Making it Safe to Use Centralized Markets: ϵ - Dominant Individual Rationality and Applications to Market Design

PRELIMINARY AND INCOMPLETE

Benjamin N. Roth*[1] and Ran I. Shorrer†[2]

[1]Department of Economics, MIT
[2]Department of Economics, The Pennsylvania State University

October 13, 2016

Abstract

We take the view that centralizing a market is akin to designing a mechanism to which people may voluntarily sign away their decision rights. Volunteers send a message to the mechanism which can credibly act on their behalf but those who retain their decision rights act on their own accord in a decentralized manner. In this general setting we ask when it is possible to guarantee agents will use a mechanism despite the voluntary nature of participation. Our first result is negative and derives from adverse selection concerns. Near any game with at least one pure strategy equilibrium we prove there is another game in which no mechanism can eliminate the equilibrium of the original game.

In light of this we offer a new desideratum for mechanism and market design which we term ϵ - *dominant individual rationality.* After noting its robustness, we establish two positive results about centralizing large markets. The first offers a novel justification for stable matching mechanisms and an insight to guide their design to achieve ϵ - dominant individual rationality. Our second result demonstrates that in large games, any mechanism with the property that every player wants to use it conditional on sufficiently many others using it can be modified to satisfy ϵ - dominant individual rationality while preserving its behavior conditional on sufficiently many players using it. The modification relies on a class of mechanisms we refer to as *random threshold mediators*, and resembles insights from the differential privacy literature.

*benroth@mit.edu

†rshorrer@gmail.edu. This material is based on work supported by the NSF Graduate Research Fellowhip under Grant No. 1122374. We would like to thank Daron Acemoglu, Itai Ashlagi, Abhijit Banerjee, Elchanan Ben-Porath, Laura Doval, Drew Fudenberg, Sergiu Hart, Nicole Immorlica, Scott Kominers, Brendan Lucier, Eric Maskin, Assaf Romm, Alvin Roth and Robert Townsend for helpful discussions. All errors are our own.

1 Introduction

Market designers spend a great deal of attention on the study of safety in centralized markets. Safety may be formalized in at least two distinct ways: safety for participants to state their true preferences (i.e. incentive compatibility) and safety in participating in the centralized market rather than in existing institutions the centralized market was meant to displace. The literature on this latter form of safety is sparse, and largely focuses on applications in which it is possible to create *an* equilibrium in which players find it worthwhile to participate in the centralized market. This is in stark contrast to the literature on incentive compatibility in market design, which primarily focuses on dominant-strategy incentive compatibility – a guarantee that it is safe to tell the truth about one's preferences regardless of what other players do. Similarly, in a setting where a centralized market is introduced to displace previously decentralized interactions, it may be especially important to identify mechanisms that guarantee the participation of relevant agents irrespective of what other agents do.[1]

Section 2 outlines our formal approach to the study of centralized market design. We analyze a general model of mechanism design in the presence of a pre-existing game. The pre-existing game specifies the players, feasible actions, and payoffs. A designer may introduce a mechanism (alternatively referred to as a centralized market), to which players may sign away their decision rights (i.e. participate in the centralized market). Players who do so select a message to send to the mechanism, which then acts on their behalf in a pre-specified manner as a function of the whole set of messages it receives. The mechanism is voluntary in the sense that players may act decentrally by choosing one of their original actions, and the mechanism cannot condition the actions of centralized participants on those of the players who act decentrally.

This may be a sensible approach for studying many market design applications. For instance, hospitals and residents who use the centralized clearinghouse known as the National Residency Matching Program are committed to follow through with their assigned match.[2] However there is no legal barrier preventing members of either side of the market from declining to use the mechanism and finding a match outside of the clearinghouse. This was becoming increasingly common in the early 1990s before the clearinghouse was redesigned to better accommodate couples seeking two jobs.[3]

Outside of market design many other centralized markets fall within this framework. Centralized financial markets such as the New York Stock Exchange elicit supply and demand curves from traders and have the authority to implement complex market clearing trades on their behalf. However traders have and frequently exercise the ability to execute trades outside of the centralized market, perhaps to the detriment of other users.[4] Crowd funding platforms like Kickstarter and online marketplaces like Groupon offer to centralize transactions (financing new projects and purchasing goods and services) such that users can offer to participate conditional on sufficient participation by others. But users retain the option to transact outside of these centralized platforms.

[1] This is also in line with the Wilson Doctrine, which dictates that mechanisms should be "detail free" (see Wilson (1987)).

[2] See http://www.nrmp.org/policies/the-match-commitment.

[3] See Kojima et al. (2013)

[4] See Patterson (2013).

Within this framework we study a variety of desiderata which may ensure that players will use the centralized mechanism. In Section 3 we explore the criterion that a mechanism should induce a game in which all equilibria result in a desirable outcome. However, we prove a negative result about the robustness of this criterion. Near any game with a pure strategy equilibrium there is another game in which no mechanism can eliminate the equilibrium of the original game. This can be viewed as a negative result about robust full implementation in games where participation in the mechanism is voluntary.[5] We expand on this point in Section 3.

In Section 4 we identify a desideratum we term ϵ - *dominant individual rationality* (ϵ - dominant IR). A mechanism is ϵ - dominant individually rational if, given any decentralized strategy any player may consider, there is an alternative centralized strategy that acheives at most ϵ less utility relative to the decentralized strategy no matter what strategy profile other players follow, and performs strictly better for strategy profiles in which sufficiently many players participate in the centralized market. Notably, this definition does not specify that there is a centralized strategy that an agent can follow that performs almost as well as *all* decentralized strategies regardless of the strategies others follow. In fact, we show that this latter criterion may be unachievable in many settings of interest. While ϵ - dominant individual rationality leaves open the possibility that for any given centralized strategy a player may follow, she may ex-post regret not having followed another decentralized strategy, it still offers a compelling justification for players to use the centralized market. For any decentralized strategy a player is considering following *ex-ante*, a centralized action exists that ϵ - dominates it. And for $\epsilon = 0$, the definition implies that each decentralized strategy is dominated by a centralized one.

We establish a number of positive results about ϵ - dominant individual rationality. First, it is robust in the sense that any mechanism that is ϵ - dominant IR will continue to be so for all nearby games. Second, we show it can be achieved in two applications of interest.

In Section 5 we study a large two sided matching market and show that with simple modification, any stable matching mechanism can be ϵ - dominant IR (with ϵ becoming arbitrarily small for increasingly large markets). The result arises from the fact that for proposer i, the decentralized strategy of making an offer to player j is ϵ - dominated by submitting any preference list to the centralized mechanism in which j is listed first. If j would have rejected i's decentralized offer, then instead using the centralized mechanism can only benefit i. If j would have accepted i's decentralized offer, then j must prefer i to the stable match k, suggested for j by the mechanism. So if i were to enter the centralized market and list j first, i and j would block the tentative matching between j and k. While this intuition does not guarantee that i and j will be matched by the stable matching mechanism, arguments similar to Immorlica and Mahdian (2005) and Kojima and Pathak (2009) lead to an approximately guarantee large markets.

In Section 6 we show that in arbitrary large games, any mechanism with the property that every player wants to use it conditional on sufficiently many other players using it, can be modified to be ϵ - dominant IR in such a way as to preserve the behavior of the mechanism conditional on a sufficiently large fraction of players participating (again, such that ϵ can become arbitrarily small for increasingly large markets). The intuition for the

[5] See Bergemann and Morris (2003).

result relies on a threshold property of the modified mechanism, such that the mechanism elicits a default strategy from every participant and plays the default strategy unless a sufficiently large number of players choose to participate. Conditional on sufficiently many players participating, the modified mechanism behaves exactly as the original mechanism would have. However, this modification on its own may not make the mechanism ϵ - dominant IR for small ϵ because players may anticipate strategy profiles in which their decision to enter the centralized market is pivotal in determining whether other players use their default action or the action specified by the mechanism. We solve this using a technique inspired by the differential privacy literature; we add a random perturbation to the threshold, so that given any strategy profile the probability that any agent is pivotal is vanishingly small in large markets.

Important to note in both applications, while the ϵ regret allowed by ϵ - dominant IR can be made arbitrarily small in large markets, the benefit each agent receives from using the mechanism in any strategy profile with sufficiently many others also participating does not vanish with the size of the market. Thus, in the large market limit, every decentralized strategy is strictly dominated by a centralized strategy offered by the mechanism.

Our paper fits into several literatures. Maskin (1999) is the seminal paper in implementation theory in games of complete information, which now forms a large body of work asking the question in what settings a social choice function can be implemented uniquely under a variety of solution concepts. Postlewaite and Schmeidler (1986) and Jackson (1991) find analogous necessary and sufficient conditions for implementation in games of incomplete information. One recent paper that asks which actions in a pre-existing game form can be implemented uniquely is Kar et al. (2010). We refer interested readers to Jackson (2001) and Maskin and Sjöstrom (2002) for two surveys of this literature.

There is a growing literature on robust partial and full implementation of social choice functions (see e.g. Bergemann and Morris (2003)). Most closely related within this literature is Oury and Tercieux (2012) who study when, in the canonical environment, a social choice function can be partially implemented in a given environment and in all nearby environments. They derive restrictive necessary and sufficient conditions that are closely related to the conditions of full implementability in rationalizable strategies.

Our model closely resembles that of Kearns et al. (2015), Myerson (1991), Kalai et al. (2010), Forges (2013) and especially Monderer and Tennenholtz (2009) and Ashlagi et al. (2007). These papers study the set of equilibria implementable via some mechanism in a one shot game and by and large obtain permissive answers. All of these ask questions about the existence of an equilibrium outcome, while we focus on equilibrium uniqueness and related criteria.

Lastly, there has been a recent interest in applied models in which players can choose not to participate in particular mechanisms. Sönmez (1999) and Kesten (2012), among others, study a form of manipulation in school choice problems in which schools and students prearrange their matches prior to use of a centralized mechanism. Ekmekci and Yenmez (2014) study a complete information model of school choice in which schools may opt out of the central matching mechanism. They show that the Deferred Acceptance mechanism has no equilibrium in undominated strategies that attracts all schools to participate and propose modifications. We reach different results using large market models, with incomplete information about preferences, where the concerns they focus on are minimal.

2 The Model

The Underlying Game: Let $\Gamma := \langle N, \{A_i, \Theta_i, u_i\}_{i\in N}, q\rangle$ be a normal form game of (potentially) incomplete information, where $N = \{1, \cdots, N\}$ is the finite set of players, and for every $i \in N$, Θ_i is a set of types, A_i is a finite set of actions, and $u_i : \Theta \times A \to \mathbb{R}$ is a Bernoulli utility function, where $\Theta = \prod_{i\in N} \Theta_i$ and $A = \prod_{i\in N} A_i$. Last, q is a probability distribution over Θ from which each player derives their conditional beliefs using Bayes law. Mixed strategy action profiles are defined in the usual way and we assume that agents are expected utility maximizers. We write $X_S = \prod_{i\in S} X_i$, and for a vector x we write x_i to refer to its i'th component, $x_{-i} = (x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_N)$ and $x_S = \prod_{i\in S} x_i$.

Mechanisms: A *mechanism* is a tuple $M = \langle R, (\mu_S)_{S\subseteq N}\rangle$, where $R = \prod_{i\in N} R_i$ is a finite *message space* and for every subset $S \subseteq N$, μ_S is a mapping from messages in R_S to distributions over action profiles in A_S.[6]

The key novelty is that players are not required (but have the option) to sign away their decision rights to the mechanism, which is akin to participating in the mechanism in the standard analysis. Those that sign away their decision rights specify some message $r_i \in R_i$. For each possible set S of players that sign away their decision rights to the mechanism, $\mu_S : R_S \to \Delta(A_S)$ is a mapping from instructions that the players in S send the mechanism to a mixed strategy profile on behalf of those players. Note that, because using the mechanism is voluntary, a different outcome function μ_S must be specified for each possible coalition S that might sign away their decision rights.

A note about the functionality of mechanisms is in order. A mechanism is a form of coalitional commitment and coordination. Players who sign away their decision rights to the mechanism are committed to following through with the mechanism's recommendation, and the mechanism can be used to coordinate players actions when the state is incompletely known by some players. Finally, our notation implies that mechanisms can implement correlated distributions over action profiles.

The Induced Game: A mechanism M applied to Γ induces a simultaneous move game $\Gamma^M := \langle N, \{A_i \cup R_i, \Theta_i, \bar{u}_i\}_{i\in N}, q\rangle$. Here, $\mathbb{E}\left[\bar{u}_i(r_S, a_{N\setminus S}, \theta)|\theta_i\right] = \mathbb{E}\left[u_i(\mu_S(r_S), a_{N\setminus S}, \theta)|\theta_i\right]$ and all other objects are defined in the natural way.

In words the induced game is as follows. Players simultaneously choose whether or not to relinquish their decision rights to the mechanism. A player i who signs away his or her decision rights also sends an instruction r_i from the set R_i. Instructions are mapped in a pre-specified manner into actions on behalf of the participating players as determined by the relevant outcome function μ_S, corresponding to the coalition S that signed away their decision rights. All players in $N \setminus S$ who opt to keep their decision rights choose their own actions. Then outcomes are realized. Note that the mechanism can condition the actions of participating players on the messages of other participating players, but cannot condition the actions of participating players on the actions of those who choose to retain their decision rights and act on their own behalf.

[6] The solution concept we use in the Section 3 is *trembling hand equilibrium* which is defined over finite action spaces. The assumption of finite action and message spaces can be relaxed with a suitable generalization of trembling hand equilibrium to games with infinitely many actions (see Simon and Stinchcombe (1995)).

3 A Negative Result for Robust Implementation

As discussed in the introduction, a natural desideratum for a mechanism to guarantee a good outcome is that all equilibria in the game it induces result in that good outcome. In this section we argue that this desideratum suffers from a lack of robustness – near any game with a pure strategy equilibrium is another game in which no mechanism can satisfy this criterion unless it implements the original equilibrium. Readers interested only in ϵ -dominant individual rationality and its applications may skip directly to Section 4.

Before a formal statement of our negative result we need several preliminary definitions. Our equilibrium concept is *trembling hand equilibrium* (henceforth "equilibrium"). It requires that each player is best responding to a sequence of profiles of totally mixed strategies that converge to the equilibrium strategy profile. See Selten (1975) for a formal description. We utilize a solution concept with trembles because the question at hand is about uniqueness of equilibrium outcomes and we want to rule out equilibria where no agent signs away their decision rights out of the expectation that no one else will.

Next we define what it means for a mechanism to implement a social choice function (a function $f : \Theta \to \Delta(A)$) uniquely.[7]

Definition 1. The social choice function (SCF) $f(\cdot)$ can be *implemented uniquely* if there exists a mechanism M that induces a game Γ^M all of whose equilibrium outcomes result in $f(\theta)$.

Finally we must define what it means for one game to be near another.

Definition 2. We say a game $\Gamma' := \left\langle N', \{A'_i, \Theta'_i, u'_i\}_{i\in N}, q'\right\rangle$ *is within* ϵ of a game $\Gamma := \left\langle N, \{A_i, \Theta_i, u_i\}_{i\in N}, q\right\rangle$ if

- $N = N', A_i = A'_i \ \forall i.$
- $u_i(a, \theta) = u'_i(a, \theta) \ \forall a \in A, \theta \in \Theta, i \in N.$
- $\Theta_i \subseteq \Theta'_i$ and there is an event E with probability $q'(E) \geq 1 - \epsilon$ such that $q(\theta) = q'(\theta|E) \forall \theta \in \Theta$.
- $\max\limits_{i,a,\theta'_i\in\Theta'_i} u'_i(a, \theta') \leq \max\limits_{i,a,\theta_i\in\Theta_i} u_i(a, \theta) + \epsilon$ and $\min\limits_{i,a,\theta'_i\in\Theta'_i} u'_i(a, \theta') \geq \min\limits_{i,a,\theta_i\in\Theta_i} u_i(a, \theta) - \epsilon.$

That is, a game Γ' is within ϵ of a game Γ if the players and actions are the same, and there is an event E with probability at least $1 - \epsilon$ in game Γ' such that conditional on E the distribution of payoff types in Γ' is the same as the distribution of payoff types in Γ. For payoff type profiles that occur with positive probability in both games, utilities are the same, and utilities are bounded for types that can be realized in Γ' that cannot be realized in Γ.

Theorem 1. *Consider any game* $\Gamma := \left\langle N, \{A_i, \Theta_i, u_i\}_{i\in N}, q\right\rangle$ *with at least one pure strategy equilibrium* σ*. For all* ϵ *there exists a game* Γ' *within* ϵ *of* Γ *such that any SCF* f *that is uniquely implementable in* Γ' *must satisfy* $\sigma(\theta) = f(\theta)$ *for all* $\theta \in \Theta$.

[7] While we restrict attention to social choice *functions* for simplicity, Theorem 1 below is easily extended to accommodate social choice *correspondences*.

Proof. See appendix. □

The intuition for the proof can be seen through the following example.

Example 1.

	C	D	S
C	2,2	0,2.5	0,0
D	2.5,0	1,1	0,0
S	0,0	0,0	0,0

Figure 1a

		g ($\sqrt{1-\epsilon}$)			b ($1-\sqrt{1-\epsilon}$)		
		C	D	S	C	D	S
g ($\sqrt{1-\epsilon}$)	C	2,2	0,2.5	0,0	-1,2	-3,2.5	-3,0
	D	2.5,0	1,1	0,0	-.5,0	-2,1	-3,0
	S	0,0	0,0	0,0	0,0	0,0	0,0
		C	D	S	C	D	S
b ($1-\sqrt{1-\epsilon}$)	C	2,-1	0,-.5	0,0	2,2	0,2.5	0,0
	D	2.5,-3	1,-2	0,0	2.5,0	1,1	0,0
	S	0,-3	0,-3	0,0	0,0	0,0	0,0

Figure 1b

The game matrix in Figure 1a represents a two player game of complete information. Restricting the game to C and D forms a prisoner's dilemma. If a player chooses S, corresponding to "stay home," he guarantees himself a payoff of 0. The game has two pure strategy equilibria, (D, D) and (S, S).

The game matrices in Figure 1b represent a modification of the game in Figure 1a. Figure 1b represents a 2 player simultaneous move game in which each player has private information about his own type, (g)ood or (b)ad, with types drawn independently, and with good types drawn with probability $\sqrt{1-\epsilon}$.

If both players are good their payoffs are the same as in Figure 1a. However conditional on facing a bad player, a good player guarantees her highest possible payoff by playing S. Bad players face the same payoffs as in Figure 1a regardless of the type of their opponent. Note that the games in figures 1a and 1b are within ϵ of one another.

For illustrative purposes we focus on the social choice function that implements the utilitarian optimum, $f(\theta) = (C, C)\ \forall\theta$. We argue that for any mechanism that induces a game one of whose equilibria features the outcome (C, C) whenever both players are good, there is also an equilibrium where all good players opt to keep their decision rights and play S.

We construct a strategy profile in which the good players keep their decision rights and play S. Bad players either use the mechanism with positive probability or "tremble" to use it more often than good players. While the mechanism may be able to generate incentive compatible Pareto improvements conditional on both players being good, good players don't find it worthwhile to deviate to signing away their decision rights. This is because conditional on the knowledge that her opponent has signed away his decision rights, a good player is confident that her opponent is bad, in which case she is already receiving her highest possible payoff by playing S.

Note that the games in Figure 1a and Figure 1b can be made arbitrarily close. This failure of unique implementation resembles an adverse selection problem; all participants who enter the market are bad. The proof in the appendix formalizes and extends this logic to any game with a pure strategy equilibrium.

4 ϵ - Dominant Individual Rationality

Given that unique implementation suffers from non robustness, we offer a new desideratum for mechanisms in this framework. Let Σ be a subset of the set of all strategies $S := \Pi_{i \in N}(\sigma_i : \Theta_i \to \Delta(A_i \cup R_i))$.

Definition 3. We say a mechanism is ϵ - *dominant individually rational with respect to* $\Sigma \subseteq S$ if for all $i \in N$, for all $\theta_i \in \Theta_i$, and for all $a_i \in A_i$, there is a centralized strategy r_{a_i} such that:

1. (ϵ - *Regret*): $\mathbb{E}u(r_{a_i}, \sigma_{-i}, \theta_i) > \mathbb{E}u(a_i, \sigma_{-i}, \theta_i) - \epsilon$ for all $\sigma_{-i} \in \Sigma_{-i}$
2. (k - *Attraction*): There exists a $k < 1$ and $\bar{u} > 0$ such that $\mathbb{E}u(r_{a_i}, \sigma_{-i}, \theta_i) > \mathbb{E}u(a_i, \sigma_{-i}, \theta_i) + \bar{u}$ when $\sigma_{-i} \in \Sigma_{-i}$ and $\frac{\#\{j \neq i \in N : \sigma_j(\theta_j) \in \Delta(R_j)\ \forall \theta_j \in \Theta_j\}}{n} > k$.[8]

A mechanism is ϵ - dominant individually rational with respect to Σ if, for every player i and every decentralized action a_i, there is a mediated action r_{a_i} such that as long as all other players choose strategies in Σ, r_{a_i} guarantees player i an expected payoff at least within ϵ of a_i. Furthermore, if sufficiently many players use centralized actions with certainty, then r_{a_i} guarantees i a strictly better expected payoff than a_i.

A few remarks are in order. We define a mechanism to be ϵ - dominant IR *with respect to a particular subset of strategies* Σ because in some applications, agents need not worry about strategies that are deemed infeasible. In our two-sided matching application below, for example, we will restrict attention to anonymous and undominated strategies, which may be sensible in a setting where proposers cannot coordinate their offers based on the names of receivers.

It is also important to note that we define ϵ - dominant IR with respect to strategies that other players may follow, rather than the actions they take. ϵ - dominant IR allows for the possibility that a player i will regret participating in the centralized market after discovering the types of players he is interacting with and the actions they follow, but rules out the

[8]The expectations above are taken with respect to the types of other players, mixed strategies they may follow, and randomization induced by the mediator on behalf of players who sign away their decision rights.

possibility of (more than ϵ) regret before discovering these realizations but after discovering the strategies of his opponents.

The guarantee of ϵ - dominant IR is that for every decentralized action a_i in A_i, player i has a centralized action r_{a_i} that performs at most ϵ worse regardless of what strategies other players follow. Importantly, however, it *does not* guarantee that there is a single centralized strategy r_i that performs at least within ϵ of the payoff of every decentralized strategy regardless of what others do. We formalize this stronger notion in the next definition.

Definition 4. We say a mechanism is ϵ - *strongly dominant individually rational with respect to* $\Sigma \subseteq S$ if for all $i \in N$, for all $\theta_i \in \Theta_i$, there is a centralized strategy r_i such that

1. $\mathbb{E}u(r_i, \sigma_{-i}, \theta_i) > \mathbb{E}u(a_i, \sigma_{-i}, \theta_i) - \epsilon$ for all $\sigma_{-i} \in \Sigma_{-i}$, and for all $a_i \in A_i$
2. There exists a $k < 1$ and $\bar{u} > 0$ such that $\mathbb{E}u(r_i, \sigma_{-i}, \theta_i) > \mathbb{E}u(a_i, \sigma_{-i}, \theta_i) + \bar{u}$ for all $a_i \in A_i$ when $\sigma_{-i} \in \Sigma_{-i}$ and $\frac{\#\{j \neq i \in N : \sigma_j(\theta_j) \in \Delta(R_j)\ \forall \theta_j \in \Theta_j\}}{n} > k$.

Note any mechanism which is ϵ - strongly dominant IR is also ϵ - dominant IR. An ϵ - strongly dominant individually rational mechanism further assures participants that by choosing their centralized strategy r_i, they won't regret not having chosen any decentralized strategy, regardless of the strategies other players follow. In contrast, a mechanism that is merely ϵ - dominant individually rational allows for the possibility that after having chosen a centralized strategy and observing the strategies that other players are following, player i may regret not having chosen another decentralized strategy. The guarantee made by an ϵ - dominant IR mechanism is that a player i considering any particular decentralized strategy a_i can safely switch to the centralized strategy r_{a_i} without experiencing more than ϵ regret. While the guarantee made by an ϵ strongly dominant IR mechanism is more appealing, in Section 6 we show that in many situations it is unattainable for sufficiently small ϵ.

We close this section with an observation about the robustness of ϵ - dominant individual rationality.

Theorem 2. *Consider any game* Γ *and any mechanism* M *that is* ϵ *- dominant IR with respect to* Σ*. Then there exists an* $\eta > 0$ *such that for any* Γ' *within* η *of* Γ*,* M *is still* ϵ *- dominant IR with respect to* Σ *for any player* i *of type* $\theta_i \in \Theta_i$*.*[9]

The above theorem is reassuring given the lack of robustness noted for unique implementation. The proof, in the appendix, proceeds by noting that any strategy in Γ' induces an expected payoff for player i very near to the corresponding strategy in Γ. Thus for Γ' sufficiently near Γ, the bounds stipulated in Definition 3 will still be satisfied.

[9]Note this is a slight abuse of notation as ϵ- dominant individual rationality was not defined with respect to individual players and types. What we mean is that the conditions stipulated in Definition 3 are satisfied for all players i of types $\theta_i \in \Theta_i$ in the game induced by applying mechanism M to Γ'. Further, Σ is not a well defined set on Γ'. We intend for Σ to represent the set of strategies which when projected onto Θ are in Σ.

5 An Application to Two Sided Matching Markets

In this section we discuss an application to large two sided matching markets. We aim to capture several insights that may guide matching market design. First, standard implementations of centralized markets may be prone to adverse selection issues. Second, these concerns can be alleviated using simple techniques already observed in practice, which make the mechanism ϵ - dominant individually rational. Lastly, while it is natural to attribute the success of stable matching markets to the fact that they implement stable matches, we will argue that they have another desirable feature that may contribute to their success. Stability is paramount in assuring ϵ - dominant individual rationality; it guarantees that any potential match a proposer could acquire decentrally will very likely also be attainable through the centralized match.

5.1 A Decentralized Matching Market

We begin with a model of a decentralized, one to one matching market adapted from the large market model of Immorlica and Mahdian (2005) and Kojima and Pathak (2009) to a non-cooperative environment.[10]

Players and Actions: We let the set of players $N = F \cup W$ be divided into disjoint sets of n firms (proposers) and n workers (receivers).[11] A firm's action corresponds to making an offer to a worker. With slight abuse of notation we write $A_f = W \cup \{f\}$, where the choice of $a_f = f$ corresponds to not making any worker an offer. A worker's action corresponds to an acceptance policy, which is a function from the set of any subset of firms who may have chosen him to a selection among them. That is $A_w = \{G_w : 2^F \to F \cup \{w\}\}$.

A matching of a subset of workers and firms $\nu : \tilde{W} \cup \tilde{F} \to \tilde{W} \cup \tilde{F}$ is a function from a subset of workers $\tilde{W} \subseteq W$ and firms $\tilde{F} \subseteq F$ to itself such that (i) $\nu(w) = w$ if $\nu(w) \notin \tilde{F}$ for all workers w, (ii) $\nu(f) = f$ if $\nu(f) \notin \tilde{W}$ for all firms f and (iii) $\nu(w) = f$ if and only if $\nu(f) = w$. A strategy profile results in a matching ν where a worker w and firm f are matched if $a_f = w$ and $a_w(F') = f$ where $F' \equiv \{f' : a_{f'} = w\}$ is the set of firms who selected worker w.

Preferences and Information Structure: Each worker is given a popularity weight $p_w > 0$ about which he or she is privately informed. Each firm has a random preference list of size $K < n$ drawn in the following manner.

- Step 1: Draw a worker w at random from all workers with weights p_w. This is f's first choice worker.
- Step k: If firm f has K workers on his list, terminate the algorithm. Else, draw a worker w at random from all workers with weights p_w . If w has been selected in a previous step, discard him and repeat this step. If w has never been selected in a previous step, this is f's k'th choice.

[10] Our analysis relies heavily on there being a small core, but much less on other details of the environment. As such it seems likely that the intuitions developed in this section would extend to other matching models with small cores such as e.g. Ashlagi et al. (2016).

[11] The balanced number of agents on each side is of no material importance, and is done only to economize on notation c.f. Ashlagi et al. (2016) and Coles et al. (2016).

Workers are endowed with a preference list in a similar fashion, with the only difference being that they have full preference lists (of length n).

Cardinal preferences for all firms (workers) are given by a function from the rank of their partner to $(0, 1)$, $u_f : \{1, \dots, k\} \to (0, 1)$ ($u_w : \{1, \dots, n\} \to (0, 1)$). For all firms (workers) we assume $u_f(w) = \frac{(n- \text{ rank of } w)}{n}$ ($u_w(f) = \frac{(n- \text{ rank of } f)}{n}$), and normalize the utility of being unmatched, and the utility of being matched to someone who was not ranked, to 0.

As stated above, we assume that p_i is agent i's private information. However we assume that the distributions D_W and D_F from which p_w and p_f are drawn are common knowledge. We assume D_W and D_F are symmetric, such that given any permutation $\pi(\cdot)$ of firms or workers, $D \circ \pi(\cdot) = D(\cdot)$. Thus players understand that their preferences may be correlated, but do not know which firms and workers are popular before preferences are drawn.

Finally, similar to Kojima and Pathak (2009), we make the following assumption on the support of D_W and D_F.

Assumption 1 (Market Thickness)**.** *There is some constant $C < \infty$ such that for every realization of D_W (D_F) and every worker w and w' (firm f and f'), $\frac{p_w}{p_{w'}} < C$ ($\frac{p_f}{p_{f'}} < C$).*

The above assumption requires that the most popular worker (firm) cannot be arbitrarily more popular than the least popular worker (firm). We make this assumption so that, as Kojima and Pathak show, not only do most workers and firms have a unique stable match, but that without knowledge of others' preferences, *all* firms and workers have an arbitrarily high probability of having a unique stable match as the market becomes large.

Lastly, we say a matching ν is *stable* if (i) every worker w prefers $\nu(w)$ to being unmatched and similarly for firms, and (ii) there is no pair (w, f) such that w prefers f to $\nu(w)$ and f prefers w to $\nu(f)$. If there is such a pair, we say that (w, f) *blocks* the matching ν.

Strategies: In the analysis that follows we restrict attention to anonymous, undominated strategies. Anonymity implies that agents only consider strategies that condition their offer or acceptance decisions on their own preferences. We maintain anonymity so that there are potential gains from coordinating the market through a mechanism, and to ensure that in all strategy profiles each agent is arbitrarily likely to have a unique stable match given stated preferences as the market becomes large.

5.2 Stable Matching Mechanisms

We now proceed to describe a *Stable Matching Mechanism* - a mechanism that captures the main features of many centralized matching markets - and the game it induces. The message space R_i for a worker w or a firm f is the set of all strict preference orderings over the other side of the market (of size K for firms and n for workers).[12] For any set of delegators $S = S_W \cup S_F \subseteq W \cup F$, $\mu_S(r_{S_W}, r_{S_F})$ determines a stable matching among the participants in S, ν_S, makes offers on behalf of each firm to the worker assigned to them, and accepts these offers on behalf of the corresponding firms.[13] Thus any firm can make a decentralized

[12]This can be relaxed to requiring that firms list a constant fraction λ of all workers and workers list any number of firms up to K.

[13]That is the mechanism selects a strategy $G_w : 2^F \to F \cup w$ for each participating worker w that accepts his stable match f whenever it is among the set of offers he receives.

offer or it can opt to enter the centralized market and participate in the stable matching mechanism. Similarly any worker can stay out of the market and accept any offer he receives, or he can enter the market and commit to accepting a centralized offer he receives.

In this setting, the mechanism has the potential to solve the coordination problems present in the decentralized market. Firms who use the mechanism are guaranteed to make non overlapping offers, and so every firm stands to benefit from coordination. Formally, we have the following observation:

Observation 1. *In a market of any size $n > 2$, and for any stable matching mechanism M, there exists a $k < 1$ such that M satisfies k - attraction.*

As such one may hope that for sufficiently large markets, all stable matching mechanisms are ϵ - dominant individually rational with respect to anonymous, undominated strategies for arbitrarily small ϵ. Our first result is that this is not the case.

Proposition 1. *There exists an $\epsilon > 0$ such that for any matching market, no stable matching mechanism is ϵ - dominant individually rational with respect to anonymous, undominated strategies.*

The proof is relegated to the appendix. While many strategy profiles may be used to demonstrate this failure of ϵ - dominant IR, one instructive profile resembles an adverse selection problem. Popular firms and workers fear that only unpopular counterparts will enter the centralized market. Thus by entering the centralized market themselves, they are dooming themselves to be matched to a low ranking counterpart with high probability. In contrast, had they stayed out of the centralized market they would have had a high probability to be matched with a counterpart they prefer.

That centralized markets can suffer from adverse selection is unsurprising. What may be surprising is that a simple adjustment to the centralizing mechanism can eliminate this concern and guarantee participation.

5.3 The Modified Stable Matching Mechanism

This section highlights two simple ideas. First, by allowing workers who join the centralized market to accept offers from either part of the market, rather than forcing them to accept offers from firms who also joined the centralized market, entering becomes a dominant strategy for the workers. Second, conditional on all workers using the centralized market *and the fact that the mechanism implements a stable outcome given stated preferences*, all firms may be enticed to use the centralized market as well. This is due to the fact that in a large market where most participants have at most one stable partner, the mechanism guarantees that with high probability they can be matched centrally to any worker they could have achieved decentrally by entering the centralized market and listing that worker first. A somewhat subtle logic underlies this second observation and understanding it is the goal of this section.

We now introduce the *modified stable matching mechanism*. The message space R_i is the same as above; workers and firms who sign away their decision rights to the mechanism communicate their preference list. The mechanism calculates a stable match among the players who signed away their decision rights and makes offers on behalf of firms to their assigned stable match. However, the mechanism does not force workers to accept their assigned stable partner. Instead, workers are allowed to accept their most preferred offer regardless of whether the firm who made that offer was part of the centralized match.[14] We are now ready to state our main result.

Theorem 3. *For every $\epsilon > 0$ there exists an $\bar{n}$ such that for any matching market of size $n > \bar{n}$, the modified stable matching mechanism is ϵ - dominant IR with respect to anonymous, undominated strategies.*

Proof. See appendix. □

The logic behind this result follows a cost-benefit analysis. As we argued above, allowing workers to accept their best offer from any firm makes it a dominant strategy for them to enter; workers who enter don't change the behavior of firms in the decentralized market and make themselves eligible to receive offers from firms in the centralized market. Hence for workers, the benefit to entering is the possibility of garnering an additional offer and there is no cost.

For firms, we perform the cost benefit analysis on a particular pair of strategies. We compare f's payoff from the decentralized strategy of making an offer to worker w, with its payoff from the centralized strategy of signing away its decision rights and submitting its true preference ordering but with w is listed first. The benefit of the centralized strategy is clear. If w would reject f's decentralized offer, there is some chance that it is because w received a more preferred centralized offer, in which case the mechanism will have f make an offer to someone else instead. Thus, the mechanism satisfies k - attraction for firms.

Relative to its decentralized strategy, the cost of f's centralized strategy is that w may have accepted it had it proposed decentrally, but the mechanism will match f with someone it disprefers to w (or no one at all). A first sign that this cost is small comes directly from the definition of stability. For an arbitrary strategy profile, let $\tilde{F}$ be the set of firms other than f who sign away their decision rights to the mechanism. If w would have accepted f's decentralized offer, it means that he prefers f to whatever centralized match $\tilde{f} \in \tilde{F}$ he received. Thus w and f block any tentative matching that pairs w with $\tilde{f}$ and f with someone it likes less than w. The primary wrinkle in this line of logic is that by joining the centralized match, f may displace another firm that would trigger a chain of displacements that would finally cause w to be matched to some new firm $\bar{f}$ that he prefers to both $\tilde{f}$ and f.

The bulk of the proof is devoted to showing that cost imposed on f of this wrinkle is small. The proof relies crucially on the small core result of Immorlica and Mahdian (2005) that as n becomes large most firms and workers have at most one stable partner. However, our Theorem 3 does not follow immediately from their result. From a result similar to Immorlica and Mahdian's we know that in the centralized market that includes $W \cup \tilde{F}$, worker w has at most one stable partner $\tilde{f}$ with high probability. We also know that in the market that includes $W \cup \tilde{F} \cup f$, w has at most one stable partner with high probability. Their results do not directly imply, however, that w is matched with f with high probability in the latter market.[15] We show this to be the case, with the following argument.

[14]That is, the mechanism chooses a strategy $G : 2^F \to F$ on behalf of each worker w such that the firm f selected is the one that worker w said he most preferred from the subset that made him offers.

[15]We further remark that their results do not apply to our setting as ours is a setting of strategic behavior,

1. First we rely on the small core result similar to that of Immorlica and Mahdian (2005) that an arbitrary worker $\tilde{w}$ has at most one stable partner with high probability in the market that includes $W \cup \tilde{F}$.[16]

2. Next we invoke a result of Demange et al. (1987) that workers who have at most one stable match cannot manipulate the firm optimal stable matching mechanism.

3. We now consider the firm optimal stable match in the centralized market that includes $W \cup \tilde{F} \cup f$, which can be implemented via the firm proposing deferred acceptance algorithm. The output of the algorithm is insensitive to the order in which proposals are made, so we run the algorithm such that all firms in $\tilde{F}$ continue to make offers until all of them are either matched or have exhausted their list of acceptable matches before f makes any offers. At this point, the tentative matching is the firm optimal stable matching in the market that includes only $W \cup \tilde{F}$. Next f proposes to workers in order of its preference list until one of them prefers it to his tentative match. By assumption, worker w, the first worker who firm f proposes to, will hold its offer.

4. If upon further proposals in the algorithm w receives an offer from some firm $\bar{f}$ whom he prefers to f, we argue that he also could have profitably manipulated the firm optimal stable matching mechanism in the market that included only $W \cup \tilde{F}$ - by truncating his preference list right above his previous tentative match. By step 2, this means that he had more than one stable partner in the market that includes only $W \cup \tilde{F}$, which we know occurs with arbitrarily small probability.

5. Thus, with arbitrarily high probability, F will be matched with w in the firm optimal stable matching in the market that includes $W \cup \tilde{F} \cup f$. We now reapply the small core result from step 1 to this extended market to show that with high probability $\tilde{w}$ and f are paired in all stable matchings.

The intuition seems not to rely on specific details of the market. By allowing workers to accept offers from anyone regardless of their participation, it becomes a dominant strategy for them to enter. And in markets with small cores, stable matching mechanisms make it safe for firms to enter the market as well because, with high probability, entry won't cause a firm to lose anyone she may have attained via a decentralized offer.

In the appendix we show that this intuition does not hold vacuously. In other commonly used but unstable mechanisms such as the Boston Mechanism or the Rank Sum Mechanism, firms may be matched to partners they disprefer to those they could have attained decentrally. Intuitively, these two mechanisms privilege workers who list the firm highly regardless of the firm's preferences.

One last note is in order about the policy suggestion of allowing workers to accept offers from anyone in the market regardless of participation in the centralized mechanism. This almost exactly mirrors a policy suggestion made by Niederle et al. (2006) when considering

and as such we need to verify that the set of stable matchings is small with high probability given reported preferences arising from any anonymous and undominated strategy profile.

[16]Note this is not implied by their results because the set of firms participating in the market may be small and thus existing large market results do not immediately apply.

how to centralize the job market for new gastroenterologists. In particular they proposed doctors should not be bound to offers they receive from any hospital weakly prior to the date of the centralized match, regardless of the form in which the offer was delivered. They understood that this would make it a dominant strategy for doctors to use the centralized mechanism. The analysis above illuminates the role of stable matching in encouraging the hospitals to follow suit.

6 Threshold Mechanisms in Large Market Games

In this section we argue that in large games, if a mechanism satisfies k - attraction, then an intuitive modification allows it to satisfy ϵ - dominant individual rationality, while preserving the mechanism's behavior in the event that sufficiently many players use it. That is, given a mechanism for which it is in every player's best interest to use it conditional on sufficiently many others using it as well, we present a modification that is ϵ - dominant IR, where ϵ goes to zero as the market becomes large. We also show that the same cannot be said for ϵ - strong dominant individual rationality.

The idea highlighted in this section relies on what we term *threshold mechanisms*. A threshold mechanism is one that acts on players' behalves if and only if sufficiently many sign away their decision rights. Else it implements pre-specified default actions on each player's behalf. Mechanisms like this arise in many natural environments. At its inception, Groupon allowed sellers to offer discounts, but to condition them on sufficient demand. Similarly, Kickstarter allows users to volunteer financing for entrepreneurs if sufficiently many others do so, and otherwise returns their money as well. For an example in the market design literature, the NRMP has a clause that the centralized match will be cancelled in the event that fewer than 70% of all applicants use it. We will show that, with minor modification, this class of mechanisms is quite powerful.

Participation in a threshold mechanism entails a potential tradeoff. On the one hand, entry allows agents to condition their behavior on the behavior of others. On the other hand, agents may fear that by entering themselves, they are pivotal in triggering other participants thresholds and are thus affecting the behavior of the other participants in undesirable ways. We show that a modified threshold mechanism, which adds a small amount of noise to the threshold, eliminates this tradeoff. By participating in the mechanism agents can condition their actions on their peers' actions while the worry that they are pivotal vanishes.[17]

Formally, we are interested in arbitrary games with utility functions bounded between 0 and 1. In addition to a standard report, a *threshold mechanism* $M = \langle R, \{\mu_S\}_{S \subseteq N} \rangle$ elicits a default action a_i. That is, $R_i = \hat{R}_i \times A_i$. Let S be the set of players who sign away their decision rights, and for each $i \in S$ let $r_i = (\hat{r}_i, a_i)$. Let $x = |S|$ be the number of players who sign away their decision rights. Then

$$\mu_S(r_S)_i = \begin{cases} \hat{\mu}_S(\hat{r}_S)_i \text{ if } x \geq \bar{x} \\ a_i \text{ if } x < \bar{x} \end{cases}$$

[17]Such randomness is not typical in formal rules of real-life mechanisms, but could be thought of as additional randomness that our models do not capture (e.g. would a seller on Groupon actually purchase 3 coupons himself if the cut-off is almost achieved?)

That is, if $x \geq \bar{x}$ then the mechanism plays an action on behalf of each player $i \in S$ that is a function of the messages that all players in S submitted to the mediator. Crucially, if the threshold $\bar{x}$ is not met, then no player conditions his or her action on any other player's message.

A *noisy threshold mechanism* has the same message space, but also draws a random number $y \sim F$ for an arbitrary distribution F. It then generates a new "effective threshold" $\tilde{x} = \bar{x} + y$. And then it applies the same μ_S from above to the effective threshold $\tilde{x}$.

We are now ready to state our next result.

Theorem 4. *For all $\epsilon > 0$ and $k < 1$ there is an $\bar{n}$ such that for any $n > \bar{n}$ and any mechanism $M = \langle R, \{\mu_S\}_{S \subseteq N} \rangle$ which satisfies k - Attraction when applied to a game Γ with $|N| = n$, there is a noisy threshold mechanism $\bar{M} = \langle \bar{R}, \{\bar{\mu}_S\}_{S \subseteq N} \rangle$ which is ϵ - dominant IR and satisfies $\bar{\mu}_N = \mu_N$.*

Proof. See appendix. □

The logic of this result is as follows. We assume that M has the property that for sufficiently large k, each player i strictly prefers to play a centralized strategy if at least a fraction k of other players use centralized strategies. Let $\bar{R} = (R \times A)$. We take F to be the uniform distribution with support on $[-\sqrt{n}, 0]$ and n to be large enough that $\frac{\sqrt{n}}{n} < 1 - k$, and the threshold to be any $\bar{x} \in (k + \frac{\sqrt{n}}{n}, 1)$

For any centralized strategy, the probability that a particular player i triggers the threshold is at most $\frac{1}{\sqrt{n}}$. Now consider any decentralized strategy a_i and the centralized strategy $\bar{r}_{a_i} = (r_{a_i}, a_i)$. We once again perform a cost-benefit analysis.

- For strategy profiles in which sufficiently few people sign away their decision rights such that the probability that i's threshold is met is 0, the two strategies are payoff equivalent.

- For strategy profiles in which sufficiently many people sign away their decision rights such that i's threshold is met with probability 1 regardless of i's participation, i strictly prefers the centralized strategy by assumption.

- For strategy profiles in which there is positive probability that i's participation is pivotal in triggering at least one other player's threshold, i's loss in this case is bounded below by $\frac{1}{\sqrt{n}}$.

Thus the cost of the centralized strategy relative to the decentralized strategy is at most $\frac{1}{\sqrt{n}}$ which vanishes as n becomes large. Note that an ordinary threshold mechanism would not have this property. For certain strategy profiles, agent i would be pivotal with probability 1 and thus his cost from participation would not be vanishing.

This result resembles insights from several literatures. Levine and Pesendorfer (1995) note that large market models where any agent can only observe aggregate actions with noise resemble continuum models in the limit, while large market models where agents can perfectly observe aggregate actions do not. There as well, the intuition comes from the fact that agents can be pivotal actors in the model without noise.

Also closely related is the literature on differential privacy, which explores the different ways in which adding noise to the output of algorithms makes it safe for participants to truthfully report inputs. For instance, Hsu et al. (2016) show that by calculating general equilibrium prices and then adding a small random perturbation in a large economy, it becomes an approximately dominant strategy for agents to report their true demand functions. And in a model closely related to this one, Kearns et al. (2015) show that in several classes of large market games in which equilibria can sensibly be perturbed, players can be induced to report their private information to a mechanism who then advises them to take actions that correspond to a slightly perturbed ex-post equilibrium. Their result differs from ours in that the perturbation in our model makes it safe to enter the centralized market regardless of what others do, while the perturbation in their model makes participation an approximate equilibrium without guaranteeing their desired equilibrium will be chosen.

We close this section with a negative result about ϵ - strong dominant individual rationality. While the above result pertains to all large games, random matching games are a natural sub class to study large market limits. Formally, for our final result we are interested in the following class of random matching games. Let $\Gamma := \langle N, A = \{A_i\}_{i\in N}, u = \{u_i\}_{i\in N}\rangle$ be a normal form game of complete information, with players N, actions A, and utilties u as defined in Section 2. Let $\Gamma^n := \langle n \times N, A, u\rangle$ be a game with n copies of each player in Γ, and where payoffs to each player are determined by choosing n uniformly random subsets of players such that each subset has a copy of each player and applying the utility functions u_i to the actions of the players in each subset. We are now ready to state our final result.

Theorem 5. *Consider any game Γ which does not have an equilibirum in dominant strategies. Then there exists an $\epsilon > 0$ such that for all n, no mechanism applied to Γ^n is ϵ - strong dominant individually rational.*

Proof. See appendix. □

The proof of the above result considers strategy profiles where player i chooses the centralized strategy r_i that supposedly ϵ - dominates his decentralized strategies, and the other players all choose decentralized strategies. In such a case, the maximal regret is clearly bounded from below by the worst case regret player i suffers in the purely decentralized game, and this is non vanishing in large markets for any game Γ where some player i does not have a dominant strategy.

7 Discussion

References

Itai Ashlagi, Yash Kanoria, and Jacob Leshno. 2016. Unbalanced Random Matching Markets: the Stark Effect of Competition. *Journal of Political Economy* (2016).

Itai Ashlagi, Dov Monderer, and Moshe Tennenholtz. 2007. Mediators in Position Auctions. *Proceedings of the 8th ACM Conference on Electronic Commerce* (2007).

Dirk Bergemann and Stephen Morris. 2003. Robust Mechanism Design. *Mimeo: Yale University* (2003).

Peter Coles, Yannai Gonczarowski, and Ran Shorrer. 2016. Strategic Behavior in Unbalanced Matching Markets. *Working Paper* (2016).

Gabrielle Demange, David Gale, and Marilda Sotomayor. 1987. A further note on the stable matching problem. *Discrete Applied Mathematics* 16, 3 (1987), 217–222.

Mehmet Ekmekci and Bumin Yenmez. 2014. Integrating Schools for Centralized Admissions. *Working Paper* (2014).

Francoise Forges. 2013. A Folk Theorem for Bayesian Games with Commitment. *Games and Economic Behavior* 78 (2013), 64–71.

Justin Hsu, Zhiyi Huang, Aaron Roth, and Steven Wu. 2016. Jointly Private Convex Programming. In *SODA*.

Nicole Immorlica and Mohammad Mahdian. 2005. Marriage, Honesty, and Stability. In *SODA*.

Mathew Jackson. 1991. Bayesian Implementation. *Econometrica* 59, 2 (March 1991), 461–477.

Mathew Jackson. 2001. A Crash Course in Implementation Theory. *Social Choice and Welfare* 18 (2001), 655–708.

Adam Kalai, Ehud Kalai, Ehud Lehrer, and Dov Samet. 2010. A Commitment Folk Theorem. *Games and Economic Behavior* 69, 1 (2010), 127–137.

Anirban Kar, Indrajit Ray, and Roberto Serrano. 2010. A Difficulty in Implementing Correlated Equilibrium Distributions. *Games and Economic Behavior* 69 (2010), 189–193.

Michael Kearns, Mallesh Pai, Ryan Rogers, Aaron Roth, and Jonathan Ullman. 2015. Robust Mediators in Large Games. *Working Paper* (2015).

Onur Kesten. 2012. On Two Kinds of Manipulation for School Choice Problems. *Economic Theory* 53, 3 (2012), 677–693.

Fuhito Kojima and Parag Pathak. 2009. Incentives and Stability in Large Two-Sided Matching Markets. *American Economic Review* 99, 3 (2009), 608–627.

Fuhito Kojima, Parag Pathak, and Alvin E. Roth. 2013. Matching with Couples: Stability and Incentives in Large Markets. *Quartery Journal of Economics* 128, 4 (2013), 1585–1632.

David K. Levine and Wolfgang Pesendorfer. 1995. When Are Agents Negligible? *American Economic Review* (1995), 1160–1170.

Eric Maskin. 1999. Nash Equilibrium and Welfare Optimality. *Review of Economic Studies* 66 (1999), 23–38.

Eric Maskin and Tomas Sjöstrom. 2002. *Handbook of Social Choice and Welfare*. Chapter 5, 237–288.

Dov Monderer and Moshe Tennenholtz. 2009. Strong Mediated Equilibrium. *Artificial Intelligence* 173, 1 (2009), 180–195.

Roger Myerson. 1991. *Game Theory: Analysis of Confllict*.

Muriel Niederle, Deborah D. Proctor, and Alvin E. Roth. 2006. What Will Be Needed for the New Gastroenterology Fellowship Match to Succeed? *Gastroenterology* 130, 1 (2006), 218–224.

Marion Oury and Olivier Tercieux. 2012. Continuous Implementation. *Econometrica* 80, 4 (2012), 1605–1637.

Scott Patterson. 2013. *Dark pools: The rise of the machine traders and the rigging of the US stock market*. Crown Pub.

Andrew Postlewaite and David Schmeidler. 1986. Implementation in Differential Information Economies. *Journal of Economic Theory* 39, 1 (1986), 14–33.

Reinhart Selten. 1975. Reexamination of the perfectness concept for equilibrium points in extensive games. *International journal of game theory* 4, 1 (1975), 25–55.

Leo K. Simon and Maxwell B. Stinchcombe. 1995. Equilibrium Refinement for Infinite Normal-Form Games. *Econometrica* (1995), 1421–1443.

Tayfun Sönmez. 1999. Can Pre-arranged Matches Be Avoided in Two-Sided Matching Markets? *Journal of Economic Theory* 86, 1 (1999), 148–156.

Robert Wilson. 1987. Game Theoretic Analysis of Trading Processes. In *Advances in Economic Theory*, T. Bewley (Ed.). Cambridge University Press.