

Matching Patterns when Group Size Exceeds Two

By CHRISTIAN AHLIN*

We study one-sided matching when groups with $n > 2$ members are being formed. Type-complementarity rules out all but the rank-ordered grouping. Type-substitutability (for example, matching to share risk) rules out much less. It requires that every two groups must be “intertwined”, in that each dominates the other at some rank. Intertwined matching is necessary and, in one context, sufficient for any grouping to be a potential equilibrium. But there are many intertwined matching patterns when $n > 2$. Thus, substitutability can be observationally similar to complementarity; we demonstrate this by showing that dyadic regressions can register intertwined (negative assortative) matching as homogeneous matching.

JEL: C78

Keywords: Matching, Group Formation, Risk Sharing, Dyadic Regression, Positive Assortative Matching, Negative Assortative Matching

Models of frictionless, one-sided matching under transferable utility have been used in a wide variety of contexts. Examples include skill matching and underdevelopment (Kremer, 1993); skill diversity and trade patterns (Grossman and Maggi, 2000); microcredit group formation (Ghatak, 1999, 2000, and Ahlin, 2015); matching to share risk (Schulhofer-Wohl, 2006);¹ and matching informed experts (Chade and Eeckhout, 2014).

Some stark results on matching into size-two groups have emerged from this and related literature (see especially Legros and Newman, 2002). If types are complements in producing group output (group payoff function is supermodular), the unique matching equilibrium involves perfectly homogeneous matching: both agents in a group have the same type. If types are substitutes in group output (group payoff function is submodular), the unique matching equilibrium involves a specific form of heterogeneous matching, “onion-style” or “median”: each agent matches with a type from the complementary percentile, e.g. a 95th percentile type with a 5th percentile type. Thus, both substitutability and complementarity

* Department of Economics, Michigan State University, 486 West Circle Dr, East Lansing, MI 48824; +1 517 3558306; ahlin@msu.edu. I thank Ashley Ahlin, Alfredo Burlando, Pierre-Andre Chiappori, Vivek Dhand, Jan Eeckhout, Sergei Izmalkov, Patrick Legros, Andrew Newman, Debraj Ray, Katerina Sherstyuk, Myrna Wooders, and workshop/seminar participants at ThReD-Oslo, Universitat Pompeu Fabra, New Economic School, and University of Oregon for valuable input. All errors are mine.

¹Schulhofer-Wohl (2006) shows that in a risk-sharing game where players differ only in risk aversion, even though utility is not transferable in any single state of the world, expected utility has a representation that is transferable iff preferences are of the ISHARA form. He analyzes a two-sided setting, but his results are straightforwardly extendable to the one-sided case with groups of size two.

have powerful implications when group size is two: they pin down matching patterns, *regardless* of the distribution of types and payoff function details.

However, size-two groups are often counterfactual – many firms or production teams have more than two workers, microcredit groups have more than two members, and risk-sharing networks have more than two households. Empirical work is thus often forced to extrapolate the results from two-person contexts to data featuring larger groups. It would clearly be helpful to understand how matching patterns generalize to groups of larger size.

One example is matching into production teams or firms. This subject is taken up by Grossman and Maggi (2000) and applied to trade patterns, but under the potentially restrictive assumption that production teams all have two members. Another illustrative example is matching for risk-sharing. The theoretical literature on how households that are heterogeneous in risk preferences match in order to share risk² studies two-sided matching, e.g. men and women marrying to share risk. Still open is the question of who matches with whom in a context that seems at least as relevant empirically, one-sided matching into risk-sharing groups of more than two households.

This paper studies one-sided matching of agents with one-dimensional types into groups of any fixed size $n \geq 2$, providing necessary and sufficient conditions for matching patterns to be observable in equilibrium, and demonstrating empirical implications.

It is straightforward to see how matching patterns under complementarity could generalize to the case of $n > 2$. Indeed, we show that complementarity dictates matching in the same way when $n > 2$ as when $n = 2$: in any equilibrium, groups are rank-ordered (Proposition 1), and perfectly homogeneous if there is a continuum of agents (Proposition 2). These predictions hold regardless of the distribution of agent types.

It is harder to conjecture how median or onion-style matching generalizes to $n > 2$. We do in fact find that for substitutability, the matching picture is dramatically different when $n > 2$. Onion-style, median, and even heterogeneous matching are no longer necessarily accurate descriptions of the matching pattern. What generalizes is “intertwined” matching: in any pair of equilibrium groups, each group strictly dominates the other at some rank (Proposition 3). That is, if L and M are equilibrium groups, there exist ranks i and i' such that the i th-ranked type in L is strictly higher than the i th-ranked type in M , and the i' th-ranked type in L is strictly lower than the i' th-ranked type in M .³

Intertwined matching is not only necessary under substitutability, it can be sufficient. We provide one context in which *any* intertwined matching pattern

²See, e.g., Schulhofer-Wohl (2006), Legros and Newman (2007), Attanasio et al. (2012), and Chiappori and Reny (2016). Schulhofer-Wohl is the only transferable utility framework, and thus closest to our model. Genicot and Ray (2003) address a complementary question, the stable size of risk-sharing groups composed of ex ante identical households; our study bypasses this issue with the assumption of a fixed, exogenous group size.

³An exception exists when the groups are identical, or nearly identical.

can characterize the unique equilibrium grouping, depending on the distribution of types (Proposition 4); and we provide a broader context in which at least a large subset of intertwined matching patterns can characterize an equilibrium, depending on the type distribution (Proposition 5).

An implication of these results is that many matching patterns are observable in equilibrium under substitutability when $n > 2$. This follows because there are many intertwined matching patterns, and more so for larger n . In the narrower context (of Proposition 4), a fraction $n/(n+1)$ of all possible matching patterns is intertwined and may characterize the unique equilibrium. In the broader context (of Proposition 5), when $n = 10$ for example, there are at least 19,000 intertwined matching patterns that may characterize the equilibrium.

Thus, in some settings when $n > 2$, substitutability by itself does relatively little to pin down the matching pattern. The final main result (Proposition 6) shows under fairly general conditions⁴ that, with $n > 2$, equilibrium matching patterns under substitutability *always* depend on the distribution of types – there is never a matching pattern that characterizes an equilibrium regardless of the type distribution. This contrasts sharply with the unique matching-pattern results for $n = 2$, and for complementarity with any n .

Thus, complementarity and substitutability have very different predictive power if groups are larger than two; rationalizable matching patterns under substitutability are numerous, and equilibrium patterns depend critically on the distribution of types.

The results are significant theoretically for several reasons. They provide a first characterization of matching for risk-sharing when groups have some fixed size greater than two. In particular, the substitutability results apply directly to a one-sided version of the Schulhofer-Wohl (2006) model, and characterize a set of matching patterns that can all be equilibria depending on the distribution of risk preferences in the matching population. They also pinpoint the core prediction in matching under substitutability, showing that it is not heterogeneity or matching around the center, but intertwined matching. Finally, they make clear that pinning down matching patterns requires assumptions on the distribution of types under substitutability, but not under complementarity.

There are also significant implications for empirical work. Since it is compatible with many different groupings, substitutability can “look like” complementarity, at least using common empirical techniques that focus on homogeneity/heterogeneity of groups.

We illustrate this with simulations featuring dyadic regressions, which are often used to understand group formation patterns, often with groups larger than two.⁵ Dyadic regressions essentially measure whether groups are homogeneous or heterogeneous by type. When $n = 2$ there is a tight connection between group

⁴One required condition is that group size is not too big, i.e. less than the total number of types, so that groups must differ.

⁵For example, see Fafchamps and Gubert (2007), Gine et al. (2010), Arcand and Fafchamps (2012), Attanasio et al. (2012), and Barr, Dekker, and Fafchamps (2012).

homogeneity and positive assortative matching, and between group heterogeneity and negative assortative matching; but as our theoretical results uncover, there need be no such connection when $n > 2$. Indeed, we find in simulations that matching patterns produced by substitutability and characterized by negative assortative matching can show up, surprisingly, as *homogeneous* matching in dyadic regressions. That is, types with statistically similar characteristics can match together under both complementarity and substitutability. This shows the potential pitfalls in extrapolating from the $n = 2$ case, and calls into question the ability of reduced-form techniques focused on within-group homogeneity/heterogeneity, like dyadic regressions, to identify the nature of the matching pattern. Structural estimation that uses an explicit payoff function and information on types may be needed instead.

A few others have studied matching under *complementarity* when group size is greater than two. Most similarly, Legros and Newman (2002) state that several of their results on two-person homogeneous matching generalize to larger group size; and Durlauf and Seshadri (2003) characterize efficient one-sided group formation under complementarity and arbitrary fixed group size. We add here an explicit set of results on equilibrium matching under complementarity for any $n \geq 2$ – a minor contribution, though, given the existing results and conjectures in the literature.⁶

The *substitutability* results, however, are the main contribution of the paper. We know of only two other papers characterizing one-sided matching under substitutability with group size greater than two – and this is not the primary focus of either. Saint-Paul (2001) analyzes matching under complementarity, substitutability, and hybrid cases, with *fixed-measure* group size. The main related result is that, in a substitutability context similar to our Propositions 4 and 5, every group will have the same average type. In independent work, Chade and Eeckhout (2014) also study matching under substitutability, with *finite* group size. They prove that rank-ordered (positive assortative) matching is not an equilibrium. They also show how the equilibrium match can be derived, and demonstrate the precise sense in which it will involve spreading types evenly across groups. The current paper is the first to provide necessary and sufficient conditions characterizing the matching patterns that can result under substitutability when $n > 2$ – we show that significantly more can be ruled out than rank-ordering, but at least in some cases, quite a lot remains. This paper also adds to the literature by demonstrating the empirical implications of these characterizations, and by calling into question the prevalent assumption that there is a clear correspondence between how homogeneous/heterogeneous matching is and whether matching is positive/negative assortative.

The baseline model and complementarity results are in section I. Substitutabil-

⁶In the case of n -sided matching (in contrast to the one-sided, n -person matching of this paper), the classic positive assortative matching result for 2-sided marriage of Becker (1973) has been generalized by Lin (1992) and Sherstyuk (1999).

ity results are in section II. Section III presents simulation results and empirical implications. Section IV provides discussion and concluding remarks. Proofs are in the Appendix.

I. Model and Complementarity

A population of agents matches to form n -person groups, where each agent joins exactly one group and $n \geq 2$ is a fixed integer. Examples include workers sorting into firms, or into production teams within firms; firms forming alliances; and households forming microcredit groups to obtain joint liability loans, or households forming groups to share risk.

Each agent has a type $p \in \mathcal{P}$, where \mathcal{P} is a bounded subset of \mathfrak{R} . Type could capture human capital or ability, risk aversion, risk or return of income streams or of projects needing funding, or firm size or reputation.⁷

Unless otherwise stated, results apply to both cases where the population of agents is finite or a continuum. However, we sometimes focus on one of two cases: **Case F**: the population consists of a finite number of agents kn , for some integer $k \geq 2$.

Case C: the population consists of a continuum of agents with types drawn from \mathcal{P} according to a continuous, strictly increasing distribution function F , with \mathcal{P} convex. Following Legros and Newman (2002), we assume there is a continuum of agents of each type in \mathcal{P} .

The group payoff function, $\Pi : \mathcal{P}^n \rightarrow \mathfrak{R}$, is assumed twice continuously differentiable and symmetric, meaning invariant to any permutation of the n types. Types are said to be **complements (substitutes)** if the group payoff function exhibits strictly positive (strictly negative) second-order cross-partials everywhere on its domain. Given twice continuous differentiability, type complementarity (substitutability) is equivalent to strict supermodularity (submodularity) of Π .⁸ Utility is assumed fully transferable, so the n agents in a group are able to share their group payoff in any way.

A **group** G is a vector of n agent types, written $G = (p_1^G, p_2^G, \dots, p_n^G) \in \mathcal{P}^n$;

⁷Thus, for tractability, type is one-dimensional. Type multi-dimensionality must be left for future work.

⁸Super/submodularity are defined in the Appendix. For many of our results, differentiability is not needed, only super/submodularity. In the context of two-person matching, Legros and Newman (2002) consider generalizations of these ideas of complementarity and substitutability, e.g. that do not require differentiability or symmetry.

Similar terminology has been used in a different context, that of two-sided, many-to-one matching, by Kelso and Crawford (1982), Roth and Sotomayor (1990), Hatfield and Milgrom (2005), and Echenique (2012). In particular, they define “substitutable preferences” as a type of preferences of agents on the side of the market that can match with multiple partners from the other side. There are significant differences between this concept and the substitutability concept we study, particularly that they are applied to different objects (one side’s preferences versus overall match surplus) and are typically used for different purposes (guaranteeing existence versus characterizing the matching pattern). Nonetheless, our preliminary exploration suggests that the two concepts can place similar restrictions on payoff functions in some contexts, but orthogonal restrictions in others. The relationship between these two concepts may warrant further exploration.

equivalently, it can be represented as a multiset with n elements drawn from \mathcal{P} .⁹ Throughout the paper, the type indices in any vector or multiset of types will be understood to correspond to rank-ordering; for example, $p_1^G \leq p_2^G \leq \dots \leq p_n^G$ in any group G .

A **grouping** is a set of groups in which all agents in the population belong to exactly one group, i.e. such that the number or measure of each type of agent across groups in the grouping is consistent with the population number or measure of agents of each type. A **grouping of P** , where P is a multiset of kn agent types, is a set of k groups (G^1, G^2, \dots, G^k) satisfying $P = \uplus_{j=1}^k G^j$.¹⁰ Essentially, it is a grouping of kn agents, one of each type in P .

An **equilibrium (core) grouping** is a grouping in which payoffs exist for each agent in the population that a) are feasible, meaning the sum of agent payoffs in each equilibrium group does not exceed that group's payoff; and b) cannot be blocked by any n agents reorganizing into a group so that each achieves a strictly higher payoff.¹¹ Equilibrium groupings are always efficient, a fact that can be useful for characterizing matching patterns:

LEMMA 1: *Any equilibrium grouping is efficient.*

See Appendix for this and all remaining proofs.

It will be helpful to be able to identify grouping *patterns* precisely, independently of the specific distribution of types. Define **group-pattern** \mathcal{G} as an n -element subset of $\{1, 2, \dots, 2n\}$, and a **grouping-pattern** $\mathcal{M} \equiv \{\mathcal{G}^1, \mathcal{G}^2\}$ as a pair of disjoint group-patterns.¹² Next, consider two groups L and M , with $P = L \uplus M = \{p_1^P, p_2^P, \dots, p_{2n}^P\}$ being the multiset of the $2n$ types of the agents in the two groups.¹³ The two groups L and M are said to **fit** grouping-pattern $\{\mathcal{G}^1, \mathcal{G}^2\}$ if their respective types are drawn from P based on the respective ranks in $\{\mathcal{G}^1, \mathcal{G}^2\}$, that is, if $[L = (p_{\mathcal{G}^1_1}^P, p_{\mathcal{G}^1_2}^P, \dots, p_{\mathcal{G}^1_n}^P)$ and $M = (p_{\mathcal{G}^2_1}^P, p_{\mathcal{G}^2_2}^P, \dots, p_{\mathcal{G}^2_n}^P)]$ or $[L = (p_{\mathcal{G}^2_1}^P, p_{\mathcal{G}^2_2}^P, \dots, p_{\mathcal{G}^2_n}^P)$ and $M = (p_{\mathcal{G}^1_1}^P, p_{\mathcal{G}^1_2}^P, \dots, p_{\mathcal{G}^1_n}^P)]$. In the former case, group L is said to **fit** group-pattern \mathcal{G}^1 and group M is said to **fit** group-pattern \mathcal{G}^2 . For example, given that $p_1 \leq p_2 \leq \dots \leq p_{12}$, groups $L = (p_1, p_4, p_7, p_8)$ and $M = (p_2, p_3, p_5, p_6)$ fit grouping-pattern $\{(1, 4, 7, 8), (2, 3, 5, 6)\}$, as do groups $L' = (p_2, p_6, p_{10}, p_{12})$ and $M' = (p_4, p_5, p_7, p_9)$.

When types are substitutes and $n = 2$, the unique equilibrium involves onion-style matching (Grossman and Maggi, 2000). This implies that every pair of

⁹A multiset is a generalization of the set concept that allows multiple instances of the same element. The equivalence between the vector and multiset representations is due to the assumption of symmetry in this paper, i.e. all types enter the group payoff function symmetrically – thus, the order of types in the vector representation is irrelevant. A group will refer interchangeably to a vector or multiset, depending on context.

¹⁰The operation \uplus is the union operation generalized for multisets. It includes all instances of all types from the united multisets in the resulting multiset. For example, $\{1, 1, 2\} \uplus \{2, 3, 4\} = \{1, 1, 2, 2, 3, 4\}$.

¹¹We use the same basic setup as Legros and Newman (2002), who also draw on Kaneko and Wooders (1986); see their papers for more details.

¹²For example, let $n = 3$. Then $(1, 4, 6)$ and $(1, 3, 5)$ are two group-patterns, and $\{(1, 4, 6), (2, 3, 5)\}$ and $\{(1, 3, 5), (2, 4, 6)\}$ are two grouping-patterns.

¹³Recall that $p_1^P \leq p_2^P \leq \dots \leq p_{2n}^P$ is understood.

equilibrium groups can be written as $\{(p_1, p_4), (p_2, p_3)\}$, for some $p_1 \leq p_2 \leq p_3 \leq p_4$; equivalently, every pair of equilibrium groups fits the grouping-pattern $\{(1, 4), (2, 3)\}$. When types are complements and $n = 2$, in any equilibrium every pair of groups can be written as $\{(p_1, p_2), (p_3, p_4)\}$, for some $p_1 \leq p_2 \leq p_3 \leq p_4$; equivalently, every pair of equilibrium groups fits the grouping-pattern $\{(1, 2), (3, 4)\}$. The remarkable result is that under complementarity or substitutability when $n = 2$, every pair of equilibrium groups fits the same grouping-pattern – regardless of the distribution of types.

A. Complementarity Results

We say that two groups L and M are **rank-ordered** if $p_n^L \leq p_1^M$ or $p_n^M \leq p_1^L$, or equivalently, if they fit the grouping-pattern $\{(1, 2, \dots, n), (n+1, n+2, \dots, 2n)\}$. A *grouping* is **rank-ordered** if every pair of groups in the grouping is rank-ordered. Thus, no groups overlap in a rank-ordered grouping. As in Legros and Newman (2002), define a *grouping* as **segregated** if every group in the grouping contains n agents of the same type.

PROPOSITION 1: *Assume types are complements. Any equilibrium grouping is rank-ordered.*

In the finite case, this result implies that any equilibrium group formation is unique and simple: the highest n types in the first group, the next highest n in the next group, and so on.¹⁴ Whether $n = 2$ or $n > 2$ is clearly irrelevant.

As a simple example, consider six agents with unique types: $p_1 < p_2 < p_3 < p_4 < p_5 < p_6$. If $n = 2$, there are fifteen potential groupings. Only one of them is rank-ordered, and the rest are ruled out by Proposition 1:

(p_5, p_6)	(p_5, p_6)	(p_5, p_6)	(p_4, p_6)	(p_4, p_5)	(p_4, p_6)	(p_4, p_6)	(p_3, p_6)
(p_3, p_4)	(p_2, p_4)	(p_2, p_3)	(p_2, p_3)	(p_2, p_3)	(p_3, p_5)	(p_2, p_5)	(p_2, p_5)
(p_1, p_2)	(p_1, p_3)	(p_1, p_4)	(p_1, p_5)	(p_1, p_6)	(p_1, p_2)	(p_1, p_3)	(p_1, p_4)
(p_3, p_6)	(p_3, p_5)	(p_4, p_5)	(p_4, p_5)	(p_3, p_5)	(p_3, p_4)	(p_3, p_4)	
(p_2, p_4)	(p_2, p_4)	(p_3, p_6)	(p_2, p_6)	(p_2, p_6)	(p_2, p_6)	(p_2, p_5)	
(p_1, p_5)	(p_1, p_6)	(p_1, p_2)	(p_1, p_3)	(p_1, p_4)	(p_1, p_5)	(p_1, p_6)	

It follows that, in any equilibrium, every pair of groups fits the grouping-pattern $\{(1, 2), (3, 4)\}$.

¹⁴This is true of any equilibrium grouping, but an equilibrium grouping may or may not exist in the case of a finite population. We have examples of non-existence, and existence proofs for some settings. See further discussion of existence in the Conclusion.

If instead $n = 3$, there are ten groupings. Again, only one is rank-ordered, so the rest are ruled out by Proposition 1:

$$\begin{array}{ccccc}
 (p_4, p_5, p_6) & \cancel{(p_3, p_5, p_6)} & \cancel{(p_3, p_4, p_6)} & \cancel{(p_3, p_4, p_5)} & \cancel{(p_2, p_5, p_6)} \\
 (p_1, p_2, p_3) & \cancel{(p_1, p_2, p_4)} & \cancel{(p_1, p_2, p_5)} & \cancel{(p_1, p_2, p_6)} & \cancel{(p_1, p_3, p_4)} \\
 \cancel{(p_2, p_4, p_6)} & \cancel{(p_2, p_4, p_5)} & \cancel{(p_2, p_3, p_6)} & \cancel{(p_2, p_3, p_5)} & \cancel{(p_2, p_3, p_4)} \\
 \cancel{(p_1, p_3, p_5)} & \cancel{(p_1, p_3, p_6)} & \cancel{(p_1, p_4, p_5)} & \cancel{(p_1, p_4, p_6)} & \cancel{(p_1, p_5, p_6)}
 \end{array}$$

It follows that, in any equilibrium, every pair of groups fits the grouping-pattern $\{(1, 2, 3), (4, 5, 6)\}$. In general, any pair of groups in an equilibrium grouping under complementarity fits the grouping-pattern $\{(1, 2, \dots, n), (n+1, n+2, \dots, 2n)\}$.

When there is a continuum of agents, rank-ordering is squeezed to perfect group homogeneity, i.e. segregation:

PROPOSITION 2: *Assume types are complements. In the case of a continuum of agents (case C), there is a unique equilibrium grouping, and it is segregated.*

This result too is no different whether $n = 2$ or $n > 2$. It is also strong. Existence and uniqueness of a homogeneous matching equilibrium obtain, regardless of group size and independent of the type distribution and details of the production function.

The results of this section are included for completeness rather than novelty, though the existence proof of Proposition 2 appears new and relies explicitly on supermodularity.¹⁵

II. Substitutability

It is not obvious how to generalize substitutability's onion-style or median matching to the case of $n > 2$. If n is even, a direct generalization could hold, where the highest $n/2$ types match with the lowest $n/2$ types, the next highest $n/2$ types match with the next lowest $n/2$, and so on. In this case, every pair of groups would fit matching-pattern $\{(1, 2, \dots, n/2, 3n/2+1, 3n/2+2, \dots, 2n), (n/2+1, n/2+2, \dots, 3n/2)\}$, whether $n = 2$ or $n > 2$. But is this grouping an equilibrium? And what matching pattern would work for n odd?

A. Necessity of "Intertwined" Matching

Let $\mathcal{N} \equiv \{1, 2, \dots, n\}$. Say that two groups L and M are **intertwined** if there exist $i, i' \in \mathcal{N}$ such that L strictly dominates M at rank i ($p_i^L > p_i^M$) and M strictly dominates L at rank i' ($p_{i'}^M > p_{i'}^L$). A grouping-pattern is intertwined under the same conditions, i.e. if each group-pattern strictly dominates the other

¹⁵The results of Kaneko and Wooders (1996) also establish equilibrium existence in the continuum setting, of an approximately feasible core.

at some rank.¹⁶ Clearly, to say two groups are intertwined is to say more than that they are not rank-ordered (see examples below).

Define two groups L and M as nearly rank-wise identical, or **nearly identical**, if they have the same value at $n - 1$ or n ranks, i.e. if there exists an $i^* \in \mathcal{N}$ such that $p_i^L = p_i^M$ for $i \in \mathcal{N} \setminus \{i^*\}$.¹⁷

Say that a *grouping* is **intertwined** if every pair of groups in the grouping is intertwined or nearly identical, and a grouping is **fully intertwined** if every pair of groups in the grouping is intertwined.

PROPOSITION 3: *Assume types are substitutes. Any equilibrium grouping is intertwined.*

Thus, substitutability of types implies that in equilibrium, every pair of groups is intertwined, i.e. each group dominates the other at some rank. The only possible exception occurs when the two groups are almost identical.

Consider again the simple example of six uniquely-typed agents. If $n = 2$, only one grouping is intertwined – the onion-style one – and the rest are ruled out by Proposition 3:¹⁸

(p₅,p₆)	(p₅,p₆)	(p₅,p₆)	(p₄,p₆)	(p₄,p₅)	(p₄,p₆)	(p₃,p₆)	(p₃,p₅)
(p₃,p₄)	(p₂,p₄)	(p₂,p₃)	(p₂,p₃)	(p₂,p₃)	(p₃,p₅)	(p₂,p₄)	(p₂,p₄)
(p₁,p₂)	(p₁,p₃)	(p₁,p₄)	(p₁,p₅)	(p₁,p₆)	(p₁,p₂)	(p₁,p₅)	(p₁,p₆)
(p₄,p₆)	(p₃,p₆)	(p₄,p₅)	(p₄,p₅)	(p₃,p₅)	(p₃,p₄)	(p ₃ ,p ₄)	
(p₂,p₅)	(p₂,p₅)	(p₃,p₆)	(p₂,p₆)	(p₂,p₆)	(p₂,p₆)	(p₂,p₅)	
(p₁,p₃)	(p₁,p₄)	(p₁,p₂)	(p₁,p₃)	(p₁,p₄)	(p₁,p₅)	(p₁,p₆)	

In fact, it is straightforward to see that onion-style matching and intertwined matching are equivalent when $n = 2$, for any even number of agents.¹⁹

But consider $n = 3$. The necessity of intertwined matching rules out only five

¹⁶Groups $L = (1, 2, 5, 7)$ and $M = (1, x, 4, 7)$ are intertwined iff $2 < x < 5$, for then group M dominates L at the second rank and group L dominates group M at the third rank. (If $x < 2$, L dominates M at every rank; if $x > 5$, M dominates L at every rank.)

¹⁷Group $L = (1, 2, 5, 7)$ is nearly identical to $M = (1, x, 5, 7)$ iff $1 \leq x \leq 5$. If $5 < x \leq 7$, for example, then $p_2^M = 5 \neq 2 = p_2^L$ and $p_3^M = x \neq 5 = p_3^L$; i.e. the groups differ at more than one rank, ranks two and three.

¹⁸In the first row of groupings, the first and second groups are not intertwined, while in the second row, the second and third groups are not intertwined, with one noted exception.

¹⁹Requiring simply that no two groups be rank-ordered in any grouping would rule out only nine of the fifteen groupings and thus would not uniquely predict onion-style matching even when $n = 2$. (It would not rule out the six groupings in which types p_1 , p_2 , and p_3 are in different groups.)

In the $n = 3$ example next, requiring simply that no two groups be rank-ordered would rule out only one of the ten groupings.

groupings:

$$\begin{array}{ccccc}
 \begin{array}{c} \cancel{(p_4, p_5, p_6)} \\ \cancel{(p_1, p_2, p_3)} \\ (p_3, p_4, p_5) \\ (p_1, p_2, p_6) \\ A \end{array} &
 \begin{array}{c} \cancel{(p_3, p_5, p_6)} \\ \cancel{(p_1, p_2, p_4)} \\ (p_2, p_4, p_5) \\ (p_1, p_3, p_6) \\ B \end{array} &
 \begin{array}{c} \cancel{(p_3, p_4, p_6)} \\ \cancel{(p_1, p_2, p_5)} \\ (p_2, p_3, p_6) \\ (p_1, p_4, p_5) \\ C \end{array} &
 \begin{array}{c} \cancel{(p_2, p_5, p_6)} \\ \cancel{(p_1, p_3, p_4)} \\ (p_2, p_3, p_5) \\ (p_1, p_4, p_6) \\ D \end{array} &
 \begin{array}{c} \cancel{(p_2, p_4, p_6)} \\ \cancel{(p_1, p_3, p_5)} \\ (p_2, p_3, p_4) \\ (p_1, p_5, p_6) \\ E \end{array}
 \end{array}
 \tag{1}$$

The groupings not ruled out (A–E) are all intertwined, i.e. each group dominates the other at some rank. Evidently, when $n > 2$ intertwining is significantly less restrictive than rank-ordering. If intertwining is not only necessary for a grouping to be an equilibrium, but also sufficient (under some set of types), then substitutability by itself rules out significantly less than complementarity.

The following corollaries show that nearly identical groups, which are not ruled out by Proposition 3, do not occur in some standard contexts.

COROLLARY 1: *Assume types are substitutes. In any equilibrium grouping of a continuum of agents (case C), two groups picked at random are intertwined with probability one.*

COROLLARY 2: *Assume types are substitutes. Any equilibrium grouping of a finite number of agents (case F), no two of whom have the same type, is fully intertwined.*

In both cases, group size is not too large relative to the total number of types in the population, so groups must differ.

However, when group size is so large that every group can contain the entire distribution of types – in contrast to much of the matching literature and the main contributions we follow, e.g. Grossman and Maggi (2000) and Legros and Newman (2002) – matching under substitutability may be characterized more by *identical* groups than by *intertwined* groups. This is true in Saint-Paul’s (2001) model, in which groups are fixed-measure continua and thus can all replicate the population distribution of types. It is also true of the fractional assignment version of Chade and Eeckhout’s (2014) model, in which agents can divide their time in any way between groups, so that again each group can mimic the population of types. Intertwined matching becomes more of a necessity when groups are of moderate size, and as a result cannot all perfectly mirror the population.

B. Sufficiency of “Intertwined” Matching

We have shown that intertwined matching is necessary for equilibrium; is it sufficient? Put differently, are there intertwined groupings that can be ruled out by substitutability alone? This section addresses these questions with three main results, obtained under progressively more general conditions.

SUM-BASED PAYOFF FUNCTION, $2n$ TYPES. — We first provide a setting in which any intertwined grouping pattern may be the unique equilibrium, depending on the distribution of types. The existence of such a setting guarantees that nothing beyond intertwined matching is generically implied by substitutability.

Assume that the payoff function can be written

$$(2) \quad \Pi(p_1, p_2, \dots, p_n) = \sum_{i \in \mathcal{N}} q(p_i) + h \left(\sum_{i \in \mathcal{N}} p_i \right)$$

for some continuously differentiable function q and twice continuously differentiable function h . The key feature is that the interaction between types in the payoff function comes through a function of the sum of types;²⁰ hence, we call this a **sum-based payoff function**. Types are substitutes iff $h''(\cdot) < 0$ and complements iff $h''(\cdot) > 0$.

The sum-based payoff function guarantees equilibrium existence in a subcase of both the finite and continuum case:

LEMMA 2: *Assume a sum-based payoff function and that types are substitutes. Any grouping in which every group has the same sum of agent types is an equilibrium.*

This is shown by constructing equilibrium payoffs for each agent type in such a grouping. It aids in proving our main result in this setting:

PROPOSITION 4: *Assume a sum-based payoff function and that types are substitutes. Fix any intertwined grouping-pattern $\mathcal{M} = \{\mathcal{G}^1, \mathcal{G}^2\}$. There exists a $\mathcal{P} = \{p_1, p_2, \dots, p_{2n}\}$ with $p_i < p_{i'}$ for $i < i'$, such that*

- the unique equilibrium grouping of \mathcal{P} fits grouping-pattern \mathcal{M} .*
- the unique equilibrium grouping of an equal number or measure of agents of each type in \mathcal{P} is a replication of the grouping in a), i.e. half of the groups fit \mathcal{G}^1 and half fit \mathcal{G}^2 .*

Thus, in this context and for any $n \geq 2$, any intertwined grouping-pattern may characterize the unique equilibrium grouping of $2n$ agents – it all depends on the set of types. The intertwined property is thus both necessary and sufficient for a grouping-pattern to be observable as an equilibrium. A novel implication of this result is that, unlike under complementarity or under $n = 2$, under substitutability with $n > 2$ the matching pattern can depend on the distribution of types.

As an example, consider the $n = 3$ case (expression (1) above). Proposition 4 guarantees that with a sum-based payoff function, each of the five intertwined groupings (A-E) is the unique equilibrium grouping of six agents for some set of

²⁰This is essentially the class of payoff functions that Saint-Paul (2001) analyzes, and that Chade and Eeckhout (2014) analyze most extensively.

types $\mathcal{P} = \{p_1, p_2, \dots, p_6\}$.²¹ Generalizing to an equal number or measure of each of six types, Proposition 4 guarantees existence of a \mathcal{P} that makes any one of these five groupings the basis for the unique equilibrium grouping – in that half of the groups are identical to one group, half to the other.

This result sheds light on an unsolved problem in the risk-sharing literature by showing how fixed-size risk-sharing groups form when group size is greater than two. Adapting the setup of Schulhofer-Wohl (2006) to our context, one can show that utility is transferable, the risk-sharing group payoff function is sum-based, and agent types (risk aversion) are substitutes, under similar assumptions.²² Thus, Proposition 4 applies, and for the type structure analyzed, any intertwined grouping-pattern of the $2n$ types may characterize a unique equilibrium grouping.

Thus, in this setting with $2n$ agents, any intertwined grouping-pattern is the basis for a potential unique equilibrium grouping. How many intertwined grouping-patterns are there?

LEMMA 3: *There are $\binom{2n}{n}/2$ total grouping-patterns and $\frac{n-1}{n+1}\binom{2n}{n}/2$ intertwined grouping-patterns. Thus, a fraction $\frac{n-1}{n+1}$ of all grouping-patterns is intertwined.*

Combining with Proposition 4, a fraction $\frac{n-1}{n+1}$ of all grouping patterns of the $2n$ types may characterize a unique equilibrium. Applying this again to risk-sharing group formation, we see that if $n > 2$, at least half of all grouping-patterns of the $2n$ types form the basis for a potential unique equilibrium grouping, and for n large, nearly all grouping-patterns may be observed.

In sum, here all that substitutability rules out by itself is non-intertwined matching; and this is not saying very much as group size moves away from two.

SUM-BASED PAYOFF FUNCTION, kn OR A CONTINUUM OF TYPES. — Does substitutability imply more than intertwined matching in cases other than that of $2n$ agent types? Perhaps most interestingly, does a greater number of types, groups, and/or agents allow some intertwined groupings to be ruled out? Our conjecture is no, but we do not have a complete answer to this. However, we are able to identify a large *subset* of intertwined grouping-patterns that characterize potential equilibria even with a continuum of types, groups, and agents.

To do so, consider two groups L and M , and let $P = L \uplus M$ be the multiset containing the $2n$ agent types from the two groups. L and M are **block-intertwined** if they are intertwined *and* there exists a partition of P into J multisets (**blocks**) such that: 1) each block $j \in \{1, \dots, J\}$ contains an even number of types, $2\nu_j$ say;

²¹This follows because all five intertwined groupings (A-E) fit an intertwined grouping-pattern. E.g., A fits $\{(1, 2, 6), (3, 4, 5)\}$, B fits $\{(1, 3, 6), (2, 4, 5)\}$, and so on.

Specifically, if \mathcal{P} satisfies $p_6 - p_5 > p_4 - p_1$, grouping A is uniquely efficient, and (by Lemma 1) the only possible equilibrium; if $p_4 - p_1 > p_6 - p_5 > p_4 - p_3, p_2 - p_1$, it is grouping B; if $p_4 - p_3 > p_6 - p_5, p_2 - p_1$, it is grouping C; if $p_6 - p_3 > p_2 - p_1 > p_4 - p_3, p_6 - p_5$, grouping D; and if $p_2 - p_1 > p_6 - p_3$, grouping E.

²²Details available on request.

2) the blocks are rank-ordered, i.e. $p_{2\nu_j}^j \leq p_1^{j'}$ or $p_{2\nu_{j'}}^{j'} \leq p_1^j$ for all $j, j' \in \{1, \dots, J\}$, $j \neq j'$; and 3) for each $j \in \{1, \dots, J\}$, group L contains either the ν_j largest types or the ν_j smallest types in block j . Block-intertwining applies to grouping-patterns in the same way as to pairs of groups.

To illustrate block-intertwining, consider $n = 4$ and some groupings of $\mathcal{P} = \{p_1, p_2, \dots, p_8\}$:

$$(3) \quad \begin{array}{cccc} L = & (p_4, p_5, p_6 \mid p_7) & (p_2 \mid p_3, p_4 \mid p_8) & (p_3, p_5, p_6, p_7) & (p_2, p_3, p_4, p_6) \\ M = & (p_1, p_2, p_3 \mid p_8) & (p_1 \mid p_5, p_6 \mid p_7) & (p_1, p_2, p_4, p_8) & (p_1, p_5, p_7, p_8) \\ & 1 & 2 & 3 & 4 \end{array}$$

(Dividers are used to highlight the block structure.) Grouping 1 is block-intertwined, since it is intertwined and has the following block structure: the first block contains the lowest six types, of which the three highest go in group L ; and the second block contains the highest two types, of which the lowest goes in L . Grouping 2 is also block-intertwined, since it is intertwined and has the following block structure: the first block contains the lowest two types, with the highest going in group L ; the second block contains the middle four types, with the lowest two going in L ; and the third block contains the highest two types, with the highest going in L . Grouping 3 is similar to grouping 1, but differs critically in that L does not contain the highest (or lowest) three types of the first block. Similarly, grouping 4 is intertwined but not block-intertwined.²³

It turns out that any block-intertwined grouping-pattern can form the basis for an equilibrium, in that all pairs of groups in the equilibrium fit this grouping-pattern.

PROPOSITION 5: *Assume a sum-based payoff function and that types are substitutes. Fix any **block-intertwined** grouping-pattern $\mathcal{M} = \{\mathcal{G}^1, \mathcal{G}^2\}$.*

a) *For any $k \geq 2$, there exists a $\mathcal{P} = \{p_1, p_2, \dots, p_{kn}\}$ with $0 < p_1 < p_2 < \dots < p_{kn} < 1$, and an equilibrium grouping of \mathcal{P} , in which every pair of groups uniquely fits grouping-pattern \mathcal{M} .*

b) *There exists a (strictly increasing, continuous) distribution function F over $\mathcal{P} = [0, 1]$ and an equilibrium grouping of a continuum of agents with this distribution of types, such that with probability one, two groups sampled from this grouping uniquely fit grouping-pattern \mathcal{M} .*

That is, a subset of intertwined grouping-patterns – all *block-intertwined* ones – form the basis for potential equilibria for any finite number of types, or a continuum. This result guarantees that the sufficiency of at least one form of intertwining is not an artefact of the case of $2n$ types, but survives even when there are many types.

²³Groupings 3 and 4 are the only two intertwined groupings of \mathcal{P} that are not block-intertwined; the other nineteen are. When $n = 2$ or $n = 3$, every intertwined grouping of $2n$ agents is block-intertwined.

To illustrate, let $n = 4$ and $k = 3$, and consider block-intertwined grouping-pattern $\{(1, 2, 3, 8), (4, 5, 6, 7)\}$. (Grouping 1 in expression (3) above fits this grouping-pattern.) Let

$$\mathcal{P} = \left\{ \underbrace{\frac{1}{72}, \frac{3}{72}, \frac{5}{72}}_A, \underbrace{\frac{7}{72}, \frac{9}{72}, \frac{11}{72}}_B, \underbrace{\frac{13}{72}, \frac{15}{72}, \frac{17}{72}}_C \mid \underbrace{\frac{27}{72}}_C, \underbrace{\frac{45}{72}}_B, \underbrace{\frac{63}{72}}_A \right\},$$

where a divider is used to highlight the block structure. Note that the grouping with the lowest three types matched with the highest (group A), the next lowest three types matched with the second highest (group B), etc., has all groups with equal sums of types, and thus is an equilibrium (Lemma 2). Clearly, every pair of groups in this equilibrium fits grouping-pattern $\{(1, 2, 3, 8), (4, 5, 6, 7)\}$. Consider also block-intertwined grouping-pattern $\{(1, 5, 6, 7), (2, 3, 4, 8)\}$. (See grouping 2 in equation (3) above.) Let

$$(4) \quad \mathcal{P} = \left\{ \underbrace{\frac{2}{36}}_A, \underbrace{\frac{6}{36}}_B, \underbrace{\frac{10}{36}}_C \mid \underbrace{\frac{13}{36}, \frac{15}{36}}_C, \underbrace{\frac{17}{36}, \frac{19}{36}}_B, \underbrace{\frac{21}{36}, \frac{23}{36}}_A \mid \underbrace{\frac{26}{36}}_A, \underbrace{\frac{30}{36}}_B, \underbrace{\frac{34}{36}}_C \right\}.$$

The marked grouping has all groups with equal sums of types, and thus is an equilibrium; and every pair of groups in this equilibrium fits grouping-pattern $\{(1, 5, 6, 7), (2, 3, 4, 8)\}$.

Moving beyond $k = 3$ groups to any k , even a continuum, is straightforward.²⁴ The proof of Proposition 5 provides general constructions for the finite and continuum case for any n and any block-intertwined grouping-pattern.

Thus, under a sum-based payoff function and substitutability, any block-intertwined grouping-pattern can characterize an equilibrium of any number of agents kn , or even of a continuum, depending on the distribution of types. How many block-intertwined grouping-patterns are there?

LEMMA 4: *There are $\sum_{j=1}^{n-1} \binom{n-1}{j} (2^j - 1)$ block-intertwined grouping-patterns.*

That is, the number of block-intertwined grouping-patterns increases rapidly in n – it more than doubles from n to $n + 1$, and exceeds 19,000 when $n = 10$. Thus Proposition 5 implies that, given substitutability, sum-based payoffs, and $n = 10$, before knowing \mathcal{P} , there are at least 19,000 candidate equilibrium groupings, and at least 19,000 potential grouping-patterns for two groups chosen at random – for *any* number of groups, even a continuum. From an empirical perspective, this result makes clear that matching patterns remain far from pinned down by substitutability alone, but will heavily depend on the distribution of types, even

²⁴In the above examples, one would simply keep the same block intervals – $[0, 1/4]$ and $[1/4, 1]$ in the first case, $[0, 1/3]$, $[1/3, 2/3]$, and $[2/3, 1]$ in the second case – and the same fraction of types within each block interval, scaling up the total number of types to fit the number of groups while keeping the types evenly spaced and centered within blocks in the manner followed above.

with many types and agents.²⁵

ANY PAYOFF FUNCTION, kn TYPES. — The results so far uncover a stark contrast between the predictive power of substitutability and that of complementarity. They have all been obtained assuming a sum-based payoff function, however. Here, this assumption is discarded to show the generality of the idea that the matching pattern depends critically on the distribution of types, under substitutability when $n > 2$. Twice continuous differentiability and symmetry of the payoff function continues to be assumed.

Necessary first is to generalize the notion of grouping-pattern to apply to any number of groups k , rather than only $k = 2$. To do so, define $\mathcal{N}_k \equiv \{1, 2, \dots, kn\}$. Define a **k-group-pattern** as an n -element subset of \mathcal{N}_k , and a **k-grouping-pattern** as a set of k disjoint group-patterns of \mathcal{N}_k . Let $M = \{G^1, G^2, \dots, G^k\}$ be a grouping of kn agents, and $P = \uplus_{j=1}^k G^j = \{p_1^P, p_2^P, \dots, p_{kn}^P\}$ be the multiset of the kn types of the agents in the k groups.²⁶ Grouping M is said to **fit** k-grouping-pattern $\mathcal{M} = \{\mathcal{G}^1, \mathcal{G}^2, \dots, \mathcal{G}^k\}$, if M 's groups' respective types are drawn from P based on \mathcal{M} 's group-patterns' respective ranks, that is, if $G^i = (p_{\mathcal{G}_1^i}^P, p_{\mathcal{G}_2^i}^P, \dots, p_{\mathcal{G}_n^i}^P)$ for all $i \in \{1, \dots, k\}$ (for some permutation of group indices). For example, with $n = 4$ and given that $p_1 \leq p_2 \leq \dots \leq p_{12}$, grouping $M = \{(p_1, p_4, p_7, p_{10}), (p_2, p_3, p_{11}, p_{12}), (p_5, p_6, p_8, p_9)\}$ fits the 3-grouping-pattern $\mathcal{M} = \{(1, 4, 7, 10), (2, 3, 11, 12), (5, 6, 8, 9)\}$.

PROPOSITION 6: *Assume types are substitutes. Fix any number of groups $k \geq 2$, group size $n \geq 3$, and k-grouping-pattern \mathcal{M} . There exists a $\mathcal{P} = \{p_1, p_2, \dots, p_{kn}\}$ with $0 < p_1 < p_2 < \dots < p_{kn} < 1$, such that the grouping of \mathcal{P} that fits \mathcal{M} is neither efficient nor an equilibrium.*

Thus, under substitutability and $n > 2$, when the number of population types exceeds group size, there exists no single grouping-pattern that characterizes an equilibrium *regardless* of the distribution of types; every grouping-pattern fails in at least some circumstances. Hence, equilibrium matching patterns always depend on the type distribution. This is in contrast with complementarity, where the same k-grouping-pattern is efficient regardless of the distribution of types, and only it can be observed in equilibrium: the rank-ordered one, $\{(1, 2, \dots, n), (n+1, n+2, \dots, 2n), \dots, (nk-n+1, nk-n+2, \dots, nk)\}$. It also contrasts with substitutability when $n = 2$, where only the onion-style k-grouping-pattern is efficient and observable in equilibrium, regardless of the distribution of types: $\{(1, 2k), (2, 2k-1), \dots, (k, k+1)\}$.

The intuition for this result is clearest with two groups. In any intertwined grouping-pattern, one group-pattern dominates the other at multiple ranks when

²⁵It remains an open question whether intertwined grouping-patterns that are not block-intertwined can be ruled out by substitutability when there are enough types.

²⁶Recall that $p_1^P \leq p_2^P \leq \dots \leq p_{kn}^P$ is understood.

$n \geq 3$.²⁷ But, a set of types exists in which the remaining ranks are of negligible importance, and thus the grouping that fits this grouping-pattern is arbitrarily close to non-intertwined. Another grouping can thus be found that dominates this one.

To illustrate, consider the intertwined groupings of the $n = 3$ example:

$$\begin{array}{ccccc} (p_3, p_4, p_5) & (p_2, p_4, p_5) & (p_2, p_3, p_6) & (p_2, p_3, p_5) & (p_2, p_3, p_4) \\ (p_1, p_2, p_6) & (p_1, p_3, p_6) & (p_1, p_4, p_5) & (p_1, p_4, p_6) & (p_1, p_5, p_6) \\ A & B & C & D & E \end{array} .$$

Note that in groupings A and B, the first group dominates the second at ranks one and two. Thus, if $p_1 < p_2 < \dots < p_5 = p_6$, A and B are not intertwined and by submodularity produce lower payoffs than C and D, which are intertwined. Note also that in groupings D and E, the second group dominates the first at ranks two and three. Thus, if $p_1 = p_2 < p_3 < \dots < p_6$, D and E are not intertwined and produce lower payoffs than B and C. And in grouping C, the first group dominates the second at ranks one and three; thus, if $p_1 < p_2 < p_3 = p_4 < p_5 < p_6$, C is not intertwined and produces lower payoffs than B and D. By continuity, in each case inefficiency of the grouping is robust to the types assumed equal being slightly different instead. This argument can be extended by induction to any intertwined grouping pattern for any $n \geq 3$ and $k \geq 2$.

In sum, the matching pattern under complementarity is robustly distribution-free, while the matching pattern under substitutability when $n > 2$ is robustly distribution-dependent, at least when group size is moderate.

C. Revisiting the Generalization

As the results make clear, matching patterns under complementarity generalize in the obvious way from $n = 2$ to $n > 2$. Thus, all the terminology from the $n = 2$ case extends to higher dimensions: rank-ordered matching, segregation, homogeneous matching, and positive assortative matching.

But, the results also show that matching patterns under substitutability generalize in a non-obvious way. Clearly, tight predictions about observable matching patterns are lost. But among the multiple observable matching patterns, does the $n = 2$ terminology still apply?

“Onion-style matching” does not. Even when n is even, so that onion-style matching can straightforwardly generalize – the top $n/2$ matching with the bottom $n/2$, and so on – this matching-pattern is not always an equilibrium or efficient, by Proposition 6. As a concrete example, Proposition 5 guarantees that an equilibrium can exist with all pairs of groups fitting the matching-pattern $\{(1, 5, 6, 7), (2, 3, 4, 8)\}$ – far from “onion-style”.²⁸

²⁷This is not true when $n = 2$, which is why the logic used here does not apply to that case.

²⁸See the example population and equilibrium grouping in equation (4), after Proposition 5. The

“Median matching” (Legros and Newman, 2002) also no longer applies, though a generalization does. Specifically, the property that all groups match around a common type in any equilibrium does extend to the general $n \geq 2$ case; however, the type around which all groups match need not be the median. Denote $p_{[x]}$ as the type at the x th quantile; e.g. $p_{[1/2]}$ is the median type.

PROPOSITION 7: *Assume types are substitutes. In any equilibrium grouping, there exists a type $\tilde{p} \in \mathfrak{R}$, such that for every group G in this grouping, $p_1^G \leq \tilde{p} \leq p_n^G$. Assuming a sum-based payoff function, for any number or measure of groups, this type \tilde{p} may be unique and as low as $p_{[1/n]}$ or as high as $p_{[1-1/n]}$.*

In other words, the type that all groups match around may be as low as the $(1/n)$ th quantile and as high as the $(1 - 1/n)$ th quantile. When $n = 2$, this isolates the median type, but the type matched about can come from a widening swath of the distribution as n gets larger.

“Heterogeneous matching” also seems less applicable in the $n > 2$ case. For example, the grouping $\{(p_1, p_2, p_3, p_4, p_{10}), (p_5, p_6, p_7, p_8, p_9)\}$ is arguably quite homogeneous; however, it is intertwined and observable in equilibrium (only) when types are substitutes. The next section uses empirical simulation to demonstrate that type substitutability can give rise to statistically homogeneous matching.

“Negative assortative matching” occurs when $n = 2$ if, the higher an agent’s type, the lower is the type of the agent he matches with. A generalization to $n > 2$ would be that agents with higher types match with other agents having lower *average* type. This generalization does accurately describe the outcome in some of the sum-based payoff settings analyzed in this paper, e.g. those of Propositions 4 and 5 – for, higher types matching with lower average types is a direct consequence of all groups having equal sums of types. However, it is not clear whether this correlation holds in general.

Finally, “intertwined matching” is a new characterization, but it is the aspect of the matching pattern that most clearly generalizes from the $n = 2$ to the $n > 2$ case – certainly as a necessary condition for equilibrium, and in some cases, a sufficient condition for the possibility of equilibrium.

III. Empirical Implications

The theoretical results have empirical implications for groups with more than two members. In particular, they suggest that there is no clear correspondence between homogeneity/heterogeneity of matching and type complementarity/substitutability, or between homogeneity/heterogeneity of matching and positive/negative assortative matching. So many matching patterns are compatible with substitutability (and negative assortative matching) that some of them may appear quite homogeneous – see Section II.

generalized onion-style matching-pattern $\{(1, 2, 7, 8), (3, 4, 5, 6)\}$ would also characterize an equilibrium in this example; but if the lowest three types were changed to $\{1/36, 5/36, 9/39\}$, then $\{(1, 5, 6, 7), (2, 3, 4, 8)\}$ would still characterize an equilibrium while $\{(1, 2, 7, 8), (3, 4, 5, 6)\}$ would not.

The empirical relevance is clear: on the basis of a test for group homogeneity vs. heterogeneity, one cannot draw certain conclusions about type complementarity vs. substitutability, or positive vs. negative assortative matching.

However, much empirical work on group formation is in fact focused on tests for homogeneity versus heterogeneity. This is certainly true of the dyadic regression, a common technique for understanding patterns of group formation.²⁹ In a dyadic regression, the unit of observation is a *pair* of individuals, or “dyad”. Applied to group formation, the dependent variable is typically an indicator for whether both individuals in the pair belong to the same group (“**co-group**”). Independent variables typically capture dis/similarity of key characteristics of individuals in the pair. This allows the data to show whether similarity or dissimilarity in types predicts individuals co-grouping – i.e. whether matching is homogeneous or heterogeneous.³⁰ Often, grouping based on similarity (homogeneous matching) is taken as evidence of positive assortative matching, while grouping based on dissimilarity (heterogeneous matching) is taken as evidence of negative assortative matching. But these conclusions can be mistaken.

Consider the following simulated population of 1500 villages, each containing 50 individuals who form 5 groups of size 10. The distribution of the 50 individual types in each village follows a discrete approximation of³¹

$$(5) \quad F(p) = \begin{cases} 9p & \text{for } p \in [0, 1/10] \\ \frac{p+8}{9} & \text{for } p \in [1/10, 1] \end{cases} .$$

We consider four different matching patterns. First, we assume complementarity of types, and thus the rank-ordered grouping. Second, we assume substitutability and a sum-based group payoff function. One can verify that all pairs of groups fit the grouping-pattern $\{(1, 2, \dots, 9, 20), (10, 11, \dots, 19)\}$ in the unique equilibrium within each village.³² Third, we consider random matching, where each potential grouping is equally likely. Fourth, we consider the *alternating* grouping, in which every pair of groups fits grouping-pattern $\{(1, 3, 5, \dots, 19), (2, 4, 6, \dots, 20)\}$. The random and alternating groupings are included as benchmarks, the alternating grouping since it may reasonably be classified as heterogeneous matching (though it is not intertwined).

A random sample of 10 individuals is drawn from each village, and each individual’s type and group membership under each of the four matching patterns are recorded. Next, dyadic data are created by forming an observation for each pair of sampled individuals from the same village. Since there are 45 ($= \binom{10}{2}$)

²⁹For example, see Fafchamps and Gubert (2007), Gine et al. (2010), Arcand and Fafchamps (2012), Attanasio et al. (2012), and Barr, Dekker, and Fafchamps (2012).

³⁰This technique is applicable for groups of any size, and is often applied to groups larger than two.

³¹That is, $[0, 1/10]$ is partitioned into 45 identical intervals and $[1/10, 1]$ is partitioned into 5 identical intervals; the fifty types are the fifty intervals’ midpoints.

³²That is, the lowest nine types match with the highest type, the second lowest nine types match with the second highest type, and so on.

unique pairings of sampled individuals in each village, and 1500 villages, there are 67,500 dyadic observations. The key dependent variables are d_{ij}^{Cmp} , d_{ij}^{Sub} , d_{ij}^{Rnd} , and d_{ij}^{Alt} , indicators for whether individuals i and j are in the same group, under the four respective groupings: complementarity, substitutability, random, and alternating. The independent variable is $|p_i - p_j|$, the dissimilarity between the types of the individuals in the pair. A positive (negative) coefficient would indicate that dissimilar (similar) types of individuals are more likely to co-group, i.e. that matching is heterogeneous (homogeneous). We estimate the logit model, once each for the four dependent variables, including a constant and clustering standard errors at the village level.³³

The above simulation is repeated 1000 times, and the results are reported in Table 1. Random matching typically produces an estimate not statistically different from zero, as expected. Rank-ordered matching, i.e. complementarity, always produces a negative and statistically significant coefficient, showing that similarity in types predicts co-grouping, i.e. matching is homogeneous. Surprisingly, however, intertwined matching, i.e. substitutability, also always produces a negative and significant coefficient – smaller in magnitude but still substantial. Evidently, substitutability and intertwined matching can also be associated with statistically homogeneous matching. Finally, the alternating grouping, which is neither intertwined nor rank-ordered, nearly always produces a positive and significant coefficient – demonstrating that detecting heterogeneous matching is possible in this context.

TABLE 1 ABOUT HERE

This example makes clear that a negative coefficient in a dyadic group membership regression is not necessarily evidence of positive assortative matching. Here, matching is clearly negative assortative under substitutability, in the sense that the higher one's type, the lower the average type of fellow group members – yet the coefficient is negative.³⁴

More generally, substitutability and intertwining allow for enough possibilities that some of them involve statistically homogeneous matching, in the sense that type similarity predicts co-grouping. Reduced-form empirical techniques that detect homogeneity/heterogeneity of matching cannot identify the underlying forces governing matching, since along this dimension, substitutability can look like complementarity.³⁵

Ways out of this empirical dilemma can be found by testing features of the matching pattern other than group homogeneity/heterogeneity. One could test

³³This corrects for correlated error terms within villages due to the dyadic structure of the data.

³⁴This result becomes apparent only when group size exceeds two. A similar set of simulations with $n = 2$ – where each village's 50 individuals form 25 groups of size 2 – always produces positive coefficients, i.e. heterogeneous matching, under substitutability.

³⁵Again, there is an asymmetry. Substitutability is compatible with both homogeneous and heterogeneous matching, so cannot be ruled in or out. By contrast, complementarity is likely compatible only with homogeneous matching, so a finding of heterogeneous matching may allow complementarity to be ruled out.

the prediction of generalized positive and negative assortative matching: the higher an agent's type, the higher (or lower) the average type in his group. However, it is an open question whether this correlation always holds under intertwined matching. A second strategy is to focus on the predictions of rank-ordering and intertwining, directly comparing pairs of groups in the same matching universe (e.g. village). However, random matching is much more likely to be intertwined than rank-ordered, so the power of this test differs across hypotheses. The most attractive approach seems to be structural estimation of the group payoff function. The maximum score matching estimator proposed by Fox (2010, 2016) could be employed. It requires group payoff functional form assumptions, and allows estimation of the parameters of that function, up to scale. The estimation is based on choosing parameters that most frequently give observed groupings higher payoffs than feasible alternative groupings. Using these estimates, one can check the sign of the cross-partial of the group payoff function, and thus verify whether matching is based on type-complementarity or type-substitutability.³⁶

IV. Discussion and Conclusion

We have characterized equilibrium matching patterns with any fixed group size. In the substitutability case, matching must be intertwined in equilibrium. Conversely, many intertwined matching patterns – in at least one context, *any* intertwined matching pattern – may be the equilibrium under substitutability, depending on the distribution of agent types. As group size grows, complementarity of types continues to make unique predictions, while substitutability by itself allows for a rapidly growing set of potential equilibria. Thus, the focus on group size of two costs significant generality in the case of substitutability, and masks an asymmetry between substitutability and complementarity in predictive power for group formation. This asymmetry means that substitutability can be observationally similar to complementarity on some dimensions, e.g. group homogeneity/heterogeneity, and may mean structural methods are required to identify matching patterns.

Two of our results (Propositions 1 and 3) show what an equilibrium must look like *if* it exists. This is justified in the continuum case, since existence is guaranteed in general by prior results (Kaneko and Wooders, 1996, under approximate feasibility) and, for complementarity, by our Proposition 2. In the finite case, however, there are examples of non-existence, at least for complementarity. However, we also have existence proofs for some finite subcases (building on Sherstyuk's (1999) results), and we do demonstrate existence for all other relevant results (Propositions 2, 4, 5, and³⁷ 7). While fully mapping out existence conditions in the finite case is beyond the scope of this paper, we have demonstrated that existence obtains – and thus our finite-case characterizations are applicable – in

³⁶See Ahlin (2016) for an application of this approach in the microcredit context.

³⁷Proposition 6 is about non-existence rather than existence.

a number of settings.³⁸

Our analysis maintains the assumption that only groups of fixed size n may form. The results apply straightforwardly if group size is mandated by technology, e.g. in certain production team settings. They also directly apply to settings where group size is set by an outside identity, e.g. a microfinance institution.³⁹ The results allow equilibria to be characterized for each n ; combined with further assumptions on the group payoff function, they thus allow the optimal group size to be traced out (Ahlin, 2015). The results are not directly applicable, however, if group size is endogenously determined through the matching process. However, it would be straightforward to show that the necessary matching pattern characterizations of Propositions 1 and 3 must hold for whatever equilibrium group sizes obtain. This suggests that the analysis here of matching with heterogeneous types could potentially be combined with analysis of group stability (e.g., Genicot and Ray, 2003) to jointly endogenize group size and matching patterns. At any rate, while we do not endogenize group size, we view the current work as an important step in understanding matching, and useful to future work with endogenous group size.

³⁸Further, even in cases where an equilibrium grouping does not exist in the finite case, the characterizations of Propositions 1 and 3 still hold for any *efficient* grouping – which is of interest in its own right.

³⁹For example, the Grameen Bank in Bangladesh stipulated that microcredit groups contain five members.

Appendix

Definition. A function $f : D \subseteq \mathfrak{R}^n \rightarrow \mathfrak{R}$ is **supermodular** if for any $x, y \in D$, $f(x \wedge y) + f(x \vee y) \geq f(x) + f(y)$, where $x \wedge y$ and $x \vee y$ denote the component-wise minimum and maximum of x and y , respectively. It is *strictly* supermodular if the inequality is strict for any $x, y \in D$ such that $x \not\leq y$ and $y \not\leq x$. The function f is (strictly) **submodular** if $-f$ is (strictly) supermodular.

Notation. For group G , let $\Pi^G \equiv \Pi(p_1^G, p_2^G, \dots, p_n^G)$.

Proof of Lemma 1. Assume the opposite, that there is a grouping M that is both an equilibrium grouping and does not maximize total surplus. Since M is an equilibrium, we may fix payoffs for each agent that support M as an equilibrium grouping. Since M does not maximize total surplus, another grouping M' exists that produces higher surplus. Then either a) there exists at least one group G in grouping M' whose group payoff is strictly higher than the sum of payoffs of its agents in equilibrium grouping M , or b) there does not exist such a group. But in case a), the n agents in group G could block equilibrium M by organizing themselves into one group and achieving strictly higher payoffs for all; this contradicts M being an equilibrium grouping. And in case b), the sum of all agent payoffs in equilibrium grouping M is at least as high as the sum of all group payoffs in grouping M' , contradicting the fact that M' produces higher total surplus than M . Thus, the hypothesis must be false, implying that any equilibrium grouping maximizes total surplus. ■

Proof of Proposition 1. Suppose that two groups in an equilibrium grouping, L and M , are not rank-ordered. This implies that $p_n^L > p_1^M$ and $p_n^M > p_1^L$ and thus, letting vectors $L' = (p_1^L, p_2^L, \dots, p_n^L)$ and $M' = (p_n^M, p_{n-1}^M, \dots, p_1^M)$, $L' \not\leq M'$ and $M' \not\leq L'$. Then if $L'' = L' \wedge M'$ and $M'' = L' \vee M'$,

$$\Pi^{L''} + \Pi^{M''} > \Pi^{L'} + \Pi^{M'} = \Pi^L + \Pi^M,$$

the inequality from strict supermodularity of Π and the equality from symmetry of Π . Since L'' and M'' represent an alternative, feasible grouping of the $2n$ agents that produces higher total payoffs, this contradicts L and M being equilibrium groups – at least one of the groups L'' and M'' earns more in total for its members than they earned in the original grouping, so all agents in L'' or M'' can be made strictly better off by defecting. Hence, the hypothesis is wrong; any two equilibrium groups are rank-ordered. ■

Proof of Proposition 2. First, we show that any grouping that is not segregated is not an equilibrium. If one equilibrium group, L say, is not homogeneous, then $p_1^L < p_n^L$. Since \mathcal{P} is convex and each type has positive measure, there must be another equilibrium group, M say, with an agent of type $p \in (p_1^L, p_n^L)$. But then L and M are not rank-ordered, contradicting Proposition 1.

Next, we show that the segregated grouping, with agents sharing group payoffs

equally, is an equilibrium. By Corollary 1 of Prat (2002),⁴⁰ given supermodularity of Π , for any (p_1, p_2, \dots, p_n) :

$$\Pi(p_1, p_2, \dots, p_n) \leq \sum_{i=1}^n \frac{\Pi(p_i, \dots, p_i)}{n}.$$

This guarantees that the payoff from any potentially deviating group (the left-hand side) is no greater than the sum of the agents' equilibrium payoffs (the right-hand side). ■

Proof of Proposition 3. Suppose that two groups in an equilibrium grouping, L and M , are neither intertwined nor nearly identical. Since they are not intertwined, one group weakly dominates the other at every rank; that is, relabeling groups if need be, $p_i^L \leq p_i^M$ for all $i \in \mathcal{N}$. Since L and M are not nearly identical, at least two of these inequalities are strict, say for $j, k \in \mathcal{N}$. Let vectors $L' = (p_1^L, \dots, p_{j-1}^L, p_j^M, p_{j+1}^L, \dots, p_n^L)$ and $M' = (p_1^M, \dots, p_{j-1}^M, p_j^L, p_{j+1}^M, \dots, p_n^M)$. Note that $L = L' \wedge M'$ and $M = L' \vee M'$. Also, $M' \not\geq L'$ (since $p_j^L < p_j^M$) and $L' \not\geq M'$ (since $p_k^L < p_k^M$). Thus,

$$\Pi^{L'} + \Pi^{M'} > \Pi^L + \Pi^M,$$

since Π is strictly submodular. Since L' and M' represent an alternative, feasible grouping of the $2n$ agents that produces higher total payoffs, this contradicts L and M being equilibrium groups, as argued in the Proof of Proposition 1. Hence, the hypothesis is wrong; any two equilibrium groups are nearly identical or intertwined. ■

Proof of Corollary 1. Fix any group G . Since there are no mass points in the distribution of types, the maximum measure of groups that can be formed that are nearly identical to G is zero. The result then follows from Proposition 3. ■

Proof of Corollary 2. This follows from Proposition 3, since no two groups can be nearly identical when all agents have distinct types. ■

Proof of Lemma 2. Fix a grouping in which every group has the same sum of types, call it Σ . It suffices to give payoffs for every type that are feasible in this grouping and that deter re-grouping. Given any sum-based payoff function written as in equation (2), let payoffs for an agent of type p_i be

$$a(p_i) = q(p_i) + h(\Sigma)/n + (p_i - \Sigma/n)h'(\Sigma).$$

Feasibility for any equilibrium group G^* requires $\sum_{i \in \mathcal{N}} a(p_i^{G^*}) \leq \Pi^{G^*}$, i.e.

$$\sum_{i \in \mathcal{N}} \left[q(p_i^{G^*}) + h(\Sigma)/n + (p_i^{G^*} - \Sigma/n) h'(\Sigma) \right] \leq \sum_{i \in \mathcal{N}} q(p_i^{G^*}) + h \left(\sum_{i \in \mathcal{N}} p_i^{G^*} \right).$$

⁴⁰See also Meyer and Mookherjee (1987), Proposition 1.

This is easily verified given that $\sum_{i \in \mathcal{N}} p_i^{G^*} = \Sigma$. Next, no set of n agents should be able to re-group to achieve higher payoffs. Sufficient is that for arbitrary group G , $\sum_{i \in \mathcal{N}} a(p_i^G) \geq \Pi^G$, i.e.

$$\sum_{i \in \mathcal{N}} [q(p_i^G) + h(\Sigma)/n + (p_i^G - \Sigma/n) h'(\Sigma)] \geq \sum_{i \in \mathcal{N}} q(p_i^G) + h\left(\sum_{i \in \mathcal{N}} p_i^G\right) .$$

Letting $S^G \equiv \sum_{i \in \mathcal{N}} p_i^G$ be the sum of types in group G , this is equivalent to

$$h(\Sigma) - \Sigma h'(\Sigma) \geq h(S^G) - S^G h'(\Sigma) ,$$

which holds since strict concavity of h (implied by substitutability) guarantees that $h(S^G) - S^G h'(\Sigma)$ is maximized at $S^G = \Sigma$. ■

Proof of Proposition 4. The proof is constructive – it provides a set of $2n$ distinct types, \mathcal{P} , such that the grouping of \mathcal{P} that fits \mathcal{M} has the same sum of types, and is thus an equilibrium (by Lemma 2).

Fix an intertwined grouping-pattern, $\mathcal{M} = \{\mathcal{G}^1, \mathcal{G}^2\}$. Let m_i denote the square root of the i th prime number: $m_1 = \sqrt{2}$, $m_2 = \sqrt{3}$, $m_3 = \sqrt{5}$, and so on.⁴¹ The critical property here of the m_i 's is incommensurability among themselves, i.e. one cannot produce any m_i by rational-coefficient linear combinations of other m_i 's. Let

$$\Sigma_{\mathcal{G}^1} \equiv \sum_{i \in \mathcal{G}^1} m_i , \quad \Sigma_{\mathcal{G}^2} \equiv \sum_{i \in \mathcal{G}^2} m_i , \quad D \equiv \Sigma_{\mathcal{G}^1} - \Sigma_{\mathcal{G}^2} .$$

Without loss, let $\Sigma_{\mathcal{G}^1} \geq \Sigma_{\mathcal{G}^2}$ (relabeling if need be) so that $D \geq 0$.

For $g = 1, 2$, let \mathcal{G}^g be written $\{\mathcal{G}_1^g, \mathcal{G}_2^g, \dots, \mathcal{G}_n^g\}$ with $\mathcal{G}_i^g < \mathcal{G}_{i'}^g$ for $i < i'$. Let i^* be the maximum (reverse) rank at which \mathcal{G}^2 dominates \mathcal{G}^1 : $i^* \equiv \max i \in \mathcal{N}$ such that $\mathcal{G}_i^2 > \mathcal{G}_i^1$. Since \mathcal{G}^1 and \mathcal{G}^2 are intertwined, i^* exists. Note also that $\mathcal{G}_{i^*}^2 = 2i^*$, and for $i \in \mathcal{N}$ and $g \in \{1, 2\}$, $\mathcal{G}_i^g > 2i^*$ only if $i > i^*$. We now define \mathcal{P} : $\mathcal{P} = \{p_1, p_2, \dots, p_{2n}\}$, where

$$p_i = \begin{cases} m_i & \text{if } i < 2i^* \\ m_i + D & \text{if } i \geq 2i^* \end{cases} .$$

Clearly $p_i < p_{i'}$ if $i < i'$. Now,

$$\sum_{i \in \mathcal{G}^2} p_i = \sum_{i \in \mathcal{G}^2} m_i + (n - i^* + 1)D = \Sigma_{\mathcal{G}^2} + (n - i^* + 1)D ,$$

⁴¹The proof is complicated by the goal of ensuring the constructed grouping is the unique way to achieve equal sums, and hence the unique equilibrium. Without this goal, the m_i 's could be replaced with i 's.

because all types of rank i^* and higher in the group corresponding to \mathcal{G}^2 get D added. Similarly,

$$\sum_{i \in \mathcal{G}^1} p_i = \sum_{i \in \mathcal{G}^1} m_i + (n - i^*)D = \Sigma_{\mathcal{G}^1} + (n - i^*)D ,$$

because all types higher than rank i^* in the group corresponding to \mathcal{G}^1 (if any exist, i.e. if $i^* < n$) get D added. Thus,

$$\sum_{i \in \mathcal{G}^1} p_i - \sum_{i \in \mathcal{G}^2} p_i = \Sigma_{\mathcal{G}^1} - \Sigma_{\mathcal{G}^2} - D = 0 .$$

That is, the grouping of \mathcal{P} that fits \mathcal{M} involves two groups with the same sum of types, and is thus an equilibrium by Lemma 2. Now, given an equal number or measure of each of the types in \mathcal{P} , all groups in the grouping with half of the groups fitting \mathcal{G}^1 and half fitting \mathcal{G}^2 have equal sums of types, and by Lemma 2, this is an equilibrium grouping.

The remainder of the proof guarantees uniqueness by showing this is the unique efficient grouping. Using the above expressions with $z \equiv n - i^*$, one derives the sum of types in both groups as

$$\Sigma = \sum_{i \in \mathcal{G}^1} p_i = \sum_{i \in \mathcal{G}^2} p_i = (z + 1) \sum_{i \in \mathcal{G}^1} m_i - z \sum_{i \in \mathcal{G}^2} m_i .$$

Consider arbitrary group G , defined by a function $\phi : \{1, 2, \dots, 2n\} \rightarrow \{0, 1, 2, \dots, n\}$ which gives the number of agents of each type $p_i \in \mathcal{P}$ in the group G . Of course, $\sum_{i=1}^{2n} \phi(i) = n$. Let z' be the number of agents in G with D added, i.e. $z' = \sum_{i=i^*}^{2n} \phi(i)$. The sum of types in group G is

$$S^G = \sum_{i=1}^{2n} \phi(i)m_i + z'D = \sum_{i=1}^{2n} \phi(i)m_i + z' \sum_{i \in \mathcal{G}^1} m_i - z' \sum_{i \in \mathcal{G}^2} m_i .$$

Given incommensurability of the m_i 's, $S^G = \Sigma$ iff the coefficient on each m_i is the same in S^G and Σ . That is, examining the above two equations,

$$S^G = \Sigma \quad \iff \quad \phi(i) = \begin{cases} z + 1 - z' & i \in \mathcal{G}^1 \\ z' - z & i \in \mathcal{G}^2 \end{cases} .$$

Clearly $z' = z$ or $z' = z + 1$ is needed to keep $\phi(i)$ non-negative for all $i \in \mathcal{N}_2$; but if $z' = z$, G is the group that fits \mathcal{G}^1 , and if $z' = z + 1$, G is the group that fits \mathcal{G}^2 . Thus, the only groups that can be assembled from agents with types in \mathcal{P} and that have sum of types equal to Σ are the two groups in the grouping of

\mathcal{P} that fits \mathcal{M} ; call them G^1 and G^2 .

Now consider any grouping M' involving a strictly positive number/measure of groups not equal to G^1 or G^2 . By the uniqueness result of the previous paragraph, these groups have sums of types not equal to Σ , and in fact bounded away by some strictly positive quantity since the set of all n -person groups that can be formed from \mathcal{P} is finite. Given substitutability, which implies concavity of the sum-based payoff function, the sum of group payoffs is higher in the grouping where all groups have equal sums of types than in any grouping where a positive number/measure of groups have sums of types bounded away from Σ (some higher and some lower, since the average sum of types in a group is Σ in any grouping). Thus, grouping M' is less efficient than the equal-sum grouping, and by Lemma 1, not an equilibrium. Thus, any grouping with a positive number/measure of groups not equal to G^1 or G^2 cannot be an equilibrium. Obviously, given equal representation of the types in \mathcal{P} , any grouping involving only G^1 or G^2 must involve an equal number of each. All groupings have thus been ruled out except the replication of the grouping of \mathcal{P} that fits \mathcal{M} . ■

Proof of Lemma 3. For group size n , the total number of grouping-patterns of $\{1, 2, \dots, 2n\}$ is $\binom{2n}{n}/2$. The division by 2 is because the grouping-pattern is the same whether a given group-pattern is labeled \mathcal{G}^1 or \mathcal{G}^2 ; hence $\binom{2n}{n}$ counts each grouping-pattern twice.

To establish the number and fraction of intertwined grouping-patterns as claimed in the Lemma, it suffices to show that the number of non-intertwined grouping-patterns is $\binom{2n}{n}/(n+1)$. Any grouping-pattern can be expressed uniquely in a $2 \times n$ matrix as follows: each group-pattern is placed on a single row in increasing order, with the group-pattern containing 1 in the first row. (Without this normalization, each grouping-pattern has two such matrix expressions.) Clearly, any such grouping-pattern is non-intertwined iff the matrix is monotone increasing going down each column. Thus, the number of non-intertwined grouping-patterns is equal to the number of ways to construct a $2 \times n$ matrix of $2n$ ordered numbers that is monotonically increasing within each row and column. This is the n th Catalan number: $\binom{2n}{n}/(n+1)$. (See Dowling and Shier, 2000, pp. 145-147, especially Example 11.) ■

Proof of Proposition 5. Fix any block-intertwined grouping-pattern $\mathcal{M} = \{\mathcal{G}^1, \mathcal{G}^2\}$; relabel if necessary so that $1 \in \mathcal{G}^1$. Since \mathcal{M} is block-intertwined, there exist $J \in \{2, 3, \dots, n\}$ rank-ordered blocks, such that for any $j \in \{1, 2, \dots, J\}$, there are $2\nu_j$ integers in block j , and the highest ν_j integers in block j belong to one group-pattern and the lowest ν_j integers to the other. Let $\sigma_j = 1$ if \mathcal{G}^2 contains the highest integers from block j , and $\sigma_j = -1$ if \mathcal{G}^2 contains the lowest. It is clear that any block-intertwined grouping-pattern of \mathcal{N}_2 is uniquely identified by J , $\{\nu_j\}_{j=1}^J$, and $\{\sigma_j\}_{j=1}^J$. Define $V_0 \equiv 0$ and $V_j \equiv \sum_{x=1}^j \nu_x$ for $j \in \{1, 2, \dots, J\}$, so

that $V_J = n$. The following quantities will be useful:

$$S_{\nu^2}^{+1} \equiv \sum_{\substack{j=1 \\ \sigma_j=+1}}^J \nu_j^2, \quad S_{\nu}^{+1} \equiv \sum_{\substack{j=1 \\ \sigma_j=+1}}^J \nu_j, \quad S_{\nu^2}^{-1} \equiv \sum_{\substack{j=1 \\ \sigma_j=-1}}^J \nu_j^2, \quad S_{\nu}^{-1} \equiv \sum_{\substack{j=1 \\ \sigma_j=-1}}^J \nu_j;$$

$$\Psi \equiv S_{\nu^2}^{+1} S_{\nu}^{-1} + S_{\nu^2}^{-1} S_{\nu}^{+1};$$

$$b_j \equiv \frac{\nu_j S_{\nu^2}^{-\sigma_j}}{\Psi}, \quad j \in \{1, 2, \dots, J\}; \quad B_0 \equiv 0; \quad B_j \equiv \sum_{x=1}^j b_x, \quad j \in \{1, 2, \dots, J\}.$$

In the **finite case** with kn agents, define $\Delta_j \equiv b_j/(k\nu_j)$ and let agent i 's type satisfy

$$p_i = B_{j-1} + (i - kV_{j-1} - 1/2)\Delta_j,$$

for $i \in \{kV_{j-1} + 1, kV_{j-1} + 2, \dots, kV_j\}, \quad j \in \{1, 2, \dots, J\}.$

One can verify that for a given (block) j , there are $k\nu_j$ evenly spaced, unique types in the interval (B_{j-1}, B_j) ; thus there are kn unique types overall. Also, $p_i < p_{i'}$ if $i < i'$. Now consider the following grouping, where $g \in \{1, 2, \dots, k\}$ indexes groups. Defining

$$x_{jg} = \begin{cases} kV_{j-1} + \nu_j(g-1) & \text{if } \sigma_j = +1 \\ kV_j - \nu_j g & \text{if } \sigma_j = -1 \end{cases} \quad \text{and}$$

(A1)

$$Z_{jg} = \underbrace{[p_{x_{jg}+1}, p_{x_{jg}+2}, \dots, p_{x_{jg}+\nu_j}]}_{1 \times \nu_j}, \quad j \in \{1, 2, \dots, J\},$$

group g is then

(A2)

$$\underbrace{(Z_{1g}, Z_{2g}, \dots, Z_{Jg})}_{1 \times n}.$$

One can check that this is a valid grouping, as each group has ν_j types from interval (B_{j-1}, B_j) , $j \in \{1, 2, \dots, J\}$, and no two groups have any of the same types from this interval. One can also check that every pair of groups $g, g' \in \{1, 2, \dots, k\}$, $g < g'$, uniquely fits grouping-pattern \mathcal{M} . If $\sigma_j = +1$, group 1 has the lowest ν_j types from interval (B_{j-1}, B_j) , group 2 has the next lowest ν_j types, and so on, while if $\sigma_j = -1$, the ordering is reversed. Thus, if $\sigma_j = +1$, group g 's ν_j types in (B_{j-1}, B_j) are all smaller than group g' 's ν_j types in (B_{j-1}, B_j) , and the reverse if $\sigma_j = -1$.

Finally, we show the sum of types in all groups is equal, so that by Lemma 2, this grouping is an equilibrium. Using equations (A1)-(A2), we have that the sum of types in group g is

$$\begin{aligned}
& \sum_{j=1}^J \sum_{m=1}^{\nu_j} p_{x_{jg}+m} = \sum_{\substack{j=1 \\ \sigma_j=+1}}^J \sum_{m=1}^{\nu_j} \{B_{j-1} + \Delta_j[\nu_j(g-1) + m - 1/2]\} \\
& + \sum_{\substack{j=1 \\ \sigma_j=-1}}^J \sum_{m=1}^{\nu_j} \{B_{j-1} + \Delta_j[\nu_j(k-g) + m - 1/2]\} \\
& = \mathcal{X} + \left[\sum_{\substack{j=1 \\ \sigma_j=+1}}^J \sum_{m=1}^{\nu_j} \Delta_j \nu_j g - \sum_{\substack{j=1 \\ \sigma_j=-1}}^J \sum_{m=1}^{\nu_j} \Delta_j \nu_j g \right] \\
& = \mathcal{X} + \frac{g}{k\Psi} \left[S_{\nu^2}^{-1} \sum_{\substack{j=1 \\ \sigma_j=+1}}^J \nu_j^2 - S_{\nu^2}^{+1} \sum_{\substack{j=1 \\ \sigma_j=-1}}^J \nu_j^2 \right] = \mathcal{X},
\end{aligned}$$

where \mathcal{X} is a quantity that does not depend on g , the third equality uses the definition of Δ_j , and the last equality uses the definitions of $S_{\nu^2}^{+1}$ and $S_{\nu^2}^{-1}$ to equate the bracketed term to zero. Thus the sum does not depend on g , i.e. all groups have the same sum of types.

In the **continuum case**, let types be drawn from $\mathcal{P} = [0, 1]$ according to density function

$$f(p) = f_j \equiv \frac{\Psi}{nS_{\nu^2}^{-\sigma_j}} \quad \text{for } p \in [B_{j-1}, B_j], \quad j \in \{1, 2, \dots, J\}.$$

This density function is piecewise flat, taking on one value for segments that will correspond to blocks where \mathcal{G}^1 dominates ($\sigma_j = -1$) and another for segments that will correspond to blocks where \mathcal{G}^2 dominates ($\sigma_j = +1$). Note also that $B_J = 1$; and $b_j f_j = \nu_j/n$, so ν_j/n is the mass of types in $[B_{j-1}, B_j]$. The corresponding distribution function is

$$F(p) = f_j(p - B_{j-1}) + V_{j-1}/n \quad \text{for } p \in [B_{j-1}, B_j], \quad j \in \{1, 2, \dots, J\}.$$

One can check that F is continuous and strictly increasing, and that $F(B_j) = V_j/n$. Letting $p_{[x]}$ denote the type at the x th quantile, we have

$$(A3) \quad p_{[x]} = \frac{x - V_{j-1}/n}{f_j} + B_{j-1} \quad \text{for } x \in [V_{j-1}/n, V_j/n], \quad j \in \{1, 2, \dots, J\},$$

also continuous and strictly increasing.

Now consider the following grouping, where $g \in [0, 1/n]$ indexes groups. Letting

$$(A4) \quad x_{jg} = \begin{cases} V_{j-1}/n + \nu_j g & \text{if } \sigma_j = +1 \\ V_j/n - \nu_j g & \text{if } \sigma_j = -1 \end{cases}$$

and $Z_{jg} = \underbrace{(p_{[x_{jg}]}, \dots, p_{[x_{jg}]})}_{1 \times \nu_j}, \quad j \in \{1, 2, \dots, J\},$

group g is then as is equation (A2).

This is a valid grouping, i.e. the measures of types in groups adds up to the measures of types overall. One can see this since the fraction of types allocated to filling block j across all groups is ν_j/n (clear from equation (A4) because x_{jg} ranges evenly over percentiles $[V_{j-1}/n, V_j/n]$ as g ranges over $[0, 1/n]$); further, this fraction ν_j/n of types is allocated evenly across the ν_j slots in block j (see Z_{jg} in equations (A2) and (A4)), so that each slot in block j is filled with a measure $1/n$ of the total types.

Further, two groups chosen at random from this grouping uniquely fit grouping-pattern \mathcal{M} . Let $g, g' \in [0, 1/n]$, $g \leq g'$, denote two groups chosen at random; with probability one, $g \neq g'$, and $g, g' \in (0, 1/n)$, so consider this case. Note that $x_{jg}, x_{jg'} \in (V_{j-1}/n, V_j/n)$; also note that $x_{jg} < x_{jg'}$ if $\sigma_j = +1$ and $x_{jg} > x_{jg'}$ if $\sigma_j = -1$. Thus, we have the same block structure as \mathcal{M} : blocks are of the same size as in \mathcal{M} , since vector Z_{jg} has length ν_j ; there is rank-ordering across blocks, i.e. $x_{jg}, x_{jg'} < x_{j'g}, x_{j'g'}$ for all $j, j' \in \{1, 2, \dots, J\}$ with $j < j'$; and there is the correct ordering within blocks, i.e. group g has the smallest ν_j types in block j if $\sigma_j = 1$ and the largest if $\sigma_j = -1$.

Finally, the sum of types in all groups is equal, so that by Lemma 2, this grouping is an equilibrium. Using equations (A2)-(A3), we have that the sum of types in group g is

$$\begin{aligned} \sum_{j=1}^J \nu_j p_{[x_{jg}]} &= \sum_{\substack{j=1 \\ \sigma_j=+1}}^J \nu_j \left[\frac{\nu_j g}{f_j} + B_{j-1} \right] + \sum_{\substack{j=1 \\ \sigma_j=-1}}^J \nu_j \left[\frac{\nu_j/n - \nu_j g}{f_j} + B_{j-1} \right] = \\ \mathcal{X} + \sum_{\substack{j=1 \\ \sigma_j=+1}}^J \frac{\nu_j^2 g}{f_j} - \sum_{\substack{j=1 \\ \sigma_j=-1}}^J \frac{\nu_j^2 g}{f_j} &= \mathcal{X} + \frac{gn}{\Psi} \left[S_{\nu^2}^{-1} \sum_{\substack{j=1 \\ \sigma_j=+1}}^J \nu_j^2 - S_{\nu^2}^{+1} \sum_{\substack{j=1 \\ \sigma_j=-1}}^J \nu_j^2 \right] = \mathcal{X}, \end{aligned}$$

where \mathcal{X} is a quantity that does not depend on g , the third equality uses the fact that $b_j f_j = \nu_j/n$ and substitutes in for f_j , and the last equality uses the definitions of $S_{\nu^2}^{+1}$ and $S_{\nu^2}^{-1}$ to equate the bracketed term to zero. Note that the sum does not depend on g , i.e. all groups have the same sum of types. ■

Proof of Lemma 4. Any block-intertwined grouping-pattern is uniquely iden-

tified by a number of blocks $J \in \{2, 3, \dots, n\}$ ($J = 1$ is impossible given intertwining), J sizes for the J blocks, and J dominance indicators for whether the second group-pattern dominates the first in block $j \in \{1, 2, \dots, J\}$. Again, normalize so that the second group-pattern dominates the first in block 1. Note that the number of block-intertwined grouping-patterns with J blocks is the product of the number of ways of assigning the J block sizes (partitioning n pairs of integers into J blocks) times the number of ways of setting the J dominance indicators. On the latter, there are 2^{J-1} ways of setting the dominance indicators for the $J - 1$ blocks other than block 1. All of these ways guarantee intertwining with one exception, the one in which all indicators are the same as block 1's. Thus, there are $2^{J-1} - 1$ intertwined ways of setting the dominance indicators. On the former, imagine the n pairs of integers lined up in order, with $n - 1$ borders between pairs; marking $J - 1$ borders creates a partition of the n pairs into J blocks. Thus, there are $\binom{n-1}{J-1}$ different ways of assigning the J block sizes. Each different way of partitioning the n pairs into J blocks can be paired with each different pattern of the $J - 1$ dominance indicators, so that the total number of block-intertwined grouping-patterns with J blocks is $\binom{n-1}{J-1}(2^{J-1} - 1)$. Summing over the different potential numbers of blocks gives the total number of block-intertwined grouping-patterns as

$$\sum_{J=2}^n \binom{n-1}{J-1} (2^{J-1} - 1) = \sum_{j=1}^{n-1} \binom{n-1}{j} (2^j - 1).$$

■

Proof of Proposition 6. Assume types are substitutes, and fix any $n \geq 3$. Let Ω_k be the set of all sets of kn unique types drawn from $(0, 1)$.

We proceed by induction on k . Let $k = 2$, and fix any intertwined grouping-pattern $\mathcal{M} = \{\mathcal{G}^1, \mathcal{G}^2\}$ of \mathcal{N}_2 ; relabel if necessary so that $1 \in \mathcal{G}^1$. For $g = 1, 2$, let \mathcal{G}^g be written $\{\mathcal{G}_1^g, \mathcal{G}_2^g, \dots, \mathcal{G}_n^g\}$ with $\mathcal{G}_i^g < \mathcal{G}_{i'}^g$ for $i < i'$. We will find a $\mathcal{P} \in \Omega_2$ such that the grouping of \mathcal{P} that fits \mathcal{M} is not efficient. There are three cases to consider. Note that $\mathcal{G}_1^1 = 1$, so in all cases $\mathcal{G}_1^1 < \mathcal{G}_1^2$; the cases differ in which group dominates in the second and third ranks.

Case 1. Assume \mathcal{G}^2 dominates \mathcal{G}^1 at both ranks 1 and 2: $\mathcal{G}_1^1 < \mathcal{G}_1^2$ and $\mathcal{G}_2^1 < \mathcal{G}_2^2$. Then there exists a $i^* \in \{2, \dots, n-1\}$ such that \mathcal{G}^2 dominates \mathcal{G}^1 at ranks $\{1, \dots, i^*\}$, but not at rank i^*+1 . Thus, $\mathcal{G}_i^1 < \mathcal{G}_i^2$ for all $i \in \{1, \dots, i^*\}$, and $\mathcal{G}_{i^*}^2 = 2i^*$. Now fix any set of types $\{p_1, p_2, \dots, p_{2i^*}\}$ such that $0 < p_1 < p_2 < \dots < p_{2i^*} < 1$. By strict submodularity of Π

$$\begin{aligned} & \Pi(p_{\mathcal{G}_1^1}, p_{\mathcal{G}_2^1}, \dots, p_{\mathcal{G}_{i^*}^1}, \underbrace{p_{2i^*}, \dots, p_{2i^*}}_{n-i^*}) + \Pi(p_{\mathcal{G}_1^2}, p_{\mathcal{G}_2^2}, \dots, p_{\mathcal{G}_{i^*}^2}, \underbrace{p_{2i^*}, \dots, p_{2i^*}}_{n-i^*}) < \\ & \Pi(p_{\mathcal{G}_1^2}, p_{\mathcal{G}_2^1}, \dots, p_{\mathcal{G}_{i^*}^1}, \underbrace{p_{2i^*}, \dots, p_{2i^*}}_{n-i^*}) + \Pi(p_{\mathcal{G}_1^1}, p_{\mathcal{G}_2^2}, \dots, p_{\mathcal{G}_{i^*}^2}, \underbrace{p_{2i^*}, \dots, p_{2i^*}}_{n-i^*}). \end{aligned}$$

Hence, by continuity of Π , there exists an $\epsilon > 0$ such that $p_{2i^*} + 2n\epsilon < 1$ and

$$\begin{aligned} & \Pi(p_{\mathcal{G}_1^1}, p_{\mathcal{G}_2^1}, \dots, p_{\mathcal{G}_{i^*}^1}, p_{2i^*} + \epsilon \mathcal{G}_{i^*+1}^1, p_{2i^*} + \epsilon \mathcal{G}_{i^*+2}^1, \dots, p_{2i^*} + \epsilon \mathcal{G}_n^1) + \\ & \Pi(p_{\mathcal{G}_1^2}, p_{\mathcal{G}_2^2}, \dots, p_{\mathcal{G}_{i^*}^2}, p_{2i^*} + \epsilon \mathcal{G}_{i^*+1}^2, p_{2i^*} + \epsilon \mathcal{G}_{i^*+2}^2, \dots, p_{2i^*} + \epsilon \mathcal{G}_n^2) < \\ & \Pi(p_{\mathcal{G}_1^1}, p_{\mathcal{G}_2^1}, \dots, p_{\mathcal{G}_{i^*}^1}, p_{2i^*} + \epsilon \mathcal{G}_{i^*+1}^1, p_{2i^*} + \epsilon \mathcal{G}_{i^*+2}^1, \dots, p_{2i^*} + \epsilon \mathcal{G}_n^1) + \\ & \Pi(p_{\mathcal{G}_1^2}, p_{\mathcal{G}_2^2}, \dots, p_{\mathcal{G}_{i^*}^2}, p_{2i^*} + \epsilon \mathcal{G}_{i^*+1}^2, p_{2i^*} + \epsilon \mathcal{G}_{i^*+2}^2, \dots, p_{2i^*} + \epsilon \mathcal{G}_n^2). \end{aligned}$$

Fix such an ϵ and let $\mathcal{P} = (p_1, p_2, \dots, p_{2i^*}, p_{2i^*} + (2i^* + 1)\epsilon, p_{2i^*} + (2i^* + 2)\epsilon, \dots, p_{2i^*} + 2n\epsilon)$. Clearly $\mathcal{P} \in \Omega_2$. Note that the left-hand side of the above inequality is the sum of payoffs from the grouping of \mathcal{P} that fits \mathcal{M} . (Recall that $\mathcal{G}_{i^*}^1 < \mathcal{G}_{i^*}^2 = 2i^*$.) But the inequality guarantees a different grouping of \mathcal{P} is more efficient.

Case 2. Assume \mathcal{G}^1 dominates \mathcal{G}^2 at ranks 2 and 3: $\mathcal{G}_1^1 < \mathcal{G}_1^2$, $\mathcal{G}_2^1 > \mathcal{G}_2^2$, $\mathcal{G}_3^1 > \mathcal{G}_3^2$. Clearly $\mathcal{G}_1^2 = 2$ and $\mathcal{G}_2^2 = 3$. There exists a $i^* \in \{3, \dots, n\}$, such that \mathcal{G}^1 dominates \mathcal{G}^2 at all ranks $\{2, \dots, i^*\}$, but not at rank $i^* + 1$ (if it exists). Thus, $\mathcal{G}_i^2 < \mathcal{G}_i^1$ for all $i \in \{2, \dots, i^*\}$, and $\mathcal{G}_{i^*}^1 = 2i^*$. Now fix any set of types $\{p_1, p_2, \dots, p_{2i^*}\}$ such that $0 < p_1 < p_2 < \dots < p_{2i^*} < 1$. By strict submodularity of Π ,

$$\begin{aligned} & \Pi(p_1, p_{\mathcal{G}_2^1}, p_{\mathcal{G}_3^1}, \dots, p_{\mathcal{G}_{i^*}^1}, \underbrace{p_{2i^*}, \dots, p_{2i^*}}_{n-i^*}) + \Pi(p_1, p_{\mathcal{G}_2^2}, p_{\mathcal{G}_3^2}, \dots, p_{\mathcal{G}_{i^*}^2}, \underbrace{p_{2i^*}, \dots, p_{2i^*}}_{n-i^*}) < \\ & \Pi(p_1, p_{\mathcal{G}_2^2}, p_{\mathcal{G}_3^2}, \dots, p_{\mathcal{G}_{i^*}^1}, \underbrace{p_{2i^*}, \dots, p_{2i^*}}_{n-i^*}) + \Pi(p_1, p_{\mathcal{G}_2^1}, p_{\mathcal{G}_3^1}, \dots, p_{\mathcal{G}_{i^*}^2}, \underbrace{p_{2i^*}, \dots, p_{2i^*}}_{n-i^*}). \end{aligned}$$

Hence, by continuity of Π , there exists an $\epsilon > 0$ such that $p_{2i^*} + 2n\epsilon < 1$, $p_1 + 2\epsilon < p_3$, and

$$\begin{aligned} & \Pi(p_1 + \epsilon \mathcal{G}_1^1, p_{\mathcal{G}_2^1}, p_{\mathcal{G}_3^1}, \dots, p_{\mathcal{G}_{i^*}^1}, p_{2i^*} + \epsilon \mathcal{G}_{i^*+1}^1, p_{2i^*} + \epsilon \mathcal{G}_{i^*+2}^1, \dots, p_{2i^*} + \epsilon \mathcal{G}_n^1) + \\ & \Pi(p_1 + \epsilon \mathcal{G}_1^2, p_{\mathcal{G}_2^2}, p_{\mathcal{G}_3^2}, \dots, p_{\mathcal{G}_{i^*}^2}, p_{2i^*} + \epsilon \mathcal{G}_{i^*+1}^2, p_{2i^*} + \epsilon \mathcal{G}_{i^*+2}^2, \dots, p_{2i^*} + \epsilon \mathcal{G}_n^2) < \\ & \Pi(p_1 + \epsilon \mathcal{G}_1^1, p_{\mathcal{G}_2^2}, p_{\mathcal{G}_3^2}, \dots, p_{\mathcal{G}_{i^*}^1}, p_{2i^*} + \epsilon \mathcal{G}_{i^*+1}^1, p_{2i^*} + \epsilon \mathcal{G}_{i^*+2}^1, \dots, p_{2i^*} + \epsilon \mathcal{G}_n^1) + \\ & \Pi(p_1 + \epsilon \mathcal{G}_1^2, p_{\mathcal{G}_2^1}, p_{\mathcal{G}_3^1}, \dots, p_{\mathcal{G}_{i^*}^2}, p_{2i^*} + \epsilon \mathcal{G}_{i^*+1}^2, p_{2i^*} + \epsilon \mathcal{G}_{i^*+2}^2, \dots, p_{2i^*} + \epsilon \mathcal{G}_n^2). \end{aligned}$$

Fix such an ϵ and let $\mathcal{P} = (p_1 + \epsilon, p_1 + 2\epsilon, p_3, p_4, \dots, p_{2i^*}, p_{2i^*} + (2i^* + 1)\epsilon, p_{2i^*} + (2i^* + 2)\epsilon, \dots, p_{2i^*} + 2n\epsilon)$. Clearly $\mathcal{P} \in \Omega_2$. Note that the left-hand side of the above inequality is the sum of payoffs from the grouping of \mathcal{P} that fits \mathcal{M} . (Recall that $\mathcal{G}_1^2 = 2$, $\mathcal{G}_2^2 = 3$, and $\mathcal{G}_{i^*}^2 < \mathcal{G}_{i^*}^1 = 2i^*$.) But the inequality guarantees a different grouping of \mathcal{P} is more efficient.

Case 3. Assume \mathcal{G}^1 dominates \mathcal{G}^2 at rank 2, but \mathcal{G}^2 dominates \mathcal{G}^1 at rank 3: $\mathcal{G}_1^1 < \mathcal{G}_1^2$, $\mathcal{G}_2^1 > \mathcal{G}_2^2$, $\mathcal{G}_3^1 < \mathcal{G}_3^2$. Clearly $\mathcal{G}_1^2 = 2$, $\mathcal{G}_2^2 = 3$, $\mathcal{G}_2^1 = 4$, and $\mathcal{G}_3^1 = 5$. There exists a $i^* \in \{3, \dots, n\}$, such that \mathcal{G}^2 dominates \mathcal{G}^1 at all ranks $\{3, \dots, i^*\}$, but not at rank $i^* + 1$ (if it exists). Thus, $\mathcal{G}_i^1 < \mathcal{G}_i^2$ for all $i \in \{3, \dots, i^*\}$, and $\mathcal{G}_{i^*}^2 = 2i^*$. Now fix any set of types $\{p_1, p_2, \dots, p_{2i^*}\}$ such that $0 < p_1 < p_2 < \dots < p_{2i^*} < 1$.

By strict submodularity of Π ,

$$\begin{aligned} & \Pi(p_{\mathcal{G}_1^1}, p_3, p_{\mathcal{G}_3^1}, p_{\mathcal{G}_4^1}, \dots, p_{\mathcal{G}_{i^*}^1}, \underbrace{p_{2i^*}, \dots, p_{2i^*}}_{n-i^*}) + \Pi(p_{\mathcal{G}_1^2}, p_3, p_{\mathcal{G}_3^2}, p_{\mathcal{G}_4^2}, \dots, p_{\mathcal{G}_{i^*}^2}, \underbrace{p_{2i^*}, \dots, p_{2i^*}}_{n-i^*}) < \\ & \Pi(p_{\mathcal{G}_1^1}, p_3, p_{\mathcal{G}_3^2}, p_{\mathcal{G}_4^1}, \dots, p_{\mathcal{G}_{i^*}^1}, \underbrace{p_{2i^*}, \dots, p_{2i^*}}_{n-i^*}) + \Pi(p_{\mathcal{G}_1^2}, p_3, p_{\mathcal{G}_3^1}, p_{\mathcal{G}_4^2}, \dots, p_{\mathcal{G}_{i^*}^2}, \underbrace{p_{2i^*}, \dots, p_{2i^*}}_{n-i^*}) . \end{aligned}$$

Hence, by continuity of Π , there exists an $\epsilon > 0$ such that $p_{2i^*} + 2n\epsilon < 1$, $p_3 + 4\epsilon < p_5$, and

$$\begin{aligned} & \Pi(p_{\mathcal{G}_1^1}, p_3 + \epsilon \mathcal{G}_2^1, p_{\mathcal{G}_3^1}, p_{\mathcal{G}_4^1}, \dots, p_{\mathcal{G}_{i^*}^1}, p_{2i^*} + \epsilon \mathcal{G}_{i^*+1}^1, p_{2i^*} + \epsilon \mathcal{G}_{i^*+2}^1, \dots, p_{2i^*} + \epsilon \mathcal{G}_n^1) + \\ & \Pi(p_{\mathcal{G}_1^2}, p_3 + \epsilon \mathcal{G}_2^2, p_{\mathcal{G}_3^2}, p_{\mathcal{G}_4^2}, \dots, p_{\mathcal{G}_{i^*}^2}, p_{2i^*} + \epsilon \mathcal{G}_{i^*+1}^2, p_{2i^*} + \epsilon \mathcal{G}_{i^*+2}^2, \dots, p_{2i^*} + \epsilon \mathcal{G}_n^2) < \\ & \Pi(p_{\mathcal{G}_1^1}, p_3 + \epsilon \mathcal{G}_2^1, p_{\mathcal{G}_3^2}, p_{\mathcal{G}_4^1}, \dots, p_{\mathcal{G}_{i^*}^1}, p_{2i^*} + \epsilon \mathcal{G}_{i^*+1}^1, p_{2i^*} + \epsilon \mathcal{G}_{i^*+2}^1, \dots, p_{2i^*} + \epsilon \mathcal{G}_n^1) + \\ & \Pi(p_{\mathcal{G}_1^2}, p_3 + \epsilon \mathcal{G}_2^2, p_{\mathcal{G}_3^1}, p_{\mathcal{G}_4^2}, \dots, p_{\mathcal{G}_{i^*}^2}, p_{2i^*} + \epsilon \mathcal{G}_{i^*+1}^2, p_{2i^*} + \epsilon \mathcal{G}_{i^*+2}^2, \dots, p_{2i^*} + \epsilon \mathcal{G}_n^2) . \end{aligned}$$

Fix such an ϵ and let $\mathcal{P} = (p_1, p_2, p_3 + 3\epsilon, p_3 + 4\epsilon, p_5, p_6, \dots, p_{2i^*}, p_{2i^*} + (2i^* + 1)\epsilon, p_{2i^*} + (2i^* + 2)\epsilon, \dots, p_{2i^*} + 2n\epsilon)$. Clearly $\mathcal{P} \in \Omega_2$. Note that the left-hand side of the above inequality is the sum of payoffs from the grouping of \mathcal{P} that fits \mathcal{M} . (Recall that $\mathcal{G}_1^2 = 2$, $\mathcal{G}_2^2 = 3$, $\mathcal{G}_2^1 = 4$, $\mathcal{G}_3^1 = 5$, and $\mathcal{G}_{i^*}^1 < \mathcal{G}_{i^*}^2 = 2i^*$.) But the inequality guarantees a different grouping of \mathcal{P} is more efficient.

Since the three cases are exhaustive, we have thus shown that for arbitrary intertwined grouping-pattern \mathcal{M} of \mathcal{N}_2 , a set of types \mathcal{P} exists such that the grouping of \mathcal{P} corresponding to \mathcal{M} is neither efficient nor (by Lemma 1) an equilibrium. The Proof of Proposition 3 also makes clear that for any $\mathcal{P} \in \Omega_2$, any grouping of \mathcal{P} corresponding to a *non*-intertwined grouping-pattern of \mathcal{N}_2 is neither efficient nor an equilibrium.

For the inductive step, assume for some $k \geq 2$ that for any k -grouping-pattern \mathcal{M} of \mathcal{N}_k , there exists a $\mathcal{P} \in \Omega_k$ such that the grouping of \mathcal{P} that fits \mathcal{M} is not efficient. It is sufficient to show that for any $(k+1)$ -grouping-pattern \mathcal{M} of \mathcal{N}_{k+1} , there exists a $\mathcal{P} \in \Omega_{k+1}$ such that the grouping of \mathcal{P} that fits \mathcal{M} is not efficient and not an equilibrium.

Fix any $(k+1)$ -grouping-pattern of \mathcal{N}_{k+1} , $\mathcal{M} = \{\mathcal{G}^1, \mathcal{G}^2, \dots, \mathcal{G}^k, \mathcal{G}^{k+1}\}$. By the inductive hypothesis, there is a $\tilde{\mathcal{P}} \in \Omega_k$ and a $(k+1)$ -grouping-pattern of \mathcal{N}_{k+1} , $\mathcal{M}' = \{\mathcal{G}'^1, \mathcal{G}'^2, \dots, \mathcal{G}'^k, \mathcal{G}^{k+1}\}$, such that the grouping of $\tilde{\mathcal{P}}$ that fits $\{\mathcal{G}^1, \mathcal{G}^2, \dots, \mathcal{G}^k\}$ produces strictly less than the grouping of $\tilde{\mathcal{P}}$ that fits $\{\mathcal{G}'^1, \mathcal{G}'^2, \dots, \mathcal{G}'^k\}$.⁴² Fix such

⁴²This is a slight abuse of notation: $\{\mathcal{G}^1, \mathcal{G}^2, \dots, \mathcal{G}^k\}$ is not a k -grouping-pattern as we defined it, since it is a collection of kn distinct integers from \mathcal{N}_{k+1} , not from \mathcal{N}_k ; the same is true of $\{\mathcal{G}'^1, \mathcal{G}'^2, \dots, \mathcal{G}'^k\}$, which contains the same kn integers. However, it is straightforward to extend the definition of k -grouping-pattern to refer to any distinct, positive kn integers rather than the first positive kn integers; and the concept of fitting a k -grouping-pattern to be based only on the rankings in the grouping-pattern. This is omitted for brevity.

a $\tilde{\mathcal{P}}$. Since $\tilde{\mathcal{P}}$ contains kn unique types from $(0, 1)$, one may choose a $\hat{\mathcal{P}} \in \Omega_1$, such that $\hat{\mathcal{P}} \cap \tilde{\mathcal{P}} = \emptyset$ and that, relative to $\mathcal{P} \equiv \tilde{\mathcal{P}} \cup \hat{\mathcal{P}}$, the group $\hat{\mathcal{P}}$ fits \mathcal{G}^{k+1} (by adding n new types from $(0, 1)$ into $\tilde{\mathcal{P}}$ such that the ranks of the n added types in the resulting set of types are the ranks from \mathcal{G}^{k+1}). Clearly $\mathcal{P} \in \Omega_{k+1}$. Let M (M') be the grouping of \mathcal{P} that fits \mathcal{M} (\mathcal{M}'). Now M produces strictly less than M' , because the groups $1-k$ in M produce strictly less than groups $1-k$ in M' , as already established, while group $k+1$ produces (and is) the same in both groupings. Thus M is not efficient, and by Lemma 1 not an equilibrium. ■

Proof of Proposition 7. Let M be an equilibrium grouping, $\underline{p} = \sup_{G \in M} p_1^G$, and $\bar{p} = \inf_{G \in M} p_n^G$. If $\underline{p} > \bar{p}$, then there exist groups G_1 and G_2 in M such that $p_n^{G_1} < p_1^{G_2}$; but then G_2 and G_1 are neither intertwined nor nearly identical, contradicting proposition 3. So, $\underline{p} \leq \bar{p}$, and for any $\tilde{p} \in [\underline{p}, \bar{p}]$, every equilibrium group G has $p_1^G \leq \underline{p} \leq \tilde{p}$ and $p_n^G \geq \bar{p} \geq \tilde{p}$.

To show that \tilde{p} may be as low as $p_{[1/n]}$, assume a sum-based payoff function and let there be kn agents: $k-1$ of type p_l , $(n-1)(k-1)$ of type p_h , and n of type p_m , where $p_l < p_m < p_h$ and $p_m = p_h(n-1)/n + p_l/n$. By Lemma 2, it is an equilibrium grouping to have $k-1$ groups each with $n-1$ high types and 1 low type, and 1 group with all medium types. (All groups have the same sum of types, $np_m = (n-1)p_h + p_l$.) The unique value for the type about which all groups match in this grouping is p_m . Note that since p_m is greater than $k-1$ low types and less than $(k-1)(n-1)$ high types, $p_{[x]} = p_m$ for

$$x \in \left[\frac{1 - \frac{1}{k}}{n}, \frac{1 + \frac{n-1}{k}}{n} \right],$$

and thus $p_{[1/n]} = p_m$ for any k . Showing that \tilde{p} may be as high as $p_{[1-1/n]}$ in the finite case is done symmetrically, and the continuum case for both bounds is done analogously with an ϵ -measure of medium types, as $\epsilon \rightarrow 0$. ■

REFERENCES

- Ahlin, Christian.** 2015. "The Role of Group Size in Group Lending." *Journal of Development Economics*, 115: 140–155.
- Ahlin, Christian.** 2016. "Group Lending, Matching Patterns, and the Mystery of Microcredit: Evidence from Thailand." Working Paper.
- Arcand, Jean-Louis, and Marcel Fafchamps.** 2012. "Matching in community-based organizations." *Journal of Development Economics*, 98(2): 203–219.
- Attanasio, Orazio, Abigail Barr, Juan Camilo Cardenas, Garance Genicot, and Costas Meghir.** 2012. "Risk Pooling, Risk Preferences, and Social Networks." *American Economic Journal: Applied Economics*, 4(2): 134–167.
- Barr, Abigail, Marleen Dekker, and Marcel Fafchamps.** 2012. "Who Shares Risk with Whom under Different Enforcement Mechanisms?" *Economic Development and Cultural Change*, 60(4): 677–706.
- Becker, Gary S.** 1973. "A Theory of Marriage: Part I." *Journal of Political Economy*, 81(4): 813–846.
- Chade, Hector, and Jan Eeckhout.** 2014. "Matching Information." Working Paper.
- Chiappori, Pierre-Andre, and Philip J. Reny.** 2016. "Matching to Share Risk." *Theoretical Economics*, 11(1): 227–251.
- Dowling, Thomas A., and Douglas R. Shier.** 2000. "Special Sequences." In *Handbook of Discrete and Combinatorial Mathematics.*, ed. Kenneth H. Rosen, John G. Michaels, Jonathan L. Gross, Jerrold W. Grossman and Douglas R. Shier. CRC Press.
- Durlauf, S. N., and A. Seshadri.** 2003. "Is Assortative Matching Efficient?" *Economic Theory*, 21(2-3): 475–493.
- Echenique, Federico.** 2012. "Contracts versus Salaries in Matching." *American Economic Review*, 102(1): 594–601.
- Fafchamps, Marcel, and Flore Gubert.** 2007. "The Formation of Risk Sharing Networks." *Journal of Development Economics*, 83(2): 326–350.
- Fox, Jeremy T.** 2010. "Identification in matching games." *Quantitative Economics*, 1(2): 203–254.
- Fox, Jeremy T.** 2016. "Estimating Matching Games with Transfers." Working Paper.

- Genicot, Garance, and Debraj Ray.** 2003. "Group Formation in Risk-Sharing Arrangements." *Review of Economic Studies*, 70(1): 87–113.
- Ghatak, Maitreesh.** 1999. "Group Lending, Local Information and Peer Selection." *Journal of Development Economics*, 60(1): 27–50.
- Ghatak, Maitreesh.** 2000. "Screening by the Company you Keep: Joint Liability Lending and the Peer Selection Effect." *Economic Journal*, 110(465): 601–631.
- Giné, Xavier, Pamela Jakiela, Dean Karlan, and Jonathan Morduch.** 2010. "Microfinance Games." *American Economic Journal: Applied Economics*, 2(3): 60–95.
- Grossman, Gene M, and Giovanni Maggi.** 2000. "Diversity and Trade." *American Economic Review*, 90(5): 1255–1275.
- Hatfield, John William, and Paul R. Milgrom.** 2005. "Matching with Contracts." *American Economic Review*, 95(4): 913–935.
- Kaneko, Mamoru, and Myrna Holtz Wooders.** 1986. "The Core of a Game with a Continuum of Players and Finite Coalitions: The Model and Some Results." *Mathematical Social Sciences*, 12(2): 105–137.
- Kaneko, Mamoru, and Myrna Holtz Wooders.** 1996. "The Nonemptiness of the f -Core of a Game without side payments." *International Journal of Game Theory*, 25(2): 245–258.
- Kelso, Jr., Alexander S., and Vincent P. Crawford.** 1982. "Job Matching, Coalition Formation, and Gross Substitutes." *Econometrica*, 50(6): 1483–1504.
- Kremer, Michael.** 1993. "The O-Ring Theory of Economic Development." *Quarterly Journal of Economics*, 108(3): 551–575.
- Legros, Patrick, and Andrew F. Newman.** 2002. "Monotone Matching in Perfect and Imperfect Worlds." *Review of Economic Studies*, 69(4): 925–942.
- Legros, Patrick, and Andrew F. Newman.** 2007. "Beauty is a beast, frog is a prince: Assortative matching with nontransferabilities." *Econometrica*, 75(4): 1073–1102.
- Lin, Y. Joseph.** 1992. "The Becker-Brock efficient matching theorem with a supermodular technology defined on an n -lattice." *Mathematical Social Sciences*, 24: 105–109.
- Meyer, Margaret A., and Dilip Mookherjee.** 1987. "Incentives, Compensation, and Social Welfare." *Review of Economic Studies*, 54(2): 209–226.
- Prat, Andrea.** 2002. "Should a team be homogeneous?" *European Economic Review*, 46(7): 1187–1207.

- Roth, Alvin E., and Marilda A. Oliveira Sotomayor.** 1990. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. New York:Cambridge University Press.
- Saint-Paul, Gilles.** 2001. "On the Distribution of Income and Worker Assignment under Intrafirm Spillovers, with an Application to Ideas and Networks." *Journal of Political Economy*, 109(1): 1–37.
- Schulhofer-Wohl, Sam.** 2006. "Negative Assortative Matching of Risk-Averse Agents with Transferable Expected Utility." *Economics Letters*, 92(3): 383–388.
- Sherstyuk, Katerina.** 1999. "Multisided Matching Games with Complementarities." *International Journal of Game Theory*, 28(4): 489–509.

TABLE .1—DYADIC REGRESSIONS

Village matching population	50
Group size	10
Number of groups per village	5
Village sample size	10
Type distribution	See $F(p)$ in Eq. (5)
d_{ij}^{Cmp} regressed on $ p_i - p_j $	Complementarity
Average coefficient (across simulations)	-1.822
Standard deviation of coefficients	(0.103)
Percent negative and significant at 5 percent	100 percent
d_{ij}^{Sub} regressed on $ p_i - p_j $	Substitutability
Average coefficient (across simulations)	-0.479
Standard deviation of coefficients	(0.056)
Percent negative and significant at 5 percent	100 percent
d_{ij}^{Alt} regressed on $ p_i - p_j $	Alternating
Average coefficient (across simulations)	0.189
Standard deviation of coefficients	(0.038)
Percent positive and significant at 5 percent	99.6 percent
d_{ij}^{Rnd} regressed on $ p_i - p_j $	Random
Average coefficient (across simulations)	-0.001
Standard deviation of coefficients	(0.042)
Percent significant at 5 percent	4.8 percent
Dyadic observations per simulation	67,500
Villages per simulation	1500
Number of simulations	1000

Note: This Table reports author's simulation in which types in each village follow a discrete approximation to the distribution in equation (5). The logit specification is run with a constant and type dissimilarity ($|p_i - p_j|$) as the independent variable. Four different dependent variables are used, indicators for whether individuals i and j in the same village co-group under complementarity (i.e. rank-ordered matching, d_{ij}^{Cmp}), substitutability (d_{ij}^{Sub}), the alternating grouping (d_{ij}^{Alt}), and random matching (d_{ij}^{Rnd}), respectively. For each specification, we report the average estimated coefficient on $|p_i - p_j|$ across 1000 simulations, the standard deviation of these 1000 estimates, and the percent of these estimates that are significant at the 5 percent level based on standard errors clustered at the village level.