DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Essays on Microeconomic Theory

Citation
Romm, Assaf. 2015. Essays on Microeconomic Theory. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:17467296

Terms of Use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Essays on Microeconomic Theory

A dissertation presented

by

Assaf Romm

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Business Economics

Harvard University

Cambridge, Massachusetts

March 2015

© 2015 Assaf Romm

All rights reserved.

Dissertation Advisors:
Professor Alvin E. Roth
Professor Drew Fudenberg

Author:
Assaf Romm

Essays on Microeconomic Theory

Abstract

This thesis contains three chapters related to the microeconomic interactions in markets. The first paper deals with markets with many participants, and in which monetary transfers are allowed, and studies core convergence. The second paper considers reputation building in time-limited negotiations. The third paper studies two-sided markets with no monetary transfers, governed by stable matching mechanisms.

Contents

List of Tables

List of Figures

Acknowledgments

The author thanks Itai Ashlagi, Elchanan Ben-Porath, Holke Brammer, Peter Coles, Mira Frick, Drew Fudenberg, Jerry Green, Sergiu Hart, Avintan Hassidim, Johannes Hörner, Yuhta Ishii, Sangram Kadam, Scott D. Kominers, Yuichiro Kamada, David Laibson, Jacob Leshno, Déborah Marciano-Romm, Eric Maskin, Morgan McClellon, Motty Perry, Alvin E. Roth, Ran Shorrer, Elyashiv Wiedman, and seminar and conference participants at Bar-Ilan University, Ben-Gurion University, Haifa University, Harvard University, the Hebrew University, IDC Herzliya, London Business School, the Technion, and Tel-Aviv University, for helpful discussions and suggestions. The author also thanks the Hebrew University department of economics and the Center for the Study of Rationality for their hospitality.

To Dvorah, my loving wife, for her continuous support,

Introduction

This thesis is about markets and analyzing them at different levels. Each chapter of the thesis represents an approach that is useful under certain information constraints, but they do have a shared goal which is determining who gets what in the market. Due to the different environments considered, the specificity of the answers also varies and ranges from a general statement such as "the short side is likely to gain more under a core allocation" to detailed predictions such as "Player 1 gets a payoff of x with probability that tends to 1 under any Sequential equilibrium".

The first chapter ("An Approximate Law of One Price in Random Assignment Games") takes the broadest perspective by looking at markets from an ex-ante point of view, before players' preferences and tastes are known to anyone. At this stage, only very limited information is available to the modeler, and therefore the solution concept of choice is the core. Not only that, but due to the stochastic nature of the environment, the only meaningful predictions are probabilistic. In this transferable utility setting we show that as the market gets larger, the salaries that workers get become more and more similar, as do the payoffs of firms (under any core allocation). This property implies core shrinkage in both balanced and unbalanced markets, and provides a strong prediction on surplus distribution in unbalanced markets.

The second chapter ("Building Reputation at the Edge of the Cliff") is on the other extreme, as it assumes complete knowledge of players' payoffs and strategic situation. The only uncertainty in this model is related to whether a player is a commitment type or not. These commitment types are used as a technical tool to explain reputation formation in a simplified two-player interaction with a deadline. The main prediction in this context is a strong deadline effect, i.e., players tend to delay their interaction until very near the deadline, and this delay may often lead to inefficient outcomes (for example, a negotiation that ends with a disagreement when the deadline is hit).

The third chapter ("Implications of Capacity Reduction and Entry in Many-to-One Stable Matching") takes the middle road by considering two-sided many-to-one matching markets without transfers in which players' identities are already known (including their preferences on being matched with agents from the other side of the market). Nevertheless, players are not taken to be strategic agents, but rather the set of stable matchings is examined under different assumptions on players' participation in the market. For example, it is shown how to identify a set of players that become strictly better off following entry on the other side of the market, regardless of the stable matching that is being selected before and after the entry takes places. Similar results are established for the case of capacity reduction and using a truncation strategy. These results extend known result for the one-to-one case.

Chapter 1

An Approximate Law of One Price in Random Assignment Games[1]

1.1 Introduction

The "law of one price" asserts that homogeneous goods must sell for the same price across locations and vendors. This basic postulate is assumed in much of the economic literature, and its origins can be traced to Adam Smith's discussion on arbitrage (Smith, 1776, e.g., Book I, Chapter V). While many (sometimes consistent) deviations from this "law" have been observed and documented in the real world (see, for example, Lamont and Thaler, 2003, and references therein), it remains an interesting and useful building block in economic theory, and serves as a benchmark for empirical studies. A crucial underlying assumption used in arguing for the validity of the law of one price is the homogeneity of goods and buyers: buyers do not care which of the goods they end up buying, or which seller they are buying it from, nor do sellers care about the identity of the buyers. In other words, any two instances of the good are perfect (or at least near-perfect) substitutes for the buyers, as are any two buyers from any seller's point of view.

However, there are many markets in which the assumption of homogeneity is highly

[1]Co-authored with Dr. Avinatan Hassidim, Bar Ilan University, Department of Computer Science

implausible. For example, in labor markets there are some workers are skilled and some unskilled, and similarly some firms are generally considered better places to work. In addition to these measurable quality differences, workers may exhibit heterogeneous preferences over being employed by different firms, due to personal likes and dislikes, location, values, and a variety of other individually determined factors. Firms may also have diverse preferences over workers, and may, for example, favor workers who seem to share their vision or fit well within their corporate culture. Similarly, in markets where buyers and sellers have heterogeneous preferences over trading with the other side, the law of one price generally should not hold.

This paper makes the formal claim that even in the presence of heterogeneous preferences, an approximate version of the law remains valid, and the approximation improves as the market grows large. We focus on labor markets as our leading example, and argue that a likely outcome of the market is that workers who are roughly equally skilled receive similar wages, and firms of similar quality garner similar profits. Because of the inherent heterogeneity in firms' and workers' preferences, the law of one price holds only approximately, with some workers being paid more than their peers with identical levels of human capital.

To prove this result we use the assignment game model of Shapley and Shubik (1971) in which there is a finite set of firms and a finite set of workers, and each firm is looking to hire exactly one worker in exchange for a negotiable salary. Each firm has a (possibly different) value for hiring each of the workers, and each worker has a (possibly different) reservation value for working for each of the firms, and utilities are assumed to be linear in money. Since transfers are freely allowed, we can describe the net productivity of each firm-worker pair by a single number, and we assume that this productivity is separable in the firm's quality, the worker's human capital level, and an idiosyncratic noise element that is independently and identically distributed according to some bounded distribution.[2]

[2]This assumption is similar in spirit to the one made in many papers in auction theory, where bidders' valuations are assumed to be heterogeneous and determined according to some random distribution. However, unlike most of the literature on auction theory, we do not wish to study the effects of the random generation on agents' beliefs and equilibrium behavior. Instead we take a different approach and characterize the likely outcomes in a typical complete information matching market created in that manner.

We then provide a probabilistic analysis of the core of the game, and show that with high probability the differences in the payoffs of agents on the same side of the market behave like $\frac{\log n}{n}$, where n is the size of the market (Theorem 1). We also prove that this bound is tight (Theorem 2).

The fact that there are heterogeneous preferences in the market also implies that there are good and bad matchings between firms and workers, and that there is a surplus that is created by matching the right worker to the right firm.[3] Our approximate law of one price helps us to analyze the distribution of this surplus between firms and workers in balanced and unbalanced markets. In an unbalanced market with more workers than firms, at least one worker will be left out, and that worker will be willing to transact with any matched firm even for a minuscule gain. This constrains the profit of the worker matched to any firm that has good idiosyncratic fit with the unmatched worker, and by the approximate law of one price, the rest of the agents on the long side will necessarily make very small profits as well (Corollary 4). This argument shows why most of the surplus goes to market participants on the short side, despite the assumed idiosyncratic nature of pairwise productivities. In a balanced market we show that the surplus can be distributed in a variety of ways (Corollary 3).

These two results extend our economic intuitions about competition and surplus distribution in markets for homogeneous goods. If there are 10 farmers trying to sell 10 bushels of wheat to 9 identical buyers, and each of the buyers is interested in buying exactly one bushel of wheat and is willing to pay up to \$100 for it, then the price of wheat will be \$0, and each buyer's welfare is \$100. In a market with 10 farmers and 10 buyers, the price of wheat can be as high as the buyers' willingness to pay.

As mentioned earlier, some of our results rely heavily on two assumptions: separability of production factors and boundedness of the idiosyncratic noise factor. We relax the first assumption by considering a model with a Cobb–Douglas productivity function, in which

[3] One interpretation of the productivities appearing in our model is to think of them as actual output of workers, which is likely to be affected by heterogeneous person-organization fit. See Kristof-Brown and Guay (2011) for a recent survey of most of the important contributions to the literature on this issue.

the firm's quality and the worker's human capital level are complements. We prove that in this model the efficient assignment is with high probability approximately assortative (Lemma 5), and recover the approximate law of one price (Theorem 6). This analysis reveals that the argument for an approximate law of one price is at least to some extent robust to other forces in the market, such as efficiently matching good workers with good firms (and vice versa).

We conclude by focusing on the boundedness assumption and show that it cannot be dispensed with. We consider a model with exponential noise and show that the differences in workers' payoffs do not vanish as the market grows (Proposition 8). Nevertheless, we do present computer simulations and a partial argument for why surplus distribution under exponential noise may present similar properties to surplus distribution under bounded noise (Theorem 10).

The rest of the paper is organized as follows. Section 1.2 reviews the literature related to our paper. Section 1.3 introduces the model and the formal notation. Section 1.4 contains the statement and the proof of the main result, as well as the tightness result, and an analysis of surplus distribution. Section 1.5 discusses the extension of the main result to a market with interaction terms in the joint productivity of firms and workers. Section 1.6 presents some results related to unbounded noise distributions. Section 1.7 provides simulation results, and Section 1.8 concludes.

1.2 Related Literature

Assignment games were first introduced by Shapley (1955). Shapley and Shubik (1971) thoroughly analyze them and show that the core can be described as the set of solutions to a linear program dual to the optimal assignment problem, and that it is therefore nonempty, compact, and convex. They also prove that it contains two special allocations: a firm-optimal and a worker-optimal core allocation. Demange and Gale (1985) extend the analysis and show, among other things, that the core has a lattice structure. They also point to the nonmanipulability by workers of the worker-optimal core allocation. Assignment games bear a great resemblance to the very familiar assortative matching model of Becker (1981), with the main difference being the lack of agreement of agents on one side over the ranking of agents on the other side in the more general assignment game model. In a slightly different interpretation, Demange *et al.* (1986) use the assignment game framework to describe auctions of heterogeneous items with unit demand bidders (with this interpretation in mind, core allocations are equivalent to Walrasian equilibria, and therefore our results provide insight into revenue acquired by multiple auctioneers under different market conditions).

Within the literature that focuses on assignment games, a paper related to ours is Kanoria *et al.* (2014). They too study a random version of the assignment game and show core convergence in the sense of agents getting similar payoffs across different core allocations. The most striking difference between the models is that in theirs each agent has a type (out of a finite set of fixed types), and agents' preferences depend only on the type of the agent to which they are matched, whereas in our model each agent may have a ranking over individual agents on the other side of the market. Other relevant papers within this literature are those that study the size of the core (in deterministic assignment games) such as Quint (1987) who defines two measures for core elongation and shows the relation between them, and Núñez and Rafels (2008) who investigate the dimension of the core based on the entries in the productivity matrix.

Several recent empirical works estimate a model similar to ours (and even more closely related to Kanoria *et al.*, 2014), with the caveat of using an extreme value distribution for the idiosyncratic component. Choo and Siow (2006) consider marital behavior in the United States and estimate a model in which each agent has a type, and idiosyncratic preferences over being matched with any type of agent on the other side of the market. Similarly, Botticini and Siow (2008) study whether there are increasing returns to scale in marriage markets, and Chiappori *et al.* (2011) study the marital college premium.

From a broader point of view, this paper belongs to the theoretical literature on matching in two-sided markets. This literature gained prominence in the 1960s and early 1970s following the publication of the seminal papers by Gale and Shapley (1962) and Shapley and Shubik (1971), and research remained mostly divided (with some notable exceptions) into two parallel strands: with and without transferable utility (i.e., money). The bulk of the literature on matching markets without transfers, also known as the marriage market model (in the one-to-one case) and the college admissions model (in the many-to-one case), is focused on studying theory related to markets with fixed preferences, often under the additional assumption of complete information. Within this realm, two important papers for our discussion are Crawford and Knoer (1981) and Kelso and Crawford (1982). These papers describe the detailed connection between marriage markets and assignment games, and point to an auction process similar to the deferred-acceptance algorithm that produces an approximation to a side-optimal core allocation.[4] We employ a similar auction process in the proof of our lower bound of variation in workers' salaries (Theorem 2).

The past two decades have seen the emergence of more models that allow for stochastic markets and incomplete information. This new focus has revealed to market designers that some of the subtleties related to small markets may very well become negligible once we consider large "likely" markets. Yet the works on large markets most relevant to our present study were already written in the 1970s by Wilson (1972) and Knuth (1976), and were extensively developed by Pittel (1989, 1992). These papers analyze marriage markets with preferences that are determined uniformly at random and show that in a situation in which the number of men is equal to the number of women, with high probability the proposing side's (in a deferred acceptance algorithm) mean rank of partners behaves like $\log n$, whereas the other side's mean rank of partners behaves like $\frac{n}{\log n}$. This particular strand of the literature remained dormant for almost three decades, but several papers have recently used similar methods. Ashlagi *et al.* (2013) show that in unbalanced random marriage markets with high probability under any stable matching the short side's mean rank of partners behaves like $\log n$, whereas the long side's mean rank of partners behaves

[4]For further generalizations of the marriage market model and the assignment game model see, for example, the works by Hatfield and Milgrom (2005), Ostrovsky (2008), Hatfield *et al.* (2013), and references therein.

like $\frac{n}{\log n}$. Coles *et al.* (2014) and Coles and Shorrer (2014) employ these results to study aspects of strategic behavior in marriage markets with incomplete information. Lee (2014) and Lee and Yariv (2014) assume that preferences are derived from underlying cardinal utilities and study the issues of core convergence and efficiency, respectively.

Using somewhat different methods, but still trying to explain core convergence using different modes of competition, Immorlica and Mahdian (2005) explain in a breakthrough paper why in a large random marriage market with one of the sides having rank-ordered lists of bounded length and with incomplete information, truth-telling become an approximately dominant strategy. Kojima and Pathak (2009) extend this result to the college admissions model, and Storms (2013) extends it to many-to-one markets with substitutable preferences.[5] Kojima *et al.* (2013) use a similar strategy to prove that in a market with "not too many" couples, a stable matching exists despite the complementarities imposed by couples' preferences. Ashlagi *et al.* (forthcoming) further improve this result, show that stability is also implied for groups that can contain more than two members, and provide a counterexample to the case of a similar number of singles and couples.

Technically, our analysis is also related to what is known in the operations research and computer science literature as the random linear sum assignment problem. Specifically, two results that are used repeatedly in our proofs are the calculation of the limit value of a large random assignment game (Aldous, 2001), and the bounding of the minimal productivity in the optimal assignment (Frieze and Sorkin, 2007). For a more exhaustive survey of the random linear sum assignment problem (and closely related problems) see Krokhmal and Pardalos (2009).

1.3 Model and Notation

Consider a sequence of markets $\{M^n\}_{n=1}^{\infty}$, such that each market can be described as $M^n = (F^n, W^n, q^n, h^n, \alpha^n)$, where F^n is a set of firms of size n, W^n is a set of workers of size

[5] Related analysis was also applied by Manea (2009), Che and Kojima (2010) and Kojima and Manea (2010) to the problem of optimal object assignments.

$n + k(n)$, with $k(n) \in \mathbb{N}$ and $k(n) = O(n)$,[6] q^n is a vector of qualities related to firms in F^n, h^n is a vector of human capital levels related to workers in W^n, and α^n is an $|F^n| \times |W^n|$ real matrix representing the value of pairs of firms and workers. We assume throughout that each element of α^n can be described as

$$\alpha_{ij}^n = u\left(q_i^n, h_j^n\right) + \varepsilon_{ij}^n,$$

where u is the part of the production function that depends only on the firm's quality and the worker's human capital level, and ε_{ij}^n is idiosyncratic noise representing the productivity related to the identities of the firm and the worker. ε_{ij}^n is independently and identically distributed according to the cumulative distribution function G which has a continuous and strictly positive probability density function g.[7]

For technical purposes we will assume (unless otherwise noted) that the elements of the vectors h^n are identically and independently distributed on the interval $\left[\underline{h}, \overline{h}\right]$ according to the cumulative distribution function H. If $\underline{h} \neq \overline{h}$ we will also require H to have positive and continuous density on this interval. This assumption can easily be relaxed, but it is kept for clarity. Note that it does not hold for the specific distribution we use in Appendix A.3.

- **The separable case:** $u(q,h) = q + h$.

- **The interactive case:** $u(q,h) = q^\gamma h^{1-\gamma}$.

Note that while q and h appear without a transformation in both cases, any continuous transformation can be applied directly to their distributions. Therefore, the word "separable" accurately describes the domain of the first case. We also distinguish between several possible assumptions on G:

- **Bounded noise:** G is bounded on the interval $[0,1]$ ($G(1) = 1$).

- **Unbounded noise:** There exists no $c \in \mathbb{R}$ such that $G(c) = 1$.

[6] The latter assumption is introduced for mathematical convenience.

[7] In fact for our results to hold we only need that the density be continuous near its supremum.

- **Exponential noise:** $G = \mathrm{Exp}(1)$ (special case of unbounded noise).

We prove our main result for the separable case with bounded noise, and extend it (under a certain technical assumption to be mentioned later) to the interactive case with bounded noise. We show that an approximate law of one price (properly formulated) does not hold in general for unbounded noise. Nevertheless, we explain why we believe some of our surplus distribution results do hold (in a weak form), at least for the case of exponential noise.

In market M^n, the value of a coalition of firms and workers S is given by

$$\mathrm{v}(S) = \max \left[\alpha^n_{i_1 j_1} + \alpha^n_{i_2 j_2} + \cdots + \alpha^n_{i_l j_l} \right],$$

where the maximum is taken over all arrangements of $2l$ distinct agents, $f^n_{i_1}, \ldots, f^n_{i_l} \in S \cap F^n$, $w^n_{j_1}, \ldots, w^n_{j_l} \in S \cap W^n$, $l \leq \min \{|S \cap F^n|, |S \cap W^n|\}$. An **allocation** is denoted by (μ, u, v) with μ being a matching of firms to workers and vice versa, and u and v being payoff vectors for the firms and workers, respectively. We refer to u as firms' "profits," and to v as workers' "salaries." Formally, $\mu \colon F^n \cup W^n \to F^n \cup W^n \cup \{\emptyset\}$, and satisfies

1. $\forall f \in F^n : \mu(f) \in W^n \cup \{\emptyset\}$,
2. $\forall w \in W^n : \mu(w) \in F^n \cup \{\emptyset\}$, and
3. $\forall f \in F^n, w \in W^n : \mu(f) = w \iff \mu(w) = f$.

An allocation is a **core allocation** if no coalition can deviate and split the resulting value between its members such that each member of the coalition becomes strictly better off. We denote the set of core allocations of M^n by $C(M^n)$. As mentioned above, Shapley and Shubik (1971) show that the core is a nonempty compact and convex set, and that it is elongated in the sense that there is a firm-optimal core allocation in which salaries are at their lowest level among all core allocations, and a worker-optimal core allocation in which salaries are at their highest level among all core allocations.

Most of our results are going to hold for "most" realizations of some stochastic matrices and vectors. We often use the technical term *with high probability* (or *whp* for short) to mean that some result holds for the sequences of markets M^n with probability $1 - O\left(\frac{1}{n}\right)$.

Whenever it is not mentioned, the term refers to realizations of the stochastic matrices α^n as well as the quality vectors q^n and h^n. However, in some places we explicitly mention that the term refers only to α^n or only to the quality vectors.

1.4 An approximate law of one price

This section presents our main result, which shows that in the separable case with bounded noise there cannot be too much variation in the payoffs of the agents on either side of the market. We then proceed to improve our upper bound on this variation for the special case of side-optimal core allocations, and establish a lower bound. These two proofs use a different method that relies on the salary adjustment procedure described by Crawford and Knoer (1981) and Kelso and Crawford (1982). The following subsection employs these results to characterize surplus distribution in these markets, and argues that the range of potential outcomes (i.e., payoffs in the core) crucially depends on whether the market is exactly balanced or not.[8] If it is not exactly balanced, the short side keeps most of the created surplus (or at least the surplus due to the idiosyncratic noise).

1.4.1 The main result

In order to gain some intuition into the mechanics of the proof and the argument behind it, let us first assume that the market is balanced, that all firms have the same quality, and that all workers have the same level of human capital. In this specific scenario our result implies that whp all workers (for example) should earn a very similar salary.

Suppose that worker w_1 is employed by firm f_1 and earns a salary of s_1 and worker w_2 is employed by f_2 and earns a salary of s_2. Suppose further that $s_2 > s_1$. If workers and firms were homogeneous goods, firm f_2 could offer worker w_2's job to worker w_1 for any salary strictly between s_1 and s_2. That is the usual argument for the law of one price in a two-sided market. However, it may well be that the combination of f_2 and w_2 has

[8] The term "balanced" is also used in the context of cooperative game theory to describe games with a nonempty core. This meaning is not used anywhere in this paper.

much higher productivity than f_2 and w_1, and therefore there is no mutually beneficial opportunity for f_2 and w_1. Nevertheless, we do know that there are about $n^{\frac{2}{3}}$ workers in the market such that their productivity with firm f_2 is no less than $1 - \frac{1}{n^{\frac{1}{3}}}$. For each of those workers the original argument works perfectly, and so none of these workers can be paid less than $s_2 - \frac{1}{n^{\frac{1}{3}}}$, because otherwise she and firm f_2 might deviate. Now we have a set of size $n^{\frac{2}{3}}$, each getting a salary of at least $s_2 - \frac{1}{n^{\frac{1}{3}}}$. Consequently there are about $n^{\frac{2}{3}}$ firms paying a salary of at least $s_2 - \frac{1}{n^{\frac{1}{3}}}$, and whp one of these firms, say f', is a good match with worker w_1, in the sense that their joint productivity is more than $1 - \frac{1}{n^{\frac{1}{3}}}$. By considering the possibility of deviation by f' and w_1, we reach the conclusion that $s_1 \geq s_2 - \frac{2}{n^{\frac{1}{3}}}$.

The argument used above is not quite accurate, since we do not account for the fact that firms in the intermediate set are not random, but are rather chosen in a specific way (i.e., they are matched to workers who are also productive when matched with firm f_1). The formal proof handles this issue by considering the likely expansion properties of the directed graph induced by the random productivity matrix and showing that a path must exist between f_2 and w_1.

As implied, the other difference from the informal argument above is that the proof uses the smallest possible expansion that still results in the necessary paths between all pairs of agents, i.e., a strongly connected digraph. This minimality is also formally established in our derivation of a lower bound for the variation in agents' payoffs. The technical element of the proof that allows for constructing high-probability paths is based on the result of Frieze and Sorkin (2007), which we extend here to deal with unbalanced markets as well as bounded distributions other than the uniform distribution.

The intuition behind proving the result for unbalanced markets is pretty straightforward given our understanding of how to utilize improvement paths, as previously described. We first show that whp all workers above a certain level of human capital are matched. Otherwise, one could replace a low-quality worker with a high-quality worker, and then reshuffle the matched workers such that the impact on the efficiency coming from idiosyncratic noise component will not be too substantial. We next show that the same logic that was used in the balanced case can be applied to the unbalanced case, if we focus only on agents above a certain level of human capital.

Theorem 1. *In the separable case with bounded noise, there exists* $c \in \mathbb{R}_+$ *such that whp for any* $(\mu^n, u^n, v^n) \in C(M^n)$ *we have*

1. $\forall i, j \in \{1, \dots |F^n|\} : u_i^n - u_j^n \leq \left(q_i^n - q_j^n\right) + \frac{c \log n}{n}$, *and*
2. $\forall i, j \in \{1, \dots |W^n|\}, \mu^n(w_i^n), \mu^n(w_j^n) \in F^n : v_i^n - v_j^n \leq \left(h_i^n - h_j^n\right) + \frac{c \log n}{n}$.

Proof. See Appendix A.1. □

Theorem 1 demonstrates that in a large random assignment game, all firms make approximately the same profits, and all matched workers earn approximately the same salary. In a sense, this theorem states that the core is not only elongated, as implied in Shapley and Shubik (1971) and Demange and Gale (1985), but that it is also narrow.

The bounds already provided do not leave much room for further improvements (let alone the constants used in the proof), but we still wish to verify that they are tight, at least in terms of order of magnitude. The following theorem shows that they are. We focus on balanced markets with all firms having the same quality and all workers having the same human capital level, governed by a specific core allocation, namely, the firm-optimal core allocation. We know that we can find the firm-optimal core allocation via the auction-like algorithm proposed by Crawford and Knoer (1981). When firms propose to workers, the auction process ends when all workers have received an offer. We can compute the probability that at each stage a worker who has not received an offer so far receives an offer, and then calculate the number of discrete steps required to reach the last worker. The approximation is possible thanks to our bounds from Theorem 1. This gives us a lower bound for the expected sum of workers' salaries, which implies a lower bound on what the top earner gets with high probability. Since we know the lowest earner gets zero, we are done. We note that the same procedure can also be used to provide better constants in Theorem 1 for the specific cases of the side-optimal core allocations.

Theorem 2. *In the separable case with bounded noise, if* $k(n) \equiv 0$*,* $q^n \equiv \underline{0}$ *and* $h^n \equiv \underline{0}$*, there exists* $c \in \mathbb{R}_+$ *such that whp there exist* $(\mu^n, u^n, v^n) \in C(M^n)$ *and* $i, j \in \{1, \ldots |W^n|\}$ *for which* $v_i^n - v_j^n \geq \frac{c \log n}{n}$.

Proof. See Appendix A.2 □

We conclude this subsection by suggesting an interpretation of our results in terms of the shape of the core. As mentioned above, Shapley and Shubik (1971) already noticed that the core is compact and convex, and that it is shaped like a nut, in the sense that it contains firm-optimal and worker-optimal core allocations. Our results suggest that in large markets the core tends to be almost one-dimensional in the sense that one parameter defines it up to very small perturbations. In balanced markets, once we know what is the average profits of firms, we also approximately know the average salaries of workers, and what every firm and worker makes under that core allocation. The same holds for unbalanced markets. However, as we will see in the next section, workers' salaries in unbalanced markets are in fact determined by the human capital levels of those workers who are left unmatched, and therefore the core actually has no real variation and resembles a point more than a line. An interesting exercise would be to calculate the elongation measures suggested by Quint (1987) for large unbalanced markets and show that they indeed converge to zero.

1.4.2 Surplus distribution

With the results from the previous subsection at hand, we are now ready to explore their implications for surplus distribution. However, before doing so it is important to understand how much surplus is created in a large market. Aldous (2001) proved that in a large balanced random market with all firms having a quality of zero, and all workers having a human capital level of zero, and noise being distributed according to the uniform distribution on $[0, 1]$, the expected surplus created is $n - \frac{\pi^2}{6}$. This result can be easily extended both to general bounded distributions (with positive and continuous density) and to unbalanced markets, and in general we know that the surplus to be divided between firms and workers is $\Omega(n)$. As for qualities and human capital, our analysis suggests that with high probability the workers who will take part in the optimal assignments are all those above a certain human capital level (see Lemma 36 in Appendix A.1), and so we can tell from the distribution of qualities and human capital levels what is going to be the surplus created due to those factors.

Our main result in this subsection is that when the market is exactly balanced (i.e., $k(n) \equiv 0$) the surplus that is created from the idiosyncractic matching between firms and workers can be divided in very different ways. However, in the presence of even a slight imbalance, most of the surplus related to the noise goes to the short side (the firms). This indicates that a large core is a knife-edge case that is not likely to be found in any real applications. This result is the assignment games parallel to Ashlagi *et al.* (2013), who prove that in the realm of matching without transfers a large core is only possible if the number of men and women is exactly equal, and that in unbalanced markets the short side has a big advantage in determining the resulting matching.

Corollary 3. *In the separable case with bounded noise, let $k(n) \equiv 0$ and let $\left(\mu^n, u^{n,F}, v^{n,F}\right)$ be the firm-optimal core allocation. Then there exist $c \in \mathbb{R}_+$ such that whp*

$$\forall w_j^n \in W^n : v_j^{n,F} \in \left(\left(h_j^n - \underline{h}\right) - \frac{c \log n}{n}, \left(h_j^n - \underline{h}\right) + \frac{c \log n}{n}\right).$$

Intuition for the proof. Under the firm-optimal core allocation there is at least one worker who gets a salary of exactly zero; otherwise we could reduce all salaries by a small constant without violating any of the inequalities defining the core. This worker's human capital level cannot be too high (otherwise, by the approximate law of one price, others with lower human capital levels would get negative salaries). Then, by the approximate law of one price, all workers must get only the difference between their human capital level and that worker's human capital level. For the full proof see Appendix A.2. □

A similar argument to the one we used for balanced markets can be applied to unbalanced markets. In this case, a worker who is left unmatched gets a salary of zero, and this constrains at least some of the salaries of the workers who are matched. Then, by the approximate law of one price, we get bounds on the salaries of all workers.

Corollary 4. *In the separable case with bounded noise, let $k(n) > 0$ for all n. Then there exist $c \in \mathbb{R}_+$ such that whp for all $(\mu^n, u^n, v^n) \in C(M^n)$ and for all $w_j^n \in W^n$ such that $\mu(w_j^n) \in F^n$,*

$$v_j^n \in \left(\left(h_j^n - h^n[n]\right) - \frac{c \log n}{n}, \left(h_j^n - h^n[n]\right) + \frac{c \log n}{n}\right),$$

where $h^n[n]$ signifies the n-th highest element in the vector h^n.

Proof. See Appendix A.2. □

Corollary 3 implies in particular that in a balanced market the expected division of surplus is such that the workers get the contribution of their excess human capital (above $\underline{h}$) and then only $O\left(\frac{\log n}{n}\right)$ out of the part of the surplus that is related to the noise distribution. Note that while Corollary 3 is put in terms of the firm-optimal core allocation, it is completely symmetric, and therefore the same applies to the opposite case of the worker-optimal core allocation. The convexity property of the core ensures that any compromise distribution is also possible in a core allocation. Unlike the long (and narrow) core characterization in balanced markets, Corollary 4 shows that in unbalanced markets the core quickly converges to almost a point. The resulting surplus division is such that under any core allocation, the agents on the long side (the workers) get the contribution of their excess quality (not above the lower bound of the distribution, but rather above the highest quality of an unmatched agent) plus a $O\left(\frac{\log n}{n}\right)$ fraction of the surplus created by the idiosyncratic matching.

1.5 Extension to Cobb–Douglas productivities

In the previous section we showed that an approximate law of one price holds for markets in which both firms' quality and workers' human capital affect the productivity of each matched pair, but we did not allow for any interaction between those two properties. In other words, good workers provided the same output regardless of whether they were working in a good firm or in a bad firm. While mathematically convenient, it is not a very plausible assumption. In this section we wish to relax our previous separability assumption and consider also the family of productivity functions suggested by Cobb and Douglas (1928).

Our main concern when considering interaction is that workers and firms will tend to ignore their idiosyncratic productivity noise and will match solely on the basis of their respective qualities. This is known in the economics literature as "assortative matching," and within the matching literature it is most identified with the work of Becker (1981). If firms and workers match assortatively, there will not be any chance of having an approximate version of the law of one price, since the idiosyncratic productivities can tilt the profits of matching pairs.

We find that as the market grows large (and under certain technical assumptions on the qualities of firms and workers), there is a trade-off between matching assortatively on the quality dimension and matching efficiently on the noise dimension. We define the concept of "approximately assortative matching," which means that all firms are matched to workers who have approximately the same level of human capital as the firms' quality. The fact that the matching is only *approximately* assortative and not completely assortative allows for more efficient matching in terms of idiosyncratic noise.

Definition 1. *A model exhibits approximately assortative matching if there exist* $c \in \mathbb{R}_+$ *and* $a \in (0,1)$ *such that whp for any* $(\mu^n, u^n, v^n) \in C(M^n)$ *and for any* i, j *such that* $\mu^n(f_i^n) = w_i^n$ *we have* $\left|q_i^n - h_j^n\right| \leq cn^{-a}$.

We now turn to a specific model, which we refer to as *the Cobb–Douglas benchmark model*. The Cobb–Douglas benchmark model consists of a balanced market ($k(n) \equiv 0$) in which productivities are given by $\alpha_{ij}^n = 2\sqrt{q_i^n h_j^n} + \varepsilon_{ij}^n$, and $q_k^n = h_k^n = \frac{k}{n}$, i.e., qualities of firms and human capital levels of workers are evenly spaced.

Lemma 5. *The Cobb–Douglas benchmark model exhibits approximately assortative matching.*

Proof. See Appendix A.3. □

Having established an approximately assortative matching, we can prove the approximate law of one price using the tools developed for the separable case, but not quite the same ones since we need to make sure that we limit the paths used in those proofs so that they do not go through firms or workers that have very different qualities. Even then a direct comparison between firms or between workers of different qualities is not straightforward, and so we restate our main result in terms of agents that have similar qualities.

Theorem 6. *In the Cobb–Douglas benchmark model there exist* $c_1, c_2 \in \mathbb{R}_+$ *and* $a, b \in (0,1)$ *such that whp for any* $(\mu^n, u^n, v^n) \in C(M^n)$*:*

- $\forall i, j \in \{1, \ldots, n\}$ *such that* $\left|q_i^n - q_j^n\right| \leq c_1 n^{-b}$*:* $u_i^n - u_j^n \leq c_2 n^{-a}$*, and*
- $\forall i, j \in \{1, \ldots, n\}$ *such that* $\left|h_i^n - h_j^n\right| \leq c_1 n^{-b}$*:* $v_i^n - v_j^n \leq c_2 n^{-a}$*.*

Proof. See Appendix A.3. □

1.5.1 Surplus distribution

Still focusing on the Cobb–Douglas benchmark model, it is quite clear that while the analysis of surplus distribution is not as straightforward as the separable case, it is still not much different. The rough intuition for the next result is that we can compare the salary of any worker with that of a worker who has a slightly lower or slightly higher human capital level, if both workers have a relatively high joint productivity with the firm that employs one of them. This allows us to build paths from any worker to one of the workers with the lowest human capital levels and deduct that the former can only make a salary that is the sum of the differences between productivities of workers along the path. In other words, the salary of a worker with human capital level h_j^n is roughly the integral from 0 to h_j^n of the marginal productivities of workers. Since we know that there is approximately assortative matching, we also know the quality of firms matched to workers along the path.

Corollary 7. *In the Cobb–Douglas benchmark model let* $\left(\mu^n, u^{n,F}, v^{n,F}\right)$ *be the firm-optimal core* *allocation. Then there exist $c, a \in \mathbb{R}_+$ such that*

$$\forall j \in \{1, \ldots, n\} : v_j^{n,F} \in \left(\frac{j}{n} - cn^{-a}, \frac{j}{n} + cn^{-a} \right).$$

Proof. Omitted. □

We note that the surplus created by worker w_j^n is approximately $\frac{2j}{n} + 1$, and so we learn that the workers get only the share of the surplus related to their own contribution to the correlated component, and none of the surplus related to the idiosyncratic component under the firm-optimal core allocation.

We conclude this section by noting that none of the technical steps we took seem to require balancedness. We therefore conjecture that in unbalanced markets any worker's salary under any core allocation will be bounded above by the integral of the marginal productivity from the highest human capital level of any unemployed worker to her own human capital level, plus an expression that behaves like $O\left(\frac{1}{n^a}\right)$ for some $a \in (0, 1)$. Simulation results presented in Section 1.7 also indicate that this conjecture holds.

1.6 Unbounded noise

Up until now we have established that an approximate version of the law of one price holds in two-sided economies with heterogeneous preferences. However, one of the more restrictive assumptions that we used was the boundedness of the noise distribution, which obviously leads to a relatively high concentration of "good enough" matches, and in particular allows an assignment so efficient that it misses a potential first-best only by a constant (Aldous, 2001). In this subsection we relax this assumption for the first time and try to understand what happens when the noise is unbounded. Apart from the mathematical elegance and conceptual difference of unbounded noise, understanding the implications of this concept is also important for comparing our work with some of the empirical papers on two-sided matching markets with transfers, which are based on models with unbounded noise (e.g., Choo and Siow, 2006).

When we discuss unbounded noise it is important to understand what it means to have "one price" in the market, since the average productivity may tend to infinity as the market grows large. Our interpretation is that an approximate law of one price holds if the variation among agents' profits is a vanishing fraction of the average productivity. In the bounded case, the average productivity approaches a constant, and therefore any sub-constant differences in profits are considered as an approximate law of one price. In what follows we focus on the exponential distribution, under which the average productivity behaves like $\log n$, and we show that whp there are two workers in the market whose salaries differ by $\Theta(\log n)$. Hence, we conclude that in the presence of unbounded noise the law of one price might not hold.

The intuition for our "counterexample" is that unbounded distributions with a heavy tail may create "good" outliers, i.e., agents that are highly productive compared to others, and such that agents from the other side fiercely compete to be matched with them. These agents share a significant portion of the surplus they help to create, and if they are common enough, they may offset other forces that would otherwise squeeze the surplus from their side (such as an adversarial core allocation, or a slight imbalance in favor of the other side of the market). Our example is based precisely on the existence of such agents.

Proposition 8. *In the separable case with exponential noise, let the market be balanced ($k(n) \equiv 0$), with all firms having the same qualities ($q^n \equiv \underline{0}$), and with all workers having the same human capital level ($h^n \equiv \underline{0}$). Let $\left(\mu^{n,F}, u^{n,F}, v^{n,F}\right)$ denote the firm-optimal core allocation of M^n. Then average productivity is $\Theta(\log n)$, and there exists $c \in \mathbb{R}_+$ such that whp there are two workers w_i^n and w_j^n with $\left|v_i^{n,F} - v_j^{n,F}\right| > c \log n$.*

Proof. See Appendix A.2. □

1.6.1 Surplus distribution under exponential noise

Despite the fact that the law of one price does not apply in general to unbounded noise, we would like to argue that at least some of the main conclusions, i.e., the convergence of the share of the surplus that each side gets, continues to hold to some extent. By studying simulation data carefully (see Figure 1.11 in Section 1.7), one suspects that the behavior of the workers' expected share of the surplus in a balanced market under the firm-optimal core allocation is $\Theta\left(\frac{\log\log n}{\log n}\right)$. In what follows we assume the following mathematical conjecture is true, and show that indeed the share of the surplus behaves in that manner.

Conjecture 9. *In the separable case with exponential noise, let $k(n) \equiv 0$, $q^n \equiv \underline{0}$ and $h^n \equiv \underline{0}$. Then there exists $c \in \mathbb{R}_+$ such that whp under the maximal assignment each firm is matched to one of the $c \log n$ workers who have the highest joint productivity with that firm.*

We note that Conjecture 9 parallels Theorem 2 of Frieze and Sorkin (2007), in the sense that it bounds the lowest possible element in the optimal assignment. While computer simulations suggest that it holds (see Section 1.7), we are not familiar with any work within the computer science literature or the operations research literature that tackles the problem of unbounded distributions.[9]

Theorem 10. *In the separable case with exponential noise, let $k(n) \equiv 0$, $q^n \equiv \underline{0}$, and $h^n \equiv \underline{0}$. Assume Conjecture 9 holds, and let $\psi^F(M^n) = \left(\mu^n, u^{n,F}, v^{n,F}\right)$ be the firm-optimal core allocation. Then there exists $c \in \mathbb{R}_+$ such that*

$$E\left[\frac{\sum_j v_j^{n,F}}{\sum_i u_i^{n,F} + \sum_j v_j^{n,F}}\right] \leq \frac{c \log\log n}{\log n}$$

Intuition for the proof. In a balanced market governed by the firm-optimal core allocation, a worker cannot make more than the value she creates together with the firm that employs her minus the lowest value that any other worker creates (Lemma 40). Given the assumption and the above claim, it remains to show that with high probability the lowest value created by any worker behaves like $\log n - c \log\log n$ (Lemma 41). The full proof appears in Appendix A.2. □

It is worth mentioning that by observing simulation results for unbalanced markets (Figure 1.11), one may arrive at the following conjecture.

[9] Those two literatures focus on minimizing the sum of costs, and not maximizing productivity, and therefore unbounded distributions are less intuitive.

Conjecture 11. *In the separable case with exponential noise, let $k(n) > 0$, $q^n \equiv \underline{0}$, and $h^n \equiv \underline{0}$, and let $\psi^W(M^n) = (\mu^n, u^{n,W}, v^{n,W})$ be the worker-optimal core allocation. Then there exists $c \in \mathbb{R}_+$ such that*

$$E\left[\frac{\sum_j v_j^{n,W}}{\sum_i u_i^{n,W} + \sum_j v_j^{n,W}}\right] \leq \frac{c \log\log n}{\log n}.$$

In particular, this implies that for any core mechanism (that is, any function from markets to core allocations) the expected surplus of the workers is $O\left(\frac{\log\log n}{\log n}\right)$.

We conclude this section by suggesting that although we were focused on the study of the exponential distribution, much can be inferred about other unbounded distributions. Proposition 8 provided a counterexample to a theorem that held for the bounded case. The conjectures we discussed in this subsection were strictly about the exponential distribution, but it is our belief that other distributions that have similar tail behavior will exhibit the same phenomena (see also Figure 1.12 and Figure 1.13 in Section 1.7).

1.7 Simulations

In this section we present results of computerized simulations that demonstrate how quickly the dispersion of payoffs contracts, and how this affects the market. Unless explicitly noted, figures are based on averaging 400 trials for each market size, where the size of balanced markets ranges from $(10, 10)$ to $(300, 300)$ with jumps of 5 agents on each side, and the size of unbalanced markets ranges from $(5, 6)$ to $(300, 301)$ with jumps of 5 agents on each side.

1.7.1 The separable case with bounded noise

We first focus on the benchmark case of uniform $[0, 1]$ distribution with all firms having the same quality ($q^n \equiv \underline{0}$) and all workers having the same human capital level ($h^n \equiv \underline{0}$), and study wage dispersion in balanced markets under the firm-optimal core allocation. Figure 1.1 shows that indeed in a balanced market the maximal difference between the profits of any two firms in any core allocation behaves like $\frac{\log n}{n}$, as proved by Theorem 1 and Theorem 2.[10]

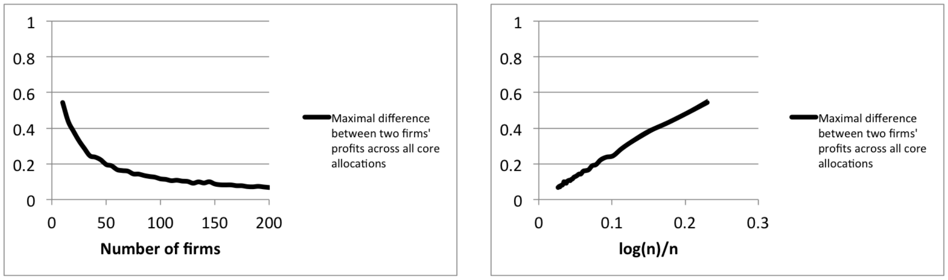


Figure 1.1: *Approximate law of one price in balanced markets*

The left panel of Figure 1.2 shows that in this case the maximum salary any worker gets under the firm-optimal core allocation also behaves like $\frac{\log n}{n}$, and the right panel of the same figure exemplifies the fact that the core in balanced markets is long, as suggested by Corollary 3.

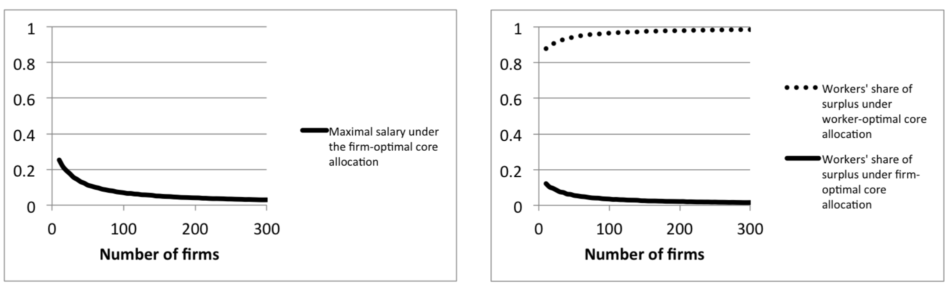


Figure 1.2: *Surplus distribution in balanced markets*

In unbalanced markets we expect the core to be much more narrow, per Corollary 4. The left panel of Figure 1.3 shows that even when the number of workers is only one more than the number of firms, the maximal salary any worker gets approaches zero rapidly, even under the worker-optimal core allocation. Furthermore, as the right panel demonstrates, in this case the workers' share in the surplus approaches 0, even under the worker-optimal core allocation. Figure 1.4 parallels Figure 4 of Ashlagi *et al.* (2013), and depicts the workers' share of the surplus when the number of workers is constant at 50, and the number of firms

[10]Figure 1.1 is based on only 25 trials for every market size, since finding the maximal difference across all core allocations requires solving $n(n-1)$ linear-programming problems.

varies from 20 to 80.

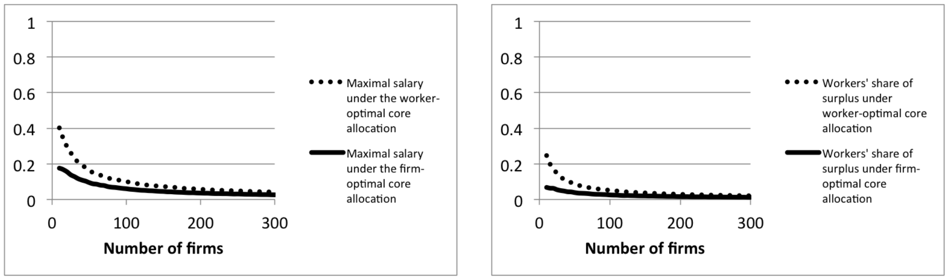


Figure 1.3: *Surplus distribution in unbalanced markets*

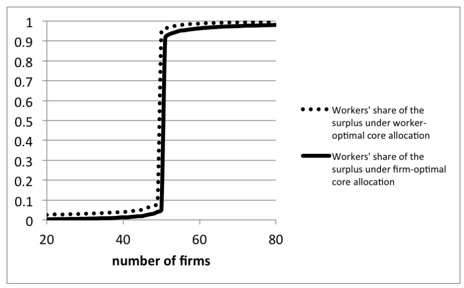


Figure 1.4: *Surplus distribution with* 50 *workers*

We now wish to verify that adding qualities to the mix does not substantially change any of these results. We let $q_i^n \sim U[0,1]$ for every i, and $h_j^n \sim U[0,1]$ for every j. In a balanced market we expect each worker to get roughly her human capital level, and for all workers to take 25% of the surplus. Under the worker-optimal core allocation we expect workers to take about 75% of the surplus. This is indeed shown in Figure 1.5. In an even slightly unbalanced market, we expect each worker to get her human capital level under any core allocation, and for the whole population of workers to take 25% of the surplus. This is demonstrated in Figure 1.6.

1.7.2 Cobb–Douglas productivity with bounded noise

We first try to demonstrate that assortative matching takes place in the model mentioned in Appendix A.3; i.e., each side of the market is characterized by evenly spaced qualities on the interval $[0,1]$, and the idiosyncratic noise is distributed according to $U[0,1]$. In Section 1.5

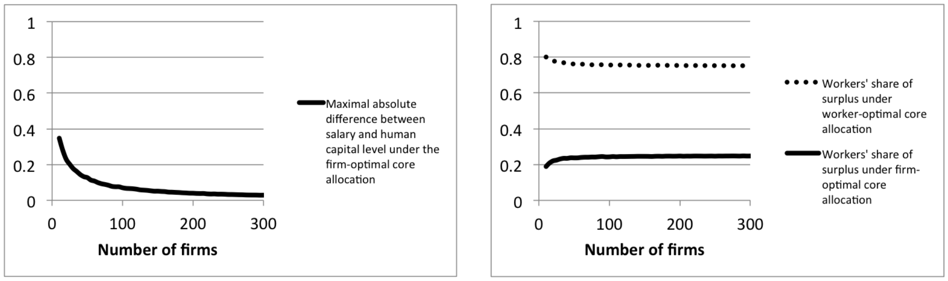


Figure 1.5: *Surplus distribution in balanced markets with qualities*

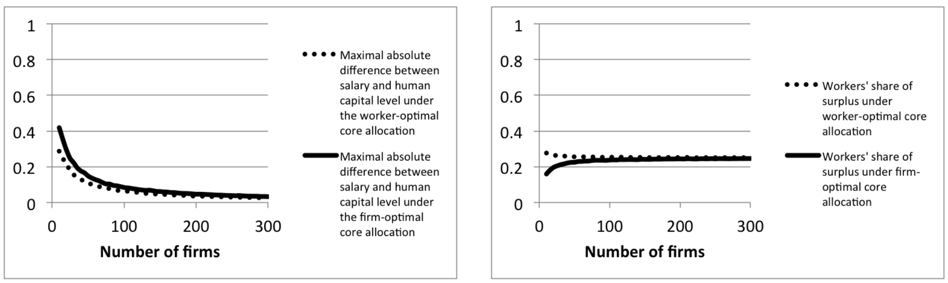


Figure 1.6: *Surplus distribution in unbalanced markets with qualities*

we proved just one aspect of assortative matching, namely, whp no firm is matched to a worker whose human capital level is substantially different from the firm's own quality (Lemma 5). The left panel of Figure 1.7 depicts the average and the maximal absolute quality difference between firms and the workers they employ under the optimal assignment. It is easy to see that these differences shrink as market size grows, and by looking at the logarithms of both axes (right panel) we can see that indeed these differences behave like a negative power of n.

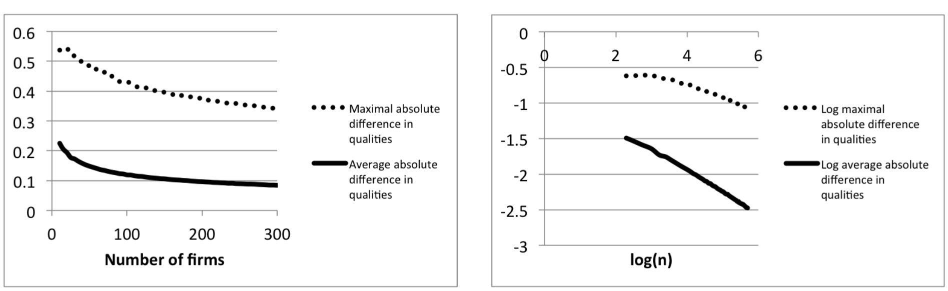


Figure 1.7: *Assortative matching when production factors are complements*

The surplus distribution described in Corollary 7 is depicted on the left panel of Figure 1.8. This panel shows the average absolute difference between workers' salaries and workers' human capital levels under the firm-optimal and the worker-optimal core allocations. The right panel supports the conjecture we raised at the end of Section 1.5 by showing the same metric in unbalanced markets for both the firm-optimal and the worker-optimal core allocations.

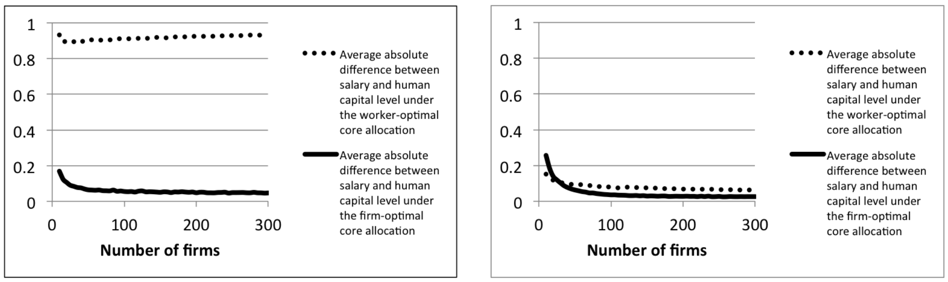


Figure 1.8: *Surplus distribution when production factors are complements*

1.7.3 Unbounded distributions

As mentioned in Section 1.6, unbounded noise distributions give rise to quite different phenomena than those mentioned with respect to bounded distributions. Figure 1.9 depicts the maximal difference between any two workers' salaries divided by the average surplus created under the optimal assignment, in a balanced market with exponential noise governed by the firm-optimal core allocation. As predicted by Proposition 8, the difference does not vanish as n gets large.

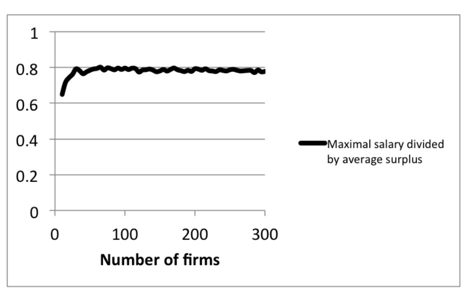


Figure 1.9: *No law of one price under Exponential distribution*

In Section 1.6 we also mentioned a conjecture about the behavior of the optimal assignment under the exponential distribution (Conjecture 9). Figure 1.10 shows that indeed it holds for medium-sized markets. The left panel of Figure 1.11 exemplifies how this conjecture translates into the conclusion of Theorem 10, and the right panel of that figure suggests that Conjecture 11 is true.

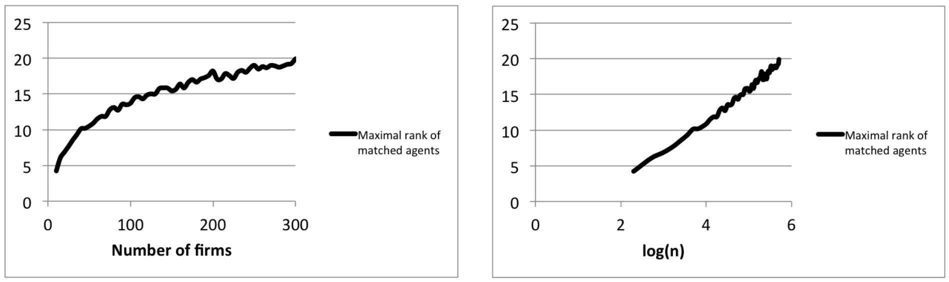


Figure 1.10: *Maximal rank of matched agents under exponential distribution*

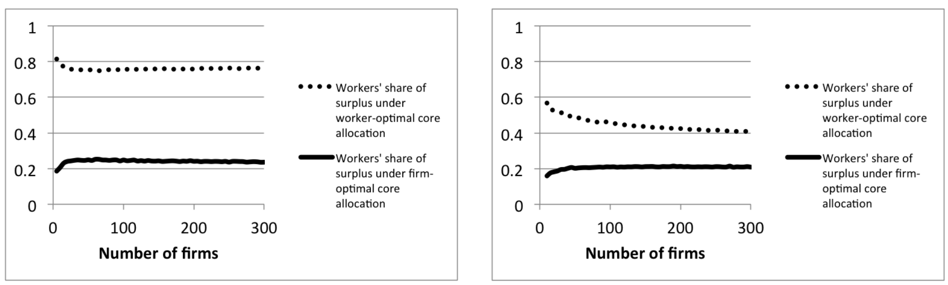


Figure 1.11: *Surplus distribution under exponential distribution (balanced and unbalanced)*

We conclude this subsection by noting that while our discussion was mostly about the exponential distribution, there are many other distributions that have similar tail behavior, and therefore are likely to exhibit the same phenomena. In particular, the extreme value distribution used in some empirical papers seems to have similar effects. Figure 1.12 parallels Figure 1.10 and shows the maximal rank of any two matched agents in a balanced market with noise distributed according to an extreme value distribution, and Figure 1.13 shows surplus distribution for both balanced and slightly unbalanced markets.

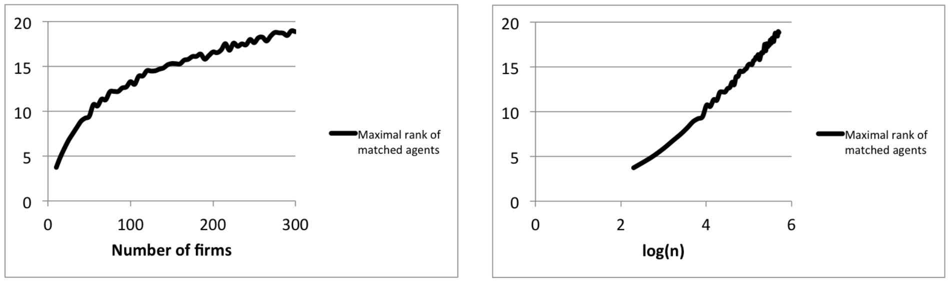


Figure 1.12: *Maximal rank of matched agents under extreme value distribution*

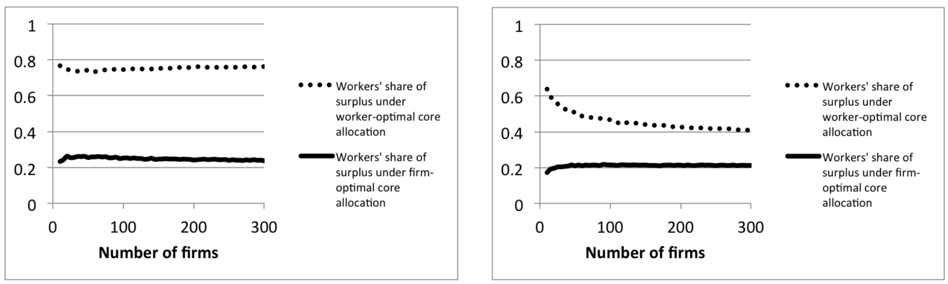


Figure 1.13: *Surplus distribution under extreme value distribution (balanced and unbalanced)*

1.8 Conclusion

During the 1980s, as it became clear that real-life centralized clearing houses could be immensely improved using intuitions gained in the study of marriage markets, the transferable utility strand of the literature became slightly neglected compared to its glorified non-transferable utility half-sibling. We decided to focus our attention in this paper on assignment games because it is our belief that they provide an excellent way to model decentralized markets, and that both strands of the matching theory literature can benefit from the continuous cross-fertilization.

We have investigated the applicability of the law of one price in two-sided matching markets with transfers, when agents have heterogeneous preferences over matching with the other side of the market. We have shown that an approximate law of one price holds, and that it implies core convergence and sharp predictions about surplus distribution in unbalanced markets. We have explained why the same kind of forces continue to work in markets in which there is interaction between the production factors, and why they fail to hold in markets in which the idiosyncratic noise is unbounded. These results indicate that only in knife-edge cases, in which the markets are exactly balanced, can we expect to see any significant variation in core outcomes.

We conclude the paper by noting that many of our assumptions were for expositional clarity only. The fact that firms had unit demand and workers supplied one unit of work is of course not crucial to our results, nor is the fact that all agents can possibly work in all the firms. The same results will hold in markets with discrete and finite demand and supply, and in markets that are less thick (at least to some extent). Nevertheless, some of the assumptions were crucial, and weakening them could lead to further understanding of markets with heterogeneous preferences. Specifically, the mechanism through which markets with unbounded noise converge remains a mystery, and the extent to which these results hold for markets with general utility functions (not quasi-linear) can be further studied. Finally, generalizing our results and the results of Ashlagi *et al.* (2013) to markets with substitutable preferences (with or without transferable utility) is another very promising direction for future research.

Chapter 2

Building Reputation at the Edge of the Cliff

2.1 Introduction

The 2012 New Year's Eve celebrations in America were somewhat clouded by the gloomy predictions of the Congressional Budget Office about an upcoming fiscal crisis that might take place in 2013. This unsettling forecast originated from the expiration of several laws, most notably the 2010 Tax Relief Act and the Budget Control Act of 2011, which entailed an increase in taxes as well as major spending cuts, leading to a sharp decline in the budget deficit. If all the changes were to go into effect simultaneously they would have induced a recession by cutting household incomes, increasing unemployment rates, and hurting both consumers' and investors' confidence in the economy. This dire situation sparked extensive media coverage that referred to the December 31 midnight deadline and the sharp decline in the budget deficit expected to ensue as the "fiscal cliff."

Preventing the fiscal cliff was supposedly a very simple task. Either the tax reliefs were to be extended, spending cuts were to be canceled, or some combination of these measures was to be taken. However, the political situation provided an extremely inconvenient environment for enacting such reforms. President Barack Obama and the Democraticcontrolled Senate disapproved of across-the-board tax cuts (as opposed to tax cuts for only the bottom 98%), and wanted to keep the spending level relatively high. The Republican-controlled House of Representatives preferred a solution that would lower spending as well as tax rates. Several proposals for amending the budget had been suggested by President Obama, House Speaker John Boehner, and others, but all were quickly rejected. In an attempt to reach an agreement, negotiations extended until the very last hours of 2012. There was some uncertainty about whether a compromise would be reached in time, and the entire bipartisan negotiation process was described by several commentators as an elaborate and dangerous high-stakes game of "chicken."[1] An agreement was finally reached just before the deadline, with legislation passing in the Senate on January 1, and in the House the following day.[2]

This paper models the negotiations as a revision game with reputation formation.[3] In the revision game model, introduced by Kamada and Kandori (2011), players prepare (pure) actions over a continuous and finite time horizon. They can change their actions only when they are called to play by stochastic Poisson processes. When the deadline is reached, the last actions prepared are used to determine the payoffs. This model also encompasses the idea of an uncertain deadline, as its effect is similar to the randomness induced by the stochastic revision opportunities.[4] We expand the model to accommodate for incomplete information, which allows us to study how adding a small probability irrational type into the game affects the equilibrium outcome.

---

[1]See, for example, Robin Harding, "US Plays Chicken on Edge of Fiscal Cliff", ft.com/world, November 11, 2012, and Robert Reich, "Cliff Hanger: Obama's Last Stand and the Republican Strategy of Fanaticism", Huffington Post, December 26, 2012.

[2]Several economists insist that the consequences of passing the deadline by several days would not have been as catastrophic as portrayed by the media; see, e.g., Baker (2013). However, our theoretical analysis applies to other situations of negotiating close to a deadline, as well as to the described situation where the utilities of the players represent the political cost of not reaching agreement for the two parties rather than the actual economic implications of the fiscal cliff.

[3]Our results also apply to a model of war of attrition with a deadline and with independent and stochastically distributed exit opportunities (see Section 2.5).

[4]Explicitly adding a deadline over which there is common uncertainty would not substantially change any of the results.

In simplifying the negotiation alternatives to a 2x2 game of opposing interests, we demonstrate that the effect of reputation-building on equilibrium outcomes can be substantial, even as the time horizon becomes infinitely long. That is, we show that, generically, reputation formation by one player prevents her opponent from achieving her most-preferred outcome (Proposition 12). Furthermore, one-sided reputation-building often leads to last-minute revisions that push some of the strategic interaction close to the deadline and induce a chance of falling over the cliff, i.e., reaching an outcome that neither side desires (Theorem 13).[5] When both parties try to build reputation, substantial delay must arise with positive probability and inefficiency is inevitable (Theorem 15). Furthermore, in this case the probability of not reaching an agreement can be non-negligible. It is important to stress that as our model contains no flow payoffs inefficiencies are caused only by ex-post Pareto inefficient outcomes, and that as a result of the discreteness of the revision phase any form of delayed action is necessarily tightly connected with inefficiency.

We provide some illustrative comparative statistics, as well as suggestive computational evidence, that help in assessing the magnitude of inefficiency. These results demonstrate that the more players are similar in strength, the more likely they are to hold to their bargaining position for a long time, leading to a deadline effect with harmful implications for the players' expected utility. In the limit case of equal strengths, the inefficiency does not vanish even as the ex-ante probability of the commitment types approaches zero.

Some methods used in the study of revision games are similar to those employed when discussing wars of attrition. We introduce a model of war of attrition over a continuous and finite horizon with Poisson arrivals and incomplete information. We prove that all sequential equilibria possess a simple structure in which one of the players uses a strategy that is completely characterized by a cutoff time, and the other player's strategy also adheres to the same cutoff time (but could be more complicated before it). This allows us to provide sharper results concerning delay and inefficiency in one-sided and two-sided

[5] It is straightforward to show existence of an equilibrium with such properties. Our contribution is in showing that every sequential equilibrium *must* exhibit the same inefficiency. The same is true for most of the results in this paper.

reputation-building scenarios.

The deadline effect that our model predicts is widely prevalent not only in political circumstances, but also in situations of a purely economic nature. It was observed in empirical data on labor strikes (Cramton and Tracy, 1992), and was replicated in lab experiments (Roth *et al.*, 1988). While the present paper stresses the role of reputation-building in inducing such an effect, other authors have suggested explanations such as irreversible commitments (Fershtman and Seidmann, 1993), private information about second-order beliefs (Feinberg and Skrzypacz, 2005), strategic delay in a bargaining process (Ma and Manove, 1993), individual deadlines (Sandholm and Vulkan, 1999), and optimism (Yildiz, 2004). Two papers that investigate this subject and have closer assumptions to ours are Fuchs and Skrzypacz (2013) and Ponsati (1995) who discuss continuous-time bargaining with private information and a deadline. Both papers point to a positive mass of agreements at the deadline, which is hard to interpret as inefficiency without a proper discretization of their continuous-time models.

This paper belongs to a growing body of work that follows Kamada and Kandori (2011) and studies different aspects of revision games. Some of the intuition is based on the analysis of the complete-information case by Calcagno *et al.* (2014). Other papers that employ similar methods include Ishii and Kamada (2011) and Kamada and Muto (2011). A few papers use a continuous-time finite-horizon framework to explore specific topics such as bargaining (Ambrus and Lu, 2010) or online auctions (Ambrus *et al.*, 2013). The novel feature of this strand of the literature is the tractable compromise between having a continuous-time model, which often requires extreme technical effort to get rid of unwanted equilibria, and using a discrete-time model, which may add additional strategic aspects (such as the exact order of play) that have nothing to do with the main issues.[6]

This work also relates to a vast literature that explores the effects of building reputation in repeated games, following the seminal papers of Kreps and Wilson (1982a) and Milgrom and Roberts (1982). Several of the many papers on reputation, and perhaps most notably

[6]For a relevant discussion on this topic see Hendricks *et al.* (1988).

Abreu and Gul (2000), demonstrate the existence of delay and inefficiency, where delay is measured from the first period of play and inefficiency is due to forgone payoff opportunities. Our results resemble these contributions in the sense that perturbing the game by adding commitment types can significantly shift the outcome when the time horizon is long enough. Players sacrifice utility to convince their opponent of their intentions not by decreasing immediate payoffs (as in the repeated games literature), but rather by increasing the probability of reaching an inferior outcome at the deadline. In this paper, the delay is measured backward from the deadline, and reputation formation is based on reaching an ex-post inefficient outcome with positive probability.[7]

The war of attrition extension is reminiscent of Fudenberg and Tirole (1986) who study what happens in a duopoly exit situation when there are commitment types. Unlike our model which is in continuous-time but has discrete preparation opportunities, in their model players can exit at any point in time, which makes their uniqueness proof far more involved. Other related models are the continuous-time model with two-sided uncertainty of Kreps and Wilson (1982a, Section 4),[8] the bargaining model of Osborne (1985), the discrete time model with generalized reputation of Kornhauser *et al.* (1989), and Atakan and Ekmekci (2013) who show that equilibrium behavior in a repeated game with two-sided reputation-building is similar to a war of attrition.

The rest of the paper is organized as follows. Section 2.2 introduces the model and the formal notation. Section 2.3 analyzes the general form in the presence of one-sided reputation. Section 2.4 deals with the case of two-sided reputation-building and provides results concerning the induced inefficiency. Section 2.5 discusses wars of attrition with a deadline, and Section 2.6 concludes. Proofs are relegated to the appendix.

---

[7] It is possible to apply some involved transformations between the revision game model and the repeated game model, so that the probabilities of reaching a Pareto inferior outcome are reflected in flow payoffs and in the discount factor. Nevertheless, using any such transformation will strip the model, the methods, and the results of any reasonably intuitive interpretation. We choose not to pursue the search for such an equivalence any further in this paper.

[8] That model uses flow payoffs and strategies in continuous-time, and therefore has a recursive structure.

<table>
<tr><td></td><td></td><td colspan="2">Republicans</td></tr>
<tr><td></td><td></td><td>(L)arger spending</td><td>(R)educed spending</td></tr>
<tr><td rowspan="2">Democrats</td><td>Taxes (U)p</td><td>$u_1(U,L), u_2(U,L)$</td><td>$u_1(U,R), u_2(U,R)$</td></tr>
<tr><td>Taxes (D)own</td><td>$u_1(D,L), u_2(D,L)$</td><td>$u_1(D,R), u_2(D,R)$</td></tr>
</table>

Figure 2.1: *Payoff matrix for rational types*

2.2 Model

Since the fiscal cliff negotiations included many elements and neither side had full control over any of the parameters of the final proposal, we choose here to abstract away from the details and present a simplified version of the process. We consider a two-player Bayesian revision game,[9] summarized by the parameters $(T; u_1, u_2; \xi_1, \xi_2)$. For $i \in \{1,2\}$, $u_i \colon \{U,D\} \times \{L,R\} \to \mathbb{R}$ is a payoff function for the rational type of Player i (see Figure 2.1 below), and $\xi_i \in [0,1]$ is the probability that Player i is a commitment type. We refer to the normal form game associated with the payoff matrix in Figure 2.1 as the *component game*.

The underlying story is that Player 1 (the Democratic Party) has sole authority to set taxes back to their original level (before the expiration of the tax relief acts), whereas Player 2 (the Republican Party) alone can approve a spending increase. We assume that both (U,L) and (D,R) are strict Nash equilibria of the component game and that

$$u_1(U,L) > u_1(D,R) \text{ and } u_2(U,L) < u_2(D,R).$$

These assumptions represent two main features of the story. First, the game is a game of opposing interests; i.e., each party has a preferred solution to the fiscal crisis situation.[10] The Democrats favor raising taxes and keep spending high, while the Republicans would rather lower taxes and embrace the upcoming spending cuts. Second, if both parties insist on their preferred outcome, then the economy falls over the fiscal cliff. That is, if the Democrats keep

[9]Alternatively, we could have used the model of Kamada and Kandori (2011) with a different equilibrium concept that takes into account the evolving probabilities about (nonexistent) irrational types.

[10]We do not consider common interest games since the analysis of such games does not become more interesting with the introduction of reputation. In such games rational players have no incentive to communicate anything other than being rational and willing to cooperate.

taxes up and the Republicans go with the reduced spending, then the resulting scenario is undesirable for both players compared to any agreed-upon solution. Finally, if both parties concede then the combination of low taxes and high spending may drive the economy to a fiscal wall, a situation in which there is too much spending and not enough taxes to cover it, and this alternative is unattractive for both players as well.

Denote the possible types of Player i by τ_i^r (rational type) and τ_i^c (commitment type). We model the commitment type of Player 1 (Player 2) as an agent who can only prepare U (R). We often refer to U and R as the commitment actions. We stress that the rational types' payoffs are independent of their rival's type.

Following the notation used by Calcagno *et al.* (2014), players prepare actions on the interval $[-T, 0]$, and the component game is played once at time 0. At time $-T$, both players simultaneously choose the initial profile of actions. We restrict players to choose pure actions at this initial choice.[11] Between time $-T$ and 0, Players 1 and 2 are called to prepare an action according to two independent Poisson processes. For expositional purposes, we assume in the main text that the frequencies of the Poisson processes are both equal to 1; i.e., each player has on average one revision opportunity per unit of time. All the results for arbitrary revision rates are stated and proved in the appendix. Players are not informed of the realizations of their opponent's Poisson process, but are informed of the current profile at any point in time.[12] This means that players' strategies depend only on their own preparation opportunities, the prepared action profiles, and time, but not on their opponents' preparation opportunities. At $t = 0$ the action profile that has been prepared most recently determines the payoffs for the players. In order for expected payoffs and probabilities of preparing certain profiles in the revision phase to be well defined for all

[11] This restriction is mostly for clarity of exposition and focusing the analysis on what we perceive as the important part of the game, namely, the revision phase. In an earlier draft we show that most of the results continue to hold if we allow mixed action at $-T$. We can also replace this assumption by an exogenous selection of a default profile of actions (U, R).

[12] Loosely interpreted: one player cannot tell when the other player considered preparing a different action, but only when the opponent actually did prepare a different action. It is also possible to consider a model in which players are aware of opponents' revision times. However, analyzing this model becomes cumbersome very quickly due to discontinuities at the revision opportunities. See also the concluding discussion.

strategy profiles, we restrict players' strategies to be measurable with respect to the natural topologies. All the elements of the model are common knowledge, and the type of each player is private information (see the appendix for a formal definition of the strategy space).

We employ the solution concept of Sequential Equilibrium (SE) (Kreps and Wilson, 1982b), which guarantees that even off the equilibrium path a player that revised her action to D or L is believed to be rational with probability 1.[13] We consider the limit set of SE payoffs as the length of time horizon, T, approaches infinity.[14] We are interested mostly in the rational types' payoffs, and so we denote by $\phi(T; u_1, u_2; \xi_1, \xi_2)$ the set of interim SE payoffs of the profile (τ_1^r, τ_2^r). We define the revision equilibrium payoff set of $(u_1, u_2; \xi_1, \xi_2)$ by

$$\bar{\phi}(u_1, u_2; \xi_1, \xi_2) = \lim_{T' \to \infty} \phi(T'; u_1, u_2; \xi_1, \xi_2).$$

Besides the eventual payoffs of the players, we also wish to discuss cases in which a significant delay in reaching an agreement is mandated by equilibrium behavior. We distinguish between two situations in which strategic interaction is significantly delayed. We say that a vector of parameters $(u_1, u_2; \xi_1, \xi_2)$ induces substantial delay if with strictly positive probability the prepared profile does not change throughout the game until close to the deadline. We say that it exhibits last-minute strategic interaction if there is a strictly positive probability that the prepared profile changes close to the deadline. In both cases, we say that the vector of parameters induces inefficiency if there is a strictly positive probability of reaching ex-post Pareto inefficient payoffs. The following definition formalizes these statements.

Definition 2. *A vector of parameters* $(u_1, u_2; \xi_1, \xi_2)$ ***induces***

- ***substantial delay*** *if there exists a time* $-t' < 0$ *and* $\delta > 0$*, such that for every sequence*

---

[13] Alternatively, we could have modelled the commitment type as having the possibility of choosing either action, but with utilities such that U and R are dominant. In this case we would also need a refinement akin to the Intuitive Criterion of Cho and Kreps (1987), since merely using the concept of Sequential Equilibrium would not rule out "strange" beliefs on and off the equilibrium path.

[14] A different interpretation is that the time horizon is fixed, and the revision rates approach infinity while the ratio between them is preserved.

$\{T^k\}_{k=1}^{\infty}$ such that $T^k \to \infty$, and every corresponding sequence of SEs, the probability that only (U, R) is being prepared before time $-t'$ is bounded above δ as k approaches infinity.

- ***last-minute strategic interaction*** *if there exists a time $-t' < 0$ and $\delta > 0$, such that for every sequence $\{T^k\}_{k=1}^{\infty}$ such that $T^k \to \infty$, and every corresponding sequence of SEs, the probability that the prepared profile changes between time $-t'$ and time 0 is bounded above δ as k approaches infinity.*
- ***inefficiency*** *if there exists $\delta > 0$, such that for every sequence $\{T^k\}_{k=1}^{\infty}$ such that $T^k \to \infty$, and every corresponding sequence of SEs, the probability of reaching ex-post inefficient payoffs is bounded above δ as k approaches infinity.*

Note that whenever a vector of parameters induces substantial delay, it also induces last-minute strategic interaction. When a vector of parameters induces last-minute strategic interaction it necessarily induces inefficiency (this is Lemma 50 in the appendix).

2.3 One-sided reputation-building: Last-minute strategic interaction

We begin our analysis by focusing on one-sided reputation-building and its consequences. This simpler environment allows us to introduce some of the intuitions that continue to guide us when studying two-sided reputation-formation. The analysis is required for the study of two-sided reputation-formation also because it represents what happens after either of the players is revealed as a rational type. In what follows we establish first that even when the time horizon becomes long, a prior positive probability of acting "crazy" prevents an opponent from getting the entire surplus.[15]

To do this, we first rephrase and explain a useful definition of Calcagno *et al.* (2014) that summarizes several aspects of the players' bargaining power in our setting.

[15] As is the case in the complete information game (Calcagno *et al.*, 2014, Theorem 3).

Definition 3. *Player i's strength is given by*

$$s_i(u_i) \equiv \frac{|u_i(U,L) - u_i(D,R)|}{[u_i(U,L) - u_1(U,R)] + [u_i(D,R) - u_i(U,R)]}.$$

Player i is stronger than Player j if $s_i(u_i) > s_j(u_j)$. Player i's relative strength (with regard to Player j) is

$$\Delta_{ij}(u_1, u_2) = s_i(u_i) - s_j(u_j).$$

The denominator of the expression in Definition 3 can be thought of as the expected surplus of Player i when both players try to myopically best-respond over a time period, and there is an equal probability that any one of them will manage to stop before the deadline. This value is normalized by the difference between Player i's preferred component game equilibrium payoff and her less-preferred component game equilibrium payoff.[16] Player i is less strong if her preferred component game equilibrium payoff is higher, and is stronger if her less-preferred component game equilibrium payoff is higher, or if $u_i(U,R)$ is higher. Intuitively, lower payoffs for reaching the preferred outcome, and higher payoffs for reaching the less-preferred outcome and or for no agreement, all allow a player to insist longer. Note that for generic payoffs, $\Delta_{ij}(u_1, u_2) \neq 0$ and so one of the players is stronger than the other.[17]

Proposition 12. *Assume Player 1 is stronger than Player 2, $\xi_1 = 0$, and $\xi_2 > 0$; then the revision equilibrium payoff set is bounded away from Player 1's preferred outcome:*

$$u(U,L) \notin \bar{\phi}(u_1, u_2; 0, \xi_2).^{18}$$

[16]Calcagno *et al.* (2014) note that there is an equivalent and perhaps more readable form for the inverse of s_i:

$$\frac{1}{s_i(u_i)} = 1 + 2 \cdot \frac{\min\{u_i(U,L), u_i(D,R)\} - u_i(U,R)}{|u_i(U,L) - u_i(D,R)|}.$$

[17]One may be tempted to compare this strength notion with the concept of risk-dominance (Harsanyi and Selten, 1988). Generally, these two measures do not agree. For example, consider the following games:

	L	R
U	7,5	2,0
D	0,0	3,6

	L	R
U	8,8	0,5
D	0,0	5,9

Player 1 is stronger than Player 2 in both games. However, the risk-dominant equilibrium in the left-most matrix is (U, L), and the risk-dominant equilibrium in the right-most matrix is (D, R).

Proof sketch. The theorem shows that one-sided reputation-building bounds the revision equilibrium payoff set away from the stronger player's preferred outcome as the length of the time horizon tends to infinity. Intuitively, to get this outcome in the limit the rational Player 2 must prepare L on average further and further away from $t = 0$. But this implies that if the stronger player is called to prepare an action at any time before $t = 0$ and the weaker player's prepared action is the commitment action, then the stronger player will attribute a very high probability to the event that the weaker player is a commitment type, and will have a strict incentive to myopically best-respond from there on. This in turn creates an incentive for the rational type of the weaker player to imitate the commitment type, and we get a contradiction. The complete proof is in the appendix. □

To fully understand the role of the random processes in the derivation of the above result, it may be helpful to consider briefly a model in which the times of players' preparations are well known in advance. If there is no incomplete information about players' types, it is easy to see (using backward induction) that the last mover will have to prepare the action related to his less-preferred component game equilibrium. The same pair of strategies will also form a SE in an incomplete information game, as long as the probability of the other player being a commitment type is small enough. Generally, a deterministic order of play enhances such bargaining strengths as first-mover or last-mover advantage or disadvantage. When preparations are random both players can find themselves in a situation where they are the last player to prepare an action, which makes the exact order of preparations irrelevant in determining the bargaining strengths of the two players.

Proposition 12 indicates what kind of outcomes are impossible, but does not provide a full description of what does happen. For example, the expected payoffs may be bounded away from $u(U, L)$ due to some plays ending with Pareto inefficient outcomes, or the equilibrium behavior may dictate arriving at the weaker player's preferred outcome with a

[18] We write $u(U, L)$ instead of $(u_1(U, L), u_2(U, L))$, and similarly for all action pairs.

positive probability. Theorem 13 demonstrates that under certain parameters the equilibrium behavior necessarily involves last-minute strategic interaction that may lead to inefficient outcomes.

Theorem 13. *Assume Player* 1 *is stronger than Player* 2*, and* $\xi_1 = 0$*. Then there exists* $\bar{\bar{\xi}}_2 > 0$ *such that for every* $\xi_2 \in \left(0, \bar{\bar{\xi}}_2\right)$*, the parameters induce last-minute strategic interaction and inefficiency.*

Proof sketch. If the probability that Player 2 is a commitment type is small enough, then Player 1 can get a payoff strictly above $u_1(D, R)$ with high probability by never changing her action and waiting for Player 2 to best-respond. That is, some histories lead to Player 1 getting her most-preferred payoff, and Player 2 getting less than her most-preferred payoff. We know from Proposition 12 that Player 2 cannot prepare the action L too soon, and when she does not, Player 1 cannot always prepare D or else Player 2 will have a profitable deviation. This necessarily implies last-minute strategic interaction, which in turn directly causes inefficiency. The complete proof is in the appendix. □

We are inclined to interpret Theorem 13 as relating to different notions of advantage. These two flavors of strategic advantage are reminiscent of those discussed in the context of bargaining. For example, Rubinstein (1982) uses slight differences in impatience to determine the bargaining outcome, whereas Abreu and Gul (2000) suggest irrationality as an explanation for the final division of residual claims in the bargaining procedure. In our case the strength of Player 1 cannot prevent the rational type of Player 2 from imitating the commitment type and creating reputation. At the same time Player 1 cannot completely give up her strong bargaining posture too soon, as she knows that rationality dictates a change in the revisions of the rational type of Player 2 close to the deadline. As the probability of actually facing a commitment type becomes small, Player 1 has an incentive to preserve some bargaining power to the end of the revision phase. This means that for an open set of parameters, one-sided reputation-building leads to last-minute strategic interaction and inefficiency.[19] In Section 2.5 we strengthen this result for the case of War of Attrition with

[19]Strictly speaking, we have not shown that the set of parameters is open. However, looking at the proof of

Poisson arrivals (Corollary 20).

An important question, to be discussed further in the two-sided reputation formation case, is whether the effect of reputation vanishes as the probability of a commitment type tends to zero. While we do not prove that it necessarily does, we can demonstrate that the efficient outcome related to the subgame perfect equilibrium of the complete information game is a limit of a sequence of outcomes in games with diminishing probability for a commitment type. That is, we show that $\bar{\phi}$ is lower hemi-continuous in ξ_2 at 0. To prove this result we explicitly construct a specific sequence of equilibria that has a simple structure of using cutoff strategies (that is, insisting up to a certain time and then myopically best-responding from there on), and show that the related expected payoffs converge to $u(U, L)$.

Proposition 14. *Assume Player* 1 *is stronger than Player* 2*, and* $\xi_1 = 0$*. Then Player* 1*'s preferred outcome is in the limit of the revision equilibrium payoff set as*[20] $\xi_2 \to 0$*:*

$$u(U, L) \in \liminf_{\xi_2 \to 0} \bar{\phi}(u_1, u_2; 0, \xi_2).$$

2.4 Two-sided reputation-building: Falling over the cliff

As demonstrated in the previous section, one-sided reputation formation may lead to inefficiency under certain conditions. The inefficiency is realized on the equilibrium path when the Democrats insist on keeping taxes high (U) and the Republicans insist on reducing spending (R), a scenario that leads to a fiscal cliff. The basic argument was that the Republicans, who are perceived as having a positive probability of being committed to reducing spending, cannot fold too quickly, or else the entire surplus will be taken by the Democrats who are stronger along other bargaining dimensions. This section applies

Theorem 13, we can see that $\bar{\xi}_2$ can be chosen as a continuous function u_1 and u_2, and hence finding an open set of parameters is trivial.

[20] Note that $\bar{\phi}(u, 0, \xi_2)$ is a set. The operator lim inf here means that for every sequence $\xi_2^k \to 0$ we can choose equilibrium payoffs $u^k \in \bar{\phi}(u, 0, \xi_2^k)$ such that $u^k \to u(U, L)$. Note also that Proposition 14 does *not* completely characterize the limit set of the payoffs.

a similar logic to situations with two-sided reputation-formation, and describes how it necessarily induces a positive probability of falling over the fiscal cliff. That is, the probability of seeing any action prepared in the game other than the commitment actions is bounded below 1 until the very end of the game, and the expected payoffs cannot be on the Pareto frontier. We characterize the inefficiency for the special case of equilibrium in cutoff strategies, and present comparative statics and limit results. Finally, we present a variety of numerical and analytic calculations that indicate that the inefficiency can be substantial for reasonable values of the parameters.

As described above, not only does our first theorem in this section predict a shift from the complete information outcome, but it also demonstrates that with positive probability a long delay must occur before any player prepares anything but the commitment action. This delay in turn must cause inefficiency, as agreement might not be reached before the deadline. Theorem 15 therefore combines and strengthens the results of Proposition 12 and Theorem 13 for the case of two-sided reputation-building.

Theorem 15. *Assume Player* 1 *is stronger than Player* 2,[21] *and* $\xi_1, \xi_2 > 0$. *Then substantial delay and inefficiency are induced.*

The intuition for Theorem 15 is an extension of the one used to derive the one-sided reputation results. The rational type of the weaker player (in terms of payoffs) cannot give up on building reputation too quickly in equilibrium. If she does, then, by extension of Calcagno *et al.* (2014, Theorem 3), her payoff will approach her least preferred component game equilibrium payoff, while pretending to be a commitment type will guarantee a larger payoff. Given that there is a time before which the weaker player's probability of revealing her rationality is bounded below 1, there is also an earlier time before which the rational type of the stronger player cannot reveal her rationality too early either. If she does that her utility will be bounded away from her preferred outcome (by Proposition 12), and deviating

[21] This assumption is only for the sake of simplicity. The case of equal strengths is susceptible to similar analysis but requires proving first that if players are equally strong, $\xi_1 > 0$ and $\xi_2 = 0$, then equilibrium payoffs approach $u(U, L)$ as T approaches infinity.

ensures a payoff that approaches the preferred outcome. This dictates substantial delay, which necessarily leads to inefficiency.

The above theorem points to some of the problematic efficiency properties of two-sided reputation-building. The rest of this section is dedicated to uncovering the magnitude of the inefficiency, and how different parameters of the game affect it. In what follows we focus on the simple equilibrium form in which each rational type insists on her commitment action until some point in time and myopically best-responds (according to the component game's payoffs) from there on. We know that in order for these strategies to form an equilibrium, each rational type must be indifferent between both actions at the switching point, and this together with the Bayesian updating of beliefs allows us to solve for the cutoff times. The following proposition gives us some basic comparative statics regarding the extent of inefficiency that Theorem 15 predicts.

Proposition 16. *If $\xi_1, \xi_2 > 0$, both players play cutoff strategies, and Player 2's cutoff is earlier than Player 1's cutoff, then the probability of reaching an ex-post Pareto inferior outcome is*

1. *increasing in $u_1(D, R)$, and decreasing in $u_1(U, L)$ and $u_1(U, R)$,*
2. *increasing in $u_2(D, R)$ and $u_2(U, R)$, and decreasing in $u_2(U, L)$.*

It is instructive to read Proposition 16 in light of Definition 3, and see that each element of the bargaining strength affects the inefficiency in the same way it affects the relative strength, $\Delta_{12}(u_1, u_2)$. Changes in the payoffs that push the relative strength away from zero decrease the inefficiency, and changes that push it towards zero increase the inefficiency. From symmetry it follows that the maximal inefficiency is attained when $\Delta_{12}(u_1, u_2) = 0$. To get a feeling of what that maximal inefficiency might be, Theorem 17 below takes both probabilities of being a commitment type to zero.

Theorem 17.
1. *Assume Player 1 is stronger than Player 2, and both players play cutoff strategies. Then in the limit as $\xi_1 \to 0$ and $\xi_2 \to 0$, the probability of reaching ex-post inefficiency tends to zero.*

2. *Assume players are equally strong ($\Delta_{12}(u_1, u_2) = 0$), and both players play cutoff strategies. Let the sequence $(\xi_1^k, \xi_2^k)_{k=1}^{\infty}$ be such that $\lim_{k\to\infty} \xi_1^k = \lim_{k\to\infty} \xi_2^k = 0$, and $\lim_{k\to\infty} \frac{\xi_2^k}{\xi_1^k} \leq 1$. Then in the limit as $k \to \infty$, the probability of reaching ex-post inefficiency tends to*

$$\left[\frac{u_1(U,L) - u_1(D,R)}{u_1(U,L) + u_1(D,R) - 2u_1(U,R)}\right] \times \lim_{k\to\infty} \frac{\xi_2^k}{\xi_1^k}.$$

The interpretation of Theorem 17 is that as the probability of commitment types goes to zero, the inefficiency (generically) goes to zero as well. Nevertheless, if the probability that either of the sides is a commitment type does not merely serve as a refinement tool, but rather constitutes an essential part of the dynamics, then the balance of the different forces becomes important. As the second part of the theorem shows, even for relatively small probabilities of being commitment types, the inefficiency can be substantial. Roughly, if the players are close to being equally strong in terms of payoffs, then reputation effect may cause significant efficiency loss. However, if one of the players is much stronger (in terms of payoffs), reputation effects only mildly hurt efficiency. Similarly, if one player's reputation is much stronger than the other player's (the ratio $\frac{\xi_2}{\xi_1}$ is very small or very big), then the inefficiency is negligible.

To illustrate the extent to which the equilibrium result may be inefficient we solve for a variety of parameters. We first note that when the two players are equally strong we can use the formula of Theorem 17 to compute the limit inefficiency. We assume $\xi_1 = \xi_2$ and (without loss of generality) normalize $u_1(U,R)$ to be zero, and get the reduced formula

$$\text{Inefficiency} = \frac{u_1(U,L) - u_1(D,R)}{u_1(U,L) + u_1(D,R)}.$$

This implies that for a standard battle of the sexes payoff matrix, the inefficiency tends to $\frac{1}{3}$ in the limit. If the stakes are higher, say if $u_1(U,L) = u_2(D,R) = 99$ and $u_1(D,R) = u_2(U,L) = 1$, then the inefficiency is 98%. In this case equilibrium behavior dictates that both players insist on playing the commitment action until very close to time 0, and the first to be called to prepare an action chooses her less-preferred component game equilibrium and gets a payoff of 1, whereas the other player gets 99. However, the time at which players stop preparing the commitment action is so close to the deadline that only with probability 2% is either of them is going to be called, which pins their expected payoff at exactly 1.

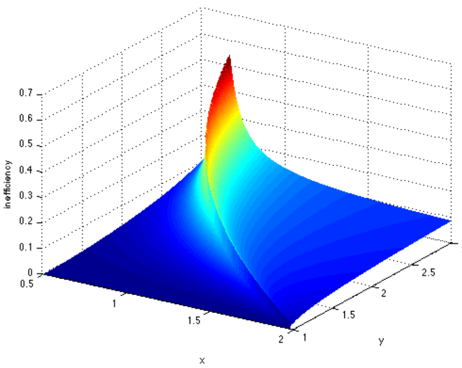
Figure 2.2 illustrates several of the properties that were just discussed. In this figure we see the extent of inefficiency for the game form presented in the top left panel for $x \in (0.5, 2)$ and $y \in (1, 3)$, and for three values of ξ_1 and ξ_2. The three shapes resemble shark fins that become narrower as ξ_1 and ξ_2 become smaller. This demonstrates that the maximal inefficiency is always at $\Delta_{12}(u_1, u_2) = 0$ and does not vanish, and that everywhere else the inefficiency tends to zero as the ex-ante probability of commitment types approaches zero. Figure 2.3 offers a different view of the same family of payoff matrices, but here the payoffs are held constant and the ex-ante probabilities of commitment types vary. A careful look at those graphs reveals that there is a trade-off between strength (in terms of payoffs) and commitment power when it comes to bargaining posture, which in turn leads to inefficiency.

2.5 Wars of attrition with Poisson arrivals

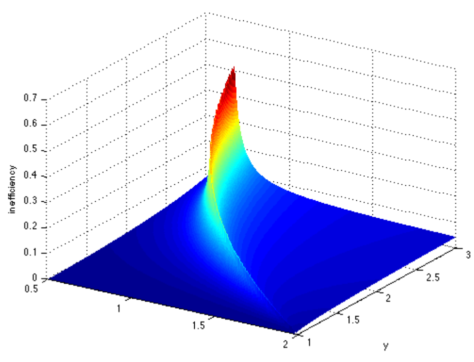
In many negotiations (including those related to the fiscal cliff), reaching a final agreement before the deadline is also a possibility. This is inherently different from the model of revision games in which even after one of the parties decides to concede, it can still return to its previous bargaining position. Assuming that a concession ends the game simplifies the model and turns it into a model of a war of attrition. Wars of attrition can generally be described as situations in which the first player to take action over some defined period of time determines the payoffs for both players. This model was first introduced by Maynard Smith (1974) to model biological situations in which two animals fight over a disputed territory. While the first animal to leave the territory receives a lower payoff than its rival, both animals suffer from the preceding fight. The framework also captures some of the economic incentives that govern firms when they are engaged in a patent race or when trying to become a monopolist in a market that cannot sustain multiple competitors. An

	L	R
U	2, 1	0, 0
D	0, 0	x, y

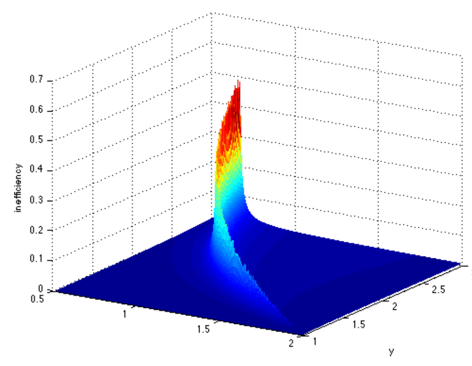
The payoff matrix



$\xi_1 = \xi_2 = 0.1$

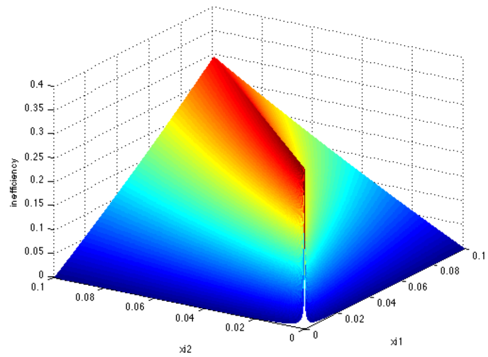


$\xi_1 = \xi_2 = 0.05$

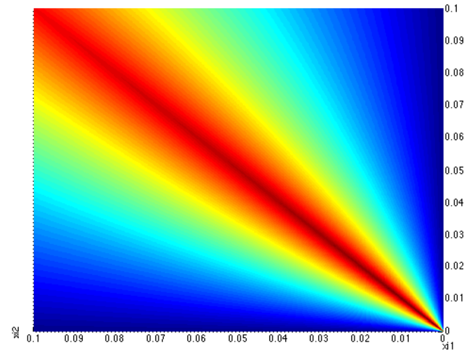


$\xi_1 = \xi_2 = 0.01$

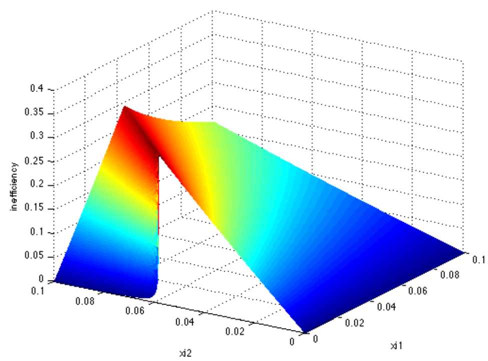
Figure 2.2: *Varying payoffs*



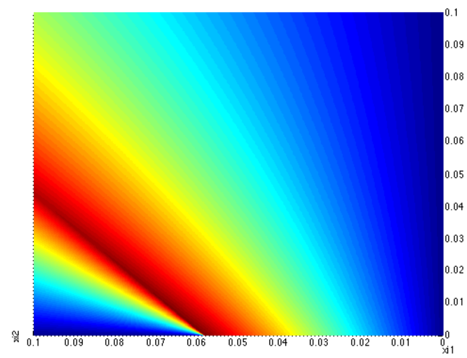
$x = 1$, $y = 2$ *(side)*



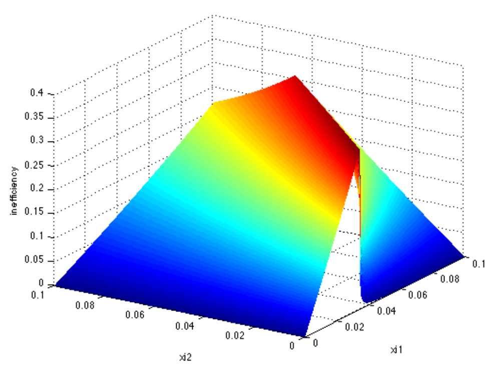
$x = 1$, $y = 2$ *(top)*



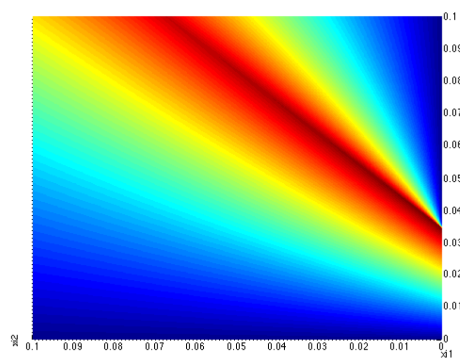
$x = 1$, $y = 1.9$ *(side)*



$x = 1$, $y = 1.9$ *(top)*



$x = 0.97$, $y = 2$ *(side)*



$x = 0.97$, $y = 2$ *(top)*

Figure 2.3: *Varying commitment probabilities*

extensive literature studies different aspects of the game.[22]

As mentioned in the Introduction, variants of wars of attrition that include reputation have been analyzed in the literature in both discrete and continuous-time. Here we introduce a new timing structure, namely, continuous and finite time, where players can exit only when being called to play by a stochastic process. We restrict ourselves to payoffs that correspond to a 2x2 opposing interests game when the stochastic processes governing the timing of the players' decisions are independent Poisson processes. One specific instance from this class of games is the "three-state example" analyzed (without reputation effects) by Kamada and Sugaya (2010, Section 6.1).

Roughly speaking, all of the previous results hold for this class of games. However, it turns out that in the war of attrition with Poisson arrivals all equilibria have the special structure of both players exiting with probability one when being called to play after some cutoff time, and at least one of the two players exiting with zero probability before that time. Furthermore, every equilibrium payoff can be mimicked using cutoff strategies for both players. We use these observations to pin down the limit results.

The fact that the strategies are simpler allows us to provide even sharper results. We continue to use the same basic definitions of types and utilities (see the appendix for formal definitions of histories, information sets, and strategies). As before, we now let $\phi^{\text{woa}}(T; u_1, u_2; \xi_1, \xi_2)$ denote the set of interim SE payoffs of the profile (τ_1^r, τ_2^r) in the war of attrition with Poisson arrivals and payoffs given by u_1 and u_2. We define $\bar{\phi}^{\text{woa}}(u_1, u_2; \xi_1, \xi_2)$ as the limit payoffs set of this game as T approaches infinity.

Lemma 18. *In the war of attrition with Poisson arrivals (and incomplete information), for any T, all sequential equilibria have the property that there exists a cutoff time $-t^*$ such that both players exit if being called to play after $-t^*$, and at least one of them exits with zero probability before $-t^*$.*

Corollary 19. *In the war of attrition with Poisson arrivals (and incomplete information), for any T all SE payoffs can be attained by SEs in which both players use cutoff strategies.*

[22] Such as random rewards (Bishop *et al.*, 1978), asymmetric equilibria (Nalebuff and Riley, 1985), effects of continuous-time (Hendricks *et al.*, 1988), strong evolutionary equilibria (Riley, 1980), and more.

The intuition for the proof of Lemma 18 is that players can use strategies that do not depend on their previous decision opportunities because the other player's inference does not depend on those opportunities. Furthermore, the nature of players' incentives is such that if one player exits at some time $-t$, and there is a strictly positive probability that the second wants to exit from then on until some future time $-t' > -t$, then after time $-t'$ the first player will definitely want to exit. This "pairwise monotonicity" property, which is precisely defined in the appendix, can be used to show that all equilibria have a cutoff time $-t^*$. Corollary 19 then follows by reducing the probability with which the "weaker" player exits before $-t^*$ to an interval just before $-t^*$, while preserving the conditions for equilibrium and giving the same expected payoffs as the original SE.

Lemma 18 and Corollary 19 represent the key difference between the analysis presented in the previous sections and the analysis of a war of attrition. Specifically, we can use Lemma 18 and Corollary 19 and repeat the same arguments from before to get sharper results. In the case of one-sided reputation formation, the results about efficiency and inefficiency under certain parameters are stronger, and equilibrium payoffs are uniquely determined for small enough values of ξ_2. Uniqueness is established by ruling out any other equilibrium payoffs using only cutoff strategies. Then the construction that appears in the proof of Proposition 14 nails down the exact values of the cutoff times. These results again demonstrate that the ability to build reputation is an extremely important property of the bargaining process.

Corollary 20. *In a war of attrition model, assume Player* 1 *is stronger than Player* 2 *and* $\xi_1 = 0$*. Then there exists* $\bar{\bar{\xi}}_2$ *such that for every* $\xi_2 \in (0, \bar{\bar{\xi}}_2)$*, the parameters induce substantial delay and inefficiency.*

Corollary 21. *In a war of attrition model, assume Player* 1 *is stronger than Player* 2*; then her preferred outcome is the unique limit of the equilibrium payoff set as the probability of Player* 2 *being the commitment type approaches zero. Formally,* $\liminf_{\xi_2 \to 0} \bar{\Phi}^{woa}(u_1, y_2; 0, \xi_2) = \{u(U, L)\}$*.*

As for the case of two-sided reputation-building, all the results from Section 2.4 hold, but whenever the restriction about cutoff strategies appears it is redundant. We here avoid repeating these results or providing their proofs, which resemble the proofs that were used for the revision game model but also employ Corollary 19.

For the sake of completeness it is useful to review how our results relate to previous models of the war of attrition with incomplete information in continuous-time. First, as mentioned before, the continuous-time framework makes proving uniqueness results far more complicated than the results we present here. Our model relies on discrete actions, and so we can more easily apply the sequential rationality property to establish uniqueness. Second, a striking feature is that the rates at which players exit in the war of attrition are similar. Since we employ cutoff strategies, players exit according to the two Poisson processes that govern their action opportunities starting from a certain time. Osborne (1985) provides an expression that amounts to exactly the same exit rate when the players are risk-neutral.

2.6 Conclusion

While the fiscal cliff negotiations ended with an agreement between the two parties, similar bargaining situations often lead to an outcome undesirable to both sides in the form of passing the deadline. While ignoring some real-life reasons that may cause this to happen (animosity, miscommunication, and so on), we provide an intuitive explanation for such events. Our prediction of non-degenerate equilibrium play with the possibility of a deadline effect stands in sharp contrast with previous results on complete information play, and suggests that reputation formation may play a crucial part in delaying meaningful bargaining in the presence of a deadline.

It is likely that the effects mentioned in this paper would only become stronger if there were an element of repeated bargaining between the parties, as is natural in the political arena. In this case, reputation formation would not only serve the present agenda, but would also be beneficial for future negotiations. A different direction in which our results can be extended is to consider a game with incomplete information on the rates of revision opportunities. In such a game each player may try to persuade her opponent that her rate of revision opportunities is low. This creates an incentive for the opponent to give up her bargaining posture faster than she would have done had she known she was facing a player who can still react with high probability before the deadline. This particular kind of incomplete information may serve as a good model of the level of interest that each of the players invests in the game, and provides yet another reason why even players who care dearly about the results of the game would like to hold their position as the deadline approaches.

Finally, it is possible to consider an alternative model in which players' revision opportunities are observed by their opponents. While non-observable arrivals of revision opportunities are slightly more intuitive, the alternative model may also be relevant for certain applications. However, once players' revision opportunities become common knowledge delicate issues of playing mixed strategies arise. Studying the structure of strategies in a war of attrition with Poisson arrivals and incomplete information under this observability condition also becomes tricky, and it remains an open question whether or not cutoff strategies can be used to exhaust the entire space of equilibrium payoffs.

Chapter 3

Implications of Capacity Reduction and Entry in Many-to-One Stable Matching

3.1 Introduction

Many economic environments can be characterized as matching markets in which agents of two different populations contract to achieve a common goal. Two prominent examples, which were extensively studied in the market design literature, are school choice problems and allocations of interns to hospitals.[1] Natural experiments with various centralized matching mechanisms have made it quite clear that the single most important property determining the success of a mechanism is its stability or lack thereof (Roth, 1991). The importance of stability drove many economists to study the properties of stable mechanisms in one-to-one matching markets, as well as more general and more realistic models of matching that allow agents to contract with multiple parties, to specify salaries or other benefits as part of the contracts, and so on.

As is by now well-known, participants in markets governed by stable mechanisms may

[1] See Abdulkadiroğlu and Sönmez (2003) and Roth (2002) for a survey of these case studies.

have an incentive to report their preferences and attributes (such as capacities) untruthfully. Roth (1982) proves that any stable mechanism can be manipulated by some participants, and Sönmez (1997) shows that every stable mechanism is manipulable via capacities.[2] We follow up on this literature by investigating the implications of capacity reductions in many-to-one markets, and show that if a capacity reduction is binding then there exists a mechanism-independent non-empty set of doctors and a related set of hospitals, such that every doctor in the first set is worse off and every hospital in the second set is better off following the capacity reduction (under any stable mechanism). Moreover, if it so happens that the hospital with the reduced capacity belongs to the set of hospitals mentioned above, then this hospital can report a lower capacity and be certain to get a better outcome, even without knowing which stable matching will be realized. That is, in this situation reporting a reduced capacity dominates truth-telling regardless of the stable mechanism that is being used. We then consider a larger domain of preferences, namely substitutable preferences that satisfy the law of aggregate demand, and prove that a capacity reduction has slightly different yet similar consequences.

We apply our results to the study of entry in many-to-one markets. While early works on entry in the matching literature focused on specific mechanisms,[3] Theorem 2.26 of Roth and Sotomayor (1990), which is an adaptation of previous work on assignment games by Mo (1988), stands out as it holds regardless of the matching process. We show that their result is false in many-to-one matching markets when a doctor enters the market. That is, there

[2]Further contributions include Konishi and Ünver (2006), who show that in a game of capacity manipulation every pure strategy equilibrium is weakly preferred by the hospitals to the outcome of any larger capacity profile, and on this matter see also Ehlers (2010), Kesten (2011), Kojima (2006), Mumcu and Saglam (2009) and Romero-Medina and Triossi (2007). Azevedo (2014) uses a continuum model (Azevedo and Leshno, 2011) to perform a more detailed analysis of hospitals' incentives to reduce capacity.

[3]For example, the results of Roth and Sotomayor (1990, Theorems 2.25 and 5.35), which are reminiscent of those by Kelso and Crawford (1982, Theorem 5) and Demange and Gale (1985, Corollary 3), show that it is possible to make weak comparisons with regard to the entire population when either the hospital-optimal or the doctor-optimal stable matching mechanisms are being used. That is, one can predict that all agents on one side of the market are made weakly better off or weakly worse off following the arrival of a new agent. By technically treating hospital's positions as separate entities (i.e., the equivalent marriage market approach), these theorems also imply that capacity reductions have a similar effect under these two mechanisms. The weak comparison results are extended to other matching environments by Crawford (1991), Hatfield and Kominers (2013) and Hatfield and Milgrom (2005, Theorem 6).

are cases in which a doctor enters the market and is matched, and yet the conclusion of the mentioned theorem about which doctors become worse off does not hold. We establish a weaker result by allowing only responsive preferences and focusing on the set of hospitals that become better off following a doctor's entry. In the case of hospitals being at full capacity prior to the entry we can also identify a non-empty related set of doctors that become worse off. We also show that the original statement does hold when a hospital joins the market.

Manipulation via truncations and dropping strategies is another strategic issue frequently discussed in the matching literature.[4] Roth and Vande Vate (1991) note that in a one-to-one matching environment it is sufficient to consider a special subclass of preference manipulations called truncation strategies, in which the hospital reports a preference that coincides with the real preference in its ordering of acceptable alternatives, but may misreport its least preferred acceptable alternative.[5] We prove that whenever a hospital uses a (binding) truncation, it harms some non-empty set of doctors and benefits a related set of hospitals. However, as soon as attention is turned to many-to-one matching, hospitals may want to resort to the broader class of dropping strategies (Kojima and Pathak, 2009). We show that even a binding dropping strategy may not have similar implications.

Our somewhat partial welfare analysis is mandated by the broad domain of matching environments to which it can be applied. For example, stability has emerged in the theoretical literature as something one may expect even in decentralized markets in which the precise matching mechanism may not be so clearly defined. Roth and Vande Vate (1990) show that reasonable random processes will converge to a stable matching, but may converge to different stable matchings at different realizations of the random process. While many of the results in the matching literature do not apply when the stable matching is stochastically determined, our results are robust even under such uncertainty. The

[4]Other forms of manipulation, which are not discussed in this paper, are via pre-arranged matches (Sönmez, 1999), creating fictitious doctor records (Afacan, 2011), and application fee manipulation (Afacan, 2012).

[5]For more on truncation strategies, see Ashlagi and Klijn (forthcoming), Coles and Shorrer (2014), Ehlers (2004), Ma (2002), Ma (2010), and Roth and Rothblum (1999).

applicability of our results to decentralized markets also explains why our analysis is useful despite the simplifying assumption of complete information over preferences, as it allows us to be agnostic about what information is available regarding the details of the matching process and which stable matching will be selected.

The paper proceeds as follows. Section 3.2 introduces the model. Section 3.3 studies the implications of capacity reduction. Section 3.4 applies them to study entry in many-to-one markets. Section 3.5 considers the effects of truncation and dropping strategies, and Section 3.6 concludes. Most proofs are relegated to the Appendix.

3.2 The model

Let D be a finite set of doctors and H a finite set of hospitals. A doctor can be matched to at most one hospital, but hospitals can be matched to several doctors. Doctors and hospitals have strict preferences over possible matchings with agents from the other side of the market, as well as remaining unmatched (which we denote by $\emptyset$). The strictness of the preferences is crucial to our results (see also the closing discussion). Each $d \in D$ is endowed with a transitive and complete preference $\succ_d$ over $H \cup \{\emptyset\}$, and every $h \in H$ is endowed with a transitive and complete preference $\succ_h$ over 2^D, i.e., all possible subsets of D. We write preferences as ordered lists and abbreviate them by omitting those elements that are unacceptable (that is, less preferred to remaining unmatched). Hospital h also has a capacity q_h, which is the cardinality of the largest acceptable set doctors.

We will denote by $\mathcal{P}$ the set of all problems, with a general element

$$P = (D, H, \{q_h\}_{h \in H}, \{\succ_i\}_{i \in D \cup H}) .$$

Throughout this paper we restrict ourselves to the analysis of two important sub-domains of preferences for the hospitals, namely the domain of responsive preferences, and the domain of substitutable preferences that satisfy the law of aggregate demand. Both of these domains have been extensively used in the literature, mostly due to their intuitive interpretations and their convenient mathematical properties when it comes to the analysis of stable matchings.

Hospital h is said to have **responsive preferences** and $\succ_h$ is said to be **responsive** if for every set of doctors $D' \subseteq D$ such that $|D'| \leq q_h$, every $d_1 \in D'$ and every $d_2 \in D \setminus D'$ we have:

1. $D' \succ_h D \cup \{d_2\} \setminus \{d_1\} \iff \{d_1\} \succ_h \{d_2\}$
2. $D' \succ_h D' \setminus \{d_1\} \iff \{d_1\} \succ_h \emptyset$

Loosely interpreted, a responsive preference is consistent with its ordering of individual doctors, and prioritizes a set with more acceptable doctors (as long as it does not exceed its capacity). We denote the subset of problems in which hospitals have responsive preferences by $\mathcal{P}^R$.

Define for any hospital $h \in H$ the choice function $C_h \colon 2^D \to 2^D$ by:

$$C_h(D') = \max_{\succ_h}\{D'' \mid D'' \subseteq D'\}$$

We say that hospital h's preference relation is **substitutable** if $\forall D'' \subseteq D' \subseteq D : D'' \cap C_h(D') \subseteq C_h(D'')$. Intuitively, a hospital's preference relation is substitutable if a doctor who is accepted remains accepted when the set of doctors under consideration shrinks. Hospital h's preference relation **satisfies the law of aggregate demand** if $\forall D'' \subseteq D' \subseteq D : |C_h(D'')| \leq |C_h(D')|$.[6] We denote the subset of problems in which hospitals have substitutable preferences that satisfy the law of aggregate demand by $\mathcal{P}^{\mathrm{SL}}$. It is easy to verify that a responsive preference is substitutable and satisfies the law of aggregate demand, and so $\mathcal{P}^R \subseteq \mathcal{P}^{\mathrm{SL}}$.

A **matching** is a function $\mu \colon D \cup H \to H \cup 2^D$ such that every doctor is either matched to one hospital or remains unmatched (denoted by $\emptyset$), every hospital is matched to a set of doctors,[7] and the matching is consistent. Formally: for every $d \in D$ we have $\mu(d) \in H \cup \{\emptyset\}$, for every $h \in H$ we have $\mu(h) \in 2^D$, and $\mu(d) = h$ if and only if $d \in \mu(h)$.

[6] For matching markets without contracts such as ours, this property was first introduced under the name *cardinal monotonicity* by Alkan (2002).

[7] Restricting the definition to only allow hospitals to be matched to sets of doctors smaller than their capacity does not affect our analysis.

A matching is individually rational if all doctors and hospitals (weakly) prefer their matching to remaining unmatched. A matching is unblocked if there is no set of doctors and a hospital such that every doctor prefers the hospital to her current match, and the hospital prefers the union of this set of doctors and some subset of its currently matched doctors to its current match.[8] A matching is stable if it is individually rational and unblocked. Formally, matching μ is **stable** if it is:

1. **Individually rational**: For every $i \in D \cup H$, $\mu(i) \succeq_i \emptyset$.[9]

2. **Unblocked**: There exist no $h \in H$, $D' \subseteq D$ and $D'' \subseteq \mu(h)$ such that for all $d' \in D'$, $h \succ_{d'} \mu(h)$ and $D' \cup D'' \succ_h \mu(h)$.

For any $P \in \mathcal{P}$, we denote by $\Psi(P)$ the set of stable matchings. A mechanism is a function from $\mathcal{P}$ to matchings. Mechanism ψ is a **stable mechanism** on sub-domain $\mathcal{P}'$ if for any $P \in \mathcal{P}'$, $\psi(P) \in \Psi(P)$, that is, its outcome for problems in $\mathcal{P}'$ is always a stable matching. Throughout the paper it will be clear what is the sub-domain currently under discussion and we will not be explicit about it when we refer to a specific stable mechanism. We denote by ψ^D the doctor-optimal stable mechanism, and by ψ^H the hospital-optimal stable mechanism. The existence of these two mechanisms is proved by Gale and Shapley (1962) for the domain of responsive preferences, and by Roth (1984b) for the domain of substitutable preferences.

We study preference domains in which the "rural hospital theorem" holds, i.e., the set of unmatched doctors remains unchanged for all stable matchings, as well as the set of doctors assigned to hospitals that fail to reach their capacity. This result was first proved by Roth (1984a, Theorem 9), and was extended to the domain of q-separable and substitutable preferences by Martínez *et al.* (2000). For a recent treatment and a survey of the different extensions see Klijn (2011). It immediately follows that given a problem $P \in \mathcal{P}^{\mathrm{SL}}$ any agent i is matched to the same number of partners under all stable matchings, and we denote this

[8]We employ here the notion of setwise stability. In the many-to-one markets with substitutable preferences that we consider, this is equivalent to pairwise stability (Blair, 1988).

[9]One can also define individual rationality for the hospitals by requiring that $\forall h \in H : \mu(h) = C_h(\mu(h))$. This yields the same stability concept and does not affect our results.

number by $m_i(P)$.

3.3 Capacity reduction

For any hospital $h \in H$ with preference relation $\succ_h$ and capacity q_h and for any $q' < q_h$, define $\succ_h^{q=q'}$ as the **preference relation derived from** $\succ_h$ **by imposing capacity** q', that is, for all $D', D'' \subseteq D$:

$$\begin{aligned} D' \succ_h^{q=q'} D'' \iff & \big(|D'| \leq q' \text{ and } |D''| \leq q' \text{ and } D' \succ_h D''\big) \text{ or} \\ & \big(|D'| \leq q' \text{ and } |D''| > q'\big) \text{ or} \\ & \big(|D'| > q' \text{ and } |D''| > q' \text{ and } D' \succ_h D''\big) \end{aligned}$$

Lemma 22. *Let $\succ_h$ and q_h be hospital h's preference relation and capacity respectively, and $\succ_h^{q=q'}$ the preference relation derived from $\succ_h$ by imposing capacity q'. Then:*[10]

1. *If $\succ_h$ is responsive, then so is $\succ_h^{q=q'}$.*[11]
2. *If $\succ_h$ is substitutable and satisfies the law of aggregate demand, then $\succ_h^{q=q'}$ satisfies the law of aggregate demand.*
3. *Even if $\succ_h$ is substitutable and satisfies the law of aggregate demand, it does not necessarily follow that $\succ_h^{q=q'}$ is substitutable.*

Our first result states that if preferences are responsive and a hospital reduces its capacity (or at least reports such a reduced capacity) beneath the number of doctors who are assigned to it in a stable matching, then it is possible to find some non-empty set of doctors and a (possibly empty) set of hospitals such that each doctor in the former set is worse off, and each hospital in the latter set is better off following the reduction. It is important to

[10]Parts of this result are similar in spirit to Mongell and Roth (1986) who demonstrate why imposing budget constraints on a firm in the model of Kelso and Crawford (1982) may cause the gross substitutes condition to stop holding for the budget-constrained firm.

[11]That is, when the capacity used in the definition of responsiveness is taken to be q'.

note that throughout this paper comparisons are strict, and so "better (worse) off" means "strictly better (worse) off". We explicitly note when the comparison is weak. Furthermore, whenever we make a comparison between outcomes it is with respect to the true capacity and full preferences (in the next section, including possible entrants), and not with respect to any partial preferences or reported capacities.

Theorem 23. *Assume $P \in \mathcal{P}^R$.*[12] *Let $h_0 \in H$ and $0 \le q' < m_{h_0}(P)$, and set*

$$P' = \left(D, H, \left(q', q_{-h_0}\right), \{\succ_i\}_{i \in D \cup H \setminus \{h_0\}} \cup \{\succ_{h_0}^{q=q'}\}\right).$$

Then there exists a non-empty subset of doctors $S \subseteq D$, such that under any stable mechanism ψ, every doctor in S is worse off and every hospital in $\left\{h \mid \psi^D(P')(h) \cap S \neq \emptyset\right\}$ is better off under $\psi(P')$ compared to $\psi(P)$.

Note that the relevant set of hospitals mentioned in the theorem includes exactly those hospitals that employ some doctors from the specified set of doctors following the capacity reduction under the doctor-optimal stable matching. The two sets are invariant under different stable mechanisms, and therefore the conclusion does not depend on the stable mechanism used. This mechanism-free welfare comparison is equivalent to the claim that all the specified doctors are worse off and all hospitals in the related set are better off under *any* stable matching following the capacity reduction compared to *any* stable matching before it.

The economic intuition and the proof of this theorem are both loosely based on Theorem 2.26 in Roth and Sotomayor (1990). For the sake of completeness we reproduce the statement of this theorem here. To make the connection clearer, the theorem is reformulated in terms of doctors and hospitals instead of men and women. Note however that the original theorem is stated in "the opposite direction" (a hospital enters the market), whereas we focus on capacity reductions that are a generalization of leaving a market.

Theorem 24 (Roth and Sotomayor, 1990, Theorem 2.26). *Assume all hospitals have a capacity*

[12] Theorem 23 holds for a more general preferences domain, namely q-separable and substitutable preferences (Martínez *et al.*, 2000). However, the proof is far more involved, and so we use the stronger assumption of responsiveness. For further details on the proof under the weaker assumptions, please contact the author.

of one. Suppose hospital h_0 is added to the market, i.e.,

$$P' = (D, H \cup \{h_0\}, (1, 1, \ldots, 1), \{\succ_{h_0}\} \cup \{\succ_i\}_{i \in D \cup H}) .$$

If $m_{h_0}(P') > 0$, then there exists a non-empty subset of doctors $S \subseteq D$, such that under any stable mechanism ψ, every doctor in S is better off and every hospital in $\{h \mid \psi^D(P)(h) \in S\}$ is worse off under $\psi(P')$ compared to $\psi(P)$.

The idea in the many-to-one case is that h_0's capacity reduction, which is in a sense opposite to h_0's entry in the one-to-one case, initiates a rejection chain. That is, a doctor that was previously matched to hospital h_0 is now "rejected" and is possibly matched to a different hospital, which in turn rejects another a doctor, and so forth until either a doctor remains unmatched or a previously empty position is filled. If it was the one-to-one case doctors on these rejection chains were all worse off, whereas hospitals were better off. However, in the many-to-one case things are not quite so simple as a hospital may employ several doctors (before and after the capacity reduction). The responsiveness assumption ensures that the same kind of welfare comparisons can be made on the (generalized) rejection chain.

Technically, the proof relies on constructing a directed graph whose vertices are the agents in the market such that there is an edge from a hospital to a doctor if they are matched under the hospital-optimal stable matching before the capacity reduction, and from a doctor to a hospital if they are matched under the doctor-optimal stable matching following the capacity reduction. The original proof (for the case of one-to-one matching) follows a path from h_0 until reaching an unmatched agent. This proof strategy cannot always work in a many-to-one environment, as paths can split and cycles can appear. Instead we consider properties of the component reachable from h_0. The implications for all stable mechanisms follow from the doctor-optimality and hospital-optimality of the matchings that were used to construct the directed graph. A similar logic applies to most of the proofs in the rest of the paper, under various adjustments to account for the different assumptions and desired conclusions.[13]

Proof of Theorem 23. Let $\mu = \psi^H(P)$ and $\mu' = \psi^D(P')$. The latter matching is well-defined because $\succ_{h_0}^{q=q'}$ is responsive, by Lemma 22. Throughout this proof "agent i is better off" means $\mu'(i) \succ_i \mu(i)$, and similarly for "indifferent", "weakly worse off", and so on.

Assume in contradiction that there exists no $S \subseteq D$ such that $S \neq \emptyset$, every doctor in S is worse off, and every hospital in $\{h \mid \mu'(h) \cap S \neq \emptyset\}$ is better off.

Construct a directed graph with vertices and edges defined as:

$$\mathcal{V} = D \cup H$$

$$\mathcal{E} = \big\{(d,h) \mid \mu'(d) = h\big\} \cup \{(h,d) \mid \mu(d) = h\}$$

Denote by $\mathcal{W}$ the set of vertices reachable from h_0, including h_0 itself (see for example Figure 3.1).

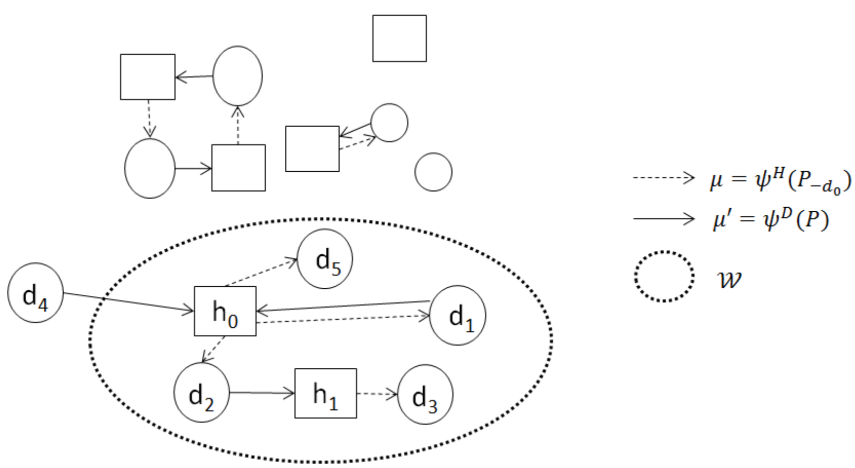


Figure 3.1: *An example of the graph used in Theorem 23*

Claim 24.1. *All doctors in D either have an outgoing edge or are indifferent.*

[13]Theorem 23 is not a direct implication of Theorem 2.26 in Roth and Sotomayor (1990) using an equivalent marriage market (see, e.g., Roth and Sotomayor, 1990, section 5.3.1). The equivalent marriage market approach will only imply that some slots of some hospitals now hold doctors that the hospital prefers, but the hospital may possibly also employ doctors that it prefers less in other slots. Similar arguments explain why the rest of the theorems in this paper are not trivially implied by Theorem 2.26 in Roth and Sotomayor (1990).

Proof. Let $d \in D$ be some doctor who is not indifferent. If d is worse off and has no outgoing edge then $S = \{d\}$ contradicts our assumption (with $\mu'(d) = \emptyset$ being the set of hospitals which are better off). Suppose d is better off. From individual rationality $\mu(d) \succeq_d \emptyset$, meaning that $\mu'(d) \neq \emptyset$ and d has an outgoing edge. ■

Claim 24.2. *All hospitals in H are weakly worse off.*

Proof. Let $h \in H$ be better off. If there exists $d \in \mu'(h)$ who is worse off, use $S = \{d\}$ to get a contradiction. On the other hand, if all doctors in $\mu'(h)$ are weakly better off we get a contradiction to the stability of μ (blocked by hospital h and doctors $\mu'(h)$). This argument applies to h_0 as well. ■

Claim 24.3. *For every hospital $h \in H$, if there is at least one doctor who is better off in $\mu'(h)$, then $|\mu(h)| = q_h$, and all doctors in $\mu(h)$ are weakly better off.*

Proof. Let $d' \in \mu'(h)$ be a doctor who is better off. If $|\mu(h)| < q_h$ then d' and h form a blocking pair for μ. Suppose $d \in \mu(h)$ is worse off. Then from the stability of μ it follows that:

$$\mu(h) \succ_h \mu(h) \cup \{d'\} \setminus \{d\},$$

and from the stability of μ' that:

$$\mu'(h) \succ_h \mu'(h) \cup \{d\} \setminus \{d'\},$$

which together contradict the responsiveness of $\succ_h$. This means that there is no doctor in $\mu(h)$ who is worse off. ■

Claim 24.4. *All doctors in $\mathcal{W} \cap D$ are weakly worse off.*

Proof. Let $D_b = \{d \in D \mid d \text{ is better off}\}$. Let $\deg_b^-(h)$ denote the number of incoming edges from doctors in D_b to hospital h, and $\deg_b^+(h)$ denote the number of outgoing edges from hospital h to doctors in D_b.

Unless $\mathcal{W} \cap D_b = \emptyset$ we can find $d' \in \operatorname{argmin}_{d \in \mathcal{W} \cap D_b} \delta(h_0, d)$, where $\delta(x, y)$ denotes the distance between nodes x and y on the graph $(\mathcal{V}, \mathcal{E})$. We denote $h' = \mu(d')$ (which is not $\emptyset$ because d' is reachable from h_0).

We claim that:

$$\deg_b^-(h') < \deg_b^+(h'). \tag{3.1}$$

To see that note first that $d' \in \mu(h')$ and so $\deg_b^+(h') \geq 1$. If $\deg_b^-(h') = 0$, then we are done. Let $n_i = |\mu(h') \cap \mu'(h')|$ denote the number of doctors who are indifferent in $\mu(h')$. If $1 \leq \deg_b^-(h') + n_i < q_{h'}$ then we can use Claim 24.3 to get Equation 3.1. If $\deg_b^-(h') + n_i = q_{h'}$ then it must be that $h' \neq h_0$ (because h_0's capacity was reduced below q_{h_0}). In this case there must be $d'' \in \mu'(h') \cap D_b$ such that $\delta(h_0, d'') < \delta(h_0, d')$, contradicting the way d' was chosen.

Putting everything together we know that:[14]

$$\begin{aligned}
&\forall h \in H : \deg_b^-(h) \leq \deg_b^+(h) && \text{(Claim 24.3)}\\
&\deg_b^-(h') < \deg_b^+(h') && \text{(Equation 3.1)}\\
&\forall d \in D_b : \deg^-(d) \leq \deg^+(d) && \text{(individual rationality)}
\end{aligned}$$

We sum over all hospitals in H to get:

$$\sum_{h \in H} \deg_b^-(h) < \sum_{h \in H} \deg_b^+(h) = \sum_{d \in D_b} \deg^-(d) \leq \sum_{d \in D_b} \deg^+(d) \leq \sum_{h \in H} \deg_b^-(h)$$

Which is a contradiction, proving that $\mathcal{W} \cap D_b$ must be empty. ■

Pick some $d \in \mu(h_0) \setminus \mu'(h_0)$, and let $h = \mu'(d)$ (exists by Claim 24.1). We get that hospital h (which is worse off by Claim 24.2) and the doctors in $\mu(h)$ (who are weakly worse off by Claim 24.4) form a blocking coalition under μ'. This concludes the contradiction argument, proving that the required S exists.

The conclusion of the theorem follows from the hospital-optimality and the doctor-

[14]For any $i \in D \cup H$, $\deg^-(i)$ denotes the indegree of i and $\deg^+(i)$ denotes the outdegree of i.

optimality of μ and μ' respectively. We showed that every $d \in S$ prefers $\mu(d)$ to $\mu'(d)$. Since the hospital-optimal stable matching is the least preferred stable matching for doctors (Knuth, 1976) we know that for any stable mechanism ψ, d weakly prefers $\psi(P)(d)$ to $\mu(d)$ and weakly prefers $\mu'(d)$ to $\psi(P')(d)$, so it must prefer $\psi(P)(d)$ to $\psi(P')(d)$. A similar argument (in the opposite direction) is true for the hospitals in $\{h \mid \psi^D(P)(h) \in S\}$. $\square$

To better understand some of the implications of Theorem 23 to manipulating capacities, it is instructive to look at the next example, in which the relevant set of hospitals contains the hospital h_0 itself.

Example Let $P \in \mathcal{P}^R$ be such that $D = \{d_1, d_2, d_3, d_4\}$, $H = \{h_0, h_1\}$, $q_{h_0} = 3$, $q_{h_1} = 2$, and the (responsive) preferences are given by:

$$\succ_{d_1} = h_1, h_0$$

$$\succ_{d_2} = h_0, h_1$$

$$\succ_{d_3} = h_0, h_1$$

$$\succ_{d_4} = h_0$$

$$\succ_{h_0} = \{d_1, d_2, d_4\}, \{d_1, d_3, d_4\}, \{d_1, d_2, d_3\}, \{d_1, d_4\}, \{d_1, d_2\}, \{d_1, d_3\},$$
$$\{d_1\}, \{d_2, d_3, d_4\}, \{d_2, d_4\}, \{d_3, d_4\}, \{d_4\}, \{d_2, d_3\}, \{d_2\}, \{d_3\}$$

$$\succ_{h_1} = \{d_2, d_3\}, \{d_1, d_2\}, \{d_2\}, \{d_1, d_3\}, \{d_3\}, \{d_1\}$$

Define $P' = \left(D, H, (1, 2), \{\succ_i\}_{i \in D \cup H \setminus \{h_0\}} \cup \{\succ_{h_0}^{q=1}\}\right)$. We have $\Psi(P) = \{\mu\}$ and $\Psi(P') = \{\mu'\}$, where μ and μ' are given by:

$$\mu = \begin{pmatrix} h_0 & h_1 \\ d_2\ d_3\ d_4 & d_1 \end{pmatrix} \qquad \mu' = \begin{pmatrix} h_0 & h_1 \\ d_1 & d_2\ d_3 \end{pmatrix}$$

Therefore h_0 should not report its true capacity in this market, because it can always obtain a better match by reporting a capacity of 1. Note that if h_0 reports a capacity of 2 and

$P'' = \Big(D, H, (2,2), \{\succ_i\}_{i \in D \cup H \setminus \{h_0\}} \cup \{\succ_{h_0}^{q=2}\}\Big)$, then we have $\Psi(P'') = \{\mu_1'', \mu_2''\}$, where:

$$\mu_1'' = \begin{pmatrix} h_0 & h_1 \\ d_1\ d_4 & d_2\ d_3 \end{pmatrix} \qquad\qquad \mu_2'' = \begin{pmatrix} h_0 & h_1 \\ d_2\ d_4 & d_1\ d_3 \end{pmatrix}$$

This means that if, for example, the hospital is uncertain about which stable matching will come about, it may want to avoid reporting a capacity of 2 in order not to lose d_3 (e.g. in case the doctor-optimal mechanism is being used). However, such hesitations should not bother the hospital when considering reporting a capacity of just one position.

Example 3.3 can also serve as an alternative and a more direct proof for (a slight modification of) Theorem 1 of Sönmez (1997). This theorem states that if there are at least two hospitals and three doctors then there exists no matching rule that is stable and non-manipulable via capacities. When using our example as the proof, the reason for non-existence of any such mechanism becomes much clearer. Instead of relying on the interaction between two hospitals' incentives under the structure imposed by the stability constraint as in the original proof, we only need to consider a rejection chain that is generated by just one hospital and the benefits it gets from inducing this chain.

Corollary 25 (Modified version of Theorem 1 of Sönmez (1997)). *Suppose there are at least two hospitals and four doctors. Then there exists no stable mechanism that is non-manipulable via capacities.*

It is certainly not the case that a hospital always wants to report a lower capacity. For example, in a market with just one hospital, reporting a lower capacity can only make the hospital weakly worse off. In this market Theorem 23 holds with the relevant set of hospitals being empty. Furthermore, it is possible that the set of hospitals mentioned in the theorem is not empty, but does not contain h_0, which means that the effect of its capacity reduction on some other hospitals is predictable, but the welfare effect on itself does not have to be. Another interesting observation is that in a setting in which doctors are in very high demand (that is, under any stable matching every doctor is employed by some hospital), it is straightforward that Theorem 23 holds with the relevant set of hospitals being non-empty.

While the earlier study of many-to-one matching focused solely on the domain of responsive preferences, in recent years the literature converged to studying a less restrictive preferences domain that still preserves the lattice structure of stable matchings, namely the domain of substitutable preferences. Most of the results on one-to-one matching were extended to this domain, with the occasional restriction of requiring preferences to satisfy the law of aggregate demand, a key property when attempting to prove extended versions of the rural hospital theorem and similar results. We now show that the result described above breaks down when the domain of preferences is extended to substitutable preferences that satisfy the law of aggregate demand. The reason for the failure is that unlike in the one-to-one case or the responsive preferences case, the matching of a hospital with a single doctor does not contain all the information about the hospital's welfare change. When hospitals offer multiple positions, a doctor taking a previously vacant position could possibly have a marginal positive direct effect on the hospital employing her, but at the same time induce a different mix of the other doctors altogether, and the latter can have a much more substantial effect on hospitals' ranking of their outcomes.

Example Let $P \in \mathcal{P}^{\mathrm{SL}}$ be such that $D = \{d_1, d_2, d_3\}$, $H = \{h_0, h_1, h_2\}$, $q_{h_0} = 1$, $q_{h_1} = 2$, $q_{h_2} = 1$, and:

$$\begin{aligned}
\succ_{d_1} &= h_0, h_1 \\
\succ_{d_2} &= h_2, h_1 \\
\succ_{d_3} &= h_1, h_2 \\
\succ_{h_0} &= \{d_1\} \\
\succ_{h_1} &= \{d_1, d_2\}, \{d_2\}, \{d_2, d_3\}, \{d_1, d_3\}, \{d_1\}, \{d_3\} \\
\succ_{h_2} &= \{d_3\}, \{d_2\}
\end{aligned}$$

Hospitals' preferences are substitutable and satisfy the law of aggregate demand.[15] Define $P' = \Big(D, H, (0,2,1), \{\succ_i\}_{i \in D \cup H \setminus \{h_0\}} \cup \{\succ_{h_0}^{q=0}\}\Big)$. One can verify that $\Psi(P) = \{\mu_1, \mu_2\}$ and

[15]However, preferences are not q-separable (Martínez *et al.*, 2000). See also Footnote 12.

$\Psi(P') = \{\mu'_1, \mu'_2\}$, where:

$$\mu_1 = \begin{pmatrix} h_0 & h_1 & h_2 \\ d_1 & d_3 & d_2 \end{pmatrix} \qquad \mu'_1 = \begin{pmatrix} & h_1 & h_2 \\ d_1 & d_3 & d_2 \end{pmatrix}$$
$$\mu_2 = \begin{pmatrix} h_0 & h_1 & h_2 \\ d_1 & d_2 & d_3 \end{pmatrix} \qquad \mu'_2 = \begin{pmatrix} & h_1 & h_2 \\ d_1 & d_2 & d_3 \end{pmatrix}$$

Thus if we define a stable mechanism ψ such that $\psi(P) = \mu_2$ and $\psi(P') = \mu'_1$ we can clearly see that an exact analog of Theorem 23 fails. This is because the only doctor who is made worse off following h_0's capacity reduction is d_1, and yet the hospital that employs d_1 under $\psi^D(P')$, namely h_1, also becomes worse off following the capacity reduction.

Inspecting Example 3.3, one may conjecture that weakening the conclusion of Theorem 23 could help in establishing a similar result even without responsiveness, and indeed this is the case.

Theorem 26. *Assume $P \in \mathcal{P}^{SL}$. Let $h_0 \in H$, $0 \leq q' < m_{h_0}(P)$, and set*

$$P' = \left(D, H, \{\succ_i\}_{i \in D \cup H \setminus \{h_0\}} \cup \{\succ_{h_0}^{q=q'}\}\right).$$

If $P' \in \mathcal{P}^{SL}$,[16] *then there exists a non-empty subset of doctors $S \subseteq D$, such that under any stable mechanism ψ, every doctor in S is worse off and every hospital in*

$$\left\{h \mid \emptyset \neq \psi^D(P')(h) \setminus \psi^H(P)(h) \subseteq S\right\}$$

is better off under $\psi(P')$ compared to $\psi(P)$.

Theorem 26 modifies the conclusion by relating a more restricted set of hospitals to the set of doctors who are made worse off following the capacity reduction. That is, the set of hospitals is now those hospitals that, except for doctors who were employed by them under the hospital-optimal stable matching prior to the reduction, employ *only* doctors from S under the doctor-optimal stable matching following the reduction. For any subset of doctors who are not indifferent between $\psi^D(P')$ and $\psi^H(P)$, this relation induces a subset

[16] This condition is required since it is not immediate (Lemma 22) and the proof uses the doctor-optimal stable matching following the manipulation, a construct that may not exist if $\succ_{h_0}^{q=q'}$ is not substitutable.

of hospitals that is weakly included in the subset mentioned in Theorem 23. That is to say:

$$\forall S' \subseteq \left\{d \mid \psi^D(P')(d) \neq \psi^H(P)(d)\right\}:$$
$$\left\{h \mid \emptyset \neq \psi^D(P')(h) \setminus \psi^H(P)(h) \subseteq S'\right\} \subseteq \left\{h \mid \psi^D(P')(h) \cap S' \neq \emptyset\right\}$$

Moreover, as Example 3.3 shows, in some cases the inclusion is strict. The restriction is necessary because when preferences are substitutable but not responsive, comparisons across individual doctors employed by a hospital become impossible, and one has to resort to comparisons between sets of doctors. When restricted to one-to-one environments the two definitions coincide and both Theorem 23 and Theorem 26 reduce to a statement equivalent to Theorem 2.26 of Roth and Sotomayor (1990).

3.4 Entry in many-to-one markets

We now turn to applying our results on capacity reduction to the study of entry in many-to-one markets. As was already briefly mentioned in the intuition for the proof of Theorem 23, a hospital's entry is the mirror image of a hospital's leaving the market, which is equivalent to reducing the hospital's capacity to zero. This relation provides us immediately some predictions that hold regardless of the stable mechanism used.

For any $P \in \mathcal{P}$ and any $d \in D$ we denote the market without d by:

$$P_{-d} = \left(D \setminus \{d\}, H, \{q_h\}_{h\in H}, \{\succ_i\}_{i\in D\setminus\{d\}} \cup \{\succ_h|_{2^{D\setminus\{d\}}}\}_{h\in H}\right),$$

where $\succ_h|_{2^{D\setminus\{d\}}} \equiv \{(S_1, S_2) \in \succ_h \mid S_1, S_2 \subseteq D \setminus \{d\}\}$. We similarly define P_{-h} for any $h \in H$.

Corollary 27. *Assume $P \in \mathcal{P}^R$. If $h_0 \in H$ and $m_{h_0}(P) > 0$ then there exists a non-empty subset of doctors $S \subseteq D$, such that under any stable mechanism ψ, every doctor in S is better off and every hospital in $\left\{h \mid \psi^D(P_{-h_0})(h) \cap S \neq \emptyset\right\}$ is worse off under $\psi(P)$ compared to $\psi(P_{-h_0})$.*

Corollary 28. *Assume $P \in \mathcal{P}^{SL}$. If $h_0 \in H$ and $m_{h_0}(P) > 0$ then there exists a non-empty subset of doctors $S \subseteq D$, such that under any stable mechanism ψ, every doctor in S is better off and every* *hospital in*

$$\left\{h \mid \emptyset \neq \psi^D(P_{-h_0})(h) \setminus \psi^H(P)(h) \subseteq S\right\}$$

is worse off under $\psi(P)$ compared to $\psi(P_{-h_0})$.

Unlike the case of a hospital's entry, we now establish that Theorem 2.26 of Roth and Sotomayor (1990) cannot be directly applied to many-to-one markets when a doctor enters the market. Recall that this theorem states that if a doctor enters the market and gets matched then there exists a non-empty set of hospitals such that every hospital in this set becomes better off following the entry under any stable mechanism. Furthermore, every doctor who was previously employed by some hospital in the mentioned set of hospitals becomes worse off following the entry. Example 3.4 describes a situation in which the one doctor that was previously in the market does not become worse off following another doctor's entry.

Example Let $P \in \mathcal{P}^R$ be such that $D = \{d_0, d_1\}$, $H = \{h_1\}$, $q_{h_1} = 2$, and the preferences are given by:

$$\succ_{d_0} = h_1$$

$$\succ_{d_1} = h_1$$

$$\succ_{h_1} = \{d_0, d_1\}, \{d_1\}, \{d_0\}$$

Note that $\Psi(P_{-d_0}) = \{\mu\}$ and $\Psi(P) = \{\mu'\}$, where μ and μ' are given by:

$$\mu = \begin{pmatrix} h_1 \\ d_1 \end{pmatrix} \qquad \mu' = \begin{pmatrix} h_1 \\ d_0\ d_1 \end{pmatrix}$$

Therefore, this example simply points to the fact that a doctor could get a vacant place without having any impact on the welfare of other doctors in the system.

While the original result cannot be fully recuperated in the many-to-one case, a partial formulation that predicts only that some hospitals will be made better off following the entry of a doctor is possible under the assumption of responsive preferences. Furthermore, if we assume all hospitals were at full capacity prior to the entry, we also get predictions on welfare effects on the entrant's side.[17]

Theorem 29. *Assume $P \in \mathcal{P}^R$. If doctor $d_0 \in D$ is such that $m_{d_0}(P) = 1$ then there exists a non-empty subset of hospitals $S \subseteq H$, such that under any stable mechanism ψ, every hospital in S is better off under $\psi(P)$ compared to $\psi(P_{-d_0})$.*

Theorem 30. *Assume $P \in \mathcal{P}^R$. If doctor $d_0 \in D$ is such that $m_{d_0}(P) = 1$ and $\forall h \in H$: $m_h(P_{-d_0}) = q_h$, then there exists a non-empty subset of hospitals $S \subseteq H$, and a non-empty subset of doctors $T \subseteq \{d \mid \psi^H(P_{-d_0})(d) \in S\}$, such that under any stable mechanism ψ, every hospital in S is better off and every doctor in T is worse off under $\psi(P)$ compared to $\psi(P_{-d_0})$.*

We conclude this section by showing that Theorem 29 and Theorem 30 fail to hold when preferences are allowed to be non-responsive, even when they are substitutable and satisfy the law of aggregate demand.

Example Let $P \in \mathcal{P}^{\mathrm{SL}}$ be such that $D = \{d_0, d_1, d_2, d_3\}$, $H = \{h_1, h_2\}$, $q_{h_1} = 2$, $q_{h_2} = 1$, and the preferences are given by:

$$\succ_{d_0} = h_1$$

$$\succ_{d_1} = h_1, h_2$$

$$\succ_{d_2} = h_2, h_1$$

$$\succ_{d_3} = h_1$$

$$\succ_{h_1} = \{d_0, d_2\}, \{d_2, d_3\}, \{d_0, d_1\}, \{d_0, d_3\}, \{d_1, d_3\}, \{d_1, d_2\},$$
$$\{d_0\}, \{d_1\}, \{d_2\}, \{d_3\}$$

$$\succ_{h_2} = \{d_1\}, \{d_2\}$$

It is easy to verify that both $\succ_{h_1}$ and $\succ_{h_2}$ are substitutable and satisfy the law of aggregate demand.[18] The stable matchings before and after d_0 enters the market are given by

[17] I thank the associate editor for suggesting this extension.

[18] In fact, $\succ_{h_1}$ is also 2-separable and $\succ_{h_2}$ is 1-separable (Martínez *et al.*, 2000).

$\Psi(P_{-d_0}) = \{\mu_1, \mu_2\}$ and $\Psi(P) = \{\mu'_1, \mu'_2\}$, where the different matchings are:

$$\mu_1 = \begin{pmatrix} h_1 & h_2 \\ d_1\ d_3 & d_2 \end{pmatrix} \qquad \mu'_1 = \begin{pmatrix} h_1 & h_2 \\ d_0\ d_1 & d_2 \end{pmatrix}$$
$$\mu_2 = \begin{pmatrix} h_1 & h_2 \\ d_2\ d_3 & d_1 \end{pmatrix} \qquad \mu'_2 = \begin{pmatrix} h_1 & h_2 \\ d_0\ d_2 & d_1 \end{pmatrix}$$

Define a stable mechanism ψ such that $\psi(P_{-d_0}) = \mu_2$ and $\psi(P) = \mu'_1$, and then both hospitals are worse off following d_0's entry.

3.5 Truncations and dropping strategies

Having dealt with the effects of capacity reduction and its potential to allow successful manipulation regardless of the stable mechanism, we now wish to study similar effects when a hospital reports a preference relation that is different than its true preference relation. For any hospital $h \in H$ with a preference relation $\succ_h$, we say that h **plays a dropping strategy** $\succ_h^{\mathbf{dr}(E)}$ for some $E \subseteq D$ if:

1. For all $D' \subseteq D$: $D' \succ_h^{\mathrm{dr}(E)} \emptyset \iff D' \succ_h \emptyset$ and $D' \cap E = \emptyset$.
2. For all $D', D'' \subseteq D$: If $D' \succ_h^{\mathrm{dr}(E)} \emptyset$ and $D'' \succ_h^{\mathrm{dr}(E)} \emptyset$, then $D' \succ_h^{\mathrm{dr}(E)} D'' \iff D' \succ_h D''$.

In other words, a hospital playing a dropping strategy submits its true preference over some subset of its acceptable doctors. Note that this definition coincides with the definition of Kojima and Pathak (2009) for the domain of responsive preferences. We say that h **plays a truncation strategy** $\succ_h^{\mathbf{tr}(d)}$ (or simply that $\succ_h^{\mathrm{tr}(d)}$ is a **truncation** of $\succ_h$) if $\succ_h^{\mathrm{tr}(d)}$ is derived from $\succ_h$ by "truncating" all doctors below and including $d \in D$. Formally, we require:

1. For all $D' \subseteq D$: $D' \succ_h^{\mathrm{tr}(d)} \emptyset \iff D' \succ_h \emptyset$ and $\forall d' \in D'$: $\{d'\} \succ_h \{d\}$.
2. For all $D', D'' \subseteq D$: If $D' \succ_h^{\mathrm{tr}(d)} \emptyset$ and $D'' \succ_h^{\mathrm{tr}(d)} \emptyset$, then $D' \succ_h^{\mathrm{tr}(d)} D'' \iff D' \succ_h D''$.

That is, $\succ_h^{\mathrm{tr}(d)}$ is the same as $\succ_h$, except it does not accept any subset of doctors containing any doctor who is weakly less preferred to d. It is easy to see that a truncation strategy is also a dropping strategy.

Observation 31. *Let $\succ_h$ be hospital h's preference relation. If $\succ_h$ is responsive, then for any $d \in D$, $\succ_h^{tr(d)}$ is also responsive.*

Our next theorem states that if hospital h_0 reports a truncated preference such that some doctor who was previously assigned to h_0 under the hospital-optimal stable matching is now unacceptable according to the truncated preference, then a similar conclusion to the one that appears in Theorem 23 holds.[19]

Theorem 32. *Assume $P \in \mathcal{P}^R$.*[20] *Let $h_0 \in H$, $\bar{d} \in D$, and $d^* \in \psi^H(P)(h_0)$ be such that $\{\bar{d}\} \succeq_{h_0} \{d^*\}$, and $P' = \left(D, H, \{\succ_i\}_{i \in D \cup H \setminus \{h_0\}} \cup \{\succ_{h_0}^{tr(\bar{d})}\}\right)$. Then there exists a non-empty subset of doctors $S \subseteq D$, such that under any stable mechanism ψ, every doctor in S is worse off and every hospital in $\left\{h \mid \psi^D(P')(h) \cap S \neq \emptyset\right\}$ is better off under $\psi(P')$ compared to $\psi(P)$.*

The intuition here resembles the one for Theorem 23, as a hospital that uses a truncation strategy practically does something which is very much like reducing its capacity. Rejecting the less attractive doctors mimics what would have happened, for example, in a deferred acceptance algorithm (with either doctors or hospitals proposing), and again causes rejection chains that may lead to more preferred doctors being available for the hospital to hire.

Theorem 32 uses truncations, which play an important role in preferences manipulation in one-to-one markets. Specifically, in these markets a player trying to act strategically to manipulate the results of the stable mechanism can restrict herself to the class of truncation strategies. However, in many-to-one matching truncations do not exhaust the space of strategies that may lead to better outcomes. Kojima and Pathak (2009, Lemma 1) show that using dropping strategies a hospital can mimic all the manipulations that are possible by reporting a reduced capacity and some preference relation over individual doctors. It turns

[19] Note that for the case of doctors truncating their preferences, truncating above the doctor-optimal stable match always results in remaining single, and so it is equivalent to leaving the market. Truncating below the doctor-optimal stable match does not give rise to any consistent welfare predictions (i.e., independently of the mechanism).

[20] Theorem 32 does not hold under the weaker assumption of q-separable and substitutable preferences. See Example 3.5.

out there is no immediate counterpart for Theorem 32 that holds for dropping strategies. To see that, consider the following counterexample:

Example Let $P \in \mathcal{P}^R$ be such that $D = \{d_1, d_2\}$, $H = \{h_0, h_1\}$, $q_{h_0} = q_{h_1} = 1$, and the preferences are given by:

$$\succ_{d_1} = h_0, h_1$$

$$\succ_{d_2} = h_0, h_1$$

$$\succ_{h_0} = \{d_1\}, \{d_2\}$$

$$\succ_{h_1} = \{d_2\}, \{d_1\}$$

Suppose that for some reason h_0 considers playing the dropping strategy:

$$\succ'_{h_0} = \{d_2\}$$

Define $P' = \Big(D, H, \{q_h\}_{h \in H}, \{\succ_{d_1}, \succ_{d_2}, \succ'_{h_0}, \succ_{h_1}\}\Big)$. It is easy to verify that $\Psi(P) = \{\mu\}$ and $\Psi(P') = \{\mu'\}$, where:

$$\mu = \begin{pmatrix} h_0 & h_1 \\ d_1 & d_2 \end{pmatrix} \qquad \mu' = \begin{pmatrix} h_0 & h_1 \\ d_2 & d_1 \end{pmatrix}$$

This implies that the non-empty set of doctors in the theorem must be $S = \{d_1\}$. However, this means that the relevant set of hospitals contains only h_1, which is worse off following the manipulation. Thus, the theorem must be revised in some way if it is to be applied to dropping strategies. Note that despite the fact that in this particular example playing the dropping strategy hurts h_0, in all other results in this paper the hospital that reduced its capacity or reports different preferences can be worse off, better off or indifferent.

Another reasonable question is whether it is possible to extend the theorem to the case of substitutable preferences that satisfy the law of aggregate demand in a manner similar to Theorem 26. The answer is again negative, as the following example demonstrates.

Example Let $P \in \mathcal{P}^{\mathrm{SL}}$ be such that $D = \{d_1, d_2, d_3, d_4\}$, $H = \{h_0, h_1\}$, $q_{h_0} = q_{h_1} = 2$, and the preferences are given by:

$$\succ_{d_1} = h_1, h_0$$

$$\succ_{d_2} = h_1, h_0$$

$$\succ_{d_3} = h_0, h_1$$

$$\succ_{d_4} = h_0, h_1$$

$$\begin{aligned}\succ_{h_0} =& \{d_1, d_2\}, \{d_1, d_3\}, \{d_1, d_4\}, \{d_2, d_3\}, \{d_2, d_4\},\\ & \{d_3, d_4\}, \{d_3\}, \{d_4\}, \{d_1\}, \{d_2\}\end{aligned}$$

$$\begin{aligned}\succ_{h_1} =& \{d_3, d_4\}, \{d_1, d_3\}, \{d_2, d_3\}, \{d_1, d_4\}, \{d_2, d_4\},\\ & \{d_1, d_2\}, \{d_3\}, \{d_4\}, \{d_1\}, \{d_2\}\end{aligned}$$

These preferences are not only substitutable and satisfy the law of aggregate demand, they are also q-separable. However, it is important to note that $\succ_{h_0}$ is not responsive. Consider the truncation $\succ_{h_0}^{\mathrm{tr}(d_1)}$, which is responsive and is therefore substitutable and satisfies the law of aggregate demand. Let $P' = \left(D, H, \{q_h\}_{h \in H}, \{\succ_i\}_{i \in D \cup H \setminus \{h_0\}} \cup \{\succ_{h_0}^{\mathrm{tr}(d_1)}\}\right)$ be the problem in which h_0 truncates, and note that $\Psi(P) = \{\mu_1, \mu_2\}$ and $\Psi(P') = \{\mu'\}$, where:

$$\mu_1 = \begin{pmatrix} h_0 & h_1 \\ d_1\ d_2 & d_3\ d_4 \end{pmatrix} \qquad \mu' = \begin{pmatrix} h_0 & h_1 \\ d_3\ d_4 & d_1\ d_2 \end{pmatrix}$$

$$\mu_2 = \begin{pmatrix} h_0 & h_1 \\ d_3\ d_4 & d_1\ d_2 \end{pmatrix}$$

This means that there exists a stable mechanism ψ such that $\psi(P) = \mu_1$ and $\psi(P') = \mu'$, and all doctors are better off following the manipulation.

3.6 Conclusion

This paper explored welfare consequences of capacity reductions, entries and truncations in many-to-one matching markets governed by stable mechanisms. We showed that there are situations in which even if participants have imperfect information about the matching process, they may profitably manipulate by reducing capacity. It should be emphasized that one key argument for reducing capacity is gaining the ability to propose positions to doctors while bypassing the centralized mechanism (e.g. market unraveling or off-shore hiring). This paper does not incorporate the advantage gained by freeing up additional positions. It does show that doing so may create a second-order effect that also works to the hospital's benefit, and therefore encourages rather than inhibits the process of market unraveling.

We assumed throughout that hospitals have strict preferences over acceptable doctors and vice-versa. While this assumption could be defended in some markets, it is quite problematic when discussing school choice, where students are frequently assigned the same priority by the schools and the indifferences are motivated by moral considerations and not by lack of sufficient information. The fact that mechanisms can break ties in different ways implies that the structure of stable matchings is a bit different, and this prevents us from applying the same techniques. Note however that in this case even comparisons that take the mechanism to be a fixed one usually hold only under a specific tie-breaking rule, so it is possible that we should be less ambitious when trying to come up with predictions that hold across all stable mechanisms as well.

Finally, throughout our analysis we used the language of the college admissions model (applied to hospitals and doctors). However, the market design literature has recently witnessed the emergence of several generalized matching models.[21] One may wonder to which extent our results continue to hold in these more sophisticated environments. First, all the results hold for the many-to-one generalized matching with contracts framework of Hatfield and Milgrom (2005).[22] This implies that the results we described are also true for matching with salaries and with other properties that may be embedded in the contracts.[23] Furthermore, while it is probably true that similar results will continue to hold in other

[21] Some of those can be found in Hatfield and Milgrom (2005), Echenique and Oviedo (2006), Hatfield *et al.* (2013) and Ostrovsky (2008).

[22] For details please contact the author.

[23] For a comparison of the generality of matching with contracts and matching with just salaries, see Echenique (forthcoming).

matching frameworks as long as we assume everybody has responsive preferences, it is also quite easy to see that in many-to-many matching environment with substitutable preferences one will run into problems trying to generalize our results, by extending Example 3.4 (in which a doctor enters the market). However, it is my belief that similar predictions are still possible in supply networks that have a pyramid structure, i.e., in which each firm can sell to multiple clients, but can have only one supplier.[24]

[24] In this suggested framework, a manipulation by one of the first-layer suppliers has predictable outcomes on a set of firms (intermediate or final users) and a related set of firms, which are the immediate suppliers of the first set.

References

Abdulkadiroğlu, A. and Sönmez, T. (2003). School choice: A mechanism design approach. *American Economic Review*, **93** (3), 729–747.

Abreu, D. and Gul, F. (2000). Bargaining and reputation. *Econometrica*, **68** (1), 85–117.

Afacan, M. O. (2011). Fictitious student creation incentives in school choice problems, mimeo.

— (2012). Application fee manipulations in matching markets. *Journal of Mathematical Economics*.

Aldous, D. J. (2001). The ζ (2) limit in the random assignment problem. *Random Structures & Algorithms*, **18** (4), 381–418.

Alkan, A. (2002). A class of multipartner matching markets with a strong lattice structure. *Economic Theory*, **19** (4), 737–746.

Ambrus, A., Burns, J. and Ishii, Y. (2013). Gradual bidding in ebay-like auctions, working paper.

— and Lu, S. E. (2010). A continuous-time model of multilateral bargaining, working paper.

Ashlagi, I., Braverman, M. and Hassidim, A. (forthcoming). Stability in large matching markets with complementarities. *Operations Research*.

—, Kanoria, Y. and Leshno, J. D. (2013). Unbalanced random matching markets. In *ACM Conference on Electronic Commerce*, pp. 27–28.

— and Klijn, F. (forthcoming). Manipulability in matching markets: conflict and coincidence of interests. *Social Choice and Welfare*.

Atakan, A. E. and Ekmekci, M. (2013). A two-sided reputation result with long-run players. *Journal of Economic Theory*, **148** (1), 376–392.

Azevedo, E. M. (2014). Imperfect competition in two-sided matching markets. *Games and Economic Behavior*, **83**, 207–223.

— and Leshno, J. D. (2011). The college admissions problem with a continuum of students. In *Proceedings of the 12th ACM conference on Electronic commerce*, ACM, pp. 333–334.

Baker, D. (2013). The fiscal cliff crisis and the real economic crisis in the united states. *Intereconomics*, **1**, 66–68.

Becker, G. (1981). *A Treatise on the Family*. Cambridge, Harvard University Press.

Bishop, D. T., Cannings, C. and Maynard Smith, J. (1978). The war of attrition with random rewards. *Journal of Theoretical Biology*, **74** (3), 377–388.

Blair, C. (1988). The lattice structure of the set of stable matchings with multiple partners. *Mathematics of operations research*, pp. 619–628.

Botticini, M. and Siow, A. (2008). Are there increasing returns in marriage markets?, university of Toronto. Mimeo.

Calcagno, R., Kamada, Y., Lovo, S. and Sugaya, T. (2014). Asynchronicity and coordination in common and opposing interest games. *Theoretical Economics*, **9** (2), 409–434.

Che, Y.-K. and Kojima, F. (2010). Asymptotic equivalence of probabilistic serial and random priority mechanisms. *Econometrica*, **78** (5), 1625–1672.

Chiappori, P. A., Salanié, B. and Weiss, Y. (2011). Partner choice and the marital college premium, working paper.

Cho, I.-K. and Kreps, D. M. (1987). Signaling games and stable equilibria. *The Quarterly Journal of Economics*, **102** (2), 179–222.

Choo, E. and Siow, A. (2006). Who marries whom and why. *Journal of Political Economy*, **114** (1), 175–201.

Cobb, C. W. and Douglas, P. H. (1928). A theory of production. *American Economic Review*, **18** (1), 139–165.

Coles, P., Gonczarowski, Y. and Shorrer, R. (2014). Strategic behavior in unbalanced matching markets, mimeo.

— and Shorrer, R. (2014). Optimal truncation in matching markets. *Games and Economic Behavior*, **87**, 591–615.

Cramton, P. and Tracy, J. S. (1992). Strikes and holdouts in wage bargaining: Theory and data. *American Economic Review*, **82** (1), 100–121.

Crawford, V. P. (1991). Comparative statics in matching markets. *Journal of Economic Theory*, **54** (2), 389 – 400.

— and Knoer, E. M. (1981). Job matching with heterogeneous firms and workers. *Econometrica*, **49** (2), 437–450.

Demange, G. and Gale, D. (1985). The strategy structure of two-sided matching markets. *Econometrica*, **53** (4), 873–888.

—, — and Sotomayor, M. (1986). Multi-item auctions. *Journal of Political Economy*, **94** (4), 863–872.

Echenique, F. (forthcoming). Contracts vs. salaries in matching. *American Economic Review*.

— and Oviedo, J. (2006). A theory of stability in many-to-many matching markets. *Theoretical Economics*, **1** (2), 233–273.

Ehlers, L. (2004). In search of advice for participants in matching markets which use the deferred-acceptance algorithm. *Games and Economic Behavior*, **48** (2), 249 – 270.

— (2010). Manipulation via capacities revisited. *Games and Economic Behavior*, **69** (2), 302 – 311.

Fanning, J. (2013). Reputational bargaining and deadlines, unpublished manuscript.

Feinberg, Y. and Skrzypacz, A. (2005). Uncertainty about uncertainty and delay in bargaining. *Econometrica*, **73** (1), 69–91.

Fershtman, C. and Seidmann, D. J. (1993). Deadline effects and inefficient delay in bargaining with endogenous commitment. *Journal of Economic Theory*, **60** (2), 306–321.

Frieze, A. and Sorkin, G. B. (2007). The probabilistic relationship between the assignment and asymmetric traveling salesman problems. *SIAM Journal on Computing*, **36** (5), 1435–1452.

Fuchs, W. and Skrzypacz, A. (2013). Bargaining with deadlines and private information. *American Economic Journal: Microeconomics*, **5** (4), 219–243.

Fudenberg, D. and Tirole, J. (1986). A theory of exit in duopoly. *Econometrica*, **54** (4), 943–960.

Gale, D. and Shapley, L. S. (1962). College admissions and the stability of marriage. *American Mathematical Monthly*, **69** (1), 9–15.

Harsanyi, J. C. and Selten, R. (1988). *A General Theory of Equilibrium Selection in Games*. Cambridge, MA: The MIT Press.

Hassidim, A. and Romm, A. (2014). An approximate "law of one price" in random assignment games, working paper.

Hatfield, J. W. and Kominers, S. D. (2013). Vacancies in supply chain networks. *Economics Letters*.

—, —, Nichifor, A., Ostrovsky, M. and Westkamp, A. (2013). Stability and competitive equilibrium in trading networks. *Journal of Political Economy*, **121** (5), 966–1005.

— and Milgrom, P. R. (2005). Matching with contracts. *American Economic Review*, **95** (4), 913–935.

Hendricks, K., Weiss, A. and Wilson, C. (1988). The war of attrition in continuous time with complete information. *International Economic Review*, **29** (4), 663–680.

Immorlica, N. and Mahdian, M. (2005). Marriage, honesty, and stability. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, Society for Industrial and Applied Mathematics, pp. 53–62.

Ishii, Y. and Kamada, Y. (2011). The effect of correlated inertia on coordination, working paper.

Kamada, Y. and Kandori, M. (2011). Revision Games, working paper.

— and Muto, N. (2011). Multi-agent search with deadline, working paper.

— and Sugaya, T. (2010). Asynchronous Revision Games with Deadline: Unique Equilibrium in Coordination Games, working paper.

Kanoria, Y., Saban, D. and Sethuraman, J. (2014). The size of the core in assignment markets, arXiv preprint arXiv:1407.2576.

Kelso, J., Alexander S. and Crawford, V. P. (1982). Job matching, coalition formation, and gross substitutes. *Econometrica*, **50** (6), 1483–1504.

Kesten, O. (2011). On two kinds of manipulation for school choice problems. *Economic Theory*, pp. 1–17, 10.1007/s00199-011-0618-6.

Klijn, F. (2011). A many-to-many rural hospital theorem, mimeo.

Knuth, D. E. (1976). *Mariages stables et leurs relations avec d'autres problèmes combinatoires*. Presses de l'Université de Montréal.

Kojima, F. (2006). Mixed strategies in games of capacity manipulation in hospital-intern markets. *Social Choice & Welfare*, **27** (1), 25 – 28.

— and Manea, M. (2010). Incentives in the probabilistic serial mechanism. *Journal of Economic Theory*, **145** (1), 106–123.

— and Pathak, P. A. (2009). Incentives and stability in large two-sided matching markets. *American Economic Review*, **99** (3), 608–627.

—, — and Roth, A. E. (2013). Matching with couples: stability and incentives in large markets. *Quarterly Journal of Economics*, **128** (4), 1585–1632.

Konishi, H. and Ünver, M. (2006). Games of capacity manipulation in hospital-intern markets. *Social Choice & Welfare*, **27** (1), 3 – 24.

Kornhauser, L., Rubinstein, A. and Wilson, C. (1989). Reputation and patience in the 'war of attrition'. *Economica*, **56** (221), 15–24.

Kreps, D. M. and Wilson, R. (1982a). Reputation and imperfect information. *Journal of Economic Theory*, **27** (2), 253–279.

— and — (1982b). Sequential equilibria. *Econometrica*, **50** (4), 863–894.

Kristof-Brown, A. and Guay, R. P. (2011). Person–environment fit. In S. Zedeck (ed.), *American Psychological Association Handbook of Industrial and Organizational Psychology*, vol. 3, American Psychological Association, pp. 1–50.

Krokhmal, P. A. and Pardalos, P. M. (2009). Random assignment problems. *European Journal of Operational Research*, **194** (1), 1–17.

Lamont, O. A. and Thaler, R. H. (2003). Anomalies: The law of one price in financial markets. *Journal of Economic Perspectives*, **17** (4), 191–202.

Lee, S. (2014). Incentive compatibility of large centralized matching markets, mimeo.

— and Yariv, L. (2014). On the efficiency of stable matchings in large markets, mimeo.

Ma, C.-T. A. and Manove, M. (1993). Bargaining with deadlines and imperfect player control. *Econometrica*, **61** (6), 1313–1339.

Ma, J. (2002). Stable matchings and the small core in nash equilibrium in the college admissions problem. *Review of Economic Design*, **7** (2), 117.

— (2010). The singleton core in the college admissions problem and its application to the national resident matching program (nrmp). *Games and Economic Behavior*, **69** (1), 150 – 164, special Issue In Honor of Robert Aumann.

Manea, M. (2009). Asymptotic ordinal inefficiency of random serial dictatorship. *Theoretical Economics*, **4** (2), 165–197.

Martínez, R., Massó, J., Neme, A. and Oviedo, J. (2000). Single agents and the set of many-to-one stable matchings. *Journal of Economic Theory*, **91** (1), 91–105.

Maynard Smith, J. (1974). The theory of games and the evolution of animal conflicts. *Journal of Theoretical Biology*, **47** (1), 209–221.

Milgrom, P. and Roberts, J. (1982). Predation, reputation, and entry deterrence. *Journal of Economic Theory*, **27** (2), 280–312.

Mo, J. (1988). Entry and structures of interest groups in assignment games. *Journal of Economic Theory*, **46** (1), 66 – 96.

Mongell, S. J. and Roth, A. E. (1986). A note on job matching with budget constraints. *Economics Letters*, **21**, 135–138.

Mumcu, A. and Saglam, I. (2009). Games of capacity allocation in many-to-one matching with an aftermarket. *Social Choice & Welfare*, **33** (3), 383 – 403.

Nalebuff, B. and Riley, J. (1985). Asymmetric equilibria in the war of attrition. *Journal of Theoretical Biology*, **113** (3), 517–527.

Núñez, M. and Rafels, C. (2008). On the dimension of the core of the assignment game. *Games and Economic Behavior*, **64** (1), 290–302.

Osborne, M. J. (1985). The role of risk aversion in a simple bargaining model. In A. E. Roth (ed.), *Game Theoretic Models of Bargaining*, Cambridge University Press, pp. 181–213.

Ostrovsky, M. (2008). Stability in supply chain networks. *American Economic Review*, **98** (3), 897–923.

Pittel, B. (1989). The average number of stable matchings. *SIAM Journal on Discrete Mathematics*, **2** (4), 530–549.

— (1992). On likely solutions of a stable marriage problem. *Annals of Applied Probability*, **2** (2), 358–401.

Ponsati, C. (1995). The deadline effect: A theoretical note. *Economics Letters*, **48** (3), 281–285.

Quint, T. (1987). *Elongation of the core in an assignment game*. Tech. rep., DTIC Document.

Riley, J. G. (1980). Strong evolutionary equilibrium and the war of attrition. *Journal of Theoretical Biology*, **82** (3), 383–400.

Romero-Medina, A. and Triossi, M. (2007). Games of capacities: A (close) look to nash equilibria, mimeo.

Roth, A. E. (1982). The economics of matching: Stability and incentives. *Mathematics of Operations Research*, **7** (4), 617–628.

— (1984a). The evolution of the labor market for medical interns and residents: A case study in game theory. *Journal of Political Economy*, **92** (6), 991–1016.

— (1984b). Stability and polarization of interests in job matching. *Econometrica: Journal of the Econometric Society*, pp. 47–57.

— (1991). A natural experiment in the organization of entry-level labor markets: Regional markets for new physicians and surgeons in the united kingdom. *American Economic Review*, **81** (3), 415–440.

— (2002). The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica*, **70** (4), 1341–1378.

—, Murnighan, J. K. and Schoumaker, F. (1988). The deadline effect in bargaining: Some experimental evidence. *American Economic Review*, **78** (4), 806–823.

— and Rothblum, U. G. (1999). Truncation strategies in matching markets-in search of advice for participants. *Econometrica*, **67** (1), 21–43.

— and Sotomayor, M. A. O. (1990). *Two-sided matching : a study in game-theoretic modeling and analysis*. Econometric Society monographs ; no. 18, New York: Cambridge University Press.

— and Vande Vate, J. H. (1990). Random paths to stability in two-sided matching. *Econometrica*, pp. 1475–1480.

— and — (1991). Incentives in two-sided matching with random stable mechanisms. *Economic Theory*, **1** (1), 31–44.

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica*, **50** (1), 97–109.

Sandholm, T. and Vulkan, N. (1999). Bargaining with deadlines. In *IAAI '99 Proceedings of the Sixteenth National Conference on Artificial Intelligence*, pp. 44–51.

Shapley, L. S. (1955). *Markets as cooperative games*. Rand Corporation.

— and Shubik, M. (1971). The assignment game i: The core. *International Journal of Game Theory*, **1** (1), 111–130.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. George Routledge and Sons, retrieved Sep. 8, 2014, via Google Books.

Sönmez, T. (1997). Manipulation via capacities in two-sided matching markets. *Journal of Economic Theory*, **77** (1), 197 – 204.

Sönmez, T. (1999). Can pre-arranged matches be avoided in two-sided matching markets? *Journal of Economic Theory*, **86** (1), 148 – 156.

Storms, E. (2013). Incentives and manipulation in large market matching with substitutes, mimeo.

Wästlund, J. (2005). The variance and higher moments in the random assignment problem, mimeo.

Wilson, L. B. (1972). An analysis of the stable marriage assignment algorithm. *BIT Numerical Mathematics*, **12** (4), 569–575.

Yildiz, M. (2004). Optimism, deadline effect, and stochastic deadlines, working Paper.

Appendix A

Appendix to Chapter 1

A.1 Proof of Theorem 1

Proof. For simplicity, the proof uses results that were proven for the uniform noise distribution $G = U[0,1]$. However, all claims hold for more general distributions. A complete proof for general distributions (albeit one that provides slightly less tight bounds and only deals with balanced markets) can be found in the working paper version of the present work (Hassidim and Romm, 2014).

The general structure of the proof is as follows.

1. Given an arbitrary vector of workers' human capital, show that whp (relevant to the distribution of $\left\{\varepsilon_{ij}^{n}\right\}$) there are only finitely many workers above a certain human capital level who are unemployed, and similarly finitely many workers below a different human capital level who are employed (Lemma 33).

2. Based on the previous step, show that a version of the result of Frieze and Sorkin (2007) holds, but with some restrictions on its applicability to workers (Lemma 34).

3. Show that in fact whp all workers above a certain human capital level are employed, and all workers below a certain human capital level are not employed (Lemma 35).

4. Improve the applicability of Lemma 34 to workers (Lemma 36).

5. Put everything together with the intuition presented in the main text to complete the proof.

For a given a vector of human capital levels h^n (of length $n + k(n)$), let us denote by $h^n[m]$ the m-th highest value.

Lemma 33. *For any $\epsilon > 0$ there exist $M \in \mathbb{N}$ such that*

1. *whp there are at most M workers with a human capital level greater than $h^n[n] + \epsilon$ who are unemployed under the optimal assignment for M^n;*
2. *whp there at most M workers with a human capital level less than $h^n[n] - \epsilon$ who are employed under the optimal assignment for M^n.*

Proof. Denote by V^n_{opt} the value resulting from the optimal assignment in M^n, and by V^n_{bound} the value resulting from optimally assigning the top n workers (in terms of human capital level) to the n available firms. From Aldous (2001) we know that

$$\lim_{n\to\infty} E[V^n_{\text{bound}}] = \sum_{i=1}^{n} q_i^n + \sum_{j=1}^{n} h^n[j] + \left(n - \frac{\pi^2}{6}\right). \tag{A.1}$$

Taking q^n and h^n as given, we know from Wästlund (2005) that

$$\text{Var}[V^n_{\text{bound}}] = \frac{4\xi(2) - 4\xi(3)}{n} + O\left(\frac{1}{n^2}\right) \approx \frac{1.7715}{n} + O\left(\frac{1}{n^2}\right). \tag{A.2}$$

By approximating the limit in (A.1), bounding the variance in (A.2), and using Markov inequality:

$$\Pr\left(V^n_{\text{bound}} \le \sum_{i=1}^{n} q_i^n + \sum_{j=1}^{n} h^n[j] + (n-2)\right) \le \frac{13.6}{n}.$$

This also implies that whp

$$V^n_{\text{opt}} \ge \sum_{i=1}^{n} q_i^n + \sum_{j=1}^{n} h^n[j] + (n-2).$$

Now assume that there are M workers with a human capital level greater than or equal to $h^n[n] + \epsilon$ who do not participate in the optimal assignment (or alternatively that there are

M workers with a human capital level less than or equal to $h^n[n] - \epsilon$ who do participate in the optimal assignment). It must be that

$$V^n_{\text{opt}} \leq \sum_{i=1}^{n} q_i^n + \sum_{j=1}^{n} h^n[j] - M\epsilon + n,$$

and therefore

$$M \leq \left\lceil \frac{2}{\epsilon} \right\rceil.$$

■

Now, given some arbitrary matchings $\{\mu^n\}$, construct digraphs $\mathcal{G}^n = (\mathcal{V}^n, \vec{E}^n)$, with $\mathcal{V}^n = F^n \cup W^n$ and

$$\begin{aligned}
\vec{E}^n = & \left\{ (w_j^n, f_i^n) \mid \mu^n(f_i^n) = w_j^n \right\} \\
& \cup \left\{ (f_i^n, w_j^n) \mid w_j^n \in N^n_{h^n[n]+\epsilon, 40+M}(f_i^n) \right\} \\
& \cup \left\{ (f_i^n, w_j^n) \mid f_i^n \in N^n_{h^n[n]+\epsilon, 40+M}(w_j^n) \right\},
\end{aligned}$$

where $N^n_{x,k}(f_i^n)$ represent the top k workers in terms of idiosyncratic fit to f_i^n) (i.e., ε_{ij}^n) out of those workers who have a human capital level above x, and similarly for $N^n_{x,k}(w_j^n)$. We call the edges from F^n to W^n "forward edges" and the edges from W^n to F^n "backward edges." The weight on each forward edge (f_i^n, w_j^n) is ε_{ij}^n (and not α_{ij}^n).

Lemma 34. *If $\underline{h} \neq \overline{h}$,*[1] *there exists $c \in \mathbb{R}_+$ such that whp there is an alternating path between every two firms with the sum of weights on the forward edges being less than or equal to $\frac{c \log n}{n}$. Similarly, there is an alternating path from any matched worker to any worker with a human capital level above $h^n[n] + \epsilon$ with the sum of weights on the forward edges being less than or equal to $\frac{c \log n}{n}$.*

Proof. First, let us choose $\epsilon > 0$ such that whp the number of workers with a human capital level above $h^n[n] + \epsilon$ is no less than $0.99n$. To see that this is possible, let us denote by ν^n

[1] This lemma also holds (with the proper adjustments) for the case where all workers have the same human capital level, but we omit the proof here since it can easily be recovered using the arguments presented in the more complicated case.

the fraction of workers who are unassigned in M^n, i.e., $\nu^n := \frac{k(n)}{n+k(n)}$, and let $\eta^n = H^{-1}(\nu^n)$. Let $\epsilon > 0$ be such that $\sup_{(x,y)\subseteq(\underline{h},\overline{h}),(y-x)<\epsilon} H(y) - H(y) < 0.0049$ (this is possible since we required the density to be continuous on $\left[\underline{h},\overline{h}\right]$, and it is therefore bounded). By Hoeffding's inequality whp $h^n[n] \in (\eta^n - \epsilon, \eta^n + \epsilon)$. Then, using Hoeffding's inequality again, we know that whp $0.99n$ of the workers have a human capital level above $\eta^n + 2\epsilon \geq h^n[n] + \epsilon$.

Note that whp there exists c_1 such that there is a directed path of length less than $c_1 \log n$ between any two firms, using the same argument as Frieze and Sorkin (2007, Lemma 5). It is true that in our case some of the workers do not have related backward edges (since they are unmatched), but out of those workers who are connected to forward edges (with a human capital level above $\overline{h} - \epsilon$) at most M do not have backward edges. Therefore, by pointing to $M + 40$ workers we keep the expansion rate of at least 40. We also note that some of the constants have to be changed to account for the fact that only a constant fraction of the workers are connected by forward edges, and that the number of workers is not necessarily n but could rather be greater than that as long as it is $O(n)$. We remark that ϵ must have been chosen such that a large majority of the firms will be matched to workers with human capital levels above $h^n[n] + \epsilon$; otherwise there would not necessarily be an overlap between the two "funnels" constructed in the proof.

We then use Lemma 7 of Frieze and Sorkin (2007) which works as is, except that the number 40 is replaced by $40 + M$ whenever it appears in the proof there. This completes the argument for the firms.

As for the workers, the same argument works, but we note that in order for a directed path to start from some worker, that worker must be matched, and in order for it to finish with some worker, that worker must have a human capital level above $h^n[n] + \epsilon$. ∎

Lemma 35. *If $\underline{h} \neq \overline{h}$, then there exist $c_1 \in \mathbb{R}_+$ such that*

1. *whp all workers with a human capital level greater than $h^n[n] + \frac{c_1 \log n}{n}$ are assigned under the optimal assignment for M^n;*

2. *whp no workers with a human capital level less than $h^n[n] - \frac{c_1 \log n}{n}$ are assigned under the optimal assignment for M^n.*

Proof. Let $c_1 \in \mathbb{R}_+$ be equal to $(c+2)$, where c is the constant recovered in the proof of Lemma 34. Assume on the contrary that there exists an unmatched worker w_1^n with human capital level $h_1^n > h^n[n] + \frac{c_1 \log n}{n}$. Let w_2^n be the worker with the lowest level of human capital in M^n that is matched. We want to argue that there exists a matching in which the set of matched workers is $\mu^n(F^n) \cup \{w_1^n\} \setminus \{w_2^n\}$ and that this matching gives a larger value. Replace the matching μ^n with the one in which $\mu^n(w_2^n)$ is matched with w_1^n. Note that this matching gives a value greater by $(h_1^n - h_2^n) \geq \frac{(c+2)\log n}{n}$ in human capital, but might provide us with less than optimal noise compatibility between $\mu^n(w_2^n)$ and w_1^n. Applying Lemma 34 to our new matching, find a directed path between w_1^n (which is now matched) and some worker who is also matched and who "likes" $\mu^n(w_2^n)$ (in the sense of having joint productivity greater than $1 - \frac{\log n}{n}$). Apply the directed path, in the sense that now each worker is going to be matched to the firm connected to her by a forward edge, and the last worker is connected to $\mu^n(w_2^n)$. The value of the resulting matching is at least $\text{val}(\mu^n) + \frac{(c+2)\log n}{n} - \frac{(c+1)\log n}{n} > \text{val}(\mu^n) + \frac{\log n}{n}$, a contradiction.

The exact same reasoning applies when a matched worker has a human capital level below $h^n[n] - \frac{c_1 \log n}{n}$, and is replaced by the best unmatched worker. ■

Lemma 36. *If $\underline{h} \neq \overline{h}$, there exist $c, c_1 \in \mathbb{R}_+$ such that whp there is an alternating path from any matched worker to any worker with a human capital level above $h^n[n] + \frac{c_1 \log n}{n}$ with the sum of weights on the forward edges being less than or equal to $\frac{c \log n}{n}$.*

Proof. Use the same logic of Lemma 34 but replace ϵ with $\frac{c_1 \log n}{n}$, which will work by virtue of Lemma 35. ■

To complete the proof, let us first consider the firms. By Lemma 34 whp for every $i, j \in \{1, \ldots, |F^n|\}$ there exists an alternating path on $\mathcal{G}^n$ (induced by μ^n, the optimal assignment for M^n). Suppose one such path is $(f_i^n, w_1^n, f_1^n, w_2^n, f_2^n, \ldots, w_k^n, f_j^n)$. Since μ^n is a core allocation, it must be that $u_i^n + v_1^n \geq \alpha_{i1}^n$, and therefore

$$u_i^n \geq \alpha_{i1}^n - v_1^n \geq q_i^n + h_1^n + (1 - \varepsilon_{i1}^n) - (\alpha_{11}^n - u_1^n) \geq u_1^n + (q_i^n - q_1^n) - \varepsilon_{i1}^n.$$

Similarly we get

$$\begin{aligned}
&u_i^n \geq u_1^n + (q_i^n - q_1^n) - \varepsilon_{i1}^n, \\
&u_1^n \geq u_2^n + (q_1^n - q_2^n) - \varepsilon_{12}^n, \\
&\ldots \\
&u_k^n \geq u_j^n + \left(q_k^n - q_j^n\right) - \varepsilon_{kj}^n.
\end{aligned}$$

Stacking all of those together we have

$$u_i^n \geq u_j^n + \left(q_i^n - q_j^n\right) - \sum \varepsilon_{xy}^n,$$

where the last sum goes over all the firms that alternate on the path, and therefore

$$u_i^n \geq u_j^n + \left(q_i^n - q_j^n\right) - \frac{c \log n}{n}.$$

Reordering terms we get

$$u_j^n - u_i^n \leq \left(q_j^n - q_i^n\right) + \frac{c \log n}{n},$$

which is exactly what we wanted.

As for the workers, we need to be slightly more careful. The same line of reasoning tells us that whp for any matched worker w_i^n and any worker w_j^n with a human capital level above $h^n[n] + \frac{c_1 \log n}{n}$ (as in Lemma 36) we have

$$v_i^n - v_j^n \leq \left(h_i^n - h_j^n\right) + \frac{c \log n}{n}.$$

However, we also want to account for matched workers with a human capital level in the interval $\left(h^n[n] - \frac{c_1 \log n}{n}, h^n[n] + \frac{c_1 \log n}{n}\right)$. Let w_i^n be some matched worker and let w_j^n be a matched worker in that interval. Since whp there are $\Theta(n)$ workers with human capital levels above $h^n[n] + \epsilon$ (for any constant ϵ), then whp one of them, say w_k^n, is a good match for $\mu^n(w_j^n)$ in the sense that their joint idiosyncratic noise is above $1 - \frac{c_2 \log n}{n}$ for some constant c_2. Consider now a path that goes from w_i^n to w_k^n (whp such a path exists) and then continues to $\mu^n(w_j^n)$ and to w_j^n, and perform the same calculation as before. $\square$

A.2 Other proofs

A.2.1 Proof of Theorem 2

Lemma 37. *Let $Z = \sum_{k=1}^{n} X_k$ where each X_k is a geometric variable with stopping probability $p_k = \frac{ck}{n^3}$. Then whp $Z > \frac{1}{16c} n^3 \log n$.*

Proof. Let n' be the largest number smaller than n such that $\sqrt{n}$ is an integer, i.e., $n' = \left(\lceil \sqrt{n} \rceil\right)^2$. Z dominates $Z' = \sum_{k=1}^{\sqrt{n'}} \sum_{l=1}^{\sqrt{n'}} X'_{kl}$, where each X'_{kl} is a geometric variable with stopping probability $p_{kl} = \frac{ck\sqrt{n'}}{n^3}$. Note that $X'_{kl} > \frac{1}{2p_{kl}}$ with probability $1 - (1 - p_{kl})^{\frac{1}{2p_{kl}}} \approx 1 - e^{-\frac{1}{2}} > 0.39$, and so using Hoeffding's inequality

$$\Pr\left(\sum_{l=1}^{\sqrt{n'}} X'_{kl} > \frac{1}{4}\sqrt{n'} \cdot \frac{1}{2p_{kl}}\right) > 1 - e^{-2(0.39-0.25)^2\sqrt{n'}} > 1 - e^{-0.03\sqrt{n'}}.$$

Therefore

$$\Pr\left(Z' > \sum_{k=1}^{\sqrt{n'}} \frac{\sqrt{n'}}{8p_{kl}}\right) \geq \left(1 - e^{-0.03\sqrt{n'}}\right) > 1 - n' e^{-0.03\sqrt{n'}}.$$

So with high probability

$$Z > \sum_{k=1}^{\sqrt{n'}} \frac{\sqrt{n'}}{8p_{kl}} = \frac{1}{8c} n^3 \sum_{k=1}^{\sqrt{n'}} \frac{1}{k} \approx \frac{1}{8c} n^3 \log \sqrt{n} = \frac{1}{16c} n^3 \log n$$

$\blacksquare$

Proof. For the sake of simplicity let us focus on the case of $G = U[0,1]$. Let us take the variant of the approximation algorithm suggested by Crawford and Knoer (1981) to solve a generalized version of the assignment game, in which firms are ordered from f_1^n to f_n^n, and at each round only the lowest-number firm that still wants to propose actually proposes.

Take the step size to be $\epsilon = \frac{1}{n^3}$. We want to bound the minimal number of steps through the entire algorithm.

We note that when it is firm f_i^n's to propose, and its previous aspiration level (i.e., the maximal utility it would get by giving some worker her current salary was $\overline{u}_i$, and if for all unmatched workers $w_j^n \in W^n$ we have $\varepsilon_{ij}^n \notin [\overline{u}_i - \epsilon, \overline{u}_i)$, then some worker's salary increases by ϵ. The conditional probability of ε_{ij}^n not being in $[\overline{u}_i - \epsilon, \overline{u}_i)$ is $1 - \frac{\epsilon}{\overline{u}_i}$. We know that in the firm-optimal core allocation at least one worker gets a salary of zero, and from Theorem 1 we learn that all workers get no more than $\frac{c \log n}{n}$. Combining this with the results of Frieze and Sorkin (2007) gives us that whp $\overline{u}_i \geq 1 - \frac{c \log n}{n}$ for some constant $c \in \mathbb{R}_+$. Therefore the conditional probability mentioned before is at least $1 - \frac{\epsilon}{1 - \frac{c \log n}{n}} > 1 - 1.01\epsilon$.[2] This implies that when there are $n - k + 1$ ($k > 1$) still unemployed workers, the probability of raising the salary of one of the employed workers by ϵ is at least

$$(1 - 1.01\epsilon)^{n-k+1} > 1 - 1.01(n - k + 1)\epsilon,$$

and the probability of employing a still unemployed worker is at most $1.01(n - k + 1)\epsilon = \frac{1.01(n-k+1)}{n^3}$. By Lemma 37, whp there are going to be at least $\frac{1}{16.16} n^3 \log n$ steps, and multiplying by ϵ we get that whp the sum of workers' salaries is at least $\frac{1}{16.16} \log n$. This implies also that whp at least one of the workers has a salary that is at least $\frac{1}{16.06} \cdot \frac{\log n}{n}$. As mentioned before, in each realization one of the workers has a salary of zero. Together this means that whp there are two workers such that the difference between their salaries is $\frac{1}{16.06} \cdot \frac{\log n}{n}$, and we are done. □

A.2.2 Proof of Corollary 3

Proof. As mentioned in the intuition for the proof, there must be at least one worker whose salary is exactly zero. If $\underline{h} \neq \overline{h}$, let $c_1 \in \mathbb{R}_+$ be such that for large enough n, $H\left(\underline{h} + \frac{c_1 \log n}{n}\right) > \frac{\log n}{n}$ (such c_1 exists since H has positive and continuous density at $\underline{h}$). It

[2] We assume that $G = U[0,1]$. When $G \neq U[0,1]$ we have to approximate the density near the upper bound, and rely on Theorem 1 to approximate the conditional probability of choosing a still unmatched worker.

follows that the probability of having at least one worker with a human capital level below $\frac{c_1 \log n}{n}$ is at least

$$1 - \left(1 - \frac{\log n}{n}\right)^n \approx 1 - e^{-\log} = 1 - \frac{1}{n}.$$

Let c_2 be the constant we arrived at in the proof of Theorem 1, if the worker who gets zero salary has a human capital level above $\frac{(c_1+c_2)\log n}{n}$; then Theorem 1 implies that any worker with a human capital level lower than $\frac{c_1 \log n}{n}$ gets a negative salary. Therefore, with high probability the worker getting a zero salary must have human capital level below $\frac{(c_1+c_2)\log n}{n}$. It follows from Theorem 1 that whp for every worker w_j^n

$$v_j^{n,F} \in \left(h_j^n - \frac{(c_1 + 2c_2)\log n}{n}, h_j^n + \frac{c_2 \log n}{n}\right).$$

By taking $c = c_1 + 2c_2$ we reach the desired conclusion. □

A.2.3 Proof of Corollary 4

Proof. We prove this corollary separately for the case of workers who have the same human capital level and for the case of workers with different human capital levels. In the first case ($h^n \equiv \underline{0}$) we recall that the same line of proof used in Lemma 34 could have shown us that in this case the approximate law of one price holds for *any* two workers (and not just two *matched* workers). The proof follows immediately from Theorem 1 by comparing any matched worker to one of the unmatched workers (whose salary is 0).

In the second case ($\underline{h} \neq \overline{h}$), note that there exists $c_1 \in \mathbb{R}_+$ such that whp all workers with a human capital level below $h^n[n] - \frac{c_1 \log n}{n}$ are unmatched (Lemma 35). Let $c_2 \in \mathbb{R}_+$ be such that whp there exists a worker w_j^n with a human capital level $h_j^n \in \left(h^n[n] - \frac{c_1 \log n}{n}, h^n[n] - \frac{(c_1+c_2)\log n}{n}\right)$. Note that there exists $c_3 \in \mathbb{R}_+$ such that whp this worker has a good match with one of the matched firms; i.e., there exists f_i^n such that $\varepsilon_{ij}^n > 1 - \frac{c_3 \log n}{n}$. It follows that the worker w_k^n employed by that firm gets no more than $\left(h_k^n - h_j^n\right) + \frac{c_3 \log n}{n} \leq \left(h_k^n - h^n[n]\right) + \frac{(c_1+c_2+c_3)\log n}{n}$. Now set $c = c_1 + c_2 + c_3 + c_4$, where c_4 is the constant provided by Theorem 1, and we get the desired result using Theorem 1 (comparing matched workers to w_k^n). □

A.2.4 Proof of Proposition 8

Lemma 38.

$$E\left[\sum_{f_i^n=\mu^n\left(w_j^n\right)} \alpha_{ij}^n\right] \le n(\log n+1)$$

Proof. We want to show that for every worker the expected value of the maximal element in the relevant column of the productivity matrix α^n equals $\log n$. To see that, note first that the minimal element is distributed according to an exponential distribution with parameter n (think of the first arrival of one of n identical arrivals). Due to the memorylessness property of exponential random variables, the difference between the first minimal element and the second minimal element is distributed like an exponential distribution with parameter $n-1$, and so on. This implies that the expected value of the largest element is

$$\frac{1}{n}+\frac{1}{n-1}+\frac{1}{n-2}+\cdots+\frac{1}{2}+1 \le \log n+1.$$

■

Lemma 39.

$$E\left[\sum_{f_i^n=\mu^n\left(w_j^n\right)} \alpha_{ij}^n\right] \ge 0.99 n \log n$$

Proof. Let μ be a matching that results from running a greedy algorithm: firm 1 picks the worker it likes best, then firm 2 picks a worker from those remaining, and so on. The expected value of μ is

$$E\left[\sum_{f_i^n=\mu\left(w_j^n\right)} \alpha_{ij}^n\right] = E\left[\max\left\{X_{1,1},\ldots,X_{1,n}\right\}\right]+E\left[\max\left\{X_{2,1},\ldots,X_{2,n-1}\right\}\right]+$$

$$\cdots+E\left[X_{n,1}\right],$$

where $\{X_{i,j}\}$ are i.i.d. Exp(1). Therefore

$$E\left[\sum_{f_i^n=\mu\left(w_j^n\right)} \alpha_{ij}^n\right] = \sum_{i=0}^{n-1} [\log(n-i)+1] \approx n \log n.$$

The result then follows from the optimality of μ^n. ■

Proof. The first claim follows from Lemma 38 and Lemma 39. For the second claim, let p^n denote the probability that for a given firm $f_i^n \in F^n$ there exists a worker $w_j^n \in W^n$ such that $\alpha_{ij}^n > 1.1 \log n$ and $\max_{k \neq j} \alpha_{ik}^n < \log n$. Then

$$p^n = n \cdot e^{-1.1 \log n} \cdot \left(1 - e^{-\log n}\right)^{n-1} = \frac{1}{n^{0.1}} \cdot \left(1 - \frac{1}{n}\right)^{n-1} \approx \frac{1}{en^{0.1}}.$$

This specifically implies that for any $\epsilon > 0$ whp there are $\Omega\left(n^{0.9-\epsilon}\right)$ firms that meet the above condition. If the same worker is the outlier in any two of these firms, then this worker must get paid at least $0.1 \log n$ under any core allocation. Since there are $\Omega\left(n^{1.8-2\epsilon}\right)$ pairs, we get that there are many workers who get paid $\Theta(\log n)$. Finally, at least one worker's salary is 0 under the firm-optimal core allocation, and so we are done. □

A.2.5 Proof of Theorem 10

Lemma 40. *In an arbitrary balanced market with productivity matrix α^n, let $(\mu^n, u^{n,F}, v^{n,F})$ be a the firm-optimal core allocation. If $f_i^n = \mu^n\left(w_j^n\right)$ then*

$$v_j^{n,F} \leq \left(\alpha_{ij}^n - \min_{f_k^n=\mu^n(w_l^n)} \alpha_{kl}^n\right).$$

Proof. Let $\underline{\alpha} := \min_{f_k^n=\mu^n(w_l^n)}$. Consider a core allocation (μ', u', v') for a modified productivity matrix $\alpha' = \alpha^n - \underline{\alpha}$. It is trivial that $\mu' = \mu^n$. Since this is a core allocation it must be that $\forall i : u_i' \geq 0$, which means that $\forall f_i^n = \mu(w_j^n) : v_j' \leq \alpha_{ij}' = \alpha_{ij} - \underline{\alpha}$. Define $u_i = u_i' + \underline{\alpha}$ and $v = v'$, and note that (μ, u, v) is a core allocation for α since all the constraints defining the core are preserved when we restore the constant. The result follows immediately from the worker-pessimality of the firm-optimal core allocation. ■

Lemma 41. *If Conjecture 9 holds, then there exists $c \in \mathbb{R}_+$ such that whp*

$$\min_{f_i^n = \mu^n(w_j^n)} \alpha_{ij}^n \geq \log n - \log\log n - \log c.$$

Proof. Let the constant used in Conjecture 9 be c_1, and let $c = c_1 + 3$. The probability that the $c_1 \log n$ highest element out of n exponential random variables will be lower than $\log n - \log\log n - \log c$ equals

$$\begin{aligned}
P = & \sum_{m=n-c_1\log n+1}^{n} \binom{n}{m} \left(1 - e^{-\log n + \log(c\log n)}\right)^m \left(e^{-\log n+\log(c\log n)}\right)^{n-m} \\
\leq & c_1 \log n \cdot \binom{n}{c_1\log n - 1} \left(1 - \frac{c\log n}{n}\right)^{n-c_1\log n+1} \left(\frac{c\log n}{n}\right)^{c_1\log n - 1} \\
\leq & c_1 \log n \left(\frac{en}{c_1\log n - 1}\right)^{c_1\log n-1} \left(1 - \frac{c\log n}{n}\right)^{n} \left(\frac{c\log n}{n}\right)^{c_1\log n-1} \\
\leq & c_1 \log n \left(\frac{e}{c}\right)^{c_1\log n-1} e^{-c\log n} = c_1 \log n \left(\frac{e}{c}\right)^{c_1\log n-1} \frac{1}{n^c} \\
\leq & \frac{c_1(c_1+3)}{e} \frac{\log n}{n^3} \leq \frac{1}{n^2},
\end{aligned}$$

where the transition in the fourth line is by Stirling's approximation, and the one in the fifth line uses $c = c_1 + 3$. Therefore the probability that after taking the $c_1 \log n$ highest element out of n exponential random variables n times the minimal value is lower than $\log n - \log\log n - \log c$ is bounded above by

$$1 - \left(1 - \frac{1}{n^2}\right)^n \approx 1 - \left(1 - \frac{1}{n} + O\left(\frac{1}{n^2}\right)\right) = \frac{1}{n} + O\left(\frac{1}{n^2}\right).$$

Conjecture 9 ensures that whp μ^n does not assign any firm to a worker who is ranked below $c_1 \log n$, and therefore whp the claim holds. ■

Proof of Theorem 10. Given Lemma 38, Lemma 40, and Lemma 41, we know that

$$E\left[\sum_j v_j^{n,F}\right] \leq n(\log n + 1) - \left(1 - \frac{c_1}{n}\right) \cdot n \cdot (\log n - c\log\log n),$$

where c_1 is such that the statement in Lemma 41 holds with probability greater than $1 - \frac{c_1}{n}$.

This implies that

$$E\left[\sum_j v_j^{n,F}\right] \le cn\log\log n + n + c_1\log n - \frac{cc_1\log\log n}{n}.$$

Finally, use Lemma 39 to complete the proof. □

A.3 Analysis of the Cobb–Douglas benchmark model

This appendix demonstrates how one can get results similar to Theorem 1 in the presence of interaction between firms' quality and workers' human capital level.

A.3.1 Sketch of proof of Lemma 5

Given any $n, d, m \in \mathbb{N}$, let

$$\mathrm{Sym}\,(n,d,m) := \{\sigma \in \mathrm{Sym}(n) \mid |\{i \mid \sigma(i) - i \ge d\}| \ge m\}\,,$$

where $\mathrm{Sym}(n)$ is the symmetric group of size n. That is, $\mathrm{Sym}(n,d,m)$ is the set of all permutations σ of the set $\{1,\ldots,n\}$ such that there are at least m elements such that the difference between their images and themselves is equal to or larger than d. For any $\sigma \in \mathrm{Sym}(n)$ we let

$$\mathrm{val}(\sigma) := \frac{\sum_{i=1}^n \sqrt{i\cdot\sigma(i)}}{n}.$$

Lemma 42. *If $\sigma \in \mathrm{Sym}(n,d,m)$, and there exist $i < j$ such that $\sigma(i) > \sigma(j)$ and $\sigma(i) - i < d$, then there exists $\sigma' \in \mathrm{Sym}(n,d,m)$ such that $\mathrm{val}(\sigma') > \mathrm{val}(\sigma)$.*

Proof. Consider $\sigma' \in \mathrm{Sym}(n,d,m)$ defined by

$$\sigma'(k) = \begin{cases} \sigma(j) & \text{if } k = i, \\ \sigma(i) & \text{if } k = j, \\ \sigma(k) & \text{otherwise.} \end{cases}$$

We get that

$$\begin{aligned}\mathrm{val}(\sigma') - \mathrm{val}(\sigma) &= \frac{1}{n}\left(\sqrt{i\sigma(j)} + \sqrt{j\sigma(i)} - \sqrt{i\sigma(i)} - \sqrt{j\sigma(j)}\right) \\ &= \frac{1}{n}\left(\sqrt{i} - \sqrt{j}\right)\left(\sqrt{\sigma(j)} - \sqrt{\sigma(i)}\right) > 0\end{aligned}$$

■

Lemma 43. *If $\sigma \in \mathrm{Sym}(n, d, m)$, and there exist $i < j$ such that $\sigma(i) > \sigma(j)$ and $\sigma(j) - j \geq d$, then there exists $\sigma' \in \mathrm{Sym}(n, d, m)$ such that $\mathrm{val}(\sigma') > \mathrm{val}(\sigma)$.*

Proof. The proof is similar to the proof of Lemma 42. The only difference is that now $\sigma' \in \mathrm{Sym}(n, d, m)$ because $\sigma'(i) - i = \sigma(j) - i > \sigma(j) - j \geq d$ and $\sigma'(j) - j = \sigma(i) - j > \sigma(j) - j \geq d$. ■

Lemma 44. *If $\sigma \in \mathrm{Sym}(n, d, m)$, $m > 0$, and $n - \sigma^{-1}(n) < d$, then there exists $\sigma' \in \mathrm{Sym}(n, d, m)$ such that $\mathrm{val}(\sigma') > \mathrm{val}(\sigma)$.*

Proof. Let n' be the largest number such that $(n' - 1) - \sigma^{-1}(n' - 1) \geq d$ (such n' exists since $m > 1$). Denote $k := n' - \sigma^{-1}(n' - 1)$. If there exists i such that $i > k$ and $\sigma(i) - i > d$, then by Lemma 43 we are done. If $\sigma(n') \neq n'$, then by a simple counting argument there exist $i < j$ such that $\sigma(i) > \sigma(j)$ and $\sigma(i) - i < d$, and then by Lemma 42 we are done. Similarly, if $\sigma(n' - k + 1) > n' - k$, we can again find $i < j$ such that $\sigma(i) > \sigma(j)$ and $\sigma(i) - i < d$, and be done by Lemma 42. Define $\sigma' \in Sym(n)$ as

$$\sigma'(i) = \begin{cases} n' - 1 & \text{if } i = n', \\ n' & \text{if } i = n' - k + 1, \\ \sigma(n' - k + 1) & \text{if } i = n' - k, \\ \sigma(i) & \text{otherwise.} \end{cases}$$

We now have:

$$\begin{aligned}
\operatorname{val}\sigma' - \operatorname{val}\sigma =& \frac{1}{n}\Bigg(\left((n'-k) + \sqrt{(n'-k+1)n'} + \sqrt{n'(n'-1)} \right) \\
& - \left(\sqrt{(n'-k)(n'-1)} + \sqrt{(n'-k+1)(n'-k)} + n' \right) \Bigg) \\
=& \frac{n'}{n}\Bigg(\left(1 - \frac{k}{n'} + \sqrt{1 - \frac{k-1}{n'}} + \sqrt{1 - \frac{1}{n'}} \right) \\
& - \left(\sqrt{1 - \frac{k}{n'}}\sqrt{1 - \frac{1}{n'}} + \sqrt{1 - \frac{k-1}{n'}}\sqrt{1 - \frac{k}{n'}} + 1 \right) \Bigg) \\
=& \frac{n'}{n}\left(-\frac{k}{n'} + \left(\sqrt{1 - \frac{1}{n'}} + \sqrt{1 - \frac{k-1}{n'}} \right) \left(1 - \sqrt{1 - \frac{k}{n'}} \right) \right) \\
=& \frac{k}{n}\left(\frac{\sqrt{1 - \frac{1}{n'}} + \sqrt{1 - \frac{k-1}{n'}}}{1 + \sqrt{1 - \frac{k}{n'}}} - 1 \right) \\
=& \frac{k}{n\left(1 + \sqrt{1 - \frac{k}{n'}}\right)} \left(\mathscr{F}(k-1, n') - \mathscr{F}(0, n) \right),
\end{aligned}$$

where $\mathscr{F}(t, n') = \sqrt{1 - \frac{t}{n'}} - \sqrt{1 - \frac{t+1}{n'}}$. Note that

$$\frac{\partial \mathscr{F}(t, n')}{\partial t} = -\frac{1}{2n'\sqrt{1 - \frac{t}{n'}}} + \frac{1}{2n'\sqrt{1 - \frac{t+1}{n'}}} = \frac{\sqrt{1 - \frac{t}{n'}} - \sqrt{1 - \frac{t+1}{n'}}}{2n'\sqrt{1 - \frac{t}{n'}}\sqrt{1 - \frac{t+1}{n'}}},$$

and therefore $\frac{\partial'}{\partial t} > 0$ for all $t \in (0, n'-1)$. This implies that $\operatorname{val}\sigma' > \operatorname{val}\sigma$ as required. ∎

Lemma 45. *For all* $\sigma \in \operatorname{Sym}(n, d, m)$, $\operatorname{val}(\sigma) \leq \frac{n+1}{2} - \frac{md^2}{8n(n+d+1)}$.

Proof. Let $\overline{\operatorname{val}}(n, d, m) := \max_{\sigma \in \operatorname{Sym}(n,d,m)} \operatorname{val}(\sigma)$. Given some $\sigma_1 \in \operatorname{Sym}(n, d, m)$ such that $\operatorname{val}(\sigma_1) = \overline{\operatorname{val}}(n, d, m)$, we can define $\sigma_2 \in \operatorname{Sym}(n+d+1, d, m)$ by

$$\sigma_2(i) = \begin{cases} \sigma_1(i) & \text{if } i \leq n, \\ i & \text{if } i > n. \end{cases}$$

Following the same logic used in the proof of Lemma 44, there exists $\sigma_3 \in \mathrm{Sym}(n+d+1, d, m)$ such that $\sigma_3(n+1) = n+d+1$, $\sigma_3(i) = i-1$ for $i \in \{n+2, \ldots, n+d+1\}$, and $\mathrm{val}(\sigma_3) > \mathrm{val}(\sigma_2)$. However, this also implies that there exists $\sigma_4 \in \mathrm{Sym}(n+d+1, d, m)$ such that σ_4 is identical to σ_3 for inputs larger than n, and is identical to a permutation $\sigma_4 \in \mathrm{Sym}(n, d, m-1)$ that achieves $\overline{\mathrm{val}}(n, d, m-1)$ for inputs smaller than or equal to n. It follows that

$$\overline{\mathrm{val}}(n,d,m) = \mathrm{val}(\sigma_1) = \frac{1}{n}\left((n+d+1)\,\mathrm{val}(\sigma_2) - \sum_{i=n+1}^{n+d+1} i\right) \leq$$
$$\frac{1}{n}\left((n+d+1)\,\mathrm{val}(\sigma_4) - \sum_{i=n+1}^{n+d+1} i\right) =$$
$$\overline{\mathrm{val}}(n,d,m-1) - \frac{1}{n}\left(\sum_{i=n+1}^{n+d+1} i - \sum_{i=n+1}^{n+d} \sqrt{i(i+1)} - \sqrt{(n+1)(n+d+1)}\right).$$

Now note that

$$\sum_{i=n+1}^{n+d+1} i - \sum_{i=n+1}^{n+d} \sqrt{i(i+1)} - \sqrt{(n+1)(n+d+1)} =$$
$$n+d+1 - \sqrt{(n+1)(n+d+1)} - \sum_{i=n+1}^{n+d}\left(\sqrt{i(i+1)} - i\right) =$$
$$\frac{(n+d+1)^2 - (n+1)(n+d+1)}{n+d+1+\sqrt{(n+1)(n+d+1)}} - \sum_{i=n+1}^{n+d} \frac{i}{\sqrt{i(i+1)}+i} \geq$$
$$\frac{(n+d+1)d}{n+d+1+\sqrt{(n+1)(n+d+1)}} - \frac{d}{2} =$$
$$\frac{d}{2}\left(\frac{(n+d+1) - \sqrt{(n+1)(n+d+1)}}{(n+d+1) + \sqrt{(n+1)(n+d+1)}}\right) =$$
$$\frac{d}{2}\left(\frac{d(n+d+1)}{\left((n+d+1)+\sqrt{(n+1)(n+d+1)}\right)^2}\right) \geq \frac{d^2}{8(n+d+1)}.$$

Therefore

$$\overline{\mathrm{val}}(n,d,m) \leq \overline{\mathrm{val}}(n,d,m-1) - \frac{d^2}{8n(n+d+1)},$$

and

$$\overline{\mathrm{val}}(n,d,m) \leq \overline{\mathrm{val}}(n,d,0) - \frac{md^2}{8n(n+d+1)}.$$

To complete the proof, note that for $m = 0$ we know by Lemma 42 that $\overline{\mathrm{val}}(n,d,0) =$

$\sum_{i=1}^{n} \frac{i}{n} = \frac{n+1}{2}$. ∎

Let μ^n be an assignment for a certain market M^n; we denote

$$\mathrm{val}(\mu^n, M^n) := \sum_{\mu^n(f_i^n)=w_j^n} 2\sqrt{q_i^n h_j^n} + \varepsilon_{ij}^n.$$

Lemma 46. *Let μ^n be an assignment for M^n such that*

$$\left|\left\{i : \left|q_i^n - h_{\mu^n(i)}^n\right| > n^{b-1}\right\}\right| \geq n^a,$$

for some $a, b \in (0,1)$; then there exists $c \in \mathbb{R}_+$ such that

$$\mathrm{val}\left(\mu^n, M^n\right) \leq 2n + 1 - cn^{a+2b-2}$$

Proof. Without loss of generality, assume that more than half of the firms in the set $\left\{i : \left|q_i^n - h_{\mu^n(i)}^n\right| > n^{b-1}\right\}$ are matched with workers whose human capital level exceeds the firms' quality. Now the maximal value is given when all the firms fit workers perfectly in terms of the idiosyncratic component (i.e., $\varepsilon_{ij}^n = 1$), and then Lemma 45 bounds the sum of the interactive components, and we get

$$\mathrm{val}\left(\mu^n, M^n\right) \leq 2\left(\frac{n+1}{2} - \frac{\left(\frac{1}{2}n^a\right)\left(n^b\right)^2}{8n(n + n^{b-1} + 1)}\right) + n \leq 2n + 1 - \frac{1}{8.01}n^{a+2b-2}.$$

∎

Lemma 47. *Let $\{\mu^n\}$ be a sequence of optimal assignments for M^n. Then there exists $c \in \mathbb{R}_+$ and $\gamma \in (0,1)$ such that whp* $\mathrm{val}\left(\mu^n, M^n\right) \geq n - cn^{\gamma}$.

Proof sketch. Use a greedy algorithm that divides the firms and workers into layers according to their quality/human capital level, where each layer contains $n^{1/3}$ firms/workers. Then perform an optimal assignment within each layer based only on the noise dimension. The result approximates the efficiency on both dimensions, and gives a lower bound on the efficient assignment. ∎

Sketch of proof of Lemma 5. We deduce from Lemma 46 and Lemma 47 that for $a+2b-2>\gamma$ it must be that $\left|\left\{i \mid \left(q_i^n - h_{\mu^n(i)}^n\right) > n^{b-1}\right\}\right| < n^a$. Now assume to the contrary that there is a firm that is matched under the optimal assignment to a worker who has a human capital level far higher than the firm's quality (by "far higher" we mean $n^{\delta-1}$ for some $\delta \in (0,1)$), and show (using a somewhat involved counting argument) that there must be another firm that is matched to a worker with a human capital level far lower than the firm's own quality, and such that a switch between the workers employed by those two firms would yield an efficiency gain of $c_1 n^{2-2\delta}$ on the quality dimension for some $c_1 \in \mathbb{R}_+$. Then for each new match, try to find an alternating path (in the spirit of Theorem 1) to fix the efficiency on the noise dimension. This leads to an overall improvement in efficiency, which leads to a contradiction. □

A.3.2 Sketch of proof of Theorem 6

Sketch of proof. The proof follows immediately from Lemma 5 and similar arguments to those used in Theorem 1, applied within a band of qualities of width $\Theta\left(n^{-b}\right)$. □

Appendix B

Appendix to Chapter 2

B.1 Formal Definition of the Extensive-Form Game

Revision games: The set of players is $N = \{1, 2\}$ and the sets of types are $\mathcal{T}_i = \{\tau_i^r, \tau_i^c\}$. The set of possible histories is

$$\mathcal{H} = \{\emptyset\} \cup \left\{ -t, (\tau_1, \tau_2, \mathcal{O}_1, \mathcal{O}_2, x) \;\middle|\; \begin{array}{l} \tau_i \in \mathcal{T}_i, \mathcal{O}_i \subseteq [-T, 0], -T \in \mathcal{O}_i, \\ x\colon [-T, -t) \to \{U, D\} \times \{L, R\} \end{array} \right\}.$$

Here $\mathcal{O}_i$ represents Player i's revision opportunities, and $x(-t')$ represents the prepared profile at time $-t'$. For almost all realizations $\mathcal{O}_1$ and $\mathcal{O}_2$ will be finite. The players who take action at h are given by the function $P\colon \mathcal{H} \to 2^{N \cup \{\text{Nature}\}}$:

$$P(h) = \begin{cases} \{\text{Nature}\} & \text{if } h = \emptyset \\ \{i \mid -t \in \mathcal{O}_i\} & \text{if } h = (-t, \tau_1, \tau_2, \mathcal{O}_1, \mathcal{O}_2 x) \end{cases}.$$

The information sets partition for Player i is given by $\mathcal{I}_i$ whose typical element is

$$\mathcal{I}_i\left(-\tilde{t}, \tilde{\tau}_i, \tilde{\mathcal{O}}_i, \tilde{x}\right) =$$
$$\left\{ (-t, \tau_1, \tau_2, \mathcal{O}_1, \mathcal{O}_2, x) \;\middle|\; -t = -\tilde{t}, \tau_i = \tilde{\tau}_i, \mathcal{O}_i|_{[-T,-t]} = \tilde{\mathcal{O}}_i, x|_{[-T,-t)} = \tilde{x} \right\}.$$

The set of information sets in which Player i takes an action is

$$\mathcal{J}_i = \left\{ \mathcal{I}_i \left(-\tilde{t}, \tilde{\tau}_i, \tilde{\mathcal{O}}_i, \tilde{x} \right) \in \mathcal{I}_i \mid -\tilde{t} \in \tilde{\mathcal{O}}_i \right\}.$$

At history $\emptyset$ Nature chooses τ_1, τ_2, $\mathcal{O}_1$, and $\mathcal{O}_2$ independently according to the probabilities ξ_1 and ξ_2 (for τ_1^c and τ_2^c respectively) and according to the distributions of the Poisson processes with frequencies λ_1 and λ_2 (to determine $\mathcal{O}_1$ and $\mathcal{O}_2$ respectively). The game then moves immediately to the history $(\tau_1, \tau_2, \mathcal{O}_1, \mathcal{O}_2, -T, x^{\emptyset})$, with the "empty" function $x^{\emptyset} \colon \emptyset \to \{U, D\} \times \{L, R\}$. Following that, time moves continuously and whenever a player is called to play the available actions are

$$A_i \left(\mathcal{I}_i \left(-\tilde{t}, \tilde{\tau}_i, \tilde{\mathcal{O}}_i, \tilde{x} \right) \right) = \begin{cases} \{U, D\} & \text{if } i = 1, \tilde{\tau}_i = \tau_i^r \\ \{U\} & \text{if } i = 1, \tilde{\tau}_i = \tau_i^c \\ \{L, R\} & \text{if } i = 2, \tilde{\tau}_i = \tau_i^r \\ \{R\} & \text{if } i = 2, \tilde{\tau}_i = \tau_i^c \end{cases}$$

We denote also $A_1 = \{U, D\}$ and $A_2 = \{L, R\}$.

A feasible strategy for Player i is $\sigma_i \colon \mathcal{J}_i \to \Delta A_i$ such that

1. $\operatorname{supp} \left(\sigma_i \left(\mathcal{I}_i \left(-\tilde{t}, \tilde{\tau}_i, \tilde{\mathcal{O}}_i, \tilde{x} \right) \right) \right) \subseteq A_i \left(\mathcal{I}_i \left(-\tilde{t}, \tilde{\tau}_i, \tilde{\mathcal{O}}_i, \tilde{x} \right) \right)$

2. $\forall a_i \in A_i : \sigma_i \left(\mathcal{I}_i \left(-T, \tau_i^r, \{-T\}, x^{\emptyset} \right) \right) [a_i] \in \{0, 1\}$

As mentioned we restrict players' strategies to be measurable with respect to the natural topologies.

The state variable x is determined by the realizations of players' strategies. For any time $-t' < -t$, let $-t_i'' = \max\{-\tau \in \mathcal{O}_i \mid -\tau < -t\}$, and let α_i be the realized action of Player i at $-t_i''$, then $x(-t') = (\alpha_1, \alpha_2)$.

The terminal histories are $\mathcal{Z} = \{(-t, \tau_1, \tau_2, \mathcal{O}_1, \mathcal{O}_2, x) \in \mathcal{H} \mid -t = 0\}$, and the payoffs at history $(0, \tau_1, \tau_2, \mathcal{O}_1, \mathcal{O}_2, x)$ are given by $u\left(x(0)\right)$, where u is a function from $\{U, D\} \times \{L, R\}$ to $\mathbb{R}^2$.

Wars of attrition with Poisson arrivals: In the case of the war of attrition with Poisson arrivals the definition of the game is slightly simpler.

The set of possible histories is now given by

$$\mathcal{H} = \{\emptyset, D, L\} \cup \{(-t, \tau_1, \tau_2, \mathcal{O}_1, \mathcal{O}_2) \mid \tau_i \in \mathcal{T}_i, \mathcal{O}_i \subseteq [-T, 0], -T \in \mathcal{O}_i\}$$

The information sets partition for Player i is given by $\mathcal{I}_i$ whose typical element is

$$\mathcal{I}_i\left(-\tilde{t}, \tilde{\tau}_i, \tilde{\mathcal{O}}_i\right) = \left\{(-t, \tau_1, \tau_2, \mathcal{O}_1, \mathcal{O}_2) \;\middle|\; -t = -\tilde{t}, \tau_i = \tilde{\tau}_i, \mathcal{O}_i|_{[-T,-t]} = \tilde{\mathcal{O}}_i\right\}.$$

The rest of the details are identical to those given for revision games, except that once Player 1 chooses action D or Player 2 chooses action L the game moves to the relevant history (D or L). The terminal histories are $\mathcal{Z} = \{D, L\} \cup \{(-t, \tau_1, \tau_2, \mathcal{O}_1, \mathcal{O}_2) \in \mathcal{H} \mid -t = 0\}$, and payoffs are given by

$$u^{\text{woa}}(h) = \begin{cases} u(D, R) & \text{if } h = D \\ u(U, L) & \text{if } h = L \\ u(U, R) & \text{otherwise} \end{cases}$$

with u defined as before.

B.2 Model with Heterogeneous Revision Rates

The main text ignored the possibility of the two players having different revision rates. Here we formalize all the necessary definitions to deal with heterogeneous revision rates. All the proofs below are provided for the more general case of different frequencies.[1] The game is now summarized by the parameters $(T; u_1, u_2; \xi_1, \xi_2; \lambda_1, \lambda_2)$, where λ_i is the frequency of the Poisson process governing Player i's revisions. We denote by $\phi(T; u_1, u_2; \xi_1, \xi_2; \lambda_1, \lambda_2)$ the set of interim SE payoffs of the profile (τ_1^r, τ_2^r), and define the revision equilibrium payoff

[1]With the exception of Theorem 17.

set of $(u_1, u_2; \xi_1, \xi_2; \lambda_1, \lambda_2)$ by

$$\bar{\phi}(u_1, u_2; \xi_1, \xi_2, \lambda_1, \lambda_2) = \lim_{T' \to \infty} \phi(T'; u_1, u_2; \xi_1, \xi_2; \lambda_1, \lambda_2).$$

We extend Definition 2 trivially.

Lastly, the definition of strength becomes slightly more complicated.

Definition 4. *Player i's strength is given by*

$$s_i(u_i; \lambda_1, \lambda_2) \equiv \frac{\lambda_{3-i} \cdot |u_i(U,L) - u_i(D,R)|}{\lambda_2 \left[u_i(U,L) - u_i(U,R)\right] + \lambda_1 \left[u_i(D,R) - u_i(U,R)\right]}.$$

Player i is stronger than Player j if

$$s_i(u_i; \lambda_1, \lambda_2) > s_j(u_j; \lambda_1, \lambda_2).$$

Player i relative strength (with regard to Player j) is

$$\Delta_{ij}(u_1, u_2; \lambda_1, \lambda_2) = s_i(u_i; \lambda_1, \lambda_2) - s_j(u_j; \lambda_1, \lambda_2).$$

Note that a player is stronger if her revision frequency is low or if her opponent's revision frequency is high. The relative frequency of play allows a player to commit and thus makes her a stronger competitor (a player whose frequency of play is much lower than her opponent is practically a Stackelberg leader).

B.3 Proofs

All proofs are for the general model (with heterogeneous revision rates). Whenever the statement of the result changes because of that, the new formulation is explicitly brought. Whenever the statement is the same as in the main text, it is omitted.

B.3.1 Proof of Proposition 12

Proposition 48. *Assume Player* 1 *is stronger than Player* 2*,* $\xi_1 = 0$*, and* $\xi_2 > 0$*; then the revision equilibrium payoff set is bounded away from Player* 1*'s preferred outcome:*

$$u(U,L) \notin \bar{\phi}(u_1,u_2;0,\xi_2;\lambda_1,\lambda_2).$$

Proof. Let $\{T^k\}_{k=1}^{\infty}$ be a sequence of horizons such that $T^k \to \infty$ and let $\{\hat{\sigma}^k\}_{k=1}^{\infty}$ be a corresponding sequence of equilibria. We let P^k denote the probability measure induced by Nature's moves (i.e., the lottery over types and the stochastic Poisson processes governing players' revision opportunities) and by equilibrium strategies $\hat{\sigma}^k$.

Choose $K^A > 0$, $-t^A < 0$ and $\delta^A > 0$ such that

1. $\delta^A < \frac{1}{3}\left(u_2(D,R)\right) - u_2(U,L)),$
2. $e^{-\frac{1}{2}\lambda_1 t^A}\left(u_2(D,R) - u_2(U,R)\right) < \delta^A,$
3. for any $k > K^A$, given that it is common knowledge that Player 2 is rational by time $-t^A$, the expected continuation payoffs induced by equilibrium strategies $\hat{\sigma}^k$ are below $u_2(U,L) + \delta^A$ for Player 2.

The reason we can find K^A, $-t^A$, and δ^A that meet these conditions is due to Lemma 49 below, the proof of which is a straightforward extension of the proof of Theorem 3 of Calcagno *et al.* (2014) and it is omitted.

Lemma 49. *Assume Player* 1 *is stronger than Player* 2*, and* $\xi_2 = 0$*. Then the revision equilibrium payoff set contains only Player* 1*'s preferred outcome:*

$$\bar{\phi}(u_1,u_2;\xi_1,0;\lambda_1,\lambda_2) = \{u(U,L)\}.$$

Assume to the contrary that $u(U,L) \in \bar{\phi}(u_1,u_2;0,\xi_2;\lambda_1,\lambda_2)$. This implies that

$$\lim_{k\to\infty} P^k\left(E_L(-t^A) \mid E_{2r}\right) = 1,$$

where

$$E_{2r} = \{\text{Player 2 is rational}\}$$
$$E_L(-t) = \{L \text{ is prepared before time } -t\}$$

(or else the equilibrium result will be bounded away from the intended limit). Then for any $\delta' > 0$ there exists $K' > K^A$ such that for any $k > K'$ Player 2's expected utility under $\hat{\sigma}^k$ can be bounded above by

$$(1-\delta') \cdot \left(u_2(U,L) + \delta^A\right) + \delta' \cdot u_2(D,R) \leq u_2(U,L) + \delta^A + \delta' u_2(D,R). \tag{B.1}$$

We can take $\delta' = \frac{\delta^A}{u_2(D,R)}$ and get that there exists $K' > 0$ such that for any $k > K'$ Player 2's expected utility is bounded above by $u_2(U,L) + 2\delta^A$.

The rational Player 2 can deviate to the strategy in which she prepares R until time $-\frac{t^A}{2}$, and then best-responds to the prepared profile. Note that

$$\lim_{k\to\infty} P^k\left(E_{2r}^c \mid \left[E_L(-t^A)\right]^c\right) = \lim_{k\to\infty} \frac{\xi_2}{\xi_2 + (1-\xi_2)\cdot P^k\left(\left[E_L(-t^A)\right]^c \mid E_{2r}\right)} = 1.$$

And therefore for large enough k Player 1's strategy from $-t^A$ onward will be to prepare D conditional on Player 2 never preparing L. This will ensure the rational Player 2 an expected utility bounded below by

$$\left(1 - e^{-\frac{1}{2}\lambda_1 t^A}\right) u_2(D,R) + e^{-\frac{1}{2}\lambda_1 t^A} u_2(U,R) > u_2(D,R) - \delta^A. \tag{B.2}$$

By the definition of δ^A, (B.1), and (B.2), it follows that Player 2 has a profitable deviation and we get a contradiction. □

B.3.2 Proof of Theorem 13

Proof. Let $\bar{\bar{\xi}}_2$ be implicitly defined by

$$\frac{\bar{\bar{\xi}}_2}{1-\bar{\bar{\xi}}_2} = \Delta_{12}(u_1,u_2;\lambda_1,\lambda_2) \cdot \frac{\lambda_1 u_1(D,R) + \lambda_2 u_1(U,L) - (\lambda_1+\lambda_2) u_1(U,R)}{(\lambda_1+\lambda_2) u_1(D,R)}.$$

The reasons for using this definition will become clear in the proof.

Given some $\xi_2 \in \left(0, \bar{\bar{\xi}}_2\right)$, let $\{T^k\}_{k=1}^{\infty}$ be a sequence of horizons such that $T^k \to \infty$ and let $\{\hat{\sigma}^k\}_{k=1}^{\infty}$ be a corresponding sequence of equilibria. We let P^k denote the probability measure induced by Nature's moves (i.e., the lottery over types and the stochastic Poisson processes governing players' revision opportunities) and by equilibrium strategies $\hat{\sigma}^k$.

Note that the rational type of Player 2 must myopically best-respond after time $-t_2^*$, where $-t_2^*$ is implicitly defined by

$$u_2(U,L) = \left(1 - e^{-(\lambda_1+\lambda_2)t_2^*}\right) \cdot \frac{\lambda_1 u_2(D,R) + \lambda_2 u_2(U,L)}{\lambda_1 + \lambda_2} + e^{-(\lambda_1+\lambda_2)t_2^*} u_2(U,R),$$

or more simply

$$e^{-(\lambda_1+\lambda_2)t_2^*} = s_2(u_2; \lambda_1, \lambda_2).$$

Let $-t^B$ be such that $-t^B \leq \min\{-t^A, -t^*2\}$, with $-t^A$ being defined as in Proposition 12. Using the notation defined in the proof of Proposition 12 we know that

$$\lim_{k\to\infty} P^k \left(E_L(-t^B) \mid E_{2r}\right) < 1.^2$$

Assume to the contrary that there is no last-minute strategic interaction. This implies that the probability of the profile of prepared actions changing in the interval $(-t^B, 0)$ goes to zero as k goes to infinity. Let $E_{(a_1,a_2)}$ denote the event that (a_1, a_2) is the prepared action profile at time $-t^B$.

Claim 49.1. $\lim_{k\to\infty} P^k(E_{(U,R)} \mid E_{2r}) = \lim_{k\to\infty} P^k(E_{(D,L)} \mid E_{2r}) = 0.$

Proof. If the action profile prepared at time $-t^B$ is either (U, R) or (D, L) there is a constant probability that Player 2 will be called to play after time $-t_2^*$, and will best-respond to the prepared action of Player 1, which contradicts the lack of last-minute strategic interaction. ■

Claim 49.2. $\lim_{k\to\infty} P^k(E_{(U,L)} \mid E_{2r}) > 0.$

[2] For convenience, and without loss of generality, we will assume throughout the proof that all the sequences of probabilities we mention converge.

Proof. If $\lim_{k\to\infty} P^k(E_{(U,L)} \mid E_{2r}) = 0$ then for any $\delta > 0$, Player 1's utility for large enough k is bounded above by $u_1(D,R) + \delta$. However, Player 1 can deviate to the strategy of playing only U until time $-t_2^*$ and then best-responding, so that she will get a payoff bounded below by

$$(1-\xi_2)\cdot\left(\left(1-e^{-(\lambda_1+\lambda_2)t_2^*}\right)\cdot\frac{\lambda_1 u_1(D,R)+\lambda_2 u_1(U,L)}{\lambda_1+\lambda_2}+\right.$$
$$\left.e^{-(\lambda_1+\lambda_2)t_2^*}u_1(U,R)\right)+\xi_2\cdot\left(\left(1-e^{-\lambda_1 t_2^*}\right)u_1(D,R)+e^{-\lambda_1 t_2^*}u_1(U,R)\right)>$$
$$\frac{1-\xi_2}{\lambda_1+\lambda_2}\cdot\Big(\lambda_1 u_1(D,R)+\lambda_2 u_1(U,L)-s_2(u_2;\lambda_1,\lambda_2)\cdot$$
$$(\lambda_1 u_1(D,R)+\lambda_2 u_1(U,L)-(\lambda_1+\lambda_2)u_1(U,R))\Big)=$$
$$\frac{1-\xi_2}{\lambda_1+\lambda_2}\cdot\Big(\lambda_1 u_1(D,R)+\lambda_2 u_1(U,L)+(\Delta_{12}(u_1,u_2;\lambda_1,\lambda_2)-s_1(u_1;\lambda_1,\lambda_2))\cdot$$
$$(\lambda_1 u_1(D,R)+\lambda_2 u_1(U,L)-(\lambda_1+\lambda_2)u_1(U,R))\Big)=(1-\xi_2)u_1(D,R)+$$
$$(1-\xi_2)\Delta_{12}(u_1,u_2;\lambda_1,\lambda_2)\cdot\frac{\lambda_1 u_1(D,R)+\lambda_2 u_1(U,L)-(\lambda_1+\lambda_2)u_1(U,R)}{\lambda_1+\lambda_2}>$$
$$(1-\xi_2)u_1(D,R)+\xi_2 u_1(D,R)=u_1(D,R),$$

where the last inequality comes from ξ_2 being strictly less than $\bar{\bar{\xi}}_2$ and the definition of $\bar{\bar{\xi}}_2$. The fact that we have a lower bound that is strictly above $u_1(D,R)$ implies that we can select δ small enough such that the deviation will be profitable for Player 1 for large enough k.

■

Claim 49.3. $\lim_{k\to\infty} P^k\left(E_{(D,R)} \mid \left[E_L(-t^B)\right]^c \cap E_{2r}\right) < 1$.

Proof. Suppose $\lim_{k\to\infty} P^k\left(E_{(D,R)} \mid \left[E_L(-t^B)\right]^c \cap E_{2r}\right) = 1$. We know from Claim 49.2 that for every $\delta' > 0$ there exists K' such that for every $k > K'$ Player 2's payoffs from equilibrium $\hat{\sigma}^k$ are bounded from above by

$$u_2(D,R)-\left(\lim_{k\to\infty}P^k\left(E_{(U,L)}\mid E_{2r}\right)-\delta'\right)\cdot\left(u_2(D,R)-u_2(U,L)\right).$$

Player 2 can deviate and always prepare R and for any $\delta'' > 0$ there exists K'' such that for any $k > K''$ Player 2's payoffs from this deviation will be bounded from below by

$$u_2(D,R) - \delta'' \cdot u_2(U,R). \tag{B.3}$$

If we pick δ' and δ'' to be small enough, and take $K = \max\{K', K''\}$, Player 2 has a profitable deviation for all $k > K$. ■

From Claim 49.3 and Proposition 12 we know that

$$\lim_{k\to\infty} P^k\left(E_{(U,R)} \mid E_{2r}\right) = \lim_{k\to\infty} P^k\left(E_{(U,R)} \mid \left[E_L(-t^B)\right]^c \cap E_{2r}\right) \cdot$$
$$\lim_{k\to\infty} P^k\left(\left[E_L(-t^B)\right]^c \mid E_{2r}\right) > 0,$$

which contradicts Claim 49.1. This concludes the contradiction argument and proves that the parameters induce last-minute strategic interaction. The proof is completed using Lemma 50 below. □

Lemma 50. *If parameter vector $(u_1, u_2; \xi_1, \xi_2; \lambda_1, \lambda_2)$ induces last-minute strategic interaction, then it induces inefficiency.*

Proof. Given the vector of parameters, we know that there exists $-t' < 0$ and $\delta > 0$ such that for every sequence $\{T^k\}_{k=1}^{\infty}$ such that $T^k \to \infty$, and every corresponding sequence of SEs, the probability that the prepared profile changes between time $-t'$ and time 0 is bounded above δ as k approaches infinity. Note that two of the four possible action profiles are inefficient and there is no way to change from one efficient profile to the other without passing through an inefficient profile. Therefore, there is a probability of at least δ of reaching an inefficient profile at some time between $-t'$ and 0. It is therefore possible to bound the inefficiency from below by δ times the probability that no player moves from the time an inefficient profile was reached until 0, that is, $\delta' = \delta \cdot e^{-(\lambda_1+\lambda_2)t'}$. □

B.3.3 Proof of Proposition 14

Proposition 51. *Assume Player* 1 *is stronger than Player* 2*, and* $\xi_1 = 0$*. Then Player* 1*'s preferred outcome is in the limit of the revision equilibrium payoff set as* $\xi_2 \to 0$*:*

$$u(U,L) \in \liminf_{\xi_2 \to 0} \bar{\phi}(u_1, u_2; 0, \xi_2; \lambda_1, \lambda_2).$$ [3]

Proof. This proposition is a specific case of Theorem 17 (part 1), whose proof can be found below. □

B.3.4 Proof of Theorem 15

Proof. We will first show that substantial delay is induced, and this in turn will imply last-minute strategic interaction and inefficiency (via Lemma 50).

Let $\{T^k\}_{k=1}^{\infty}$ be a sequence of horizons such that $T^k \to \infty$ and let $\{\hat{\sigma}^k\}_{k=1}^{\infty}$ be a corresponding sequence of equilibria. We let P^k denote the probability measure induced by Nature's moves (i.e., the lottery over types and the stochastic Poisson processes governing players' revision opportunities) and by equilibrium strategies $\hat{\sigma}^k$.

Choose $K^A > 0$, $-t^A < 0$ and $\delta^A > 0$ such that

1. $\delta^A < \frac{1}{3}(1-\xi_1)\left(u_2(D,R)) - u_2(U,L)\right)$,
2. $e^{-\frac{1}{2}\lambda_1 t^A}\left(u_2(D,R) - u_2(U,R)\right) < \delta^A$,
3. $e^{-\frac{1}{2}\lambda_2 t^A}\left(u_2(U,L) - u_2(U,R)\right) < \delta^A$,
4. for any $k > K^A$, given that it is common knowledge that Player 2 is rational by time $-t^A$, the expected continuation payoffs induced by equilibrium strategies $\hat{\sigma}^k$ are above $u_1(U,L) - \delta^A$ for Player 1 and below $u_2(U,L) + \delta^A$ for Player 2.

The reason we can find K^A, $-t^A$, and δ^A that meet these conditions is due to Lemma 49.

[3] See Footnote 20.

We describe several events using the following notation:

$$E_{1r} = \{\text{Player 1 is rational}\}$$
$$E_{2r} = \{\text{Player 2 is rational}\}$$
$$E_r = E_{1r} \cap E_{2r}$$
$$E_D(-t) = \{D \text{ is prepared before time } -t\}$$
$$E_L(-t) = \{L \text{ is prepared before time } -t\}$$

Assume to the contrary that no substantial delay is induced. This implies that the probability that by time $-t^A$ the only profile that was prepared was (U,R) conditional on both players being rational converges to zero as $k \to \infty$. Formally:

$$\lim_{k\to\infty} P^k \left(\left[E_D(-t^A)\right]^c \cap \left[E_L(-t^A)\right]^c \mid E_r \right) = 0.^4 \tag{B.4}$$

Claim 51.1. $\lim_{k\to\infty} P^k \left(\left[E_L(-t^A)\right]^c \mid E_r \right) > 0.$

Proof. Suppose that $\lim_{k\to\infty} P^k \left(\left[E_L(-t^A)\right]^c \mid E_r \right) = 0$; then for any $\delta' > 0$ there exists $K' > K^A$ such that for any $k > K'$ Player 2's expected utility under $\hat{\sigma}^k$ can be bounded above by

$$\begin{aligned} &\xi_1 \cdot u_2(U,L) + (1-\xi_1) \cdot \left((1-\delta') \cdot \left(u_2(U,L) + \delta^A \right) + \delta' \cdot u_2(D,R) \right) \\ &\leq u_2(U,L) + \delta^A + \delta' u_2(D,R). \end{aligned} \tag{B.5}$$

We can take $\delta' = \frac{\delta^A}{u_2(D,R)}$ and get that there exists $K' > 0$ such that for any $k > K'$ Player 2's expected utility is bounded above by $u_2(U,L) + 2\delta^A$.

The rational Player 2 can deviate to the strategy in which she prepares R until time $-\frac{t^A}{2}$,

[4] For convenience, and without loss of generality, we will assume throughout the proof that all the sequences of probabilities we mention converge.

and then best-responds to the prepared profile. Note that

$$\lim_{k\to\infty} P^k \left(E_{2r}^c \mid \left[E_L(-t^A)\right]^c \cap E_{1r}\right) =$$
$$\lim_{k\to\infty} \frac{\xi_2}{\xi_2 + (1-\xi_2)\cdot P^k\left(\left[E_L(-t^A)\right]^c \mid E_r\right)} = 1.$$

And therefore for large enough k the rational Player 1's strategy from $-t^A$ onward will be to prepare D conditional on Player 2 never preparing L. This will ensure the rational Player 2 an expected utility bounded below by

$$\begin{aligned}
&\xi_1 \cdot \left(\left(1-e^{-\frac{1}{2}\lambda_2 t^A}\right) u_2(U,L) + e^{-\frac{1}{2}\lambda_2 t^A} u_2(U,R)\right) + \\
&(1-\xi_1)\cdot\left(\left(1-e^{-\frac{1}{2}\lambda_1 t^A}\right) u_2(D,R) + e^{-\frac{1}{2}\lambda_1 t^A} u_2(U,R)\right) > \\
&\xi_1 \cdot u_2(U,L) + (1-\xi_1)\cdot u_2(D,R) - \delta^A.
\end{aligned} \tag{B.6}$$

From the definition of δ^A, (B.5), and (B.6), it follows that Player 2 has a profitable deviation, and we get a contradiction. ■

Following Claim 51.1, let $\bar{p} \equiv \lim_{k\to\infty} P^k\left(\left[E_L(-t^A)\right]^c \mid E_r\right)$. Choose $K^B > K^A$, $-t^B \le -t^A$ and $\delta^B > 0$ such that

1. for any $k > K^B$, given that it is common knowledge by time $-t \le -t^B$ that Player 1 is rational, and that Player 1 believes that Player 2 is a commitment type with probability at least ξ_2, the expected continuation payoffs induced by equilibrium strategies $\hat{\sigma}^k$ are below $u_1(U,L) - \delta^B$ for Player 1,
2. $e^{-\frac{1}{2}\lambda_1 t^B}\left(u_1(U,L) - u_1(U,R)\right) < \frac{1}{2}(1-\xi_2)\bar{p}\delta^B$.

We can select such K^B, t^B, and δ^B as shown by Theorem 13 (note that the method of proof there implied that payoffs can be bounded for all times before $-t^B$ simultaneously).

Claim 51.2. $\lim_{k\to\infty} P^k\left(E_D(-t^B) \mid \left[E_L(-t^B)\right]^c \cap E_r\right) < 1.$

Proof. Note that Claim 51.1 and (B.4) imply that

$$\lim_{k\to\infty} P^k\left(E_D(-t^B) \cap \left[E_L(-t^B)\right]^c \mid E_r\right) > 0.$$

Assume to the contrary that $\lim_{k\to\infty} P^k\left(E_D(-t^B) \mid \left[E_L(-t^B)\right]^c \cap E_r\right) = 1$. Then

$$\lim_{k\to\infty} P^k\left(\left[E_L(-t^B)\right]^c \mid E_r\right) = \lim_{k\to\infty} P^k\left(E_D(-t^B) \cap \left[E_L(-t^B)\right]^c \mid E_r\right) > 0.$$

This in turn implies that for any $\delta' > 0$ there exists $K' > K^B$ such that the rational type of Player 1's expected utility under $\hat{\sigma}^k$ can be bounded above by

$$\xi_2 \cdot u_1(D,R) + (1-\xi_2)\cdot\Bigg(\left(1 - \lim_{k\to\infty} P^k\left(\left[E_L(-t^B)\right]^c \mid E_r\right) + \delta'\right) u_1(U,L) +$$
$$\left(\lim_{k\to\infty} P^k\left(\left[E_L(-t^B)\right]^c \mid E_r\right) - \delta'\right)\cdot\left(u_1(U,L) - \delta^B\right)\Bigg).$$

And if we take $\delta' < \frac{1}{2}\lim_{k\to\infty} P^k\left(\left[E_L(-t^B)\right]^c \mid E_r\right)$ and remember that

$$\lim_{k\to\infty} P^k\left(\left[E_L(-t^B)\right]^c \mid E_r\right) \geq \bar{p},$$

we get an upper bound of

$$\xi_2 \cdot u_1(D,R) + (1-\xi_2)\cdot u_1(U,L) - \frac{1}{2}(1-\xi_2)\bar{p}\delta^B. \tag{B.7}$$

Player 1 can deviate to the strategy in which she prepares U until time $-\frac{t^B}{2}$, and then best-responds to the prepared profile. Note that

$$\lim_{k\to\infty} P^k\left(E_{1r}^c \mid \left[E_D(-t^B)\right]^c \cap \left[E_L(-t^B)\right]^c \cap E_{2r}\right) =$$
$$\lim_{k\to\infty} \frac{\xi_2}{\xi_2 + (1-\xi_2)\cdot P^k\left(\left[E_D(-t^B)\right]^c \mid \left[E_L(-t^B)\right]^c \cap E_r\right)} = 1.$$

This deviation will ensure the rational Player 1 an expected payoff bounded below by

$$\begin{aligned}
&\xi_2 \cdot \left(\left(1 - e^{-\frac{1}{2}\lambda_1 t^B}\right) u_1(D,R) + e^{-\frac{1}{2}\lambda_2 t^B} u_1(U,R)\right) + \\
&(1-\xi_2)\cdot\left(\left(1 - e^{-\frac{1}{2}\lambda_2 t^B}\right) u_1(U,L) + e^{-\frac{1}{2}\lambda_2 t^B} u_1(U,R)\right) > \\
&\xi_2 \cdot u_1(D,R) + (1-\xi_2)\cdot u_1(U,L) - \frac{1}{2}(1-\xi_2)\bar{p}\delta^B.
\end{aligned} \tag{B.8}$$

From (B.7) and (B.8) it follows that Player 1 has a profitable deviation, and we get a contradiction. ■

We have

$$\lim_{k\to\infty} P^k\left(\left[E_L(-t^B)\right]^c \mid E_r\right) \geq \bar{p} > 0 \qquad \text{(Claim 51.1, } -t^B < -t^A\text{)}$$

$$\lim_{k\to\infty} P^k\left(\left[E_D(-t^B)\right]^c \mid \left[E_L(-t^B)\right]^c \cap E_r\right) > 0 \qquad \text{(Claim 51.2)}$$

which together imply that

$$\lim_{k\to\infty} P^k\left(\left[E_D(-t^B)\right]^c \cap \left[E_L(-t^B)\right]^c \mid E_r\right) > 0,$$

and therefore substantial delay is induced. □

B.3.5 Proof of Proposition 16

Proof. In an equilibrium of the prescribed form with cutoff times $-t_1^*$ and $-t_2^*$, where $-t^*2 \leq -t_1^*$, the following three equations must hold.

Indifference of Player 1: (B.9)

$$u_1(D,R) = (1-q(t_2^*-t_1^*))\cdot\left[\left(1-e^{-(\lambda_1+\lambda_2)t_1^*}\right)\cdot\frac{\lambda_2 u_1(U,L)+\lambda_1 u_1(D,R)}{\lambda_1+\lambda_2}+ e^{-(\lambda_1+\lambda_2)t_1^*}\cdot u_1(U,R)\right] + q(t_2^*-t_1^*)\cdot\left[\left(1-e^{-\lambda_1 t_1^*}\right)u_1(D,R)+e^{-\lambda_1 t_1^*}u_1(U,R)\right].$$

Indifference of Player 2: (B.10)

$$u_2(U,L) = \left(1-e^{-\lambda_2(t_2^*-t_1^*)}\right)u_2(U,L)+e^{-\lambda_2(t_2^*-t_1^*)}\times \left[(1-\xi_1)\cdot\left[\left(1-e^{-(\lambda_1+\lambda_2)t_1^*}\right)\cdot\frac{\lambda_2 u_2(U,L)+\lambda_1 u_2(D,R)}{\lambda_1+\lambda_2}+ e^{-(\lambda_1+\lambda_2)t_1^*}u_2(U,R)\right]+\xi_1\cdot\left[\left(1-e^{-\lambda_2 t_1^*}\right)u_2(U,L)+e^{-\lambda_2 t_1^*}u_2(U,R)\right]\right].$$

Bayesian updating: (B.11)

$$q(\bar{t}) = \frac{\xi_2}{\xi_2+(1-\xi_2)e^{-\lambda_2\bar{t}}}.$$

Note first that (B.10) can be reduced to

$$u_2(U,L) = (1-\xi_1)\cdot\Big[\Big(1-e^{-(\lambda_1+\lambda_2)t_1^*}\Big)\cdot\frac{\lambda_2 u_2(U,L)+\lambda_1 u_2(D,R)}{\lambda_1+\lambda_2}+ \\ e^{-(\lambda_1+\lambda_2)t_1^*}u_2(U,R)\Big]+\xi_1\cdot\Big[\Big(1-e^{-\lambda_2 t_1^*}\Big)u_2(U,L)+e^{-\lambda_2 t_1^*}u_2(U,R)\Big] \tag{B.12}$$

In equilibrium, $-t_1^*$ is nailed down by (B.12), and so any change to Player 1's payoffs does not affect $-t_1^*$, and changes only $-t_2^*$. Assume that strategies defined by cutoffs $(-t_1^*, -t_2^*)$ form an equilibrium, and examine (B.9). Note that from the structure imposed on the payoffs it is always true that $u_1(D,R)$ is greater than the second part of the RHS of (B.9) (the part multiplied by $q(t_2^*-t_1^*)$). The first part of the RHS of (B.9) must therefore be greater than $u_1(D,R)$. If $u_1(U,L)$ becomes larger it makes the first part of the RHS even greater, and since t_1^* does not change, this must mean that $q(t_2^*-t_1^*)$ goes up, which in turns means that t_2^* becomes smaller. The probability of reaching (U,R) is simply $e^{-(\lambda_1 t_1^*+\lambda_2 t_2^*)}$, and so if t_1^* remains the same and t_2^* becomes larger, this probability becomes smaller. A similar argument shows that raising $u_1(U,R)$ also decreases this probability. When $u_1(D,R)$ increases, the LHS of (B.9) rises more than the RHS, and then by a similar argument the probability of reaching (U,R) increases.

Dealing with changes in Player 2's payoffs is only slightly more involved. Raising $u_2(D,R)$ makes the first part of the RHS of (B.12) larger, and so it must be that $e^{-(\lambda_1+\lambda_2)t_1^*}$ becomes larger as well, and this means that t_1^* becomes smaller. Looking now at (B.9), we see that the decrease in t_1^* has two direct meanings if t_2^* is kept constant: $q(t_2^*-t_1^*)$ increases, and $e^{-(\lambda_1+\lambda_2)t_1^*}$ and $e^{-\lambda_1 t_1^*}$ both increase. This means that each part of the RHS of (B.9) becomes smaller, and the weight on the second part becomes larger. Both effects lead to the RHS becoming smaller, and to offset it, $q(t_2^*-t_1^*)$ must decrease, which implies that t_2^* decreases. So both t_1^* and t_2^* decrease, so that the probability of reaching (U,R) increases. Similar arguments show that this probability decreases with $u_2(U,L)$ and increases with $u_2(U,R)$. □

B.3.6 Proof of Theorem 17

Theorem 52. *1. Assume Player 1 is stronger than Player 2, and both players play cutoff strategies. Then in the limit as $\xi_1 \to 0$ and $\xi_2 \to 0$, the probability of reaching ex-post inefficiency tends to zero.*

2. Assume players are equally strong ($\Delta_{12}(u_1, u_2; \lambda_1, \lambda_2) = 0$), $\lambda_1 = \lambda_2 = 1$,[5] and both players play cutoff strategies. Let the sequence $(\xi_1^k, \xi_2^k)_{k=1}^{\infty}$ be such that $\lim_{k\to\infty} \xi_1^k = \lim_{k\to\infty} \xi_2^k = 0$, and $\lim_{k\to\infty} \frac{\xi_2^k}{\xi_1^k} < 1$. Then in the limit as $k \to \infty$, the probability of reaching ex-post inefficiency tends to

$$s_1(u_1; \lambda_1, \lambda_2) \times \lim_{k\to\infty} \frac{\xi_2^k}{\xi_1^k} \quad \left[= s_2(u_2; \lambda_1, \lambda_2) \times \lim_{n\to\infty} \frac{\xi_2^k}{\xi_1^k} \right].$$

Proof. Part 1: For any given ξ_1 and ξ_2, consider a SE that is defined by a pair of strategies for the rational players. Player 1 prepares U until time $-t_1^*(\xi_1, \xi_2)$, and best-responds (according to the component game's payoffs) from there on. Similarly, the rational Player 2 prepares R until time $-t_2^*(\xi_1, \xi_2)$, and best-responds (according to the component game's payoffs) from there on. In order for these two strategies to form a SE it is sufficient that three conditions are satisfied. The first is that the rational Player 2 "gives up" before Player 1 does, that is $-t_2^*(\xi_1, \xi_2) \leq -t_1^*(\xi_1, \xi_2)$.

The second condition is that Player 1 is indifferent between preparing U and D at $-t_1^*(\xi_1, \xi_2)$, conditional on the current prepared action of Player 2 being R and on her

[5] This theorem *requires* players' revision rates to be equal. This is because solving the equations involves a polynomial of degree $\frac{\lambda_1}{\lambda_2}$. Proofs for the cases where $\frac{\lambda_1}{\lambda_2} \in \{\frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 2, 3, 4\}$ can also be found. Furthermore, for the specific case of a symmetric payoff matrix, it is possible to prove this claim for arbitrary revision rates. It is a reasonable conjecture that the theorem holds even for non-symmetric payoffs and arbitrary revision rates.

playing according to the prescribed strategy at any subsequent time. This can be written as

$$
\begin{aligned}
&u_1(D,R) = (1 - q\,(t_2^*(\xi_1,\xi_2) - t_1^*(\xi_1,\xi_2))) \times \\
&\left[\left(1 - e^{-(\lambda_1+\lambda_2)t_1^*(\xi_1,\xi_2)}\right) \cdot \frac{\lambda_2 u_1(U,L) + \lambda_1 u_1(D,R)}{\lambda_1+\lambda_2} + \right. \\
&\left. e^{-(\lambda_1+\lambda_2)t_1^*(\xi_1,\xi_2)} \cdot u_1(U,R) \right] + q\,(t_2^*(\xi_1,\xi_2) - t_1^*(\xi_1,\xi_2)) \times \\
&\left[\left(1 - e^{-\lambda_1 t_1^*(\xi_1,\xi_2)}\right) \cdot u_1(D,R) + e^{-\lambda_1 t_1^*(\xi_1,\xi_2)} \cdot u_1(U,R) \right],
\end{aligned} \tag{B.13}
$$

where $q(\bar{t})$ is the posterior probability that Player 1 assigns to the event that Player 2 is the commitment type conditional on Player 2 not preparing L on an interval of length $\bar{t}$ and her playing a strategy that dictates preparing L on this interval, which is given by

$$
q(\bar{t}) = \frac{\xi_2}{\xi_2 + (1-\xi_2)e^{-\lambda_2 \bar{t}}}. \tag{B.14}
$$

Note that as $q(\cdot)$ is weakly increasing in $\bar{t}$, Player 1 strictly prefers preparing U before $-t_1^*(\xi_1,\xi_2)$, and strictly prefers preparing D after $-t_1^*(\xi_1,\xi_2)$.

The third condition is that the rational type of Player 2 is indifferent at $-t_2^*(\xi_1,\xi_2)$ conditional on Player 1's prepared action being U and her playing according to the prescribed strategy at any subsequent time:

$$
\begin{aligned}
&u_2(U,L) = \left(1 - e^{-\lambda_2(t_2^*(\xi_1,\xi_2) - t_1^*(\xi_1,\xi_2))}\right) \cdot u_2(U,L) + \\
&e^{-\lambda_2(t_2^*(\xi_1,\xi_2) - t_1^*(\xi_1,\xi_2))} \times \left[(1-\xi_1) \cdot \left[\left(1 - e^{-(\lambda_1+\lambda_2)t_1^*(\xi_1,\xi_2)}\right) \cdot \right.\right. \\
&\left. \frac{\lambda_2 u_2(U,L) + \lambda_1 u_2(D,R)}{\lambda_1+\lambda_2} + e^{-(\lambda_1+\lambda_2)t_1^*(\xi_1,\xi_2)} \cdot u_2(U,R) \right] + \\
&\left. \xi_1 \cdot \left[\left(1 - e^{-\lambda_2 t_1^*(\xi_1,\xi_2)}\right) u_2(U,L) + e^{-\lambda_2 t_1^*(\xi_1,\xi_2)} u_2(U,R) \right] \right].
\end{aligned} \tag{B.15}
$$

This means that the rational type of Player 2 weakly prefers preparing R prior to $-t_2^*(\xi_1,\xi_2)$, and weakly prefers preparing L after $-t_2^*(\xi_1,\xi_2)$ (she strictly prefers preparing L after $-t_1^*(\xi_1,\xi_2)$). It remains to show that these three conditions can be met simultaneously (as

$T \to \infty$). To see that, note first that (B.15) reduces to

$$u_2(U,L) = (1-\xi_1)\cdot\Big[\Big(1-e^{-(\lambda_1+\lambda_2)t_1^*(\xi_1,\xi_2)}\Big)\cdot \\ \frac{\lambda_2 u_2(U,L)+\lambda_1 u_2(D,R)}{\lambda_1+\lambda_2} + e^{-(\lambda_1+\lambda_2)t_1^*(\xi_1,\xi_2)}\cdot u_2(U,R)\Big]+ \\ \xi_1\cdot\Big[\Big(1-e^{-\lambda_2 t_1^*(\xi_1,\xi_2)}\Big)u_2(U,L)+e^{-\lambda_2 t_1^*(\xi_1,\xi_2)}u_2(U,R)\Big], \tag{B.16}$$

which implies that $t_1^*(\xi_1,\xi_2)$ only depends on ξ_1. Furthermore, from continuity:

$$\lim_{\xi_1\to 0} e^{-(\lambda_1+\lambda_2)t_1^*(\xi_1,\xi_2)} = \frac{\lambda_1\,(u_2(D,R)-u_2(U,L))}{\lambda_1 u_2(D,R)+\lambda_2 u_2(U,L)-(\lambda_1+\lambda_2)u_2(U,R)}.$$

Therefore $t_1^*(\xi_1,\xi_2)$ approaches a constant as $\xi_1 \to 0$, and in this equilibrium the rational Player 2 is indifferent between insisting on playing R or not along the interval $[-T,-t_1^*(\xi_1,\xi_2)]$. Knowing that $t_1^*(\xi_1,\xi_2)$ does not change with ξ_2 and converges to a constant, and looking at (B.13), we can immediately deduce that $q\,(t_2^*(\xi_2)-t_1^*)$ also converges to a constant. It follows from (B.14) that $t_2^*(\xi_1,\xi_2)$ tends to infinity as $\xi_1,\xi_2\to 0$. This means that for small enough ξ_2 the first condition is also satisfied, and ensures that the strategies we described form an equilibrium for small enough ξ_1 and ξ_2.

As mentioned above, when $\xi_1,\xi_2\to 0$, $t_1^*(\xi_1,\xi_2)$ approaches some constant, and $t_2^*(\xi_1,\xi_2)$ tends to infinity. This means that the limit of the expected payoffs (for the rational types) is $u(U,L)$, and the probability of reaching the Pareto inferior outcome approaches zero.

Part 2: Assume that $-t_2^*(\xi_1^k,\xi_2^k) \le -t_1^*(\xi_1^k,\xi_2^k)$.[6] We need to solve a set of four equations, namely, Equations (B.9), (B.12), (B.11), and $\Delta_{12}(u_1,u_2)=0$. Inputting all of these into a

[6]This is not without loss of generality, because the two players have different payoffs and different probabilities of being the commitment type. Nevertheless, we will show that this is the right assumption if $\lim_{k\to\infty}\frac{\xi_2^k}{\xi_1^k}<1$.

standard mathematical solver and then simplifying gives

$$\begin{aligned}
e^{-t_1^*(\xi_1^k,\xi_2^k)} =& -\xi_1^k s_2(u_2)\left(\frac{u_2(U,L)-u_2(U,R)}{u_2(D,R)-u_2(U,L)}\right)+ \\
&\sqrt{(1-\xi_1^k)^2 s_2(u_2)+(\xi_1^k)^2 s_2^2(u_2)\left(\frac{u_2(U,L)-u_2(U,R)}{u_2(D,R)-u_2(U,L)}\right)^2} \\
e^{-t_2^*(\xi_1^n,\xi_2^n)} =& \frac{\xi_2^k}{\xi_1^k}\times\Bigg[-\xi_1^k s_1(u_1)\left(\frac{u_1(D,R)-u_1(U,R)}{u_1(U,L)-u_1(D,R)}\right)+ \\
&\sqrt{(1-\xi_1^k)^2 s_1(u_1)+(\xi_1^k)^2 s_1^2(u_1)\left(\frac{u_1(D,R)-u_1(U,R)}{u_1(U,L)-u_1(D,R)}\right)^2}\Bigg].
\end{aligned}$$

Taking the limit as $k \to \infty$ gives us

$$\begin{aligned}
\lim_{k\to\infty} e^{-t_1^*(\xi_1^k,\xi_2^k)} =& \sqrt{s_2(u_2)} \\
\lim_{k\to\infty} e^{-t_2^*(\xi_1^k,\xi_2^k)} =& \frac{\xi_2}{\xi_1}\times\sqrt{s_1(u_1)}.
\end{aligned}$$

Given cutoff strategies the probability that both players never revise their strategies is $e^{-t_1^*(\xi_1^k,\xi_2^k)} \times e^{-t_2^*(\xi_1^k,\xi_2^k)}$, and this is also the probability of reaching an ex-post inefficient outcome. Because $s_1(u_1) = s_2(u_2)$ we get in the limit exactly

$$s_1(u_1;\lambda_1,\lambda_2)\times\lim_{k\to\infty}\frac{\xi_2^k}{\xi_1^k}\quad\left[= s_2(u_2;\lambda_1,\lambda_2)\times\lim_{k\to\infty}\frac{\xi_2^k}{\xi_1^k}\right]$$

B.3.7 Proof of Lemma 18

Proof. Let us denote

$$\sigma_i(-t) = E_j\left[\sigma_i\left(\mathcal{I}_i\left(-t,\tau_i^r,\mathcal{O}_i'\right)\right) \mid -t\in\mathcal{O}_i'\right],$$

where the expectation is taken with respect to Player j's beliefs on the possible realizations of revision opportunities for Player i up until time $-t$ (that is, it is a simple expectation derived from the definition of the Poisson process), such that $-t$ itself is a revision opportunity as well.

We wish to show that any pair of strategies played in equilibrium satisfies some kind of restricted “pairwise-monotonicity,” or formally

$$(\sigma_i(-t) > 0) \wedge (-t < -t') \wedge \left(\int_{-t}^{-t'} \sigma_j(\tau)\,\mathrm{d}\tau > 0\right) \Longrightarrow \sigma_i(-t') = 1.$$

Note that if $\sigma_i(-t) > 0$, then Player i's continuation payoff at $-t$ if not exiting is less than or equal to her payoff if exiting (the continuation payoff cannot rely on previous revision opportunities because they do not affect Player j's behavior). The continuation payoff at time $-t$ is a convex combination of (1) the expected payoff in case Player i is called to make a decision on the interval $(-t, -t']$ and exits (in which case she gets the same payoff), (2) the expected payoff in case Player j is called to make a decision on the interval $(-t, -t']$ and exits, and (3) the continuation payoff at t'. Since the probability of Player j exiting is at least $e^{-\lambda_i(t-t')}\left(1 - e^{-\lambda_j(t-t')}\right) \cdot \int_{-t}^{-t'} \sigma_j(\tau)\,\mathrm{d}\tau > 0$, the expected continuation payoff at time $-t'$ must be strictly lower than the payoff from exiting. This means that Player i exits at $-t'$, i.e., $\sigma_i(-t') = 1$.[7]

Finally, define

$$-t^* = \limsup \left\{ -t \;\middle|\; \min_i \int_{-T}^{-t} \sigma_i(\tau)\,\mathrm{d}\tau = 0 \right\}.$$

It is immediate from the definition that at least one of the players exits with zero probability before $-t^*$.

Assume (without loss of generality) that $\int_{-T}^{-t^*} \sigma_2(\tau)\,\mathrm{d}\tau \geq \int_{-T}^{-t^*} \sigma_1(\tau)\,\mathrm{d}\tau$. To see that for any $-t' > -t^*$ we have $\sigma_1(-t') = \sigma_2(-t') = 1$, consider first how the pairwise-monotonicity property works for Player 2. Let $\epsilon > 0$ be such that $\epsilon < \int_{-t^*}^{-t'} \sigma_1(\tau)\,\mathrm{d}\tau$, and let $-t \in [-T, -t^* + \epsilon)$ be such that $\sigma_2(-t) > 0$ (exists from definition of $-t^*$). This implies that $\sigma_2(-t') = 1$. This is true for arbitrary $-t' > -t^*$, so $\sigma_2(-t) = 1$ for all $-t \in (-t^*, 0]$. Now let $-t \in (-t^*, -t^* + \epsilon)$ be such $\sigma_1(-t) > 0$, and since $\int_{-t}^{-t'} \sigma_2(\tau)\,\mathrm{d}\tau > 0$ the pairwise-monotonicity property again implies $\sigma_1(-t') = 1$. □

[7] We here use the sequential rationality property of the equilibrium; the last step would not have worked for a Nash equilibrium concept.

B.3.8 Proof of Corollary 19

Proof. We wish to modify the strategies to get a new pair of strategies that are essentially the same as the old pair, and still constitute an equilibrium. Similar to the proof of Lemma 18, let

$$\sigma_i'\left(\mathcal{I}_i\left(-t, \tau_i^r, \mathcal{O}_i\right)\right) = E_j\left[\sigma_i\left(\mathcal{I}_i\left(-t, \tau_i^r, \mathcal{O}_i'\right)\right) \mid -t \in \mathcal{O}_i'\right].$$

Note that the fact that Player i could have played any continuation strategy at time $-t$ without Player j knowing about it implies that she must be indifferent between playing σ_i and σ_i'. Furthermore, σ_j is a best-response to σ_i', because nothing was changed with respect to Player j's beliefs or expected continuation payoffs. For the rest of the proof we abbreviate and write $\sigma_i'(-t)$ instead of $\sigma_i'\left(\mathcal{I}_i\left(-t, \tau_i^r, \mathcal{O}_i\right)\right)$.

Define

$$-t_i^* \equiv \inf\left\{-t \in [-T, 0] \;\middle|\; \forall -t' \in (-t, 0] : \int_{-t}^{-t'} \sigma_i'(\tau)\,\mathrm{d}\tau > 0\right\}.$$

That is, $-t_i^*$ is the earliest point from which Player i has a strictly positive probability of exiting (it could be that $-t_i^* = 0$). We know from Lemma 18 that there are only two possible equilibrium structures:

1. $-t_2^* \leq -t_1^*$, and both players exit starting from $-t_2^*$.

2. $-t_2^* > -t_1^*$, and both players exit starting from $-t_1^*$.

Without loss of generality, consider the first case, and define $-\hat{t}_2^*$ as

$$\hat{t}_2^* = t_1^* + \int_{-t_2^*}^{-t_1^*} \sigma_2'(\tau)\,\mathrm{d}\tau.$$

We claim that the strategies in which the rational type of Player 1 exits from $-t_1^*$, and the rational type of Player 2 from $-\hat{t}_2^*$, constitute a SE. To see that, note that Player 1's incentives are essentially the same, but they might have "shifted," because now the relation between the exact time and the probability is different. However, for every posterior probability of Player 2 being the rational type, the same probability that Player 2 is going to exit until time $-t_1^*$ remains the same, and so the fact that the previous pair of strategies was a SE implies that the new strategy is a best-response as well. □

B.3.9 Proof of Corollary 20

Proof. The proof follows the same lines as the proof of Theorem 13, and we omit some of the technical arguments and definitions.

Let $\{T^k\}_{k=1}^{\infty}$ be a sequence of horizons such that $T^k \to \infty$ and let $\{\hat{\sigma}^k\}_{k=1}^{\infty}$ be a corresponding sequence of equilibria. We let P^k denote the probability measure induced by Nature's moves (i.e., the lottery over types and the stochastic Poisson processes governing players' exit opportunities) and by equilibrium strategies $\hat{\sigma}^k$.

Following Corollary 19 we know that there is a sequence of equilibria in cutoff strategies, denoted by $\{\bar{\sigma}^k\}_{k=1}^{\infty}$, such that the expected payoffs from $\bar{\sigma}^k$ are the same as the expected payoffs from $\hat{\sigma}^k$. Let $\bar{\sigma}^k$ be defined by the cutoffs $\left(-\bar{t}_1^k, -\bar{t}_2^k\right)$.

If $\lim_{k\to\infty} \bar{t}_2^k = \infty$,[8] then Player 2's payoffs must approach $u_2(D, R)$. If they are not, then Player 2 can deviate to the strategy of never exiting, thus convincing Player 1 that Player 2 is the commitment type, and getting a payoff that approaches $u_2(D, R)$. This implies that Player 1's payoffs must approach $u_1(D, R)$, but then, as in the proof of Theorem 13, Player 1 can deviate to the strategy of exiting only after $-\bar{t}_2^k$ and get a payoff that is strictly greater than $u_1(D, R)$ (for small enough selection of $\bar{\bar{\xi}}_2$). We reach a contradiction, implying that $\lim_{k\to\infty} \bar{t}_2^k < \infty$. Because of the same deviation, it cannot be that $\lim_{k\to\infty} \bar{t}_1^k = \infty$, and so we get that the sequence $\{\hat{\sigma}^k\}_{k=1}^{\infty}$ exhibits substantial delay.

Moving back to the sequence $\{\bar{\sigma}^k\}_{k=1}^{\infty}$, it must be that the common cutoff time, which we will denote by $-\hat{t}^k$, does not approach infinity either, and there is a positive probability of reaching it. This means that the parameters induce substantial delay, and inefficiency follows. □

[8] For convenience, and without loss of generality, we will assume throughout the proof that all the sequences we mention converge (in the broad sense).

B.3.10 Proof of Corollary 21

Corollary 53. *In a war of attrition model, assume Player* 1 *is stronger than Player* 2*; then her preferred outcome is the unique limit of the equilibrium payoff set as the probability of Player* 2 *being the commitment type approaches zero. Formally,* $\liminf_{\xi_2 \to 0} \bar{\Phi}^{woa}(u_1, u_2; 0, \xi_2; \lambda_1, \lambda_2) = \{u(U, L)\}$.

Proof. The existence proof is very similar to (an abbreviated version of) the proof of Theorem 17. Following Corollary 19, we need only to show that there is no other pair of cutoff strategies that forms a SE. We can rule out this possibility by showing that there are no such equilibria with $-t_1^* < -t_2^*$ for small enough ξ_2 and for large enough T.[9] The other case, in which $-t_2^* < -t_1^*$, is already nailed down in the calculations appearing in the proof for Theorem 17. To see that indeed there are no equilibria of the former type, suppose that Player 1 starts exiting first. Then the rational Player 2 must be indifferent at $-t_2^*$, which gives

$$u_2(U, L) = \left(1 - e^{-(\lambda_1+\lambda_2)t_2^*}\right) \cdot \frac{\lambda_2 u_2(U, L) + \lambda_1 u_2(D, R)}{\lambda_1 + \lambda_2} + e^{-(\lambda_1+\lambda_2)t_1^*} \cdot u_2(U, R)$$

or

$$e^{-t_2^*} = \sqrt[\lambda_1+\lambda_2]{\frac{\lambda_1 \left(u_2(D, R) - u_2(U, L)\right)}{\lambda_1 u_2(D, R) + \lambda_2 u_2(U, L) - (\lambda_1 + \lambda_2) u_2(U, R)}}. \tag{B.17}$$

We also know that Player 1 is indifferent at $-t_2^*$ (because she is indifferent along the interval $(-t_1^*, -t_2^*)$), that is,

$$u_1(D, R) = (1 - \xi_2)\left[\left(1 - e^{-(\lambda_1+\lambda_2)t_2^*}\right) \frac{\lambda_2 u_1(U, L) + \lambda_1 u_1(D, R)}{\lambda_1 + \lambda_2} + \right.$$
$$\left. e^{-(\lambda_1+\lambda_2)t_2^*} u_1(U, R)\right] + \xi_2 \left[\left(1 - e^{-\lambda_1 t_2^*}\right) u_1(D, R) + e^{-\lambda_1 t_2^*} u_1(U, R)\right].$$

---

[9] Both t_1^* and t_2^* may be functions of ξ_2 and T. We omit this dependence in our notation except where it is crucial for understanding.

If we define $\overline{t_2} \equiv \lim_{\xi_2 \to 0} \lim_{T \to \infty} t_2^*(\xi_2, T)$, taking the last expression to the limit gives us

$$e^{-\overline{t_2}} = \sqrt[\lambda_1+\lambda_2]{\frac{\lambda_2\left(u_1(U,L) - u_1(D,R)\right)}{\lambda_1 u_1(D,R) + \lambda_2 u_1(U,L) - (\lambda_1+\lambda_2) u_1(U,R)}}. \tag{B.18}$$

Turning back to the assumption that Player 1 is stronger than Player 2, (B.17) together with the limit in (B.18) yield a contradiction, as needed when ξ_2 tends to zero. $\square$

Appendix C

Appendix to Chapter 3

C.1 Proofs

C.1.1 Proof of Lemma 22

1. This part is immediate from the definitions and is omitted.

2. Let $\succ_h$ be a substitutable preference that satisfies the law of aggregate demand,[1] and assume in contradiction that $\succ_h^{q=q'}$ does not satisfy the law of aggregate demand. Then there are two sets of doctors $D' \subseteq D''$ such that $|C_h^{q=q'}(D')| > |C_h^{q=q'}(D'')|$, where $C_h^{q=q'}$ is the choice function related to the preference $\succ_h^{q=q'}$.
   Let $E = C_h\left(C_h^{q=q'}(D') \cup C_h^{q=q'}(D'')\right)$. Then from the law of aggregate demand (applied to $\succ_h$) we have $|E| \geq \left|C_h^{q=q'}(D')\right|$, which means $E \nsubseteq C_h^{q=q'}(D'')$, and there exists some $d \in E \setminus C_h^{q=q'}(D'')$. Since $\{d\} \cup C_h^{q=q'}(D'') \subseteq D''$ and $|C_h^{q=q'}(D'')| < q'$, then

[1]The substitutability assumption is imperative. A counterexample when substitutability is not assumed is the following preference relation:

$$\succ = \{d_1, d_2, d_3\}, \{d_1\}, \{d_1, d_3\}, \{d_2, d_3\}, \{d_3\}$$

which satisfies the law of aggregate demand, but after imposing capacity of 2 the resulting preference does not satisfy the law of aggregate demand.

$C_h\left(\{d\} \cup C_h^{q=q'}(D'')\right) = C_h^{q=q'}(D'')$. Putting all together we get:

$$d \in \left(\{d\} \cup C_h^{q=q'}(D'')\right) \cap C_h\left(C_h^{q=q'}(D') \cup C_h^{q=q'}(D'')\right), \text{ and}$$
$$d \notin C_h\left(\{d\} \cup C_h^{q=q'}(D'')\right)$$

which contradicts the substitutability of $\succ_h$.

3. Consider the following example. Let $D = \{d_1, d_2, d_3, d_4\}$ and $H = \{h\}$. Hospital h's preferences over doctors are given by:

$$\begin{aligned}\succ_h = &\{d_1, d_2, d_3, d_4\}, \{d_1, d_2, d_3\}, \{d_1, d_2, d_4\}, \{d_1, d_3, d_4\},\\ &\{d_2, d_3, d_4\}, \{d_1, d_2\}, \{d_3, d_4\}, \{d_1, d_3\}, \{d_1, d_4\}, \{d_2, d_3\},\\ &\{d_2, d_4\}, \{d_1\}, \{d_2\}, \{d_3\}, \{d_4\}\end{aligned}$$

This preference is substitutable and satisfies the law of aggregate demand (h never rejects any doctors). However, imposing a capacity 2 on $\succ_h$ gives us the following substitutability violation:

$$\{d_2, d_3, d_4\} \cap C_h^{q=2}(D) = \{d_2\} \nsubseteq \{d_3, d_4\} = C_h^{q=2}(\{d_2, d_3, d_4\})$$

C.1.2 Proof of Theorem 26

Let $\mu = \psi^H(P)$ and $\mu' = \psi^D(P')$. Throughout this proof "agent i is better off" means $\mu'(i) \succ_i \mu(i)$, and similarly for "indifferent", "weakly worse off", and so on. Assume in contradiction that there exists no $S \subseteq D$ such that $S \neq \emptyset$ and every doctor in S is worse off, and every hospital in $\{h \mid \emptyset \neq \mu'(h) \setminus \mu(h) \subseteq S\}$ is better off.

Construct a directed graph with vertices $\mathcal{V} = D \cup H$, and edges

$$\mathcal{E} = \left\{(d, h) \mid \mu'(d) = h\right\} \cup \left\{(h, d) \mid \mu(d) = h\right\}.$$

We prove a series of claims that will enable us eventually to show that the number of outgoing edges is strictly larger than the number of incoming edges (in the entire graph), thus reaching a contradiction.

Note that Claims 24.1 and 24.2 continue to hold, but Claim 24.3 relied on responsiveness and does not hold here.

Claim 53.1. *For every hospital $h \in H \setminus \{h_0\}$ that is not indifferent there is at least one doctor in $\mu(h)$ who is better off.*

Proof. If not then h (which is worse off by Claim 24.2) and the doctors in $\mu(h)$ block μ'. ■

Claim 53.2. *For every hospital $h \in H$, if there exists some $d \in \mu'(h)$ who is worse off, then $\mu'(h) \setminus \mu(h) = \{d\}$.*

Proof. If $|\mu'(h) \setminus \mu(h)| > 1$, then $S = \{d\}$ provides a contradiction. ■

Claim 53.3. *For every hospital $h \in H$, the number of doctors in $\mu'(h)$ who are better off is less or equal to the number of doctors in $\mu(h)$.*

Proof. If $h = h_0$ this is immediate from the assumption on q'. For other hospitals, denote by $\tilde{D}$ the set of doctors in $\mu'(h)$ who are better off. From the substitutability of $\succ_h$ and $\tilde{D} \subseteq \mu'(h)$ it is immediate that $C_h(\tilde{D}) = \tilde{D}$, and from the stability of μ it is also true that $C_h(\tilde{D} \cup \mu(h)) = \mu(h)$. Then the law of aggregate demand implies that $\left|\tilde{D}\right| \leq |\mu(h)|$. ■

Claim 53.4. *For every agent $i \in D \cup H$ the number of incoming edges is less or equal to the number of outgoing edges:* $\deg^-(i) \leq \deg^+(i)$.

Proof. If the agent is a doctor, this is immediate from Claim 24.1. Suppose the agent is $h \in H$. If $\deg^-(h) = 0$, the conclusion is immediate. If there exists a doctor $d \in \mu'(h)$ who is worse off, then by Claim 53.2 $|\mu'(h) \setminus \mu(h)| = 1$, but we also know from Claim 53.1 that $|\mu(h) \setminus \mu'(h)| \geq 1$, and so we have $\deg^-(h) = |\mu'(h) \cap \mu(h)| + 1 \leq \deg^+(h)$. If there exists one doctor $d \in \mu'(h)$ who is better off, then by Claim 53.3 the conclusion is correct. And finally, if all the doctors in $\mu'(h)$ are indifferent, then $\deg^-(h) \leq \deg^+(h)$. ■

Sum the indegrees and the outdegrees of all agents. From Claim 53.4 we know that $\sum_{i \in D \cup H} \deg^-(i) \leq \sum_{i \in D \cup H} \deg^+(i)$. Moreover, since by assumption $\deg^-(h_0) \leq q' <$

$m_{h_0}(P) = \deg^+(h_0)$, the inequality is in fact strict. This concludes the contradiction argument, proving that the required S exists.

The conclusion of the theorem follows from the hospital-optimality and the doctor-optimality of μ and μ' respectively in a way similar to the proof of Theorem 23. The polarization of interests for the optimal/pessimal stable matchings under substitutable preferences is proved by Roth (1984b, Theorem 3).

C.1.3 Proof of Theorem 29

Let $\mu = \psi^H(P_{-d_0})$ and $\mu' = \psi^D(P)$. Throughout this proof "agent i is better off" means $\mu'(i) \succ_i \mu(i)$, and similarly for "indifferent", "weakly worse off", and so on. Assume in contradiction that all hospitals in H are weakly worse off.

Construct a directed graph with vertices $\mathcal{V} = D \cup H$, and edges $\mathcal{E} = \{(h,d) \mid \mu(d) = h\} \cup \{(d,h) \mid \mu'(d) = h\}$.

Claim 53.5. *For every hospital $h \in H$, if there is at least one doctor who is better off in $\mu'(h) \setminus \{d_0\}$, then $|\mu(h)| = q_h$, and all doctors in $\mu(h)$ are weakly better off.*

Proof. Let $d' \in \mu'(h) \setminus \{d_0\}$ be better off. If $|\mu(h)| < q_h$ then h and d' form a blocking pair under μ, and therefore $|\mu(h)| = q_h$. Suppose $d \in \mu(h)$ is worse off. From the stability of μ we have:

$$\mu(h) \succ_h \mu(h) \cup \{d'\} \setminus \{d\},$$

and from the stability of μ' that:

$$\mu'(h) \succ_h \mu'(h) \cup \{d\} \setminus \{d'\},$$

which together contradict the responsiveness of $\succ_h$. Hence no such d exists, and all doctors in $\mu(h)$ are weakly better off. ■

Claim 53.6. *Let $h' = \mu'(d_0)$, then the number of doctors who are better off in $\mu(h')$ is strictly larger than the number of doctors who are better off in $\mu'(h') \setminus \{d_0\}$.*

Proof. First, note that there must be at least one doctor in $\mu(h')$ who is better off, or otherwise h' (which is worse off by our contradiction assumption) and $\mu(h')$ form a blocking coalition under μ'. If all doctors in $\mu'(h') \setminus \{d_0\}$ are weakly worse off then we are done, and if some are better off, then use Claim 53.5. ■

Let $D_b \subseteq D \setminus \{d_0\}$ denote the set of doctors who are better off (d_0 not included), $\deg_b^-(h)$ denote the number of incoming edges from doctors in D_b to hospital h, and $\deg_b^+(h)$ denote the number of outgoing edges from hospital h to doctors in D_b. We know that:

$$\forall h \in H : \deg_b^-(h) \leq \deg_b^+(h) \qquad \text{(Claim 53.5)}$$

$$\deg_b^-\left(\mu'(d_0)\right) < \deg_b^+\left(\mu'(d_0)\right) \qquad \text{(Claim 53.6)}$$

$$\forall d \in D_b : \deg^-(d) \leq \deg^+(d) \qquad \text{(individual rationality)}$$

We sum over all hospitals to get:

$$\sum_{h \in H} \deg_b^-(h) < \sum_{h \in H} \deg_b^+(h) = \sum_{d \in D_b} \deg^-(d) \leq \sum_{d \in D_b} \deg^+(d) \leq \sum_{h \in H} \deg_b^-(h)$$

We reached a contradiction and therefore there exists some non-empty set $S \subseteq H$ of hospitals that are better off.

The conclusion of the theorem follows from the hospital-optimality and the doctor-optimality of μ and μ' respectively, in a way similar to the proof of Theorem 23.

C.1.4 Proof of Theorem 30

Let $\mu = \psi^H(P_{-d_0})$ and $\mu' = \psi^D(P)$. Throughout this proof "agent i is better off" means $\mu'(i) \succ_i \mu(i)$, and similarly for "indifferent", "weakly worse off", and so on. Assume in contradiction that there exists no $S \subseteq H$ and $T \subseteq \{d \mid \mu(d) \in S\}$ such that $S \neq \emptyset$, $T \neq \emptyset$, every hospital in S is better off and every doctor in T is worse off.

Construct a directed graph with vertices $\mathcal{V} = D \cup H$, and edges $\mathcal{E} = \{(h,d) \mid \mu(d) = h\} \cup \{(d,h) \mid \mu'(d) = h\}$. As in the proof of Theorem 29, let $D_b \subseteq D \setminus \{d_0\}$ denote the set of doctors who are better off (d_0 not included), $\deg_b^-(h)$ denote the number of incoming edges from doctors in D_b to hospital h, and $\deg_b^+(h)$ denote the number of outgoing edges from hospital h to doctors in D_b.

Claim 53.7. *There exists $h_0 \in H$ such that $\deg_b^-(h_0) < \deg_b^+(h_0)$.*

Proof. Let h_0 be some hospital that is better off (by Theorem 29). If there exists a doctor $d \in \mu(h_0)$ who is worse off, then $S = \{h_0\}$ and $T = \{d\}$ provide a contradiction. It must be then that all doctors in $\mu(h_0)$ are weakly better off. Since by assumption $|\mu(h_0)| = m_{h_0}(P) = q_{h_0}$, it follows that $\deg_b^-(h_0) \leq \deg_b^+(h_0)$. If $d_0 \in \mu'(h_0)$ then we are done. If $d_0 \notin \mu'(h_0)$, and all doctors in $\mu'(h_0)$ are weakly better off, then we get a contradiction to the stability of μ through the blocking coalition composed of h_0 and $\mu'(h_0)$. ■

Putting everything together we know that:

$$\forall h \in H : \deg_b^-(h) \leq \deg_b^+(h) \qquad \text{(Claim 53.5)}$$

$$\exists h_0 \in H : \deg_b^-(h_0) < \deg_b^+(h_0) \qquad \text{(Claim 53.7)}$$

$$\forall d \in D_b : \deg^-(d) \leq \deg^+(d) \qquad \text{(individual rationality)}$$

We sum over all hospitals to get:

$$\sum_{h\in H} \deg_b^-(h) < \sum_{h\in H} \deg_b^+(h) = \sum_{d\in D_b} \deg^-(d) \leq \sum_{d\in D_b} \deg^+(d) \leq \sum_{h\in H} \deg_b^-(h)$$

We reached a contradiction and therefore the required S and T do exist.

The conclusion of the theorem follows from the hospital-optimality and the doctor-optimality of μ and μ' respectively, in a way similar to the proof of Theorem 23.

C.1.5 Proof of Theorem 32

The method of proof is very similar to the one used in proving Theorem 23. We repeat the construction done there and recall that $\mu = \psi^H(P)$ and $\mu' = \psi^D(P')$ (the existence of μ' is guaranteed from Observation 31). Note that Claims 24.1 and 24.2 continue to hold.

Claim 53.8. *For every hospital $h \in H$, if there is at least one doctor who is better off in $\mu'(h)$, then $|\mu(h)| = q_h$, and all doctors in $\mu(h)$ are weakly better off.*

Proof. For every $h \neq h_0$ the proof is the same as the proof of Claim 24.3. For h_0, let $d' \in \mu'(h_0)$ be better off. If $|\mu(h_0)| < q_{h_0}$ then d' and h_0 form a blocking pair for μ, hence $|\mu(h_0)| = q_{h_0}$. Suppose $d \in \mu(h_0)$ is worse off. From the stability of μ we have:

$$\mu(h_0) \succ_{h_0} \mu(h_0) \cup \{d'\} \setminus \{d\},$$

and from stability of μ' we have:

$$\mu'(h_0) \succ_{h_0}^{\mathrm{tr}(\bar{d})} \mu'(h_0) \cup \{d\} \setminus \{d'\},$$

and since $\succ_{h_0}^{\mathrm{tr}(\bar{d})}$ is a truncation, these two statements together contradict the responsiveness of $\succ_{h_0}$. Hence no such d exists and all doctors in $\mu(h_0)$ are weakly better off. ∎

Claim 53.9. *All doctors in $\mathcal{W} \cap D$ are weakly worse off.*

Proof. Let $D_b = \{d \in D \mid d \text{ is better off}\}$. Let $\deg_b^-(h)$ denote the number of incoming edges from doctors in D_b to hospital h, and $\deg_b^+(h)$ denote the number of outgoing edges from hospital h to doctors in D_b.

Unless $\mathcal{W} \cap D_b = \emptyset$ we can find $d' \in \operatorname{argmin}_{d \in \mathcal{W} \cap D_b} \delta(h_0, d)$, where $\delta(x, y)$ denotes the distance between nodes x and y on the graph $(\mathcal{V}, \mathcal{E})$. We denote $h' = \mu(d')$. We claim that:

$$\deg_b^-(h') < \deg_b^+(h'). \tag{C.1}$$

Note that $d' \in \mu(h')$ and so $\deg_b^+(h') \geq 1$. If $\deg_b^-(h') = 0$, then we are done. Let $n_i = |\mu(h') \cap \mu'(h')|$ denote the number of doctors who are indifferent in $\mu(h')$. If $1 \leq \deg_b^-(h') + n_i < q_{h'}$ then we can use Claim 53.8 to prove the strict inequality. Finally, if $1 \leq \deg_b^-(h') + n_i = q_{h'}$ we need to distinguish between two cases. If $h' = h_0$ then we know all doctors in $\mu'(h_0)$ are weakly better off. However, there must be at least one doctor $d'' \in \mu'(h_0) \setminus \mu(h_0)$ such that $d'' \succ_{h_0} d^* \in \mu(h_0)$ (by the assumption on the truncation), and so we get a contradiction to the stability of μ. If $h' \neq h_0$, then there must be $d'' \in \mu'(h') \cap D_b$ such that $\delta(h_0, d'') < \delta(h_0, d')$, contradicting the way d' was chosen.

Putting everything together we know that:

$$\forall h \in H : \deg_b^-(h) \leq \deg_b^+(h) \qquad \text{(Claim 53.8)}$$

$$\deg_b^-(h') < \deg_b^+(h') \qquad \text{(Equation C.1)}$$

$$\forall d \in D_b : \deg^-(d) \leq \deg^+(d) \qquad \text{(individual rationality)}$$

We sum over all hospitals in H to get:

$$\sum_{h\in H} \deg_b^-(h) < \sum_{h\in H} \deg_b^+(h) = \sum_{d\in D_b} \deg^-(d) \leq \sum_{d\in D_b} \deg^+(d) \leq \sum_{h\in H} \deg_b^-(h)$$

Which is a contradiction, proving that $\mathcal{W} \cap D_b$ must be empty. ■

Pick some $d \in \mu(h_0) \setminus \mu'(h_0)$, and let $h = \mu'(d)$ (exists by Claim 24.1). We get that hospital h (which is worse off by Claim 24.2) and the doctors in $\mu(h)$ (who are weakly worse off by Claim 53.9) form a blocking coalition under μ'. This concludes the contradiction argument, proving that the required S exists.

The conclusion of the theorem follows from the hospital-optimality and the doctor-optimality of μ and μ' respectively, in a way similar to the proof of Theorem 23.