assumptions of expected utility maximization and Bayesian updating, the only additional assumption of the model is that the decision maker prefers higher expected payoffs. As special cases of the model, we characterize a representation with multiplicative costs of information and a representation with a constrained information set.

### 2.1.1 Why does this matter?

The paper of Caplin and Dean (2015) investigates the particular case of the model studied here, in which costs are additive. Our model should be viewed as a direct generalization of this contribution. This being said, the work is motivated by classical economic environments. There are several standard economic frameworks in which we would not expect costs to be additive.

The multiplicative cost model is a particularly interesting case. There are several standard economic environments in which we would expect costs to arise multiplicatively, rather than additively. Let us state three particularly compelling environments:

1. The cost associated with the acquired information may be ascribed directly to a time delay. In standard models of discounting, delay enters payoffs multiplicatively. Thus, in a model where different information arises with different delays, we would expect the behavior under consideration to approximately match the the multiplicative model.

2. The cost associated with the acquired information may accrue because of some probability of an absolute breakdown. For example, eliciting the information may involve some type of illegal activity; if the acquirer is caught, she gets nothing. In this case, the probability of not being caught enters multiplicatively into a (risk-neutral) individual's utility.

3. A more straightforward example involves the individual directly contracting with an outside provider of information, who insists on a profit-sharing agreement with the individual,

where the share of profits asked for can depend on the information sold. In such a setup, costs obviously enter multiplicatively.

More broadly, the class of nonseparable costly information acquisition models nest behavior generated when there are potentially multiple components of the cost of information acquisition. For example, the decision maker may incur *both* an additively separable cost to access an information structure, as well as a discounting cost from time delay.[1] Caplin and Dean (2015) provide a revealed preference test for costly information acquisition when costs are additively separable. When there is only a cost from discounting, we show that behavior is characterized by a condition basically derived from the Homothetic Axiom of Revealed Preference (See Varian (1983a)).

### 2.1.2 Methods and related literature

We take a revealed preference approach that builds on the recent contribution of Caplin and Dean (2015). In particular, the model considers a decision maker facing actions with state-contingent payoffs. The decision maker chooses an information structure and makes stochastic choices conditioning on the signal received from the information structure. Using state dependent stochastic choices, there is a natural revealed information structure that facilitates the analysis. Our main result characterizes the general model of costly information acquisition with an axiom on expected payoffs that resembles the Generalized Axiom of Revealed Preference[2]. To emphasize the potential interaction between expected utility and cost, we refer to such model as a *nonseparable* costly information acquisition model.

Our results generalize the *No Improving Attention Cycles* condition of Caplin and Dean (2015) in the same way that the Generalized Axiom of Revealed Preference generalizes the cyclic monotonicity condition of Rockafellar (1966) or the condition of Koopmans and Beckmann

[1]The exact content of this particular example remains unknown.

[2]See Houthakker (1950), Richter (1966), Varian (1982), and Chambers and Echenique (2016).

(1957). [3] Importantly cyclic monotonicity is equivalent to rationalization via a quasi-linear utility function, which imposes cardinal restrictions on consumption data. Thus, Caplin and Dean (2015) reflects a type of cardinal model, while the model here is ordinal. Using the intuition from the consumer problem, we show that data consistent with nonseparable costly information preferences can be taken to satisfy quasiconcavity and weak Blackwell monotonicity without loss of generality.

The characterizations here exploit results and intuition from classical consumer theory. An experiment, or signal, is a probability distribution over posteriors (as in Blackwell (1953)). Mathematically, up to a normalization, probability measures and normalized *price* vectors can be viewed as the same object. In the consumer setting, expenditure is computed as the inner-product of price and quantity demanded. Similarly, the ex-ante payoff from the experiment can be computed as the inner-product of the information structure and posterior value function. Thus, the ex-ante payoff can be treated as *wealth*.

The similarity between standard consumer theory and costly information acquisition extends beyond the correspondence of primitives. The nonseparable model of costly information acquisition is defined as a preference that is increasing in ex-ante payoff and depends on the information structure. Using the relation above, one notes the similarity to the *indirect utility function* that is increasing in wealth and depends on prices. In standard consumer theory, the *direct utility* of a consumption bundle is obtained from indirect utility through minimization over price vectors. The setting of costly information differs since the direct utility of a decision problem is defined through maximization over information structures. While the optimization principle differs, the same underlying duality holds which leads to the characterization by a condition resembling the General Axiom of Revealed Preference.

While we have highlighted the similarity to standard consumer theory, there are some technical differences. Most importantly, the information structures and posterior value functions

[3] See also Brown and Calsamiglia (2007) and Chambers, Echenique, and Saito (2016) for variants of this condition in an explicit revealed preference framework.

are objects in infinite dimensional vector spaces. Thus, our proofs utilize the general results on quasi-concave duality that have been fruitfully studied by Chateauneuf and Faro (2009) and Cerreia-Vioglio et al. (2011a, 2011b). However, once one makes this connection the results follow by leveraging existing revealed preference and duality techniques. As such, the paper also serves as a didactic exercise.

This paper is related to other works on costly information acquisition and boundedly rational behavior. Costly information acquisition has received study from a revealed preference perspective in Caplin and Martin (2015) and Caplin and Dean (2015). Boundedly rational behavior has been studied with revealed preference conditions in Fudenberg, Iijima, and Strzalecki (2015), and Allen and Rehbeck (2016).

The paper proceeds as follows. Section 2.2 presents the notation and some useful facts. Section 2.3.1 introduces and characterizes the nonseparable costly information acquisition model. Section 2.3.3 presents a variant of the model whereby choice of information structure is costless, but is constrained to lie in some unknown set. Section 2.3.2 presents a model with a multiplicative cost of information acquisition. Section 2.4 relates the conditions to those in Caplin and Dean (2015), provides examples of behavior allowed by the various models, provides guidance on out of sample prediction, and discusses some limitations. Finally, Section 2.6 contains our concluding remarks. Proofs are relegated to the appendix.

## 2.2 Preliminaries

### 2.2.1 Notation

We study a decision maker facing actions with state-contingent payoffs.[4] Notation is consistent with Caplin and Dean (2015) whenever possible for ease of comparison. We study a variety of models that are increasing in ex-ante payoff and satisfy Bayes' law. A decision

[4]The ideas discussed here are broader if one considers general mappings over posteriors.

maker chooses actions whose outcome depends on a finite number of states of the world. Let $\Omega$ denote a finite set of states. Let $X$ denote a set of outcomes. Therefore, the set of all actions (state-contingent outcomes) is $X^{\Omega}$.

The set of all finite decision problems is given by $\mathcal{A} = \{A \subset X^{\Omega} \mid |A| < \infty\}$. As in Caplin and Dean (2015) we investigate the situation in which a researcher has a state dependent stochastic choice dataset from decision problems in $\mathcal{A}$. For $A \in \mathcal{A}$, $\Delta(A)$ refers to the set of probability distributions over actions in $A$.

**Definition 2.2.1.** A *state dependent stochastic choice dataset* is a finite collection of decision problems $\mathcal{D} \subset \mathcal{A}$ and a related set of state dependent stochastic choice functions $\mathcal{P} = \{P_A\}_{A \in \mathcal{D}}$ where $P_A : \Omega \to \Delta(A)$. Denote the probability of choosing an action $a$ conditional on state $\omega$ in decision problem $A$ as $P_A(a \mid \omega)$.

We assume that the prior beliefs of the decision maker $\mu \in \Gamma = \Delta(\Omega)$ are known. Moreover, we assume that the utility index $u : X \to \mathbb{R}$ is a known function.

The following example illustrates the notations and primitives of the model. Throughout the paper, we will build on this example in order to illustrate the different testable implications for various models of costly information acquisition.

**Example 1.** *The set of states is $\Omega = \{\omega_1, \omega_2\}$, and the prior is given by $\mu = \left(\frac{1}{2}, \frac{1}{2}\right)$. There are two menus $A = \{a, b\}$ and $A' = \{a', b'\}$. Let the utilities from actions in menu $A$ and $A'$ take the following values:*

$$u(a(\omega)) = \begin{cases} 0 & \text{if } \omega = \omega_1 \\ 2 & \text{if } \omega = \omega_2 \end{cases} \quad u(b(\omega)) = \begin{cases} 2 & \text{if } \omega = \omega_1 \\ 0 & \text{if } \omega = \omega_2 \end{cases}$$
$$u(a'(\omega)) = \begin{cases} 0 & \text{if } \omega = \omega_1 \\ 10 & \text{if } \omega = \omega_2 \end{cases} \quad u(b'(\omega)) = \begin{cases} 10 & \text{if } \omega = \omega_1 \\ 0 & \text{if } \omega = \omega_2 \end{cases}.$$

*The choice probabilities are given by*

$$P_A(a \mid \omega_1) = \frac{2}{10} \qquad P_A(a \mid \omega_2) = \frac{8}{10}$$
$$P_{A'}(a' \mid \omega_1) = \frac{3}{10} \qquad P_{A'}(a' \mid \omega_2) = \frac{7}{10}$$

*where the choice of $b$ and $b'$ are given by the complementary probabilities.*

We take an abstract approach to modeling the choice of an information structure. Each subjective signal is identified with its associated posterior beliefs $\gamma \in \Gamma$. Thus, an information structure is given by a finite support distribution over $\Gamma$ that satisfies Bayes' law.

**Definition 2.2.2.** The set of *information structures*, $\Pi$, comprises all Borel probability distributions over $\Gamma$, $\pi \in \Delta(\Gamma)$, that have finite support and satisfy Bayes' law. A distribution over posteriors satisfies Bayes' law if the distribution over posteriors is a mean-preserving spread of the prior $\mu$ denoted as

$$E_\pi[\gamma] = \sum_{\gamma \in \text{Supp}(\pi)} \gamma\pi(\gamma) = \mu$$

where $\pi(\gamma) = \Pr(\gamma \mid \pi) = \sum_{\omega\in\Omega} \mu(\omega)\pi(\gamma \mid \omega)$. [5]

We now provide definitions necessary to discuss the ex-ante payoff. The ex-ante payoff is the utility an individual expects to receive for a given information structure. Given a utility index, each decision problem $A \in \mathcal{A}$ induces a posterior value function, $f_A : \Gamma \to \mathbb{R}$, which maps posterior beliefs $\gamma$ to the maximal utility possible from $A$ under posterior $\gamma$. Formally, for any decision problem $A$ and posterior belief $\gamma$

$$f_A(\gamma) = \max_{a\in A} \sum_{\omega\in\Omega} \gamma(\omega)u(a(\omega)).$$

[5] A similar notion of Bayesian plausibility is commonly used in the Bayesian persuasion literature. See, for example, Kamenica and Gentzkow (2011).

**Definition 2.2.3.** We denote the *ex-ante payoff* induced by an information structure $\pi \in \Pi$ as

$$\pi \cdot f_A = \sum_{\gamma \in \text{Supp}(\pi)} \pi(\gamma) f_A(\gamma)$$

where $\pi(\gamma) = \Pr(\gamma \mid \pi) = \sum_{\omega \in \Omega} \mu(\omega)\pi(\gamma \mid \omega)$.

This inner-product representation of ex-ante payoff is intuitive since $f_A$ is a continuous function on $\Gamma$ and the set of continuous functions on $\Gamma$ is topologically dual to the set of countably additive Borel measures on $\Gamma$ (Aliprantis and Border (2006), Theorem 14.15).

### 2.2.2 Revealed Information Structures

While we present several models of costly information acquisition, the analysis relies on the recovery of a *revealed information structure* from the state dependent stochastic choice data. Using the procedure from Caplin and Dean (2015), we associate each chosen action to a subjective information state. The revealed information structure may not be identical to the true information structure. However, the revealed information structure is a garbling (as defined in Blackwell (1953)) of the true information structure. The relationship between the true information structures and revealed information structures allows us to order the information structures and deduce conditions on revealed information. Without further delay, we define *revealed posteriors* and *revealed information structures*.

**Definition 2.2.4.** Given $\mu \in \Gamma$ , $A \in \mathcal{D}$, $P_A \in \mathcal{P}$, and $a \in \text{Supp}(P_A)$, the *revealed posterior* $\bar{\gamma}_A^a \in \Gamma$ is defined as

$$\begin{aligned} \bar{\gamma}_A^a(\omega) &= \Pr(\omega \mid a \text{ is chosen from } A) \\ &= \frac{\mu(\omega) P_A(a \mid \omega)}{\sum_{\nu \in \Omega} \mu(\nu) P_A(a \mid \nu)}. \end{aligned}$$

**Definition 2.2.5.** Given $\mu \in \Gamma$ , $A \in \mathcal{D}$, and $P_A \in \mathcal{P}$, the revealed information structure $\bar{\pi}_A \in \Pi$ is defined by

$$\bar{\pi}_A(\gamma \mid \omega) = \sum_{\{a \in \text{Supp}(P_A) \mid \bar{\gamma}_A^a = \gamma\}} P_A(a \mid \omega)$$

and induces a revealed distribution on posteriors $\bar{\pi}_A$ such that

$$\bar{\pi}_A(\gamma) = \sum_{\omega \in \Omega} \mu(\omega) \bar{\pi}_A(\gamma \mid \omega).$$

The revealed information structure for decision problem $A$ is a finite probability measure over the revealed posteriors.

**Example 1** (Continued)**.** *The choices in Example 1 generate the following revealed posteriors*

$$\begin{aligned}
\bar{\gamma}_A^a &= \left(\frac{2}{10}, \frac{8}{10}\right) \quad ; \quad \bar{\gamma}_A^b = \left(\frac{8}{10}, \frac{2}{10}\right) \\
\bar{\gamma}_{A'}^{a'} &= \left(\frac{3}{10}, \frac{7}{10}\right) \quad ; \quad \bar{\gamma}_{A'}^{b'} = \left(\frac{7}{10}, \frac{3}{10}\right).
\end{aligned}$$

*Each revealed posterior has the same probability of occurring so that*

$$\bar{\pi}_A(\bar{\gamma}_A^a) = \bar{\pi}(\bar{\gamma}_A^b) = \bar{\pi}(\bar{\gamma}_{A'}^{a'}) = \bar{\pi}(\bar{\gamma}_{A'}^{b'}) = \frac{1}{2}.$$

*The optimal decision rules for these posteriors give*

$$\begin{aligned}
f_A(\bar{\gamma}_A^a) = f_A(\bar{\gamma}_A^b) = 1.6 \quad &; \quad f_A(\bar{\gamma}_{A'}^{a'}) = f_A(\bar{\gamma}_{A'}^{b'}) = 1.4 \\
f_{A'}(\bar{\gamma}_{A'}^{a'}) = f_{A'}(\bar{\gamma}_{A'}^{b'}) = 7 \quad &; \quad f_{A'}(\bar{\gamma}_A^a) = f_{A'}(\bar{\gamma}_A^b) = 8
\end{aligned}$$

As mentioned before, we use the notion of garbling to partially order information structures.

**Definition 2.2.6.** The information structure $\pi \in \Pi$ (with posteriors $\gamma^j$) is a garbling of $\rho \in \Pi$

(with posteriors $\eta^i$) if there exists a $|\text{Supp}(\rho)| \times |\text{Supp}(\pi)|$ matrix $\mathbf{B}$ with non-negative entries such that for all $i \in \{1, \ldots, |\text{Supp}(\rho)|\}$ we have $\sum_{\gamma^j \in \text{Supp}(\pi)} b^{i,j} = 1$ and for all $\gamma^j \in \text{Supp}(\pi)$ and $\omega \in \Omega$ that

$$\pi(\gamma^j \mid \omega) = \sum_{\eta^i \in \text{Supp}(\rho)} b^{i,j} \rho(\eta^i \mid \omega).$$

In other words, $\pi$ a garbling of $\rho$ if there is a stochastic matrix $\mathbf{B}$ that can be applied to $\rho$ that yields $\pi$. We present two important properties about garblings that we use extensively in the analysis.

Lastly, an information structure $\pi$ is consistent with stochastic choice function $P_A$, if $P_A$ can be generated by the decision maker making optimal choice under information structure $\pi$.

**Definition 2.2.7.** For $\pi \in \Pi$ and $P_A \in \mathcal{P}$, we say $\pi$ is *consistent* with $P_A$ if there exists a choice function $C_A : \text{Supp}(\pi) \to \Delta(A)$ such that for all $\gamma \in \text{Supp}(\pi)$,

$$C_A(a \mid \gamma) > 0 \quad \Rightarrow \quad \sum_{\omega \in \Omega} \gamma(\omega) u(a(\omega)) \geq \sum_{\omega \in \Omega} \gamma(\omega) u(b(\omega)) \quad \text{for all } b \in A$$

and for all $\omega \in \Omega$ and $a \in A$

$$P_A(a \mid \omega) = \sum_{\gamma \in \text{Supp}(\pi)} \pi(\gamma \mid \omega) C_A(a \mid \gamma).$$

The next three intermediate results regarding revealed information structures will be used extensively in the analysis.

**Lemma 3.** *If $\pi$ is consistent with $P_A$, then $\bar{\pi}_A$ is a garbling of $\pi$.*

Lemma 3 is proved in Caplin and Dean (2015). The lemma says that if an information structure is consistent with the state dependent stochastic choice dataset, then the revealed information structure is a garbling.

**Lemma 4.** *Given a decision problem $A \in \mathcal{A}$ and $\pi, \rho \in \Pi$ with $\pi$ a garbling of $\rho$, then*

$$\rho \cdot f_A \geq \pi \cdot f_A.$$

Lemma 4 follows straightforwardly from Blackwell's theorem (Blackwell 1953). Blackwell's theorem establishes the notion that some information structures are "more valuable" than others. In particular, if $\pi$ is a garbling of $\rho$, then $\rho$ yields weakly higher ex-ante payoff in any decision problem.

**Lemma 5.** *For all decision problems $A, B \in \mathcal{D}$ if $\pi_A$ is an information structure consistent with choice data $P_A$, then $f_B \cdot \pi_A \geq f_B \cdot \bar{\pi}_A$ and $f_A \cdot \pi_A = f_A \cdot \bar{\pi}_A$*

Lemma 5 follows since the two information structures $\pi_A$ and $\bar{\pi}_A$ induce the same state dependent choices, so their ex ante payoffs should be identical.

## 2.3 Characterizing Costly Information Models

In this section, we introduce three models of costly information acquisition. The nonseparable information cost model is the most general: it assumes only that the decision maker prefers higher expected payoffs from choices, and that more informative signals are more costly. Both the multiplicative cost model and the constrained information model are special cases of the nonseparable model.

### 2.3.1 Nonseparable Information Cost

We place minimal restrictions on a decision maker's preferences on information structures. The only condition we impose is that preferences are monotone increasing in ex-ante payoff.

**Definition 2.3.1.** Given $\mu \in \Gamma$ and $u : X \to \mathbb{R}$, a state dependent stochastic choice dataset $(\mathcal{D}, \mathcal{P})$ has a *nonseparable costly information representation* if there exists a function $V : \mathbb{R} \times \Pi \to \mathbb{R} \cup \{-\infty\}$, information structures $\{\pi_A\}_{A \in \mathcal{D}}$, and choice functions $\{C_A\}_{A \in \mathcal{D}}$ such that:

1. Monotonicity: For all $\pi \in \Pi$ and for all $t, s \in \mathbb{R}$, if $t < s$ and $V(t, \pi) > -\infty$, then $V(t, \pi) < V(s, \pi)$.

2. Non-triviality: For all $t \in \mathbb{R}$, there exists $\pi_t \in \Pi$ such that $V(t, \pi_t) > -\infty$.

3. Information is optimal: For all $A \in \mathcal{D}$, $\pi_A \in \arg\max_{\pi \in \Pi} V(\pi \cdot f_A, \pi)$ .

4. Choices are optimal: For all $A \in \mathcal{D}$, the choice function $C_A : \text{Supp}(\pi_A) \to \Delta(A)$ is such that given $a \in A$ and $\gamma \in \text{Supp}(\pi_A)$ with $C_A(a \mid \gamma) = \Pr_A(a \mid \gamma) > 0$, then

$$\sum_{\omega \in \Omega} \gamma(\omega) u(a(\omega)) \geq \sum_{\omega \in \Omega} \gamma(\omega) u(b(\omega)) \quad \text{for all} \quad b \in A.$$

5. The data are matched: For all $A \in \mathcal{D}$, given $\omega \in \Omega$ and $a \in A$,

$$P_A(a \mid \omega) = \sum_{\gamma \in \text{Supp}(\pi_A)} \pi_A(\gamma \mid \omega) C_A(a \mid \gamma).$$

The above definition is a large class of preferences. However, it allows for the presence of unknown discounting and additively separable information costs. We give some examples of functions nested in this class below.

**Example 2.** *We give a special case of $V$ that allows for both unknown discounting from acquiring information and unknown additively separable costs. Consider when the function $V$ takes the form*

$$V(\pi \cdot f_A, \pi) = \delta(\pi)(\pi \cdot f_A) - K(\pi)$$

*where $\delta(\pi) \in [0, 1]$ gives the fraction of expected utility lost from discounting and $K(\pi)$ specifies* *the cost of accessing the information. We note the similarity to the polar form from Gorman (1953) which has been characterized using revealed preference by Cherchye et al. (2016).*

**Example 3.** *We consider the special case of a non-separable costly information acquisition given by*

$$V(\pi \cdot f_A, \pi) = \Phi(\pi \cdot f_A) - K(\pi)$$

*where $\Phi : \mathbb{R} \to \mathbb{R}$ is an increasing transformation of the expected utility and $K(\pi)$ is the cost of accessing information. This example takes the utils from expected utility and transforms them to the same units as the cost function. While this is cosmetically similar to the model by Caplin and Dean (2015), the characterization there does not apply.*

**Example 4.** *A transformation of utils from expected utility may also be pertinent in the presence of discounting. This is represented as*

$$V(\pi \cdot f_A, \pi) = \delta(\pi)\Phi(\pi \cdot f_A) - K(\pi)$$

*where $\Phi : \mathbb{R} \to \mathbb{R}$ is an increasing transformation of the expected utility, $\delta(\pi) \in [0,1]$ gives the fraction of utils lost from acquiring information, and $K(\pi)$ is the cost of accessing information.*

We now define the properties that completely characterize the model. The first condition is similar to the generalized axiom or revealed preference.

**Condition 1** (Generalized Axiom of Costly Information (GACI))**.** *We say the dataset $(\mathcal{D}, \mathcal{P})$ satisfies GACI if for all sequences $(\bar{\pi}_{A_1}, f_{A_1}), \ldots, (\bar{\pi}_{A_k}, f_{A_k})$ with $A_i \in \mathcal{D}$ for which $\bar{\pi}_{A_i} \cdot f_{A_i} \leq \bar{\pi}_{A_i} \cdot f_{A_{i+1}}$ for all i (with addition modulo k), then equality holds throughout.*

Comparing this condition to GARP, we see that the $\bar{\pi}$ play a role similar to prices and the $f$ terms play a role similar to consumption bundles albeit with the inequality reversed. The GACI condition rules out the possibility of cycles in ex-ante payoff across different decision problems.

Using this condition, we invoke a version of Afriat's theorem (see Chambers and Echenique (2016)).

**Lemma 6** (Afriat's Theorem)**.** *Let $\mathcal{D}$ be finite. For all $(A,B) \in \mathcal{D}^2$, let $\alpha_{A,B} \in \mathbb{R}$. If for all $A \in \mathcal{D}$ one has $\alpha_{A,A} = 0$ and for any sequence $A_1, A_2, \ldots, A_k \in \mathcal{D}$ with $\alpha_{A_i,A_{i+1}} \leq 0$ (with addition mod k) for all i it follows that $\alpha_{A_i,A_{i+1}} = 0$ for all i, then there exist numbers $U_A$ and $\lambda_A > 0$ such that for all $(A,B) \in \mathcal{D}^2$, $U_A \leq U_B + \lambda_B \alpha_{B,A}$.*

The other condition that characterizes the nonseparable costly information representation is the no improving action switches (NIAS) condition. This condition was first examined in the study of Bayesian decision makers in Caplin and Martin (2015).

**Condition 2** (No Improving Action Switches (NIAS))**.** *Given $\mu \in \Gamma$ and $u : X \to \mathbb{R}$, a dataset $(\mathcal{D}, \mathcal{P})$ satisfies NIAS if, for every $A \in \mathcal{D}$, $a \in Supp(P_A)$, and $b \in A$,*

$$\sum_{\omega \in \Omega} \mu(\omega) P_A(a \mid \omega)(u(a(\omega)) - u(b(\omega))) \geq 0$$

As we show in Theorem 5 below, the combination of GACI and NIAS completely characterizes the model of nonseparable costly information acquisition; moreover, one can impose additional properties on the nonseparable costly information representation. These conditions are monotonicity, quasiconcavity, and a normalization property on the function $V(\cdot,\cdot)$.

**Condition 3.** *The function $V : \mathbb{R} \times \Pi \to \mathbb{R} \cup \{-\infty\}$ satisfies weak monotonicity in information if for any $t \in \mathbb{R}$ and $\pi, \rho \in \Pi$ with $\pi$ a garbling of $\rho$, then*

$$V(t,\rho) \leq V(t,\pi).$$

The monotonicity condition says that if one adds noise to a signal $\rho$, then the noisier signal is cheaper. This is one definition of monotonicity and it agrees with the notion of informativeness introduced in Blackwell (1953).

**Condition 4.** *The function* $V : \mathbb{R} \times \Pi \to \mathbb{R} \cup \{-\infty\}$ *is quasiconcave if for any* $(t_1, \pi_1), (t_2, \pi_2) \in \mathbb{R} \times \Pi$ *and* $\lambda \in [0,1]$,

$$V(\lambda t_1 + (1-\lambda)t_2, \lambda\pi_1 + (1-\lambda)\pi_2) \geq \min\{V(t_1, \pi_1), V(t_2, \pi_2)\}.$$

This condition says if there is a mixture between two ex-ante payoffs and information structures, then the utility of the mixture is weakly higher than the worst case of the two environments. In particular, this implies quasiconcavity in information structures if one sets $t_1 = \pi_1 \cdot f$ and $t_2 = \pi_2 \cdot f$.

**Condition 5.** *Define* $\pi_0$ *as the information structure with* $\pi_0(\mu|\omega) = 1$ *for all* $\omega \in \Omega$. *The function* $V : \mathbb{R} \times \Pi \to \mathbb{R} \cup \{-\infty\}$ *satisfies the normalization if* $V(0, \pi_0) = 0$.

The normalization condition says that utility is normalized to zero when the ex-ante payoff is zero and an individual does not update their prior.

**Theorem 5.** *Given* $\mu \in \Gamma$ *and* $u : X \to \mathbb{R}$, *the dataset* $(\mathcal{D}, \mathcal{P})$ *has a nonseparable costly information representation if and only if it satisfies GACI and NIAS. Moreover, if GACI and NIAS are satisfied, then one can find a* $V$ *that rationalizes the data that satisfies Conditions 3, 4 and 5.*[6]

While we characterize a general model, we show that it is without loss to assume an individual's payoff is quasiconcave in the information structure for a fixed level of expected utility. Quasiconcavity might be interpreted as an informal statement that more informative structures are more costly to achieve. This is not meant in a Blackwell sense. Rather, given two information structures with known costs, taking a convex combination of them leads to a structure which is less costly than the highest cost of the two. The combination structure is intuitively less

[6]As an obvious consequence of Theorem 5, the model is also empirically equivalent to a model in which there is an endogenous (possibly singleton) set $H$ of hidden actions. In particular, the model $\pi \in \max_{h \in H} V(u \cdot \pi, h, \pi)$ is equivalent to ours since one could choose a set $H$ to be a singleton. Thus, unlike Machina (1984), adding the potential for hidden actions does not change the content of observable behavior, and hence is non-testable. This is also true of the model in Caplin and Dean (2015).

informative. While this property is certainly intuitive, the result is mathematical and owes to the structure of data and the same phenomenon whereby Afriat determined that convexity (as a property of preferences over consumption space) is non-testable.

**Example 1** (Continued)**.** *One can verify from this information that NIAS holds. To test whether the stochastic choice pattern can be rationalized by the nonseparable costly information acquisition model, it remains to verify that GACI holds. To this end, observe that*

$$\bar{\pi}_A \cdot f_A = 1.6 \quad ; \quad \bar{\pi}_{A'} \cdot f_A = 1.4$$

$$\bar{\pi}_{A'} \cdot f_{A'} = 7 \quad ; \quad \bar{\pi}_A \cdot f_{A'} = 8.$$

*Now, since*

$$\bar{\pi}_A \cdot f_A < \bar{\pi}_A \cdot f_{A'} \quad \textit{and} \quad \bar{\pi}_{A'} \cdot f_A < \bar{\pi}_{A'} \cdot f_{A'},$$

*there are no cycles that violate GACI. The stochastic choice pattern can be rationalized by the nonseparable costly information acquisition model.*

### 2.3.2 Multiplicative Information Cost

We now study a multiplicative costly information representation. In this representation, the cost is interpreted as losing a fraction of the ex-ante payoff. We interpret this cost as resulting from discounting due to unobserved delay when acquiring information.

**Definition 2.3.2.** Given $\mu \in \Gamma$ and $u : X \to \mathbb{R}_+$, a state dependent stochastic choice dataset $(\mathcal{D}, \mathcal{P})$ has a *multiplicative costly information representation* if there exists a function $R : \Pi \to \mathbb{R}_+$, information structures $\{\pi_A\}_{A\in\mathcal{D}}$, and choice functions $\{C_A\}_{A\in\mathcal{D}}$ such that:

1. Non-triviality: There exists $\pi \in \Pi$ such that $R(\pi) > 0$.

2. Information is optimal: For all $A \in \mathcal{D}$,

   $\pi_A \in \arg\max_{\pi\in\Pi} [R(\pi)(\pi \cdot f_A)]$.

3. Choices are optimal: For all $A \in \mathcal{D}$, the choice function $C_A : \text{Supp}(\pi_A) \to \Delta(A)$ is such that given $a \in A$ and $\gamma \in \text{Supp}(\pi_A)$ with $C_A(a \mid \gamma) = \Pr_A(a \mid \gamma) > 0$, then

$$\sum_{\omega \in \Omega} \gamma(\omega) u(a(\omega)) \geq \sum_{\omega \in \Omega} \gamma(\omega) u(b(\omega)) \quad \text{for all} \quad b \in A.$$

4. The data are matched: For all $A \in \mathcal{D}$, given $\omega \in \Omega$ and $a \in A$,

$$P_A(a \mid \omega) = \sum_{\gamma \in \text{Supp}(\pi_A)} \pi_A(\gamma \mid \omega) C_A(a \mid \gamma).$$

We note that one difference in the statement of the multiplicative costly information representation is that the utility index $u$ is required to be non-negative. While this is more restrictive than the other cases, this is a common property of multiplicative representations. For example, Chateauneuf and Faro (2009) make such an assumption. The condition that characterizes the multiplicative costly information representation is a version of the homothetic axiom of revealed preference; see Varian (1983a).[7]

**Condition 6** (Homothetic Axiom of Costly Information (HACI))**.** *Given data set $(\mathcal{D}, \mathcal{P})$, define $\mathcal{D}_0 = \{A \in \mathcal{D} \mid \sum_{\omega \in \Omega} \mu(\omega) u(a(\omega)) = 0 \text{ for all } a \in A\}$. We say the dataset $(\mathcal{D}, \mathcal{P})$ satisfies HACI if for all sequences $(\bar{\pi}_{A_1}, f_{A_1}), \ldots, (\bar{\pi}_{A_k}, f_{A_k})$ with $A_i \in \mathcal{D} \backslash \mathcal{D}_0$, that $\prod_{i=1}^{k} \frac{\bar{\pi}_{A_i} \cdot f_{A_{i+1}}}{\bar{\pi}_{A_i} \cdot f_{A_i}} \leq 1$ (with addition modulo k).*

HACI is essentially the homothetic axiom of revealed preference restricted to decision problems that give positive ex-ante payoff. The decision problems that give zero ex-ante payoff are removed since they can be trivially rationalized and they would create an indeterminate fraction.

As in the case of the nonseparable costly information representation, we are able to put additional properties on the function $R$. We find that $R$ respects monotonicity with respect to the

[7]It can also be derived as a relatively easy corollary from the work of Rochet (1987).

Blackwell partial order, is concave, and satisfies a normalization property. We now define these properties and then give a statement of the theorem.

**Condition 7.** *The function $R : \Pi \to \mathbb{R}_+$ satisfies weak monotonicity in information if $\rho, \pi \in \Pi$ with $\pi$ a garbling of $\rho$,*

$$R(\pi) \geq R(\rho).$$

**Condition 8.** *The function $R : \Pi \to \mathbb{R}_+$ is concave in information structures if for for any $\pi_1, \pi_2 \in \Pi$ and $\lambda \in [0,1]$,*

$$R(\lambda\pi_1 + (1-\lambda)\pi_2) \geq \lambda R(\pi_1) + (1-\lambda)R(\pi_2).$$

**Condition 9.** *Define $\pi_0$ as the information structure with $\pi_0(\mu|\omega) = 1$ for all $\omega \in \Omega$. The function $R$ satisfies normalization if $R(\pi_0) = 1$ and $R : \Pi \to [0,1]$.*

**Theorem 6.** *Given $\mu \in \Gamma$ and $u : X \to \mathbb{R}_+$, the dataset $(\mathcal{D}, \mathcal{P})$ has a multiplicative costly information representation if and only if it satisfies HACI and NIAS. Moreover, if HACI and NIAS are satisfied, then one can find an $R$ that rationalizes the data and satisfies Conditions 7, 8, and 9.*

**Example 1** (Continued)**.** *In additional to the general nonseparable cost model, the stochastic choice data in Example 1 can also be rationalized by the multiplicative cost model, which is a special case of the nonseparable cost model. In fact, the data satisfies HACI since*

$$\left(\frac{\bar{\pi}_A \cdot f_{A'}}{\bar{\pi}_A \cdot f_A}\right)\left(\frac{\bar{\pi}_{A'} \cdot f_{A'}}{\bar{\pi}_{A'} \cdot f_A}\right) = \left(\frac{80}{16}\right)\left(\frac{14}{70}\right) = 1.$$

Lastly, note that one could re-parameterize $R(\pi)$ to be $(1-K(\pi))$ where $K : \Pi \to [0,1]$ to interpret the costs as a fraction of ex-ante payoff. Alternatively, one could re-parameterize $R(\pi)$ to a function $\delta^{T(\pi)}$ where $T(\pi) \geq 0$ represents time delay.

### 2.3.3 Constrained Information Acquisition

The previous section studies nonseparable costly information acquisition, but there are other structures on preferences that are of interest. We consider when an individual is constrained to choose an information structure from a fixed set of information structures. The interpretation is that the decision maker does not have access to the full set of information structures when updating the prior, but all available information structures are costless.

**Definition 2.3.3.** Given $\mu \in \Gamma$ and $u : X \to \mathbb{R}$, a state dependent stochastic choice dataset $(\mathcal{D}, \mathcal{P})$ has a *constrained costly information representation* if there exists a set $\Pi_c \subseteq \Pi$ of available information structures, information structures $\{\pi_A\}_{A \in \mathcal{D}}$, and choice functions $\{C_A\}_{A \in \mathcal{D}}$ such that:

1. Non-triviality: The set $\Pi_c \neq \emptyset$.

2. Information is optimal: For all $A \in \mathcal{D}$, $\pi_A \in \arg\max_{\pi \in \Pi_c} \pi \cdot f_A$ .

3. Choices are optimal: For all $A \in \mathcal{D}$, the choice function $C_A : \text{Supp}(\pi_A) \to \Delta(A)$ is such that given $a \in A$ and $\gamma \in \text{Supp}(\pi_A)$ with $C_A(a \mid \gamma) = \Pr_A(a \mid \gamma) > 0$, then

$$\sum_{\omega \in \Omega} \gamma(\omega) u(a(\omega)) \geq \sum_{\omega \in \Omega} \gamma(\omega) u(b(\omega)) \quad \text{for all} \quad b \in A.$$

4. The data are matched: For all $A \in \mathcal{D}$, given $\omega \in \Omega$ and $a \in A$,

$$P_A(a \mid \omega) = \sum_{\gamma \in \text{Supp}(\pi_A)} \pi_A(\gamma \mid \omega) C_A(a \mid \gamma).$$

A constrained costly information structure is characterized by a condition similar to the Weak Axiom of Cost Minimization (Varian 1984). Using this intuition, the revealed information structures are analogous to inputs of production and $f_A$ are analogous to prices of inputs. To avoid confusion with the Weak Axiom of Revealed Preference, we call this the Binary Axiom of Costly Information.

**Condition 10** (Binary Axiom of Costly Information (BACI))**.** *The dataset $(\mathcal{D},\mathcal{P})$ satisfies BACI if for all $A,B \in \mathcal{D}$, it follows that*

$$\bar{\pi}_A \cdot f_A \geq \bar{\pi}_B \cdot f_A.$$

Similar to the nonseparable case, additional structure can be placed on a constrained costly information representation without restricting observable behavior. Using standard arguments, the constraint set $\Pi_c$ can be made convex.

**Theorem 7.** *Given $\mu \in \Gamma$ and $u : X \to \mathbb{R}$, the dataset $(\mathcal{D},\mathcal{P})$ has a constrained costly information representation if and only if it satisfies BACI and NIAS. Moreover, if BACI and NIAS are satisfied, then one can find a convex set $\Pi_c$ that rationalizes the data with a constrained costly information representation.*

**Example 1** (Continued)**.** *The stochastic choice data in Example 1 cannot be rationalized by the constrained costly information model, since the dataset violates BACI: $\bar{\pi}_{A'} \cdot f_{A'} < \bar{\pi}_A \cdot f_{A'}$.*

## 2.4 Relationship Among Models

The nonseparable information cost model generalizes all three alternative models of costly information acquisition. The constrained model, by contrast, is the most special one, and is a special case of both the additive and multiplicative models: it can be regarded as a multiplicative model with function $R(.)$ equals to 1 on $\Pi_c$ and 0 everywhere else; alternatively, it can also be regarded as an additive model where the additive cost function $K(.)$ equals to 0 on $\Pi_c$ and $+\infty$ everywhere else. Figure 2.1 below summarizes the relationship among various models of costly information acquisition.

As a point of reference, next we examine in details how the nonseparable costly information representation relates to the additive costly information representation in Caplin and Dean (2015).

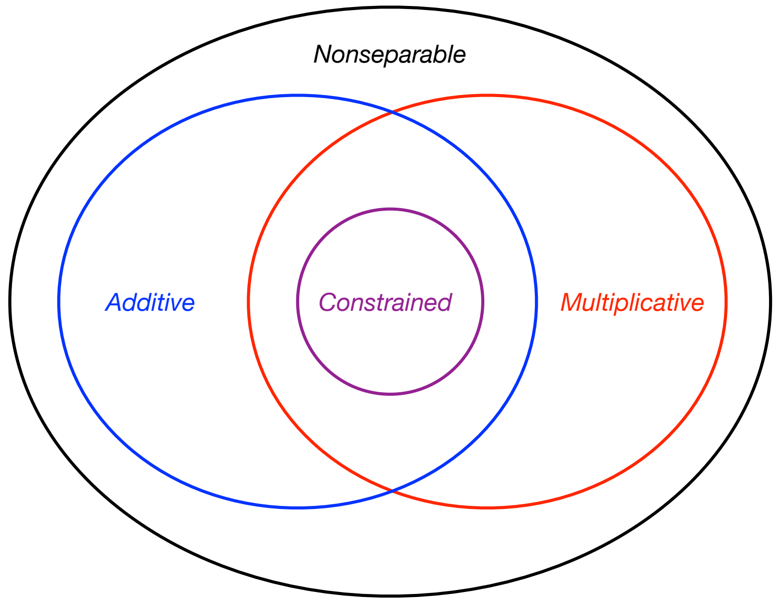


**Figure 2.1**: Relationship Among Models

We first review the definition of an additive costly information model, and show that the nonseparable model generalizes the additive model. We then show that one particular limitation of the additive model is that it forbids individuals from choosing less information whenever the menu provides more return to information, even if menus that generate higher returns might also entails higher costs for information.

### 2.4.1 Additive Information Cost Model

**Definition 2.4.1.** Given $\mu \in \Gamma$ and $u : X \to \mathbb{R}$, a state dependent stochastic choice dataset $(\mathcal{D}, \mathcal{P})$ has an *additive costly information representation* if there exists a function $K : \Pi \to \bar{\mathbb{R}} \cup \{\infty\}$, information structures $\{\pi_A\}_{A \in \mathcal{D}}$, and choice functions $\{C_A\}_{A \in \mathcal{D}}$ such that:

1. Non-triviality: There exists $\pi \in \Pi$ such that $K(\pi) < \infty$.

2. Information is optimal: For all $A \in \mathcal{D}$, $\pi_A \in \arg\max_{\pi \in \Pi} [\pi \cdot f_A - K(\pi)]$.

3. Choices are optimal: For all $A \in \mathcal{D}$, the choice function $C_A : \text{Supp}(\pi_A) \to \Delta(A)$ is such that

given $a \in A$ and $\gamma \in \text{Supp}(\pi_A)$ with $C_A(a \mid \gamma) = \Pr_A(a \mid \gamma) > 0$, then

$$\sum_{\omega \in \Omega} \gamma(\omega) u(a(\omega)) \geq \sum_{\omega \in \Omega} \gamma(\omega) u(b(\omega)) \quad \text{for all} \quad b \in A.$$

4. The data are matched: For all $A \in \mathcal{D}$, given $\omega \in \Omega$ and $a \in A$,

$$P_A(a \mid \omega) = \sum_{\gamma \in \text{Supp}(\pi_A)} \pi_A(\gamma \mid \omega) C_A(a \mid \gamma).$$

Caplin and Dean (2015) showed that an additive costly information representation is characterized by the NIAS condition and a no improving attention cycles (NIAC) condition. The NIAC condition is defined below.

**Condition 11** (No Improving Attention Cycles (NIAC))**.** *Given $\mu \in \Gamma$ and $u : X \to \mathbb{R}$, a dataset $(\mathcal{D}, \mathcal{P})$ satisfies NIAC if for all sequences $(\bar{\pi}_{A_1}, f_{A_1}), \ldots, (\bar{\pi}_{A_k}, f_{A_k})$ with $A_i \in \mathcal{D}$, then*

$$\sum_{i=1}^{k} \bar{\pi}_{A_i} \cdot f_{A_i} \geq \sum_{i=1}^{k} \bar{\pi}_{A_{i+1}} \cdot f_{A_i}$$

*where addition of the indices is modulo k.*

The interpretation of NIAC is that one cannot cycle through the information structures and improve the ex-ante payoff. From the definition of NIAC and GACI, it is easy to see that if a dataset satisfies NIAC, then the dataset also satisfies GACI with equality.

**Proposition 1.** *If the dataset $(\mathcal{D}, \mathcal{P})$ satisfies NIAC, then it also satisfies GACI.*

The proof, which is provided in the appendix, is standard: a violation of GACI implies the existence of a sequence $\bar{\pi}_{A_i} \cdot f_{A_i} \leq \bar{\pi}_{A_i} \cdot f_{A_{i+1}}$, where, say, $\bar{\pi}_{A_k} \cdot f_{A_k} < \bar{\pi}_{A_k} \cdot f_{A_1}$. Subtracting obtains that for each $i$, $\bar{\pi}_{A_i} \cdot (f_{A_i} - f_{A_{i+1}}) \leq 0$, with one inequality strict, whereby $\sum_i \bar{\pi}_{A_i} \cdot (f_{A_i} - f_{A_{i+1}}) < 0$. Rearranging terms now obtains a violation of NIAC. We come back to Example 1 again as an illustration.

**Example 1** (Continued). *Observe that the stochastic choice function cannot be rationalized by the additive cost model:*

$$\bar{\pi}_A \cdot f_A + \bar{\pi}_{A'} \cdot f_{A'} = 8.6 < 9.4 = \bar{\pi}_A \cdot f_{A'} + \bar{\pi}_{A'} \cdot f_A$$

*so that NIAC fails.*

### 2.4.2 Gross Return from Information

We note that an additively separable model forbids an individual from choosing a less informative information structure when there are "higher gross return from information", while this is allowed under the nonseparable cost model: this flexibility may be relevant if a menu generating higher gross return from information may at the same time entail more cost to information (for example, when the cost of information is the discounting incurred from waiting). We formally define "higher returns to information" below.

**Definition 2.4.2.** Menu $A$ provides a *higher gross return from information* than menu $B$ if for any information structure $\pi$ and $\pi'$ a garbling of $\pi$ with $\pi' \neq \pi$, we have [8]

$$\pi \cdot f_A - \pi' \cdot f_A > \pi \cdot f_B - \pi' \cdot f_B.$$

We establish that an individual with an additive costly information representation can never choose a less informative information structure when faced with a menu that has a higher gross return from information.

**Proposition 2.** *Suppose $\mathcal{D} = \{A, B\}$ for dataset $(\mathcal{D}, \mathcal{P})$ with menu $A$ providing a higher gross return from information than menu $B$. If $\bar{\pi}_A$ is a garbling of $\bar{\pi}_B$, then the choice data violates NIAC and thus cannot be generated by an additive costly information representation.*

[8] This definition is non-vacuous. In fact, it can be shown that menu $A$ provides a higher gross return from information than menu $B$ if and only if $f_A = f_B + g$ where $g$ is a strictly convex function.

The next result shows that a nonseparable model, on the contrary, always accommodates this behavior if the menu that provides a higher gross return from information also yields higher utility for any posterior.

**Proposition 3.** *Suppose $\mathcal{D} = \{A, B\}$ for dataset $(\mathcal{D}, \mathcal{P})$ with menu $A$ providing a higher gross return from information than menu $B$. If $\bar{\pi}_A$ is a garbling of $\bar{\pi}_B$, NIAS is satisfied, and $f_A > f_B$[9], then this dataset is rationalized by a nonseparable costly information representation.*

## 2.5 Out of Sample Prediction

One may wonder what type of data will violate GACI, or in other words, the extent to which the nonseparable model puts meaningful constraints on choice behavior. In this section, we first demonstrate the restrictions on choice probabilities for a specific two state environment, with a uniform prior and menus of two acts. This simple environment allows us to obtain a closed-form expression for the restrictions on choice probabilities. We then provide a numerical example as a further illustration.

### 2.5.1 The $2 \times 2$ Case

Let the states be given by $\Omega = \{\omega_1, \omega_2\}$. Let the menus be denoted $A = \{a, b\}$ and $A' = \{a', b'\}$. Assume without loss that $u(a(\omega_1)) > u(b(\omega_1))$ and $u(b(\omega_2)) > u(a(\omega_2))$. Similarly, assume that $u(a'(\omega_1)) > u(b'(\omega_1))$ and $u(b'(\omega_2)) > u(a'(\omega_2))$.

As is shown in Caplin and Dean (2015), NIAS on menu $A$ in this environment is equivalent to

$$P_A(a \mid \omega_1) \geq \max \begin{cases} \frac{u(b(\omega_2)) - u(a(\omega_2))}{u(a(\omega_1)) - u(b(\omega_1))} P_A(a \mid \omega_2), \\ \frac{u(b(\omega_2)) - u(a(\omega_2))}{u(a(\omega_1)) - u(b(\omega_1))} P_A(a \mid \omega_2) + \frac{u(a(\omega_1)) + u(a(\omega_2)) - u(b(\omega_1)) - u(b(\omega_2))}{u(a(\omega_1)) - u(b(\omega_1))} \end{cases}.$$

[9]We say $f_A > f_B$ if $f_A(\gamma) > f_B(\gamma)$ for all $\gamma \in \Gamma$.

A similar condition for menu $A'$ is equivalent to the satisfaction of NIAS there.

We focus on the case when decisions are aligned: we say the decisions are aligned if

$$a = \arg\max_{c \in \{a,b\}} \sum_{\omega \in \{\omega_1, \omega_2\}} \bar{\gamma}_{A'}^{a'}(\omega) u(c(\omega)),$$
$$b = \arg\max_{c \in \{a,b\}} \sum_{\omega \in \{\omega_1, \omega_2\}} \bar{\gamma}_{A'}^{b'}(\omega) u(c(\omega)),$$

and similar conditions hold for choices from $A'$ using the revealed information structure $\bar{\pi}_A$.

Essentially, decisions are aligned if the decision maker will choose action $a$ from menu $A$, if he used the information structure from $A'$ and received the signal that would have led him to choose $a'$. This assumption is made to make the algebra tractable. The same assumption is also implicitly assumed in Caplin and Dean (2015).

Now, there is a violation of GACI if

$$\bar{\pi}_A \cdot f_A \leq \bar{\pi}_A \cdot f_{A'} \quad \text{and} \quad \bar{\pi}_{A'} \cdot f_{A'} \leq \bar{\pi}_{A'} \cdot f_A$$

with one inequality strict. Under the above assumptions of NIAS and aligned choices, a *violation* of GACI is equivalent to the choice probabilities simultaneously satisfying the following two inequalities:

$$\begin{aligned} P_A(a \mid \omega_1)\Delta_1 + P_A(a \mid \omega_2)\Delta_2 &\leq \beta \\ P_{A'}(a' \mid \omega_1)\Delta_1 + P_{A'}(a' \mid \omega_2)\Delta_2 &\geq \beta \end{aligned} \tag{2.1}$$

where

$$\Delta_1 = u(a(\omega_1)) - u(a'(\omega_1)) + u(b'(\omega_1) - u(b(\omega_1))$$

$$\Delta_2 = u(a(\omega_2)) - u(a'(\omega_2)) + u(b'(\omega_2) - u(b(\omega_2))$$

$$\beta = u(b'(\omega_1)) + u(b'(\omega_2)) - u(b(\omega_1)) - u(b(\omega_2)).$$

Therefore, any probabilities that satisfy these inequalities with at least one strict inequality violate a nonseparable costly information representation.

In general, suppose one has a menu $M \in \mathcal{A}$ such that $M \notin \mathcal{D}$. If the dataset $\mathcal{D}$ satisfies NIAS and GACI, we can use the information to place bounds on the information structures that are consistent with the model using the restrictions of GACI and NIAS. The full set of restrictions is given by a *supporting set* as defined in Varian (1984).

Denote the set of information structures that support the menu $M$ that are consistent with GACI and NIAS by

$$S_{GACI}(M) = \{\pi_M \in \Pi \mid \{(\bar{\pi}_A, f_A)\}_{A \in \mathcal{D}} \cup (\pi_M, f_M) \text{ satisfies NIAS and GACI}\}.$$

This set places restrictions on $\pi_M$ that can be translated to restrictions on individual state dependent stochastic choices. It is easy to define supporting sets for multiplicatively separable, additively separable, and constrained costly information representation. While the supporting set is often difficult to compute, it provides the full set of $\pi_M$ consistent with a given representation.

### 2.5.2 Numerical Example: GACI *vs* NIAC

Let the actions' payoffs in menus $A$ and $A'$ take the following values:

$$u(a(\omega)) = \begin{cases} 5 & \text{if } \omega = \omega_1 \\ 1 & \text{if } \omega = \omega_2 \end{cases} \qquad u(b(\omega)) = \begin{cases} 0 & \text{if } \omega = \omega_1 \\ 4 & \text{if } \omega = \omega_2 \end{cases}$$

$$u(a'(\omega)) = \begin{cases} 4 & \text{if } \omega = \omega_1 \\ 2 & \text{if } \omega = \omega_2 \end{cases} \qquad u(b'(\omega)) = \begin{cases} 1 & \text{if } \omega = \omega_1 \\ 3 & \text{if } \omega = \omega_2 \end{cases} .$$

Substituting the above utility numbers into inequalities (2.1), we can see that the choice probabilities from $A$ and $A'$ violate GACI if and only if:

$$P_A(a \mid \omega_1) + P_A(b \mid \omega_2) \leq 1 \text{ and } P_{A'}(a' \mid \omega_1) + P_{A'}(b' \mid \omega_2) \geq 1. \tag{2.2}$$

On the other hand, substituting the above utility numbers into inequality (5) from Caplin and Dean (2015), we see that a violation of NIAC for this decision problem is equivalent to

$$P_A(a \mid \omega_1) + P_A(b \mid \omega_2) - [P_{A'}(a' \mid \omega_1) + P_{A'}(b' \mid \omega_2)] \leq 0. \tag{2.3}$$

By comparing (2.2) and (2.3) above, it is straightforward to see that in this numerical example, a violation of GACI implies a violation of NIAC, but not the other way around.

### 2.5.3 Limitations

The revealed preference conditions for costly information acquisition often provide interesting bounds and intuition for these models. Moreover, we note that an additive costly information representation has the property of being translation invariant in ex-ante payoff. Similarly, a multiplicative costly information representation has the property of being scale invariant in ex-ante payoff.

One may want to look at choices from menus of this type to violate an additively separable or multiplicatively separable costly information representation respectively. However, a dataset with menus that are additive utility translations of one another always satisfy NIAC. Similarly, a dataset with menus that are scale shifts of one another always satisfy HACI. To study these questions, we provide two definitions. For a menu $A = \{a_1, \ldots, a_n\} \in \mathcal{A}$ and $c \in \mathbb{R}$ let $A + c = \{a'_1, \ldots, a'_n\}$ be the menu that adds a constant utility $c$ to each act. That is, $u(a'_i(\omega)) = u(a_i(\omega)) + c$ for $i = 1, \ldots, n$ and all $\omega \in \Omega$. Similarly, let $cA = \{ca_1, \ldots, ca_n\}$ be the menu where the utility of all acts is multiplied by $c$, so $u(a'_i(\omega)) = cu(a_i(\omega))$ for $i = 1, \ldots, n$ and all $\omega \in \Omega$.

**Proposition 4.** *Let $\mu \in \Gamma$ and $u : X \to \mathbb{R}$. If the dataset $(\mathcal{D}, \mathcal{P})$ satisfies NIAS, $\mathcal{D} = \{A + c_1, A + c_2, \ldots, A + c_M\}$, and for all $m = 1, \ldots, M$ that $c_m \in \mathbb{R}$, then the dataset is rationalized by the additive costly information representation.*

**Proposition 5.** *Let $\mu \in \Gamma$ and $u : X \to \mathbb{R}_+$. Suppose the dataset $(\mathcal{D}, \mathcal{P})$ satisfies NIAS, $\mathcal{D} = \{c_1A, c_2A, \ldots, c_MA\}$, and and for all $m = 1, \ldots, M$ that $c_m \in \mathbb{R}_+$, then the dataset is rationalized by the multiplicative costly information representation.*

### 2.5.4 Unknown utility, unknown prior

The model we fleshed out requires utility and the prior to be known (in fact, we use a "reduced-form" model where signals can be written as they are only because the prior is known).

That said, even if the utility is unknown, some implications *may* be derived. As a general rule, if utility is totally unrestricted, the model has no content. This is a relatively standard observation, and owes to the fact that complete indifference can rationalize everything. On the other hand, in our abstract model, it makes sense to ask that utility lies in some *set*, $\mathcal{U}$. The notion of a value function $f_A$ now necessarily also depends on $u \in \mathcal{U}$.

Then an obvious violation of our model occurs when GACI is violated for each $u \in \mathcal{U}$.

There are nontrivial examples of such violations; simply fix some $u$ and find a violation of GACI for that $u$. Then by continuity, there is an open neighborhood $\mathcal{U}$ for which GACI is also violated.

Working out the more complete implications is a nontrivial task, but one we feel may bear some fruit.

On the other hand, suppose the prior is not known. This problem appears to be much more complicated. It is plausible that we may be able to address this question via duality techniques, but this is not certain. Two things are certain: we would require a more general model of signal structures (in which case the "reduced form" of the distribution over distributions is not observed), and, we can no longer use our knowledge of $u$ and $\mu$ to guide us as to whether the expected utility of $A$ is larger than the expected utility of $B$; these must be inferred. An interesting study of a related question is due to De Oliveira and Lamba (2018).

## 2.6 Conclusion

In this paper, we provide revealed preference characterizations for several models of costly information acquisition. The most general form allows for costs from time delay in addition to an additively separable cost. The characterization of these models follows directly from classical revealed preference theory. We also provide examples showing how the information acquisition differs across models.

## 2.7 Acknowledgement

Chapter 2, in full, is currently being prepared for submission for publication of the material. Chambers, Christopher P; Liu, Ce; Rehbeck, John. "Costly Information Acquisition". The dissertation author was a primary author of this paper.

# Chapter 3

# A Test for Risk-Averse Expected Utility

## 3.1 Introduction

The recent contribution of Kubler, Selden, and Wei (2014) provides a GARP-like test for risk-averse expected utility maximization in a contingent-consumption environment. In an environment with a single consumption good and finite states of the world, they establish an acyclicity condition on observed data which is both necessary and sufficient for a finite list of observed price and consumption pairs to be consistent with the hypothesis of expected utility maximization. Thus, their paper provides a counterpart of the classical work of Afriat (1967) with the added restriction that rationalizations be risk-averse expected utility.

As Kubler, Selden, and Wei (2014) note, their test is universal in nature, removing all existential quantification. Their test amounts to verifying that the product of certain cycles of risk-neutral prices be bounded above by one. Our aim in this note is to provide a *different* universal test. Our test should be distinguished from the Kubler, Selden, and Wei (2014) test in three ways. First, it applies to any finite number of consumption goods, whereas the test of Kubler, Selden, and Wei (2014) only applies for a single consumption good. Secondly, our test is intimately tied to the classical von Neumann-Morgenstern axioms of expected utility theory, and thus has a simple economic intuition. On the other hand, our test involves universal quantification over a potentially infinite number of objects, while the test in Kubler, Selden, and Wei (2014) can be reduced to universal quantification over a finite set.

We emphasize that what we mean by *test* is a method for falsifying the model with directly observable data. In other words, we say a model is *testable* if whenever data are inconsistent with the model, they can be demonstrated to be inconsistent. In this sense of the term test, a demonstration is distinct from an algorithm which would find this falsifying certificate. Hence, a test in our sense is not intended to be useful from a computational perspective, and as far as we can tell, ours is not in general. Indeed; there are already practical algorithms for determining when the expected utility model is falsified in our context. Rather, such a test is important for understanding the economic content of the model, by specifying a condition stated in terms of data alone, which does not reference unobservable concepts such as utilities or marginal rates of substitution. As a point of comparison, the work of Richter (1966) can be understood as providing the testable restrictions of the preference maximization hypothesis; however, no general algorithm would exist in Richter's case either.[1]

Our test is perhaps most closely related to an early revealed preference test of expected utility due to Fishburn (1975). Fishburn constructs a test for an abstract environment of choice over lotteries with finite support. In his setting, one observes a finite set of binary comparisons; some are weak, and some are strict. Fishburn provides necessary and sufficient conditions for there to exist an expected utility ranking which extends the observed binary comparisons. Imagine that we observe lottery $l_k$ weakly preferred to lottery $l'_k$ for $k = 1, \ldots, g$, and $l_k$ strictly preferred to $l'_k$ for $k = g+1, \ldots, K$. Fishburn establishes that these observations are consistent with expected utility maximization if there is no probability distribution over $\{1, \ldots, K\}$ which puts positive probability on $\{g+1, \ldots, K\}$, and for which the mixture of the $l_k$'s under this probability distribution is equal to the mixture of the $l'_k$'s. Fishburn's test can be viewed as claiming that the

[1] In the special case where budgets are given by linear inequalities and preference satisfies monotonicity, an algorithm exists for Richter's test, namely the Afriat test. Here we refer to the abstract budget environment.

smallest possible extension of the observed relations satisfying both independence and transitivity leads to no contradiction. We stress that Fishburn's test also presents with no algorithm: no recipe is given for finding the probability distribution.

In our case, we have $n$ commodities, and a finite set of states $\Omega = \{\omega | 1, 2 \ldots, S\}$. We observe a finite list of prices and contingent consumption bundles chosen at those prices $(x^k, p^k)$, $k \in \{1, \ldots K\}$. Consumption in state $\omega$ at observation $k$ is of the form $x^k_\omega \in \mathbb{R}^n_+$. Probabilities over $\Omega$ are known and are given by the full support distribution $\pi$.

We first ask: What could reveal a violation of the joint hypothesis of expected utility and risk aversion in this context? There are only a finite set of states of the world, with known probabilities, but if the choices *were* rationalizable by an expected utility preference, there would be a natural extension to a preference over the set of all simple lotteries. One such violation would look like the following: suppose that for each $x^k$, there is some $y^k$ which is feasible at prices $p^k$. In other words, the induced lottery $l_{x^k}$ is *revealed preferred* to the induced lottery $l_{y^k}$. And suppose that there is some $g$ for which $y^g$ is strictly cheaper than $x^g$ at prices $p^g$. In other words, the induced lottery $l_{x^g}$ is *revealed strictly preferred* to the induced lottery $l_{y^g}$. Now, suppose we can find, for each $k$, a lottery $l'_k$ which is a mean-preserving spread of $l_{y^k}$. If the data were rationalizable by a risk-averse expected utility preference, the lottery $l_{x^k}$ would be preferred to $l'_k$ for all $k$ (and $l_{x^g}$ would be strictly preferred to $l'_g$).

We now have a set of $K$ pairs of lotteries $(l_{x^k}, l'_k)$ which could be obtained in the preceding fashion. These data can be tested with Fishburn's condition. If, in fact, they violate Fishburn's condition, then we know that the original data cannot be expected utility rationalizable.

So far this is very simple. However, in the demand setting, for each observation $(p^k, x^k)$, there are usually infinitely many candidates for the above $y^k$, and for each $y^k$, an infinite number of possible mean-preserving spreads $l'_k$. This would result in an infinite number of possible $\{(l_{x^k}, l'_k)\}_{k=1}^K$ sets. While the Fishburn condition is sufficient to ensure each $\{(l_{x^k}, l'_k)\}_{k=1}^K$ set has its own preference extension, it has nothing to say about whether or not there is a *single* preference extension for the infinitely many revealed preference relations.

In fact, what we show is the following: If the data are not risk-averse expected utility rationalizable, then there exists at least one set, $\{(l_{x^k}, l'_k)\}_{k=1}^K$, as above, that violates Fishburn's condition. In addition, they can be chosen to violate Fishburn's condition in a very stark way: one must only test the uniform lottery over $\{1, \ldots, K\}$.

Moreover, the support of each $l'_k$ can be chosen to consist only of consumption that was actually observed demanded at some state; *i.e.* the support can be chosen amongst elements of the form $x^k_\omega$. This resonates with the idea from Polisson et al. (2015), who observe that in order to rationalize data, it is both necessary and *sufficient* to maintain consistency on the set of minimally extended "imaginary" data, constructed from those actually observed. However, while Polisson et al. (2015) is concerned with developing Afriat-style algorithms (see Afriat (1967)) for testing decision models with money lotteries, our focus is developing universal statements about data from lotteries of general consumption bundles, which provides direct falsification of the expected utility model under risk aversion.

It is important to note that due to the infinite nature of our test, our contribution lies not in providing a procedure to be implemented to check actual data; for such a test, the readers are directed to the work by Green and Srivastava (1986). Instead, the main contribution of our test is that it extends the intuition of the Fishburn test to demand-based observations: whenever the smallest possible extension of the observed relations satisfying both independence and transitivity leads to no contradiction, the data are rationalizable by risk-averse expected utility preference. In addition, the test by Green and Srivastava involves theoretical objects that are not directly observable, while our conditions directly characterize exactly which types of data are ruled out by the hypothesis of expected utility maximization, and thus can be interpreted as its UNCAF axiomatization, when observations are made in a demand-based framework.[2]

[2]UNCAF stands for *universal negation of conjunction of atomic formulas*. Chambers, Echenique, and Shmaya (2014) demonstrate that theories which make no non-empirical predictions are exactly those which have UNCAF axiomatizations.

The idea of the proof is remarkably simple, and is a simple restatement of the dual set of linear inequalities stemming from the Afriat-style inequalities of Green and Srivastava (1986) or Varian (1983b).

A host of other interesting papers have recently studied choice data in the context of expected utility maximization. In particular, Echenique and Saito (2015) investigates the subjective expected utility version of the model, which forms a kind of analogue of the Kubler, Selden, and Wei (2014) test. It would be interesting to propose a test of our structure in the subjective expected utility framework. Epstein (2000) investigates the empirical content of the notion of probabilistic sophistication (due to Machina and Schmeidler (1992)), providing a test which can refute the hypothesis.

## 3.2 The Model

We assume that there is a finite state space $\Omega = \{\omega | 1, 2, \ldots, S\}$ and a finite collection of consumption goods, labeled $1, 2, \ldots, N$. The agent is given an objective probability distribution over states $\pi \in \Delta(\Omega)$, where for all $\omega \in \Omega$, $Pr(\omega) = \pi_\omega > 0$. An observation is a pair $(p, x)$, where $p \in \mathbb{R}^{SN}_{++}$ is a list of the prices of all $N$ consumption goods under all $S$ possible states, and $x \in \mathbb{R}^{SN}_{+}$ details the purchased amount of each consumption good under each state of the world.[3] We assume that our *data set* $\mathcal{D}$ consists of a $K$ tuple of $(x, p)$ pairs, i.e. $\mathcal{D} = \{(x^k, p^k)^K_{k=1}\}$. $K$ is assumed finite.

[3] As usual, $\mathbb{R}_{++}$ denotes the positive reals, and $\mathbb{R}_{+}$ the nonnegative reals.

In particular,

$$x^k = \begin{bmatrix} x_1^k \\ \vdots \\ x_\omega^k \\ \vdots \\ x_S^k \end{bmatrix} \quad p^k = \begin{bmatrix} p_1^k \\ \vdots \\ p_\omega^k \\ \vdots \\ p_S^k \end{bmatrix}$$

and

$$x_\omega^k = \begin{bmatrix} x_{\omega,1}^k \\ \vdots \\ x_{\omega,N}^k \end{bmatrix} \quad p_\omega^k = \begin{bmatrix} p_{\omega,1}^k \\ \vdots \\ p_{\omega,N}^k \end{bmatrix}$$

where for all $\omega, k, n$, $x_{\omega,n}^k \geq 0$ and $p_{\omega,n}^k > 0$. Each $x^k$ is referred to as a *contingent consumption bundle*, and $x_\omega^k$ a *state-specific consumption bundle*. We use $C = \mathbb{R}_+^{NS}$ to denote the set of all contingent consumption bundles.

We say that $\mathcal{D}$ is *risk-averse expected utility rationalizable* if there exists a concave, continuous, and increasing $u : \mathbb{R}_+^N \to \mathbb{R}$ for which for all $k$, $x^k$ solves

$$\max_{x \in \mathbb{R}_+^{SN}} \sum_\omega \pi_\omega u(x_\omega)$$

subject to $p^k \cdot x \leq p^k \cdot x^k$.[4]

Given a data set $\mathcal{D}$, we collect all the state-specific consumption bundles $x_\omega^k$ observed in the data:

$$\mathcal{X} = \{x \in \mathbb{R}_+^N | x = x_\omega^k \text{ for some } k \text{ and } \omega \text{where } (x^k, p^k) \in \mathcal{D}\}.$$

Denote the set of all simple lotteries on $\mathbb{R}_+^N$ with finite support by $\Delta_s(\mathbb{R}_+^N)$. Denote the set of all

[4]We take increasing to mean that if $x \geq y$ and $x \neq y$, then $u(x) > u(y)$.

lotteries on $\mathcal{X}$ by $\Delta(\mathcal{X})$. Note that $\Delta(\mathcal{X}) \subseteq \Delta_s(\mathbb{R}^N_+)$.

Any contingent consumption bundle $x^k \in C$ induces an element $l_{x^k} \in \Delta(\mathcal{X})$, which places probability $\pi_\omega$ on $x^k_\omega$. As such, a pair of revealed preference relations $\succeq^C$ and $\succ^C$ can be defined on $\Delta(\mathcal{X})$:

For $x, y \in C$, $l_x \succeq^C l_y$ if $x = x^k$ for some $(x^k, p^k) \in \mathcal{D}$ and $p^k \cdot y \leq p^k \cdot x$. For $x, y \in C$, $l_x \succ^C l_y$ if $x = x^k$ for some $(x^k, p^k) \in \mathcal{D}$ and $p^k \cdot y < p^k \cdot x$. $\succeq^C$ is intended to represented a revealed weak preference and $\succ^C$ a revealed strict preference.

Moreover, to test the hypothesis of risk aversion, it is natural to extend the above revealed preference relations to $\Delta_s(\mathbb{R}^N_+)$. For example, suppose that $l_x \succeq^C l_y$, and $l \in \Delta_s(\mathbb{R}^N_+)$ can be obtained by a sequence of mean-preserving spreads of $l_y$.[5] If our decision maker's behavior is consistent with risk-averse expected utility maximization, it follows that $l_x$ should also be preferred to $l$. These ideas motivate the following definitions.

For $l, l' \in \Delta_s(\mathbb{R}^N_+)$, $l \succeq^{m.p.s.} l'$ if $l'$ can be obtained by a series of mean-preserving spreads of $l$. Define the pair of binary relations $\succeq^R$ and $\succ^R$ on $\Delta_s(\mathbb{R}^N_+)$ by

$$l \succeq^R l'' \text{ if there exists } l' \text{ such that } l \succeq^C l' \succeq^{m.p.s.} l''$$

and

$$l \succ^R l'' \text{ if there exists } l' \text{ such that } l \succ^C l' \succeq^{m.p.s.} l''$$

If the agent's behavior is consistent with risk-averse expected utility maximization, the pair of relations $\succeq^R, \succ^R$ will necessarily satisfy Fishburn's condition on $\Delta_s(\mathbb{R}^N_+)$; *i.e.* if $l_k \succeq^R l'_k$ for $k = 1, \ldots, g$, and $l_k \succ^R l'_k$ for $k = g+1, \ldots, K$, then there are no $\{\mu_i\}_{i=1}^K \subseteq \mathbb{R}^K_+$, with $\sum_{k=g+1}^K \mu_k > 0$, and $\sum_1^K \mu_k l_k = \sum_1^K \mu_k l'_k$. As we show in our main result, it turns out that a *sufficient* condition for the data $\mathcal{D}$ to conform with risk aversion and expected utility maximization is that the restriction

[5] That is, if there exists a random variable $\varepsilon$ such that $l \overset{d}{=} l_y + \varepsilon$ with $E(\varepsilon | l_y) = 0$. "$\overset{d}{=}$" here means "has the same distribution as". See Rothschild and Stiglitz (1970) for more details.

of $\succeq^R, \succ^R$ to $\Delta(\mathcal{X})$ satisfies Fishburn's condition.

*Theorem* 1. For every data set $\mathcal{D} = \{(x^k, p^k)_{k=1}^K\}$, the following are equivalent:

I For any $\{l'_k\}_{k=1}^K \subseteq \Delta(\mathcal{X})$ for which $l_{x^k} \succeq^R l'_k$ for all $k$, there is no $\{\mu_k\}_{k=1}^K \subseteq \mathbb{R}_+^K$ for which $\sum_{\{k: l_k \succ^R l'_k\}} \mu_k > 0$ and $\sum_1^K \mu_k l_{x^k} = \sum_1^K \mu_k l'_k$.

II Suppose that for each $k \in \{1, \ldots, K\}$ and $\omega \in \Omega$, $S^k_\omega : \{1, \ldots, K\} \times \Omega \to \mathbb{R}_+$ is a function, such that for all $k, \omega$, $\sum_{g,\tau} S^k_\omega(g, \tau) = \pi_\omega = \sum_{g,\tau} S^g_\tau(k, \omega)$. If, in addition, for all $k$,

$$p^k \cdot x^k \geq p^k \cdot \left( \frac{\sum_g \sum_\tau S^k_\omega(g,\tau) x^g_\tau}{\pi_\omega} \right)_{\omega \in \Omega}$$

then there is no $k$ for which $p^k \cdot x^k > p^k \cdot \left( \frac{\sum_g \sum_\tau S^k_\omega(g,\tau) x^g_\tau}{\pi_\omega} \right)_{\omega \in \Omega}$.[6]

III For all $\omega, \tau \in \Omega$ and $k, g \in \{1, ..., K\}$ there exist $u^k_w, u^g_\tau \geq 0$ and $\lambda_k, \lambda_g > 0$ s.t. $u^k_\omega \leq u^g_\tau + \lambda_g \frac{p^g_\tau}{\pi_\tau} \cdot (x^k_\omega - x^g_\tau)$.[7]

IV Data set $\mathcal{D}$ is risk-averse expected utility rationalizable.

Before proceeding, we comment on cases I and II, which are our contribution. Case I considers the smallest possible preference extension "consistent" with the data, risk-aversion, and the expected utility hypothesis. It claims that if this extension is meaningfully defined; in that we cannot derive that a lottery $l$ is strictly preferred to itself, then the data are expected utility rationalizable. Importantly, we only need to consider lotteries whose support are actual observed consumption bundles. This can be seen as a natural analogue of Fishburn's condition as applied to $l_{x^k}$ and $l'_k$.

[6] $\left( \frac{\sum_g \sum_\tau S^k_\omega(g,\tau) x^g_\tau}{\pi_\omega} \right)_{\omega \in \Omega} = \left( \frac{\sum_g \sum_\tau S^k_1(g,\tau) x^g_\tau}{\pi_1}, \ldots, \frac{\sum_g \sum_\tau S^k_S(g,\tau) x^g_\tau}{\pi_S} \right)$ *i.e.* $\frac{\sum_g \sum_\tau S^k_\omega(g,\tau) x^g_\tau}{\pi_\omega}$ is the consumption in state $\omega$.

[7] Green and Srivistava's proof of this statement assumes the non-emptyness of $u$'s superdifferential over $\mathbb{R}^n_+$; however, it is easy to modify their proof even with empty superdifferential on the boundary. Essentially, whenever $x^g$ is known to be a utility maximizer, we can always find $\nabla u(x^g_\tau)$ in the superdifferential of $u$ for which $\nabla u(x^g_\tau) = \lambda_g \frac{p^g_\tau}{\pi_\tau}$ (see Theorem 28.3 in Rockafellar (1997)). So $u^k_\omega \leq u^g_\tau + \nabla u(x^g_\tau) \cdot (x^k_\omega - x^g_\tau) = u^g_\tau + \lambda_g \frac{p^g_\tau}{\pi_\tau} \cdot (x^k_\omega - x^g_\tau)$.

Case II demonstrates a dual system of linear inequalities to the inequalities of case III, which was derived previously by Green and Srivastava (1986). The interpretation of the terms $S^k_\omega$ is as a system of probability weights. To obtain some intuition on Case II, suppose that the inequalities therein are satisfied, then one can find a contradiction as follows: For each $k$, by demand behavior, the inequalities in Case II imply that the lottery $l_{y^k}$ induced by the contingent consumption bundle $\left(\frac{\sum_g \sum_\tau S^k_\omega(g,\tau)x^g_\tau}{\pi_\omega}\right)_{\omega\in\Omega}$ is revealed weakly worse than the lottery $l_{x^k}$ induced by $x^k$, with strict preference for at least one $k$. Observe that $l_{y^k}$ is a lottery that places probability $\pi_\omega$ on $\frac{\sum_g \sum_\tau S^k_\omega(g,\tau)x^g_\tau}{\pi_\omega}$. Since $\sum_{g,\tau} S^k_\omega(g,\tau) = \pi_\omega$, simple algebra (included in the proof) shows that the lottery $l'_k$, which places probability weight $S^k_\omega(g,\tau)$ on $x^g_\tau$, is a mean-preserving spread of $l_{y^k}$. If the data were really consistent with the hypothesis of *risk-averse* expected utility maximization, transitivity would imply that for each $k$, the lottery $l'_k$ should be worse than the lottery $l_{x^k}$, strictly so for at least one $k$. We now have in total $K$ revealed preference relations between the pairs of lotteries $l_{x^k}$ and $l'_k$. As we demonstrate in the proof, the condition $\sum_{g,\tau} S^g_\tau(k,\omega) = \pi_\omega$ then allows us to find a violation by applying the condition from Fishburn (1975) on the lotteries $l_{x^k}$ and $l'_k$ across all $k$.

The following example illustrates the theorem.

**Example 5.** *Consider the case $k \in \{1,2\}$, $\Omega = \{1,2\}$ and $N = 2$: There are 2 observations, each consisting of the price and purchased quantity for the consumptions good under 2 possible states of the world. Suppose each of the two states are equally likely; $\pi_1 = \pi_2 = .5$. Suppose we observe:*

$$(x^1, p^1) = \left( \begin{bmatrix} 0 \\ 0 \\ 10 \\ 5 \end{bmatrix}, \begin{bmatrix} 5 \\ 10 \\ 5 \\ 10 \end{bmatrix} \right) \quad (x^2, p^2) = \left( \begin{bmatrix} 4 \\ 2 \\ 6 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 8 \\ 5 \\ 10 \end{bmatrix} \right)$$

*In this case there is no violation of GARP. However, since the hypothesis of risk-averse EU preference is stronger than than GARP, we show that this case still violates our conditions.*

***Violation of Statement I:*** *The induced lotteries by $x^1$ and $x^2$ are $l_{x^1} = ((10,5),1/2;(0,0),1/2)$ and $l_{x^2} = ((4,2),1/2;(6,3),1/2)$, respectively. To see that this is a violation of statement I, consider contingent consumption bundles $y^1 = y^2 = ((5,2.5);(5,2.5))$ which induce $l_{y^1} = l_{y^2} = ((5,2.5),1)$. Clearly $p^1 \cdot x^1 \geq p^1 \cdot y^1$, and $p^2 \cdot x^2 > p^2 \cdot y^2$. So by definition $l_{x^1} \succeq^C l_{y^1}$ and $l_{x^2} \succ^C l_{y^2}$.*

*Observe that the lottery $l'_1 = ((4,2),1/2;(6,3),1/2)$ is a mean-preserving spread of $l_{y^1}$ and the lottery $l'_2 = ((10,5),1/2;(0,0),1/2)$ is a mean-preserving spread of $l_{y^2}$. By definition $l_{x^1} \succeq^R l'_1$ and $l_{x^2} \succ^R l'_2$. However,*

$$\frac{1}{2}l_{x^1} + \frac{1}{2}l_{x^2} = \frac{1}{2}l'_1 + \frac{1}{2}l'_2$$

*This constitutes a violation of Statement I.*

***Violation of Statement II:***

*Set $S_1^1(2,1) = S_1^2(1,1) = \frac{1}{5}$, $S_1^1(2,2) = S_2^2(1,1) = \frac{3}{10}$, $S_2^1(2,1) = S_1^2(1,2) = \frac{3}{7}$, and $S_2^1(2,2) = S_2^2(1,2) = \frac{1}{14}$.*

*To solve:*

$$\begin{bmatrix} 0 \\ 0 \\ 10 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 10 \\ 5 \\ 10 \end{bmatrix} > 2 \cdot \begin{bmatrix} 5 \\ 10 \\ 5 \\ 10 \end{bmatrix} \cdot \begin{bmatrix} S_1^1(2,1)*4 + S_1^1(2,2)*6 \\ S_1^1(2,1)*2 + S_1^1(2,2)*3 \\ S_2^1(2,1)*4 + S_2^1(2,2)*6 \\ S_2^1(2,1)*2 + S_2^1(2,2)*3 \end{bmatrix}$$

$$\begin{bmatrix} 4 \\ 2 \\ 6 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 8 \\ 5 \\ 10 \end{bmatrix} \geq 2 \cdot \begin{bmatrix} 4 \\ 8 \\ 5 \\ 10 \end{bmatrix} \cdot \begin{bmatrix} S_1^2(1,1)*0 + S_1^2(1,2)*10 \\ S_1^2(1,1)*0 + S_1^2(1,2)*5 \\ S_2^2(1,1)*0 + S_2^2(1,2)*10 \\ S_2^2(1,1)*0 + S_2^2(1,2)*5 \end{bmatrix}$$

$$S_1^1(2,1) + S_1^1(2,2) = S_2^1(2,1) + S_2^1(2,2) = \frac{1}{2}$$

A couple of observations are in order. It can be shown that both (I) and (II) of our properties imply GARP. Suppose by means of contradiction that GARP is violated, *i.e.* that there are contingent consumption bundles $z^{k_1}, \ldots, z^{k_m}$ such that $p^{k_1} \cdot z^{k_1} \geq p^{k_1} \cdot z^{k_2}, p^{k_2} \cdot z^{k_2} \geq p^{k_2} \cdot z^{k_3}, \ldots, p^{k_m} \cdot z^{k_m} > p^{k_m} \cdot z^{k_1}$, where without loss we may assume there is no repetition in the cycle. This implies $l_{z^{k_1}} \succeq^C l_{z^{k_2}} \succeq^C \ldots \succeq^C l_{z^{k_m}} \succ^C l_{z^{k_1}}$.

To see that (I) implies GARP, observe that since $\succeq^C$ implies $\succeq^R$ and $\succ^C$ implies $\succ^R$, we have $l_{z^{k_1}} \succeq^R l_{z^{k_2}} \succeq^R \ldots l_{z^{k_m}} \succ^R l_{z^{k_1}}$. Let $l_{x^i} = l_{z^{k_i}}$ and $l'_i = l_{z^{k_{i+1}}}$ as in property (I), then a uniform distribution $\mu$ over the indices $i = 1, 2, \ldots m$ constitutes a violation of (I).

For (II), consider the following set of $S_\omega^k(g,\tau)$'s in property II: For $k = k_i$ for some $i$ (that is, if $k$ shows up in the cycle)

$$S_\omega^{k_i}(g,\tau) = \begin{cases} \pi_\omega & \text{if } g = k_{i+1} \text{ and } \tau = \omega \\ 0 & \text{otherwise} \end{cases}$$

and for $k \neq k_i$ for any $i$ ($k$ not in the cycle)

$$S_\omega^{k}(g,\tau) = \begin{cases} \pi_\omega & \text{if } g = k \text{ and } \tau = \omega \\ 0 & \text{otherwise} \end{cases}$$

Then the cycle condition gives a violation of property (II), a contradiction.

Finally, we wish to emphasize that the result is by no means a trivial consequence of Fishburn (1975). In his paper, he also considers the issue of testing the consistency of revealed preference relations with functional restrictions on the von Neumann-Morgenstern utility index (as we wish to test for concavity and monotonicity). Specifically, he wants to test when observed data are consistent with the utility index $u$ belonging to some convex cone $\mathcal{U}$. Again, he assumes a finite number of relations (which does not hold in our context). A natural guess is that if $l_k$ is revealed weakly preferred to $l'_k$ for $k = 1, \ldots, g$ and revealed strictly preferred to $l'_k$ for $k = g+1, \ldots, K$, then if there is $\mu \in \Delta(K)$ for which $\mu(\{g+1, \ldots, K\}) > 0$ and $u \cdot \left(\sum_k \mu_k l'_k\right) \geq u \cdot (\sum_k \mu_k l_k ß)$ for all $u \in \mathcal{U}$, then the observed data are inconsistent with expected utility maximization with utility index $u \in \mathcal{U}$[8]. In our case, for example, we would consider the cone of concave, nondecreasing and locally non-satiated functions; the claim would then be that $\sum_k \mu_k l'_k$ second order stochastically dominates $\sum_k \mu_k l_k$. Of course, the existence of such a $\mu$ refutes the hypothesis of expected utility rationalization with $u \in \mathcal{U}$, but for technical reasons, the converse statement need not hold in general (it would hold, for example, if the cone $\mathcal{U}$ were polyhedral, which is not the case here). However, we are able to show that owing to the special structure of linear pricing, a converse statement along the lines of this idea does in fact hold in the demand-based environment. In fact, it holds *even though observed revealed preference relations are infinite*.

*Proof.* **(III $\Leftrightarrow$ IV)**

The equivalence of III and IV is due to Green and Srivastava (1986).

[8]Here we continue to use $x$ and $z$ for lotteries, and dot product for integration with respect to measures.

**(II $\Leftrightarrow$ III)**

We proceed to show that II and III are equivalent. To this end, observe that III does not hold if and only if there is no solution to the following linear system.[9] $Ab \geq 0$ and $\lambda \gg 0$, where

$$b = \begin{bmatrix} u_1^1 \\ u_2^1 \\ \vdots \\ u_S^K \\ \lambda \end{bmatrix} \quad \lambda = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_K \end{bmatrix}$$

and $A$ is equal to the top two quadrants of the matrix below:

$$T = \begin{array}{cc} & \begin{array}{cccccccccccc} u_1^1 & \cdots & u_\omega^k & \cdots & u_\tau^g & \cdots & u_S^K & \lambda_1 & \cdots & \lambda_k & \cdots & \lambda_K \end{array} \\ \begin{array}{c} \eta_{1,1,1,1} \\ \vdots \\ \eta_{k,\omega,g,\tau} \\ \vdots \\ \hline \vdots \\ \eta'_k \\ \vdots \end{array} & \left[ \begin{array}{ccccccc|ccccc} 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & -1 & \cdots & 0 & 0 & \cdots & \frac{p_\omega^k}{\pi_k} \cdot (x_\tau^g - x_\omega^k) & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots \\ \hline \vdots & & \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots & \vdots & & \vdots & & \vdots \end{array} \right] \end{array}$$

By construction of $T$ and a standard theorem of the alternative (see for example Mangasarian (1994) p. 30), the nonexistence of $b, \lambda$ such that $Ab \geq 0$ and $\lambda \gg 0$, is equivalent to the existence of $\eta \geq 0$ such that $T'\eta \leq 0$, where

[9] Vector inequalities are $x \geq y$ if $x_i \geq y_i$ for all $i$ and $x \gg y$ if $x_i > y_i$ for all $i$.

$$\eta = \begin{bmatrix} \eta_{1,1,1,1} \\ \vdots \\ \eta_{K,S,K,S} \\ \eta' \end{bmatrix} \eta' = \begin{bmatrix} \eta'_1 \\ \vdots \\ \eta'_K \end{bmatrix}$$

such that at least one $\eta'_k > 0$.

This is equivalent to

$$\sum_{\omega} \sum_{(g,\tau)\neq(k,\omega)} \eta_{k,\omega,g,\tau} \frac{p^k_\omega}{\pi_\omega} \cdot (x^g_\tau - x^k_\omega) \leq 0 \, \forall k \tag{3.1}$$

with strict inequality for at least one $k$, and

$$\sum_{(g,\tau)\neq(k,\omega)} \eta_{k,\omega,g,\tau} = \sum_{(g,\tau)\neq(k,\omega)} \eta_{g,\tau,k,\omega} \, \forall k, \omega \tag{3.2}$$

We claim that a solution to systems (3.1) and (3.2), implies the existence of $\gamma_{k,\omega,g,\tau} \geq 0$ so that

$$\sum_{\omega} \sum_{(g,\tau)} \gamma_{k,\omega,g,\tau} \frac{p^k_\omega}{\pi_\omega} \cdot (x^g_\tau - x^k_\omega) \leq 0 \, \forall k \tag{3.3}$$

$$\sum_{(g,\tau)} \gamma_{k,\omega,g,\tau} = \sum_{(g,\tau)} \gamma_{g,\tau,k,\omega} = \pi_\omega \, \forall k, \omega \tag{3.4}$$

with at least one inequality in (3.3) being strict, effectively showing (3.3) and (3.4) are equivalent to (3.1) and (3.2).

To see this, list the $\eta_{k,\omega,g,\tau}$'s from systems (3.1) and (3.2) as in Figure 3.1 (Notice that system (3.2) ensures that columns and rows passing through the same diagonal element, like the column and row in red and blue boxes, sum up to the same number.) We now construct a

$$
\begin{array}{cc}
 & \begin{array}{ccccccccccc} 1,1 & & 1,\omega & & 1,S & & k,\omega & & K,S \end{array} \\
\begin{array}{c} 1,1 \\ \\ 1,\omega \\ \\ 1,S \\ \\ k,\omega \\ \\ K,S \end{array} &
\begin{bmatrix}
\eta_{1,1,1,1} & \cdots & \boxed{\eta_{1,1,1,\omega}} & \cdots & \eta_{1,1,1,S} & \cdots & \eta_{1,1,k,\omega} & \cdots & \eta_{1,1,K,S} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \cdots \\
\eta_{1,\omega,1,1} & \cdots & \eta_{1,\omega,1,\omega} & \cdots & \eta_{1,\omega,1,S} & \cdots & \eta_{1,\omega,k,\omega} & \cdots & \eta_{1,\omega,K,S} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \cdots \\
\eta_{1,S,1,1} & \cdots & \eta_{1,S,1,\omega} & \cdots & \eta_{1,S,1,S} & \cdots & \eta_{1,S,k,\omega} & \cdots & \eta_{1,S,K,S} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \cdots \\
\eta_{k,\omega,1,1} & \cdots & \eta_{k,\omega,1,\omega} & \cdots & \eta_{k,\omega,1,S} & \cdots & \eta_{k,\omega,k,\omega} & \cdots & \eta_{k,\omega,K,S} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \cdots \\
\eta_{K,S,1,1} & \cdots & \eta_{K,S,1,\omega} & \cdots & \eta_{K,S,1,S} & \cdots & \eta_{K,S,k,\omega} & \cdots & \eta_{K,S,K,S}
\end{bmatrix}
\end{array}
$$

**Figure 3.1**: $\eta$ matrix

new matrix, say, $\Lambda$, with generic element $\lambda_{k,\omega,g,\tau}$ by raising all diagonal entries of the $\eta$ matrix, leaving all remaining entries the same, so that there is some $M > 0$ for which $\sum_{(g,\tau)} \eta_{k,\omega,g,\tau} = \sum_{(g,\tau)} \eta_{g,\tau,k,\omega} = M\pi_\omega$.[10] Since $p^k_\omega \cdot (x^k_\omega - x^k_\omega) = 0$, and since the diagonal element shows up both in the column and and row, the resulting $\eta$ matrix satisfies (3.3) (with $\lambda$'s in place of $\eta$'s), and the first equality in system (3.4). Finally, the $\gamma$ terms are constructed by dividing each element of the matrix $\Lambda$ by $M$.

Rearranging inequalities (3.3) gives

$$
\begin{aligned}
\sum_\omega \sum_g \sum_\tau \gamma_{k,\omega,g,\tau} \frac{p^k_\omega}{\pi_\omega} \cdot (x^g_\tau - x^k_\omega) &= \sum_\omega p^k_\omega \cdot \Big(\sum_g \sum_\tau \frac{\gamma_{k,\omega,g,\tau} x^g_\tau}{\pi_\omega} - \sum_g \sum_\tau \frac{\gamma_{k,\omega,g,\tau} x^k_\omega}{\pi_\omega}\Big) \\
&= \sum_\omega p^k_\omega \cdot \Big(\sum_g \sum_\tau \frac{\gamma_{k,\omega,g,\tau} x^g_\tau}{\pi_\omega} - x^k_\omega\Big) \le 0
\end{aligned}
$$

with at least one strict inequality. The second equality follows from (3.4). This together with (3.4) establishes the equivalence of II and III, by taking $S^k_\omega(g,\tau) = \gamma_{k,\omega,g,\tau}$.

[10]One simple way of doing this is to pick $M$ large enough so that $\min_\omega \pi_\omega M > \max_\omega \sum_{(g,\tau)\neq(k,\omega)} \eta_{k,\omega,g,\tau}$.

**(IV $\Rightarrow$ I)**

That IV implies I is straightforward. Let $u : \mathbb{R}^N_+ \to \mathbb{R}$ be any concave, nondecreasing and locally non-satiated utility function. For lottery $l$, let $u \cdot l$ denote the expected utility of $l$, $\sum_{x \in l} l(x)u(x)$.

Suppose that $\mathcal{D}$ is risk-averse expected utility rationalizable by $u$, and suppose by means of contradiction that statement I is not true

For all $l'_k \in \Delta_s(\mathbb{R}^N_+)$, $l_{x^k} \succeq^R l'_k$ implies $u \cdots l_{x^k} \geq u \cdot l'_k$, and $l_{x^k} \succ^R l'_k$ implies $u \cdot l_{x^k} > u \cdot l'_k$. Since expected utility is linear in lottery mixtures, we have that $u \cdot (\sum_1^K \mu_k l_{x^k}) > u \cdot (\sum_1^K \mu_k l'_k)$, a contradiction to $\sum_1^K \mu_k l_{x^k} = \sum_1^K \mu_k l'_k$.

**(I $\Rightarrow$ II)**

We now show that I implies II. Suppose by means of contradiction that there is a solution to the system listed in II. We will show that this implies I is false. Let

$$y^k = \left( \frac{\sum_g \sum_\tau S^k_\omega(g,\tau) x^g_\tau}{\pi_\omega} \right)_{\omega \in \Omega}$$

By II, we have $p^k \cdot x^k \geq p^k \cdot y^k \; \forall k$ with $>$ for at least one $k$. By definition of $\succeq^C$, $l_{x^k} \succeq^C l_{y^k}$, with $\succ^C$ for at least one $k$.

Next, observe that $l_{y^k}$ places probability $\pi_\omega$ at $\frac{\sum_g \sum_\tau S^k_\omega(g,\tau) x^g_\tau}{\pi_\omega}$ for each $\omega$. Let $l'_k$ be the lottery that puts probability $\sum_\omega S^k_\omega(g,\tau)$ on $x^g_\tau$. Since $\sum_{g,\tau} S^k_\omega(g,\tau) = \pi_\omega$, $l'_k$ can be obtained from $l_{y^k}$ by spreading, for each $\omega$, the probability $\pi_\omega$ placed on $\frac{\sum_g \sum_\tau S^k_\omega(g,\tau) x^g_\tau}{\pi_\omega}$ to probabilities $S^k_\omega(g,\tau)$'s on $x^g_\tau$'s, $(g,\tau) \in \{1, \cdots, K\} \times \Omega$. Moreover, $\frac{\sum_g \sum_\tau S^k_\omega(g,\tau) x^g_\tau}{\pi_\omega}$ is a weighted average of the $x^g_\tau$'s by weights $S^k_\omega(g,\tau)$'s. So for each $\omega$ the spread described above is a mean-preserving spread in the sense of Rothschild and Stiglitz (1970), and $l'_k$ can be obtained from $l_{y^k}$ by a finite number of mean-preserving spread.

By definition of $\succeq^R$, we have obtained lotteries $l_{x^k}$ and $l'_k$ such that $l_{x^k} \succeq^R l'_k \; \forall k$, with $\succ^R$

for at least one $k$. In order to contradict I, it only remains now to find $\{\mu_k\}_{k=1}^K$ such that $\mu_k \geq 0$, $\sum_{\{k:l_k \succ^R l'_k\}} \mu_k > 0$ and $\sum_1^K \mu_k l_{x^k} = \sum_1^K \mu_k l'_k$. As it turns out, it suffices to take $\mu_k = \frac{1}{K}$ for each $k$:

The lottery $\sum_{k=1}^K \frac{1}{K} l'_k$ places probability $\frac{1}{K} \sum_k \sum_\omega S_\omega^k(g,\tau) = \frac{\pi_\tau}{K}$ on each $x_\tau^g$, $(g,\tau) \in \{1,\cdots,K\} \times \Omega$, while the lottery $\sum_{k=1}^K \frac{1}{K} l_{x^k}$, places $\frac{\pi_\tau}{K}$ on each $x_\tau^g$. So $\sum_{k=1}^K \frac{1}{K} l'_k = \sum_{k=1}^K \frac{1}{K} l_{x^k}$. This constitutes a contradiction to I (in particular, the contradiction comes in the form of a uniform distribution over the observations $1,\ldots,K$).

$\square$

## 3.3 Conclusion

We have developed a universal test for the risk-averse expected utility environment with many commodities. Of interest for future research would be an analogous test in the subjective expected utility context, following the work of Echenique and Saito (2015). The difficulty inherent in this approach rests in the fact that the inequalities in III of Theorem 1 are polynomial, rather than linear. While we have some conjectures on what might be an appropriate test, these are very speculative.

A final remark is in order. Observe that when $|\Omega| = 1$ (and hence $\pi_\omega = 1$ for $\omega$ for which $\Omega = \{\omega\}$), we are back to the environment of Afriat (1967). In such an environment, the function $S$ referenced in Theorem 1, condition II can be taken to be a function of $\{1,\ldots,K\}$ alone. And condition II in this case tells us that $\sum_k S_k(l) = \sum_k S_l(k) = 1$ for each $l$; in other words, viewing $S$ as a matrix, the matrix is *bistochastic*. Now, one of the contributions of Afriat (1967) is that condition II is necessary and sufficient for concave rationalization when the matrix $S$ is restricted to be a *permutation* matrix; that is, a matrix consisting solely of zeroes and ones. Of course, it is well-known that the permutation matrices are the extreme points of the set of bistochastic matrices (this is the celebrated theorem of Birkhoff (1946) and Neumann (1953)). A natural conjecture is that a similar statement may hold here; that it is enough to check the extreme points of the set of $S$ functions satisfying condition II of Theorem 1.

## 3.4 Acknowledgement

Chapter 3, in full, is a reprint of the material as it appears in Journal of Economic Theory, 2016. Chambers, Christopher P; Liu, Ce; Martinez, Seung-Keun. “A Test for Risk-Averse Expected Utility”. The dissertation author was a primary author of this paper. The copyright of this article is held by Elsevier.

# Bibliography


Abreu, Dilip, Prajit K Dutta, and Lones Smith. 1994. "The folk theorem for repeated games: a NEU condition". *Econometrica* 62 (4): 939–948.

Afriat, Sidney N. 1967. "The construction of utility functions from expenditure data". *International economic review* 8 (1): 67–77.

Akbarpour, Mohammad, Shengwu Li, and Shayan Oveis Gharan. 2017. "Thickness and information in dynamic matching markets".

Ali, Nageeb, and Ce Liu. 2018. "Conventions and Coalitions in Repeated Games".

Aliprantis, Charalambos D, and Kim Border. 2006. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer Science & Business Media.

Allen, Roy, and John Rehbeck. 2016. "Revealed preference analysis of characteristics in discrete choice". Working Paper.

Anderson, Ross, Itai Ashlagi, David Gamarnik, and Yash Kanoria. 2015. "A dynamic model of barter exchange". In *Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms*, 1925–1933. Society for Industrial and Applied Mathematics.

Azevedo, Eduardo M, and John William Hatfield. 2018. "Existence of equilibrium in large matching markets with complementarities".

Baccara, Mariagiovanna, SangMok Lee, and Leeat Yariv. 2015. "Optimal dynamic matching".

Bando, Keisuke. 2012. "Many-to-one matching markets with externalities among firms". *Journal of Mathematical Economics* 48 (1): 14–20.

Banerjee, Suryapratim, Hideo Konishi, and Tayfun Sönmez. 2001. "Core in a simple coalition formation game". *Social Choice and Welfare* 18 (1): 135–153.

Bernheim, B Douglas, and Sita N Slavov. 2009. "A solution concept for majority rule in dynamic settings". *The Review of Economic Studies* 76 (1): 33–62.

Bindman, Andrew B, Dennis Keane, and Nicole Lurie. 1990. "A public hospital closes: impact on patients' access to care and health status". *Jama* 264 (22): 2899–2904.

Birkhoff, Garrett. 1946. "Tres observaciones sobre el álgebra lineal". *Univ. Nac. Tucumán Rev. Ser. A* 5:147–151.

Blackwell, David. 1953. "Equivalent comparisons of experiments". *The annals of mathematical statistics* 24 (2): 265–272.

Brown, Donald J, and Caterina Calsamiglia. 2007. "The nonparametric approach to applied welfare analysis". *Economic Theory* 31 (1): 183–188.

Caplin, Andrew, and Mark Dean. 2015. "Revealed preference, rational inattention, and costly information acquisition". *The American Economic Review* 105 (7): 2183–2203.

Caplin, Andrew, and Daniel Martin. 2015. "A testable theory of imperfect perception". *The Economic Journal* 125 (582): 184–202.

Cerreia-Vioglio, Simone, Fabio Maccheroni, Massimo Marinacci, and Luigi Montrucchio. 2011a. "Complete monotone quasiconcave duality". *Mathematics of Operations Research* 36 (2): 321–339.

— . 2011b. "Uncertainty averse preferences". *Journal of Economic Theory* 146 (4): 1275–1330.

Chambers, Christopher P, and Federico Echenique. 2016. *Revealed Preference Theory*. Vol. 56. Cambridge University Press.

Chambers, Christopher P, Federico Echenique, and Kota Saito. 2016. "Testing theories of financial decision making". *Proceedings of the National Academy of Sciences* 113 (15): 4003–4008.

Chambers, Christopher P., Federico Echenique, and Eran Shmaya. 2014. "The axiomatic structure of empirical content". *American Economic Review* 104 (8): 2303–2319.

Chateauneuf, Alain, and José Heleno Faro. 2009. "Ambiguity through confidence functions". *Journal of Mathematical Economics* 45 (9): 535–558.

Che, Yeon-Koo, Jinwoo Kim, and Fuhito Kojima. 2017. "Stable Matching in Large Economies".

Cherchye, Laurens, Ian Crawford, Bram De Rock, and Frederic Vermeulen. 2016. "Gorman revisited: nonparametric conditions for exact linear aggregation". *SERIEs* 7 (2): 203–220.

Chwe, Michael Suk-Young. 1994. "Farsighted coalitional stability". *Journal of Economic theory* 63 (2): 299–325.

Corbae, Dean, Ted Temzelides, and Randall Wright. 2003. "Directed matching and monetary exchange". *Econometrica* 71 (3): 731–756.

Damiano, Ettore, and Ricky Lam. 2005. "Stability in dynamic matching markets". *Games and Economic Behavior* 52 (1): 34–53.

De Oliveira, Henrique, and Rohit Lamba. 2018. “Rationalizing dynamic choices”. Working paper.

Department of Justice. 2010. *Complaint: U.S. v. Adobe Systems, Inc, et al*. https://www.justice.gov/atr/case-document/file/483451/download.

Doval, Laura. 2018. *A theory of stability in dynamic matching markets*. Tech. rep. mimeo.

Du, Songzi, and Yair Livne. 2016. “Rigidity of Transfers and Unraveling in Matching Markets”.

Dutta, Bhaskar, and Jordi Massó. 1997. “Stability of matchings when individuals have preferences over colleagues”. *Journal of Economic Theory* 75 (2): 464–475.

Echenique, Federico, and Kota Saito. 2015. “Savage in the market”. *Econometrica* 83 (4): 1467–1495.

Echenique, Federico, and M Bumin Yenmez. 2005. “A solution to matching with preferences over colleagues”.

Epstein, Larry G. 2000. “Are probabilities used in markets?” *Journal of Economic Theory* 91 (1): 86–90.

Fishburn, Peter C. 1975. “Separation theorems and expected utilities”. *Journal of Economic Theory* 11 (1): 16–34.

Flanagan, Francis. 2015. “Contracts vs. preferences over colleagues in matching”. *International Journal of Game Theory* 44 (1): 209–223.

Fragiadakis, Daniel, and Peter Troyan. 2017. “Improving matching under hard distributional constraints”. *Theoretical Economics* 12 (2): 863–908.

Fudenberg, Drew, Ryota Iijima, and Tomasz Strzalecki. 2015. “Stochastic choice and revealed perturbed utility”. *Econometrica* 83 (6): 2371–2409.

Fudenberg, Drew, and Eric Maskin. 1986. “The folk theorem in repeated games with discounting or with incomplete information”. *Econometrica*: 533–554.

Gale, David, and Lloyd S Shapley. 1962. “College admissions and the stability of marriage”. *The American Mathematical Monthly* 69 (1): 9–15.

Gomes, Armando, and Philippe Jehiel. 2005. “Dynamic processes of social and economic interactions: On the persistence of inefficiencies”. *Journal of Political Economy* 113 (3): 626–667.

Gorman, William M. 1953. “Community preference fields”. *Econometrica: journal of the Econometric Society*: 63–80.

Goto, Masahiro, Fuhito Kojima, Ryoji Kurata, Akihisa Tamura, and Makoto Yokoo. 2017. "Designing matching mechanisms under general distributional constraints". *American Economic Journal: Microeconomics* 9 (2): 226–62.

Green, Richard C., and Sanjay Srivastava. 1986. "Expected utility maximization and demand behavior". *Journal of Economic Theory* 38 (2): 313–323.

Hadley, Jack, and K Nair. 1991. "The impact of rural hospital closure on Medicare beneficiaries' access to hospital care". *Washington, DC: Georgetown University Center for Health Policy Studies*.

Hatfield, John William, and Paul R Milgrom. 2005. "Matching with contracts". *American Economic Review* 95 (4): 913–935.

Houthakker, Hendrik S. 1950. "Revealed preference and the utility function". *Economica* 17 (66): 159–174.

Kadam, Sangram V, and Maciej H Kotowski. 2018a. "Multiperiod Matching". *International Economic Review* 59 (4): 1927–1947.

— . 2018b. *Time horizons, lattice structures, and welfare in multi-period matching markets*. Tech. rep.

Kamada, Yuichiro, and Fuhito Kojima. 2015. "Efficient matching under distributional constraints: Theory and applications". *American Economic Review* 105 (1): 67–99.

— . 2018. "Stability and strategy-proofness for matching with constraints: A necessary and sufficient condition". *Theoretical Economics* 13 (2): 761–793.

— . 2017. "Stability concepts in matching under distributional constraints". *Journal of Economic Theory* 168:107–142.

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian persuasion". *The American Economic Review* 101 (6): 2590–2615.

Kelso, Alexander S, and Vincent P Crawford. 1982. "Job matching, coalition formation, and gross substitutes". *Econometrica: Journal of the Econometric Society*: 1483–1504.

Kominers, Scott Duke. 2010. "Matching with preferences over colleagues solves classical matching". *Games and Economic Behavior* 68 (2): 773–780.

Konishi, Hideo, and Debraj Ray. 2003. "Coalition formation as a dynamic process". *Journal of Economic theory* 110 (1): 1–41.

Koopmans, Tjalling C, and Martin Beckmann. 1957. "Assignment problems and the location of economic activities". *Econometrica*: 53–76.

Kotowski, Maciej H. 2015. *A note on stability in one-to-one, multi-period matching markets*. Tech. rep.

Kubler, Felix, Larry Selden, and Xiao Wei. 2014. "Asset demand based tests of expected utility maximization". *American Economic Review* 104 (11): 3459–3480.

Kurino, Morimitsu. 2009. "Credibility, efficiency, and stability: A theory of dynamic matching markets". *Jena economic research papers, JENA* 7:41.

Leshno, Jacob D. 2017. "Dynamic matching in overloaded waiting lists".

Machina, Mark J. 1984. "Temporal risk and the nature of induced preferences". *Journal of economic theory* 33 (2): 199–231.

Machina, Mark J., and David Schmeidler. 1992. "A more robust definition of subjective probability". *Econometrica* 60 (4): 745–780.

Mangasarian, Olvi L. 1994. *Nonlinear Programming*. Classics in Applied Mathematics. Society for Industrial / Applied Mathematics. ISBN: 9780898713411.

Neumann, John von. 1953. "A certain zero-sum two-person game equivalent to the optimal assignment problem". *Contributions to the Theory of Games* 2:5–12.

Newton, Jonathan, and Ryoji Sawa. 2015. "A one-shot deviation principle for stability in matching problems". *Journal of Economic Theory* 157:1–27.

Niederle, Muriel, and Leeat Yariv. 2009. *Decentralized matching with aligned preferences*. Tech. rep. National Bureau of Economic Research.

Polisson, Matthew, John Quah, and Ludovic Renou. 2015. "Revealed preference tests over risk and uncertainty". Working paper.

Prada, Juan David. 2011. "A note on concavity, homogeneity and non-Increasing returns to scale"". *Economics Bulletin* 31 (1): 100–105.

Pycia, Marek. 2012. "Stability and preference alignment in matching and coalition formation". *Econometrica* 80 (1): 323–362.

Pycia, Marek, and M Bumin Yenmez. 2017. "Matching with Externalities".

Ray, Debraj. 2007. *A Game-Theoretic Perspective on Coalition Formation*. New York, NY: Oxford University Press.

Ray, Debraj, and Rajiv Vohra. 1999. "A Theory of Endogenous Coalition Structures". *Games and Economic Behavior* 26:286336.

— . 1997. "Equilibrium binding agreements". *Journal of Economic theory* 73 (1): 30–78.

Richter, Marcel K. 1966. “Revealed preference theory”. *Econometrica* 34 (3): 635–645.

Rochet, Jean-Charles. 1987. “A necessary and sufficient condition for rationalizability in a quasi-linear context”. *Journal of mathematical Economics* 16 (2): 191–200.

Rockafellar, Ralph. 1966. “Characterization of the subdifferentials of convex functions”. *Pacific Journal of Mathematics* 17 (3): 497–510.

Rockafellar, Tyrrell. 1997. *Convex Analysis*. 28. Princeton University Press.

Rosenbach, Margo L, and Debra A Dayhoff. 1995. “Access to care in rural America: impact of hospital closures”. *Health Care Financing Review* 17 (1): 15.

Roth, Alvin E. 1991. “A natural experiment in the organization of entry-level labor markets: regional markets for new physicians and surgeons in the United Kingdom”. *The American economic review*: 415–440.

— . 1986. “On the allocation of residents to rural hospitals: a general property of two-sided matching markets”. *Econometrica: Journal of the Econometric Society*: 425–427.

Roth, Alvin E, and Marilda Sotomayor. 1992. “Two-sided matching”. *Handbook of game theory with economic applications* 1:485–541.

Rothschild, Michael, and Joseph E. Stiglitz. 1970. “Increasing risk: I. A definition”. *Journal of Economic theory* 2 (3): 225–243.

Roy, René. 1947. “La distribution du revenu entre les divers biens”. *Econometrica*: 205–225.

Sasaki, Hiroo, and Manabu Toda. 1996. “Two-sided matching problems with externalities”. *Journal of Economic Theory* 70 (1): 93–108.

Ünver, M Utku. 2010. “Dynamic kidney exchange”. *The Review of Economic Studies* 77 (1): 372–414.

U.S. Government Accountability Office. 1991. “Federal Efforts Should Target Areas Where Closures Would Threaten Access to Care”. *Washington, DC: US General Accounting Office*: 91–41.

Varian, Hal R. 1983a. “Non-parametric tests of consumer behaviour”. *The review of economic studies* 50 (1): 99–110.

Varian, Hal R. 1983b. “Nonparametric tests of models of investor behavior”. *Journal of Financial and Quantitative Analysis* 18 (3): 269–278.

Varian, Hal R. 1982. “The nonparametric approach to demand analysis”. *Econometrica*: 945–973.

— . 1984. “The nonparametric approach to production analysis”. *Econometrica*: 579–597.

# Appendix A

# Proofs in Chapter 1

## A.1 Proofs in Section 1.2

**Lemma 7.** *Let $\mu$ be a self-enforcing matching process. For any institution $I_k$, the set of recommended stage-game payoffs from $\mu$ across all possible ex post histories, $\{u_k(\mu(h)) : h \in \mathcal{H}\}$, is bounded.*

*Proof.* Fix an institution $I_k$ and let $\mu = (\psi, \{\xi_l\}_{l=1}^{K})$ be a self-enforcing matching process.

I first show the set $\{u_k(\mu(h)) : h \in \mathcal{H}\}$ is bounded from above. Note that by the definition of self-enforcing matching process, at every ex post history $h \in \mathcal{H}$, the recommended matching $\mu(h)$ must satisfy $v_j(\mu(h)) \geq 0$ for all $a_j \in \mathcal{A}$. In particular, $v_j(I_k, \psi(I_k|h)) + \xi_{kj}(h) \geq 0$ for all $a_j \in \psi(I_k|h)$. This implies that, at every ex post history $h \in \mathcal{H}$,

$$\begin{aligned} u_k(\mu(h)) &= u_k(\psi(I_k|h)) - \sum_{a_j \in \psi(I_k|h)} \xi_{kj}(h) \\ &\leq u_k(\psi(I_k|h)) + \sum_{a_j \in \psi(I_k|h)} v_j(I_k, \psi(I_k|h)) \\ &\leq \max_{B \subseteq \mathcal{A}} u_k(B) + \sum_{a_j \in B} v_j(I_k, B) \equiv \widehat{b}_k \end{aligned}$$

So $\{u_k(\mu(h)) : h \in \mathcal{H}\}$ is bounded above by $\widehat{b}_k$.

In order to show $\{u_k(\mu(h)) : h \in \mathcal{H}\}$ is bounded from below. I will first establish that the set of continuation values for $I_k$ across all ex ante histories is bounded from above, while the set of continuation values following ex post histories is bounded from below, which then delivers the desired claim.

Fix any ex ante history $\widehat{h} \in \overline{\mathcal{H}}$, the continuation matching process at $\mu|_{\widehat{h}}$ satisfies

$$u_k(\mu|_{\widehat{h}}(h)) = u_k(\mu(\widehat{h}, h)) \leq \widehat{b}_k$$

for all ex post histories $h \in \mathcal{H}$. Since the above inequality holds for all $h$, it must hold for every stage-game matching along every possible outcome path generated by $\mu|_{\widehat{h}}$. So the continuation value $u_k(\mu|_{\widehat{h}})$, as an expectation over the discounted sum of these stage-game payoffs, must satisfy $u_k(\mu|_{\widehat{h}}) \leq \widehat{b}_k$. Since this holds uniformly for every $\widehat{h} \in \overline{\mathcal{H}}$, the set $\{u_k(\mu|_{\overline{h}}) : \overline{h} \in \overline{\mathcal{H}}\}$ is bounded above.

Next I show the set $\{u_k(\mu|_h) : h \in \mathcal{H}\}$ is bounded from below. In particular, I will show that for all $h \in \mathcal{H}$,

$$u_k(\mu|_h) \geq \min_{\{A \subseteq \mathcal{A} : |A| \leq \sum_{l=1}^{K} q_l\}} \max_{B \subseteq \mathcal{A} \setminus A} u_k(B) + \sum_{a_j \in B} v_j(I_k, B) \equiv \underline{b}_k$$

Suppose by contradiction that $u_k(\mu|_{\widetilde{h}}) < \underline{b}_k$ for some $\widetilde{h} \in \mathcal{H}$. We will show that $I_k$ has a feasible profitable deviation plan from $\mu|_{\widetilde{h}}$, which is a contradiction to $\mu$ being a self-enforcing matching process.

Consider the following deviation plan $(d'_k, \xi'_k)$ from $\mu|_{\widetilde{h}} = (\widetilde{\Psi}, \{\widetilde{\xi}_l\}_{l=1}^{K})$ : for any $h \in \mathcal{H}$,

$$d'_k(h) = \underset{B \subseteq \mathcal{A} \setminus (\cup_{l=1}^{K} \widetilde{\Psi}(I_l|h))}{\arg\max} u_k(B) + \sum_{a_j \in B} v_j(I_k, B).$$

If $d'_k(h) \neq \emptyset$, then

$$\xi'_{kj}(h) = \begin{cases} -v_j(I_k, d'_k(h)) + \frac{1}{2|d'_k(h)|}[\underline{b}_k - u_k(\mu|_{\widetilde{h}})] & \text{if } a_j \in d'_k(h); \\ 0 & \text{otherwise.} \end{cases}$$

If $d'_k(h) = \emptyset$, define

$$\xi'_{kj}(h) = 0 \ \ \forall a_j \in \mathcal{A}.$$

Note that by construction,

$$\begin{aligned} & u_k(d'_k(h)) + \sum_{a_j \in d'_k(h)} v_j(I_k, d'_k(h)) \\ = & \max_{B \subseteq \mathcal{A} \setminus (\cup_{l=1}^{K} \widetilde{\psi}(I_l|h))} u_k(B) + \sum_{a_j \in B} v_j(I_k, B) \\ \geq & \min_{\{A \subseteq \mathcal{A} : |A| \leq \sum_{l=1}^{K} q_l\}} \max_{B \subseteq \mathcal{A} \setminus A} u_k(B) + \sum_{a_j \in B} v_j(I_k, B) = \underline{b}_k. \end{aligned} \tag{A.1}$$

The inequality above follows from $\left|(\cup_{l=1}^{r} \widetilde{\psi}(I_l|h))\right| \leq \sum_{l=1}^{K} q_l$.

We first verify that the deviation plan $(d'_k, \xi'_k)$ is feasible. By construction, at every ex post history $h \in \mathcal{H}$, $d'_k(h) \subseteq \mathcal{A} \setminus (\cup_{l=1}^{K} \widetilde{\psi}(I_l|h))$. Since every agent in $a_j \in \mathcal{A} \setminus \cup_{l=1}^{K} \widetilde{\psi}(I_l|h))$ is unmatched, we have $v_j(\mu|_{\widetilde{h}}(h)) = 0$ for all $a_j \in \mathcal{A} \setminus \cup_{l=1}^{K} \widetilde{\psi}(I_l|h))$. So $v_j(\mu_{\widetilde{h}}(h)) = 0$ for all $a_j \in d'_k(h)$.

Meanwhile, at every ex post history $h$ and for all $a_j \in d'_k(h)$,

$$\begin{aligned} & v_j(I_k, d'_k(h)) + \xi'_{kj}(h) \\ = & v_j(I_k, d'_k(h)) - v_j(I_k, d'_k(h)) + \frac{1}{2\left|d'_k(h)\right|}[\underline{b}_k - u_k(\mu|_{\widetilde{h}})] \\ = & \frac{1}{2\left|d'_k(h)\right|}[\underline{b}_k - u_k(\mu|_{\widetilde{h}})] > 0 = v_j(\mu|_{\widetilde{h}}(h)) \end{aligned}$$

So at every possible ex post history $h$, every agent in $d'_k(h)$ finds himself strictly better off by joining the deviation, which ensures the feasibility of the deviation $(d'_k, \xi'_k)$.

To see that $(d'_k, \xi'_k)$ is profitable, observe that at every ex post history $h$, institution $I_k$'s stage-game payoff from the manipulated static matching $[\mu|_{\widetilde{h}}(h), (d'_k(h), \xi'_k(h))]$ is

$$\begin{aligned}
& u_k(d'_k(h)) - \sum_{a_j \in d'_k(h)} \xi'_{kj}(h) \\
= {} & u_k(d'_k(h)) + \sum_{a_j \in d'_k(h)} v_j(I_k, d'_k(h)) - \frac{1}{2}[\underline{b}_k - u_k(\mu|_{\widetilde{h}})] \\
\geq {} & \underline{b}_k - \frac{1}{2}[\underline{b}_k - u_k(\mu|_{\widetilde{h}})] \\
= {} & \frac{1}{2}\underline{b}_k + \frac{1}{2}u_k(\mu|_{\widetilde{h}}) > \ u_k(\mu|_{\widetilde{h}}).
\end{aligned}$$

The second inequality above follows from inequality (A.1). Since this is true for every ex post history $h$, it must hold along every possible outcome paths generated by the manipulated matching $[\mu|_{\widetilde{h}}, (d'_k, \xi'_k)]$. Therefore, $I_k$'s total expected discounted payoff from the deviation plan satisfies

$$u_k([\mu|_{\widetilde{h}}, (d'_k, \xi'_k)]) = \frac{1}{2}\underline{b}_k + \frac{1}{2}u_k(\mu|_{\widetilde{h}}) > u_k(\mu|_{\widetilde{h}}).$$

The deviation plan $(d'_k, \xi'_k)$ is both feasible and profitable for institution $I_k$, which is a contradiction to the self-enforcement of $\mu$. So $u_k(\mu|_h) \geq \underline{b}_k$ for all $h \in \mathcal{H}$. The set $\{u_k(\mu|_h) : h \in \mathcal{H}\}$ is bounded from below.

Now, at every ex post history $\widetilde{h} \in \mathcal{H}$, we have

$$u_k(\mu|_{\widetilde{h}}) = (1-\delta)u_k(\mu(\widetilde{h})) + \delta u_k(\mu|_{\widetilde{h}, \mu(\widetilde{h})}),$$

or

$$u_k(\mu(\widetilde{h})) = \frac{u_k(\mu|_{\widetilde{h}}) - \delta u_k(\mu|_{\widetilde{h}, \mu(\widetilde{h})})}{1-\delta}.$$

$u_k(\mu|_{\widetilde{h}})$ is an element in $\{u_k(\mu|_h) : h \in \mathcal{H}\}$, which is bounded from below by $\underline{b}_k$; $u_k(\mu|_{\widetilde{h}, \mu(\widetilde{h})})$ is an element in $\{u_k(\mu|_{\overline{h}}) : \overline{h} \in \overline{\mathcal{H}}\}$, which is bounded from above by $\widehat{b}_k$. $\{u_k(\mu(h)) : h \in \mathcal{H}\}$ is therefore bounded uniformly from below. $\square$

**Lemma 8.** *Let $\mu$ be a matching process that satisfies $v_j(\mu(h)) \geq 0$ for all $a_j \in \mathcal{A}$ and all $h \in \mathcal{H}$. If $I_k$ has a profitable, feasible deviation plan from $\mu$, then $I_k$ must have one such plan $(d'_k, \xi'_k)$ that has bounded per-period payoffs, i.e. the set $\left\{u^t_k\big([\mu,(d'_k,\xi'_k)]\big) : t \geq 0\right\}$ is bounded.*

*Proof.* Let $(d'_k, \xi'_k)$ be a feasible, profitable deviation plan from $\mu$ for institution $I_k$. First we show that $\left\{u^t_k\big([\mu,(d'_k,\xi'_k)]\big) : t \geq 0\right\}$ is bounded from above.

By the feasibility of the deviation plan $(d'_k, \xi'_k)$, we have, for all $h \in \mathcal{H}$,

$$v_j(I_k, d'_k(h)) + \xi'_{kj}(h) \geq v_j(\mu(h)) \geq 0 \text{ for all } a_j \in d'_k(h) \cap \psi(I_k|h)$$

and

$$v_j(I_k, d'_k(h)) + \xi'_{kj}(h) > v_j(\mu(h)) \geq 0 \text{ for all } a_j \in d'_k(h) \backslash \psi(I_k|h).$$

So

$$v_j(I_k, d'_k(h)) + \xi'_{kj}(h) \geq 0 \text{ for all } a_j \in d'_k(h).$$

for all $h \in \mathcal{H}$. This implies

$$\begin{aligned} u_k\big([\mu,(d'_k,\xi'_k)](h)\big) &= u_k(d'_k(h)) - \sum_{a_j \in d'_k(h)} \xi'_{kj}(h) \\ &\leq u_k(d'_k(h)) + \sum_{a_j \in d'_k(h)} v_j(I_k, d'_k(h)) \\ &\leq \max_{B \subseteq \mathcal{A}} u_k(B) + \sum_{a_j \in B} v_j(I_k, B) \equiv \widehat{b}_k \end{aligned}$$

for all $h \in \mathcal{H}$, so

$$u^t_k([\mu,(d'_k,\xi'_k)]) = \mathbb{E}^{\mu}_{\mathcal{H}_t}\Big[u_k\big([\mu,(d'_k,\xi'_k)](h)\big)\Big] \leq \widehat{b}_k \quad \text{for all } t.$$

The set $\{u_k^t([\mu,(d_k',\xi_k')]) : t \geq 0\}$ is bounded from above by $\widehat{b}_k$.

If $\{u_k^t([\mu,(d_k',\xi_k')]) : t \geq 0\}$ is also bounded from below, there is nothing left to prove. If, however, $\{u_k^t([\mu,(d_k',\xi_k')]) : t \geq 0\}$ is not bounded from below, I will show that it is possible to construct another feasible, profitable deviation plan $(d_k'',\xi_k'')$ such that $\{u_k^t([\mu,(d_k'',\xi_k'')]) : t \geq 0\}$ is bounded from below.

Suppose $\{u_k^t([\mu,(d_k',\xi_k')]) : t \geq 0\}$ is not bounded from below, then the set

$$\left\{t \geq 0 : u_k^t([\mu,(d_k',\xi_k')]) < -\frac{\delta\widehat{b}_k}{1-\delta}\right\}$$

must be nonempty. Set

$$\bar{t} = \min\left\{t \geq 0 : u_k^t([\mu,(d_k',\xi_k')]) < -\frac{\delta\widehat{b}_k}{1-\delta}\right\}.$$

By construction,

$$u_k^t([\mu,(d_k',\xi_k')]) \geq -\frac{\delta\widehat{b}_k}{1-\delta}$$

for all $0 \leq t < \bar{t}$, and

$$(1-\delta)u_k^{\bar{t}}([\mu,(d_k',\xi_k')]) + \delta\widehat{b}_k < 0, \tag{A.2}$$

Inequality (A.2) implies that if institution $I_k$ follows the stage-game matchings prescribed by $[\mu,(d_k',\xi_k')]$ in period $\bar{t}$, even if $I_k$ obtains its highest possible continuation payoff $\widehat{b}_k$ from period $(\bar{t}+1)$ onwards, it would still be better off to instead remain autarkic from period $\bar{t}$. Therefore, $I_k$ must have a feasible and even more profitable deviation plan from $\mu$ by enforcing autarky from period $\bar{t}$. Formally, define the deviation plan $(d_k'',\xi_k'')$ by

$$d_k''(h) = \begin{cases} d_k'(h) & \text{if } h \in \cup_{t=0}^{\bar{t}-1}\mathcal{H}_t \\ \emptyset & \text{if } h \in \cup_{t=\bar{t}}^{\infty}\mathcal{H}_t \end{cases}$$

and

$$\xi_k''(h) = \begin{cases} \xi_k'(h) & \text{if } h \in \cup_{t=0}^{\bar{t}-1}\mathcal{H}_t \\ 0 & \text{if } h \in \cup_{t=\bar{t}}^{\infty}\mathcal{H}_t \end{cases}$$

By construction, $(d_k'', \xi_k'')$ is feasible. Moreover,

$$\begin{aligned}
u_k([\mu,(d_k'',\xi_k'')]) &= (1-\delta)\Big[\sum_{t=0}^{\bar{t}-1}\delta^t u_k^t\big([\mu,(d_k'',\xi_k'')]\big) + \delta^{\bar{t}}\cdot 0\Big] \\
&> (1-\delta)\Big[\sum_{t=0}^{\bar{t}-1}\delta^t u_k^t\big([\mu,(d_k'',\xi_k'')]\big)\Big] + \delta^{\bar{t}}\Big[(1-\delta)u_k^{\bar{t}}\big([\mu,(d_k',\xi_k')]\big) + \delta\widehat{b}_k\Big] \\
&= (1-\delta)\Big[\sum_{t=0}^{\bar{t}-1}\delta^t u_k^t\big([\mu,(d_k',\xi_k')]\big)\Big] + \delta^{\bar{t}}\Big[(1-\delta)u_k^{\bar{t}}\big([\mu,(d_k',\xi_k')]\big) + \delta\widehat{b}_k\Big] \\
&= (1-\delta)\Big[\sum_{t=0}^{\bar{t}-1}\delta^t u_k^t\big([\mu,(d_k',\xi_k')]\big) + \delta^{\bar{t}} u_k^{\bar{t}}\big([\mu,(d_k',\xi_k')]\big) + \delta^{\bar{t}+1}\frac{\widehat{b}_k}{1-\delta}\Big] \\
&= (1-\delta)\Big[\sum_{t=0}^{\bar{t}}\delta^t u_k^t\big([\mu,(d_k',\xi_k')]\big) + \sum_{t=\bar{t}+1}^{\infty}\delta^t\widehat{b}_k\Big] \\
&\geq (1-\delta)\Big[\sum_{t=0}^{\infty}\delta^t u_k^t\big([\mu,(d_k',\xi_k')]\big)\Big] = u_k([\mu,(d_k',\xi_k')])
\end{aligned}$$

The second line above follows from inequality (A.2), the third from the construction of $(d_k'', \xi_k'')$, the last from the fact that $\big\{u_k^t\big([\mu,(d_k',\xi_k')]\big) : t \geq 0\big\}$ is bounded above by $\widehat{b}_k$. Since $(d_k', \xi_k')$ is profitable, it follows that $u_k([\mu,(d_k'',\xi_k'')]) > u_k([\mu,(d_k',\xi_k')]) > u_k(\mu)$, so $(d_k'', \xi_k'')$ is both feasible and profitable.

Lastly, by construction, $\big\{u_k^t\big([\mu,(d_k'',\xi_k'')]\big) : t \geq 0\big\}$ is bounded from below by $\min\{-\frac{\delta\widehat{b}_k}{1-\delta}, 0\}$. This completes the proof. □

**Lemma 9.** *Automaton representation*

## A.2 Proofs in Section 1.3

**Lemma 10.** *Let $U^* = \left\{\mathbb{E}_\lambda[u(\phi)] : \lambda \in \Lambda^*\right\}$. For every vector $u \in U^*$, there exist vectors $\{u^k : k \in \kappa(\mathcal{I}\backslash\mathcal{T})\} \subseteq U^*$ such that*

$$u_k^k < u_k$$

*for all $k \in \kappa(\mathcal{I}\backslash\mathcal{T})$, and*

$$u_k^k < u_k^{k'}$$

*for all $k \neq k' \in \kappa(\mathcal{I}\backslash\mathcal{T})$.*

*Proof.* The proof relies on similar techniques as in Abreu, Dutta, and Smith (1994). Whenever possible, I will cite intermediate results in Abreu, Dutta, and Smith (1994) without reproducing their proofs. Compared to Abreu, Dutta, and Smith (1994), the main difference is that instead of imposing NEU, it is shown that a matching environment always satisfies NEU.

I will first establish that the set $\Phi_\mathcal{T}$ satisfies what Abreu, Dutta, and Smith (1994) called the non-equivalent utilities (NEU) condition for institutions in $\mathcal{I}\backslash\mathcal{T}$. For every $k \in \kappa(\mathcal{T})$, there exists $a_{j(k)} \in \mathcal{A}\backslash\mathcal{T}$ such that $v_{j(k)}(I_k) > 0$. Define the assignment $\phi^0$ by

$$\phi^0(I_l) = \begin{cases} \widehat{A}_l & \text{if } I_l \in \mathcal{T} \\ \emptyset & \text{otherwise} \end{cases}$$

For each $k \in \kappa(\mathcal{I}\backslash\mathcal{T})$, define the assignment $\phi^k$ by

$$\phi^k(I_l) = \begin{cases} \widehat{A}_l & \text{if } I_l \in \mathcal{T} \\ \{a_{j(k)}\} & \text{if } I_l = I_k \\ \emptyset & \text{otherwise} \end{cases}$$

Clearly $\phi^0 \in \Phi_\mathcal{T}$. Since $v_{j(k)}(I_k) > 0$ for all $k \in \kappa(\mathcal{I}\backslash\mathcal{T})$, it follows that $\phi^k \in \Phi_\mathcal{T}$ for all $k \in \kappa(\mathcal{I}\backslash\mathcal{T})$

as well.

There are at least two elements in $\kappa(I \backslash \mathcal{T})$. For each pair of $k \neq k' \in \kappa(I \backslash \mathcal{T})$, $u_k(\phi^0) = u_k(\phi^{k'}) = 0$, while by strictness of institution preferences, $u_{k'}(\phi^0) = 0 \neq u_{k'}(\{a_{j(k')}\}) = u_{k'}(\phi^{k'})$. Therefore, for each pair $k \neq k' \in \kappa(I \backslash \mathcal{T})$, there do not exist scalars $\alpha > 0$ and $\beta$ such that $u_k(\phi) = \alpha u_{k'}(\phi) + \beta$ for all $\phi \in \Phi_{\mathcal{T}}$. This verifies the NEU condition in Abreu, Dutta, and Smith (1994) for indices in $\kappa(I \backslash \mathcal{T})$.

Since $\text{co}(U_{\mathcal{T}}) = \text{co}(\{u(\phi) : \phi \in \Phi_{\mathcal{T}}\})$, Lemma 1 and Lemma 2 in Abreu, Dutta, and Smith (1994) then ensure the existence of vectors $\{\widehat{u}^k : k \in \kappa(I \backslash \mathcal{T})\} \subseteq \text{co}(U_{\mathcal{T}})$ that satisfy $\widehat{u}^k_k < \widehat{u}^{k'}_k$ for all $k \neq k' \in \kappa(I \backslash \mathcal{T})$. Since $\{u(\phi) : \phi \in \Phi_{\mathcal{T}}\}$ is a finite set, for each $k \in \kappa(I \backslash \mathcal{T})$, there exists an element $\overline{u}^k$ in $\{u(\phi) : \phi \in \Phi_{\mathcal{T}}\}$ that minimizes $I_k$'s payoff.

For an arbitrary vector $u \in U^*$, define

$$u^k = \varepsilon(1-\eta)\overline{u}^k + \eta\varepsilon\widehat{u}^k + (1-\varepsilon)u$$

for each $k \in \kappa(I \backslash \mathcal{T})$. Observe that if $\varepsilon\eta > 0$, then $u^k_k < u^{k'}_k$; for all $0 < \varepsilon < 1$ and $0 < \eta < 1$, $u^k \in \text{co}(U_{\mathcal{T}})$; for small enough $\varepsilon > 0$, $u^k_k > \underline{u}^{\mathcal{T}}_k$; and finally, for small enough $\eta > 0$, it must be true that $u^k_k < u_k$. Therefore, there must exist $\{u^k : k \in \kappa(I \backslash \mathcal{T})\} \subseteq U^*$ such that

$$u^k_k < u_k$$

for all $k \in \kappa(I \backslash \mathcal{T})$, and

$$u^k_k < u^{k'}_k$$

for all $k \neq k' \in \kappa(I \backslash \mathcal{T})$. □

**Lemma 11.** *For each institution $I_k \in I \backslash \mathcal{T}$, there exists $\underline{\phi}_k \in \Phi_{\mathcal{T}}$ such that $u_k(\underline{\phi}_k) = \underline{u}^{\mathcal{T}}_k$, and*

$$\max_{B \in D^{\mathcal{A}}_k(\underline{\phi}_k)} u_k(B) = \underline{u}^{\mathcal{T}}_k$$

**Lemma 12.** *For each institution $I_k \in \mathcal{T}$ and every $\phi \in \Phi_{\mathcal{T}}$,*

$$\max_{B \in D_k^{\mathcal{A}}(\phi)} u_k(B) = u_k(\phi)$$

**Lemma 13.** *Let $\Phi' \subseteq \Phi$ be a subset of static assignments, and $\psi : \mathcal{H}^F \to \Phi'$ be a self-enforcing assignment process. Fix an institution $I_k \in I$. If for every $\phi \in \Phi'$, there exists a subset of agents $A^\phi \in D_k^{\mathcal{A}}(\phi)$ that satisfies $u_k(A^\phi) \geq \underline{u}_k$, then*

$$u_k(\psi) \geq \underline{u}_k$$

*Proof.* Suppose by contradiction that $\psi$ is a self-enforcing assignment process, but $u_k(\psi) < \underline{u}_k$. I will show that institution $I_k$ has a feasible profitable deviation plan from $\psi$, so $\psi$ must not be self-enforcing. Consider the following deviation plan $d'_k$ from $\psi$: for every ex post history $h \in \mathcal{H}^F$, define

$$d'_k(h) = A^{\psi(h)}.$$

By assumption, $\psi(h) \in \Phi'$ for all $h \in \mathcal{H}^F$, so $d'_k$ is well-defined and feasible. To see that $d'_k$ is profitable, observe that at every ex post history $h$, institution $I_k$'s stage-game payoff from the manipulated static assignment $[\psi(h), (I_k, d'_k(h))]$ is $u_k(A^{\psi(h)}) \geq \underline{u}_k > u_k(\psi)$. Since this is true for every ex post history $h$, $I_k$'s total discounted payoff from the deviation plan satisfies

$$u_k([\psi, (I_k, d'_k)]) > u_k(\psi).$$

The deviation plan $d'_k$ is both feasible and profitable for institution $I_k$, which is a contradiction to the self-enforcement of $\psi$. So $u_k(\psi) \geq \underline{u}_k$.

□

**Proof of Theorem 2.** Let $\{\underline{\phi}_k : k \in \kappa(I \backslash \mathcal{T})\}$ be the static assignments as constructed in

Lemma 11. Fix $\lambda^0 \in \Lambda^*$ and define $u^0 = \mathbb{E}_{\lambda^0}[u(\phi)]$. By Lemma 10, there exist vectors $\{u^k : k \in \kappa(I\backslash\mathcal{T})\} \subseteq U^*$, such that

$$u_k^k < u_k^0$$

for every $k \in \kappa(I\backslash\mathcal{T})$, and

$$u_k^k < u_k^{k'}$$

for all $k, k' \in \kappa(I\backslash\mathcal{T})$, $k \neq k'$. For each $k \in \kappa(I\backslash\mathcal{T})$, let $\lambda^k \in \Lambda^*$ be the distribution over $\Phi_{\mathcal{T}}$ that give rise to the payoff vectors $u^k$.

By Lemma 9, it suffices to consider the matching process represented by the automaton $(\Theta, p^0, f, \gamma)$, where

- $\Theta = \left\{\theta(e,\phi) : e \in \kappa(I\backslash\mathcal{T}) \cup \{0\}, \phi \in \Phi_{\mathcal{T}}\right\} \cup \left\{\underline{\theta}(k,t) : k \in \kappa(I\backslash\mathcal{T}),\ 0 \leq t < L\right\}$ is the set of all possible states;

- $p^0$ is the initial distribution over states, which satisfies $p^0(\theta(0,\phi)) = \lambda^0(\phi)$ for all $\phi \in \Phi_{\mathcal{T}}$;

- $f : \Theta \to \Phi$ is the output function, where $f(\theta(e,\phi)) = \phi$ and $f(\underline{\theta}(k,t)) = \underline{\phi}_k$;

- $\gamma : \Theta \times \Phi \to \Delta(\Theta)$ is the transition function. For states $\{\underline{\theta}(k,t) | 0 \leq t < L-1\}$, $\gamma$ is defined as

$$\gamma\big(\underline{\theta}(k,t), \phi'\big) = \begin{cases} \underline{\theta}(k',0) & \text{if } \phi' \neq \underline{\phi}_k;\ \phi' = [\underline{\phi}_{k'}, (I_{k'}, B)] \text{ for some } k' \in \kappa(I\backslash\mathcal{T}) \\ & \text{and } B \in D_{k'}^{\mathcal{A}}(\underline{\phi}_{k'}) \\ \underline{\theta}(k,t+1) & \text{otherwise} \end{cases}$$

For states $\underline{\theta}(k, L-1)$, the transition is defined as

$$\gamma\big(\underline{\theta}(k,L-1),\phi'\big) = \begin{cases} \underline{\theta}(k',0) & \text{if } \phi' \neq \underline{\phi}_k;\ \phi' = [\underline{\phi}_{k'},(I_{k'},B)] \text{ for some } k' \in \kappa(I\backslash\mathcal{T}) \\ & \text{and } B \in D^{\mathcal{A}}_{k'}(\underline{\phi}_{k'}) \\ p^k & \text{otherwise} \end{cases}$$

where $p^k$ is the distribution over states that satisfies $p^k(\theta(k,\phi)) = \lambda^k(\phi)$ for all $k \in \kappa(I\backslash\mathcal{T})$ and $\phi \in \Phi_{\mathcal{T}}$.

For states $\theta(e,\phi)$, the transition is

$$\gamma\big(\theta(e,\phi),\phi'\big) = \begin{cases} \underline{\theta}(k',0) & \text{if } \phi' \neq \phi;\ \phi' = [\phi,(I_{k'},B)] \text{ for some } k' \in \kappa(I\backslash\mathcal{T}) \text{ and } B \in D^{\mathcal{A}}_{k'}(\phi) \\ p^e & \text{otherwise} \end{cases}$$

Note that owing to the identifiability of deviating institution, for any $\theta \in \Theta$ and assignment $\phi' \neq f(\theta)$ which can result from an institution's deviation, we can uniquely identify the institution, so the transition above is well-defined. Any $\phi' \neq f(\theta)$ that cannot possibly result from a deviation by an institution is ignored by the transition.

The assignment process represented by the above automaton randomizes over $\Phi_{\mathcal{T}}$ according to $\lambda^0$ in every period. It remains to check that it is self-enforcing, or equivalently,

1. every agent's payoff is greater than or equal to 0 in all automaton states,
2. no institution has any profitable one-shot deviation in any automaton state,
3. for every institution, the stage-game payoffs across all automaton states are bounded.

Point 3 above follows because the set of states $\Theta$ is finite. In every state $\theta \in \Theta$, the recommended assignment $f(\theta) \in \Phi_{\mathcal{T}}$ is individually rational for all agents, so point 1 follows as well. It remains to verify point 2.

For every state $\theta$, I use $U(\theta) = (U_1(\theta), \ldots, U_K(\theta))$ to denote the discounted expected payoff profile for the institutions in state $\theta$. By construction,

$$U(\theta(e,\phi)) = (1-\delta)u(\phi) + \delta u^e, \text{ for all } e \in \kappa(I \backslash \mathcal{T}) \cup \{0\},$$

and

$$U(\underline{\theta}(k,t)) = (1-\delta^{L-t})u(\underline{\phi}_k) + \delta^{L-t}u^k, \text{ for all } k \in \kappa(I \backslash \mathcal{T}), 0 \leq t \leq L-1.$$

In addition, since $u_l(\phi) = u_l(\widehat{A}_l)$ for every $I_l \in \mathcal{T}$ and all $\phi \in \Phi_{\mathcal{T}}$, the above equalities simplify to

$$U_l(\theta(e,\phi)) = U_l(\underline{\theta}(k,t)) = u_l(\widehat{A}_l) \tag{A.3}$$

for every $I_l \in \mathcal{T}$.

I now verify no institution has profitable one shot deviations in any automaton states.

For states $\left\{\theta(e,\phi) : e \in \kappa(I \backslash \mathcal{T}) \cup \{0\}, \phi \in \Phi_{\mathcal{T}}\right\}$: there are three cases to consider.

**Case 1:** $I_{k'} \in \mathcal{T}$. By A.3, the continuation value of $I_{k'}$ is identical across states. By Lemma 12, no feasible deviation for $I_{k'}$ can improve its stage-game payoff. $I_{k'}$ does not have any profitable one shot deviation.

**Case 2:** $I_{k'} \in I \backslash \mathcal{T}$, $k' \neq e$. Choose a number $Z > \sup_{\{\phi \in \Phi, k \in \kappa(I \backslash \mathcal{T})\}} u_k(\phi)$. For any assignment $\phi \in \Phi_{\mathcal{T}}$, the manipulated assignment resulting from a feasible deviation must still be an assignment in $\Phi$. So $Z$ is larger than the payoff any institution can obtain in any feasible deviation from an assignment in $\Phi_{\mathcal{T}}$.

Consider a one-shot deviation $(I_{k'}, B)$ by institution $I_{k'}$. Without deviation, $I_{k'}$ has value $(1-\delta)u_{k'}(\phi) + \delta u^e_{k'}$. After deviation, $I_{k'}$ yields less than

$$(1-\delta)Z + \delta U_{k'}(\underline{\theta}(e,0)) = (1-\delta)Z + \delta(1-\delta^L)\underline{u}^{\mathcal{T}}_{k'} + \delta^{L+1}u^{k'}_{k'}$$

There is no profitable one-shot deviation for $I_{k'}$ if

$$(1-\delta)u_{k'}(\phi)+\delta u^e_{k'} \geq (1-\delta)Z+\delta(1-\delta^L)\underline{u}^{\mathcal{T}}_{k'}+\delta^{L+1}u^{k'}_{k'}$$

As $\delta \to 1$, the LHS converges to $u^e_{k'}$ while the RHS converges to $u^{k'}_{k'}$. By construction, $u^e_{k'} > u^{k'}_{k'}$. It follows that such deviations are not profitable for $\delta$ high enough.

**Case 3:** $I_{k'} \in I\backslash\mathcal{T}$, $k' = e$. Without deviation, $I_{k'}$ has value $(1-\delta)u_{k'}(\phi)+\delta u^{k'}_{k'}$. After deviation, $I_{k'}$ yields less than

$$(1-\delta)Z+\delta U_{k'}(\underline{\theta}(k',0)) = (1-\delta)Z+\delta(1-\delta^L)\underline{u}^{\mathcal{T}}_{k'}+\delta^{L+1}u^{k'}_{k'}.$$

There is no profitable one-shot deviation for $I_{k'}$ if

$$(1-\delta)u_{k'}(\phi)+\delta u^{k'}_{k'} \geq (1-\delta)Z+\delta(1-\delta^L)\underline{u}^{\mathcal{T}}_{k'}+\delta^{L+1}u^{k'}_{k'}.$$

The inequality is equivalent to

$$Z-u_{k'}(\phi) \leq \delta(1+\ldots+\delta^{L-1})[u^{k'}_{k'}-\underline{u}^{\mathcal{T}}_{k'}]$$

By construction, $u^{k'}_{k'}-\underline{u}^{\mathcal{T}}_{k'} > 0$. Choose $L$ large enough so that $L(u^{k'}_{k'}-\underline{u}^{\mathcal{T}}_{k'}) > Z-u_{k'}(\phi)$. As $\delta \to 1$, the LHS remains unchanged while the RHS converges to $L(u^{k'}_{k'}-\underline{u}^{\mathcal{T}}_{k'})$, so such deviations are not profitable for $\delta$ high enough.

For states $\{\underline{\theta}(k,t) : k \in \kappa(I\backslash\mathcal{T}), 0 \leq t \leq L-1\}$: there are three cases to consider.

**Case 1:** $I_{k'} \in \mathcal{T}$. By A.3, the continuation value of $I_{k'}$ is identical across states. By Lemma 12, no feasible deviation for $I_{k'}$ can improve its stage-game payoff. $I_{k'}$ does not have any profitable one shot deviation.

**Case 2:** $I_{k'} \in I \backslash \mathcal{T}$, $k' \neq k$. Without deviation, institution $I_{k'}$ has payoff

$$(1-\delta^{L-t})u_{k'}(\underline{\phi}_k)+\delta^{L-t}u_{k'}^{k}$$

With any deviation, $I_{k'}$ has payoff less than

$$(1-\delta)Z+\delta(1-\delta^{L})\underline{u}_{k'}^{\mathcal{T}}+\delta^{L+1}u_{k'}^{k'}$$

There is no profitable one-shot deviation for $I_{k'}$ if

$$(1-\delta^{L-t})u_{k'}(\underline{\phi}_k)+\delta^{L-t}u_{k'}^{k} \geq (1-\delta)Z+\delta(1-\delta^{L})\underline{u}_{k'}^{\mathcal{T}}+\delta^{L+1}u_{k'}^{k'}$$

Observe that as $\delta \to 1$, the LHS converges to $u_{k'}^{k}$ for all $t$ such that $0 \leq t \leq L$, while the RHS converges to $u_{k'}^{k'}$. By construction $u_{k'}^{k} > u_{k'}^{k'}$. So the above inequality holds for sufficiently high $\delta$.

**Case 3:** $I_{k'} \in I \backslash \mathcal{T}$, $k' = k$. Without deviation, institution $I_{k'}$ has payoff

$$(1-\delta^{L-t})\underline{u}_{k'}^{\mathcal{T}}+\delta^{L-t}u_{k'}^{k'}.$$

When deviating from $\underline{\phi}_{k'}$, by Lemma 11, $I_k$'s stage-game payoff is at most $\underline{u}_k^{\mathcal{T}}$. So $I_{k'}$'s discounted expected payoff from deviation is at most

$$(1-\delta)\underline{u}_{k'}^{\mathcal{T}}+\delta(1-\delta^{L})\underline{u}_{k'}^{\mathcal{T}}+\delta^{L+1}u_{k'}^{k'}=(1-\delta^{L+1})\underline{u}_{k'}^{\mathcal{T}}+\delta^{L+1}u_{k'}^{k'}$$

Institution $I_{k'}$ has no profitable deviation if

$$(1-\delta^{L-t})\underline{u}_{k'}^{\mathcal{T}}+\delta^{L-t}u_{k'}^{k'} \geq (1-\delta^{L+1})\underline{u}_{k'}^{\mathcal{T}}+\delta^{L+1}u_{k'}^{k'}, \tag{A.4}$$

or

$$u_{k'}^{k'} \geq \underline{u}_{k'}^{\mathcal{T}},$$

which is true by construction. So $I_{k'}$ has no profitable one-shot deviation.

We have verified that there is no profitable one-shot deviation in any states of the automaton. This completes the proof. $\square$

**Proof of Theorem 1.** To prove the result, it is sufficient to show that if $\psi$ is a self-enforcing assignment process, then

1. $\psi(h) \in \Phi_{\mathcal{A}}$ at every ex post history $h \in \mathcal{H}^F$;
2. $\psi(I_k|h) = \widehat{A}_k$ for every $I_k \in \mathcal{T}$ at every ex post history $h \in \mathcal{H}^F$; and
3. $u_k(\psi) \geq \underline{u}_k^{\mathcal{T}}$ for every $I_k \in I \backslash \mathcal{T}$.

I now establish these claims in order.

1. This follows from the definition of self-enforcing matching process.
2. If $\mathcal{T} = \emptyset$, there is nothing to prove. Suppose $\mathcal{T} = \{(I_{k_1}, \widehat{A}_{k_1}), \ldots, (I_{k_G}, \widehat{A}_{k_G})\}$. The proof proceeds by induction.

   First I establish that in every self-enforcing assignment process $\psi$, $\psi(I_{k_1}|h) = \widehat{A}_{k_1}$ for all $h \in \mathcal{H}^F$. Suppose by contradiction that $\overline{\psi}(I_{k_1}|\widetilde{h}) \neq \widehat{A}_{k_1}$ for some self-enforcing assignment process $\overline{\psi}$ at some $\widetilde{h} \in \mathcal{H}^F$.

   By the construction of $\mathcal{T}$, $u_{k_1}(B) \leq u_{k_1}(\widehat{A}_{k_1})$ for all $B \subseteq \mathcal{A}$. By the strictness of institutions' preferences, $u_{k_1}(\overline{\psi}(I_{k_1}|\widetilde{h})) < u_{k_1}(\widehat{A}_{k_1})$. In addition, since $u_{k_1}(\widehat{A}_{k_1})$ is the highest possible stage-game payoff for $I_{k_1}$, $u_{k_1}(\psi) \leq u_{k_1}(\widehat{A}_{k_1})$ for every assignment process $\psi$. Together these imply

$$\begin{aligned} u_{k_1}(\overline{\psi}|_{\widetilde{h}}) &= (1-\delta)u_{k_1}(\overline{\psi}(I_{k_1}|\widetilde{h})) + \delta u_{k_1}(\overline{\psi}|_{\widetilde{h},\overline{\psi}(\widetilde{h})}) \\ &< (1-\delta)u_{k_1}(\widehat{A}_{k_1}) + \delta u_{k_1}(\widehat{A}_{k_1}) = u_{k_1}(\widehat{A}_{k_1}) \end{aligned} \tag{A.5}$$

Since $\overline{\psi}$ is a self-enforcing assignment process, $\overline{\psi}|_{\widetilde{h}}$ must also be self-enforcing, so $\overline{\psi}|_{\widetilde{h}}(h) \in \Phi_{\mathcal{A}}$ at every $h \in \mathcal{H}^F$. In addition, for every $\phi \in \Phi_{\mathcal{A}}$, by the construction of $\mathcal{T}$ and the strictness of agents' preferences, $v_j(I_{k_1}) > v_j(\phi)$ for every $a_j \in \widehat{A}_{k_1} \setminus \phi(I_{k_1})$. So $\widehat{A}_{k_1} \in D^{\mathcal{A}}_{k_1}(\phi)$ for all $\phi \in \Phi_{\mathcal{A}}$. By Lemma 13,

$$u_{k_1}(\overline{\psi}|_{\widetilde{h}}) \geq u_{k_1}(\widehat{A}_{k_1}) \tag{A.6}$$

Inequalities (A.5) and (A.6) cannot be true at the same time, a contradiction. So $\overline{\psi}(I_{k_1}|h) = \widehat{A}_{k_1}$ for all $h \in \mathcal{H}^F$.

Suppose it has been shown that in every self-enforcing assignment process $\psi$, $\psi(I_{k_i}|h) = \widehat{A}_{k_i}$ for $i = 1, \ldots, g-1$ at every ex post history $h \in \mathcal{H}^F$. I show that this implies that in every self-enforcing assignment process $\psi$, $\psi(I_{k_g}|h) = \widehat{A}_{k_g}$ at every ex post history $h \in \mathcal{H}^F$. Suppose by contradiction that $\overline{\psi}(I_{k_g}|\widetilde{h}) \neq \widehat{A}_{k_g}$ for some self-enforcing assignment process $\overline{\psi}$ and some $\widetilde{h} \in \mathcal{H}^F$.

Since by the inductive hypothesis $\overline{\psi}(I_{k_i}|h) = \widehat{A}_{k_i}$ for all $1 \leq i \leq g-1$ and all $h \in \mathcal{H}^F$, it must be that $\overline{\psi}(I_{k_g}|h) \subseteq \mathcal{A} \setminus \cup_{i=1}^{g-1} \widehat{A}_{k_i}$ at all $h \in \mathcal{H}^F$. From the construction of $\mathcal{T}$, $u_{k_g}(\overline{\psi}(I_{k_g}|h)) \leq u_{k_g}(\widehat{A}_{k_g})$ for all $h \in \mathcal{H}^F$, it follows that at every ex post history $h \in \mathcal{H}^F$,

$$u_{k_g}(\overline{\psi}|_h) \leq \widehat{A}_{k_g}$$

In addition, by the strictness of institutions preferences, $\overline{\psi}(I_{k_g}|\widetilde{h}) \neq \widehat{A}_{k_g}$ implies that

$u_{k_g}(\overline{\psi}(I_{k_g}|\widetilde{h})) < u_{k_g}(\widehat{A}_{k_g})$. Together these imply

$$\begin{aligned} u_{k_g}(\overline{\psi}|_{\widetilde{h}}) &= (1-\delta)u_{k_g}(\overline{\psi}(I_{k_g}|\widetilde{h})) + \delta u_{k_g}(\overline{\psi}|_{\widetilde{h},\overline{\psi}(\widetilde{h})}) \\ &< (1-\delta)u_{k_g}(\widehat{A}_{k_g}) + \delta u_{k_g}(\widehat{A}_{k_g}) = u_{k_g}(\widehat{A}_{k_g}) \end{aligned} \tag{A.7}$$

Since $\overline{\psi}$ is a self-enforcing assignment process, $\overline{\psi}|_{\widetilde{h}}$ must also be self-enforcing. Let $\Phi^g = \{\phi \in \Phi_{\mathcal{A}} : \phi(I_{k_i}) = \widehat{A}_{k_i} \text{ for } i = 1, \ldots, g-1\}$. By the inductive hypothesis, $\overline{\psi}|_{\widetilde{h}}(h) \in \Phi^g$ at every $h \in \mathcal{H}^F$. In addition, for every $\phi \in \Phi^g$, by the construction of $\mathcal{T}$ and the strictness of agents' preferences, $v_j(I_{k_g}) > v_j(\phi)$ for every $a_j \in \widehat{A}_{k_g} \backslash \phi(I_{k_g})$. So $\widehat{A}_{k_g} \in D^{\mathcal{A}}_{k_g}(\phi)$ for all $\phi \in \Phi^g$. By Lemma 13,

$$u_{k_1}(\overline{\psi}|_{\widetilde{h}}) \geq u_{k_1}(\widehat{A}_{k_1}) \tag{A.8}$$

Inequalities (A.7) and (A.8) cannot be true at the same time, a contradiction. So $\overline{\psi}(I_{k_g}|h) = \widehat{A}_{k_g}$ for all $h \in \mathcal{H}^F$, completing the induction argument.

3. Let $\psi$ be a self-enforcing assignment process. $\psi(h) \in \Phi_{\mathcal{T}}$ for all ex post histories $h \in \mathcal{H}^F$.

   For every institution $I_k \in I \backslash \mathcal{T}$, from the definition of $\underline{u}^{\mathcal{T}}_k$, there exists $A^{\phi} \in D^{\mathcal{A}}_k(\phi)$ for every $\phi \in \Phi_{\mathcal{T}}$ such that

$$u_k(A^{\phi}) \geq \underline{u}^{\mathcal{T}}_k.$$

   By Lemma 13, $u_k(\psi) \geq \underline{u}^{\mathcal{T}}_k$ for every $I_k \in I \backslash \mathcal{T}$.

□

**Proof of Corollary 1.** Suppose all agents share a common ordinal ranking $I_{k_1} \succ I_{k_2} \succ \ldots \succ I_{k_K}$ over the institutions. For each institution $I_{k_g}$, $1 \leq g \leq K$, define

$$\widehat{A}_{k_g} = \underset{B \subseteq \mathcal{A} \backslash \cup_{i=1}^{g-1} \widehat{A}_{k_i}}{\arg\max}\ u_{k_g}(B)$$

Then $\mathcal{T} = \{(I_{k_1}, \widehat{A}_{k_1}), \ldots, (I_{k_K}, \widehat{A}_{k_K})\}$ is the top coalition sequence of the matching market. Let $\phi_{\mathcal{T}}$ be the static assignment that satisfies $\phi_{\mathcal{T}}(I_k) = \widehat{A}_k$ for $k = 1, \ldots, K$.

**Claim 1.** $\phi_{\mathcal{T}}(I_k) = \widehat{A}_k$ *is the unique static stable matching.*

The proof is by induction. It is clear that in every static stable assignment, $I_{k_1}$ must be matched to $\widehat{A}_{k_1}$, otherwise $\widehat{A}_{k_1}$ is a feasible profitable deviation for $I_{k_1}$. Suppose $I_{k_1}, \ldots, I_{k_{g-1}}$ is matched to $\widehat{A}_{k_1}, \ldots, \widehat{A}_{k_{g-1}}$, respectively, in every static assignment. Let $\phi$ be an arbitrary static stable assignment. I prove that $\phi(I_{k_g}) = \widehat{A}_{k_g}$.

Suppose by contradiction that $\phi(I_{k_g}) \neq \widehat{A}_{k_g}$, then by the inductive hypothesis, $\psi(a_j) \in I \backslash \{I_{k_1}, \ldots, I_{k_{g-1}}\}$ for every $a_j \in \widehat{A}_{k_g} \backslash \phi(I_{k_g})$. By the definition of top coalition sequence and the strictness of agents' preferences, $v_j(I_{k_g}) > v_j(\phi)$ for every $a_j \in \widehat{A}_{k_g} \backslash \phi(I_{k_g})$, so $\widehat{A}_{k_g} \in D^{\mathcal{A}}_{k_g}(\phi)$. Similarly, by the inductive hypothesis, $\phi(I_{k_g}) \subseteq \mathcal{A} \backslash \cup_{i=1}^{g-1} \widehat{A}_{k_i}$, so $u_{k_g}(\widehat{A}_{k_g}) > u_{k_g}(\phi)$ by the definition of a top coalition sequence and the strictness of institutions' preferences. This implies that $\widehat{A}_{k_g}$ is a profitable and feasible deviation, a contradiction to $\phi$ being a stable assignment. So $\phi(I_{k_g}) = \widehat{A}_{k_g}$. This completes the inductive step.

**Claim 2.** *There is a unique self-enforcing assignment process* $\psi_{\mathcal{T}}$*, where* $\psi_{\mathcal{T}}(h) = \phi_{\mathcal{T}}$ *for all* $h \in \mathcal{H}^F$.

Since $I \subseteq \mathcal{T}$, $\Phi_{\mathcal{T}} = \{\phi : \phi(I_k) = \widehat{A}_k \ \forall k = 1, \ldots, K\}$. So $\phi_{\mathcal{T}}$ is the unique element of the set $\Phi_{\mathcal{T}}$. The claim then follows from Theorem 1. $\square$

## A.3 Proofs in Section 1.4

**Definition A.3.1.** Fix the hierarchy $O = \{\mathcal{P}_1, \ldots, \mathcal{P}_G, \mathcal{R}\}$. For each $I_k \in \mathcal{P}_g, g = 1, \ldots, G$, define $\mathcal{P}(I_k) \equiv \cup_{i=g+1}^{G} \mathcal{P}_i \cup \{I_l \in \mathcal{P}_g : l < k\}$

**Definition A.3.2.** Fix the hierarchy $O = \{\mathcal{P}_1, \ldots, \mathcal{P}_G, \mathcal{R}\}$ and a set of agents $C$. For each $I_k \in \mathcal{P}_g, g = 1, \ldots, G$, the sets of agent $A_k^O(C)$ is defined by

$$A_k^O(C) = \underset{B \subseteq \mathcal{A} \setminus \left[C \cup \left(\cup_{I_l \in \mathcal{P}(I_k)} A_l^O(C)\right)\right]}{\arg\max} \left\{u_k(B) + \sum_{a_j \in B} v_j(I_k, B)\right\}$$

**Definition A.3.3.** Fix the hierarchy $O = \{\mathcal{P}_1, \ldots, \mathcal{P}_G, \mathcal{R}\}$ and a set of agents $C$. For each $I_k \in \mathcal{P}_g, g = 1, \ldots, G$, define

$$\pi_k^O(C) = u_k(A_k^O(C)) + \sum_{a_j \in A_k^O(C)} v_j(I_k, A_k^O(C))$$

**Lemma 14.** *For each $I_k \in \cup_{g=1}^G \mathcal{P}_g$, and all $C \subseteq \mathcal{A}$ such that $|C| \leq Q(\mathcal{R})$,*

$$\pi_k^O(C) = \widehat{\pi}_k \tag{A.9}$$

*In addition, for every $I_k \in \mathcal{R}$,*

$$\widehat{\pi}_k > \underline{\pi}_k(\mathcal{R})$$

*Proof.* Fix any $I_k \in \mathcal{P}_g$ for some $g = 1, \ldots, G$. By the construction of $O$, we have $\widehat{\pi}_k = \underline{\pi}_k(I \setminus \cup_{i=1}^{g-1} \mathcal{P}_i)$. Recall that

$$\begin{aligned}
\pi_k^O(C) &= u_k(A_k^O(C)) + \sum_{a_j \in A_k^O(C)} v_j(I_k, A_k^O(C)) \\
&= \max_{B \subseteq \mathcal{A} \setminus \left[C \cup \left(\cup_{I_l \in \mathcal{P}(I_k)} A_l^O(C)\right)\right]} \left\{u_k(B) + \sum_{a_j \in B} v_j(I_k, B)\right\} \\
&\leq \max_{B \subseteq \mathcal{A}} \left\{u_k(B) + \sum_{a_j \in B} v_j(I_k, B)\right\} \\
&= \widehat{\pi}_k
\end{aligned} \tag{A.10}$$

Note that since $C \le Q(\mathcal{R})$ and $A_l^O(C) \le q_l$ for all $I_l \in \cup_{g=1}^{G} \mathcal{P}_g$, we have

$$\left| C \cup \left( \cup_{I_l \in \mathcal{P}(I_k)} A_l^O(C) \right) \right| \le Q(\mathcal{R}) + \sum_{i=g}^{G} Q(\mathcal{P}_i) = Q(I \setminus \cup_{i=1}^{g-1} \mathcal{P}_i)$$

So

$$\begin{aligned}
\pi_k^O(C) &= \max_{B \subseteq \mathcal{A} \setminus \left[ C \cup \left( \cup_{I_l \in \mathcal{P}(I_k)} A_l^O(C) \right) \right]} \left\{ u_k(B) + \sum_{a_j \in B} v_j(I_k, B) \right\} \\
&\ge \min_{\{A \subseteq \mathcal{A} : |A| \le Q(I \setminus \cup_{i=1}^{g-1} \mathcal{P}_i)\}} \max_{B \subseteq \mathcal{A} \setminus A} \left\{ u_k(B) + \sum_{a_j \in B} v_j(I_k, B) \right\} \qquad \text{(A.11)} \\
&= \underline{\pi}_k(I \setminus \cup_{i=1}^{g-1} \mathcal{P}_i)
\end{aligned}$$

Combining inequalities (A.10) and (A.11), we have

$$\widehat{\pi}_k \ge \pi_k^O(C) \ge \underline{\pi}_k(I \setminus \cup_{i=1}^{g-1} \mathcal{P}_i) = \widehat{\pi}_k,$$

so $\pi_k^O(C) = \widehat{\pi}_k$.

For every $I_k \in \mathcal{R}$, $\widehat{\pi}_k \ge \underline{\pi}_k(\mathcal{R})$. If $\widehat{\pi}_k = \underline{\pi}_k(\mathcal{R})$, then by the construction of $O$, $I_k \notin \mathcal{R}$, a contradiction. So $\widehat{\pi}_k > \underline{\pi}_k(\mathcal{R})$. □

I use $\kappa(\mathcal{R}) \equiv \{k : I_k \in \mathcal{R}\}$ to denote the indices of institutions that are in $\mathcal{R}$, and $\Gamma(\mathcal{R}) \equiv \left\{ \sigma : \{1, 2, \ldots, |\kappa(\mathcal{R})|\} \to \kappa(\mathcal{R}) \right\}$ to denote the set of ordering of the numbers in $\kappa(\mathcal{R})$. For all $k \in \kappa(\mathcal{R})$, define

$$\widehat{\Gamma}_k(\mathcal{R}) = \{\sigma \in \Gamma(\mathcal{R}) : \sigma(1) = k\}.$$

$\widehat{\Gamma}_k(\mathcal{R})$ is the set of permutations in $\Gamma(\mathcal{R})$ that ranks $k$ in the first place.

**Definition A.3.4.** Given a permutation $\sigma \in \Gamma(\mathcal{R})$, the sets of agent $\{A_k^\sigma : k \in \kappa(\mathcal{R})\}$ is defined by

$$A_{\sigma(l)}^\sigma = \operatorname*{arg\,max}_{B \subseteq \mathcal{A} \setminus \cup_{i=1}^{l-1} A_{\sigma(i)}^\sigma} \left\{ u_{\sigma(l)}(B) + \sum_{a_j \in B} v_j(I_{\sigma(l)}, B) \right\}$$

for each $1 \leq l \leq |\kappa(\mathcal{R})|$.

**Definition A.3.5.** For each institution $I_k \in \mathcal{R}$ and $\sigma \in \Gamma(\mathcal{R})$, define

$$\pi_k(\sigma) = u_k(A_k^\sigma) + \sum_{a_j \in A_k^\sigma} v_j(I_k, A_k^\sigma)$$

**Lemma 15.** *For any institution $I_k \in \mathcal{R}$, and all $\sigma \in \widehat{\Gamma}_k(\mathcal{R})$,*

$$\pi_k(\sigma) = \widehat{\pi}_k$$

*Proof.* For all $\sigma \in \widehat{\Gamma}_k(\mathcal{R})$, we have $\sigma(1) = k$. By definition,

$$\begin{aligned} A_k^\sigma = A_{\sigma(1)}^\sigma &= \underset{B \subseteq \mathcal{A} \setminus \cup_{i=1}^{1-1} A_{\sigma(i)}^\sigma}{\arg\max} \left\{ u_{\sigma(1)}(B) + \sum_{a_j \in B} v_j(I_{\sigma(1)}, B) \right\} \\ &= \underset{B \subseteq \mathcal{A}}{\arg\max} \left\{ u_k(B) + \sum_{a_j \in B} v_j(I_k, B) \right\}, \end{aligned}$$

so

$$\begin{aligned} \pi_k(\sigma) &= u_k(A_k^\sigma) + \sum_{a_j \in A_k^\sigma} v_j(I_k, A_k^\sigma) \\ &= \max_{B \subseteq \mathcal{A}} \left\{ u_k(B) + \sum_{a_j \in B} v_j(I_k, B) \right\} = \widehat{\pi}_k \end{aligned}$$

□

**Lemma 16.** *For any $\sigma \in \Gamma(\mathcal{R})$ and any institution $I_k \in \mathcal{R}$,*

$$\pi_k(\sigma) \geq \underline{\pi}_k(\mathcal{R})$$

*Proof.* For every institution $I_k$ and permutation $\sigma \in \Gamma(\mathcal{R})$, define $C_k(\sigma) = \cup_{\{1 \leq \sigma^{-1}(l) < \sigma^{-1}(k)\}} A_l^\sigma$.

Then

$$A_k^{\sigma} = \underset{B \subseteq \mathcal{A} \setminus C_k(\sigma)}{\arg\max} \left\{ u_k(B) + \sum_{a_j \in B} v_j(I_k, B) \right\}$$

By definition,

$$\begin{aligned}
\pi_k(\sigma) &= u_k(A_k^{\sigma}) + \sum_{a_j \in A_k^{\sigma}} v_j(I_k, A_k^{\sigma}) \\
&= \max_{B \subseteq \mathcal{A} \setminus C_k(\sigma)} \left\{ u_k(B) + \sum_{a_j \in B} v_j(I_k, B) \right\}
\end{aligned}$$

Note that for all $\sigma \in \Gamma(\mathcal{R})$ and $k \in \kappa(\mathcal{R})$, $|C_k(\sigma)| \leq Q(\mathcal{R})$. So

$$\begin{aligned}
\pi_k(\sigma) &= \max_{B \subseteq \mathcal{A} \setminus C_k(\sigma)} \left\{ u_k(B) + \sum_{a_j \in B} v_j(I_k, B) \right\} \\
&\geq \min_{\{A \subseteq \mathcal{A} : |A| \leq Q(\mathcal{R})\}} \max_{B \subseteq \mathcal{A} \setminus A} \left\{ u_k(B) + \sum_{a_j \in B} v_j(I_k, B) \right\} \\
&= \underline{\pi}_k(\mathcal{R})
\end{aligned}$$

□

**Proof of Theorem 4.** There are two cases to consider.

**Case 1:** $\mathcal{R} = \emptyset$. It suffices to check that $M^{\circ}$ is nonempty, as any degenerate lottery on $M^{\circ}$ will be in $\Lambda^{\circ}$. Define the matching $m^* = (\phi^*, \{w_k^*\}_{k=1}^K)$ by

$$\phi^*(I_k) = A_k^{O}(\emptyset) \;\; \forall k = 1, \ldots, K$$

and

$$w^*_{kj} = \begin{cases} -v_j(I_k, A^O_k(\emptyset)) & \text{if } a_j \in A^O_k(\emptyset) \\ 0 & \text{otherwise} \end{cases}$$

By construction,

$$u_k(m^*) = \pi^O_k(\emptyset)$$

for all $1 \leq k \leq K$.

Since $\mathcal{R} = \emptyset$, we have $r = 0$, and $|\emptyset| = 0 = Q(\mathcal{R})$. By Lemma 14, $u_k(m^*) = \pi^O_k(\emptyset) = \widehat{\pi}_k$ for all $1 \leq k \leq K$. In addition, all employed agents have zero payoff, so $m^*$ is individually rational for agents. Therefore, $m^* \in M^\circ$.

**Case 2:** $\mathcal{R} \neq \emptyset$. First I show that $M^\circ$ is nonempty. For each $k \in \kappa(\mathcal{R})$, fix

$$\widehat{\sigma}_k \in \widehat{\Gamma}_k(\mathcal{R})$$

and define the matchings $\widehat{m}_k = (\widehat{\phi}^k, \{\widehat{w}^k_l\}_{l=1}^K)$ by

$$\widehat{\phi}^k(I_l) = \begin{cases} A^{\widehat{\sigma}_k}_l & \text{if } I_l \in \mathcal{R} \\ A^O_l(\cup_{i \in \kappa(\mathcal{R})} A^{\widehat{\sigma}_k}_i) & \text{if } I_l \in \cup_{g=1}^G \mathcal{P}_g \end{cases}$$

and

$$\widehat{w}^k_{lj} = \begin{cases} -v_j(I_l, \widehat{\phi}_k(I_l)) & \text{if } a_j \in \widehat{\phi}_k(I_l) \\ 0 & \text{otherwise} \end{cases}$$

By construction, for each $\widehat{m}_k$ and every $I_{k'} \in \cup_{g=1}^G \mathcal{P}_g$,

$$u_{k'}(\widehat{m}_k) = \pi^O_{k'}(\cup_{l \in \kappa(\mathcal{R})} A^{\widehat{\sigma}_k}_l) = \widehat{\pi}_{k'}$$

where the second equality follows from Lemma 14 and the fact that $\left|\cup_{l \in \kappa(\mathcal{R})} A^{\widehat{\sigma}_k}_l\right| \leq Q(\mathcal{R})$. In addition, in each $\widehat{m}_k$, all agents are getting zero payoff. So $\{\widehat{m}_k\}_{k\in\kappa(\mathcal{R})} \subseteq M^\circ$.

To see $\Lambda^\circ$ is nonempty, first observe that for each $k \in \kappa(\mathcal{R})$,

$$\begin{aligned} u_k(\widehat{m}_k) &= u_k(A_k^{\widehat{\sigma}_k}) - \sum_{a_j \in A_k^{\widehat{\sigma}_k}} \widehat{w}_{kj}^k \\ &= u_k(A_k^{\widehat{\sigma}_k}) + \sum_{a_j \in A_k^{\widehat{\sigma}_k}} v_j(I_k, A_k^{\widehat{\sigma}_k}) \\ &= \pi_k(\widehat{\sigma}_k) = \widehat{\pi}_k > \underline{\pi}_k(\mathcal{R}) \end{aligned}$$

The first two equalities above follow from the definition of the matching $\widehat{m}_k$, the third equality follows from the definition of $\pi_k(.)$, the fourth follows from Lemma 15. Lastly, $\widehat{\pi}_k > \underline{\pi}_k(\mathcal{R})$ follows from Lemma 14.

Second, for all $k, k' \in \kappa(\mathcal{R})$, $k \neq k'$,

$$\begin{aligned} u_{k'}(\widehat{m}_k) &= u_{k'}(A_{k'}^{\widehat{\sigma}_k}) - \sum_{a_j \in A_{k'}^{\widehat{\sigma}_k}} \widehat{w}_{k'j}^k \\ &= u_{k'}(A_{k'}^{\widehat{\sigma}_k}) + \sum_{a_j \in A_{k'}^{\widehat{\sigma}_k}} v_j(I_{k'}, A_{k'}^{\widehat{\sigma}_k}) \\ &= \pi_{k'}(\widehat{\sigma}_k) \geq \underline{\pi}_{k'}(\mathcal{R}) \end{aligned}$$

where the last inequality follows from Lemma 16.

Fix a set of weights $\{\lambda_l : l \in \kappa(\mathcal{R})\}$ that satisfy

$$\lambda_l > 0 \;\; \forall l \in \kappa(\mathcal{R}), \text{ and } \sum_{l \in \kappa(\mathcal{R})} \lambda_l = 1.$$

Let $\lambda^0 \in \Delta(M^\circ)$ be the lottery over $M^\circ$ that is defined by

$$\lambda^0 = \sum_{l \in \kappa(\mathcal{R})} \lambda_l \widehat{m}_l$$

I will show $\mathbb{E}_{\lambda^0}[u_k(m)] > \underline{\pi}_k(\mathcal{R})\ \forall k \in \kappa(\mathcal{R})$, which proves $\lambda^0 \in \Lambda^\circ$ and therefore the nonemptyness of $\Lambda^\circ$.

Fix any $k \in \kappa(\mathcal{R})$, then $u_k(\widehat{m}_l) \geq \underline{\pi}_k(\mathcal{R})$ for all $l \in \kappa(\mathcal{R})$. In addition, $u_k(\widehat{m}_k) = \widehat{\pi}_k > \underline{\pi}_k(\mathcal{R})$. So $\mathbb{E}_{\lambda^0}[u_k(m)] = \sum_{l \in \kappa(\mathcal{R})} \lambda_l u_k(\widehat{m}_l) > \underline{\pi}_k(\mathcal{R})$, and $\lambda^0 \in \Lambda^\circ$. □

**Lemma 17.** *For every $m \in M^\circ$ and every $I_k \in \cup_{g=1}^G \mathcal{P}_g$,*

$$\sup_{(A'_{k'}, \zeta'_{k'}) \in D_{k'}(m)} u_{k'}(m, [A'_{k'}, \zeta'_{k'}]) = u_{k'}(m),$$

*so no institution in $\cup_{g=1}^G \mathcal{P}_g$ has profitable feasible deviations.*

*Proof.* Consider any feasible deviation $(A'_k, \zeta'_k) \in D_k(m)$. For the deviation to be feasible, each $a_j \in A'_k$ must be individually rational, so $v_j(I_k, A'_k) + \zeta'_{kj} \geq 0$. This implies

$$\begin{aligned} u_k(A'_k) - \sum_{a_j \in A'_k} \zeta'_{kj} &\leq u_l(A'_k) + \sum_{a_j \in A'_k} v_j(I_k, A'_k) \\ &\leq \max_{A \subseteq \mathcal{A}} u_k(A) + \sum_{a_j \in A} v_j(I_k, A) \\ &= \widehat{\pi}_k = u_k(m) \end{aligned}$$

□

**Lemma 18.** *Suppose $\mathcal{R} = \emptyset$, then every matching in $m \in M^\circ$ is a static stable matching*

*Proof.* Since $\mathcal{R} = \emptyset$, $\cup_{g=1}^G \mathcal{P}_g = I$. By Lemma 17, no institutions in $I$ has any profitable feasible deviations. By construction, every agent $a_j \in \mathcal{A}$ has utility $v_j(m) \geq 0$, so no agent has any profitable deviation. The matching $m$ is a static stable matching. □

**Lemma 19.** *Suppose $\mathcal{R} \neq \emptyset$, then there exist static matchings $\{\underline{m}_k\}_{k\in\kappa(\mathcal{R})} \subseteq M^\circ$ such that*

1. *for all $k \in \kappa(\mathcal{R})$, $u_k(\underline{m}_k) \leq \underline{\pi}_k(\mathcal{R})$;*

2. *for all $k \in \kappa(\mathcal{R})$,*

$$\sup_{(A'_k,\zeta'_k)\in D_k(\underline{m}_k)} u_k(\underline{m}_k,[A'_k,\zeta'_k]) \leq \underline{\pi}_k(\mathcal{R});$$

*Proof.* For each $k \in \kappa(\mathcal{R})$, I use $\underline{A}_k$ to denote the set of agents

$$\underline{A}_k = \underset{\{A\subseteq\mathcal{A}:|A|\leq Q(\mathcal{R})\}}{\arg\min} \max_{B\subseteq\mathcal{A}\setminus A} \Big\{u_k(B) + \sum_{a_j\in B} v_j(I_k,B)\Big\},$$

and $b_k$ to denote the value

$$b_k = \max_{B\subseteq\mathcal{A}} \Big\{u_k(B) + \sum_{a_j\in B} v_j(I_k,B)\Big\} - \underline{\pi}_k(\mathcal{R}).$$

Note that $b_k \geq 0$ for all $k \in \kappa(\mathcal{R})$. For each institution $k \in \kappa(\mathcal{R})$, let $\{\underline{A}_k^l\}_{l\in\kappa(\mathcal{R})}$ be a partition of $\underline{A}_k$ such that $\left|\underline{A}_k^l\right| \leq q_l$ for all $l \in \kappa(\mathcal{R})$. This is possible because $|\underline{A}_k| \leq Q(\mathcal{R})$ by construction. Define the static matching $\underline{m}_k = (\underline{\phi}^k, \{\underline{\zeta}_l^k\}_{l=1}^K)$ by

$$\underline{\phi}^k(I_l) = \begin{cases} \underline{A}_k^l & \text{if } l \in \mathcal{R} \\ A_l^O(\underline{A}_k) & \text{if } l \in \cup_{g=1}^G \mathcal{P}_g \end{cases}$$

and

$$\underline{\zeta}_{lj}^k = \begin{cases} b_k - v_j(I_l,\underline{\phi}^k(I_l)) & \text{if } l \in \mathcal{R}\text{, and } a_j \in \underline{\phi}^k(I_l) \\ -v_j(I_l,\underline{\phi}^k(I_l)) & \text{if } l \in \cup_{g=1}^G \mathcal{P}_g\text{, and } a_j \in \underline{\phi}^k(I_l) \\ 0 & \text{otherwise} \end{cases}$$

By construction, $u_l(\underline{m}_k) = \pi_l^O(\underline{A}_k)$ for all $l \in \cup_{g=1}^G \mathcal{P}_g$. Since $|\underline{A}_k| \leq Q(\mathcal{R})$, by Lemma 14, we have $u_l(\underline{m}_k) = \pi_l^O(\underline{A}_k) = \widehat{\pi}_l$ for all $l \in \cup_{g=1}^G \mathcal{P}_g$, so $\underline{m}_k \in M^\circ$.

1. For each $l \in \mathcal{R}$, if $\underline{A}_k^k = \emptyset$, then

$$\begin{aligned} u_k(\underline{m}_k) &= u_k(\emptyset) \\ &\leq \max_{\{B: B \subseteq \mathcal{A} \backslash A\}} \left\{ u_k(B) + \sum_{a_j \in B} v_j(I_k, B) \right\} \end{aligned}$$

   for all $A$ such that $|A| \leq Q(\mathcal{R})$. So

$$\begin{aligned} u_k(\underline{m}_k) &\leq \min_{\{A \subseteq \mathcal{A}: |A| \leq Q(\mathcal{R})\}} \max_{B \subseteq \mathcal{A} \backslash A} \left\{ u_k(B) + \sum_{a_j \in B} v_j(I_k, B) \right\} \\ &= \underline{\pi}_k(\mathcal{R}) \end{aligned}$$

   If $\underline{A}_k^k \neq \emptyset$, then

$$\begin{aligned} u_k(\underline{m}_k) &= u_k(\underline{A}_k^k) - \sum_{a_j \in \underline{A}_k^k} \underline{\zeta}_{kj}^k \\ &= u_k(\underline{A}_k^k) + \sum_{a_j \in \underline{A}_k^k} v_j(I_k, \underline{A}_k^k) - \left| \underline{A}_k^k \right| \cdot b_k \\ &\leq u_k(\underline{A}_k^k) + \sum_{a_j \in \underline{A}_k^k} v_j(I_k, \underline{A}_k^k) - b_k \\ &= u_k(\underline{A}_k^k) + \sum_{a_j \in \underline{A}_k^k} v_j(I_k, \underline{A}_k^k) - \max_{B \subseteq \mathcal{A}} \left\{ u_k(B) + \sum_{a_j \in B} v_j(I_k, B) \right\} + \underline{\pi}_k(\mathcal{R}) \\ &\leq \underline{\pi}_k(\mathcal{R}) \end{aligned}$$

   The first two equalities above follow from the definition of the matching $\underline{m}_k$; the first inequality follows from $\left| \underline{A}_k^k \right| \geq 1$ and the fact that $b_k \geq 0$.

2. Consider any feasible deviation $(A_k', \zeta_k') \in D_k(\underline{m}_k)$. Suppose $A_k' \subseteq \mathcal{A} \backslash \underline{A}_k$. By feasibility, each $a_j \in A_k'$ must at least find the deviation individually rational, so $v_j(I_k, A_k') + \zeta_{kj}' \geq 0$.

This implies

$$\begin{aligned}
u_k(A'_k) - \sum_{a_j \in A'_k} \zeta'_{kj} &\leq u_k(A'_k) + \sum_{a_j \in A'_k} v_j(I_k, A'_k) \\
&\leq \max_{B \subseteq \mathcal{A} \setminus \underline{A}_k} u_k(B) + \sum_{a_j \in B} v_j(I_k, B) \\
&= \underline{\pi}_k(\mathcal{R})
\end{aligned}$$

where the second inequality above follows from the fact that $A'_k \subseteq \mathcal{A} \setminus \underline{A}_k$, and the equality follows the definition of $\underline{A}_k$.

Suppose instead $A'_k \nsubseteq \mathcal{A} \setminus \underline{A}_k$. Fix $a_i \in A'_k \cap \underline{A}_k$. By the construction of $\underline{m}_k$, $v_i(\underline{m}_k) = b_k$. For the deviation to be feasible, $a_i$ must obtain a payoff weakly higher than in $\underline{m}_k$ (and strictly higher than in $\underline{m}_k$ if $a_i \notin \underline{A}^k_k$) , so $v_i(I_k, A'_k) + \zeta'_{ki} \geq b_k$, or $-\zeta'_{kj} \leq v_i(I_k, A'_k) - b_k$; at the same time, every other agent $a_j \in A'_k$ needs to at least find the deviation individually rational, so $v_j(I_k, A'_k) + \zeta'_{kj} \geq 0$ for all $a_j \in A'_k$, $a_j \neq a_i$. We have

$$\begin{aligned}
u_k(A'_k) - \sum_{a_j \in A'_k} \zeta'_{kj} &= u_k(A'_k) - \zeta'_{ki} - \sum_{a_j \in A'_k, a_j \neq a_i} \zeta'_{kj} \\
&\leq u_k(A'_k) + \sum_{a_j \in A'_k} v_j(I_k, A'_k) - b_k \\
&= u_k(A'_k) + \sum_{a_j \in A'_k} v_j(I_k, A'_k) - \max_{B \subseteq \mathcal{A}} \left\{ u_k(B) + \sum_{a_j \in B} v_j(I_k, B) \right\} + \underline{\pi}_k(\mathcal{R}) \\
&\leq \underline{\pi}_k(\mathcal{R}).
\end{aligned}$$

We have shown that $u_k(\underline{m}_k, [A'_k, \zeta'_k]) \leq \underline{\pi}_k(\mathcal{R})$ regardless of whether $A'_k \subseteq \mathcal{A} \setminus \underline{A}_k$, so

$$\sup_{(A'_k, \zeta'_k) \in D_k(\underline{m}_k)} u_k(\underline{m}_k, [A'_k, \zeta'_k]) \leq \underline{\pi}_k(\mathcal{R}).$$

□

**Lemma 20.** *Let $U^\circ = \left\{\mathbb{E}_\lambda[u(m)] : \lambda \in \Lambda^\circ\right\}$. If $|\mathcal{R}| \geq 2$, for every vector $u \in U^\circ$, there exist vectors $\{u^k : k \in \kappa(\mathcal{R})\} \subseteq U^\circ$ such that*

$$u_k^k < u_k$$

*for all $k \in \kappa(\mathcal{R})$, and*

$$u_k^k < u_k^{k'}$$

*for all $k \neq k' \in \kappa(\mathcal{R})$.*

*Proof.* The proof of this lemma is based on Abreu, Dutta, and Smith (1994). Compared to Abreu, Dutta, and Smith (1994), the changes are:

1. instead of imposing NEU condition like Abreu, Dutta, and Smith (1994), it is shown that the matching environment always satisfies NEU;
2. since the set $M^\circ$ can be potentially infinite, a different method of construction for $u^k$ is needed;
3. the current proof operate in a cooperative game environment without action spaces; and
4. only a subset of indices, $\kappa(\mathcal{R})$, are concerned.

Whenever possible, I will cite intermediate results in Abreu, Dutta, and Smith (1994) without reproducing their proofs. I will first establish that the set $M^\circ$ satisfies what Abreu, Dutta, and Smith (1994) called the non-equivalent utilities (NEU) condition for institutions in $\mathcal{R}$. Fix an arbitrary $m^0 = (\phi^0, \{\zeta_l^0\}_{l=1}^K) \in M^\circ$ and define

$$b_k = \max_{B \subseteq \mathcal{A}} \left\{ u_k(B) + \sum_{a_j \in B} v_j(I_k, B) \right\} - \underline{\pi}_k(\mathcal{R}).$$

Note that $b_k > 0$ for all $k \in \kappa(\mathcal{R})$. For each $k \in \kappa(\mathcal{R})$, define the matching $m^k = (\phi^k, \{\zeta_l^k\}_{l=1}^K)$ by $\phi^k = \phi^0$, and

$$\zeta_{lj}^k = \begin{cases} \zeta_{kj}^0 + b_k & \text{if } l = k \text{ and } a_j \in \phi^k(I_k) \\ \zeta_{lj}^0 & \text{otherwise} \end{cases}$$

Now I establish a few properties about the matchings $m^k$.

Clearly, $v_j(m^k) = v_j(m^0)$ for all $j \in \mathcal{A} \backslash \phi^k(I_k)$, and $v_j(m^k) > v_j(m^0)$ for all $j \in \phi^k(I_k)$, so $m^k$ is individually rational for all agents. In addition, $u_k(m^k) = u_k(m^0) = \widehat{\pi}_k$ for all $I_k \in \cup_{g=1}^G \mathcal{P}_g$. It follows that $m^k \in M^\circ$ for all $k \in \kappa(\mathcal{R})$.

By the construction of $\{m^k : k \in \kappa(\mathcal{R})\}$, for each $k \in \kappa(\mathcal{R})$,

$$\begin{aligned} u_k(m^k) &= u_k(\phi^k(I_k)) - \sum_{a_j \in \phi^k(I_k)} \zeta_{kj}^k \\ &= u_k(\phi^0(I_k)) - \sum_{a_j \in \phi^0(I_k)} \zeta_{kj}^0 - \left|\phi^0(I_k)\right| \cdot b_k \\ &\leq u_k(\phi^0(I_k)) - \sum_{a_j \in \phi^0(I_k)} \zeta_{kj}^0 - b_k \\ &\leq u_k(\phi^0(I_k)) + \sum_{a_j \in \phi^0(I_k)} v_j(I_k, \phi^0(I_k)) - b_k \\ &= u_k(\phi^0(I_k)) + \sum_{a_j \in \phi^0(I_k)} v_j(I_k, \phi^0(I_k)) - \max_{B \subseteq \mathcal{A}} \left\{ u_k(B) + \sum_{a_j \in B} v_j(I_k, B) \right\} + \underline{\pi}_k(\mathcal{R}) \\ &\leq \underline{\pi}_k(\mathcal{R}). \end{aligned} \tag{A.12}$$

The first two equalities follow from the definition of $m^k$; the first inequality follows because $b_k > 0$; the second inequality follows $m^0$ is individually rational for agents and therefore $\zeta_{kj}^0 + v_j(I_k, \phi^0(I_k)) \geq 0$; the third equality follows from the definition of $b_k$; the last inequality follows by construction.

For each pair of $k \neq k' \in \kappa(\mathcal{R})$, $u_k(m^0) = u_k(m^{k'})$, while $u_{k'}(m^0) > u_{k'}(m^{k'})$. Therefore, for each pair $k \neq k' \in \kappa(\mathcal{R})$, there do not exist scalars $\alpha > 0$ and $\beta$ such that $u_k(m) = \alpha u_{k'}(m) + \beta$

for all $m \in M^\circ$. This verifies the NEU condition in Abreu, Dutta, and Smith (1994) for indices in $\kappa(\mathcal{R})$.

Lemma 1 and Lemma 2 in Abreu, Dutta, and Smith (1994) then ensure the existence of vectors $\{\widehat{u}^k : k \in \kappa(\mathcal{R})\} \subseteq \{\mathbb{E}_\lambda[u(m)] : \lambda \in \Delta(M^\circ)\}$ that satisfy $\widehat{u}^k_k < \widehat{u}^{k'}_k$ for all $k \neq k' \in \kappa(\mathcal{R})$.

For an arbitrary vector $u^0 \in U^\circ$, define

$$u^k = \varepsilon(1-\eta)u(m^k) + \eta\varepsilon\widehat{u}^k + (1-\varepsilon)u$$

for each $k \in \kappa(\mathcal{R})$.

First observed that by the construction of $\{m^k : k \in \{0\} \cup \kappa(\mathcal{R})\}$, $u_k(m^k) < u_k(m^0) = u_k(m^{k'})$, so $u^k_k < u^{k'}_k$ if $\varepsilon\eta > 0$. Second, since $\{m^k : k \in \kappa(\mathcal{R})\} \subseteq M^\circ$, $u^k \in \mathrm{co}(U_\mathcal{T})$ for all $0 < \varepsilon < 1$ and $0 < \eta < 1$. Third, for small enough $\varepsilon > 0$, $u^k_k > \underline{\pi}_k(\mathcal{R})$. Finally, by inequality (A.12), for small enough $\eta > 0$, it must be true that $u^k_k < u_k$. Therefore, there must exist $\{u^k : k \in \kappa(\mathcal{R})\} \subseteq U^\circ$ such that

$$u^k_k < u_k$$

for all $k \in \kappa(\mathcal{R})$, and

$$u^k_k < u^{k'}_k$$

for all $k \neq k' \in \kappa(\mathcal{R})$. □

**Lemma 21.** *If a matching $m = (\phi, \{\zeta_l\}_{l=1}^K) \in M$ is individually rational for all agents, then $u_k(m) \leq \widehat{\pi}_k$ for all $I_k \in I$*

*Proof.* Since $m$ is individually rational for agents, it must be that $v_j(I_k, \phi(I_k)) + \zeta_{kj} \geq 0$, or equivalently, $-\zeta_{kj} \leq +v_j(I_k, \phi(I_k))$ for all $a_j \in \phi(I_k)$. So

$$
\begin{aligned}
u_k(m) &= u_k(\phi(I_k)) - \sum_{a_j \in \phi(I_k)} \zeta_{kj} \\
&\leq u_k(\phi(I_k)) + \sum_{a_j \in \phi(I_k)} v_j(I_k, \phi(I_k)) \\
&\leq \max_{B \subseteq \mathcal{A}} u_k(B) + \sum_{a_j \in B} v_j(I_k, B) \\
&= \widehat{\pi}_k.
\end{aligned}
$$

$\square$

**Proof of Theorem 3.** There are three cases to consider.

**Case 1:** $\mathcal{R} = \emptyset$. By Lemma 18, all $\lambda \in \Lambda^\circ$ are lotteries over static stable matchings. The matching process $\mu^*$ defined by

$$\mu^*(\widehat{h}) = \lambda^0 \;\; \forall \widehat{h} \in \overline{\mathcal{H}}$$

is a self-enforcing matching process.

**Case 2:** $|\mathcal{R}| = 1$. Suppose $\mathcal{R} = \{I_k\}$. Fix any $\lambda^0 \in \Lambda^\circ$ and define $u^0 \equiv \mathbb{E}_{\lambda^0}[u(m)]$. Let $\underline{m}_k \in M^\circ$ be the static matching constructed in Lemma 19. Note that $\underline{m}_k \in M^\circ$, and

$$u_k(\underline{m}_k) \leq \underline{\pi}_k(\mathcal{R}) \tag{A.13}$$

Let $\rho \in (0,1)$ be such that

$$(1-\widetilde{\rho})u_k(\underline{m}_k) + \widetilde{\rho} u_k^0 > \underline{\pi}_k(\mathcal{R}), \tag{A.14}$$

for all $\widetilde{\rho} \in [\rho, 1]$. When the post-minmaxing phase payoffs are discounted at $\rho$, inequality (A.14) implies that compared to obtaining its minmax value forever, institution $I_k$ would rather live with the lower payoff $u_k(\underline{m}_k)$ during the minmax phase, with the promise of transitioning back into the lottery $\lambda^0$;

The above inequalities holds at $\rho = 1$, so there exists a value of $\rho \in (0,1)$ such that the inequality holds for all $\widetilde{\rho} \in [\rho, 1]$. Let $L(\delta) \equiv \left\lceil \frac{\log\rho}{\log\delta} \right\rceil$ where $\lceil\cdot\rceil$ is the ceiling function. Below, we use the property that $\lim_{\delta\to 1} \delta^{L(\delta)} = \rho$.[1]

By Lemma 9, it suffices to consider the matching process represented by the automaton $(\Theta, p^0, f, \gamma)$, where

- $\Theta = \{\theta(m) : m \in M^\circ\} \cup \{\underline{\theta}(t) : 0 \le t < L(\delta)\}$ is the set of all possible states;
- $p^0$ is the initial distribution over states, which satisfies $p^0(\theta(m)) = \lambda^0(m)$ for all $m \in M^\circ$;
- $f : \Theta \to M$ is the output function, where $f(\theta(m)) = m$ and $f(\underline{\theta}(t)) = \underline{m}_k$;
- $\gamma : \Theta \times M \to \Delta(\Theta)$ is the transition function. For states $\{\underline{\theta}(t) | 0 \le t < L(\delta) - 1\}$, $\gamma$ is defined as

$$\gamma\big(\underline{\theta}(t), m'\big) = \begin{cases} \underline{\theta}(0) & \text{if } m' \neq \underline{m}_k;\ m' = [\underline{m}_k, (I_k, A'_k, \zeta'_k)] \\ \underline{\theta}(t+1) & \text{otherwise} \end{cases}$$

For states $\underline{\theta}(L(\delta)-1)$, the transition is defined as

$$\gamma\big(\underline{\theta}(L(\delta)-1), m'\big) = \begin{cases} \underline{\theta}(0) & \text{if } m' \neq \underline{m}_k;\ m' = [\underline{m}_k, (I_k, A'_k, \zeta'_k)] \\ p^0 & \text{otherwise} \end{cases}$$

For states $\theta(m)$, the transition is

$$\gamma\big(\theta(m), m'\big) = \begin{cases} \underline{\theta}(0) & \text{if } m' \neq m;\ m' = [m, (I_k, A'_k, \zeta'_k)] \\ p^0 & \text{otherwise} \end{cases}$$

[1]To see why, observe that $\delta^{L(\delta)} \in [\delta^{\frac{\log\rho}{\log\delta}}, \delta^{\frac{\log\rho}{\log\delta}+1}] = [\delta\rho, \rho] \to \{\rho\}$ as $\delta \to 1$

Note that owing to the identifiability of deviating institution, for any $\theta \in \Theta$ and matching $m \neq f(\theta)$ which can (but not necessarily does) result from $I_k$'s deviation, we can uniquely identify $I_k$ as the deviating institution, so the transition above is well-defined. Any $m \neq f(\theta)$ that cannot possibly result from a deviation by $I_k$ is ignored by the transition.

The matching process represented by the above automaton randomizes over $M^\circ$ according to $\lambda^0$ in every period. It remains to check that it is self-enforcing, or equivalently,

1. every agent's payoff is greater than or equal to 0 in all automaton states,

2. no institution has any profitable one-shot deviation in any automaton state,

3. for every institution, the stage-game payoffs across all automaton states are bounded.

Point 3 above follows by the construction of $\lambda^0$. In every state $\theta \in \Theta$, the recommended assignment $f(\theta) \in M^\circ$ is individually rational for all agents, so point 1 follows as well. It remains to verify point 2.

For every state $\theta$, I use $U(\theta) = (U_1(\theta), \ldots, U_K(\theta))$ to denote the discounted expected payoff profile for the institutions in state $\theta$. By construction,

$$U_k(\theta(m)) = (1-\delta)u(m) + \delta u^0,$$

and

$$U_k(\underline{\theta}(t)) = (1-\delta^{L(\delta)-t})u(\underline{m}_k) + \delta^{L(\delta)-t}u^0$$

In addition, since $u_l(m) = \widehat{\pi}_l$ for every $I_l \neq I_k$ and all $m \in M^\circ$, the above equalities simplify to

$$U_l(\theta(m)) = U_l(\underline{\theta}(t)) = \widehat{\pi}_l \tag{A.15}$$

for every $I_l \neq I_k$.

I now verify no institution $I_{k'}$ has profitable one shot deviations in any automaton states.

For states $\{\theta(m) : m \in M^\circ\}$: there are two cases to consider.

Case a: $I_{k'} \neq I_k$. By A.15, the continuation value of $I_{k'}$ is identical across states. By Lemma 17, no feasible deviation for $I_{k'}$ can improve its stage-game payoff. $I_{k'}$ does not have any profitable one shot deviation.

Case b: $I_{k'} = I_k$. Without deviation, $I_k$ has value $(1-\delta)u_k(m) + \delta u_k^0$. After deviation, $I_k$ yields less than

$$(1-\delta)Z + \delta(1-\delta^{L(\delta)})u_k(\underline{m}_k) + \delta^{L(\delta)+1}u_k^0.$$

There is no profitable one-shot deviation for $I_k$ if

$$(1-\delta)u_k(m) + \delta u_k^0 \geq (1-\delta)Z + \delta(1-\delta^{L(\delta)})u_k(\underline{m}_k) + \delta^{L(\delta)+1}u_k^0.$$

The inequality is equivalent to

$$(1-\delta)(Z - u_k(m)) \leq \delta(1-\delta^{L(\delta)})[u_k^0 - u_k(\underline{m}_k)]$$

As $\delta \to 1$, the LHS converges to 0 while the RHS converges to $(1-\rho)(u_k^0 - u_k(\underline{m}_k))$. By construction, $0 < (1-\rho)(u_k^0 - u_k(\underline{m}_k))$, so such deviations are not profitable for $\delta$ high enough.

For states $\{\underline{\theta}(t) : 0 \leq t \leq L(\delta)-1\}$: there are two cases to consider.

Case a: $I_{k'} \neq I_k$. By A.15, the continuation value of $I_{k'}$ is identical across states. By Lemma 17, no feasible deviation for $I_{k'}$ can improve its stage-game payoff. $I_{k'}$ does not have any profitable one shot deviation.

Case b: $I_{k'} \in \mathcal{R}$, $k' = k$. Without deviation, institution $I_k$ has payoff

$$(1-\delta^{L(\delta)-t})u_k(\underline{m}_k) + \delta^{L(\delta)-t}u_k^0$$

With any deviation, $I_k$ has payoff less than

$$(1-\delta)\underline{\pi}_k(\mathcal{R})+\delta(1-\delta^{L(\delta)})u_k(\underline{m}_k)+\delta^{L(\delta)+1}u_k^0$$

There is no profitable one-shot deviation for $I_k$ if, for all $0 \le t \le L(\delta)$,

$$(1-\delta^{L(\delta)-t})u_k(\underline{m}_k)+\delta^{L(\delta)-t}u_k^0 \ge (1-\delta)\underline{\pi}_k(\mathcal{R})+\delta(1-\delta^{L(\delta)})u_k(\underline{m}_k)+\delta^{L(\delta)+1}u_k^0$$

Since $u_k(\underline{m}_k) \le \underline{\pi}(\mathcal{R}) < u_k^k$, the LHS of the above inequality is monotonically increasing in $t$. As a result, to show that the above inequality holds for all $0 \le t \le L(\delta)$, it is sufficient to show that it holds for $t = 0$.

When $t = 0$, the above inequality is equivalent to

$$(1-\delta^{L(\delta)})\underline{u}_k(\underline{m}_k)+\delta^{L(\delta)}u_k^0 \ge \underline{u}_k(\mathcal{R})$$

By inequality (A.14), this is satisfied for $\delta$ suffifiently close to 1.

**Case 3:** $|\mathcal{R}| \ge 2$. Fix any $\lambda^0 \in \Lambda^\circ$. Let $\{\underline{m}_k\}_{k\in\kappa(\mathcal{R})} \subseteq M^\circ$ be the static matchings as constructed in Lemma 19. Note that $\{\underline{m}_k\}_{k\in\kappa(\mathcal{R})} \subseteq M^\circ$, and

$$u_k(\underline{m}_k) \le \underline{\pi}_k(\mathcal{R}) \tag{A.16}$$

for $k \in \kappa(\mathcal{R})$. Define $u^0 \equiv \mathbb{E}_{\lambda^0}[u(m)]$. By Lemma 20, there exist payoff vectors $\{u^k\}_{k\in\kappa(\mathcal{R})} \subseteq U^\circ$ such that $u_k^k < u_k^0$ for all $k \in \kappa(\mathcal{R})$, and $u_k^k < u_k^{k'}$ for all $k \ne k'$, $k, k' \in \kappa(\mathcal{R})$. For each $k \in \kappa(\mathcal{R})$, let $\lambda^k \in \Lambda^\circ$ be the distribution over $M^\circ$ that give rise to the payoff vector $u^k$.

Given $\{u^k\}_{k\in\kappa(\mathcal{R})}$ and $\{\underline{m}_k\}_{k\in\kappa(\mathcal{R})}$, let $\rho \in (0,1)$ be such that

1. for every $k \in \kappa(\mathcal{R})$ and every $\widetilde{\rho} \in [\rho, 1]$,

$$(1-\widetilde{\rho})u_k(\underline{m}_k) + \widetilde{\rho}u_k^k > \underline{\pi}_k(\mathcal{R}), \tag{A.17}$$

and

2. for all $k, k' \in \kappa(\mathcal{R})$, $k' \neq k$, and every $\widetilde{\rho} \in [\rho, 1]$,

$$(1-\widetilde{\rho})u_k(\underline{m}_{k'}) + \widetilde{\rho}u_k^{k'} > (1-\widetilde{\rho})u_k(\underline{m}_k) + \widetilde{\rho}u_k^k. \tag{A.18}$$

When the post-minmaxing phase payoffs are discounted at $\rho$, inequality (A.17) implies that compared to obtaining its minmax value forever, institution $I_k$ would rather live with the lower payoff $u_k(\underline{m}_k)$ during the minmax phase, with the promise of transitioning into its institution specific punishment; inequality (A.18) implies that institution $I_{k'}$ is willing to bear the cost of minmaxing institution $I_k$ with the promise of transitioning into institution $I_k$'s specific punishment rather than its own.

The above inequalities holds at $\rho = 1$ for each $k$ and $k' \neq k$. Since the set of players is finite, there exists a value of $\rho \in (0,1)$ such that the inequality holds for all $\widetilde{\rho} \in [\rho, 1]$, $k, k' \in \kappa(\mathcal{R})$ and $k \neq k'$. Let $L(\delta) \equiv \left\lceil \frac{\log \rho}{\log \delta} \right\rceil$ where $\lceil \cdot \rceil$ is the ceiling function. Below, we use the property that $\lim_{\delta \to 1} \delta^{L(\delta)} = \rho$.[2]

By Lemma 9, it suffices to consider the matching process represented by the automaton $(\Theta, p^0, f, \gamma)$, where

- $\Theta = \left\{\theta(e,m) : e \in \kappa(\mathcal{R}) \cup \{0\}, m \in M^\circ\right\} \cup \left\{\underline{\theta}(k,t) : k \in \kappa(\mathcal{R}),\ 0 \leq t < L(\delta)\right\}$ is the set of all possible states;
- $p^0$ is the initial distribution over states, which satisfies $p^0(\theta(0,m)) = \lambda^0(m)$ for all $m \in M^\circ$;

[2]To see why, observe that $\delta^{L(\delta)} \in [\delta^{\frac{\log \rho}{\log \delta}}, \delta^{\frac{\log \rho}{\log \delta}+1}] = [\delta\rho, \rho] \to \{\rho\}$ as $\delta \to 1$

- $f : \Theta \to M$ is the output function, where $f(\theta(e,m)) = m$ and $f(\underline{\theta}(k,t)) = \underline{m}_k$;

- $\gamma : \Theta \times M \to \Delta(\Theta)$ is the transition function. For states $\{\underline{\theta}(k,t) | 0 \le t < L(\delta) - 1\}$, $\gamma$ is defined as

$$\gamma\big(\underline{\theta}(k,t), m'\big) = \begin{cases} \underline{\theta}(k',0) & \text{if } m' \neq \underline{m}_k;\ m' = [\underline{m}_{k'}, (I_{k'}, A'_{k'}, \zeta'_{k'})] \text{ for some} \\ & k' \in \kappa(\mathcal{R}) \text{ and } (A'_{k'}, \zeta'_{k'}) \in D_{k'}(\underline{m}_{k'}) \\ \underline{\theta}(k,t+1) & \text{otherwise} \end{cases}$$

For states $\underline{\theta}(k, L(\delta)-1)$, the transition is defined as

$$\gamma\big(\underline{\theta}(k,L(\delta)-1), m'\big) = \begin{cases} \underline{\theta}(k',0) & \text{if } m' \neq \underline{m}_k;\ m' = [\underline{m}_{k'}, (I_{k'}, A'_{k'}, \zeta'_{k'})] \text{ for some} \\ & k' \in \kappa(\mathcal{R}) \text{ and } (A'_{k'}, \zeta'_{k'}) \in D_{k'}(\underline{m}_{k'}) \\ p^k & \text{otherwise} \end{cases}$$

where $p^k$ is the distribution over states that satisfies $p^k(\theta(k,m)) = \lambda^k(m)$ for all $k \in \kappa(\mathcal{R})$ and $m \in M^\circ$.

For states $\theta(e,m)$, the transition is

$$\gamma\big(\theta(e,m), m'\big) = \begin{cases} \underline{\theta}(k',0) & \text{if } m' \neq m;\ m' = [m, (I_{k'}, A'_{k'}, \zeta'_{k'})] \text{ for some} \\ & k' \in \kappa(\mathcal{R}) \text{ and } (A'_{k'}, \zeta'_{k'}) \in D_{k'}(m) \\ p^e & \text{otherwise} \end{cases}$$

Note that owing to the identifiability of deviating institution, for any $\theta \in \Theta$ and matching $m \neq f(\theta)$ which can (but not necessarily does) result from an institution's deviation, we can uniquely identify the institution, so the transition above is well-defined. Any $m \neq f(\theta)$ that cannot possibly result from a deviation by an institution is ignored by the transition.

The matching process represented by the above automaton randomizes over $M^{\circ}$ according to $\lambda^0$ in every period. It remains to check that it is self-enforcing, or equivalently,

1. every agent's payoff is greater than or equal to 0 in all automaton states,

2. no institution has any profitable one-shot deviation in any automaton state,

3. for every institution, the stage-game payoffs across all automaton states are bounded.

Point 3 above follows by the construction of $\lambda^0$. In every state $\theta \in \Theta$, the recommended assignment $f(\theta) \in M^{\circ}$ is individually rational for all agents, so point 1 follows as well. It remains to verify point 2.

For every state $\theta$, I use $U(\theta) = (U_1(\theta), \dots, U_K(\theta))$ to denote the discounted expected payoff profile for the institutions in state $\theta$. By construction,

$$U(\theta(e,m)) = (1-\delta)u(m) + \delta u^e, \text{ for all } e \in \kappa(\mathcal{R}) \cup \{0\},$$

and

$$U(\underline{\theta}(k,t)) = (1-\delta^{L(\delta)-t})u(\underline{m}_k) + \delta^{L(\delta)-t}u^k, \text{ for all } k \in \kappa(\mathcal{R}), 0 \leq t \leq L(\delta)-1.$$

In addition, since $u_l(m) = \widehat{\pi}_l$ for every $I_l \in \cup_{g=1}^{G} \mathcal{P}_g$ and all $m \in M^{\circ}$, the above equalities simplify to

$$U_l(\theta(e,m)) = U_l(\underline{\theta}(k,t)) = \widehat{\pi}_l \tag{A.19}$$

for every $I_l \in \cup_{g=1}^{G} \mathcal{P}_g$.

I now verify no institution $I_{k'}$ has profitable one shot deviations in any automaton states.

<u>For states $\left\{\theta(e,m) : e \in \kappa(\mathcal{R}) \cup \{0\}, m \in M^{\circ}\right\}$</u>: there are three cases to consider.

<u>Case a:</u> $I_{k'} \in \cup_{g=1}^{G} \mathcal{P}_g$. By A.19, the continuation value of $I_{k'}$ is identical across states. By Lemma 17, no feasible deviation for $I_{k'}$ can improve its stage-game payoff. $I_{k'}$ does not have any

profitable one shot deviation.

Case b: $I_{k'} \in \mathcal{R}$, $k' \neq e$. Choose a number $Z > \sup_{m \in M^\circ} \sup_{m' \in D_k(m)} u_k(m')$. For any matching $m \in M^\circ$, the manipulated matching resulting from a feasible deviation must still be individually rational. So $Z$ is finite by Lemma 21, and is larger than the payoff any institution can obtain in any feasible deviation from a matching in $M^\circ$.

Consider a one-shot deviation $(I_{k'}, A'_{k'}, \zeta'_{k'})$ by institution $I_{k'}$. Without deviation, $I_{k'}$ has value $(1-\delta)u_{k'}(m) + \delta u^e_{k'}$. After deviation, $I_{k'}$ yields less than

$$(1-\delta)Z + \delta U_{k'}(\underline{\theta}(e,0)) = (1-\delta)Z + \delta(1-\delta^{L(\delta)})u_{k'}(\underline{m}_{k'}) + \delta^{L(\delta)+1}u^{k'}_{k'}$$

There is no profitable one-shot deviation for $I_{k'}$ if

$$(1-\delta)u_{k'}(m) + \delta u^e_{k'} \geq (1-\delta)Z + \delta(1-\delta^{L(\delta)})u_{k'}(\underline{m}_{k'}) + \delta^{L(\delta)+1}u^{k'}_{k'}$$

As $\delta \to 1$, the LHS converges to $u^e_{k'}$ while the RHS converges to $(1-\rho)u_{k'}(\underline{m}_{k'}) + \rho u^{k'}_{k'}$. By construction, $u^e_{k'} > u^{k'}_{k'}$ and $u^{k'}_{k'} > \underline{\pi}_{k'}(\mathcal{R})$ for all $k'$ such that $k' \in \kappa(\mathcal{R})$ and $k' \neq e$. From A.16, it follows that $u^e_{k'} > u^{k'}_{k'} > (1-\rho)\underline{\pi}_{k'}(\mathcal{R}) + \rho u^{k'}_{k'} \geq (1-\rho)u_{k'}(\underline{m}_{k'}) + \rho u^{k'}_{k'}$, so such deviations are not profitable for $\delta$ high enough.

Case c: $I_{k'} \in \mathcal{R}$, $k' = e$. Without deviation, $I_{k'}$ has value $(1-\delta)u_{k'}(m) + \delta u^{k'}_{k'}$. After deviation, $I_{k'}$ yields less than

$$(1-\delta)Z + \delta U_{k'}(\underline{\theta}(k',0)) = (1-\delta)Z + \delta(1-\delta^{L(\delta)})u_{k'}(\underline{m}_{k'}) + \delta^{L(\delta)+1}u^{k'}_{k'}.$$

There is no profitable one-shot deviation for $I_{k'}$ if

$$(1-\delta)u_{k'}(m) + \delta u^{k'}_{k'} \geq (1-\delta)Z + \delta(1-\delta^{L(\delta)})u_{k'}(\underline{m}_{k'}) + \delta^{L(\delta)+1}u^{k'}_{k'}.$$

The inequality is equivalent to

$$(1-\delta)(Z-u_{k'}(m)) \leq \delta(1-\delta^{L(\delta)})[u_{k'}^{k'}-u_{k'}(\underline{m}_{k'})]$$

As $\delta \to 1$, the LHS converges to 0 while the RHS converges to $(1-\rho)(u_{k'}^{k'}-u_{k'}(\underline{m}_{k'}))$. By construction, $0<(1-\rho)(u_{k'}^{k'}-u_{k'}(\underline{m}_{k'}))$ for $k \in \kappa(\mathcal{R})$, so such deviations are not profitable for $\delta$ high enough.

<u>For states $\{\underline{\theta}(k,t) : k \in \kappa(\mathcal{R}), 0 \leq t \leq L(\delta)-1\}$</u>: there are three cases to consider.

<u>Case a:</u> $I_{k'} \in \cup_{g=1}^{G} \mathcal{P}_g$. By A.19, the continuation value of $I_{k'}$ is identical across states. By Lemma 17, no feasible deviation for $I_{k'}$ can improve its stage-game payoff. $I_{k'}$ does not have any profitable one shot deviation.

<u>Case b:</u> $I_{k'} \in \mathcal{R}$, $k' \neq k$. Without deviation, institution $I_{k'}$ has payoff

$$(1-\delta^{L(\delta)-t})u_{k'}(\underline{m}_k)+\delta^{L(\delta)-t}u_{k'}^{k}$$

With any deviation, $I_{k'}$ has payoff less than

$$(1-\delta)Z+\delta(1-\delta^{L(\delta)})u_{k'}(\underline{m}_{k'})+\delta^{L(\delta)+1}u_{k'}^{k'}$$

There is no profitable one-shot deviation for $I_{k'}$ if

$$(1-\delta^{L(\delta)-t})u_{k'}(\underline{m}_k)+\delta^{L(\delta)-t}u_{k'}^{k} \geq (1-\delta)Z+\delta(1-\delta^{L(\delta)})u_{k'}(\underline{m}_{k'})+\delta^{L(\delta)+1}u_{k'}^{k'} \tag{A.20}$$

We prove that this inequality is satisfied if $\delta$ is sufficiently high. Examining the LHS of inequality (A.20), observe that for all $t$ such that $0 \le t \le L(\delta)$,

$$\lim_{\delta \to 1} \left[ (1 - \delta^{L(\delta)-t}) u_{k'}(\underline{m}_k) + \delta^{L(\delta)-t} u_{k'}^k \right] = \lim_{\delta \to 1} \left[ \left(1 - \frac{\rho}{\delta^t}\right) u_{k'}(\underline{m}_k) + \frac{\rho}{\delta^t} u_{k'}^k \right]$$
$$= (1 - \widetilde{\rho}) u_{k'}(\underline{m}_k) + \widetilde{\rho} u_{k'}^k$$

for some $\widetilde{\rho} \in [\rho, 1]$.[3]

Examining the RHS of inequality (A.20), observe that

$$\lim_{\delta \to 1} \left[ (1-\delta)Z + \delta(1 - \delta^{L(\delta)}) u_{k'}(\underline{m}_{k'}) + \delta^{L(\delta)+1} u_{k'}^{k'} \right] = \lim_{\delta \to 1} \left[ (1 - \delta^{L(\delta)}) u_{k'}(\underline{m}_{k'}) + \delta^{L(\delta)} u_{k'}^{k'} \right]$$
$$= (1-\rho) u_{k'}(\underline{m}_{k'}) + \rho u_{k'}^{k'}$$
$$\le (1-\widetilde{\rho}) u_{k'}(\underline{m}_{k'}) + \widetilde{\rho} u_{k'}^{k'},$$

where the first equality follows from taking limits, the second from $\lim_{\delta \to 1} \delta^{L(\delta)} = \rho$, and the weak inequality follows from $\widetilde{\rho} \ge \rho$ and $u_{k'}(\underline{m}_{k'}) \le \underline{\pi}_{k'}(\mathcal{R}) < u_{k'}^{k'}$. Since $\widetilde{\rho} \in [\rho, 1]$, inequality (A.18) delivers that $(1-\widetilde{\rho}) u_{k'}(\underline{m}_k) + \widetilde{\rho} u_{k'}^k$ is strictly higher than $(1-\widetilde{\rho}) u_{k'}(\underline{m}_{k'}) + \widetilde{\rho} u_{k'}^{k'}$. This guarantees that inequality (A.20) holds for sufficiently high $\delta$.

<u>Case c:</u> $I_{k'} \in \mathcal{R}$, $k' = k$. Without deviation, institution $I_k$ has payoff

$$(1 - \delta^{L(\delta)-t}) u_k(\underline{m}_k) + \delta^{L(\delta)-t} u_k^k$$

With any deviation, $I_k$ has payoff less than

$$(1-\delta)\underline{\pi}_k(\mathcal{R}) + \delta(1 - \delta^{L(\delta)}) u_k(\underline{m}_k) + \delta^{L(\delta)+1} u_k^k$$

[3] In the second equality, we use $\widetilde{\rho}$ rather than $\rho$ because $\tau$ is any integer between 0 and $L(\delta)$.

There is no profitable one-shot deviation for $I_k$ if, for all $0 \leq t \leq L(\delta)$,

$$(1-\delta^{L(\delta)-t})u_k(\underline{m}_k)+\delta^{L(\delta)-t}u_k^k \geq (1-\delta)\underline{\pi}_k(\mathcal{R})+\delta(1-\delta^{L(\delta)})u_k(\underline{m}_k)+\delta^{L(\delta)+1}u_k^k$$

Since $u_k(\underline{m}_k) \leq \underline{\pi}(\mathcal{R}) < u_k^k$, the LHS of the above inequality is monotonically increasing in $t$. As a result, to show that the above inequality holds for all $0 \leq t \leq L(\delta)$, it is sufficient to show that it holds for $t = 0$.

When $t = 0$, the above inequality is equivalent to

$$(1-\delta^{L(\delta)})\underline{u}_k(\underline{m}_k)+\delta^{L(\delta)}u_k^k \geq \underline{u}_k(\mathcal{R})$$

By inequality (A.17), this is satisfied for $\delta$ suffifiently close to 1. □

**Lemma 22.** *Let $\mu = (\psi, \{\xi_k\}_{k=1}^K)$ be a self-enforcing matching process, then $u_k(\mu) \leq \widehat{\pi}_k$ for all $I_k \in I$.*

*Proof.* By the one-shot deviation principle, $v_j(\mu(h)) \geq 0$ for all $a_j \in \mathcal{A}$, at every ex post history $h$. So by Lemma 21, $u_k(\mu(h)) \leq \widehat{\pi}_k$ at every $h \in \mathcal{H}$. Therefore, as the discounted sum of per-period utilities, $u_k(\mu) \leq \widehat{\pi}_k$ as well. □

**Lemma 23.** *Let $m = (\phi, \{\zeta_l\}_{l=1}^K)$ be a static matching that satisfies $v_j(m) \geq 0$ for all $a_j \in \mathcal{A}$. If*

$$u_k(m) = u_k(\phi(I_k)) + \sum_{a_j \in \phi(I_k)} v_j(I_k, \phi(I_k)),$$

*for some $I_k \in I$, then $v_j(m) = 0$ for all $a_j \in \phi(I_k)$.*

*Proof.* For every $a_j \in \phi(I_k)$, since $v_j(m) = v_j(I_k, \phi(I_k)) + \zeta_{kj} \geq 0$, we have $-\zeta_{kj} \leq v_j(I_k, \phi(I_k))$. Suppose by contradiction that $v_i(m) > 0$ for some $a_i \in \phi(I_k)$, then $-\zeta_{ki} < v_i(I_k, \phi(I_k))$. It follows that

$$\begin{aligned}
u_k(m) &= u_k(\phi(I_k)) + \sum_{a_j \in \phi(I_k)} \zeta_{kj} \\
&= u_k(\phi(I_k)) + \sum_{a_j \in \phi(I_k), j \neq i} \zeta_{kj} + \zeta_{ki} \\
&< u_k(\phi(I_k)) + \sum_{a_j \in \phi(I_k)} v_j(I_k, \phi(I_k)),
\end{aligned}$$

a contradiction. □

**Lemma 24.** *Let $M' \subseteq M$ be a subset of stage-game matchings, and $\mu : \mathcal{H} \to M'$ be a self-enforcing matching process. Fix an institution $I_k \in I$. If for every $m \in M'$, $I_k$ has a feasible deviation $(A_k^m, \zeta_k^m) \in D_k(m)$ that satisfies $u_k([m, (I_k, A_k^m, \zeta_k^m)]) \geq \underline{u}_k$, then*

$$u_k(\mu) \geq \underline{u}_k$$

*Proof.* Suppose by contradiction that $\mu$ is a self-enforcing matching process, but $u_k(\mu) < \underline{u}_k$. I will show that institution $I_k$ has a feasible profitable deviation plan from $\mu$, so $\mu$ must not be self-enforcing. Consider the following deviation plan $(d_k', \xi_k')$ from $\mu$: for every ex post history $h \in \mathcal{H}$, define

$$d_k'(h) = A_k^{\mu(h)} \text{ and } \xi_k'(h) = \zeta_k^{\mu}$$

By assumption, $\mu(h) \in M'$ for all $h \in \mathcal{H}$, so $(d_k', \xi_k')$ is well-defined and feasible. To see that $(d_k', \xi_k')$ is profitable, observe that at every ex post history $h$, institution $I_k$'s stage-game payoff from the manipulated static assignment $[\mu(h), (I_k, d_k'(h), \xi_k'(h))]$ is $u_k([\mu(h), (I_k, A_k^{\mu(h)}, \zeta_k^{\mu(h)})]) \geq \underline{u}_k > u_k(\mu)$. Since this is true for every ex post history $h$, $I_k$'s total discounted payoff from the deviation plan satisfies

$$u_k([\mu, (I_k, d_k', \xi_k')]) > u_k(\mu).$$

The deviation plan $(d_k', \xi_k')$ is both feasible and profitable for institution $I_k$, which is a contradiction to the self-enforcement of $\mu$. So $u_k(\mu) \geq \underline{u}_k$.

$\square$

# Appendix B

# Proofs in Chapter 2

*Proof of Theorem 5.* ($\Rightarrow$) First, we show that a nonseparable costly information representation satisfies NIAS. Fix $A \in \mathcal{D}$, $\pi_A \in \arg\max_{\pi \in \Pi} V(\pi \cdot f_A, \pi)$, and $C_A : \text{Supp}(\pi_A) \to \Delta(A)$ and $a \in \text{Supp}(P_A)$. By definition of a nonseparable costly information representation, we know that the $V(\pi_A \cdot f_A, \pi_A)$ is monotone in $\pi_A \cdot f_A$ and choices are optimal conditional on posteriors. Thus, if $a$ was chosen when $\gamma$ was realized, then the expected utility must be weakly higher for these $\gamma$. For $\gamma$ such that $C_A(a \mid \gamma) > 0$,

$$\sum_{\omega \in \Omega} \gamma(\omega) u(a(\omega)) \geq \sum_{\omega \in \Omega} \gamma(\omega) u(b(\omega)) \quad \forall b \in A.$$

The proof now follows from arguments in Caplin and Dean 2015 that are reproduced here for completeness. Recall that

$$\gamma(\omega) = \frac{\mu(\omega)\pi_A(\gamma \mid \omega)}{\sum_{\nu \in \Omega} \mu(\nu)\pi(\gamma \mid \nu)},$$

which can be substituted on both sides and the denominator cancels so

$$\sum_{\omega \in \Omega} \mu(\omega)\pi_A(\gamma \mid \omega) u(a(\omega)) \geq \sum_{\omega \in \Omega} \mu(\omega)\pi_A(\gamma \mid \omega) u(b(\omega)) \quad \forall b \in A.$$

Therefore,

$$\sum_{\gamma \in \mathrm{Supp}(\pi_A)} C_A(a \mid \gamma) \left[ \sum_{\omega \in \Omega} \mu(\omega) \pi_A(\gamma \mid \omega) u(a(\omega)) \right]$$
$$\geq \sum_{\gamma \in \mathrm{Supp}(\pi_A)} C_A(a \mid \gamma) \left[ \sum_{\omega \in \Omega} \mu(\omega) \pi_A(\gamma \mid \omega) u(b(\omega)) \right] \quad \forall b \in A$$

since $C_A(a \mid \gamma)$ are either zero or positive multiples of the earlier introduced inequalities. Next, recall from data matching that $P_A(a \mid \omega) = \sum_{\gamma \in \mathrm{Supp}(\pi_A)} \pi_A(\gamma \mid \omega) C_A(a \mid \gamma)$. Therefore, we see that

$$\begin{aligned}
\sum_{\omega \in \Omega} \mu(\omega) u(a(\omega)) P_A(a \mid \omega) &= \sum_{\omega \in \Omega} \mu(\omega) u(a(\omega)) \left[ \sum_{\gamma \in \mathrm{Supp}(\pi_A)} \pi_A(\gamma \mid \omega) C_A(a \mid \gamma) \right] \\
&= \sum_{\gamma \in \mathrm{Supp}(\pi_A)} C_A(a \mid \gamma) \left[ \sum_{\omega \in \Omega} \mu(\omega) u(a(\omega)) \pi_A(\gamma \mid \omega) \right] \\
&\geq \sum_{\gamma \in \mathrm{Supp}(\pi_A)} C_A(a \mid \gamma) \left[ \sum_{\omega \in \Omega} \mu(\omega) u(b(\omega)) \pi_A(\gamma \mid \omega) \right] \\
&= \sum_{\omega \in \Omega} \mu(\omega) u(b(\omega)) P_A(a \mid \omega)
\end{aligned}$$

where the first set of equalities follows from substitutions, the inequality follows from optimality conditional on $\gamma$, and the last equality follows from the same substitutions above. Rearranging this inequality shows that NIAS is satisfied.

Next, we show that a nonseparable costly information representation implies GACI. Observe $\arg\max_\pi V(\pi \cdot f_A, \pi) = V(\pi_A \cdot f_A, \pi_A)$ by definition. We first establish that $V(\pi_A \cdot f_A, \pi_A) > -\infty$ for all $A \in \mathcal{D}$. To see this, notice that for all $A \in \mathcal{D}$, $f_A$ is a continuous function on the compact set $\Gamma$, so $f_A$ achieves a minimum value $c_A$. By non-triviality, there exists $\pi_A^c$ such that $V(c_A, \pi_A^c) > -\infty$. Observe $\pi_A^c \cdot f_A \geq c_A$. By monotonicity, $V(\pi_A^c \cdot f_A, \pi_A^c) \geq V(c_A, \pi_A^c) > -\infty$. Since $\pi_A$ is the optimal choice, we have $V(\pi_A \cdot f_A, \pi_A) \geq V(c_A, \pi_A^c) > -\infty$.

Suppose without loss of generality that $\bar{\pi}_{A_i} \cdot f_{A_i} \leq \bar{\pi}_{A_i} \cdot f_{A_{i+1}}$ for $i = \{1, \ldots, k\}$ (with addition modulo $k$). It follows that

$$\begin{aligned} V(\pi_{A_i} \cdot f_{A_i}, \pi_{A_i}) &= V(\bar{\pi}_{A_i} \cdot f_{A_i}, \pi_{A_i}) \\ &\leq V(\bar{\pi}_{A_i} \cdot f_{A_{i+1}}, \pi_{A_i}) \\ &\leq V(\pi_{A_i} \cdot f_{A_{i+1}}, \pi_{A_i}) \leq V(\pi_{A_{i+1}} \cdot f_{A_{i+1}}, \pi_{A_{i+1}}) \end{aligned} \tag{B.1}$$

Since $V(\bar{\pi}_{A_i} \cdot f_{A_i}, \pi_{A_i}) = V(\pi_{A_i} \cdot f_{A_i}, \pi_{A_i}) > -\infty$ for all $i$, strict monotonicity in the first component of $V$ implies that the inequality in (B.1) is strict if $\bar{\pi}_{A_i} \cdot f_{A_i} < \bar{\pi}_{A_i} \cdot f_{A_{i+1}}$. Suppose there is a strict inequality in the sequence, then we obtain the contradiction $V(\pi_{A_1} \cdot f_{A_1}, \pi_{A_1}) < V(\pi_{A_1} \cdot f_{A_1}, \pi_{A_1})$. Consequently, we must have $\bar{\pi}_{A_i} \cdot f_{A_i} = \bar{\pi}_{A_i} \cdot f_{A_{i+1}}$ for all $i$.

($\Leftarrow$) The converse is a direct application of Afriat's Theorem. Let $\alpha_{A,B} = -\bar{\pi}_A \cdot (f_B - f_A)$ for all $(A,B) \in \mathcal{D}^2$. Observe that by GACI, the condition in Afriat's Theorem is satisfied. Conclude there is $U_A$ and $\lambda_A > 0$ such that for all $(A,B) \in \mathcal{D}^2$, $U_A \leq U_B - \lambda_B \bar{\pi}_B \cdot (f_A - f_B)$. Taking negatives and letting $\widetilde{U}_A = -U_A$, we have

$$\widetilde{U}_B + \lambda_B \bar{\pi}_B (f_A - f_B) \leq \widetilde{U}_A.$$

Most of the remaining construction follows Afriat's theorem directly. Let $C(\Gamma)$ be the set of continuous, convex functions on $\Gamma$. Define $U : C(\Gamma) \to \mathbb{R}$ by

$$U(f) = \max_{A \in \mathcal{D}} \widetilde{U}_A + \lambda_A \bar{\pi}_A \cdot (f - f_A)$$

Clearly, $U$ is convex, continuous, and monotone increasing[1] (as the maximum of a finite number of continuous affine functionals). For every $A \in \mathcal{D}$, $U(f_A) = \widetilde{U}_A$ by construction. Moreover, for every $A \in \mathcal{D}$, if $\bar{\pi}_A \cdot f \geq \bar{\pi}_A \cdot f_A$, then $U(f) \geq U(f_A)$, which is also straightforward by construction.

Define $V : \mathbb{R} \times \Pi \to \mathbb{R} \cup \{-\infty\}$ by $V(t, \pi) = \inf_{f \in C(\Gamma)} \{U(f) : \pi \cdot f \geq t\}$. Observe that

[1]The functional $U$ is monotone increasing in $f$ if $(f - g)(\gamma) > 0$ for all $\gamma \in \Gamma$, then $U(f) > U(g)$.

the monotonicity condition is trivially satisfied for fixed $\pi$, since a greater $t$ reduces the set of $f \in C(\Gamma)$ satisfying the inequality.

The assumption that for each $t \in \mathbb{R}$, there exists a $\pi_t \in \Pi$ such that $V(t, \pi_t) > -\infty$ is also satisfied. In fact, we will show $V(t, \bar{\pi}_A) > -\infty$ for any $t \in \mathbb{R}$ and $A \in \mathcal{D}$. For any $t \in \mathbb{R}$, let $G^t_- = \{g \in C(\Gamma) \mid U(g) \leq t\}$. $G^t_-$ is closed and convex by the continuity and convexity of $U(\cdot)$.

The set of continuous functions on $\gamma$, of which $C(\Gamma)$ is a subset, is the topological dual to the set of signed Borel measures with bounded variation over $\Gamma$ (Aliprantis and Border 2006 Theorem 14.15). Let $M(\Gamma)$ be the set of such measures on $\Gamma$.

Fix $\widehat{A} \in \mathcal{D}$. Note that for any $f \in G^t_-$ that

$$\widetilde{U}_{\widehat{A}} + \lambda_{\widehat{A}} \bar{\pi}_{\widehat{A}} \cdot (f - f_{\widehat{A}}) \leq U(f) = \max_{A \in \mathcal{D}} \widetilde{U}_A + \lambda_A \bar{\pi}_A \cdot (f - f_A) \leq t.$$

Rearranging the equation gives

$$\sup_{f \in G^t_-} \bar{\pi}_{\widehat{A}} \cdot f \leq \frac{t - \widetilde{U}_{\widehat{A}} + \lambda_{\widehat{A}} \bar{\pi}_{\widehat{A}} f_{\widehat{A}}}{\lambda_{\widehat{A}}}$$

Let $K(t) = \frac{t - \widetilde{U}_{\widehat{A}} + \lambda_{\widehat{A}} \bar{\pi}_{\widehat{A}} f_{\widehat{A}}}{\lambda_{\widehat{A}}}$. Note that the function $K$ is monotonically increasing with domain and range both spanning the reals. The function $K^{-1}$ is well-defined and monotonic, with $K^{-1}(x) > -\infty$ for all $x \in \mathbb{R}$.

It follows that $G^t_- \subseteq \{f \mid \bar{\pi}_{\widehat{A}} \cdot f \leq K(t)\}$. Note that for all $f$ such that $\bar{\pi}_{\widehat{A}} \cdot f \geq K(K^{-1}(t))$ it follows that, $U(f) \geq \widetilde{U}_{\widehat{A}} + \lambda_{\widehat{A}} \bar{\pi}_{\widehat{A}} \cdot (f - f_{\widehat{A}}) \geq K^{-1}(t)$. It follows by definition that for all $t \in \mathbb{R}$

$$V(t, \bar{\pi}_{\widehat{A}}) = \inf_{f \in C(\Gamma)} \{U(f) : \bar{\pi}_{\widehat{A}} \cdot f \geq t\} \geq K^{-1}(t) > -\infty.$$

We now assert that for all $A \in \mathcal{D}$, $\bar{\pi}_A \in \arg\max_{\pi \in \Pi} V(\pi \cdot f_A, \pi)$. First, from the monotonicity property of the $U$ function

$$V(\bar{\pi}_A \cdot f_A, \bar{\pi}_A) = \inf_{f \in C(\Gamma)} \{U(f) : \bar{\pi}_A \cdot f \geq \bar{\pi}_A \cdot f_A\}$$
$$= U(f_A)$$

Second, for any $\pi \in \Pi$, we have $V(\pi \cdot f_A, \pi) = \inf_{f \in C(\Gamma)}\{U(f) : \pi \cdot f \geq \pi \cdot f_A\} \leq U(f_A)$, since $\pi \cdot f_A \geq \pi \cdot f_A$. Therefore $V(\pi \cdot f_A, \pi) \leq V(\bar{\pi}_A \cdot f_A, \bar{\pi}_A)$ for all $\pi \in \Pi$. Therefore, the revealed information structure is optimal for $V$.[2]

The function

$$\widetilde{V}(t,\pi) = V(t,\pi) - V(0,\pi_0)$$

satisfies Condition 3, Condition 4, and Condition 5 while maintaining the other properties above. First, note that $\widetilde{V}(0,\pi_0) = V(0,\pi_0) - V(0,\pi_0) = 0$ so the normalization condition is satisfied.

Since the difference of $V$ and $\widetilde{V}$ is a constant, we can check quasiconcavity and weak monotonicity of $V$. Next, we check weak monotonicity. If $\pi$ is a garbling of $\rho$, then

$$V(t,\rho) = \inf_{f \in C(\Gamma)} \{U(f) \mid \rho \cdot f \geq t\}$$
$$\leq \inf_{f \in C(\gamma)} \{U(f) \mid \pi \cdot f \geq t\}$$
$$= V(t,\pi)$$

since $\pi \cdot f \geq t$ implies that $\rho \cdot f \geq t$ by Lemma 4 so the infimum is taken over a weakly smaller set of functions. Thus, weak monotonicity in the second argument of $V$ holds.

[2]We note that a version of Roy's identity holds (Roy (1947)). Observe that by definition of $V$, if $\pi \cdot f_A \geq w$ implies $U(f_A) \geq V(w,\pi)$. We conclude that $\pi \cdot f_A \geq \bar{\pi}_A \cdot f_A$ implies $U(f_A) \geq V(\bar{\pi}_A \cdot f_A, \pi)$. We have already shown that $U(f_A) = V(\bar{\pi}_A \cdot f_A, \bar{\pi}_A)$. Thus, if $\pi \cdot f_A \geq \bar{\pi}_A \cdot f_A$, then $V(\bar{\pi}_A \cdot f_A, \bar{\pi}_A) \geq V(\bar{\pi}_A \cdot f_A, \pi)$.

Lastly, we examine quasiconcavity of $V$. Let $(t_1,\pi_1),(t_2,\pi_2)\in\mathbb{R}\times\Pi$, then for $\lambda\in[0,1]$

$$V(\lambda t_1+(1-\lambda)t_2,\lambda\pi_1+(1-\lambda)\pi_2)=\inf_{f\in C(\Gamma)}\{U(f)\mid\lambda\pi_1\cdot f+(1-\lambda)\pi_2\cdot f\geq\lambda t_1+(1-\lambda)t_2\}.$$

Note that if $\lambda\pi_1\cdot f+(1-\lambda)\pi_2\cdot f\geq\lambda t_1+(1-\lambda)t_2$, then either $\pi_1\cdot f\geq t_1$ or $\pi_2\cdot f\geq t_2$. Therefore, for $f\in C(\Gamma)$ we have

$$\{f\mid\lambda\pi_1\cdot f+(1-\lambda)\pi_2\cdot f\geq\lambda t_1+(1-\lambda)t_2\}\subseteq\{f\mid\pi_1\cdot f\geq t_1\}\cup\{f\mid\pi_2\cdot f\geq t_2\}.$$

Therefore, the infimum of $U$ over the first set,$V(\lambda t_1+(1-\lambda)t_2,\lambda\pi_1+(1-\lambda)\pi_2)$, is greater than or equal to the infimum of $U$ over the second set, $\min\{V(t_1,\pi_1),V(t_2,\pi_2)\}$. Thus, quasiconcavity holds.

We now show data matching and choices are optimal by following Caplin and Dean 2015 and using NIAS. Next we show that there exists stochastic choice functions $\{C_A:\text{Supp}(\bar{\pi}_A)\to\Delta(A)\}_{A\in\mathcal{D}}$ that satisfy optimality and matches data.

For each $\gamma\in\text{Supp}(\bar{\pi}_A)$, define:

$$C_A(a\mid\gamma)=\begin{cases}\frac{P_A(a)}{\sum_{\{b\in A:\bar{\gamma}_A^b=\gamma\}}P_A(b)} & \text{if } \bar{\gamma}_A^a=\gamma\\ 0 & \text{otherwise}\end{cases}$$

where $P_A(a)=\sum_{\omega\in\Omega}P_A(a\mid\omega)\mu(\omega)$ is the unconditional probability of choosing action $a$ from decision problem $A$. Note the $C_A(a\mid\gamma)>0$ only if $\bar{\gamma}_A^a=\gamma$. The NIAS condition implies that

$$\sum_{\omega\in\Omega}\mu(\omega)P_A(a\mid\omega)u(a(\omega))\geq\sum_{\omega\in\Omega}\mu(\omega)P_A(b\mid\omega)u(b(\omega))$$
$$\Rightarrow\sum_{\omega\in\Omega}\bar{\gamma}_A^a(\omega)u(a(\omega))\geq\sum_{\omega\in\Omega}\bar{\gamma}_A^a(\omega)u(b(\omega))$$

The second line follows by dividing both sides by $P_A(a)$. Thus, NIAS ensures that the choices are optimal.

It remains to show that the data are matched. In other words, $P_A$ is generated from the information structure $\bar{\pi}_A$ and choices $C_A$. First, note that for any $b, b' \in A$ such that $\bar{\gamma}_A^b = \bar{\gamma}_A^{b'}$, implies that for any $\omega \in \Omega$ such that $\bar{\gamma}_A^b(\omega) > 0$, then

$$\frac{P_A(b \mid \omega)}{P_A(b' \mid \omega)} = \frac{P_A(b)}{P_A(b')}.$$

Thus, for every $\omega \in \Omega$ and $a \in A$ such that $P_A(a) > 0$, then

$$\begin{aligned}
\sum_{\gamma \in \mathrm{Supp}(\bar{\pi}_A)} \bar{\pi}_A(\gamma|\omega) C_A(a|\gamma) &= \bar{\pi}_A(\bar{\gamma}_A^a|\omega) C_A(a|\bar{\gamma}_A^a) \\
&= \sum_{\{c \in A : \bar{\gamma}_A^c = \bar{\gamma}_A^a\}} P_A(c|\omega) \frac{P_A(a)}{\sum_{\{b \in A | \bar{\gamma}_A^b = \bar{\gamma}_A^a\}} P_A(b)} \\
&= \sum_{\{c \in A | \bar{\gamma}_A^c = \bar{\gamma}_A^a\}} P_A(c|\omega) \frac{P_A(a|\omega)}{\sum_{\{b \in A | \bar{\gamma}_A^b = \bar{\gamma}_A^a\}} P_A(b|\omega)} \\
&= P_A(a|\omega).
\end{aligned}$$

Therefore, the data are matched. □

*Proof of Theorem 7.* We note that NIAS is equivalent to optimal choices and matched data. Therefore, we focus on non-triviality and optimal information.

($\Rightarrow$) Suppose the data is represented by a constrained costly information representation and for all $A \in \mathcal{D}$ that $\pi_A \in \arg\max_{\pi \in \Pi_c} \pi \cdot f_A$. Since the utility depends only on ex-ante payoff, then $\pi_A \cdot f_A = \bar{\pi}_A \cdot f_A \geq \pi_B \cdot f_A \geq \bar{\pi}_B \cdot f_A$. The first equality follows from equivalent choices, the next inequality follows from optimality, while the final inequality follows Lemma 4.

($\Leftarrow$) Suppose BACI holds. Let $\bar{\Pi}_c = \bigcup_{A \in \mathcal{D}} \{\bar{\pi}_A\}$. For $\mathcal{D}$ nonempty, $\Pi_c \neq \emptyset$.[3] Moreover,

[3] If $\mathcal{D} = \emptyset$, then let $\Pi_c = \Pi$.

for any $A, B \in \mathcal{D}$, we have

$$\bar{\pi}_A \cdot f_A \geq \bar{\pi}_B \cdot f_A.$$

In other words, for all $A \in \mathcal{D}$ we have $\bar{\pi}_A \in \arg\max_{\pi \in \bar{\Pi}_c} \pi \cdot f_A$. Therefore nontriviality and optimal information hold.

Let $\text{conv}(\bar{\Pi}_c) = \text{conv}\left(\bigcup_{A \in \mathcal{D}} \{\bar{\pi}_A\}\right)$. Here $\text{conv}(\cdot)$ represents the convex hull of information structures. For all $B \in \mathcal{D}$ let $\lambda_B \in [0,1]$ such that $\sum_{B \in \mathcal{D}} \lambda_B = 1$. Now for fixed $A \in \mathcal{D}$

$$\sum_{B \in \mathcal{D}} \lambda_B \bar{\pi}_B \cdot f_A \leq \sum_{B \in \mathcal{D}} \lambda_B \bar{\pi}_A \cdot f_A = \bar{\pi}_A \cdot f_A$$

where the inequality follows from BACI. The result holds for any fixed $A$ and convex combination so that $\bar{\pi}_A \in \arg\max_{\pi \in \text{conv}(\bar{\Pi}_c)} \pi \cdot f_A$. Thus, the constraint set can be chosen convex without loss of generality. $\square$

*Proof of Theorem 6.* We note that NIAS is equivalent to optimal choices and matched data. Therefore, we focus on non-triviality and optimal information.

($\Rightarrow$) We show that a multiplicatively costly information representation satisfies HACI. For all $A \in \mathcal{D}$, let $\pi_A \in \arg\max_{\pi \in \Pi} [R(\pi)(\pi \cdot f_A)]$.

First, we show for $A \in \mathcal{D} \setminus \mathcal{D}_0$, that $(\pi \cdot f_A) > 0$ for all information structures and $R(\pi_A) > 0$. Let $\pi_0$ denote the non-informative information structure with $\pi_0(\mu) = 1$. By assumption, $\pi_0 \cdot f_A > 0$ for any $A \in \mathcal{D} \setminus \mathcal{D}_0$. Since $f_A$ is convex, $\pi \cdot f_A > 0$ for all information structures. In particular, let $\pi' \in \Pi$ be an information structure such that $R(\pi') > 0$, then $\pi' \cdot f_A > 0$ as well. Note that such a $\pi' \in \Pi$ with $R(\pi') > 0$ exists by nontriviality. For any $A \in \mathcal{D} \setminus \mathcal{D}_0$ and for any $\pi \in \Pi$, we have $R(\pi_A)(\pi_A \cdot f_A) \geq R(\pi')(\pi' \cdot f_A) > 0$ since $\pi_A$ is the optimal information structure. Therefore, for all $A \in \mathcal{D} \setminus \mathcal{D}_0$ we have $R(\pi_A) > 0$.

Next, for any pair $A_i, A_{i+1} \in \mathcal{D} \backslash \mathcal{D}_0$, we have

$$\begin{aligned} R(\pi_{A_i})(\bar{\pi}_{A_i} \cdot f_{A_i}) &= R(\pi_{A_i})(\pi_{A_i} \cdot f_{A_i}) \\ &\geq R(\pi_{A_{i+1}})(\pi_{A_{i+1}} \cdot f_{A_i}) \\ &\geq R(\pi_{A_{i+1}})(\bar{\pi}_{A_{i+1}} \cdot f_{A_i}) > 0 \end{aligned}$$

where the equality follows from equivalent choices, the first inequality follows from optimality, the second inequality follows from Lemma 4, and the last term is greater than zero by the earlier arguments. Rearranging the end terms of the inequalities,

$$\frac{R(\pi_{A_{i+1}})(\bar{\pi}_{A_{i+1}} \cdot f_{A_i})}{R(\pi_{A_i})(\bar{\pi}_{A_i} \cdot f_{A_i})} \leq 1.$$

We can now take any cycle $A_1, \ldots, A_k \in \mathcal{D} \backslash \mathcal{D}_0$ and take products to see that costs will be removed so

$$\prod_{i=1}^{k} \frac{R(\pi_{A_{i+1}})(\bar{\pi}_{A_{i+1}} \cdot f_{A_i})}{R(\pi_{A_i})(\bar{\pi}_{A_i} \cdot f_{A_i})} = \prod_{i=1}^{k} \frac{(\bar{\pi}_{A_{i+1}} \cdot f_{A_i})}{(\bar{\pi}_{A_i} \cdot f_{A_i})} \leq 1$$

where the indices are calculated with addition modulo $k$. Let $\sigma : \{1, \ldots, k, \} \to \{1, \ldots, k\}$ where $\sigma(1) = 1, \sigma(2) = k, \sigma(3) = k-1, \ldots, \sigma(k) = 2$.[4] Therefore,

$$\prod_{i=1}^{k} \frac{\bar{\pi}_{A_{\sigma(i)}} \cdot f_{A_{\sigma(i+1)}}}{\bar{\pi}_{A_{\sigma(i)}} \cdot f_{A_{\sigma(i)}}} \leq 1$$

and HACI is satisfied.

($\Leftarrow$) Now we show from HACI that we can generate a non-trivial utility function. Following Varian 1983a, for all $A \in \mathcal{D}$ let $U_A$ be the maximum of

$$\prod_{i=1}^{k-1} \frac{\bar{\pi}_{A_i} \cdot f_{A_{i+1}}}{\bar{\pi}_{A_i} \cdot f_{A_i}} \tag{B.2}$$

[4]Note addition is still modulo $k$ in the index so $\sigma(k+1) = 1$.

where the maximization is taken over all finite sequences $\{A_i\}_{i=1}^{k-1} \subseteq \mathcal{D} \setminus \mathcal{D}_0$ with $A_k = A$. Note that if $A \in \mathcal{D}_0$ then $U_A = 0$. Since the number of menus in $\mathcal{D} \setminus \mathcal{D}_0$ is finite, the number of sequences $\{A_i\}_{i=1}^{k-1}$ not containing cycles is also finite. Moreover by HACI, the presence of any cycles in a sequence $\{A_i\}_{i=1}^{k-1}$ only decreases the value of (B.2). Therefore the maximum in (B.2) exists for each $A$. Note that $U_A > 0$ for all $A \in \mathcal{D} \setminus \mathcal{D}_0$, and for all $A, B \in \mathcal{D}$

$$U_B \geq U_A \frac{\bar{\pi}_A \cdot f_B}{\bar{\pi}_A \cdot f_A} \tag{B.3}$$

by definition.[5] Define

$$U(f) = \begin{cases} \max_{A \in \mathcal{D}} \left[ U_A \frac{\bar{\pi}_A \cdot f}{\bar{\pi}_A \cdot f_A} \right] & \text{if } f \in C_+(\Gamma) \\ +\infty & \text{otherwise} \end{cases}$$

where $C_+(\Gamma)$ are nonnegative convex continuous functions on $\Gamma$. From the definition of $U$, it is obvious that $U(\cdot)$ is homogenous of degree 1 (as the supremum of a finite number of linear functionals), and $U(f) \geq 0$ for all $f \in C(\Gamma)$. In addition, inequality (B.3) implies that for all $A \in \mathcal{D}$ that $U(f_A) = U_A$. It is also straightforward that $U$ is convex, continuous, and monotone increasing over $C_+(\Gamma)$. Finally, we have

$$U(f) \geq U(f_A) \text{ if } \bar{\pi}_A \cdot f \geq \bar{\pi}_A \cdot f_A \tag{B.4}$$

which is also straightforward by construction.

Let $M_+(\Gamma)$ be the set of non-negative Borel measures over $\Gamma$ with bounded variation. Define $V : \mathbb{R}_+ \times M_+(\Gamma) \to \mathbb{R}_+$ by $V(t, m) = \inf_{f \in C(\Gamma)} \{U(f) : m \cdot f \geq t\}$. Now, we show that

[5] We define $0 \cdot \infty = 0$ as is standard in convex analysis.

$V(\cdot,\cdot)$ is indeed of the multiplicative form. By the definition of $V$, for any $t > 0$ we have

$$
\begin{aligned}
V(t,m) &= \inf_{f \in C(\Gamma)} \{U(f) : m \cdot \frac{f}{t} \geq 1\} \\
&= \inf_{tf' \in C(\Gamma)} \{U(tf') : m \cdot f' \geq 1\} \\
&= \inf_{f' \in C(\Gamma)} \{U(tf') : m \cdot f' \geq 1\} \\
&= t \inf_{f' \in C(\Gamma)} \{U(f') : m \cdot f' \geq 1\} \\
&= t\bar{R}(m)
\end{aligned}
$$

where the first equality comes from rearrangement, the second equality comes from $f' = f/t$, the third equality comes since any $tf'$ can be expressed as a function, the fourth equality holds since $U$ is homogeneous degree 1, and the final equality holds by defining the function $\bar{R} : M_+(\Gamma) \to \mathbb{R}_+$ as $\bar{R}(m) = \inf_{f \in C(\Gamma)}\{U(f) : m \cdot f \geq 1\}$.

Next, if $t = 0$ then $V(t,m) = 0$ which is consistent with the multiplicative form. To see this, consider the constant function $f_0(\gamma) = 0$ for all $\gamma \in \Gamma$ and see that $V(0,\pi) \leq U(f_0) = 0$. Let $\widetilde{R} : \Pi \to \mathbb{R}_+$ be the restriction of $\bar{R}$ to $\Pi$.

Since $U(f) \geq 0$ for all $f \in C(\Gamma)$, we have $\widetilde{R}(\pi) = \inf_{f \in C(\Gamma)}\{U(f) : \pi \cdot f \geq 1\} \geq 0$. Moreover, we show that $\widetilde{R}(\pi) < \infty$. Consider the constant function $f_1(\gamma) = 1$ for all $\gamma \in \Gamma$ so that $\pi \cdot f_1 = 1$. Therefore, we deduce

$$
\widetilde{R}(\pi) \leq \max_{A \in \mathcal{D}} \frac{U_A}{\bar{\pi}_A \cdot f_A} < \infty.
$$

We also prove that there are $\pi \in \Pi$ such that $\widetilde{R} > 0$. For an arbitrary $A \in \mathcal{D}\backslash\mathcal{D}_0$, we have

$$\begin{aligned}
\widetilde{R}(\bar{\pi}_A) &= \frac{V(\bar{\pi}_A \cdot f_A, \bar{\pi}_A)}{\bar{\pi}_A \cdot f_A} \\
&= \frac{\inf_{f \in C(\Gamma)}\{U(f) : \bar{\pi}_A \cdot f \geq \bar{\pi}_A \cdot f_A\}}{\bar{\pi}_A \cdot f_A} \\
&= \frac{U(f_A)}{\bar{\pi}_A \cdot f_A} = \frac{U_A}{\bar{\pi}_A \cdot f_A} > 0
\end{aligned}$$

where from definitions and $\bar{\pi}_A \cdot f_A > 0$. Therefore, $\widetilde{R}(\pi_A) > 0$.

We note that if $\pi$ is a garbling of $\rho$ then $\widetilde{R}(\rho) \leq \widetilde{R}(\pi)$ since $\widetilde{R}(\pi) = \inf_{f \in C(\Gamma)}\{U(f) \mid \pi \cdot f \geq 1\}$ and $\pi \cdot f \geq 1$ implies $\rho \cdot f \geq 1$ so the infimum is over a weakly larger set. Let $\pi_0$ as the information structure with $\pi_0(\mu|\omega) = 1$ for all $\omega \in \Omega$. Since $\Pi$ is the set of information sets consistent with Bayes' Law, $\pi_0$ is a garbling of any $\pi \in \Pi$. Thus, for all $\pi \in \Pi$, $\widetilde{R}(\pi_0) \geq \widetilde{R}(\pi) > 0$. Lastly, rescale the function $\widetilde{R}(.)$ with $1/\widetilde{R}(\pi_0)$, and define

$$R(\pi) = \frac{\widetilde{R}(\pi)}{\widetilde{R}(\pi_0)}.$$

We now assert that for all $A \in \mathcal{D}$, $\bar{\pi}_A \in \arg\max_{m \in M_+(\Gamma)} V(\pi \cdot f_A, \pi)$. First, from inequality (B.4) we have

$$\begin{aligned}
V(\bar{\pi}_A \cdot f_A, \bar{\pi}_A) &= \inf_{f \in C(\Gamma)}\{U(f) : \bar{\pi}_A \cdot f \geq \bar{\pi}_A \cdot f_A\} \\
&= U(f_A)
\end{aligned}$$

Second, for any $m \in M_+(\Gamma)$, we have $V(m \cdot f_A, m) = \inf_{f \in C(\Gamma)}\{U(f) : m \cdot f \geq m \cdot f_A\} \leq U(f_A)$, since $m \cdot f_A \geq m \cdot f_A$. Therefore $V(m \cdot f_A, m) \leq V(\bar{\pi}_A \cdot f_A, \bar{\pi}_A)$ for all $m \in M_+(\Gamma)$. From this, we have that

$$\bar{\pi}_A \in \arg\max_{\pi\in\Pi} \frac{V(\pi\cdot f_A,\pi)}{\widetilde{R}(\pi_0)} = \arg\max_{\pi\in\Pi} \frac{\widetilde{R}(\pi)}{\widetilde{R}(\pi_0)}(\pi\cdot f_A)$$
$$= \arg\max_{\pi\in\Pi} R(\pi)(\pi\cdot f_A)$$

where $\bar{\pi}_A$ is am optimizer since $\bar{\pi}_A$ is optimal for $V$ over a larger set and this is a positive scaling of $V$. Therefore $\pi_A$ is optimal for the rescaled $V$ and has the multiplicative costly representation.

We note that the $R$ was already shown to satisfy weak monotonicity in information and the normalization property. The $\widetilde{R}(m)$ defined in Theorem 6 is homogenous of degree one, increasing in $m$, and quasiconcave by the same arguments used in Theorem 5. By Theorem 1 in Prada 2011, we have that $\bar{R}$ is concave. Therefore, $\widetilde{R}$ restricted to $\Pi$ is the restriction of $\bar{R}$ to a convex set and is thus concave. Finally, $R$ is concave as it is a positive re-scaling. □

*Proof of Proposition 1.* Suppose there is a sequence $(\bar{\pi}_{A_1}, f_{A_1}), \ldots, (\bar{\pi}_{A_k}, f_{A_k})$ such that $\bar{\pi}_{A_i}\cdot f_{A_i} \leq \bar{\pi}_{A_i}\cdot f_{A_{i+1}}$ for all $i$ (with addition modulo $k$). Let $\sigma : \{1,\ldots,k,\} \to \{1,\ldots,k\}$ where $\sigma(1) = 1, \sigma(2) = k, \sigma(3) = k-1, \ldots, \sigma(k) = 2$. Note that addition is still modulo $k$ in the original indexing so $\sigma(k+1) = 1$.

By NIAC, we have

$$\sum_{i=1}^{k} \bar{\pi}_{A_{\sigma(i)}}\cdot f_{A_{\sigma(i)}} \geq \sum_{i=1}^{k} \bar{\pi}_{A_{\sigma(i+1)}}\cdot f_{A_{\sigma(i)}}.$$

However, by since this is a potential GACI cycle we know that each $\bar{\pi}_{A_{\sigma(i)}}\cdot f_{A_{\sigma(i)}} \leq \bar{\pi}_{A_{\sigma(i+1)}}\cdot f_{A_{\sigma(i)}}$. The only way both can hold simultaneously is if $\bar{\pi}_{A_i}\cdot f_{A_i} = \bar{\pi}_{A_i}\cdot f_{A_{i+1}}$ for $i = 1,\ldots,k$. □

*Proof of Proposition 2.* To satisfy NIAC it is required that

$$\bar{\pi}_A\cdot f_A + \bar{\pi}_B\cdot f_B \geq \bar{\pi}_A\cdot f_B + \bar{\pi}_B\cdot f_A$$

or equivalently,

$$\bar{\pi}_B \cdot f_A - \bar{\pi}_A \cdot f_A \leq \bar{\pi}_B \cdot f_B - \bar{\pi}_A \cdot f_B$$

However, since menu $A$ provides a higher return to information than menu $B$ and $\bar{\pi}_A$ is a garbling of $\bar{\pi}_B$, then

$$\bar{\pi}_B \cdot f_A - \bar{\pi}_A \cdot f_A > \bar{\pi}_B \cdot f_B - \bar{\pi}_A \cdot f_B$$

which violates NIAC. □

*Proof of Proposition 3.* Note that GACI is violated for a dataset of two menus if and only if

$$\bar{\pi}_A \cdot f_A \leq \bar{\pi}_A f_B \quad \text{and} \quad \bar{\pi}_B \cdot f_B \leq \bar{\pi}_B f_A$$

with one inequality strict. Since $f_A > f_B$, it follows that

$$\bar{\pi}_A \cdot f_A > \bar{\pi}_A f_B$$

and there can be no violation of GACI. Since the dataset was assumed to satisfy NIAS, the data is rationalized by a nonseparable costly information representation. □

*Proof of Proposition 4.* For any sequence $(\bar{\pi}_{A_1}, f_{A_1}), \ldots, (\bar{\pi}_{A_k}, f_{A_k})$ with $A_i \in \mathcal{D}$. Note that $\bar{\pi}_{A_i} \cdot f_{A_i} = \bar{\pi}_{A_i} \cdot f_A + c_{m_i}$ and $\bar{\pi}_{A_{i+1}} \cdot f_{A_i} = \bar{\pi}_{A_{i+1}} \cdot f_A + c_{m_i}$ for some $m_i \in \{1, \ldots, M\}$. This implies that

$$\sum_{i=1}^{k} \bar{\pi}_{A_i} \cdot f_{A_i} = \sum_{i=1}^{k} \bar{\pi}_{A_i} \cdot f_A + c_{m_i} = \sum_{i=1}^{k} \bar{\pi}_{A_{i+1}} \cdot f_A + c_{m_i} = \sum_{i=1}^{k} \bar{\pi}_{A_{i+1}} \cdot f_{A_i}$$

where addition of the index is modulo $k$. Therefore, NIAC is satisfied in addition to NIAS and the dataset is rationalized by the additive costly information representation. □

*Proof of Proposition 5.* For any sequences $(\bar{\pi}_{A_1}, f_{A_1}), \ldots, (\bar{\pi}_{A_k}, f_{A_k})$ with $A_i \in \mathcal{D} \backslash \mathcal{D}_0$. Note that $\bar{\pi}_{A_i} \cdot f_{A_i} = c_{m_i} \bar{\pi}_{A_i} \cdot f_A$ and $\bar{\pi}_{A_i} \cdot f_{A_{i+1}} = c_{m_{i+1}} \bar{\pi}_{A_i} \cdot f_A$ for some $m_i, m_{i+1} \in \{1, \ldots, M\}$. This implies that

$$\prod_{i=1}^{k} \frac{\bar{\pi}_{A_i} \cdot f_{A_{i+1}}}{\bar{\pi}_{A_i} \cdot f_{A_i}} = \prod_{i=1}^{k} \frac{c_{m_i}}{c_{m_{i+1}}} = 1$$

since addition of the index is modulo $k$ and each $c_{m_i}$ term appears in the numerator and denominator. Therefore, HACI is satisfied in addition to NIAS and the dataset is rationalized by the multiplicative costly information representation. $\square$