Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Games and Economic Behavior 52 (2005) 34–53
GAMES and Economic Behavior
www.elsevier.com/locate/geb

Stability in dynamic matching markets

Ettore Damiano [a,*], Ricky Lam [b]

[a] *Department of Economics, University of Toronto, 150 St. George St., Toronto, ON, M5S 3G7, Canada*
[b] *McKinsey & Company, Chicago, IL, USA*

Received 3 September 2002

Available online 12 October 2004

Abstract

A dynamic two-sided matching market is considered. We examine two existing notions of stability—the core and recursive core—for this multi-period market and argue that they both have limitations. We define two new notions of stability and label them, *self-sustaining stability* and *strict self-sustaining stability*. Both concepts can be viewed as the recursive core with more stringent conditions for when deviating coalitions are effective. We show that these concepts overcome some of the weaknesses of the core and the recursive core. We also provide conditions for the existence of our concepts.

JEL classification: C71

Keywords: Repeated matchings; Credibility; Recursive core

1. Introduction

Many trading arrangements in the real world do not satisfy the assumptions of a Walrasian model of exchange. A special class of such arrangements is *two-sided matching markets*. These markets are characterized by two important features. First, participants belong to two disjoint sets; they cannot switch from one side of the market to the other no matter what the market condition. A second feature is the bilateral nature of exchange; the

* Corresponding author.
E-mail addresses: ettore.damiano@utoronto.ca (E. Damiano), ricky.lam@aya.yale.edu (R. Lam).

0899-8256/$ – see front matter
doi:10.1016/j.geb.2004.06.014

contrast is with centralized goods markets where the identity of one's trading partner is a matter of indifference. Examples of two-sided markets include many labor markets, as well as auction markets.

In this paper, we are concerned with a subclass of two-sided matching markets, namely those in which matches are *one-to-one*: each participant may be matched with at most one partner from the other side. Historically, the two sides of the market have been labeled the *males* and the *females*, and the model termed a *marriage market*. In many applications, a many-to-one relationship is more realistic but the issues we are concerned with can be discussed in the simpler class of one-to-one matching markets.

Any testable theory of a matching market must place some restrictions on the kind of outcomes that one expects to observe. An obvious restriction is that outcomes be "stable." In thinking about stability, we have in mind cooperative concepts similar to the *core*. An outcome that is not in the core is, by definition, susceptible to blocking by rational players. In general, there will of course be many other restrictions imposed by the incentives and rules associated with a particular trading institution. We consider the requirement of stability because it represents a minimal constraint in markets where participation is voluntary. In addition, the cooperative notion of stability only requires a very general description of the game and so is applicable to many markets, whereas issues of non-cooperative behavior depend crucially on the particular trading and information structure under consideration.

A large and very successful literature has considered stability in the special case of a static market (Roth and Sotomayor, 1990 provide an excellent summary). Existence of the core has been established and many of its interesting characteristics noted. In many markets, however, participants trade repeatedly. Indeed, investigating stability in a dynamic market is a necessary first step to studying more realistic models where agents have to learn about the values of matches as they occur over time.

When there are multiple periods, a *matching plan* specifies a partner for each participant, at each point in time. An obvious candidate notion of stability is the core over the set of feasible matching plans. The core, however, has a particularly unsatisfactory property in a dynamic game: it can admit matching plans that are not "time-consistent." Unless players can make binding agreements, elements in the core may be blocked at some later point in time.

The *recursive core*—defined by Becker and Chakrabarti (1995) in the context of a capital accumulation model—overcomes this. In our context, it requires that the continuation of a matching plan be in the core of the continuation market at all times.

Unfortunately, the recursive core is frequently empty. The reason is that it admits "incredible" deviations. In judging the stability of the grand coalition's matching plan—that is, the plan for all players—the recursive core requires that the plan be immune to blocking by coalitions at every point in time. However, no deviating coalition is subject to the same requirement.

The issue of what constitutes a "credible" deviation turns out to be very important in a dynamic matching market (in ways that are not apparent in a static market). Appropriately modifying the recursive core to deal with issues of credibility leads to a stability concept that is always non-empty.

We begin by imposing the requirement that blocking coalitions be "self-sustaining." They must choose matching plans in which no subset of the coalition can reach an agree-

ment to deviate from the deviation. The sub-coalitions have to satisfy the same requirement, and so on. This is the cooperative analogue to the concept of *coalition proofness* of Bernheim et al. (1987).[1]

We call our concept *self-sustaining stability*. By limiting the plans of deviating coalitions, we obtain a set of matching plans that is (at least weakly) larger than the recursive core. Even so, an example shows that there might be dynamic matching markets with no self-sustaining stable plan. Non-existence highlights a second aspect of credibility.

Self-sustaining stability implicitly stipulates that coalitions that deviate do so thereafter: members of the blocking coalition do not have to be immune to proposals to rejoin the grand market in some later period.[2] This amounts to a form of dynamic commitment for deviations.

In many situations, this degree of commitment is not possible. We introduce a second definition that imposes a stricter requirement for when a deviating plan is credible. It exists under general conditions. For want of a better term, we refer to it as *strict self-sustaining stability*. Under this new definition, a deviation must satisfy the conditions of self-sustaining stability, and in addition, it must be better—relative to the candidate stable plan—for each member of the deviating coalition, *at each future point in time*. A deviating coalition is now only credible if it can commit to remaining away from the grand market. This enlarges the set of stable plans for the grand coalition. Whether self-sustaining stability or its stricter version is appropriate will depend on the level of commitment available to the participants of the market.

The paper proceeds as follows. In the next section, we describe a dynamic matching market.[3] Section 3 presents the agent-form of the game and illustrates some of the issues that arise when defining stability in a dynamic matching market. In Section 4, we introduce the concept of self-sustaining stability. Sections 5 introduces the definition of strict self-sustaining stability. Section 6 discusses our concepts and relates them to other concepts in the literature. Finally, Section 7 concludes.

2. Dynamic matching markets

We denote the two disjoint, finite sides of the market, the *males* and the *females*, by $M = \{m_1, m_2, \ldots, m_{|M|}\}$ and $F = \{f_1, f_2, \ldots, f_{|F|}\}$, respectively. We will refer to the set of players, $M \cup F$, as the *grand coalition*. The market operates for T finite periods with players free to rematch at the end of each period.[4] The outcome each period is referred to as a *matching*.

[1] The self-sustaining idea can be applied to a static cooperative game. Ray (1989) showed, however, that this additional requirement does not alter the core. We show that it does matter in a dynamic game when it is imposed in conjunction with time-consistency.

[2] Of course, this criticism also applies to the core and the recursive core.

[3] A few papers in the matching literature deal with a notion of dynamics different from ours. For example, Roth and Vande Vate (1990) and Blum et al. (1997) are interested in decentralized, dynamic processes by which stable matchings can be reached. We are interested in stable matching *plans* in a multi-period market.

[4] A previous version of this paper, Damiano and Lam (2001), contains some results for the infinite-horizon case.

Definition 1. A *matching* is a one-to-one function μ satisfying the following:

$$\mu : (M \cup F) \to (M \cup F), \tag{1}$$

$$\mu(\mu(i)) = i, \tag{2}$$

$$\text{if } \mu(m) \neq m \in M, \quad \text{then } \mu(m) \in F, \tag{3}$$

$$\text{if } \mu(f) \neq f \in F, \quad \text{then } \mu(f) \in M. \tag{4}$$

$\mu(i) = i$ implies that individual i is unmatched; we will also refer to such a player as being *self-matched* or *single*. We denote the set of all possible matchings by $\mathcal{M}$. A *matching plan* is simply a matching for each period.

Definition 2. A *matching plan* is a function $\boldsymbol{\mu} : \mathbb{N}_T \to \mathcal{M}$, where $\mathbb{N}_T$ is the set of natural numbers $\{1, 2, \ldots, T\}$.

We focus on repeated matching markets. Each period, players have strict preferences over matches with members of the other side of the market. Let $\pi(\mu)$ in $\mathbb{R}^{|M \cup F|}$ be a vector of period payoffs for each player from matching with the partner specified under μ in $\mathcal{M}$. Throughout the paper, we assume that the outside option associated with being single is normalized to zero for all players. If any element of $\pi(\mu)$ is strictly negative, we say that μ is not *individually rational*.

Let $\beta \in [0, 1]$ be the discount factor and define the payoff function, $\boldsymbol{\pi}$, over matching plans, to be the sum of discounted period payoffs:[5,6]

$$\boldsymbol{\pi}(\boldsymbol{\mu}) = \sum_{t=1}^{T} \beta^{t-1} \pi\big(\boldsymbol{\mu}(t)\big). \tag{5}$$

For a subset of players S, we use a subscript to denote the above objects restricted to players in S. For example, $\mathcal{M}_S$ is the set of possible matchings among players in S.

3. Agent-form representation

The notions of stability that we will introduce are most tractable when stated in a transformed game.[7] We refer to this transformed game as the *agent-form* of the dynamic market. It is obtained by assuming that each player is represented by a sequence of agents—one for each time period of the market. We refer to a participant in the multi-period market as a *player* and reserve the term *agent* for some player at a particular time.

The notation for the agent-form is involved, but an example will follow. Define $i(t) \equiv i$ for all $i \in M \cup F$ and all $t = 1, \ldots, T$. For any $S \subseteq M \cup F$, let $S(t) = \{i(t) \mid i \in S\}$. The

[5] In our notation, μ is a particular element of $\mathcal{M}$; while a bold $\boldsymbol{\mu}$ is a mapping from time onto $\mathcal{M}$. Similarly, π is a function over the set of matchings; while bold $\boldsymbol{\pi}$ is a function over the set of matching plans.

[6] In a static market, ordinal preferences suffice for many results. We assign values to period matchings to allow payoffs to be aggregated over time.

[7] Although we will continue to discuss stability in the context of a two-sided matching market, the definitions to follow can also be applied to any dynamic cooperative game.

set of agents in the agent-form is $\boldsymbol{M} \cup \boldsymbol{F}$, where $\boldsymbol{M} = M(1) \cup M(2) \cup \cdots \cup M(T)$ and $\boldsymbol{F} = F(1) \cup F(2) \cup \cdots \cup F(T)$.

Previously, we defined a matching plan to be a function from time to a matching outcome for the players. We abuse notation and now define a matching plan $\boldsymbol{\mu}$ to be a one-to-one function satisfying the following:

$$\boldsymbol{\mu} : (\boldsymbol{M} \cup \boldsymbol{F}) \to (\boldsymbol{M} \cup \boldsymbol{F}), \tag{6}$$

$$\boldsymbol{\mu}\big(\boldsymbol{\mu}\big(i(t)\big)\big) = i(t), \tag{7}$$

$$\text{if } \boldsymbol{\mu}\big(m(t)\big) \neq m(t) \in M(t), \quad \text{then } \boldsymbol{\mu}\big(m(t)\big) \in F(t), \tag{8}$$

$$\text{if } \boldsymbol{\mu}\big(f(t)\big) \neq f(t) \in F(t), \quad \text{then } \boldsymbol{\mu}\big(f(t)\big) \in M(t). \tag{9}$$

This is the usual definition of matching in a static market—applied to agents rather than players—with the additional restriction that only couples at the same time can match.

Let $\mathcal{M}(t)$ denote the set of possible matchings between $M(t)$ and $F(t)$. The set of feasible matching plans is: $\prod_{t=1}^{T} \mathcal{M}(t) = \mathcal{M}(1) \times \mathcal{M}(2) \times \cdots \times \mathcal{M}(T)$. We denote the projection of $\boldsymbol{\mu}$ in $\prod_{t=1}^{T} \mathcal{M}(t)$ onto $\mathcal{M}(\tau)$ by $\boldsymbol{\mu}(\tau)$.[8]

As before, the vector $\pi(\boldsymbol{\mu}(t))$ in $\mathbb{R}^{|M \cup F|}$ specifies the period payoffs at time t. We abuse notation by using $\boldsymbol{\pi}$ to denote the payoff function in the agent-form game. Previously, we used $\boldsymbol{\pi}$ to denote a vector in $\mathbb{R}^{|M \cup F|}$, representing the present discounted payoffs for the players. In the agent-form, we treat each player as a sequence of agents so $\boldsymbol{\pi}$ is now a vector in $\mathbb{R}^{|\boldsymbol{M} \cup \boldsymbol{F}|}$, representing the payoffs to agents. It is defined as follows:

$$\boldsymbol{\pi}(\boldsymbol{\mu})^{M(t) \cup F(t)} = \sum_{\tau \geqslant t} \beta^{\tau - t} \pi\big(\boldsymbol{\mu}(\tau)\big), \tag{10}$$

where x^S is the projection of a vector $x \in \mathbb{R}^{|\boldsymbol{M} \cup \boldsymbol{F}|}$ onto the subspace $\mathbb{R}^{|S|}$.

For a subset of agents S from $\boldsymbol{M} \cup \boldsymbol{F}$, $\boldsymbol{\pi}_S$ and $\boldsymbol{\mu}_S$ are defined as in Eqs. (6) to (10), restricted to agents in S.

To illustrate these new definitions, consider the following 2×2 matching market. The matrix below denotes the period payoffs. The (i, j)th cell contains two numbers, being the payoffs to male m_i and female f_j, respectively, from a match with each other. Assume that this one-shot market is repeated twice with no discounting ($\beta = 1$):

	f_1	f_2
m_1	$5, -1$	$-1, 5$
m_2	$-1, 5$	$5, -1$

(11)

Since players receive a payoff of zero when they remain single, the unique element of the Gale–Shapley set (in the one-shot market) specifies that players remain self-matched: $\{\mu^{\text{single}}(i) = i,\ i \in (M \cup F)\}$, which yields $\pi(\mu^{\text{single}}) = (0, 0, 0, 0)$.

In the agent-form of the two-period repeated game, the set of male agents is $\boldsymbol{M} = \{m_1(1), m_2(1), m_1(2), m_2(2)\}$. Each male player, m_i, is represented by two agents, $m_i(1)$

[8] This notation is consistent with our previous view of a matching plan as a mapping from time to the set of possible period matchings.

and $m_i(2)$. Similarly, $\boldsymbol{F} = \{f_1(1), f_2(1), f_1(2), f_2(2)\}$. The payoffs in the agent-form game can be partially represented by the following matrix:

	$f_1(1)$	$f_2(1)$	$f_1(2)$	$f_2(2)$
$m_1(1)$			$\times$	$\times$
$m_2(1)$			$\times$	$\times$
$m_1(2)$	$\times$	$\times$	$5, -1$	$-1, 5$
$m_2(2)$	$\times$	$\times$	$-1, 5$	$5, -1$

(12)

For the agents in period 2, their payoffs are provided by the stage payoffs. "$\times$" is used to denote infeasible matches: matching plans do not allow agents from different periods to match. What about the payoff to a period 1 agent, say $m_1(1)$? It consists of two components: the utility he obtains from matching with some period 1 agent, and the utility he obtains from the match of his future self. Consider, for example, the matching plan $\boldsymbol{\mu}$ given by $\{\boldsymbol{\mu}(m_1(1)) = f_1(1), \boldsymbol{\mu}(m_2(1)) = f_2(1), \boldsymbol{\mu}(m_1(2)) = f_2(2), \boldsymbol{\mu}(m_2(2)) = f_1(2)\}$. The utility of agent $m_1(1)$ is the sum of 5—being the benefit from matching with $f_2(1)$—and $\beta \times (-1)$—being the discounted utility that agent $m_1(1)$ obtains from $m_1(2)$'s match with $f_2(2)$. Discounting in the original dynamic market has been transformed into an *externality* between agents in the agent-form.[9] These externalities flow from later times to earlier ones.

3.1. Existing notions of stability in agent-form

In the context of a single-period market, Gale and Shapley (1962) consider *stable matchings*. These are matchings that are individually rational, and that cannot be *blocked* by a male and female pair. A pair players (m, f) blocks a matching μ, if each of them prefers the other to her partner under μ. The notion of blocking can be extended, naturally, to coalitions of more than two players. The *core* of a matchings market is the set of individually rational matchings that are not blocked by any coalition of agents. It is well known that the set of stable matchings is equivalent to the core—no matching can be blocked by a larger coalition if it is not blocked by a male and female pair—and that this set is non-empty. To avoid confusion with the multi-period market, we refer to the core in the single-period market as the *Gale–Shapley set*.

An obvious notion of stability in our dynamic market is the core over the set of matching plans. In contrast to a static market, coalitions of more than a male and female pair do matter in a dynamic market. They provide the possibility of altering partners and so can achieve payoffs that a couple cannot. Next, we formally define the core over the set of matching plans in terms of blocking conditions in the agent-form. Later, we will write alternative definitions of stability in a dynamic market as modifications of the core of the agent-form game.

[9] Because the payoffs to agents in period 1 depend on the matches of agents in period 2, the payoffs cannot be completely represented in the matrix of (12).

The following operator, $O : 2^{\boldsymbol{M}\cup\boldsymbol{F}} \to 2^{\boldsymbol{M}\cup\boldsymbol{F}}$, will prove useful in the definitions that follow:

$$O(S) = \left\{ i(t) \in \boldsymbol{M} \cup \boldsymbol{F} \;\middle|\; i(\tau) \in S \text{ and } \tau \leqslant t \right\}. \tag{13}$$

For any set of agents S, $O(S)$ extends the set to include the future selves of agents in S.

Definition 3. A matching plan $\boldsymbol{\mu} \in \prod_t \mathcal{M}(t)$ is in the *core* of the dynamic market if there does not exist a coalition of agents $S \subseteq M(1) \cup F(1)$ and a feasible matching plan for $O(S)$, $\boldsymbol{\mu}_{O(S)}$, such that

$$\boldsymbol{\pi}_{O(S)}(\boldsymbol{\mu}_{O(S)})^S > \boldsymbol{\pi}(\boldsymbol{\mu})^S. \tag{14}$$

In the above definition, $(x_1, \ldots, x_n) > (y_1, \ldots, y_n)$ denotes $x_i \geqslant y_i$ for all $i = 1, \ldots, n$ and $x_i > y_i$ for some i. Blocking from the core must involve a collection of period 1 agents, together with their successors. In considering whether this coalition is effective, only the payoffs of agents in period 1 matter.[10]

For the two-period example of (11), it is easy to see that a core matching plan, $\boldsymbol{\mu}^{\text{core}}$, involves implementing the male-preferred matching $\{\mu(m_1) = f_1,\ \mu(m_2) = f_2\}$ in one period and the female-preferred matching $\{\mu(m_1) = f_2,\ \mu(m_2) = f_1\}$ in the other. There are two such plans. At the beginning of the game, these plans achieve the payoffs $\boldsymbol{\pi}(\boldsymbol{\mu}^{\text{core}}) = (4, 4, 4, 4)$, and cannot be blocked by any coalition.

Unfortunately, one would not expect to observe such an outcome in any play of the game if matching plans are not binding. In the second period, one side of the market always has an incentive to withdraw participation and to renege on the plan agreed upon in period 1.[11]

A concept that does impose stability at every point in time is the *recursive core* of Becker and Chakrabarti (1995). This concept is closely related to the *sequential core* of Gale (1982). Both are motivated by the lack of trust in a general equilibrium model. As we did for the core, we can define the recursive core of a dynamic matching market in terms of blocking conditions in the agent-form.

Definition 4. A matching plan $\boldsymbol{\mu} \in \prod_t \mathcal{M}(t)$ is in the *recursive core* of the dynamic market if at all times $t \leqslant T$, there does not exist a coalition of agents, $S \subseteq M(t) \cup F(t)$, and a feasible matching plan for $O(S)$, $\boldsymbol{\mu}_{O(S)}$, such that

$$\boldsymbol{\pi}_{O(S)}(\boldsymbol{\mu}_{O(S)})^S > \boldsymbol{\pi}(\boldsymbol{\mu})^S. \tag{15}$$

In this definition, a deviating coalition is made up of agents of the same period, $S \subseteq M(t) \cup F(t)$, together with their successors. It differs from the definition of the core in that

[10] Notice that this definition is not simply the usual definition of the "core" applied to the agent-form game. Only certain coalitional deviations are allowed—namely $O(S)$ where $S \subseteq M(1) \cup F(1)$—and only the payoffs of certain agents in such deviations matter—namely those in S.

[11] It may be thought that randomization could provide a solution to this problem. Players could agree to implement matchings based on a publicly observable flip of a coin each period. This would give each player an expected payoff equal to the payoff from $\boldsymbol{\mu}^{\text{core}}$. However, this requires that the outcome specified by the coin flip be enforceable, which is counter to the spirit of the paper. We want to place some restrictions on the matching plans which may be observed when enforceability is assumed not to be possible.

it allows $t > 1$. Like that definition, however, only the payoffs of agents in S matter. The condition of dynamic consistency in a multi-period game has been expressed as a static condition in the agent-form.

In a dynamic matching market, the recursive core demands that a matching plan be in the core at the beginning of the market, and that its continuation be in the core of the continuation market at all points in time. It is clear that the recursive core is a refinement, or a subset, of the core. In the two-period example of matrix (11), both core matching plans specify that players be matched in the second period. This is inconsistent with stability in the last period, which requires all players to remain single. The recursive core is thus empty in this example.

Once again, this result seems unsatisfactory. Intuitively, the matching plan which specifies that participants remain unmatched in both periods appears to be robust to "blocking by rational players." Yet, this matching plan is not in the recursive core. Denote this matching plan by $\boldsymbol{\mu}^{\text{single}}$. In period 2, the continuation of $\boldsymbol{\mu}^{\text{single}}$ is consistent with the recursive core since it specifies a Gale–Shapley matching in the final period. However, in period 1, $\boldsymbol{\mu}^{\text{single}}$ is blocked by the grand coalition playing a core matching plans, which we have already argued does not satisfy the requirement of time-consistency.

This highlights an inconsistency associated with the recursive core: coalitions are allowed too much freedom in choosing the deviating matching plan. In judging the original matching plan, the recursive core requires that the plan be immune to blocking by coalitions. However, no deviating group of players (including the grand coalition) is subject to the same requirement.

4. Self-sustaining stability

The example and discussion in the previous section suggest that, for blocking coalitions to be credible, they should themselves be stable against further deviations. In fact, to be consistent, we should not only demand that deviations be credible, but that any deviation from a deviation also be credible, and so on. We refer to this sequence of requirements as *self-sustainability*.

The non-cooperative concept of *coalition-proof Nash equilibria* due to Bernheim et al. (1987) is motivated by such considerations. Ray (1989) defines the cooperative analogue to coalition-proof Nash equilibria and label it the *modified core*. A matching plan is in the modified core if there does not exist any "credible" blocking coalition. Blocking coalitions are "credible" if they choose matching plans in which no subset of the coalition can reach an agreement to deviate from the deviation. The sub-coalitions have to satisfy the same requirement, and so on. The concept is formally defined inductively, beginning with the singleton coalition. Ray shows, somewhat surprisingly, that self-sustainability in a static game has no impact: the modified core is equivalent to the core.[12] We show that Ray's result no longer holds in a dynamic market when time consistency is also imposed. In a dynamic

[12] This is not true in the non-cooperative setting, where the set of coalition-proof Nash equilibria—the non-cooperative analogue of the modified core—in general differs from the set of strong Nash equilibria—the non-cooperative analogue of the core.

market, time-consistency should be a necessary condition for coalitional credibility, and not just a requirement on the grand coalition's plan. Self-sustainability, when interacted with time-consistency, does in general reduce the number of credible deviations.

Next we formally define *self-sustaining stability* (S^3). It is essentially the cooperative analogue to the concept of *perfect coalition-proofness* for extensive form games of Bernheim et al. (1987).[13] Recall that the modified core is obtained by imposing the idea of the core, not only on the grand coalition, but on deviating coalitions, as well as on deviations from deviations, and so on. In this sense, it is the "self-sustaining core." S^3 is obtained by imposing the idea of the recursive core, not just on the grand coalition, but on deviating coalitions, and on deviations from deviations, and so on. It can be viewed as the "self-sustaining recursive core." The definition below is recursive, both through the size of the coalition, as well as through time.

Definition 5.

(1) For coalitions of agents of the form $O(\{i\})$, where $i \in \boldsymbol{M} \cup \boldsymbol{F}$, the plan $\boldsymbol{\mu}(j) = j$, for all $j \in O(\{i\})$, satisfies *self-sustaining stability* with respect to i. For any coalition of agents from the final period, $S \subset (M(T) \cup F(T))$, $\boldsymbol{\mu}_S$ satisfies *self-sustaining stability* if it is in the Gale–Shapley set.[14]

(2) Consider a coalition of the form $O(S)$, where $S \subset (M(t) \cup F(t))$ for some t. Assume that self-sustaining stability has been defined for all coalitions C, where $C \subset S$ or $C \subseteq (M(\tau) \cup F(\tau))$ for some $\tau > t$. A matching plan $\boldsymbol{\mu}_{O(S)}$ is *self-sustaining-stable* with respect to S if:

   (a) There does not exist a coalition C with $C \subset S$ *or* $C \subseteq (M(\tau) \cup F(\tau)) \cap O(S)$, with a feasible matching plan $\tilde{\boldsymbol{\mu}}_{O(C)}$, which satisfies self-sustaining stability for C, such that

$$\boldsymbol{\pi}_{O(C)}(\tilde{\boldsymbol{\mu}}_{O(C)})^C > \boldsymbol{\pi}_{O(S)}(\boldsymbol{\mu}_{O(S)})^C. \tag{16}$$

   (b) There is no other matching $\boldsymbol{\mu}'_{O(S)}$ satisfying (a) such that

$$\boldsymbol{\pi}_{O(S)}\big(\boldsymbol{\mu}'_{O(S)}\big)^S > \boldsymbol{\pi}_{O(S)}(\boldsymbol{\mu}_{O(S)})^S. \tag{17}$$

   Like the recursive core, S^3 allows deviating coalitions of the form $O(S)$, where $S \subseteq (M(t) \cup F(t))$ for some t. Also, in considering whether the deviation is actually effective, only the payoffs of the agents at the time of the deviation, S, are relevant. The difference is that S^3 requires deviating plans to be self-sustaining, whereas the recursive core does not.

[13] As an aside, if the S^3 is the counterpart of perfect coalition-proof equilibria, we can think of the core as the counterpart of strong Nash equilibria. Similarly, the recursive core can be thought of as the analogue to Rubinstein's (1990) concept of *strong perfect equilibrium*.

[14] Strictly speaking, this statement is an immediate consequence of the equivalence between core and modified core—self-sustainability does not "bite" in a static market (Ray, 1989)—rather than a *definition*.

In a finite-horizon market, we can construct a matching plan in the S^3 set via backward induction.[15] The recursion begins with singleton coalitions in period T. It proceeds through the size of the coalition until the grand coalition is reached, and then considers the two-period market beginning at time $T-1$, and so on.

Applying this concept to the two-period example of (11) is relatively simple. One can verify that the unique plan in S^3 is $\boldsymbol{\mu}^{\text{single}}$ which specifies that all players remain unmatched in both periods. This is what we had claimed to be the intuitive outcome of that market. The plan survives because the deviating plan which blocks $\boldsymbol{\mu}^{\text{single}}$ from the recursive core is not admitted under S^3.

4.1. Existence of S^3

The following lemma, needed to prove the subsequent proposition, is a strengthening of the *strong stability property* for a special class of preferences. The strong stability property (see Theorem 3.4 in Roth and Sotomayor, 1990) states that unstable matchings either fail individual rationality, or are blocked by a pair of agents that would be better off under some stable matching. The lemma below strengthens the claim establishing that the blocking pair will be indeed matched under that stable matching.

Lemma 1. *In a static market, suppose that for all $S \subseteq M \cup F$, there exists a unique Gale–Shapley matching among players in S, and let $\mu^{\text{G–S}}$ be the Gale–Shapley matching for $M \cup F$. Then, for any individually rational, non-Gale–Shapley matching μ among $M \cup F$, there is a player $i \in M \cup F$ such that $\{i, \mu^{\text{G–S}}(i)\}$ blocks μ.*

Proof. The claim is obvious when either M or F is a singleton. Suppose the statement is true if the market is restricted to any coalition $S \subset M \cup F$. We prove the inductive step that the statement is true when the market consists of participants $M \cup F$.

Let μ be an individually rational, non-Gale–Shapley matching for $M \cup F$. Let $M_{\succ}$, $M_{\prec}$, and $M_{\sim}$ denote the sets of males that strictly prefer μ to $\mu^{\text{G–S}}$, strictly prefer $\mu^{\text{G–S}}$ to μ, and are indifferent between the two matchings, respectively. $F_{\succ}$, $F_{\prec}$, and $F_{\sim}$ are defined analogously. Since preferences are strict, for $i \in M_{\sim} \cup F_{\sim}$, $\mu(i) = \mu^{\text{G–S}}(i)$. Thus, $M_{\succ} \cup M_{\prec} \cup F_{\succ} \cup F_{\prec}$ are matched amongst themselves under both μ and $\mu^{\text{G–S}}$. Moreover, because $\mu^{\text{G–S}}$ is a Gale–Shapley matching, $\mu(i) \in F_{\prec}$ for all $i \in M_{\succ}$, and $\mu(i) \in M_{\prec}$ for all $i \in F_{\succ}$. If $\mu^{\text{G–S}}(i) \in M_{\prec} \cup F_{\prec}$ for any $i \in M_{\prec} \cup F_{\prec}$, then both i and $\mu^{\text{G–S}}(i)$ strictly prefer $\mu^{\text{G–S}}$ to μ and we are done. Suppose otherwise that $\mu^{\text{G–S}}(i) \in F_{\succ}$ for $i \in M_{\prec}$, and $\mu^{\text{G–S}}(i) \in M_{\succ}$ for $i \in F_{\prec}$. Consider a matching market restricted to $M_{\prec} \cup F_{\succ}$. Since μ is individually rational and $\mu^{\text{G–S}}$ is the only stable matching, $M_{\prec} \cup F_{\succ} \neq \emptyset$ by the strong stability property. Both $\mu^{\text{G–S}}$ and μ define a matching for this smaller market. Moreover, $\mu^{\text{G–S}}$ is also a Gale–Shapley matching for this smaller market. Since there is a unique Gale–Shapley matching by assumption, μ is not a Gale–Shapley matching in the smaller market. Thus, there is a player $i \in M_{\prec} \cup F_{\succ}$ such that $\{i, \mu^{\text{G–S}}(i)\}$ blocks μ. □

[15] Solving for S^3 in an infinite-horizon game is more difficult. In Damiano and Lam (2001), we employ the idea of dynamic programming to characterize the S^3 set when the horizon is infinite.

The following proposition provides conditions for the existence of S^3.

Proposition 2. *There exists a matching plan which satisfies* S^3 *if at least one of the following conditions hold*:

(a) *The discount factor* β *is sufficiently close to zero.*
(b) *There are less than, or equal to, two players on each side of the market.*
(c) *All feasible matchings are individually rational and for all subsets of players,* $S \subseteq M \cup F$*, there is a unique Gale–Shapley matching among the players.*
(d) *All players remaining single is a Gale–Shapley matching.*

Proof. (a) This is an obvious consequence of strict preferences and the fact that the Gale–Shapley set is non-empty.

(b) The claim is obvious when there are less than two players on either side of the market. The unique plan in S^3 consists of repeating the unique Gale–Shapley matching in every period. Consider a market with two players on both sides of the market. We show by induction that any plan which consists of repeating the same Gale–Shapley matching each period is in S^3. When $T = 1$ any stable matching is in S^3. For $\widetilde{T} < T$, assume that any sequence of $\widetilde{T}$ identical Gale–Shapley matchings is in S^3 for the market with $\widetilde{T}$ periods. Take $\boldsymbol{\mu}^{\text{G–S}}$ to be a sequence of T identical Gale–Shapley matchings. $\boldsymbol{\mu}^{\text{G–S}}$ can only be blocked by a coalition $O(S)$, where $S \subseteq M(1) \cup F(1)$. If S has strictly less than four agents, there is a unique matching plan $\boldsymbol{\mu}_{O(S)}$ which satisfies S^3 with respect to S, and $\boldsymbol{\mu}_{O(S)}(1) = \cdots = \boldsymbol{\mu}_{O(S)}(T)$. Thus, $O(S)$ cannot block $\boldsymbol{\mu}^{\text{G–S}}$ if $\boldsymbol{\mu}^{\text{G–S}}(1)$ is a Gale–Shapley matching. It remains to show that there cannot be a different matching plan in S^3 that Pareto dominates (with respect to $M(1) \cup F(1)$) the proposed plan. If the unique Gale–Shapley matching is for all agents to stay single, a backward induction argument shows that no agent can ever be matched in a S^3 plan. If in a Gale–Shapley matching some agent is not single, then at least one agent must be receiving his/her maximal payoff. Thus, $\boldsymbol{\mu}^{\text{G–S}}$ cannot be Pareto dominated.

(c) We show that the matching plan that specifies the unique Gale–Shapley matching $\mu^{\text{G–S}}$ in each period is the only element of the S^3 set. Notice that for markets with just one male or one female, this statement is trivially true. Now, assume that for all $S \subset M \cup F$, the unique S^3 plan for S specifies that the Gale–Shapley matching for S, $\mu_S^{\text{G–S}}$, be played every period. We need to show two things for the market with players $M \cup F$:

(i) any plan that specifies a matching $\mu \neq \mu^{\text{G–S}}$ is not in S^3;
(ii) the plan $\boldsymbol{\mu}^{\text{G–S}}(t) = \mu^{\text{G–S}}$ for all t is in S^3.

For (i): suppose $\mu \neq \mu^{\text{G–S}}$ is the last non-Gale–Shapley matching played in some matching plan $\boldsymbol{\mu}$ in S^3, and that $\bar{t}$ is the last period in which μ is played. By Lemma 1, there is a player i such that $\{i, \mu^{\text{G–S}}(i)\}$ blocks μ. Then, at time $\bar{t}$, $O(\{i, \mu^{\text{G–S}}(i)\})$ blocks $\boldsymbol{\mu}$, a contradiction.

For (ii): for any set of players $S \subset M \cup F$, the unique S^3 plan specifies the Gale–Shapley matching $\mu_S^{\text{G–S}}$ in all periods. Thus, S blocks $\boldsymbol{\mu}^{\text{G–S}}$ via an admissible plan only if

$\pi_S(\mu_S^{\text{G–S}}) > \boldsymbol{\pi}(\boldsymbol{\mu}^{\text{G–S}})^S$. This is not possible because by assumption $\mu^{\text{G–S}}$ is a Gale–Shapley matching.

(d) By backward induction, the plan which specifies that all players remain single, at all points in time, is in the S^3 set. □

If players are sufficiently impatient (condition (a)), the dynamic game effectively becomes a sequence of static markets. Non-emptiness of S^3 follows from the non-emptiness of the Gale–Shapley set. With few players (condition (b)), the number of possible deviations are sufficiently limited that any plan that repeats the same Gale–Shapley matching is in S^3.

These conditions are restrictive and it is not difficult to construct an example in which none of the conditions are satisfied, and in fact no matching plan satisfies S^3. Such an example follows. It will provide some intuition for why condition (c) implies existence. More importantly, it will also serve to motivate our definition of *strict self-sustaining stability*. The following stage-game is repeated twice, with no discounting ($\beta = 1$):[16]

	f_1	f_2	f_3	f_4
m_1	1, 1	1, 1	3, 2	2, 1
m_2	1, 1	1, 1	2, 1	3, 2
m_3	2, 3	1, 1	5, 1	1, 5
m_4	1, 2	2, 3	1, 5	5, 1

(18)

This stage game has a unique Gale–Shapley matching: $\{\mu^{\text{G–S}}(m_1) = f_3,\ \mu^{\text{G–S}}(m_2) = f_4,\ \mu^{\text{G–S}}(m_3) = f_1,\ \mu^{\text{G–S}}(m_4) = f_2\}$. Any candidate for inclusion in S^3 must specify this matching among agents in $M(2)$ and $F(2)$. Consider a matching plan that specifies the Gale–Shapley matching in both periods. Denote this plan by $\boldsymbol{\mu}^{\text{G–S}}$. It is blocked by the coalition of 8 agents $O(S)$, where $S = \{m_3(1), m_4(1), f_3(1), f_4(1)\}$, playing the following S^3 plan: $\{\boldsymbol{\mu}^{\text{switch}}_{O(S)}(m_i(1)) = f_i(1),\ i = 3, 4;\ \boldsymbol{\mu}^{\text{switch}}_{O(S)}(m_i(2)) = f_j(2),\ i, j \in \{3, 4\}$ and $i \neq j\}$. From the perspective of the dynamic market, the 4 players $\{m_3, m_4, f_3, f_4\}$ carry out one matching in period 1, and switch to another in period 2. This plan gives agents in S: $\boldsymbol{\pi}_{O(S)}(\boldsymbol{\mu}^{\text{switch}}_{O(S)})^S = (6, 6, 6, 6) > (4, 4, 4, 4) = \boldsymbol{\pi}(\boldsymbol{\mu}^{\text{G–S}})^S$. Other candidates for S^3 can be similarly eliminated.

Because the candidate S^3 plan specifies the same Gale–Shapley matching in both periods, no subcoalition can do better by using only one matching. $O(S)$ blocks the proposed plan by playing two different matchings—both of which are in the Gale–Shapley set for S. Condition (c) implies existence because it rules out this possibility.

Under the assumptions of Proposition 2, there exists a sequence of Gale–Shapley matchings which satisfies S^3. However, in general, a matching plan might belong to S^3 even if

[16] This example does not satisfy the assumption of strict preferences over period matches. It can be easily modified to satisfy strict preferences without changing any of the conclusions. We do not do so in order to simplify the presentation of the example.

it is not a sequence of Gale–Shapley matchings. We illustrate this claim with an example. Consider the following stage game repeated twice with no discounting:

	f_1	f_2	f_3
m_1	$1,5$	$5,1$	$2,2$
m_2	$-1,6$	$1,5$	$5,1$
m_3	$5,1$	$2,2$	$1,5$

(19)

The stage game has three Gale–Shapley matchings: $\{\tilde{\mu}^{\text{G–S}}(m_1) = f_1, \tilde{\mu}^{\text{G–S}}(m_2) = f_2, \tilde{\mu}^{\text{G–S}}(m_3) = f_3\}$, $\{\hat{\mu}^{\text{G–S}}(m_1) = f_2, \hat{\mu}^{\text{G–S}}(m_2) = f_3, \hat{\mu}^{\text{G–S}}(m_3) = f_1\}$, and $\{\dot{\mu}^{\text{G–S}}(m_1) = f_3, \dot{\mu}^{\text{G–S}}(m_2) = f_2, \dot{\mu}^{\text{G–S}}(m_3) = f_1\}$. Consider a fourth matching $\{\overline{\mu}(m_1) = f_1, \overline{\mu}(m_2) = f_3, \overline{\mu}(m_3) = f_2\}$ which is not Gale–Shapley because it is blocked by m_1 and f_3. The matching plan in which $\overline{\mu}$ is implemented in the first period followed by $\tilde{\mu}^{\text{G–S}}$ in the second period, yields a payoff vector $(2, 6, 3, 10, 7, 6)$ to $(m_1(1), m_2(1), m_3(1), f_1(1), f_2(1), f_3(1))$ and, it can be verified, satisfies S^3. Notice that m_1 and f_3 do not block the proposed matching plan. The gain to f_3 from matching to m_1 rather than m_2 in the first period, is smaller than her loss from matching to m_2 instead of m_3 in the second period.

5. Strict self-sustaining stability

The non-existence example of the previous section suggests a criticism of S^3: when a deviation occurs, only members of the deviating coalition may contemplate deviations from the deviation. Members of the deviating coalition are prevented from forming a pact to deviate further with someone not included in the coalition.[17]

In the example, the candidate plan $\boldsymbol{\mu}^{\text{G–S}}$ is blocked by a deviating plan which satisfies S^3 but may nevertheless be "incredible." At time $t = 1$, it is certainly true that the plan $\boldsymbol{\mu}^{\text{switch}}_{O(S)}$ dominates remaining in the grand market; that is, agents in S benefit from the deviation. However, at time $t = 2$, agents in $O(S) - S$ receive a payoff vector of $(1, 1, 5, 5)$. The payoffs are ordered: $m_3(2)$, $m_4(2)$, $f_3(2)$, $f_4(2)$. From the perspective of agents $\{m_3(2), m_4(2)\}$, they would have done better under the candidate plan $\boldsymbol{\mu}^{\text{G–S}}$: $\boldsymbol{\pi}_{O(S)}(\boldsymbol{\mu}^{\text{switch}}_{O(S)})^{\{m_3(2),m_4(2)\}} = (1, 1) < (2, 2) = \boldsymbol{\pi}(\boldsymbol{\mu}^{\text{G–S}})^{\{m_3(2),m_4(2)\}}$. What would stop agents $\{m_3(2), m_4(2)\}$ from reneging on the deviation agreed to by $\{m_3(1), m_4(1)\}$, and trying to return to the grand market in period $t = 2$? In the definition of S^3, blocking coalitions take the form of $O(S)$ but only the payoffs to agents in S are relevant. In terms of the dynamic market, we are allowing deviating players to commit to match amongst themselves after the deviation. When this commitment is not possible, certain deviations allowed under S^3 are not credible.

Our concept of *strict self-sustaining stability* (S^4) imposes a more stringent condition for when deviations are credible. A deviating coalition must specify a plan that satisfies the conditions of self-sustaining stability, and in addition, this plan must be better—relative to the candidate stable plan—for every agent in the coalition (not just those at the time of the deviation). In the terminology of the dynamic game, the plan has to be better for all

[17] The same criticism applies to coalition-proof Nash equilibria and to the modified core.

players of the deviating coalition *at all points in time*. Credible deviations under S^4 have to account for the possibility that players may return to the grand coalition. S^4 ensure that they have no incentive to do so, assuming that they return to the original plan.

Definition 6. A matching plan $\boldsymbol{\mu} \in \prod_t \mathcal{M}(t)$ satisfies *strict self-sustaining stability*, if there does not exist a coalition of agents $S \subseteq M(t) \cup F(t)$ together with a feasible matching plan $\boldsymbol{\mu}'_{O(S)}$ for $O(S)$, such that

$$\boldsymbol{\pi}_{O(S)}\big(\boldsymbol{\mu}'_{O(S)}\big) > \boldsymbol{\pi}(\boldsymbol{\mu})^{O(S)}, \tag{20}$$

for any $t = 1, 2, \ldots, T$.

There are two apparent differences between S^4 and S^3. The first can be seen in a comparison of the superscripts in Eq. (16) with those in (20). For a coalition $O(S)$—where $S \subseteq M(t) \cup F(t)$—to block in S^3, only agents in S have to be better off. For a coalition $O(S)$ to block in S^4, all of the agents in $O(S)$ have to benefit; the agents in S must be in agreement with all their future selves.

The second difference is that the definition of S^4 is not recursive; deviating coalitions are not required to propose a "stable" outcome. The following proposition shows that this second difference is only apparent. This proposition simplifies the use of S^4: one does not have to worry about self-sustainability if the stricter condition on deviating coalitions is imposed. Intuitively, the definition of "blocking" in S^4 is the standard one (with respect to the agent-form): all agents in the deviating coalition must do better. Because of this, we can use Ray's (1989) argument for the equivalence between the modified core and the core.

Proposition 3. *Suppose that the matching plan $\boldsymbol{\mu}$ is not in S^4. Then there is some subcoalition $S \subseteq M(t) \cup F(t)$ for some time $t \leqslant T$, and a matching $\boldsymbol{\mu}_{O(S)}$ which satisfies S^4 with respect to the agent-form game for $O(S)$, such that*

$$\boldsymbol{\pi}_{O(S)}(\boldsymbol{\mu}_{O(S)}) > \boldsymbol{\pi}(\boldsymbol{\mu})^{O(S)}. \tag{21}$$

Proof. If $\boldsymbol{\mu}$ is not in S^4, there is by definition a collection of agents $C \subseteq M(t) \cup F(t)$, for some $t \leqslant T$, and a matching $\boldsymbol{\mu}_{O(C)}$ such that:

$$\boldsymbol{\pi}_{O(C)}(\boldsymbol{\mu}_{O(C)}) > \boldsymbol{\pi}(\boldsymbol{\mu})^{O(C)}. \tag{22}$$

If $\boldsymbol{\mu}_{O(C)}$ satisfies S^4 with respect to $O(C)$, then take $S = C$ and we are done. Otherwise, there exists a $\tau \geqslant t$, a coalition of agents $C' \subset [M(\tau) \cup F(\tau)] \cap O(C)$, with $O(C') \subset O(C)$, and a feasible matching plan, $\boldsymbol{\mu}_{O(C')}$, such that:

$$\boldsymbol{\pi}_{O(C')}(\boldsymbol{\mu}_{O(C')}) > \boldsymbol{\pi}_{O(C)}(\boldsymbol{\mu}_{O(C)})^{O(C')} \geqslant \boldsymbol{\pi}(\boldsymbol{\mu})^{O(C')}. \tag{23}$$

That is, $\boldsymbol{\mu}$ is also S^4 blocked by $O(C')$ through $\boldsymbol{\mu}_{O(C')}$. Again, if $\boldsymbol{\mu}_{O(C')}$ satisfies S^4 in the game with $O(C')$, the claim in the proposition is true with $S = C'$. Otherwise we can repeat the argument for a subcoalition of agents in $O(C')$. Since remaining unmatched every period satisfies S^4 in a game with a single agent and her successors, we will eventually find a coalition of agents that blocks $\boldsymbol{\mu}$ through a plan which is consistent with S^4. □

It is important to emphasize that self-sustainability is not a vacuous requirement in the definition for S^3. It is only the additional limitation on deviating plans incorporated in S^4 which yields self-sustainability for free.

S^4 also has the desirable interpretation that it can be viewed as the core of the agent-form game, with one qualification: only coalitions of the form $O(S)$, where $S \subseteq M(t) \cup F(t)$ for some t, can deviate.

Before we discuss the existence of matching plans that satisfy S^4, we note that S^4 does rule out the "incredible" deviation in the example of matrix (18). It can be shown that the matching plan $\boldsymbol{\mu}^{\text{G–S}}$, which specifies the Gale–Shapley matching in both periods, is in the S^4 set. Rather than illustrating this, we prove, as part of the next theorem, that this is a general phenomenon: any (not necessarily identical) sequence of Gale–Shapley matchings is in S^4.

5.1. *Existence of S^4*

One desirable feature of S^4 is that it always exists in a matching market with a finite number of periods. We have the following theorem.

Theorem 4. *In a matching game with a finite number of periods, any sequence of Gale–Shapley stage matchings satisfies* S^4. *Therefore, the set of matching plans that satisfy* S^4 *is non-empty.*

Proof. Let $\boldsymbol{\mu}^{\text{G–S}}$ be a matching plan in the agent-form, such that for all t, $\boldsymbol{\mu}^{\text{G–S}}(t)$ is a Gale–Shapley matching. Suppose $\boldsymbol{\mu}^{\text{G–S}}$ is not in S^4. Then, there is a coalition of agents $S \subseteq M(t) \cup F(t)$, for some $t \leqslant T$, and a feasible matching $\boldsymbol{\mu}^{\text{block}}_{O(S)}$ for $O(S)$ such that

$$\boldsymbol{\pi}_{O(S)}\left(\boldsymbol{\mu}^{\text{block}}_{O(S)}\right) > \boldsymbol{\pi}\left(\boldsymbol{\mu}^{\text{G–S}}\right)^{O(S)}. \tag{24}$$

$\boldsymbol{\mu}^{\text{G–S}}$ specifies a Gale–Shapley matching in the last period. Thus, for any coalition of agents C in the last period and any feasible matching $\boldsymbol{\mu}_C$,

$$\boldsymbol{\pi}(\boldsymbol{\mu}_C) \not\succ \boldsymbol{\pi}\left(\boldsymbol{\mu}^{\text{G–S}}\right)^{C}. \tag{25}$$

If $C \subseteq [M(T) \cup F(T)] \cap O(S)$, (24) and (25) together imply:

$$\boldsymbol{\pi}_{O(S)}\left(\boldsymbol{\mu}^{\text{block}}_{O(S)}\right)^{C} = \boldsymbol{\pi}\left(\boldsymbol{\mu}^{\text{G–S}}\right)^{C}. \tag{26}$$

Thus, the blocking plan $\boldsymbol{\mu}^{\text{block}}_{O(S)}$ must be identical to $\boldsymbol{\mu}^{\text{G–S}}$ in the last period. We now need only to establish the inductive step that if, for some $\tau > t$ and all $C \subseteq [M(\tau) \cup F(\tau)] \cap O(S)$,

$$\boldsymbol{\pi}_{O(S)}\left(\boldsymbol{\mu}^{\text{block}}_{O(S)}\right)^{O(C)} = \boldsymbol{\pi}\left(\boldsymbol{\mu}^{\text{G–S}}\right)^{O(C)}, \tag{27}$$

then, for all $C' \subseteq [M(\tau-1) \cup F(\tau-1)] \cap O(S)$,

$$\boldsymbol{\pi}_{O(S)}\left(\boldsymbol{\mu}^{\text{block}}_{O(S)}\right)^{O(C')} = \boldsymbol{\pi}\left(\boldsymbol{\mu}^{\text{G–S}}\right)^{O(C')}. \tag{28}$$

To see why the claim is true, first notice that $O(C') \subseteq O(S)$, implies:

$$\boldsymbol{\pi}_{O(S)}\left(\boldsymbol{\mu}^{\text{block}}_{O(S)}\right)^{O(C')} \geqslant \boldsymbol{\pi}\left(\boldsymbol{\mu}^{\text{G–S}}\right)^{O(C')}. \tag{29}$$

Writing the payoff of agents at time $\tau - 1$ as the sum of their immediate payoff plus the payoff of their immediate future selves, we have

$$\boldsymbol{\pi}_{O(S)}\big(\boldsymbol{\mu}^{\text{block}}_{O(S)}\big)^{S(\tau-1)} = \pi\big(\boldsymbol{\mu}^{\text{block}}_{O(S)}(\tau-1)\big)^{S} + \boldsymbol{\pi}_{O(S)}\big(\boldsymbol{\mu}^{\text{block}}_{O(S)}\big)^{S(\tau)}, \tag{30}$$

$$\boldsymbol{\pi}\big(\boldsymbol{\mu}^{\text{G–S}}\big)^{M(\tau-1)\cup F(\tau-1)} = \pi\big(\boldsymbol{\mu}^{\text{G–S}}(\tau-1)\big) + \boldsymbol{\pi}\big(\boldsymbol{\mu}^{\text{G–S}}\big)^{M(\tau)\cup F(\tau)}. \tag{31}$$

(29)–(31) and (27) imply that for all $C' \subseteq [M(\tau-1)\cup F(\tau-1)]\cap O(S)$,

$$\pi\big(\boldsymbol{\mu}^{\text{block}}_{O(S)}(\tau-1)\big)^{C'} \geqslant \pi\big(\boldsymbol{\mu}^{\text{G–S}}(\tau-1)\big)^{C'}. \tag{32}$$

Since $\boldsymbol{\mu}^{\text{G–S}}(\tau-1)$ is a Gale–Shapley matching for $M(\tau-1)\cup F(\tau-1)$, the above cannot hold with strict inequality. From (30) and (31) we can then deduce that (28) holds. □

Note that, in general, S^4 can sustain matching plans which are not merely sequences of Gale–Shapley matchings. The following is an example:

	f_1	f_2	f_3	f_4
m_1	2, 4	1, 4	3, 2	4, 2
m_2	3, 3	2, 3	4, 1	1, 1
m_3	2, 2	4, 2	3, 4	1, 3
m_4	3, 1	1, 1	4, 2	2, 4

(33)

The above stage market has two and only two Gale–Shapley matchings. In the male-preferred one, μ^{M}, m_1, m_2, m_3, and m_4 are paired with f_4, f_1, f_2, and f_3, respectively. In the female-preferred matching, μ^{F}, each m_i is matched to f_i for i in $\{1, 2, 3, 4\}$.

Assume that the above market is repeated twice with no discounting and consider a matching plan $\boldsymbol{\mu}$ that specifies μ^{F} in period 2. In the first period, $\boldsymbol{\mu}$ specifies the unstable matching μ, where m_1, m_2, m_3, and m_4 are paired with f_4, f_3, f_2, and f_1, respectively.

In the stage market, μ is not in the Gale–Shapley set because it is blocked by m_4 and f_3. Notice, however, that the coalition of agents $O(S)$, with $S = \{m_4(1), f_3(1)\}$, does not block $\boldsymbol{\mu}$ under S^4. This is because a period 1 agent, $f_3(1)$, would not agree to the deviation, so the coalition is ineffective even under S^3.

Now, consider the coalition of agents $O(S)$, where $S = \{m_2(1), f_1(1)\}$. $O(S)$ would block $\boldsymbol{\mu}$ under S^3: $\boldsymbol{\pi}_{O(S)}(\boldsymbol{\mu}'_{O(S)})^{S} = (6, 6) > (6, 5) = \boldsymbol{\pi}(\boldsymbol{\mu})^{S}$. It does not block under S^4: $\boldsymbol{\pi}_{O(S)}(\boldsymbol{\mu}'_{O(S)})^{O(S)} = (6, 6, 3, 3) \not> (6, 5, 2, 4) = \boldsymbol{\pi}(\boldsymbol{\mu})^{O(S)}$. The payoffs are ordered: $m_2(1)$, $f_1(1)$, $m_2(2)$, $f_1(2)$. Under the candidate plan, $m_2(1)$'s payoff is $4 + 2$. Under the deviation, $m_2(1)$ receives a lower direct payoff from his match, but receives a higher externality from the match of $m_2(2)$: $m_2(1)$'s payoff is $3 + 3$. Whether the coalition $O(S)$ can credibly deviate depends on whether the player m_2 can trust f_1 to continue with the deviation in the second period. S^4 limits the trust among agents to the minimum. In particular, a proposed deviation is only credible if the deviating players are made better off at every point in time. In the example, player m_2 does not trust f_1 because $f_1(2)$ is strictly better off under $\boldsymbol{\mu}$ than under the deviating plan $\boldsymbol{\mu}'_{O(S)}$. Other deviations can be similarly eliminated.

6. Comparing definitions of stability

We have discussed five possible concepts of stability for a dynamic matching market: the core, the recursive core, the modified core, S^3, and S^4. These concepts can be summarised by Table 1.

Table 1 also contains a concept that has not yet been introduced. Blocking coalitions in the *strict core* have to satisfy the more stringent condition associated with S^4. However, neither the grand coalition's plan, nor deviating plans, have to be time consistent. The blocking requirement is made only at the beginning of the game.

In addition to categorizing these concepts, the table shows the inclusion relationships between these sets of matching plans. The recursive core is a subset of the core, because the recursive core imposes time-consistency. The recursive core is a subset of the S^3 set, because S^3 only allows for self-sustaining deviations. The S^3 set is a subset of the S^4 set, because the latter involves a stricter condition for when deviations are effective. This stricter condition also explains why the core is a subset of the strict core.

More surprisingly, there is no inclusion relationship between S^3 and the modified core, even though S^3 imposes the additional condition of dynamic consistency. The explanation lies in the interaction between self-sustainability and time-consistency. Although time-consistency tends to reduce admissible matching plans for the grand coalition, it also limits the set of coalitional deviations because of self-sustainability. The first effect tends to make the S^3 set smaller relative to the modified core, while the second tends to make it larger. A similar explanation applies to the lack of a inclusion result between the strict core and S^4.[18]

Table 1
Definitions of stability

	Does not impose time consistency		Imposes time consistency
Does not impose self-sustainability	CORE Pareto optimal	$\supseteq$	RECURSIVE CORE May be empty Pareto optimal
	$=$		$\subseteq$
Imposes self-sustainability	MODIFIED CORE	$\not\supseteq$	S^3 Non-empty under conditions May not be Pareto optimal
	$\subseteq$		$\subseteq$
Imposes "stricter" blocking condition	STRICT CORE May not be Pareto optimal	$\not\supseteq$	S^4 Non-empty
$\Rightarrow$			May not be Pareto optimal
Self-sustainability			

[18] Similar inclusion relationships apply to the non-cooperative notions of: strong Nash, perfect strong, coalition-proof, and perfect coalition-proof equilibrium.

Finally, the table notes that, in contrast to the core and the recursive core, matching plans that satisfy either S^3, S^4, or the strict core, may not be Pareto optimal with respect to players (not agents).

6.1. Other notions of credibility

The above concepts implicitly make different assumptions regarding the options that are available to deviators following a deviation from a candidate stable plan. In S^3, players in a deviating coalition are prevented from further deviating with players outside of the coalition. In S^4, players in a deviating coalition are allowed to interact with non-deviators in periods subsequent to the deviation. The outcome of this interaction is, however, limited to a return to the original plan.

In a static framework, two concepts which allow for more general deviations from deviations are Zhou's (1994) *bargaining set* and Klijn and Massó's (2003) concept of *weak stability*. These concepts allow members of the deviating coalition to form a pact to deviate further with players not included in the coalition. These concepts do not, however, require deviations from deviations to themselves be stable to yet further deviations. That is, they do not impose self-sustainability. Self-sustainability is a requirement that all coalitions be treated uniformly. It is particularly important in a dynamic environment where time-consistency is a natural minimal requirement for all coalitions—including deviations from deviations—to be credible.

Also in a static framework, Chwe's (1994) concept of *farsighted coalitional stability* allows for arbitrary deviations from deviations and incorporates strategic behavior that is similar to self-sustainability. In his concept, players may deviate in order to trigger a series of further deviations from which they will ultimately benefit. Bhattacharya (2002) proposes a modification of Chwe's (1994) stability concept by introducing the credibility requirement that only undominated deviations be considered. In a recent paper, Konishi and Ray (2003) look more closely to the idea of sequences of coalitional deviations in a dynamic model where coalitions form and break over time.

In this paper, we are interested in deviational credibility in a dynamic environment. Our concepts focus on time-consistency and self-sustainability. Though time-consistent and self-sustaining, one criticism of plans in S^3 is that deviations do not have to immune to proposals to deviate with players from outside of the deviation. S^4 partially addresses this criticism. It allows deviations to return to the original plan. If the original plan is "stable," returning to it is certainly a credibile threat that deviations have to consider.

S^4 does, however, rule out more general deviations from deviations. In non-cooperative game theory, different beliefs held by players off the equilibrium path can support different equilibrium behavior. Here, different beliefs about what further deviations are possible would give rise to concepts of stability that are different from S^4. Further, when assuming that a deviating coalition may choose to return to the original plan, S^4 is silent on whether agents who do not belong to the deviating group will consent. If, following the deviation, these agents re-match among themselves, they might not be willing to return to the original plan. These limitations notwithstanding, the simplicity of S^4—it is a simple modification of the core of the agent-form—and the existence result, make it a starting point for addressing the issue of stability in a dynamic environment. Developing a stability concept in a dynamic

market that allows for time-consistency, self-sustainability, and a more general deviations from deviations, remains a challenge.

7. Conclusion

This paper considers various notions of stability in dynamic matching markets. The dynamic nature of the market introduces a number of issues that are not present in a static model. First, time consistency is an important requirement if players cannot credibly commit to a matching plan at the outset of the game. Second, what constitutes a credible deviation can have important implications on the predictions of the model, even more so than in a static market.

We showed how S^3 may be a more appropriate stability concept than the recursive core if credibility requires deviations to be self-sustaining. If, in addition, credible deviations must be robust to proposals to rejoin the original plan, then the predictions of S^4 are more relevant. Which concept is appropriate depends essentially on the amount of commitment that is possible among the players.

The agent form representation of the dynamic game proved a powerful tool for investigating these different credibility issues within a unified framework.

Acknowledgments

Many thanks to Dirk Bergemann for introducing us to matching models and for his invaluable advice and encouragement. David Pearce, Ben Polak, Herbert Scarf, Abhijit Sengupta, as well as various seminar participants provided helpful comments. Two anonymous referees provided detailed suggestions that greatly improved the paper. The first author thankfully acknowledges financial support from the Connaught Foundation.

References

Bhattacharya, A., 2002. Coalitional stability with a credibility constraint. Math. Soc. Sci. 43, 27–44.

Becker, R.A., Chakrabarti, S.K., 1995. The recursive core. Econometrica 63, 401–423.

Bernheim, B.D., Peleg, B., Whinston, M.D., 1987. Coalition-proof Nash equilibria; I: Concepts. J. Econ. Theory 42, 1–12.

Blum, Y., Roth, A.E., Rothblum, U.G., 1997. Vacancy chains and equilibration in senior-level labor markets. J. Econ. Theory 76, 362–411.

Chwe, M.S.Y., 1994. Farsighted coalitional stability. J. Econ. Theory 63, 299–325.

Damiano, E., Lam, R., 2001. Self-sustaining stability in dynamic marriage markets. Mimeo. University of Toronto.

Gale, D., 1982. Money in Equilibrium. Cambridge Univ. Press, Cambridge.

Gale, D., Shapley, D., 1962. College admissions and the stability of marriage. Amer. Math. Monthly 69, 9–15.

Klijn, F., Massó, J., 2003. Weak stability and a bargaining set for the marriage model. Games Econ. Behav. 42, 91–100.

Konishi, H., Ray, D., 2003. Coalition formation as a dynamic process. J. Econ. Theory 110, 1–41.

Ray, D., 1989. Credible coalitions and the core. Int. J. Game Theory 18, 185–187.

Roth, A.E., Sotomayor, M.A.O., 1990. Two-sided Matching: A Study in Game-theoretic Modeling and Analysis. Cambridge Univ. Press, Cambridge.
Roth, A.E., Vande Vate, J.H., 1990. Random paths to stability in two-sided matching. Econometrica 58, 1475–1480.
Rubinstein, A., 1990. Strong perfect equilibrium in supergames. Int. J. Game Theory 9, 1–12.
Zhou, L., 1994. A new bargaining set of an N-person game and endogenous coalition formation. Games Econ. Behav. 6, 512–526.