Group Formation with Fixed Group Size: Complementarity vs Substitutability

Christian Ahlin*

March 2014

Abstract

Matching under transferable utility is well understood when groups of size $n = 2$ are being formed: Complementarity or substitutability of types in the group payoff function pins down key features of the matching pattern, whatever the distribution of types or specifics of the payoff function. But little is known about one-sided matching in the often more realistic case of groups with $n > 2$ members being formed. This subject is taken up here. Type-complementarity continues to rule out all but one grouping. Type-substitutability – which occurs, for example, in matching to share risk – rules out much less. We show it requires that in equilibrium, every two groups must be "intertwined", in that each dominates the other at some rank. Intertwined matching is necessary and, in one context we provide, sufficient for any grouping to be an equilibrium for some set of types; thus intertwined matching is all that substitutability generically predicts. But, the number of intertwined matching patterns increases rapidly in n. Thus, substitutability by itself has much less predictive power than complementarity when $n > 2$. One implication is that substitutability can be observationally similar to complementarity using common empirical techniques that detect homogeneity/heterogeneity of matching. We demonstrate this through simulated dyadic regressions, showing that statistically homogeneous groups are observable under intertwined (negative assortative) matching.

*I thank Ashley Ahlin, Alfredo Burlando, Pierre-Andre Chiappori, Vivek Dhand, Sergei Izmalkov, Patrick Legros, Andrew Newman, Debraj Ray, Katerina Sherstyuk, Myrna Wooders, and seminar participants at ThReD-Oslo, New Economic School, and University of Oregon for valuable input. All errors are mine.

1 Introduction

Models of frictionless, one-sided matching under transferable utility have been used in various contexts, often with agent types being complements or substitutes. Examples include skill matching and underdevelopment (Kremer, 1993, complementarity); skill diversity and trade patterns (Grossman and Maggi, 2000, complementarity and substitutability); microcredit group formation (Ghatak, 1999, 2000, and Ahlin, 2013, complementarity and substitutability); and matching to share risk (Schulhofer-Wohl, 2006, substitutability[1]).

Some stark results on matching have emerged from this and related literature (see especially Legros and Newman, 2002). If the group payoff function exhibits type complementarity (or supermodularity), the unique matching equilibrium involves perfectly homogeneous matching (or "segregation"): all agents in a group have the same type. If the group payoff function exhibits type substitutability (or submodularity), 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, complementarity and substitutability pin down key aspects of the matching pattern for any distribution of types and payoff function details.

However, most of the literature has stayed within the confines of two-person matches.[2] It would clearly be helpful to know how the matching results generalize to larger groups. Size-two groups are often counterfactual – e.g., often firms 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 left to apply results on two-person matches to data

[1] He focuses on the two-sided case, but results are easily extendable to the one-sided, two-person case.

[2] For complementarity, there are a few significant exceptions. Legros and Newman (2002) state that several of their results on homogeneous matching generalize to larger group size. Durlauf and Seshadri (2003) characterize *efficient* one-sided group formation under complementarity and arbitrary fixed group size. In the case of n-sided matching, the classic positive assortative matching result for (2-sided) marriage of Becker (1973) has been generalized by Lin (1992) and Sherstyuk (1999).

For substitutability, the lone exception we are aware of is Saint-Paul (2001), who characterizes matching patterns under complementarity, substitutability, and hybrid cases, with fixed-*measure* group size. His paper differs from both this paper and the other literature referenced here in that he examines groups with a continuum of members, while here the focus is on finite-member groups. See conclusion for more discussion.

featuring larger groups, without knowing how valid this approach is.

An illustrative example is matching for risk-sharing. The research on how households that are heterogeneous in risk preferences match in order to share risk (Schulhofer-Wohl, 2006; Legros and Newman, 2007; Chiappori and Reny, 2006)[3] 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 explores matching into groups of fixed size $n \geq 2$. We find that complementarity dictates matching in the same way when $n = 2$ as when $n > 2$: in any equilibrium, groups are rank-ordered, and perfectly homogeneous if there is a continuum of agents. These predictions hold regardless of the distribution of agent types.

For substitutability, however, 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 that in equilibrium, groups will be "intertwined": in any pair of groups, each group dominates the other at some rank (or is nearly identical to it). Intertwined matching is not only necessary, it can be sufficient: we provide one context in which any intertwined grouping is the (unique) equilibrium for some configuration of types, and a broader context in which a large subset of intertwined groupings can be equilibria. Since there can be many intertwined groupings, and moreso for larger n, the results make clear that substitutability by itself has significantly less predictive power than complementarity – in general, equilibrium matching patterns under substitutability depend on the distribution of types, when $n > 2$. This asymmetry between complementarity and substitutability in predictive power only appears when $n > 2$.

The results are significant theoretically for several reasons. They provide a first characterization of matching for risk-sharing when groups have more than two members. In particular, the substitutability results apply directly to a one-sided modification of the Schulhofer-Wohl

[3]Genicot and Ray (2003) address a complementary question, the stable size of risk-sharing groups composed of ex ante identical households.

(2006) model, and characterize a set of matchings 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. Hence, at least using some common empirical techniques, substitutability may be hard to rule out, while complementarity may be hard to rule in.

We illustrate this by simulations featuring dyadic regressions, which are often used to understand group formation patterns.[4] Dyadic regressions focus on whether groups are homogeneous or heterogeneous by type. When $n = 2$ there is a tight connection between group homogeneity (heterogeneity) and positive (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, similarity in characteristics can predict matching 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.

The baseline model and complementarity results are in section 2. Substitutability results are in section 3. Section 4 presents simulation results and empirical implications. Section 5 provides discussion and concluding remarks. All proofs are in the Appendix.

[4]For example, see Fafchamps and Gubert (2007), Attanasio et al. (2012), Arcand and Fafchamps (2012), Barr et al. (2012), and Gine et al. (2010).

2 Model and Complementarity

A set of agents match 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 $\Re$. 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. We often focus on one of two cases:

Case F: there is a finite number of agents equal to kn, for some integer $k \geq 2$.

Case C: there is 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, $\Pi : \mathcal{P}^n \to \Re$, is assumed twice continuously differentiable and symmetric, meaning invariant to any permutation of n types. Types are said to be **complements** (**substitutes**) if the group payoff function exhibits strictly positive (strictly negative) cross-partials everywhere on its domain. Given twice continuous differentiability, type complementarity (substitutability) is equivalent to strict supermodularity (submodularity) of Π.[5] Agent utility is assumed 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 $G = (p_1^G, p_2^G, ..., p_n^G) \in \mathcal{P}^n$; equivalently, it can be represented as a multiset with n elements drawn from $\mathcal{P}$.[6] Throughout the paper,

[5]Super/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.

[6]A 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.

the type indices in any vector or multiset of types will be understood to correspond to rank-ordering; for example, $p_1^G \leq p_2^G \leq ... \leq p_n^G$ in any group G.

A **grouping** is a set of groups in which all agents 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 total number or measure of agents of each type. An **equilibrium** (core) **grouping** is a grouping in which payoffs exist for each agent that a) are feasible, meaning the sum of agent payoffs in each group does not exceed the group payoff; and b) cannot be blocked by any n agents reorganizing into a group so that each achieves a strictly higher payoff.[7]

We say that two *groups* L and M are **rank-ordered** if $p_n^L \leq p_1^M$ or $p_n^M \leq p_1^L$, and 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 each group 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.[8] This is the result of supermodularity of the payoff function, which implies that payoffs can always be raised by taking the maximum n types and the minimum n types from any two groups. Continually applying this fact leaves the rank-ordered grouping as uniquely efficient; and any equilibrium grouping must be efficient.[9]

When there is a continuum of agents, rank-ordering is squeezed to perfect group homogeneity, i.e. segregation. To prove existence, we first establish the following lemma.

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

[8] This is true of any equilibrium grouping, but an equilibrium grouping may not exist in the finite case. While supermodularity guarantees existence in the n-sided matching case (Proposition 2 of Sherstyuk, 1999), it is not sufficient here. In fact, we have *both* examples of non-existence *and* existence proofs for some settings. Thus, existence in the finite, complementarity case depends on specifics of the set of agent types and the payoff function. See further discussion of existence in the Conclusion.

[9] Proposition 1 of Durlauf and Seshadri (2003) uses similar reasoning and derives similar conclusions about efficient groupings under supermodularity.

Lemma 1. *If $f : \Re^n \to \Re$ is symmetric and supermodular, then for any $(x_1, x_2, ..., x_n) \in \Re^n$,*

$$f(x_1, x_2, ..., x_n) \leq \frac{\sum_{i=1}^{n} f(x_i, ..., x_i)}{n}.$$

Proposition 2. *Assume types are complements. In the case of a continuum of agents (case C), the unique equilibrium grouping is segregated.*

This result is familiar, in the sense of being 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 essentially Folk Theorems included for completeness, though the existence proof of Proposition 2 is new and relies explicitly on supermodularity.[10]

3 Substitutability

Grossman and Maggi (2000) show that in the $n = 2$ case when types are substitutes, what might be called onion-style matching occurs. Every group has one member above and one member below the median, both equidistant from it in percentile terms. If $p_{[x]}$ denotes the type at the xth quantile, then every group is of the form $(p_{[x]}, p_{[1-x]})$. Schulhofer-Wohl (2006), Legros and Newman (2007), and Chiappori and Reny (2006) show a similar pattern obtains in two-sided matching for risk-sharing. Thus for $n = 2$, substitutability gives similarly strong results, pinning down the matching pattern regardless of the type distribution and production function. What happens when $n > 2$?

3.1 Necessity of "Intertwined" Matching

For any integer $k \geq 1$, let $\mathcal{N}_k \equiv \{1, 2, ..., kn\}$, so that $\mathcal{N}_1 = \{1, 2, ..., n\}$.

[10]Formal equilibrium characterizations of one-sided matching exist for $n = 2$; we have not found explicit characterizations for $n > 2$, though existing results are stated to be easily extendable, e.g. Legros and Newman (2002) (see also Lin, 1992). The results of Kaneko and Wooders (1996) also establish equilibrium existence in this context, of an "approximately feasible" core.

Say that two *groups* L and M are **intertwined** if there exist $i, i' \in \mathcal{N}_1$ such that L strictly dominates M at rank i ($p_i^L > p_i^M$) and M strictly dominates L at rank i' ($p_{i'}^M > p_{i'}^L$). 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 $i^* \in \mathcal{N}_1$ such that $p_i^L = p_i^M$ for $i \in \mathcal{N}_1 \setminus \{i^*\}$.[11]

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, in that each group dominates the other at some rank. The only possible exception occurs when the two groups are almost identical.

The reasoning is Proposition 1's in reverse. Submodularity (implied by substitutability) and symmetry imply that any two groups that can be written as the element-by-element maximum and minimum (after permutation), respectively, of two other groups can be rearranged into these other groups so as to raise payoffs. In turn, iff two groups are not intertwined can they be written as the maximum and minimum of two other groups (after permutation). The exception is where the two groups are so similar – nearly identical – that the above "rearrangement" leaves them unchanged.

The result applies for all $n \geq 2$, but consider the simplest example of $n = 2$ and four agents with unique types: $p_1 < p_2 < p_3 < p_4$. By Proposition 1, complementarity rules out two of the three ($= \binom{4}{2}/2$) possible groupings, leaving only the rank-ordered one: $\{(p_1, p_2), (p_3, p_4)\}$. By Proposition 3, substitutability also rules out two groupings, in which

[11] Groups $L = (1, 2, 5, 7)$ and $M = (1, x, 4, 7)$ are intertwined iff $2 < x < 5$. (If $x < 2$, L dominates M at every rank; the reverse is true if $x > 5$. If $x = 2$ or $x = 5$, L and M are nearly identical.) L is nearly identical to $M' = (1, x, 5, 7)$ iff $1 \leq x \leq 5$. (For example, if $5 < x \leq 7$, then $p_2^{M'} = 5 \neq 2 = p_2^L$ and $p_3^{M'} = x \neq 5 = p_3^L$.)

one group rank-wise dominates the other:

$$\begin{array}{ccc} \xcancel{\begin{matrix}(p_3,p_4)\\(p_1,p_2)\end{matrix}} & \xcancel{\begin{matrix}(p_2,p_4)\\(p_1,p_3)\end{matrix}} & \begin{matrix}(p_2,p_3)\\(p_1,p_4)\end{matrix} \end{array} \quad .$$

The only possible equilibrium grouping is the onion-style one, since only there are the two groups intertwined. In fact, onion-style matching is equivalent to the intertwined grouping when there are $2k$ distinct types forming k groups of size 2.

But consider $n = 3$ and six agents with unique types: $p_1 < p_2 < p_3 < p_4 < p_5 < p_6$. Complementarity rules out nine of the ten $(= \binom{6}{3}/2)$ groupings, again leaving only the rank-ordered one: $\{(p_1, p_2, p_3), (p_4, p_5, p_6)\}$. However, the necessity of intertwined matching rules out only five groupings, where one group rank-wise dominates the other:

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

The groupings not ruled out (A–E) are all intertwined, i.e. each group dominates the other at some rank. Evidently, when $n = 3$ intertwining is less restrictive than rank-ordering.

The following corollaries show that nearly identical groups, which are not ruled out by Proposition 3, do not occur or are the exceptional case in our 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) is fully intertwined, if no two agents have the same type.*

3.2 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? Here we provide a setting in which any intertwined grouping can be the unique equilibrium – guaranteeing that nothing beyond intertwined matching is generically implied by substitutability.

Consider the following setting. There is an equal number (in the finite case) or measure (in the continuum case) of agents of each of $2n$ unique types. The payoff function can be written

$$\Pi(p_1, p_2, ..., p_n) = \sum_{i \in \mathcal{N}_1} q(p_i) + h\left(\sum_{i \in \mathcal{N}_1} p_i\right) \tag{2}$$

for some differentiable function q and twice-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;[12] 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 proved by providing and checking equilibrium payoffs for each agent type.

The following notation enables us to differentiate the grouping patterns – defined by the ranks only – from the actual values of the p_i's. Let $\mathcal{G}$ be a **group-pattern** from $\mathcal{N}_k = \{1, 2, ..., kn\}$ if $\mathcal{G}$ is an n-element subset of $\mathcal{N}_k$.[13] Call a set of k group-patterns $\mathcal{M} \equiv \{\mathcal{G}^1, \mathcal{G}^2, ..., \mathcal{G}^k\}$ a **grouping-pattern** of $\mathcal{N}_k$ if the $\mathcal{G}^i$'s are disjoint group-patterns from $\mathcal{N}_k$.[14] The intertwining property applies to group- and grouping-patterns in the same way it does to groups and groupings (as does block-intertwining, defined below).

[12] This is essentially the class of payoff functions that Saint-Paul (2001) analyzes.

[13] As with groups, group-patterns can also be represented as length-n vectors.

[14] For example, let $n = 3$. Then $(1, 4, 6)$ and $(1, 3, 5)$ are group-patterns of $\mathcal{N}_2$; $\{(1, 4, 6), (2, 3, 5)\}$ and $\{(1, 3, 5), (2, 4, 6)\}$ are grouping-patterns of $\mathcal{N}_2$, the former intertwined, the latter not.

Given a multiset of kn agent types, $P = \{p_1^P, p_2^P, ..., p_{kn}^P\}$,[15] a **grouping of** P is a grouping of kn agents with types given by a one-to-one mapping from P. We will say a *group* G **corresponds in** P **to** group-pattern $\mathcal{G} = (\mathcal{G}_1, \mathcal{G}_2, ..., \mathcal{G}_n)$ if $G = (p_{\mathcal{G}_1}^P, p_{\mathcal{G}_2}^P, ..., p_{\mathcal{G}_n}^P)$. In other words, G corresponds in P to $\mathcal{G}$ if its types are drawn from P based on the ranks in $\mathcal{G}$.[16] We say k *groups*, $\{G^1, G^2, ..., G^k\}$, **correspond to** a grouping-pattern of $\mathcal{N}_k$, $\{\mathcal{G}^1, \mathcal{G}^2, ..., \mathcal{G}^k\}$, if (for some permutation of the G^is' indices) group G^i corresponds in $\uplus_{i=1}^k G^i$ to group-pattern $\mathcal{G}^i$ for $i \in \{1, 2, ..., k\}$.[17]

Proposition 4. *Assume a sum-based payoff function and that types are substitutes. Fix any intertwined grouping-pattern $\{\mathcal{G}^1, \mathcal{G}^2\}$ of $\mathcal{N}_2$. There exists a $\mathcal{P} = \{p_1, p_2, ..., p_{2n}\}$ with $p_i < p_{i'}$ for $i < i'$, such that the unique equilibrium grouping of an equal number or measure of agents of each type in $\mathcal{P}$ involves half of the groups corresponding in $\mathcal{P}$ to $\mathcal{G}^1$ and the other half corresponding in $\mathcal{P}$ to $\mathcal{G}^2$.*

Thus, in this context and for any $n \geq 2$, any intertwined grouping of $2n$ agents can be the unique equilibrium. As an example, consider the $n = 3$ case outlined above (expression 1). Proposition 4 guarantees that with a sum-based payoff function, any of the five intertwined groupings (A-E) is the unique equilibrium grouping of six agents for some set of types $\{p_1, p_2, ..., p_6\}$. Generalizing to an equal number or measure of each of the six types, Proposition 4 guarantees that any of these five groupings can form the basis for the unique equilibrium grouping – in that half of the groups are identical to each of the two groups.

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 – at least for

[15] Recall that $p_1^P \leq p_2^P \leq ... \leq p_{kn}^P$ is understood.

[16] For example, if $n = 3$ and $P = \{p_1, p_2, ..., p_6\}$, group (p_1, p_4, p_6) corresponds in P to group-pattern $(1, 4, 6)$. Clearly, for any $n \geq 2$, if $P = \{p_1, p_2, ..., p_{2n}\}$ and $p_i < p_{i'}$ for $i < i'$, $\mathcal{G}^1$ and $\mathcal{G}^2$ are intertwined iff the groups corresponding in P to $\mathcal{G}^1$ and $\mathcal{G}^2$ are intertwined.

[17] The operation $\uplus$ is a generalized union for multisets that includes all instances of all types from the united multisets in the resulting multiset. For example, $\{1, 1, 2\} \uplus \{2, 3\} = \{1, 1, 2, 2, 3\}$.

If $n = 3$ and $P = \{p_1, p_2, ..., p_9\}$, groups $\{(p_1, p_4, p_6), (p_2, p_3, p_5)\}$ correspond to $\mathcal{N}_2$ grouping-pattern $\{(1, 4, 6), (2, 3, 5)\}$; so do groups $\{(p_1, p_6, p_9), (p_3, p_4, p_7)\}$ (with $P' \equiv \{p_1, p_6, p_9\} \uplus \{p_3, p_4, p_7\} = \{p_1, p_3, p_4, p_6, p_7, p_9\}$, note that $(p_1, p_6, p_9) = (p_1^{P'}, p_4^{P'}, p_6^{P'})$ and $(p_3, p_4, p_7) = (p_2^{P'}, p_3^{P'}, p_5^{P'})$). The grouping $\{(p_1, p_6, p_9), (p_2, p_5, p_8), (p_3, p_4, p_7)\}$ corresponds to $\mathcal{N}_3$ grouping-pattern $\{(1, 6, 9), (2, 5, 8), (3, 4, 7)\}$.

the case of transferable utility and $2n$ types. Adapting the standard setup of Schulhofer-Wohl (2006) to our context,[18] one can show that utility is transferable, the risk-sharing group payoff function is sum-based, and agent types (risk aversion) are substitutes. Thus, Proposition 4 applies, and for the type structure analyzed, any intertwined grouping of the $2n$ types may form the basis for a unique equilibrium grouping.

More generally, under substitutability, intertwined matching is both *necessary* for a grouping to be an equilibrium (Proposition 3), and at least in one setting (the $2n$-agent subcase of Case F), *sufficient* for a grouping to be the equilibrium under some configuration of types (Proposition 4). In short, intertwined matching is all that is generically implied by substitutability.

The possibility remains, of course, that substitutability implies more than intertwined matching in cases other than that of $2n$ uniquely-typed agents. 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.[19] However, we are able to identify a subset of intertwined grouping-patterns that constitute 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) the blocks are rank-ordered; 2) for each $j \in \{1, ..., J\}$, block j contains an even number of types, $2\nu_j$ say; and 3) for each $j \in \{1, ..., J\}$, group L contains either the ν_j largest types or the ν_j smallest types

[18] Schulhofer-Wohl's results on utility transferability, submodularity, and dependence only on the sum of types extend to the one-sided, n-person matching setting of this paper, under similar assumptions; details available on request.

[19] Proposition 4's proof constructs, given arbitrary grouping-pattern $(\mathcal{G}^1, \mathcal{G}^2)$, a set $\mathcal{P}$ of $2n$ types such that the sum of types is the same in groups G^1 and G^2 corresponding in $\mathcal{P}$ to $\mathcal{G}^1$ and $\mathcal{G}^2$; this guarantees (G^1, G^2) is an equilibrium grouping of $\mathcal{P}$. Such a set of types does not always exist in the general case where there are kn uniquely-typed agents forming k groups of size n. (For example, linear algebra results can be used to show that no set of unique types gives equal sums to the groups in the intertwined grouping $\{(p_1, p_4, p_7, p_{10}), (p_2, p_3, p_{11}, p_{12}), (p_5, p_6, p_8, p_9)\}$.) Of course, failure of this proof technique need not imply the statement is false, and we know of no example of an intertwined grouping that can never be an equilibrium.

in block j. When $n = 2$ or $n = 3$, any two intertwined groups are also block-intertwined.[20] However, when $n \geq 4$, not all pairs of intertwined groups are block-intertwined.[21]

Proposition 5. *Assume a sum-based payoff function and that types are substitutes. Fix any block-intertwined grouping-pattern* $\{\mathcal{G}^1, \mathcal{G}^2\}$ *of* $\mathcal{N}_2$. *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 correspond to* $\{\mathcal{G}^1, \mathcal{G}^2\}$. *And for any* $k \geq 2$, *there exists a set* $\mathcal{P} = \{p_1, p_2, ..., p_{kn}\}$ *with* $0 < p_1 < p_2 < ... < p_{kn} < 1$, *and an equilibrium grouping of* $\mathcal{P}$ *in which every pair of groups uniquely corresponds to* $\{\mathcal{G}^1, \mathcal{G}^2\}$.

That is, a subset of intertwined grouping-patterns of $\mathcal{N}_2$ – all *block*-intertwined ones – form the basis for potential equilibria for any finite number or continuum of groups. The importance of this result is that it guarantees that the sufficiency of at least one form of intertwining is not an artefact of the two-group case, but survives even when there are many types, groups, and agents.

For example, consider $n = 3$ and block-intertwined grouping-pattern $\{(1, 2, 6), (3, 4, 5)\}$ (to which corresponds grouping A in expression 1). Let types be taken from $\mathcal{P} = [0, 1]$ according to distribution $F(p) = 2p$ for $p \in [0, 1/3]$ and $F(p) = (1 + p)/2$ for $p \in [1/3, 1]$. Let $p_{[x]}$ denote the type at the xth percentile of the distribution ($p_{[x]} = F^{-1}(x)$), let $g \in [0, 1/3]$ index groups, and consider the grouping $\{(p_{[2g]}, p_{[2g]}, p_{[1-g]})\}$. First, note that any two groups g, g' uniquely correspond to the given grouping-pattern with probability one. Second, note that the measure of types "adds up", since the types used to fill each of the

[20] For example, consider the groupings of 6 uniquely-typed agents into groups of size 3 (expression 1), and let L be the group that contains p_1. For grouping A, the block structure is $J = 2$, $\nu_1 = 2$, $\nu_2 = 1$, and L has the ν_1 smallest types from block 1 and the ν_2 largest types from block 2. Grouping E involves $J = 2$, $\nu_1 = 1$, $\nu_2 = 2$, and L has the ν_1 smallest types from block 1 and the ν_2 largest types from block 2. Groupings B-D all involve $J = 3$, $\nu_1 = \nu_2 = \nu_3 = 1$, and L has the smallest type from block 1; the only difference is in whether group L contains the largest type in block 2, block 3, or both.

[21] For example, there are 21 intertwined groupings of 8 uniquely-typed agents into 2 groups of size 4, of which 19 are block-intertwined. The two intertwined groupings that are not block-intertwined are $\{(p_1, p_2, p_4, p_8), (p_3, p_5, p_6, p_7)\}$ and $\{(p_1, p_5, p_7, p_8), (p_2, p_3, p_4, p_6)\}$. (General count formulas are provided in the next section.)

first two slots has measure 1/3 (each slot gets half of the lowest 2/3), and the last slot is filled with types of measure 1/3 (the highest 1/3). Third, one can verify that the sum of types in all groups is the same, since group g is of the form $(g, g, 1-2g)$. Hence, under substitutability and a sum-based payoff function, this grouping is an equilibrium. Consider also the grouping $\{(p_{[g]}, p_{[1/3+g]}, p_{[1-g]})\}$, $g \in [0, 1/3]$. Note that virtually every pair of groups uniquely corresponds to the same grouping-pattern as grouping B of expression 1, and that the measure of types adds up. One can also verify that if $F(p) = 4p/3$ for $p \in [0, 1/2]$ and $F(p) = (1+2p)/3$ for $p \in [1/2, 1]$, then the sum of types in all groups is the same, so again, this grouping is an equilibrium. The same can be achieved for groupings C-E in expression 1, and more generally, the Proof of Proposition 5 shows how to construct the appropriate grouping and distribution function (Case C) or set of types (Case F) to achieve these goals for any $n \geq 2$ and any block-intertwined grouping-pattern of $\mathcal{N}_2$.

Thus, under a sum-based payoff function and substitutability, any block-intertwined grouping-pattern of $\mathcal{N}_2$ can form the basis for an equilibrium of any number of agents kn, or even of a continuum, depending on the distribution of types. The question remains open whether some intertwined groupings that do not fit this block-intertwined pattern can be ruled out by substitutability in some setting.

3.3 Predictive power of Substitutability

There are clearly contexts in which substitutability has significantly weaker predictive power than complementarity. This section provides a more detailed comparison.

Our results ruling out and ruling in various groupings revolve around intertwined and block-intertwined pairs of groups, i.e. grouping-patterns of $\mathcal{N}_2 = \{1, 2, ..., 2n\}$. A first step is to count how many grouping-patterns of $\mathcal{N}_2$ are intertwined and block-intertwined.

Lemma 3. *There are $\binom{2n}{n}/2$ total grouping-patterns of $\mathcal{N}_2$, $\frac{n-1}{n+1}\binom{2n}{n}/2$ intertwined grouping-patterns of $\mathcal{N}_2$, and $\sum_{j=1}^{n-1}\binom{n-1}{j}(2^j-1)$ block-intertwined grouping-patterns of $\mathcal{N}_2$. Thus, a fraction $\frac{n-1}{n+1}$ of all grouping-patterns of $\mathcal{N}_2$ are intertwined.*

That is, the number of total groupings, intertwined groupings, and block-intertwined groupings of $2n$ uniquely-typed agents all increase rapidly in n (all more than double from n to $n+1$). Further, the larger is n, the smaller the fraction of groupings of $2n$ unique types the intertwined matching requirement marks out as unobservable in equilibrium ($\frac{2}{n+1}$). In the limit, the fraction of groupings ruled out vanishes.

Combining this result with Proposition 4 gives that under substitutability, a sum-based payoff function, and $2n$ types of equal representation, a fraction $\frac{n-1}{n+1}$ of the groupings of the $2n$ types can form the basis for a unique equilibrium. That is, in this setting almost anything goes, and increasingly so for larger n. Applying this again to risk-sharing group formation, we see that if $n \geq 5$, most groupings of $2n$ distinct risk-types form the basis for a potential unique equilibrium grouping of an equal representation of each of the $2n$ types.

Regarding the general case of kn agents of unique types forming k groups of size n, a count formula for intertwined groupings akin to Lemma 3's seems elusive.[22] However, we conjecture (and computation suggests) that the fraction of intertwined groupings with n fixed decreases in k and approaches 0, and with k fixed increases in n and approaches 1. This is intuitive, since having more groups raises the number of pairs of groups that must be intertwined, while having larger groups increases the likelihood any pair of groups is intertwined. Thus, the predictive power of intertwined matching likely depends to some degree on the balance between the number of groups and the size of groups.

Still, even if the *fraction* of (block-)intertwined groupings goes to zero as the number of types and groups increases, the *number* of (block-)intertwined groupings can still be quite large. Lemma 3 establishes that the number of block-intertwined grouping-patterns of $\mathcal{N}_2$ grows rapidly with n, exceeding 19,000 when $n = 10$. Thus Proposition 5 implies that, given substitutability, sum-based payoffs, and $n = 10$ – e.g. in matching to share risk – there are more than 19,000 potential equilibrium groupings, and more than 19,000 potential grouping-patterns for two groups chosen at random – for any number of groups, even a continuum.

[22]Lemma 3 is proved using Catalan number arguments. The k-group case could be solved with a k-dimensional generalization of the Catalan numbers that appears not to exist.

From an empirical perspective, the lack in predictive power from having a large *number* of possible groupings may present significant challenges, even if the *fraction* of permissible groupings is low.

In sum, these results uncover a stark contrast between substitutability and complementarity. Complementarity rules out all but one grouping regardless of n and k, as does substitutability when $n = 2$. However, when $n > 2$ and at least for some payoff functions, substitutability allows many possible matching patterns in equilibrium, and moreso as n increases. Only in moving away from $n = 2$ does the asymmetry in the predictive power of complementarity and substitutability emerge.[23,24]

Note that all the results on sufficiency of (block-)intertwining have been obtained assuming sum-based payoff functions. However, we are able to discard the assumption of a sum-based payoff function to show that the following is a general property of substitutability in the finite case when $n > 2$: there is no single grouping that is efficient or an equilibrium independent of the type distribution.

Proposition 6. *Assume types are substitutes. Fix any number of groups* $k \geq 2$*, group size* $n \geq 3$*, and grouping-pattern* $\mathcal{M}$ *of* $\mathcal{N}_k$*. There exists a set* $\mathcal{P}$ *of* kn *unique types drawn from* $[0, 1]$*, such that the grouping of* $\mathcal{P}$ *corresponding to* $\mathcal{M}$ *is not efficient and not an equilibrium.*

Thus, in general no one grouping is efficient or an equilibrium for all sets of types, under substitutability and $n > 2$.

[23] None of the results are about multiplicity of equilibrium given a distribution of types – they say that different equilibria may be realized under different distributions of types. However, multiplicity of equilibrium given a single distribution of types is also possible under substitutability. For example, if $n = 4$ and types are distributed uniformly on $[0, 1]$, the two groupings indexed by $g \in [0, 1/4]$, $\{p_{[g]}, p_{[1/4+g]}, p_{[3/4-g]}, p_{[1-g]}\}$ and $\{p_{[g]}, p_{[1/2-g]}, p_{[3/4-g]}, p_{[3/4+g]}\}$, are both equilibria – and there are more. Almost every pair of groups in the former grouping uniquely corresponds to $\{(1, 3, 6, 8), (2, 4, 5, 7)\}$, and in the latter grouping to $\{(1, 4, 6, 7), (2, 3, 5, 8)\}$.

[24] We have shown a certain kind of intertwined grouping, based on all pairs of groups corresponding to the same block-intertwined grouping-pattern of $\mathcal{N}_2$, to be a possible equilibrium even in the continuum. There can also be equilibrium groupings in the continuum in which positive measures of pairs of groups correspond to *different* block-intertwined grouping-patterns. For example, if $n = 3$ and types are distributed uniformly on $[0, 1]$, the grouping indexed by $g \in [0, 1/3]$, $\{p_{[3g]}, p_{[1/2-3g/2]}, p_{[1-3g/2]}\}$, is an equilibrium. A positive measure of pairs of groups uniquely corresponds to $\{(1, 4, 6), (2, 3, 5)\}$ (e.g. $g, g' \in (0, 1/9)$), to $\{(1, 4, 5), (2, 3, 6)\}$ (e.g. $g, g' \in (1/9, 2/9)$), and to $\{(1, 3, 6), (2, 4, 5)\}$ (e.g. $g, g' \in (2/9, 1/3)$), respectively.

The intuition for $k = 2$ is that, when $n > 2$, every intertwined grouping contains at least two ranks in which one group dominates the other. But, a set of types exists for which these are the most important ranks, so that this grouping is arbitrarily close to a non-intertwined grouping, and a better grouping can be found. To illustrate, consider the groupings in expression 1. If $\mathcal{P}$ is modified so that $p_5 = p_6$, A and B are not intertwined and by submodularity produce lower payoffs than C and D; if $\mathcal{P}$ is modified so that $p_1 = p_2$, D and E are not intertwined and produce lower payoffs than B and C; and if $\mathcal{P}$ is modified so that $p_3 = p_4$, C is not intertwined and produces lower payoffs than B and D. By continuity, in each case inferiority of the grouping is robust to the types assumed equal being slightly different instead.

This result is in stark contrast to complementarity, where the rank-ordered grouping-pattern is always uniquely efficient and the unique equilibrium (if an equilibrium exists). Under substitutability, however, there is a dependence on the distribution of types regardless of the details of the payoff function. In sum, it appears that the matching pattern under complementarity is robustly distribution-free, while the matching-pattern under substitutability when $n > 2$ is robustly distribution-dependent.

3.4 Terminology for Substitutability

The terminology for complementarity – rank-ordered matching, segregation, homogeneous matching, positive assortative matching – applies equally well when $n = 2$ as when $n > 2$, since the grouping patterns generalize in the obvious way. The same is not true for substitutability.

"Onion-style matching" is appropriate under substitutability when $n = 2$, but inadequate for the $n > 2$ case. Hypothetically, some generalization of onion-style matching might have held, e.g. with the maximum and minimum type in each group following an onion-style pattern – as in groupings A, B, D, and E of expression 1. But grouping C is not onion-style even in this weak sense, since one group has the lowest type and the other the highest type.

And as we know from Proposition 5, grouping C can be replicated in equilibrium for any number of groups, even a continuum, e.g. $g \in [0, 1/3]$ and $\{(p_{[g]}, p_{[2/3-g]}, p_{[2/3+g]})\}$.

"Median matching" also applies to substitutability when $n = 2$ (Legros and Newman, 2002), since every group has one type above and one type below the median. This property of all groups matching around a common type does extend to the general $n \geq 2$ case; however, the type(s) around which all groups match need not include the median.

Proposition 7. *Assume types are substitutes. In any equilibrium grouping, there exists a type $\widetilde{p} \in \Re$, such that for every group G in this grouping, $p_1^G \leq \widetilde{p} \leq p_n^G$. For any number or measure of groups, this type $\widetilde{p}$ 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. Thus, "median matching" does not extend beyond $n = 2$.

"Heterogeneous matching" also has been applied to matching under substitutability, but seems potentially inapplicable in the $n > 2$ case. For example, should the grouping $\{(p_1, p_2, p_3, p_4, p_{10}), (p_5, p_6, p_7, p_8, p_9)\}$ be considered heterogeneously or homogeneously matched? It is arguably quite homogeneous; however, it is intertwined and thus observable (only) when types are substitutes. The next section uses empirical simulation to demonstrate that homogeneous and heterogeneous matching do not always correspond to types being complements and substitutes, respectively.

"Negative assortative matching" when $n = 2$ occurs 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 generalized negative assortative matching 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 kind of pattern holds in general.

In the end, "intertwined matching" appears to be the best terminology to describe matching when types are substitutes, since intertwined matching is necessary, and at least in some cases sufficient, for any grouping to be an equilibrium under some configuration of types.

4 Empirical Implications

The theoretical results have empirical implications for groups with more than two members. Since substitutability of types can be compatible with so many different matching patterns, it may be hard to rule out empirically. Equivalently, complementarity can be hard to establish empirically, since there may be matching patterns that are "close" to rank-ordered but observable under substitutability. These points especially apply to reduced-form approaches that test for homogeneity vs. heterogeneity in matching – as discussed in Section 3.4, it is not clear that one can identify intertwined matching with heterogeneous matching and rank-ordered matching with homogeneous matching.

Consider one common technique for understanding patterns of group formation, the dyadic regression.[25] In a dyadic regression, the unit of observation is a *pair* of individuals. 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. 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

[25] For example, see Fafchamps and Gubert (2007), Attanasio et al. (2012), Arcand and Fafchamps (2012), Barr et al. (2012), and Gine et al. (2010).

follows a discrete approximation of[26]

$$F(p) = \begin{cases} 9p & \text{for } p \in [0, 1/10] \\ \frac{p+8}{9} & \text{for } p \in [1/10, 1] \end{cases} . \tag{3}$$

We consider four alternative 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 – e.g. matching to share risk. One can verify that given the assumed set of types, in the unique equilibrium within each village, all pairs of groups correspond to $\{(1, 2, ..., 9, 20), (10, 11, ..., 19)\}$. Third, we consider random matching, where each potential grouping is equally likely. Fourth, we consider the *alternating* grouping, where the types of ranks $\{1, 6, ..., 46\}$ join one group, types of ranks $\{2, 7, ..., 47\}$ join another, and so on. 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).

We draw a random sample of 10 individuals from each village, recording each individual's type and group membership under each of the four matching patterns. We next transform this to dyadic data by forming an observation for each pair of sampled individuals from the same village. Since there are 45 ($= \binom{10}{2}$) 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. We estimate the logit model, once each for the four dependent variables, including a constant and clustering standard errors at

[26] That 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.

the village level.[27]

The above simulation is repeated 1000 times, and the results are reported in Table 1, first column. 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. 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 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

Thus, substitutability and intertwining allow for enough possibilities that some of them involve homogeneous matching, in the sense that type similarity predicts co-grouping. 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.[28]

This set of simulations assumes a specific distribution of types in all villages which, by construction, always gives rise to the same equilibrium grouping patterns under substitutability. A second set of simulations makes a similar point with i.i.d. lognormal types, where equilibrium grouping patterns can and do differ across villages. Types are i.i.d. within and across villages, following $p_i = e^{2Z_i}$, Z_i standard normal. Each village contains 20 individuals, forming 2 groups[29] of size 10, and we sample both entire groups in each village.

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

[28] This distinction 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. The same is true of the next set of simulations also.

[29] We reduce the number of groups here for tractability – given the i.i.d. draws in each village, grouping

TABLE 1 – Dyadic Regressions

Village matching population	50	20
Village sample size	10	20
Group size	10	10
Number of groups	5	2
Type distribution	See $F(p)$ in Eq. 3	Lognormal, i.i.d.: $p_i = e^{2Z_i}$, $Z_i \sim N(0,1)$
d_{ij}^{Cmp} regressed on $\lvert p_i - p_j \rvert$	Complementarity	
Average coefficient (across simulations)	-1.822	-0.06386
Standard deviation of coefficients	(0.103)	(0.02377)
Percent **negative** and significant at 5%	100%	84.4%
d_{ij}^{Sub} regressed on $\lvert p_i - p_j \rvert$	Substitutability	
Average coefficient (across simulations)	-0.479	-0.00169
Standard deviation of coefficients	(0.056)	(0.00081)
Percent **negative** and significant at 5%	100%	82.0%
d_{ij}^{Alt} regressed on $\lvert p_i - p_j \rvert$	Alternating	
Average coefficient (across simulations)	0.189	0.00921
Standard deviation of coefficients	(0.038)	(0.00406)
Percent positive and significant at 5%	99.6%	84.7%
d_{ij}^{Rnd} regressed on $\lvert p_i - p_j \rvert$	Random	
Average coefficient (across simulations)	-0.001	-0.00001
Standard deviation of coefficients	(0.042)	(0.00038)
Percent significant at 5%	4.8%	2.2%
Dyadic observations per simulation	67,500	285,000
Villages per simulation	1500	1500
Number of simulations	1000	1000

Note: The two columns report two sets of simulations, one in which types in each village follow a discrete approximation to the distribution in equation 3, the other based on the lognormal distribution. In each, logit specifications are run with type dissimilarity ($|p_i - p_j|$) and a constant as the independent variables. 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% level based on standard errors clustered at the village level. Estimates in the second column are in standard deviation (of p) units.

We find similar results; see Table 1, second column. Typically, random matching produces a coefficient indistinguishable from zero, alternating matching produces a significantly positive coefficient, and complementarity and substitutability both produce significantly negative coefficients. Coefficients are smaller in these simulations (though they are in standard deviation units), and the average coefficient for substitutability is about one fortieth the size of the one for complementarity. However, the frequency of finding statistically significant negative coefficients is about the same under both complementarity and substitutability. Thus, while the evidence for positive assortative matching might not be overwhelming, one might conclude there is sufficient evidence to reject both random matching and negative assortative matching – even though matching is in fact negative assortative, or at least intertwined and driven by substitutability.

These examples[30] make clear that reduced form empirical techniques that detect homogeneity/heterogeneity of matching do not always identify the underlying forces governing matching, since along this dimension, substitutability can look like complementarity and be compatible with coefficients of any sign. However, though these techniques may not be able to rule out substitutability, in some cases – when type dissimilarity predicts co-grouping – they may be able to rule out complementarity, since its predictions are so stark.

One alternative empirical strategy is to focus on 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 appears different for different potential outcomes. A third

patterns differ across villages, and time required to find the equilibrium grouping rises rapidly with number of groups. To illustrate, there are nearly 100,000 groupings of 20 individuals into 2 groups, but nearly 1 trillion groupings of 30 individuals into 3 groups.

[30] Of course, many other examples give more expected outcomes. The examples here are meant to show possibilities, and the impossibility of firmly drawing certain conclusions based on this empirical approach.

approach is structural estimation of the group payoff function. The maximum score matching estimator proposed by Fox (2010a,b) is one such technique. 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 verify whether matching is based on type-complementarity or type-substitutability. It is clear that in the second set of simulations above, this procedure would be able to identify the complementarity vs. substitutability of the group payoff function, because identification would be based on whether the observed groupings are minimizing or maximizing the difference in sum of types across groups, which is exactly how the equilibrium groupings are chosen.

5 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. when using dyadic regressions to assess group homogeneity, and may mean structural methods are required to identify matching patterns.

It is worth comparing these results to earlier work by Saint-Paul (2001). He examines group formation where each group has a fixed-*measure continuum* of agents, under complementarity, substitutability, and hybrid cases. Given that groups are so "large", a simple solution in the substitutability case is to put all types in each group, so that all groups are identical and mimic the overall distribution of types – trivially creating an intertwined grouping. A key distinction between his work and ours (and between his work and other contributions we follow, including Grossman and Maggi, 2000, and Legros and Newman, 2002) is that we look at finite-size groups. This includes the central case where the number of types is greater than group size, and hence groups must differ. Our results contribute to understanding how matching patterns generalize in this setting.

Some of our results 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 do have existence proofs for some finite subcases (building on Sherstyuk's (1999) results), and we do demonstrate existence for our main sufficiency results (Propositions 4 and 5). While fully mapping out existence conditions in the finite case is beyond the scope of this paper, we have demonstrated that our finite-case characterizations are applicable in a number of settings.[31]

Though we consider group size to be fixed (along with the vast two-sided matching literature), it will often be a choice variable or an equilibrium outcome. The results here provide building blocks for understanding matching patterns, for any given group size. If group size is set by an outside identity, e.g. a microfinance institution, then further assumptions on the payoff function could allow the optimal size to be traced out (Ahlin, 2013). If group size is determined through the matching process, the analysis here of matching with heterogeneous types could potentially be combined with analysis of group stability (Genicot and Ray, 2003) to provide a richer model of group formation. At any rate, we view the current work as an important step in understanding matching, and likely to be useful to future work

[31]Further, even if an equilibrium grouping does not exist in the finite case, the characterizations still hold for any *efficient* grouping – which is of interest in its own right.

with endogenous group size.

References

[1] Christian Ahlin. The role of group size in group lending. Working Paper, December 2013.

[2] Jean-Louis Arcand and Marcel Fafchamps. Matching in community-based organizations. *Journal of Development Economics*, 98(2):203–219, July 2012.

[3] Orazio Attanasio, Abigail Barr, Juan Camilo Cardenas, Garance Genicot, and Costas Meghir. Risk pooling, risk preferences, and social networks. *American Economic Journal: Applied Economics*, 4(2):134–167, April 2012.

[4] Abigail Barr, Marleen Dekker, and Marcel Fafchamps. Who shares risk with whom under different enforcement mechanisms? *Economic Development and Cultural Change*, 60(4):677–706, 2012.

[5] Gary S. Becker. A theory of marriage: Part I. *Journal of Political Economy*, 81(4):813–846, July 1973.

[6] Pierre-Andre Chiappori and Philip J. Reny. Matching to share risk. Working Paper, May 2006.

[7] Thomas A. Dowling and Douglas R. Shier. Special sequences. In Kenneth H. Rosen, John G. Michaels, Jonathan L. Gross, Jerrold W. Grossman, and Douglas R. Shier, editors, *Handbook of Discrete and Combinatorial Mathematics*. CRC Press, 2000.

[8] S. N. Durlauf and A. Seshadri. Is assortative matching efficient? *Economic Theory*, 21(2-3):475–493, March 2003.

[9] Marcel Fafchamps and Flore Gubert. The formation of risk sharing networks. *Journal of Development Economics*, 83(2):326–350, July 2007.

[10] Jeremy T. Fox. Identification in matching games. *Quantitative Economics*, 1(2):203–254, November 2010a.

[11] Jeremy T. Fox. Estimating matching games with transfers. Working Paper, December 2010b.

[12] Garance Genicot and Debraj Ray. Group formation in risk-sharing arrangements. *Review of Economic Studies*, 70(1):87–113, January 2003.

[13] Maitreesh Ghatak. Group lending, local information and peer selection. *Journal of Development Economics*, 60(1):27–50, October 1999.

[14] Maitreesh Ghatak. Screening by the company you keep: Joint liability lending and the peer selection effect. *Economic Journal*, 110(465):601–631, July 2000.

[15] Xavier Giné, Pamela Jakiela, Dean Karlan, and Jonathan Morduch. Microfinance games. *American Economic Journal: Applied Economics*, 2(3):60–95, July 2010.

[16] Gene M Grossman and Giovanni Maggi. Diversity and trade. *American Economic Review*, 90(5):1255–1275, December 2000.

[17] Mamoru Kaneko and Myrna Holtz Wooders. 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, October 1986.

[18] Mamoru Kaneko and Myrna Holtz Wooders. The nonemptiness of the f-Core of a game without side payments. *International Journal of Game Theory*, 25(2):245–258, 1996.

[19] Michael Kremer. The O-Ring theory of economic development. *Quarterly Journal of Economics*, 108(3):551–575, August 1993.

[20] Patrick Legros and Andrew F. Newman. Monotone matching in perfect and imperfect worlds. *Review of Economic Studies*, 69(4):925–942, October 2002.

[21] Patrick Legros and Andrew F. Newman. Beauty is a beast, frog is a prince: Assortative matching with nontransferabilities. *Econometrica*, 75(4):1073–1102, July 2007.

[22] Y. Joseph Lin. The Becker-Brock efficient matching theorem with a supermodular technology defined on an n-lattice. *Mathematical Social Sciences*, 24:105–109, 1992.

[23] Gilles Saint-Paul. 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, February 2001.

[24] Sam Schulhofer-Wohl. Negative assortative matching of risk-averse agents with transferable expected utility. *Economics Letters*, 92(3):383–388, September 2006.

[25] Katerina Sherstyuk. Multisided matching games with complementarities. *International Journal of Game Theory*, 28(4):489–509, November 1999.

Appendix

Definition. A function $f : D \subseteq \Re^n \to \Re$ 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 $x \ngeq y$ and $y \ngeq x$. The function f is (strictly) **submodular** if $-f$ is (strictly) supermodular.

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

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 $L' = (p_1^L, p_2^L, ..., p_n^L)$ and $M' = (p_n^M, p_{n-1}^M, ..., p_1^M)$, $L' \ngeq M'$ and $M' \ngeq L'$. 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 Lemma 1. The claim is clearly true when $n = 2$, for

$$2f(x_1, x_2) = f(x_1, x_2) + f(x_2, x_1) \leq f(x_1, x_1) + f(x_2, x_2),$$

the equality by symmetry and the inequality by supermodularity of f.

Now suppose the claim holds for any symmetric and supermodular $f : \Re^n \to \Re$ and any $(x_1, x_2, ..., x_n) \in \Re^n$. By induction, it remains to show that if $g : \Re^{n+1} \to \Re$ is symmetric and supermodular, then for any $(x_1, x_2, ..., x_{n+1}) \in \Re^{n+1}$,

$$(n+1)\ g(x_1, x_2, ..., x_{n+1}) \leq \sum_{i=1}^{n+1} g(x_i, ..., x_i) \quad .$$

Fix $(X_1, X_2, ..., X_{n+1}) \in \Re^{n+1}$ and a symmetric and supermodular function $g : \Re^{n+1} \to \Re$.

First, note that $\tilde{G} : \Re^n \to \Re$ defined as $\tilde{G}(x_1, x_2, ..., x_n) \equiv g(x_1, x_2, ..., x_n, X_{n+1})$ satisfies symmetry and supermodularity, since g does. The hypothesis then gives that

$$n\ \tilde{G}(X_1, X_2, ..., X_n) \leq \sum_{i=1}^{n} \tilde{G}(X_i, ..., X_i).$$

Replacing $\tilde{G}$ with g and multiplying by $(n+1)/n$ gives

$$(n+1)\ g(X_1, X_2, ..., X_{n+1}) \leq \frac{1}{n} \sum_{i=1}^{n} (n+1)\ g(X_i, ..., X_i, X_{n+1}).$$

Next, fix an $i \in \mathcal{N}_1$ and note that there are $(n+1)$ terms in the right-hand side sum of the form $g(X_i, ..., X_i, X_{n+1})$. By symmetry of g, we can permute the arguments of these $(n+1)$ terms so that X_{n+1} appears in a different position in each term (1st, 2nd, ..., $(n+1)$st), without changing the sum's value. Finally, we can iteratively apply the supermodularity condition to pairs of these $(n+1)$ terms, always choosing pairs in which X_{n+1} is an argument, to (weakly) increase the sum. After n iterations there is one term of the form $g(X_{n+1}, ..., X_{n+1})$ and n of the form $g(X_i, ..., X_i)$. This holds for every $i \in \mathcal{N}_1$, so the above right-hand side term satisfies

$$\leq \frac{1}{n}\sum_{i=1}^{n}\left[n\ g(X_i, ..., X_i) + g(X_{n+1}, ..., X_{n+1})\right] = \sum_{i=1}^{n+1} g(X_i, ..., X_i).$$

Since $(X_1, X_2, ..., X_{n+1})$ was arbitrary, this establishes the hypothesis for $(n+1)$.

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 Lemma 1, given supermodularity of Π, for any $(p_1, p_2, ..., p_n)$:

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

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 $i \in \mathcal{N}_1$. Since L and M are not nearly identical, at least two of these inequalities are strict, say for $j, k \in \mathcal{N}_1$. Let vectors $L' = (p_1^L, ..., p_{j-1}^L, p_j^M, p_{j+1}^L, ..., p_n^L)$ and $M' = (p_1^M, ..., p_{j-1}^M, p_j^L, p_{j+1}^M, ..., p_n^M)$. Note that $L = L' \wedge M'$ and $M = L' \vee M'$. Also, $M' \ngeq L'$ (since $p_j^L < p_j^M$) and $L' \ngeq M'$ (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 $\sum_{i\in\mathcal{N}_1} a(p_i^{G^*}) \leq \Pi^{G^*}$, i.e.

$$\sum_{i\in\mathcal{N}_1} \left[q(p_i^{G^*}) + h(\Sigma)/n + \left(p_i^{G^*} - \Sigma/n\right) h'(\Sigma)\right] \leq \sum_{i\in\mathcal{N}_1} q(p_i^{G^*}) + h\left(\sum_{i\in\mathcal{N}_1} p_i^{G^*}\right) \ .$$

This is easily verified given that $\sum_{i\in\mathcal{N}_1} p_i^{G^*} = \Sigma$. Next, no set of n agents should be able to re-group to achieve higher payoffs. Sufficient is that for arbitrary group G, $\sum_{i\in\mathcal{N}_1} a(p_i^G) \geq \Pi^G$, i.e.

$$\sum_{i\in\mathcal{N}_1} \left[q(p_i^{G}) + h(\Sigma)/n + \left(p_i^{G} - \Sigma/n\right) h'(\Sigma)\right] \geq \sum_{i\in\mathcal{N}_1} q(p_i^{G}) + h\left(\sum_{i\in\mathcal{N}_1} p_i^{G}\right) \ .$$

Letting $S^G \equiv \sum_{i\in\mathcal{N}_1} p_i^G$ 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 $2n$ types $\mathcal{P}$ such that the groups corresponding in $\mathcal{P}$ to $\mathcal{G}^1$ and $\mathcal{G}^2$ have the same sum of types, so that the hypothesized grouping is an equilibrium (by Lemma 2).

Fix an intertwined grouping-pattern of $\mathcal{N}_2$, $\{\mathcal{G}^1, \mathcal{G}^2\}$. Let m_i denote the square root of the ith prime number: $m_1 = \sqrt{2}$, $m_2 = \sqrt{3}$, $m_3 = \sqrt{5}$, and so on.[32] 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

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

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

[32]The 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.

For $g = 1, 2$, let $\mathcal{G}^g$ be written $\{\mathcal{G}^g_1, \mathcal{G}^g_2, ..., \mathcal{G}^g_n\}$ with $\mathcal{G}^g_i < \mathcal{G}^g_{i'}$ for $i < i'$. Let i^* be the maximum (reverse) rank at which $\mathcal{G}^2$ dominates $\mathcal{G}^1$: $i^* \equiv \max i \in \mathcal{N}_1$ such that $\mathcal{G}^2_i > \mathcal{G}^1_i$. Since $\mathcal{G}^1$ and $\mathcal{G}^2$ are intertwined, i^* exists. Note also that $\mathcal{G}^2_{i^*} = 2i^*$, and for $i \in \mathcal{N}_1$ and $g \in \{1, 2\}$, $\mathcal{G}^g_i > 2i^*$ only if $i > i^*$. We now define $\mathcal{P}$: $\mathcal{P} = \{p_1, p_2, ..., p_{2n}\}$, where

$$p_i = \begin{cases} m_i & \text{if } i < 2i^* \\ m_i + D & \text{if } i \geq 2i^* \end{cases} \quad . \tag{4}$$

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

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

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 group corresponding in $\mathcal{P}$ to $\mathcal{G}^1$ has the same sum of types as the group corresponding in $\mathcal{P}$ to $\mathcal{G}^2$. 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 corresponding in $\mathcal{P}$ to $\mathcal{G}^1$ and half to $\mathcal{G}^2$ 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 $\phi_i : \mathcal{N}_2 \to \{0, 1, 2, ..., n\}$ 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,

$$S^G = \Sigma \qquad \Longleftrightarrow \qquad \phi_i = \begin{cases} z + 1 - z' & i \in \mathcal{G}^1 \\ z' - z & i \in \mathcal{G}^2 \end{cases} \quad .$$

Clearly $z' = z$ or $z' = z+1$ is needed to keep ϕ_i non-negative for all $i \in \mathcal{N}_2$; but if $z' = z$, G is the group corresponding in $\mathcal{P}$ to $\mathcal{G}^1$, and if $z' = z+1$, G is the group corresponding in $\mathcal{P}$ to $\mathcal{G}^2$. 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 ones corresponding in $\mathcal{P}$ to $\mathcal{G}^1$ and $\mathcal{G}^2$.

Now consider any grouping M involving a strictly positive number/measure of groups that do not correspond in $\mathcal{P}$ to $\mathcal{G}^1$ or $\mathcal{G}^2$. The previous paragraph establishes that 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 not an equilibrium – n agents exist that can achieve higher payoffs by forming a new group belonging to an efficient grouping. Thus, any grouping with a positive number/measure of groups corresponding in $\mathcal{P}$ to neither $\mathcal{G}^1$ nor $\mathcal{G}^2$ cannot be an equilibrium. Given equal representation of the types in $\mathcal{P}$, there are no groupings where a fraction $\alpha \neq 1/2$ of groups correspond in $\mathcal{P}$ to $\mathcal{G}^1$ and $1-\alpha$ to $\mathcal{G}^2$. All groupings have thus been ruled out except the one(s) in which half the groups correspond in $\mathcal{P}$ to $\mathcal{G}^1$ and half to $\mathcal{G}^2$.

Proof of Proposition 5. Fix any block-intertwined grouping-pattern of $\mathcal{N}_2$, $\mathcal{M} = \{\mathcal{G}^1, \mathcal{G}^2\}$; relabel if necessary so that $1 \in \mathcal{G}^1$. Since $\mathcal{M}$ is block-intertwined, there exist $J \in \{2, 3, ..., n\}$ rank-ordered blocks, such that for any $j \in \{1, 2, ..., J\}$, 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 $\nu_0 \equiv 0$ and $V_j \equiv \sum_{x=0}^j \nu_x$ for $j \in \{0, 1, ..., J\}$, so that $V_0 = 0$ and $V_J = n$. The following quantities will be useful:

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

In order to define the type distribution in **the continuum case**, define

$$b_j \equiv \frac{\nu_j S_{\nu^2}^{-\sigma_j}}{\Psi} \ , \ \ f_j = \frac{\Psi}{n S_{\nu^2}^{-\sigma_j}} \ , \ j \in \{1, 2, ..., J\} \ ; \qquad b_0 \equiv 0 \ ; \qquad B_j \equiv \sum_{x=0}^{j} b_x \ , \ j \in \{0, 1, ..., J\} \ .$$

Now let types be drawn from $\mathcal{P} = [0,1]$ according to density function

$$f(p) = f_j \qquad \text{for} \ \ p \in [B_{j-1}, B_j] \ , \ j \in \{1, 2, ..., J\} \ .$$

Note that the 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

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

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 xth percentile, we have

$$p_{[x]} = \frac{x - V_{j-1}/n}{f_j} + B_{j-1} \qquad \text{for } \; x \in [V_{j-1}/n, V_j/n] \; , \; j \in \{1, 2, ..., J\} \; , \tag{5}$$

also continuous and strictly increasing.

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

$$x_{jg} = \begin{cases} V_{j-1}/n + \nu_j g & \text{if } \sigma_j = +1 \\ V_j/n - \nu_j g & \text{if } \sigma_j = -1 \end{cases} \qquad \text{and} \quad Z_{jg} = (\underbrace{p_{[x_{jg}]}, ..., p_{[x_{jg}]}}_{1\,x\,\nu_j}) \quad , \quad j \in \{1, 2, ..., J\} \; , \tag{6}$$

group g is then

$$\underbrace{(\underbrace{Z_{1g}}_{1\,x\,\nu_1}, \underbrace{Z_{2g}}_{1\,x\,\nu_2}, ..., \underbrace{Z_{Jg}}_{1\,x\,\nu_J})}_{1\,x\,n} \; . \tag{7}$$

There are three things to check. First, the measures of types in groups must add up to the measures of types overall. One can see this holds, since the fraction of types allocated to filling block j across all groups is ν_j/n (clear from equation 6 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 6 and 7), so that each slot in block j is filled with a measure $1/n$ of the total types.

Second, with probability one, two groups chosen at random must correspond to 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 $j, j' \in \{1, 2, ..., J\}$ 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 $\sigma_j = -1$. This establishes that groups $\{g, g'\}$ uniquely correspond to grouping-pattern $\mathcal{M}$.

Third, we show the sum of types in all groups is equal, so that by Lemma 2, this grouping is an equilibrium. Using equations 5-7, we have that the sum of types in group g is

$$\sum_{j=1}^{J} \nu_j p_{[x_{jg}]} = \sum_{\substack{j=1 \\ \sigma_j=+1}}^{J} \nu_j \left[\frac{\nu_j g}{f_j} + B_{j-1} \right] + \sum_{\substack{j=1 \\ \sigma_j=-1}}^{J} \nu_j \left[\frac{\nu_j/n - \nu_j g}{f_j} + B_{j-1} \right] =$$

$$\sum_{\substack{j=1 \\ \sigma_j=+1}}^{J} \nu_j B_{j-1} + \sum_{\substack{j=1 \\ \sigma_j=-1}}^{J} \nu_j B_j + \frac{gn}{\Psi} \left[S_{\nu^2}^{-1} \sum_{\substack{j=1 \\ \sigma_j=+1}}^{J} \nu_j^2 - S_{\nu^2}^{+1} \sum_{\substack{j=1 \\ \sigma_j=-1}}^{J} \nu_j^2 \right] = \sum_{\substack{j=1 \\ \sigma_j=+1}}^{J} \nu_j B_{j-1} + \sum_{\substack{j=1 \\ \sigma_j=-1}}^{J} \nu_j B_j \ ,$$

where the second 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.

Turning to **the finite case**, define $\Delta_j \equiv b_j/(k\nu_j)$ and let there be kn agents, with agent i's type satisfying

$$p_i = B_{j-1} + (i - kV_{j-1} - 1/2)\Delta_j \quad \text{for } i \in \{kV_{j-1}+1, kV_{j-1}+2, ..., kV_j\} \ , \ \ j \in \{1, 2, ..., J\} \ . \tag{8}$$

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), and thus kn unique types overall. Also, $p_i < p_{i'}$ if $i < i'$. Now consider the following grouping, where $g \in \{1, 2, ..., k\}$ indexes groups. Redefining

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

$$Z_{jg} = \underbrace{[p_{x_{jg}+1}, p_{x_{jg}+2}, ..., p_{x_{jg}+\nu_j}]}_{1\,x\,\nu_j} \ , \quad j \in \{1, 2, ..., J\} \ , \tag{9}$$

group g is as in equation 7. One can check that this is a valid grouping, as each group has ν_j types from interval (B_{j-1}, B_j), $j \in \{1, 2, ..., J\}$, and no two groups have any of the same types from this interval. One can also check that every pair of groups $g, g' \in \{1, 2, ..., k\}$, $g < g'$, uniquely corresponds to 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 7-9, we have that the sum of types in group g is

$$\sum_{j=1}^{J}\sum_{m=1}^{\nu_j} p_{x_{jg}+m} =$$

$$\sum_{\substack{j=1\\ \sigma_j=+1}}^{J}\sum_{m=1}^{\nu_j}\{B_{j-1}+\Delta_j[\nu_j(g-1)+m-1/2]\} + \sum_{\substack{j=1\\ \sigma_j=-1}}^{J}\sum_{m=1}^{\nu_j}\{B_{j-1}+\Delta_j[\nu_j(k-g)+m-1/2]\} =$$

$$g\left[\sum_{\substack{j=1\\ \sigma_j=+1}}^{J}\sum_{m=1}^{\nu_j}\Delta_j\nu_j - \sum_{\substack{j=1\\ \sigma_j=-1}}^{J}\sum_{m=1}^{\nu_j}\Delta_j\nu_j\right] + \mathcal{X} = \frac{g}{k\Psi}\left[S_{\nu^2}^{-1}\sum_{\substack{j=1\\ \sigma_j=+1}}^{J}\nu_j^2 - S_{\nu^2}^{+1}\sum_{\substack{j=1\\ \sigma_j=-1}}^{J}\nu_j^2\right] = \mathcal{X}\ ,$$

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.

Proof of Lemma 3. For group size n, the total number of grouping-patterns of $\{1, 2, ..., 2n\}$ 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 of $\mathcal{N}_2$ 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 nth Catalan number: $\binom{2n}{n}/(n+1)$. (See Dowling and Shier, 2000, pp. 145-147, especially Example 11.)

Turning to the number of block-intertwined grouping-patterns, any block-intertwined grouping-pattern is uniquely identified by a number of blocks $J \in \{2, 3, ..., n\}$ ($J = 1$ 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 $j \in \{1, 2, ..., J\}$. 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 $\{\mathcal{G}^g_1,\mathcal{G}^g_2,...,\mathcal{G}^g_n\}$ with $\mathcal{G}^g_i<\mathcal{G}^g_{i'}$ for $i<i'$. We will find a $\mathcal{P}\in\Omega_2$ such that the grouping of $\mathcal{P}$ corresponding to $\mathcal{M}$ is not efficient. There are three cases to consider. Note that $\mathcal{G}^1_1=1$, so in all cases $\mathcal{G}^1_1<\mathcal{G}^2_1$; the cases differ in which group dominates in the second and third ranks.

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

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

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

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

Fix such an ϵ and let $\mathcal{P}=(p_1,p_2,...,p_{2i^*},p_{2i^*}+(2i^*+1)\epsilon,p_{2i^*}+(2i^*+2)\epsilon,...,p_{2i^*}+2n\epsilon)$. 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 $\mathcal{P}$ corresponding to $\mathcal{M}$. (Recall that $\mathcal{G}^1_{i^*}<\mathcal{G}^2_{i^*}=2i^*$.) But the inequality guarantees a different grouping of $\mathcal{P}$ is more efficient.

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

By strict submodularity of Π,

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

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

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

Fix such an ϵ and let $\mathcal{P} = (p_1+\epsilon, p_1+2\epsilon, p_3, p_4, ..., p_{2i^*}, p_{2i^*}+(2i^*+1)\epsilon, p_{2i^*}+(2i^*+2)\epsilon, ..., p_{2i^*}+2n\epsilon)$. 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 $\mathcal{P}$ corresponding to $\mathcal{M}$. (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.

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

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

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

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

Fix such an ϵ and let $\mathcal{P} = (p_1, p_2, p_3 + 3\epsilon, p_3 + 4\epsilon, p_5, p_6, ..., p_{2i^*}, p_{2i^*} + (2i^* + 1)\epsilon, p_{2i^*} + (2i^* + 2)\epsilon, ..., p_{2i^*} + 2n\epsilon)$. 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 $\mathcal{P}$ corresponding to $\mathcal{M}$. (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 not efficient; this grouping is not an equilibrium either, since any equilibrium grouping must be efficient. 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 grouping-pattern $\mathcal{M}$ of $\mathcal{N}_k$, there exists a $\mathcal{P} \in \Omega_k$ such that the grouping of $\mathcal{P}$ corresponding to $\mathcal{M}$ is not efficient. It must be shown that for any grouping-pattern $\mathcal{M}$ of $\mathcal{N}_{k+1}$, there exists a $\mathcal{P} \in \Omega_{k+1}$ such that the grouping of $\mathcal{P}$ corresponding to $\mathcal{M}$ is not efficient and not an equilibrium.

Fix any grouping-pattern of $\mathcal{N}_{k+1}$, $\mathcal{M} = \{\mathcal{G}^1, \mathcal{G}^2, ..., \mathcal{G}^k, \mathcal{G}^{k+1}\}$. By the inductive hypothesis, there is a $\widetilde{\mathcal{P}} \in \Omega_k$ and a grouping-pattern of $\mathcal{N}_{k+1}$, $\mathcal{M}' = \{\mathcal{G}'^1, \mathcal{G}'^2, ..., \mathcal{G}'^k, \mathcal{G}^{k+1}\} \neq \mathcal{M}$, such that the grouping of $\widetilde{\mathcal{P}}$ corresponding to $\{\mathcal{G}^1, \mathcal{G}^2, ..., \mathcal{G}^k\}$ produces strictly less than the grouping of $\widetilde{\mathcal{P}}$ corresponding to $\{\mathcal{G}'^1, \mathcal{G}'^2, ..., \mathcal{G}'^k\}$.[33] Fix such a $\widetilde{\mathcal{P}}$. Since $\widetilde{\mathcal{P}}$ contains kn unique types from $[0, 1]$, one may choose a $\widehat{\mathcal{P}} \in \Omega_1$, such that $\widehat{\mathcal{P}} \cap \widetilde{\mathcal{P}} = \emptyset$ and the group $\widehat{\mathcal{P}}$ corresponds in $\mathcal{P} \equiv \widetilde{\mathcal{P}} \cup \widehat{\mathcal{P}}$ to $\mathcal{G}^{k+1}$ (by adding n new types from $[0, 1]$ into $\widetilde{\mathcal{P}}$ 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 $\mathcal{P}$ corresponding to $\mathcal{M}$ ($\mathcal{M}'$). 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 hence not an equilibrium, since every equilibrium grouping must be efficient.

Proof of Proposition 7. Let M be an equilibrium grouping, $\underline{p} = \sup_{G \in M} p_1^G$, and $\overline{p} = \inf_{G \in M} p_n^G$. If $\underline{p} > \overline{p}$, 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, $\underline{p} \leq \overline{p}$, and for any $\widetilde{p} \in [\underline{p}, \overline{p}]$, every equilibrium group G has $p_1^G \leq \underline{p} \leq \widetilde{p}$ and $p_n^G \geq \overline{p} \geq \widetilde{p}$.

To show that $\widetilde{p}$ may be as low as $p_{[1/n]}$, 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 $\widetilde{p}$ 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 $\epsilon \to 0$.

[33]This is a slight abuse of notation: $\{\mathcal{G}^1, \mathcal{G}^2, ..., \mathcal{G}^k\}$ is not a 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 $\{\mathcal{G}'^1, \mathcal{G}'^2, ..., \mathcal{G}'^k\}$, which contains the same kn integers. However, it is straightforward to extend the definition of grouping-pattern to refer to any distinct, positive kn integers rather than the first positive kn integers; and the definition of correspondence to be based only on the rankings in the grouping-pattern. This is omitted for brevity.