Evolutionary Games with Group Selection

Martin Kaae Jensen* and Alexandros Rigos†

December 14, 2012

Abstract

This paper introduces two new concepts in evolutionary game theory: Nash equilibrium with Group Selection (NEGS) and Evolutionarily Stable Strategy with Group Selection (ESSGS). These concepts generalize Maynard Smith and Price (1973) to settings with arbitrary matching rules, in particular they reduce, respectively, to Nash equilibrium and ESS when matching is random. NEGS and ESSGS are to the canonical group selection model of evolutionary theory what Nash Equilibrium and ESS are to the standard replicator dynamics: any NEGS is a steady state, any stable steady state is a NEGS, and any ESSGS is asymptotically stable. We exploit this to prove what may be called "the second welfare theorem of evolution": Any evolutionary optimum will be a NEGS under some matching rule. Our results are illustrated in Hawk-Dove, Prisoners’ dilemma, and Stag Hunt games.

Keywords: Evolutionary Game Theory, Evolutionarily Stable Strategy, ESS, Group Selection, Non-random Matching.

JEL Classification Codes: C72, C73.

*Department of Economics, University of Birmingham. Email: m.k.jensen@bham.ac.uk.
†Department of Economics, University of Birmingham. Email: axr644@bham.ac.uk.
1 Introduction

The canonical evolutionary game theory model of Maynard Smith and Price (1973) plays an important role in biology, economics, political science, and other fields of inquiry. Its equilibrium concept, an evolutionarily stable strategy (ESS) describes evolutionary outcomes in environments where populations are large and matching is random.\(^1\) Since an ESS is a refinement of the Nash equilibrium concept, it obviously cannot explain any behavioral departure from purely self-serving behavior in the one-shot Nash sense. In particular it cannot account for cooperative behavior in say, a prisoners’ dilemma, or shed light on altruism more generally, nor can it account for any other non-Nash behaviors such as spite (Wilson, 1975) or costly punishment (Fehr and Gächter, 2000).

In order to explain such deviations from Nash behavior, evolutionary game theory turned to models with a finite number of agents hence departing from the first of the mentioned conditions of Maynard Smith and Price (1973). Thus in Schaffer (1988), the finite set of individuals have “market power” and can influence average fitness, while in the model preferred by Maynard Smith (1982) — namely repeated games — a few agents, usually just two, can perfectly monitor and record each others’ past actions and condition their strategies hereupon (in evolutionary theory, the repeated games approach is usually referred to as direct reciprocity). Both of these frameworks have led to an enormous body of research in economics and game theory (see e.g. Alós-Ferr and Ania, 2005; Leininger, 2006; Samuelson, 2002; Vega-Redondo, 1997, and references therein).

While evolutionary game theorists turned to finite populations, evolutionary theorists more broadly devoted as much, if not more attention to a departure from the second basic condition of Maynard Smith and Price (1973), namely the assumption that matching is random. For an account of the literature on models with non-random matching — better known as models of group selection, see e.g. Sober and Wilson (1999) or Bergström (2002). In these models, matching is typically assortative meaning in the context of a prisoners’ dilemma type environment that cooperators have higher probability of being matched with other cooperators than with defectors. This perspective is often highly realistic corresponding for example to situations where a large group of individuals cannot perfectly monitor each others’ past behaviors but receive some “revealing signals” about opponents’ types and exert some influence on who they are matched with (Maynard Smith, 1964; Wilson, 1975, 1977). Non-random matching also results if matching depends on the geographical location of individuals (Eshel et al., 1998; Nowak and May, 1992; Skyrms, 2004); or if (genetically) similar individuals match preferentially as in models of kin selection (Hamilton, 1964). When matching is non-random a variety of different groups will generally coexist at any given moment in time. For example in a prisoners’ dilemma type

\(^1\)Intuitively, random matching means that an individuals’ type has no influence on what type of individual he is likely to be matched with.
setting, some groups will consist of defectors only, some of cooperators only, and some will be mixed. Thus the average fitness will differ across groups, as will the fitness a specific type of individual obtains if he is placed in different groups. It follows that evolutionary pressure takes places not just at the individual level but also at the group level even though individuals are ultimately the fitness bearing entities.\(^2\)

Now, the existing literature on group selection is usually informal and/or deals only with special cases (typically two types who are matched pairwise and assortatively). As a basis for this paper’s main results, we begin in section 2 by laying out the canonical group selection model in a general and self-contained manner. Compared to existing literature, we add value by setting up a model that allows for arbitrary matching rules (ways to match populations into groups), any number of strategies, arbitrary group sizes, and any possible payoff structure in the group stages (i.e., any possible underlying symmetric normal form game, see section 2.1). For any reader who is unfamiliar with — or confused by — group selection theory, it is our hope that section 2 will provide an easily accessible point of entry.

A key thing to notice about the group selection model of section 2 is that it is not a game theoretic model. In the terminology of biologists it is a model of evolutionary theory, not a model of evolutionary game theory. More specifically, it is a dynamical model of selection where attention is devoted to steady states of the associated replicator dynamics. This is of course in sharp contrast to the random matching case where Nash equilibrium and ESS play central roles and allows the powerful machinery of game theory to be applied.\(^3\) The main objective of the present paper is to fill the resulting gap in the literature. Specifically, we are going to ask what game theoretic equilibrium concepts form group selection’s natural parallels to Nash equilibrium and ESS. This leads to two new equilibrium concepts, namely a Nash equilibrium with group selection (NEGS) and an evolutionarily stable strategy with group selection (ESSGS). These concepts turn out to be intuitive once the underlying evolutionary game, which we call a group selection game, is understood. Interestingly, this game turns out to be novel even from a game theoretic perspective: As in standard imperfect information games, agents make decisions without knowing with certainty the strategies pursued by opponents — all they know is the distribution of the opponents’ strategies, or to put it in the evolutionary terminology, the probabilities of ending up in any of the different kinds of groups. Crucially, these probabilities depend on the actual strategies

\(^2\)As shown by Kerr and Godfrey-Smith (2002), one may with equal formal correctness think of selection taking place at the individual or the group level. This difference in perspective has been (and is) the topic of a heated debate in evolutionary theory, a key reference here being the book “Unto Others” by Sober and Wilson (1999). As explained in section 2 we are going to take a so-called “individualist” perspective in this paper, and will not go into the more philosophical aspects of the group selection controversy.

\(^3\)The precise relationship between steady states, Nash equilibrium, and ESS is returned to in a moment and in detail in section 4. For example an ESS is a refinement of Nash equilibrium with the crucial property that any asymptotically stable state of the evolutionary (replicator) dynamics is an ESS (Weibull, 1995, chapter 3). Thus when one studies the set of ESS’s, “bad” equilibria have been removed which not surprisingly leads to stronger results.
pursued by the agents. For simplicity, imagine a large group of individuals, all of whom has a choice between two strategies, “honesty” (H) or “deception” (D). Agents must commit to a strategy before being allocated into equal-sized groups where they execute these strategies (equivalently, they choose their actions with imperfect knowledge about opponents’ actions). Given a specific matching rule (a given way to divide a population with a given fraction of H- and D-types into groups of equal size) and given that agents know the population wide composition into H and D types, each agent can calculate the probabilities of ending up in any specific kind of group as a function of the specific strategy chosen (H or D). In an NEGS, individuals’ optimal choices precisely lead to the population-wide composition into H and D types which formed the basis of their decisions in the first place. The concept of an ESSGS simply adds a “non-invasion” criterion to this Nash/fixed point criterion precisely as is the case with random matching (Maynard Smith and Price, 1973). Note than a NEGS interchangeably can be viewed as a mixed strategy pursued by all individuals or as a vector that gives the fractions of each type in equilibrium. In the previous case, a NEGS or an ESSGS may be, say (0.9, 0.1) meaning that each individual in the population will be honest with probability 90 % and deceptive with probability 10 %. With an infinite population size, this of course implies that 90 % of the population will be honest, and 10 % deceptive at any given moment in time. Intuitively, in a NEGS the deceptive individuals’ purpose is to keep the honest individuals in check (and vice versa): without a sufficiently large population of deceivers, the benefit of choosing to deceive will outweigh that of being honest because deceptive individuals will face a relatively small chance of being matched with another deceiver even though matching is assortative.

After defining group selection games and proving existence, we turn to the relationship with the dynamic evolutionary model of section 2. Thus in theorem 5 — which together with the “second welfare theorem of evolution” described below forms this paper’s main contribution — we prove that any NEGS is a steady state for the replicator dynamics, that any (Lyapunov) stable steady state for the replicator dynamics is an NEGS, and that any ESSGS is an asymptotically stable state of the replicator dynamics. These results extend existing results on Nash equilibrium and ESS (Hofbauer and Sigmund, 1998; Maynard Smith and Price, 1973; Weibull, 1995) to settings with non-random matching, and show that NEGS and ESSGS are important new evolutionary game theory concepts. Immediately, a long list of research questions report themselves in that one could attempt to “transfer” over to group selection models all of the existing results from evolutionary game theory. We shall leave the bulk of this for future research, for example we are not going to go into topics related to neutrally stable strategies, asymptotically stable sets, doubly symmetric games or the fundamental theorem of natural selection.

---

4 Obviously, the number of possible group compositions depends on the group size as well as the number of strategies. With two strategies and groups of size two, any individual can end up in precisely two different kinds of groups — one where the opponent is of the same type and one where he is not.
(for these “textbook” issues see the monographs of Hofbauer and Sigmund (1998) or Weibull (1995)). Instead we are going to focus in section 6 on a question which in some sense “ignited” this whole literature. The point of the prisoners’ dilemma is that Nash equilibrium and with it evolutionary models based on random matching may easily fail to produce outcomes that maximize average payoff/welfare in the population. The question from the group selection point of view then becomes: What types of (non-random) matching will, if any, lead to optimality? Our main result in this regard (theorem 6) might be called the “second welfare theorem of evolutionary theory” telling us that any outcome that is optimal will in fact be a NEGS under some matching rule. In a number of standard games (hawk-doves, stag hunt, prisoners’ dilemma) we proceed to characterize these matching rules and in doing so gain an understanding of when evolution in a specific situation (for a fixed matching rule and payoff structure) is likely to lead to an evolutionary optimum or not. To give an example, we show that in the Hawk-Doves model even low levels of assortativity (in the “constant index” sense of Bergström (2003)) may still lead to the evolutionary optimum if a Dove who meets a Dove receives only a “modest” gain from changing to the Hawk strategy.

The structure of the paper is as follows: Section 2 describes the general group selection model and section 3 defines group selection games, NEGS and ESSGS. Section 3 also contains some some basic results on existence and the relationship between NEGS and ESSGS. Section 4 contains our main theoretical results discussed above. Section 5 contains a number of examples, and section 6 discusses the fitness/welfare issues with basis in the aforementioned “second welfare theorem”. Finally, section 7 concludes.

2 Group Selection in Evolutionary Theory

In this section we are going to present the canonical group selection model of Wilson (1975), Maynard Smith (1964), and Wilson (1977). In the literature, this model is also called either the haystack model (Maynard Smith, 1964) or the multi-level selection model (Kerr and Godfrey-Smith, 2002). In the language of Kerr and Godfrey-Smith (2002), we adopt an “individualist perspective” meaning that in our model, fitness is assigned to individuals rather than to the groups they form. Hence, the extent to which groups are selected for, will depend on the fitnesses of the individuals they contain. Nonetheless, as was shown by Kerr and Godfrey-Smith (2002), one can formally recast this model so that groups are fitness bearing — a perspective strongly advocated in the famous book “Unto Others” by Sober and Wilson (1999). See also Maynard Smith (1998) and Okasha (2005) for more on this issue.

The traditional treatment of the group selection model is mostly informal or focuses on quite spe-
cial cases. A notable exception being Kerr and Godfrey-Smith (2002) who present a detailed analysis of the two strategies case. Our presentation below includes cases where more than two pure strategies are available to the population and provide a rigorous formalism that is suitable to analyze such models under any kind of group-formation rule while also making the model’s connection with standard replicator dynamics clear. This lays the groundwork for this paper’s main results.

Briefly, the model can be summarized as follows: At each date there is a large set of individuals, formally the continuum $I = [0, 1]$. At the beginning of each period, the agents are allocated into groups of the same finite size $n \in \mathbb{N}$. This happens in accordance with what we call a matching rule (formally defined in subsection 2.2) which is a function that maps the type frequency of the set of agents into the distribution of group types. After the $n$-sized groups are formed, the individuals in each group face a symmetric normal-form game (section 2.1). An agent’s payoff determines his fitness, i.e., the (expected) number of children the agent will send on to the next round. Crucially, in accordance with the basic premise of evolutionary game theory, agents are assumed to be hard-wired to follow the same strategy as the parent (‘like begets like’). Thus an individual who is fathered by a parent who chose strategy $j$, say, in the previous round will mechanically execute strategy $j$ in his group game, regardless of the composition of individuals in the specific group he is drawn into. Finally, after the group game stage, a new generation is born with the relative proportion of each type determined by the success (fitness) this type’s strategy enjoyed across the group games. The above process then repeats itself leading to a new generation and so on. The evolutionary outcome of this group selection process is a steady state of the resulting replicator dynamical system as described in section 2.3.

Note that — apart from our insistence that the group games can be seen as normal form games (a perspective that is alien to the existing literature) — the model is entirely non-game theoretic.

2.1 The Underlying Normal Form Group Games

Our description begins with the underlying normal form game that agents face in the group stages. Although in evolutionary models, individuals act purely mechanically and play the strategy inherited from the parent, they nonetheless participate in a standard normal form game and receive payoffs/fitnesses accordingly. We need to make this game theoretic aspect clear to set the stage for this paper’s main results.

Let $n \in \{2, 3, \ldots\}$ denote the group size so that $N = \{1, \ldots, n\}$ is the set of players in a group. A group
A game is a symmetric normal form game \( G = \langle N, M, A \rangle \) where \( M = \{1, \ldots, m\} \) is the set of pure strategies and \( A : M \times M^{n-1} \to \mathbb{R} \) is the (common, pure) payoff function. The set of all \( n \)-players, \( m \)-strategies symmetric normal form games is denoted by \( \mathcal{G}_{n,m} \). Note that by symmetry, \( A(y^i, y^{-i}) = A(y^i, \tilde{y}^{-i}) \) where \( y^i \in M \) is any pure strategy for player \( i \), and \( y^{-i}, \tilde{y}^{-i} \in M^{n-1} \) are pure strategies of \( i \)'s opponents where \( \tilde{y}^{-i} \) is any permutation of \( y^{-i} \). A (symmetric) Nash equilibrium for \( G \) is defined in the usual way as a vector \( \sigma^* \in S_m \equiv \{ \sigma \in \mathbb{R}_+^m : \sum_{j=1}^m \sigma_j = 1 \} \) such that \( A(\sigma^*, (\sigma^*, \ldots, \sigma^*)) \geq A(\sigma, (\sigma^*, \ldots, \sigma^*)) \) for all \( \sigma \in S_m \).

It is convenient to write the previous payoff structure in a way that makes explicit reference to the group structure. Call an individual who chooses pure strategy \( j \in M \) a type \( j \) individual. Due to symmetry, the payoff to such a type \( j \) individual depends only on the number of opponents in his group who play each of the \( m \) strategies (as opposed to which opponents follow what strategies). Next imagine that this type \( j \) individual finds himself in a group, group \( i \) say, consisting of \( n^i \) individuals of type 1, \( n^i_2 \) individuals of type 2, and so on up to \( n^i_m \). In this situation, the individual’s payoff will be equal to \( A(j, j^{opp}) \) where \( j^{opp} \in M^{n-1} \) is any vector of opponents’ strategies which contains \( n^i \) strategy 1 entries, \( \ldots, n^i_{j-1} \) strategy \( j-1 \) entries, \( n^i_j-1 \) strategy \( j \) entries, \( n^i_{j+1} \) strategy \( j+1 \) entries, \( \ldots, n^i_m \) strategy \( m \) entries. Crucially, we can write the payoff \( A(j, j^{opp}) \) simply as \( \gamma_{n^i} \), or even as \( A_j \) where \( i \) is the index of the specific group the individual finds himself in (as long as we keep record of the group composition \( n^i = (n^i_1, \ldots, n^i_m) \) of group \( i \)).

In this way, we can capture all of the information we need about the normal form game in a sequence \( (A^i_j) \) where \( j = 1, \ldots, m \) and \( i = 1, \ldots, \gamma_{n,m} \). Here \( \gamma_{n,m} \) is the number of different \( n \)-sized groups that can be formed with \( m \) different pure strategies. From combinatorics we know that \( \gamma_{n,m} \) precisely equals the number of multisets of cardinality \( n \) with elements taken from a set with cardinality \( m \) (see Aigner, 2007, p. 15), i.e.,

\[
\gamma_{n,m} = \frac{(n + m - 1)!}{n!(m - 1)!}
\]  

(1)

For example, \( \gamma_{2,2} = 3 \) since three different groups can be formed if the group size equals 2 and there are 2 possible strategies (these groups are, respectively, one where both are of type 1, one where both are of type 2, and one where the individuals follow different strategies).

\footnote{Letting \( \sigma^i \in S_m \equiv \{ \sigma \in \mathbb{R}_+^m : \sum_{j=1}^m \sigma_j = 1 \} \) denote a mixed strategy for player \( i \) and \( \sigma^{-i} \in S_{m-1} \) denote a mixed strategy profile of player \( i \)'s opponents, it is easy to see that \( A(\sigma^i, \sigma^{-i}) = \sum_{y \in S_m} A(y^i, y^{-i}) \prod_{k \in \mathbb{N}} \sigma^i_k \).}

\footnote{Note that since the individual himself is counted here, we necessarily have \( n^i_j \geq 1 \) (there is at least one of the individual’s own type). Of course we must also have \( \sum n^i_j = n \) and each \( n^i_j \) must be non-negative.}

\footnote{Of course, we must be a little careful here because some of these are not really properly defined. Specifically, \( A^i_j \) is not well-defined unless \( n^i_j \geq 1 \). But building this explicitly into the notation leads to unwarranted complications.}
2.2 Group Formation

We now turn to the question of how groups are formed out of each generation’s individuals. The key concept is that of a matching rule which generalizes what Kerr and Godfrey-Smith (2002, p. 484) call a “rule of group assembly” to more than 2 (pure) strategies.\(^{11}\)

A population state is the frequency distribution of the different types in the population, i.e., a vector \(x = (x_1, \ldots, x_m) \in S_m\) where \(x_1\) is the fraction of 1-strategists in the population, \(x_2\) is the fraction of 2-strategists in the population, and so on. A group state similarly represents the group frequencies and so is a vector \(g = (g_1, g_2, \ldots, g_{\gamma_{n,m}}) \in S_{\gamma_{n,m}}\) where \(g_1\) is the fraction of all groups that is of type 1, \(g_2\) the fraction of type 2 groups, and so on up to group \(\gamma_{n,m}\) which it is recalled is the number of different \(n\)-sized groups it is possible to form when there are \(m\) different strategies (section 2.1).

A matching rule is simply a function that maps a population state \(x \in S_m\) into a group state \(g \in S_{\gamma_{n,m}}\).

So a matching rule describes how a given population is allocated into groups. Of course, consistency requires that the fraction of \(j\)-type individuals allocated into the different groups equals the fraction \(x_j\) of individuals of type \(j\) that are actually present in the population. Since the proportion of \(j\)-type individuals in an \(i\)-type group by definition is \(\frac{n^i_j}{n}\), the fraction of individuals that are of type \(j\) and in \(i\)-type groups will, for any given group state \(g\), be \(\frac{n^i_j}{n} g_i\). Hence, across all groups the fraction of the population that is of type \(j\) is \(\sum_{i=1}^{\gamma_{n,m}} \frac{n^i_j}{n} g_i\). This number must then equal \(x_j\) for every individual to be allocated to one (and only one) group. In summary we have:

**Definition 1. (Matching Rules)** A matching rule is a function \(f : S_m \rightarrow S_{\gamma_{n,m}}\) such that for all \(x \in S_m\):

\[
\sum_{i=1}^{\gamma_{n,m}} \frac{n^i_j}{n} f_i(x) = x_j, \text{ for } j = 1, \ldots, m
\]

The set of matching rules in a population with \(m\) strategies and group-size \(n\) is denoted by \(\mathcal{F}_{n,m}\).

Note that if we define \(\text{supp}(j)\) to be the set of group types that contain at least one \(j\)-strategist, (2) can equivalently be written:

\[
\sum_{i \in \text{supp}(j)} \frac{n^i_j}{n} f_i(x) = x_j \quad \text{for all } j \in M
\]

Recall from above that \(\frac{n^i_j}{n} f_i(x)\) is the fraction of the total population that is of type \(j\) and is allocated to a group of type \(i\) under the matching rule \(f\). When \(x_j > 0\) we may divide this by the fraction \(x_j\) of the population that is of type \(j\) in order to get the fraction of \(j\)-type individuals that is allocated to a group of type \(i\):

\(^{11}\)The concept is also related to Bergström (2003) as returned to below.
This may be compared with Bergström (2003) who studies group selection (again in the special case \( n = m = 2 \)), and who takes the \( w^i \)'s as fundamentals instead of the matching rule. More specifically, Bergström (2003) considers the difference \( w^1 - w^2 \) and calls this difference the ‘index of assortativity’. We return to the index of assortativity in example 1.3 below.

We finish this subsection by presenting a number of concrete examples of matching rules. We shall be calling on these repeatedly throughout the rest of this paper.

**Example 1.**

1. **Complete segregation.** Different strategies do not mix. All individuals are allocated into groups with only individuals of the same type and thus all groups contain a single type of individuals each (\( n \) individuals that follow the same strategy). The group types that have \( n \) individuals of the same type get a non-negative frequency whereas all other kinds of groups get a frequency of zero. Due to the consistency requirements for matching rules, we get that the group type that contains \( n \) \( j \)-types should get a frequency of \( x^j \). So, formally, the matching rule for complete segregation is the following.

\[
\begin{align*}
    f_i(x) &= x^j , & \text{if } n^i_j = n \\
    f_i(x) &= 0 , & \text{otherwise.}
\end{align*}
\]

* e.g. When \( n = m = 2 \) the matching rule for complete segregation take the form:

\[
\begin{align*}
    f_1(x_1, x_2) &= x_1 \\
    f_2(x_1, x_2) &= 0 \\
    f_3(x_1, x_2) &= x_2.
\end{align*}
\]

2. **Random matching.** Let us define the opponent profile of a type \( j \) individual in a type \( i \) group to be the vector \( v^i_j = (v^i_1, \ldots, v^i_j, \ldots, v^i_m) \equiv (n^i_1, \ldots, n^i_j - 1, \ldots, n^i_m) \) that shows how many opponents of each type a type \( j \) individual faces when she is drawn into a group of type \( i \). Obviously, individuals of different types that face the same opponent profile will be in groups of different types. We will say that matching is random when the (ex ante) probability of an individual (conditional on her type) to end up facing a specific opponent profile is independent of her type. If this is the case, then the frequencies of group types will follow a multinomial distribution (see for example Lefebvre, 2007, p. 22):\(^{12}\)

\[
\begin{align*}
    w^i_j(x) &\equiv \frac{n^i_j}{n x^j_k \prod_k n^i_k \prod_k x^i_k} = (n - 1)! \prod_k v^i_k \prod_k x^i_k , \text{ i.e. it is independent of the individual's strategy } j.
\end{align*}
\]

\(^{12}\)To show that the property described above holds for the matching rule of equation (6), let us consider a group of type \( i \) with \( n^i_j \geq 1 \) for some \( j \in M \). Notice that a \( j \)-type in that group has \( n^i_j \) type 1 opponents, \( n^i_j - 1 \) type \( j \) opponents, \ldots, \( n^i_m \) type \( m \) opponents. So the opponent profile for a \( j \) strategist in a type \( i \) group will be \( v = (n^i_1, n^i_j - 1, \ldots, n^i_m) \). Indeed, the probability of a type \( j \) individual (conditional on her type) to end up in group with opponent profile \( v = (v_1, \ldots, v_m) \) is given by:

\[
\begin{align*}
    w^i_j(x) &\equiv \frac{n^i_j}{n x^j_k \prod_k n^i_k \prod_k x^i_k} \prod_k v^i_k \prod_k x^i_k = (n - 1)! \prod_k v^i_k \prod_k x^i_k , \text{ i.e. it is independent of the individual's strategy } j.
\end{align*}
\]
\[ f_i(x) = \frac{n!}{\prod_{j \in M} n_i^j!} \prod_{j \in M} x_j^{n_i^j}. \]  

Notice that for \( m = 2 \), the random matching rule becomes

\[ f_i(x_1, x_2) = \frac{n!}{n_1^1!(n-n_1^1)!} x_1^{n_i^1} x_2^{n-n_i^1}. \]

That is it boils down to the binomial distribution (see Kerr and Godfrey-Smith, 2002, p. 484).

3. **Constant Index of Assortativity.**

Bergström (2003) studies 2-person prisoner’s dilemma population games by using the ‘index of assortativity’ which he defines as “the difference between the probability that a C-strategist meets a C-strategist and the probability that a D-strategist meets a C-strategist”. In terms of notation used in this paper (with \( x_1 \) and \( x_2 \) denoting the proportion of cooperators and defectors in the population respectively), this means that the index of assortativity at a state \( (x_1, x_2) \) will be:

\[ \alpha(x_1, x_2) = w_1(x_1, x_2) - w_2(x_1, x_2) = \frac{f_1(x_1, x_2)}{x_1} - \frac{f_2(x_1, x_2)}{2x_2}. \]

Bergstrom goes on to analyze prisoners’ dilemma games under “assortative matching” rules that have a constant index of assortativity \( \alpha \) for all values of \( x \). As one easily verifies, the matching rule corresponding to a constant index of assortativity \( \alpha \) is:

\[
\begin{align*}
  f_1(x) &= x_1 (1 - (1-\alpha) x_2) \\
  f_2(x) &= 2(1 - \alpha) x_1 x_2 \\
  f_3(x) &= x_2 (1 - (1 - \alpha) x_1).
\end{align*}
\]

In the case of \( \alpha = 0 \) the rule coincides with the random matching rule and in the case of \( \alpha = 1 \) it coincides with the complete segregation rule (for both of these statements we of course need \( n = m = 2 \), i.e., two players and two strategies).

4. **“Almost” Constant Index of Dissociation.**

It is not possible to extend the previous constant index of assortativity rule to \( \alpha < 0 \), i.e., to dissociative matching. This is because if such a ‘constant index of dissociation’ rule would be imposed, values of the matching rule function when \( x \) is close to 0 or to 1 would be outside the boundaries allowed by condition (3).13 A simple solution to this problem is to make the value of the matching rule at such \( x \)s to be exactly on the boundary (to allow ‘as much mixing as possible’). This leads to the following matching rule where \( \beta \in [0, 1] \) is the ‘index of dissociation’

13More specifically, this happens for \( x \in (0, \frac{1}{1+\alpha}) \cup (\frac{1}{1+\alpha}, 1) \) when \( \alpha \in [-1, 0) \).
\[
\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), f_3(\mathbf{x})) = \begin{cases} 
(0, 2x_1, 1 - 2x_1), & x_1 \in \left[0, \frac{\beta}{1+\beta}\right] \\
(x_1(1 - (1 + \beta)x_2), 2(1 + \beta)x_1x_2, x_2(1 - (1 + \beta)x_1)), & x_1 \in \left[\frac{\beta}{1+\beta}, \frac{1}{1+\beta}\right] \\
(1 - 2x_2, 2x_2, 0), & x_1 \in \left[\frac{1}{1+\beta}, 1\right] 
\end{cases}
\]

2.3 Dynamics and Steady States

At this point we have defined all of the key ingredients of a group selection model: A set of agents \( I = [0, 1] \), the normal form group game \( G = \langle N, M, A \rangle \) (here \( N = \{1, \ldots, n\} \) and \( M = \{1, \ldots, m\} \) where \( n \) is the group size and \( m \) the number of strategies/types), and the matching rule \( \mathbf{f}: S_m \to S_{\gamma_{n,m}} \) which in each period allocates the newborn generation into groups (recall from section 2.1 that \( \gamma_{n,m} \) is the number of different \( n \)-sized groups that can be formed from \( m \) different strategies). A group selection model can thus be summarized by a tuple \( < I, G, \mathbf{f}> \).

We are now in a position to describe the dynamical system that constitutes the evolutionary model of group selection. Let \( \mathbf{x}^t \in S_m \) denote the (population) state at date \( t \) (the vector of frequencies of the different types at the given date, see subsection 2.2). At date \( t \), the population is allocated into groups according to the matching rule \( \mathbf{f} \), hence \( \mathbf{f}(\mathbf{x}^t) \in S_{\gamma_{n,m}} \) is the resulting group frequency distribution. Regardless of which group an individual of type \( j \) ends up in, he will mechanically follow the strategy of his type (as inherited from the parent) and fitness will be distributed accordingly. Now recall from equation (4) of section 2.2 that \( w^i_j(\mathbf{x}) = \frac{n_i^j x_j}{n} f_i(\mathbf{x}) \) is the fraction of \( j \)-type individuals that is allocated to groups of type \( i \) under the matching rule \( \mathbf{f} \) when the population state is \( \mathbf{x} \) and \( x_j > 0 \). From section 2.1 we know that the payoff/fitness of a \( j \)-type who finds himself in a group of type \( i \) is \( A^i_j \). The average fitness of a type \( j \) individual at date \( t \) is consequently \( \sum_{i \in \text{supp}(j)} w^i_j(\mathbf{x}^t) A^i_j \). This average fitness will be denoted by \( \pi_j(\mathbf{x}^t) \), and if we substitute for \( w^i_j(\mathbf{x}^t) \) it is clear that this is given by:

\[
\pi_j(\mathbf{x}) \equiv \sum_{i \in \text{supp}(j)} \frac{n_i^j}{nx_j} f_i(\mathbf{x}^t) A^i_j 
\]

Since \( \pi_j(\mathbf{x}) \) is the average fitness of a \( j \)-type, the average fitness of all types in the population will be:

\[
\bar{\pi}(\mathbf{x}) = \sum_{j=1}^{m} x_j \pi_j(\mathbf{x}) 
\]

All that now remains is to describe how these fitnesses determine the next generation. At this point we have deliberately avoided saying whether time is to be thought of as discrete or continuous. In fact, we are going to describe both, since both play important roles in the existing literature.
Beginning with the discrete time version, the well-known replicator dynamics equations (Hammerstein and Selten, 1994; Taylor and Jonker, 1978; Weibull, 1995, pp. 122-4), formalize the (sensible) notion that at time \( t+1 \) the proportion of the population that is of type \( j \) must equal the proportion of type \( j \) individuals at date \( t \) times the relative fitness of a type \( j \) individual.

**Definition 2.** The discrete time replicator dynamics of the group selection model \(< I, G, f >\) is given by the equations:

\[
x^t_{j+1} = x^t_j \frac{\pi_j(x^t)}{\bar{\pi}(x)} \quad \text{for all } j \in M.
\]

where \( \pi_j \) and \( \bar{\pi} \) were defined in equations (7) and (8), respectively.

Turning next to the continuous-time case, the definition becomes (see Hofbauer and Sigmund, 1998, p. 67; Weibull, 1995, p. 72):

**Definition 3.** The continuous time replicator dynamics of the group selection model \(< I, G, f >\) is given by the equations:

\[
\dot{x}_j = x_j (\pi_j(x) - \bar{\pi}(x)) \quad \text{for all } j \in M.
\]

where \( \pi_j \) and \( \bar{\pi} \) were defined in equations (7) and (8), respectively.

A steady state is a rest point of any of the dynamical systems (9) or (10) (clearly the steady states are the same in either case). Different notions of stability such as Lyapunov and asymptotic stability are defined as usual in both discrete and continuous time, and the associated steady states (if any) are said to be Lyapunov stable, asymptotically stable, and so on. Since any uniform population state — i.e., any state where all individuals are of the same type — will be a steady state, it is clear that stability must be considered or else the model will have no predictive power. Since stability analysis is very difficult, especially in cases with more than two strategies, the group selection model as presented is generally quite difficult to analyze.

### 3 Group Selection Games

In the previous section we described in full detail what we think is a natural generalization of the canonical group selection model of evolutionary theory. It should be noted that except when we described the group games in section 2.1, we made no mention of game theory (in fact, the only reason we did mention the normal form game was because we need it in this section). In this section we are going to shift the perspective entirely to a game theoretic one. The basic underlying object of study will remain
the same: We have a continuum \( I = [0, 1] \) that is now referred to as the set of players, we have an underlying normal form game \( G \in \mathcal{G}_{n,m} \) as described in section 2.1, and we have a matching rule \( f \in \mathcal{F}_{n,m} \) as described in section 2.2. But the “story” will be very different. All three together will define a game which we call a group selection game:

**Definition 4. (Group Selection Games)** A group selection game is a tuple \(< I, G, f >\) where \( I \) is a continuum of players, \( G \in \mathcal{G}_{n,m} \) is a symmetric normal form game, and \( f \in \mathcal{F}_{n,m} \) is a matching rule.

Here is the structure of the game: As mentioned, there is a continuum \( I = [0, 1] \) of agents. These are identical, in particular they have the same finite set of pure strategies \( M = \{1, \ldots, m\} \) given from the normal form game \( G \). The game is symmetric, so we can conveniently summarize a (pure) strategy profile by its frequency distribution \( x = (x_1, \ldots, x_m) \in S_m \) where the \( j \)’th coordinate is the fraction of the players whose strategy is \( j \in \{1, \ldots, m\} \). The individual player takes \( x \) as given and being infinitesimally small his own choice of strategy will not affect the relative proportions expressed in \( x \). Now, in one description, the game has two stages: In the first stage, players choose their strategies and in the second stage they are allocated into groups of the same finite size \( n \in \{2, 3, \ldots\} \) where they execute their strategies.\(^{14}\) What is crucial here is that agents do not know with certainty which group they will end up in when they choose their strategies. However, because the structure of the game is known (common knowledge), an agent will know the rule according to which agents are allocated into groups, and so will be able to calculate the probability of ending up in any particular type of group after a specific strategy is chosen. This brings us back to (4) of section 2.2. Recall from that section that if \( x_j > 0 \) then \( w_j^i(x) \) is the fraction of type \( j \) individuals that are allocated into groups of type \( i \) under the matching rule \( f \) (and the state \( x \)):

\[
    w_j^i(x) = \frac{n^i_j f_i(x)}{nx_j}
\]  

(11)

The case where \( x_j = 0 \) is returned to in a moment. It is clear that from an expected payoff point-of-view, \( w_j^i(x) \) is the *ex-ante probability a \( j \)-strategist has of being “drawn” into group \( i \). It follows that the expected payoff to strategy \( j \) will equal,

\[
    \pi_j(x) = \sum_{i \in \text{supp}(j)} w_j^i(x) A_j^i,
\]  

(12)

where we remind the reader that \( A_j^i \) is the payoff received from playing strategy \( j \) in a group of type \( i \) (section 2.1); and \( \text{supp}(j) \) is the set of groups that contain at least one \( j \)-strategist (section 2.2). Com-

\(^{14}\)From a game theoretic perspective, it is much more natural to think of this as a situation involving uncertainty (a type of Bayesian game). But the imperfect information perspective actually turns out to be non-standard because probabilities are endogenously determined.
paring with section 2.3, this expected payoff precisely coincides with the average fitness to a type $j$ individual in the (deterministic) evolutionary group selection model.

Now, for the previous two definitions it is required that $x_j > 0$. The definition of the $w^j_i$s in (11) and so the definition of the $\pi_j$s in (12) are extended to the boundary of $S_m$ ($\text{bd}_j(S_m) = \{x \in S_m : x_j = 0\}$) by taking $w^j_i(x) = \lim_{x_j \downarrow 0} \frac{n^j_i f_i(\tilde{x})}{n \tilde{x}_j}$ whenever $x \in \text{bd}_j(S_m)$. Evidently, we need to assume that these limits exist for these extensions to be well-defined (featured in theorem 1 below). Note that the limit $\lim_{x_j \downarrow 0} \frac{f_i(\tilde{x})}{\tilde{x}_j}$, if it exists, precisely is the $j$’th partial (upper) derivative of $f_i$, $\partial_j^+ f_i(x)$. Hence $w^j_i(x) = \frac{n^j_i}{n} \partial^+_j f_i(x)$ when $x_j = 0$.

Finishing now the description of the game, players are allowed to choose mixed strategies, i.e., their strategy set is $S_m = \{y \in \mathbb{R}_+^m : \sum_{j \in M} y_j = 1\}$. The expected payoff to a mixed strategy $y \in S_m$ is then $y \cdot \pi(x)$ where $\pi(x) \equiv (\pi_1(x), \ldots, \pi_m(x))$. Note that if all players choose the same mixed strategy $y$, the state $x$ will necessarily be equal to $y$. Thus the definition of an equilibrium follows naturally:

**Definition 5. (NEGS)** Let $<I, G, f>$ be a group selection game. A strategy $x^* \in S_m$ is a Nash Equilibrium with Group Selection (NEGS) if:

$$ x^* \cdot \pi(x^*) \geq y \cdot \pi(x^*) \quad \text{for all } y \in S_m. \quad (13) $$

The average payoff (the welfare) at a NEGS $x^*$ is denoted by $\bar{\pi}(x^*) = x^* \cdot \pi(x^*)$.

Intuitively, agents in a group selection game take the matching rule's payoff effects into account, and integrate into their optimal choices the fact that different choices of strategies are associated with different probability distributions over opponents’ strategies. In a NEGS, these probabilities are “self-fulfilling” in the sense that agents’ ex-post decisions lead to the ex-ante probabilities upon which the decisions are based. In the following sections we shall see that this concept has a very close relationship with the steady states of the canonical group selection model (section 2.3). The remainder of this section is devoted to showing that the NEGS concept is well-founded, and strengthening the equilibrium concept.

Our first result states that any group selection game has an equilibrium when certain regularity conditions are satisfied by the matching rule. Note that the differentiability requirement trivially will be satisfied if the matching rule is differentiable at the boundary of $S_m$:

---

15Note that from a formal point of view, this is actually not acceptable because the game will not be well-defined if the expected payoffs are not well-defined. In a previous version of this paper we took instead $w^j_i(x) = \limsup_{x_j \downarrow 0} \frac{n^j_i f_i(\tilde{x})}{n \tilde{x}_j}$ which is always well-defined. This, however, tends to lead to confusion. Hence the present slight violation of mathematical rigor.
Theorem 1. Let \(< I, G, f >\) be a group selection game and assume that \(f\) is continuous and that the (upper) partial derivatives \(\partial^+ f_i(x)\) exist whenever \(x_j = 0\) (for all \(j \in M\) and \(i \in \text{supp}(j)\)). Then \(< I, G, f >\) has a NEGS.

Proof. See Appendix C.

Notice that all matching rules in Example 1 satisfy the conditions of Theorem 1.

Probably the most commonly used solution concept in evolutionary game theory is that of an evolutionarily stable strategy or ESS (Maynard Smith and Price, 1973). As Maynard Smith (1982, p. 14) puts it: "If \(I\) is a stable strategy, it must have the property that, if almost all members of the population adopt \(I\), then the fitness of these typical members is greater than that of any possible mutant; otherwise, the mutant could invade the population and \(I\) would not be stable".

In the literature, ESS is usually defined in games with random matching and in the special case when \(n = 2\) (see Hofbauer and Sigmund, 1998, p. 63). An appropriate generalization of the ESS concept to include non-random matching and any number of strategies is the following.

Definition 6. (ESSGS) Let \(< I, G, f >\) be a group selection game. A strategy \(\hat{x} \in S_m\) is an Evolutionarily Stable Strategy with Group Selection (ESSGS) if for each \(y \in S_m \setminus \{\hat{x}\}\), there exists \(\bar{\epsilon}_y > 0\) such that

\[
\hat{x} \cdot \pi(\epsilon y + (1 - \epsilon)\hat{x}) > y \cdot \pi(\epsilon y + (1 - \epsilon)\hat{x}) \quad \text{for all } \epsilon \in (0, \bar{\epsilon}_y). \quad (14)
\]

As Maynard-Smith’s quote suggests, the central idea behind the ESS (and therefore the ESSGS concept) is that of non-invasion. This means that a (monomorphic) population where all individuals use an ESSGS \(\hat{x}\) cannot be successfully invaded by a small but measurable (of measure \(\epsilon\) up to \(\bar{\epsilon}_y\)) group of individuals using any other strategy \(y \in S_m\) in the sense that in the new population – composed of \(1 - \epsilon\) \(\hat{x}\)-strategists and \(\epsilon\) \(y\)-strategists – the individuals using the ESS will get higher expected payoff than the invaders (the \(y\)-strategists). This is exactly what condition (14) expresses.

The ESSGS concept is a strengthening of the NEGS concept, just as the traditional notion of an ESS is a strengthening of Nash equilibrium:

Theorem 2. Let \(< I, G, f >\) be a group selection game with \(f\) satisfying the assumptions of Theorem 1. Then any ESSGS is a NEGS.

Proof. By way of contradiction, let us assume that some \(\hat{x} \in S_m\) is an ESSGS but not a NEGS. Then, there exists some \(y \in S_m\) such that \((y - \hat{x}) \cdot \pi(\hat{x}) > 0\). But from the definition of an ESSGS, there must exist some \(\bar{\epsilon}_y \in (0, 1)\) such that for all \(\epsilon \in (0, \bar{\epsilon}_y), (y - \hat{x}) \cdot \pi(\epsilon y + (1 - \epsilon)\hat{x}) < 0\). By continuity therefore \((y - \hat{x}) \cdot \pi(\hat{x}) \leq 0\). A contradiction.
We finish this section with an alternative characterization of an ESSGS which invokes the notion of *local superiority* (see for example Weibull, 1995, p. 45) defined as follows.

**Definition 7. (Local Superiority)** A strategy \( \hat{x} \in S_m \) is called locally superior if there exists a neighborhood \( U \) of \( \hat{x} \) such that for all \( y \in U \setminus \{ \hat{x} \} \)
\[
\hat{x} \cdot \pi(y) > y \cdot \pi(y). \tag{15}
\]

**Proposition 3.** A strategy \( \hat{x} \in S_m \) is an ESSGS if and only if it is locally superior.

**Proof.** The proof is essentially identical to that of Proposition 2.6 in Weibull (1995, pp. 45–46). There are only two changes that need to be made: (i) The score function now is \( f(\epsilon, y) = (\hat{x} - y) \cdot \pi(\epsilon y + (1 - \epsilon)\hat{x}) \) and (ii) now there is *not necessarily* “at most one \( \epsilon \)” for which \( f(\epsilon, y) = 0 \). This is because the payoff function is not necessarily bilinear (that is \( \pi \) is not necessarily linear). In the case where there is more than one such \( \epsilon \), we can set \( \epsilon_0 = \min\{\epsilon \in (0, 1) | f(\epsilon, y) = 0\} \). Now everything is in place and the result carries through. \( \square \)

### 4 Group Selection and Evolutionary Game Theory

In evolutionary models with random matching, there is a clear and well-known connection between dynamic models of the replicator type and game theoretic concepts such as Nash equilibrium and evolutionary stable strategies (Hofbauer and Sigmund, 1998). The main purpose of this section, and indeed the main theoretical contribution of this paper, is to show that the previous section’s notions of a Nash equilibrium and evolutionarily stable strategy with group selection (NEGS and ESSGS, respectively) are for evolutionary models of group selection what Nash equilibria and ESS are for evolutionary models with random matching. Precisely, we are going to show that any Nash equilibrium under group selection (NEGS) is a steady state of the corresponding evolutionary dynamical system (the replicator dynamics).\(^{16}\) Furthermore, we are going to prove that any *stable steady state* of the replicator dynamics (be it Lyapunov or in the \( \omega \)-limit sense) will be a NEGS. These results directly parallel known results on models with random matching (see e.g. Chapter 7.2.1. in Hofbauer and Sigmund (1998)). Finally, we will prove that any ESSGS is asymptotically stable for the associated replicator dynamics.\(^{17}\) Again, this result transfers a well-known result from the random matching case over to models with non-random matching/group selection (see e.g. Proposition 3.10. in Weibull (1995)).

Before turning to the main results, the following observation clarifies the exact relationship between our results and the mentioned results on random matching. Precisely, our results generalize

\(^{16}\)Since the evolutionary dynamics have the same steady states in discrete and continuous time, this statement obviously applies to either.

\(^{17}\)Note that, just as in the standard case with random matching, the stability statements refer to continuous time replicator dynamics only.
existing ones, since as we now proceed to show, if matching is assumed to be random (example 1.2) in a group selection game, one precisely recoups the traditional Nash Equilibrium concept:

**Theorem 4.** Let \(< I, G, f >\) be a group selection game under random matching. Then the set of Nash equilibria with group selection coincides with the set of symmetric Nash equilibria in the underlying normal form game \(G\). Likewise, when matching is random the set of evolutionarily stable strategies with group selection coincides with the set of evolutionarily stable strategies.

**Proof.** See Appendix B.

Theorem 4 shows that NEGS unifies the treatment of models with or without random matching in evolutionary game theory. In particular, theorem 5 implies as special cases Theorem 7.2.1. in Hofbauer and Sigmund (1998) and Proposition 3.10. in Weibull (1995). But of course, the more interesting cases arise when matching is not random.

**Theorem 5.** Let \(< I, G, f >\) be a group selection game and assume that \(f\) satisfies the assumptions of Theorem 1 and consider the evolutionary steady states of the associated dynamical systems (9)-(10). Then,

1. Any NEGS is a steady state of the discrete time replicator dynamics (9) as well as the continuous time replicator dynamics (10).

2. If \(x^*\) is the \(\omega\)-limit of an orbit \(x(t)\) of the replicator dynamics (10) that lies everywhere in the interior of \(S_m\), then \(x^*\) is a NEGS.

3. If \(x^*\) is Lyapunov stable for the replicator dynamics (10), then \(x^*\) is a NEGS.

4. Assume that \(f\) is of class \(C^1\). Then if \(x^*\) is an ESSGS, it is asymptotically stable under the replicator dynamics (10).

**Proof.**

1. Let \(x^* \in S_m\) be a NEGS, \(I(x)\) ≡ \(\{j \in M | x_j > 0\}\) and \(O(x)\) ≡ \(\{j \in M | x_j = 0\}\). Then, from (13) we get for all \(y \in S_m:\) \(\sum_{j \in M} y_j \pi_j(x^*) \leq \sum_{l \in I(x^*)} x^*_l \pi_l(x^*) + \sum_{q \in O(x^*)} x^*_q \pi_q(x^*)\). Hence:

\[
\sum_{j \in M} y_j \pi_j(x^*) \leq \sum_{l \in I(x^*)} x^*_l \pi_l(x^*)
\]

Now let \(p = \arg \max_{j \in M} \pi_j(x^*)\) and \(r = \arg \max_{l \in I(x^*)} \pi_l(x^*)\). Clearly, \(\sum_{l \in I(x^*)} x^*_l \pi_l(x^*) \leq \pi_r(x^*) \leq \pi_p(x^*)\) where the second inequality holds because \(I(x^*) \subseteq M\). Hence for all \(y \in S_m:\)

\[
\sum_{j \in M} y_j \pi_j(x^*) \leq \sum_{l \in I(x^*)} x^*_l \pi_l(x^*) \leq \pi_r(x^*) \leq \pi_p(x^*)
\]
Taking \( y = (0, \ldots, 0, \frac{1}{p}, 0, \ldots, 0) \), we get \( \pi_p(x^*) \leq \sum_{i \in I(x^*)} x_i^* \pi_i(x^*) \leq \pi_r(x^*) \leq \pi_p(x^*) \) which obviously implies that \( \sum_{i \in I(x^*)} x_i^* \pi_i(x^*) = \pi_r(x^*) \). But this is only possible if \( \pi_j(x^*) = \pi_k(x^*) \) for all \( j, k \in I(x^*) \), and this in turn implies that \( \pi_j(x^*) = x^* \cdot \pi_i(x^*) \) for all \( j \in I(x^*) \). From equation (10), we therefore get \( \dot{x}_j^* = 0 \) for all \( j \in M \), i.e., \( x^* \) is a steady state.

2. Assume that \( x(t) \in \text{int} S_m \) converges to \( x^* \) and that \( x^* \) is not a NEGS. That \( x^* \) is not a NEGS means that there exists a \( j \) with \( e_j \cdot \pi(x^*) = \pi_j(x^*) > x^* \cdot \pi(x^*) \). Hence \( (\pi_j(x^*) - x^* \cdot \pi(x^*)) \geq \epsilon > 0 \) for some \( \epsilon > 0 \). Since \( x(t) \) converges and \( \pi_j \) is continuous on the interior, \( (\pi_j(x(t)) - x^* \cdot \pi(x(t))) \to 0 \) as \( t \to \infty \). This is a contradiction. Note that at the boundary, this holds because we have defined the \( \pi_j \)'s so that they are continuous onto the boundary. If we had not done that, the claim would in general be false for a vector \( x^* \) on the boundary. This same problem does not arise with random matching/in the usual replicator dynamics setting because the payoff functions trivially are continuous. This shows exactly why our “continuous extension to the boundary” is the right thing to do.

3. Precisely as in the previous proof and by continuity of the \( \pi_j \)'s we get that if \( x^* \) is not an NEGS then there exists an \( \epsilon > 0 \) such that for all \( x \) in a neighborhood of \( x^* \): \( (\pi_j(x) - x \cdot \pi(x)) \geq \epsilon > 0 \). For such \( x \), the component \( x_i \) increases exponentially which contradicts Lyapunov stability.

4. Following Weibull (1995, pp. 95–100), we will use Lyapunov’s direct method to prove the proposition. What we need is to find a scalar function \( H \) that is defined on a neighborhood \( Q \) of \( x^* \) which has the following properties: (i) \( H \) is continuously differentiable on \( Q \), (ii) \( H(x^*) = 0 \), (iii) \( H(y) > 0 \) for all \( y \in Q \setminus \{x^*\} \) and (iv) \( \dot{H}(y) = \frac{d}{dt} H(y) < 0 \) for all \( y \in Q \setminus \{x^*\} \).

Let us consider the set \( Q_{x^*} \equiv \{y \in S_m | I(x^*) \subseteq I(y)\} \) i.e. the set of all states that assign positive weights to all the pure strategies that \( x^* \) assigns positive weights. Obviously, \( x^* \in Q_{x^*} \) and \( Q_{x^*} \) is an open set (in the topology induced from \( \mathbb{R}^m \)). So, \( Q_{x^*} \) is a neighborhood of \( x^* \). We will show that the function \( H_{x^*} : Q_{x^*} \to \mathbb{R} \) defined by \( H_{x^*}(y) = \sum_{j \in I(x^*)} x_j^* \log \left( \frac{x_j^*}{y_j} \right) \) satisfies all of the above conditions (i–iv) i.e. that it is a strict local Lyapunov function on \( Q_{x^*} \).

First of all, it is easy to verify that (i) \( H_{x^*} \) is continuously differentiable on \( Q_{x^*} \) and that (ii) \( H_{x^*}(x^*) = 0 \). Now, as \( x^* \) is an ESSGS, we know that there exists a neighborhood \( U \) of \( x^* \) such that condition (14) holds for all \( y \in U \). We will consider the restriction of \( H_{x^*} \) on the set \( U \cap Q_{x^*} \), a neighborhood of \( x \). The next step is to show that \( H_{x^*} \) is strictly positive on \( U \cap Q_{x^*} \). As the function \( -\log(\cdot) \) is convex, we get from Jensen’s inequality:

\[
H_{x^*}(y) = \sum_{j \in I(x^*)} x_j^* \left( -\log \left( \frac{y_j}{x_j^*} \right) \right) \geq -\log \left( \sum_{j \in I(x^*)} x_j^* \left( \frac{y_j}{x_j^*} \right) \right) \geq -\log \left( \sum_{j \in M} y_j \right) = 0
\]
Now, in the case where $I(x^*) = I(y)$, the first inequality is strict (because of the log's strict concavity) and in the case where $I(x^*) \subset I(y)$, the second inequality is strict. In any case, we will always have that (iii) $H_{x^*}(y) > 0$ for all $y \in U \cap Q_{x^*} \setminus \{x^*\}$.

The last step is to show that $\dot{H}_{x^*}$ is negative for all $y \in U \cap Q_{x^*} \setminus \{x^*\}$. Indeed:

$$\dot{H}_{x^*}(y) = \sum_{j \in I(x^*)} \partial_j H_{x^*}(y) \dot{y}_j = -\sum_{j \in I(x^*)} \frac{x_j^*}{y_j} \dot{y}_j$$

and using equation (10), we get:

$$\dot{H}_{x^*}(y) = -x^* \cdot \pi(y) + \bar{\pi}(y) = (y - x^*) \pi(y)$$

which we know is negative because of (14). So, the final condition (iv) is satisfied. 

Remember from the discussion at the end of section 2.3, that the evolutionary/replicator dynamics model has limited predictive power without accompanying stability analysis (e.g., any uniform population state will be a steady state). This places high demands on the analyst because models with non-random matching lead to complicated non-linear dynamical systems whose stability properties are non-trivial to analyze. As we shall see in the following sections, our new game theoretic concepts (NEGS and ESSGS) to a large extent overcome these problems. The reason is the previous theorem which ensures that when we look at the set of NEGS we select all steady states for the replicator dynamics that are (Lyapunov) stable, in particular we capture any steady state for the replicator dynamics that is also an ESSGS.

5 Some Examples

In this section we analyze a number of group selection games with 2 players and 2 strategies under different matching rules. We apply a method that allows us to graphically portray matching rules and makes the process of finding NEGSs and ESSGSs as simple as finding the intersections of two curves.\textsuperscript{18}

We also provide comparative statics results for the class of matching rules with a constant index of assortativity.

5.1 Hawk-Dove/Chicken

A game often analyzed in the literature of both economics and biology is the Hawk-Dove (HD) game.\textsuperscript{19}

Players in this game have two available pure strategies: Hawk (H) and Dove (D). In our formalization,

\textsuperscript{18}The method is described in Appendix A in detail.
\textsuperscript{19}Economists usually refer to this game as Chicken rather than Hawk/Dove.
a Hawk/Dove game is a 2 × 2 game with $A_2^2 > A_1^1 > A_1^2 > A_2^1$. The payoff matrices of three Hawk-Dove games are depicted in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>D</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>50,50</td>
<td>40,80</td>
</tr>
<tr>
<td>H</td>
<td>80,40</td>
<td>0,0</td>
</tr>
</tbody>
</table>

(a) $A_1^1 + A_2^2 > 2A_1^2$

<table>
<thead>
<tr>
<th></th>
<th>D</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>50,50</td>
<td>40,60</td>
</tr>
<tr>
<td>H</td>
<td>60,40</td>
<td>0,0</td>
</tr>
</tbody>
</table>

(b) $A_1^1 + A_2^2 = 2A_1^2$

<table>
<thead>
<tr>
<th></th>
<th>D</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>50,50</td>
<td>20,60</td>
</tr>
<tr>
<td>H</td>
<td>60,20</td>
<td>0,0</td>
</tr>
</tbody>
</table>

(c) $A_1^1 + A_2^2 < 2A_1^1$

Table 1: The payoff matrices of three Hawk-Dove games.

In this game, there are three Nash Equilibria: Two asymmetric ones in pure strategies $(H, D)$ and $(D, H)$ and a symmetric one in mixed strategies where both players play Dove with probability $p_D = \frac{A_2^1 - A_3^2}{A_1^1 + A_2^2 - A_1^1 - A_2^2}$ and Hawk with probability $p_H = \frac{A_1^1 - A_2^2}{A_1^1 + A_2^2 - A_1^1 - A_2^2}$. In the group selection game the state will be summarized by $x$ which indicates the proportion of the population that follows $D$.

Equilibria of the Group Selection Game

Now, in order to find the NEGS and ESSGSs of the PD game, we follow the methodology proposed in Appendix A. The equilibrium curves of the games in Table 1 are shown in Figure 1.

![Figure 1: Equilibrium curves of the games in Table 1.](image)

Random Matching As expected, the unique equilibrium of the group selection game under the Random Matching rule yields the unique symmetric Nash equilibrium of the game where a proportion $x^* = \frac{A_1^1 - A_2^2}{A_1^1 + A_2^2 - A_1^1 - A_2^2}$ of the population play $D$.

Complete Segregation Under complete segregation, there is a unique equilibrium of the group selection game $x^* = 1$ where the whole population follows $D$.

---

20As a convention in what follows and without loss of generality we will assume that $A_1^2 \geq A_2^2$. 19
**Constant Index of Assortativity**  Under a constant index of assortativity rule (see Appendix A.1), the group selection game has a unique equilibrium given by:

\[
x^* = \begin{cases} 
\frac{A_1^3 - A_3^2}{A_2^2 + A_4 - A_1^2} & \text{if } 0 \leq \alpha < \frac{A_2^2 - A_1^2}{A_3^2 - A_2^2} \\
1 & \text{if } \frac{A_2^2 - A_1^2}{A_3^2 - A_2^2} \leq \alpha \leq 1
\end{cases}
\]  \hspace{1cm} (18)

The equilibrium-finding process is shown in Figure 2 for constant index of assortativity rules for different values of \(\alpha\). The comparative statics results are summarized in Figure 3.

---

**Welfare**

In order to conduct welfare analysis, we use the methodology described in Appendix A.3. The isogrowth diagram of a Hawk/Dove game is shown in Figure 4. The comparison of equilibrium welfare in the group selection game and the normal form game is shown in figure 5. Notice that the equilibrium welfare curve is not defined for \(x \in [0, \frac{A_1^3 - A_3^2}{A_2^2 + A_4 - 2A_1^2}]\) as these states can never be attained as equilibria of the group selection game. In all HD games, the level of equilibrium welfare is strictly increasing with the proportion of Doves in the population and thus, maximum equilibrium welfare is obtained when the equilibrium state is \(x = 1\) i.e. when the whole population follows D.

Now, in the case where \(A_1^3 + A_2^2 \leq 2A_1^1\), maximum equilibrium welfare coincides with the maximum expected payoff players using symmetric strategies can get in the normal form game (which is attained when both players play D with certainty).

In the case where \(A_1^3 + A_2^2 > 2A_1^1\), the normal form game maximum expected payoff (under symmetric strategies) is obtained if both players play D with probability \(p_D^* = \frac{A_1^2 + A_2^2 - 2A_1^1}{2(A_1^2 + A_2^1 - A_1^1 - A_2^1)}\). However, when a
matching rule that makes $x = p D^*$ an equilibrium is implemented, equilibrium welfare is reduced below $A_1^1$. This is because the proportion of Hawk-Dove pairs – which are efficient in the utilitarian sense – is reduced in favor of more Hawk-Hawk and Dove-Dove pairs which are not as efficient.

![Figure 4: Isogrowth diagram.](image)

![Figure 5: Equilibrium welfare and normal form payoff.](image)

### 5.2 Stag Hunt

Another game with interesting insights on social behavior is the Stag Hunt.\(^{21}\) In our notation a SH game will have values $A_1^1 > A_2^2 \geq A_3^2 > A_1^2$. The payoff matrices of three SH game are depicted in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>S</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>100,100</td>
<td>0,70</td>
</tr>
<tr>
<td>H</td>
<td>70,0</td>
<td>60,60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>S</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>100,100</td>
<td>0,70</td>
</tr>
<tr>
<td>H</td>
<td>70,0</td>
<td>70,70</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>S</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>100,100</td>
<td>0,80</td>
</tr>
<tr>
<td>H</td>
<td>80,0</td>
<td>70,70</td>
</tr>
</tbody>
</table>

Table 2: The payoff matrices of three SH games.

The game has three Nash equilibria, all symmetric. Two of them are in pure strategies (S,S) and (H,H) and one in mixed strategies where both players play S with probability $p_S = \frac{A_1^1 - A_2^2}{A_1^1 + A_2^1 - A_3^1}$ and H with probability $p_H = \frac{A_1^2 - A_2^1}{A_1^1 + A_2^1 - A_3^1}$. Also, we require that $A_2^2 + A_3^2 > A_1^1 + A_1^2$ so that even though the pure strategy equilibrium (S,S) is payoff dominant (i.e. it yields higher payoffs for both players), the pure strategy equilibrium (H,H) is risk dominant (i.e. if we assume that players are not sure which strategy their opponent will follow and assign equal probabilities to the two strategies, then the expected payoff from playing H exceeds the expected payoff from playing S).\(^ {22}\)

The importance of the Stag Hunt is that it shows that although the efficient outcome (S,S) is a Nash equilibrium, it may not always be selected. More than that, it has been shown that in some stochastic evolutionary models the risk dominant outcome occurs with probability 1 (Young, 1993) and that in

\(^{21}\)For an extensive analysis see Skyrms (2004).

\(^{22}\)See Carlsson and Van Damme (1993).
global games, the risk dominant outcome is the only one that survives iterative elimination of dominated strategies when noise tends to vanish (Carlsson and Van Damme, 1993). So the literature suggests that in several environments it is the risk dominant rather than the payoff dominant outcome that prevails. We show that in our model this inefficiency can be amended under matching rules with high enough assortativity.

Equilibria of the Group Selection Game

The equilibrium curves of the games in Table 2 are shown in Figure 6.

![Equilibrium curves of the games in Table 2.](image)

Figure 6: Equilibrium curves of the games in Table 2.

Random Matching As before, under the Random Matching rule, as expected, we get that there are three NEGS in the group selection game that coincide with the three Nash equilibria of the normal form game: two stable ones (ESSGSs) at $x = 0$ and $x = 1$ (where the whole population follows H and S respectively) and a NEGS which is not an ESSGS where a fraction of the population $x = \frac{A_1^1 - A_1^2}{A_1^1 - A_1^2 + A_1^2 - A_2^2}$ follows S.

Complete Segregation Under the complete segregation matching rule, there is only one NEGS where the whole population follows S ($x = 1$) and it is also an ESSGS.

Constant Index of Assortativity Under a matching rule with a constant index of assortativity $\alpha$ we have two cases depending on the value of $\alpha$:

- if $\alpha \leq \frac{A_2^2 - A_1^1}{A_1^1 - A_1^2}$ we have three NEGS: two NEGS that are also ESSGSs where everybody follows H ($x = 0$) or S ($x = 1$) and a NEGS which is not an ESSGS where a proportion of the population $x = \frac{A_1^2 - A_1^1 + A_1^1 - A_2^1}{A_1^1 + A_2^2 - A_1^1 - A_2^2}$ follows S.
• if $\alpha > \frac{A_1^3 - A_2^3}{A_1^1 - A_2^1}$ there is only one NEGS that is also an ESSGS where the whole population follows S ($x = 1$).

The equilibrium-finding process is shown in Figure 7 for constant index of assortativity rules for different values of $\alpha$. The comparative statics results are summarized in Figure 8.

Figure 7: NEGS with a constant index of assortativity.

Figure 8: NEGS as a function of the index of assortativity.

As in the case of the Hawk/Dove game, in the Stag Hunt there are some states that cannot be attained as equilibria under any matching rule. At these states, namely $x \in \left(\frac{A_1^3 - A_2^3}{2A_1^1 - A_1^1 - A_2^1}, 1\right)$, the dynamics will tend to lead the population towards $x = 1$ where they all follow S under any matching rule. So, if it happens that the system reaches one of these states, then it will be eventually brought to the state where the whole population uses the efficient strategy S.

**Risk Dominance** Notice that there is a value $\alpha^* = \frac{(A_2^3 - A_1^3) - (A_1^1 - A_2^1)}{(A_2^1 - A_1^1) + (A_1^1 - A_2^1)}$ for which the basin of attraction of the ESSGS at $x = 1$ is greater than that of the ESSGS at $x = 0$ iff $\alpha \in (\alpha^*, 1]$. We can interpret that as follows: Assume that players in the population do not know whether each of the other players is going to play S or H and so, using the principle of insufficient reason, they ascribe equal probabilities (equal to 0.5 each) to each other player following S and H.\(^{23}\) Then, if $\alpha \in (\alpha^*, 1]$ the expected payoff for a player following S is higher than his expected payoff when he follows H and so, given the aforementioned beliefs, it is a best response for all of them to follow H, leading to the state being $x = 1$. Conversely when $\alpha \in [0, \alpha^*)$.

So, in the terms described above, we can have a notion of *risk dominance* in the group selection game. Of course – having assumed that $A_2^3 + A_2^3 > A_1^1 + A_1^1$ as is usually done in Stag Hunt games – in the case where $\alpha = 0$, it is always the case that the risk dominant equilibrium is the one where the whole population follows H ($x = 0$).

\(^{23}\)See also Carlsson and Van Damme (1993).
Welfare

The isogrowth diagram of a Stag Hunt game is shown in Figure 9. The comparison of equilibrium welfare in the group selection game and the normal form game is shown in figure 10. Notice that the equilibrium welfare curve is not defined for $x \in \left(\frac{A_1^2 - A_1^3}{2A_1^1 - A_1^2}, 1\right)$ as these states can never be attained as equilibria of the group selection game. The maximum level of welfare is obtained when the equilibrium state is the one where everybody follows S ($x = 1$) and it coincides with the maximum expected payoff players using symmetric strategies can get in the normal form game.

![Figure 9: Isogrowth diagram.](image)

![Figure 10: Equilibrium welfare and normal form payoff.](image)

5.3 Prisoner's Dilemma

As a final application, we leave arguably the most analyzed game in the literature and which has served as the canonical way to model altruistic behavior: The Prisoner's Dilemma (PD). The two players involved in the game have two possible (pure) strategies each: Cooperate (C) or Defect (D). In our notation, a PD game is a game with $A_2^1 > A_1^1 > A_2^3 > A_1^2$. The payoff matrices of three PD games are shown in Table 3.

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>40,40</td>
<td>0,100</td>
</tr>
<tr>
<td>D</td>
<td>100,0</td>
<td>20,20</td>
</tr>
</tbody>
</table>

(a) $A_2^1 + A_2^3 > A_1^1 + A_1^2$   
(b) $A_2^1 + A_2^3 < A_1^1 + A_1^2$   
(c) $A_2^1 + A_2^3 = A_1^1 + A_1^2$

Table 3: The payoff matrices of three Prisoner's Dilemma games.

In any PD game, there exists a unique pure strategy Nash equilibrium (D,D) as defection strictly dominates cooperation. The outcome is far from optimal as there is an obvious Pareto improvement if we move to (C,C).
Equilibria of the Group Selection Game

The equilibrium curves the Prisoner’s Dilemma games of Table 3 are shown in Figure 11.

![Equilibrium Curves](image)

Figure 11: Equilibrium curves of the games in Table 3.

**Random Matching**  Under random matching, $\phi(x) = 2x(1-x)$. Out of the three conditions (21), (22) and (23), only (22) is satisfied for any PD game and this agrees with what we expected as the NEGS under random matching should coincide with the Nash Equilibrium i.e. all follow D ($x = 0$). It’s easy to check using condition (26) that the NEGS at $x = 0$ is also an ESSGS.

**Complete Segregation**  Under complete segregation, $\phi(x) = 0$ and so, only condition (23) is satisfied. Thus, the unique NEGS is $x = 1$, i.e. pure cooperation. This state is also an ESSGS.

**Constant Index of Assortativity**  In the case of a ‘constant index of assortativity’ rule, $\phi(x) = 2(1 - \alpha)x(1-x)$ (see example 3 in section 2.2). Depending on the value of $\alpha$, we get all three cases. As intuition would suggest, the higher the assortativity, the higher the level of cooperation in equilibrium.

1. If $A_1^2 + A_2^2 > A_1^1 + A_2^3$, then there is a unique equilibrium given by:

   $x^*(\alpha) = \begin{cases} 
   0 & \text{if } \alpha \leq \frac{A_1^2 - A_1^1}{A_2^2 - A_1^1} \\
   \frac{A_1^2 - A_1^1 + A_2^2 - A_1^1}{A_2^1 - A_1^1 + A_2^2 - A_1^1} & \text{if } \frac{A_1^2 - A_1^1}{A_2^2 - A_1^2} < \alpha < \frac{A_2^2 - A_1^1}{A_2^2 - A_1^2} \\
   1 & \text{if } \alpha \geq \frac{A_1^2 - A_1^1}{A_2^2 - A_1^2} 
   \end{cases}$

2. If $A_1^2 + A_2^2 < A_1^1 + A_2^3$ then
   
   (a) if $\alpha < \frac{A_1^2 - A_1^1}{A_2^2 - A_1^2}$, there is a unique equilibrium at $x^* = 0$ (all play D),
   
   (b) if $\alpha > \frac{A_1^2 - A_1^1}{A_2^1 - A_1^1}$, there is a unique equilibrium at $x^* = 1$ (all play C),
(c) If \( \alpha = \frac{A_2^2 - A_1^2}{A_2^2 - A_2} \) or \( \alpha = \frac{A_1^2 - A_2^2}{A_1^2 - A_1} \), there are two equilibria: one at \( x_1^* = 0 \) and one at \( x_2^* = 1 \) and

(d) If \( \frac{A_2^2 - A_1^2}{A_2^2 - A_2} < \alpha < \frac{A_1^2 - A_2^2}{A_1^2 - A_1} \), there are three equilibria: one at \( x_1^* = 0 \), one at \( x_2^* = \frac{A_2^2 - A_1^2 + A_1^2 - A_1^2}{A_1^2 - A_2^2 - A_1^2 - A_2} \), and one at \( x_3^* = 1 \).

3. If \( A_2^2 + A_1^2 = A_3^2 + A_1^2 \) then

(a) If \( \alpha < \frac{A_2^2 - A_1^2}{A_2^2 - A_2} \), there is a unique equilibrium at \( x^* = 0 \) (all play D),

(b) If \( \alpha > \frac{A_2^2 - A_1^2}{A_2^2 - A_2} \), there is a unique equilibrium at \( x^* = 1 \) (all play C) and

(c) If \( \alpha = \frac{A_2^2 - A_1^2}{A_2^2 - A_2} \), there is a continuum of equilibria. Actually, any \( x \in [0, 1] \) is an equilibrium.

The equilibrium-finding process for all three cases is shown in Figure 12 for constant index of assortativity rules with different values of \( \alpha \). The comparative statics results are summarized in Figure 13.

---

**Figure 12:** NEGS for three different cases of Prisoner's Dilemma games under a constant index of assortativity.

---

**Figure 13:** NEGS for three different cases of Prisoner's Dilemma as a function of the index of assortativity.
**Risk Dominance** In the case where \( A_1^2 + A_2^2 < A_1^1 + A_3^2 \) (where two ESSGs exist for certain values of \( \alpha \)), the risk dominant equilibrium (in the sense introduced in 5.2) is the one where all play D (\( x = 0 \) when \( \alpha < \alpha^* = \frac{(A_2^2 - A_1^2) - (A_1^1 - A_3^2)}{(A_2^2 - A_1^1) + (A_3^2 - A_2^2)} \) and the one where all play C (\( x = 1 \)) when \( \alpha > \alpha^* \), as was the case in the SH game.

Notice that unlike the HD and the SH games, in a PD game, all states can be attained as equilibria if an appropriate matching rule is selected.

**Welfare**

The isogrowth diagrams of three Prisoner’s Dilemma games are shown in Figure 14. The comparison of equilibrium welfare in the group selection game and the normal form game for each of the three cases is shown in figure 5. The maximum level of welfare is obtained when the equilibrium state is the one where all cooperate (\( x = 1 \)) and it coincides with the maximum expected payoff players using symmetric strategies can get in the normal form game when \( A_1^2 + A_2^2 > 2A_1^1 > A_1^1 + A_3^2 \). In the case where \( A_1^1 + A_3^2 > 2A_1^1 \) the maximum value of welfare in the normal form game is obtained when both players play C with probability \( p_C = \frac{A_1^1 + A_3^2 - 2A_1^1}{2(A_1^1 + A_3^2 - A_1^1 - A_2^2)} \). However, when this state is implemented as an equilibrium in the group selection game, it does not grant the players such high expected payoffs as the frequency of (C,D) or (D,C) pairs is not high enough. The implementation of an assortative matching rule can make the state an equilibrium but this happens at the expense of obtained payoff at that state. Also, if we restrict ourselves to equilibrium payoffs, then the payoff obtained at \( x = \frac{A_1^1 + A_3^2 - 2A_1^1}{2(A_1^1 + A_3^2 - A_1^1 - A_2^2)} \) is no longer the optimal payoff. Once again, utilitarian optimality is achieved when \( x = 1 \) (all cooperate) is implemented as an equilibrium.

![Isogrowth diagrams for three PD games.](image-url)
6 Group Selection and the Fitness of Populations

Group selection can explain behavioral traits such as altruism or cooperation which cannot arise in Nash equilibrium and so cannot be favored by natural selection if matching is random (see theorem 4). Importantly, such departures from egoism may be superior to the outcomes under random matching in the sense that the average fitness may be higher. The classical example here is of course the prisoners’ dilemma where the outcome of random matching yields lower average fitness than outcomes with assortative matching (see section 5.3 and also Bergström (2002)). In this section we are going to discuss these issues drawing on both the abstract results and the concrete examples of the previous sections. As will become clear, our new concepts (NEGS and ESSGS) allow us to push the discussion substantially forward in comparison with existing literature.

First, we need to define the concepts involved. Recall from section 2 that the average fitness \( \bar{\pi}(x) \) at a population state \( x \in S_m \) is given by \( \bar{\pi}(x) = \sum_{j=1}^{m} x_j \pi_j(x) \) (equation (8)). In the context of a group selection game \( <I, G, f> \), we referred instead to this as the average payoff or the welfare (see equation (12)). Since average fitness in the evolutionary model is obviously equal to average payoff in the (evolutionary) game theory model, and since by theorem 5 we know how the various equilibrium/ steady state concepts relate to each other, we need not differentiate between them in what follows. Accordingly, we use the term average fitness exclusively from now on. Average fitness at a population state \( x \) will from now on be denoted by \( \bar{\pi}_f(x) \) so as to explicitly mention the matching rule. This allows us to easily compare average fitnesses under different matching rules for a fixed underlying payoff structure/normal form game \( G \) (e.g., prisoners’ dilemma or hawk-doves).

Now as was already mentioned, random matching — or for that matter any other specifically given matching rule \( f \) — may not maximize average fitness in a NEGS \( x^* \). The prisoners’ dilemma was already mentioned above, but in the previous section we saw that the observation remains valid in other standard 2 by 2 games such as hawk-doves and stag-hunt; and it also remains valid if instead of NEGS we

\[
\text{Figure 15: Comparison between equilibrium welfare and normal form payoff in three PD games.}
\]
focus on ESSGS. Thus, evolution under non-random matching certainly does not imply fitness maximization. The interesting next question therefore is whether for a fixed underlying normal form game there exists some matching rule given which average fitness will be maximized in NEGS; and if the answer is yes, to characterize these matching rules in concrete situations. Thus in the prisoners’ dilemma, random matching is inferior in average fitness terms, but as we saw in section 5.3, a rule such as complete segregation will lead to equilibria where everybody cooperates and so to average fitness maximization. When discussing this topic it is important to understand that when \( f \) is varied, not only does the set of NEGS (and ESSGS and also, the set of steady states of the replicator dynamics) change — the average fitness \( \bar{\pi}_f(x) \) will also change at any given population state \( x \). So if some population state maximizes welfare but is not a NEGS at some matching rule \( f_1 \), it could be a NEGS at another matching rule \( f^2 \) but no longer maximize welfare! Any sensible discussion must therefore consider the joint selection of a population state and matching rule as captured by the following definition.

**Definition 8. (Evolutionary Optimum)** Let \( G \) be a normal form game. A population state \( x^* \in S_m \) together with a matching rule \( f^* \in \mathcal{F}_{n,m} \) is said to be an evolutionary optimum if \( \bar{\pi}_{f^*}(x^*) \geq \bar{\pi}_f(x) \) for all \( (x,f) \in \mathcal{E} = \{(x,f) \in S_m \times \mathcal{F}_{n,m} : x \text{ is a steady state of } (I,G,f)\} \).

Intuitively, a population state \( x^* \) and a matching rule \( f^* \) form an optimum if they lead to maximum average fitness of the population among all population state/matching rule combinations that satisfy the steady state restriction. Note that the restriction to steady states is entirely natural here: Any population state that is not a steady state under some matching rule would immediately be “destroyed” by natural selection. Given these definitions, we can now answer the previous question:

**Theorem 6.** Let \((x^*,f^*)\) be an evolutionary optimum. Then there exists a matching rule \( h \in \mathcal{F}_{n,m} \) which satisfies the assumptions of theorem 1, such that \( x^* \) is a NEGS under \( h \), and such that \((x^*,h)\) is an evolutionary optimum (in particular, \( \bar{\pi}_h(x^*) = \bar{\pi}_f(x^*) \)).

**Proof.** See Appendix D. Theorem 6 can be thought of as the “second welfare theorem of evolution” telling us that any evolutionary optimum can be “decentralized” in the evolutionary environment through some matching rule. That this should be so is easy to see in simple cases, but it is in general a surprising result. In most standard games (including the ones considered in this paper), there is a premium on coordination/uniformity, and so what is needed in order to reach an evolutionary optimum is a sufficiently high level of assortativity. In games where there is a premium on agents in a group being different — e.g.,

24Note in this connection that any uniform population state is a steady state (in fact, any uniform population state is a steady state under any matching rule).
due to specialization — it will instead be a sufficiently high degree of dissociation that leads to evolutionary optimality. To illustrate, consider first the Hawk-Doves model of section 5.1. As was shown in that section, the Hawk-Doves model has a unique evolutionary optimum, namely the state where all individuals are doves (since this state is uniform, it will be a steady state of the replicator dynamics under any matching rule). As was also shown in section 5.1, the doves only outcome is not a NEGS for all matching rules, however. Specifically, (18) shows that only when matching is sufficiently assortative will a uniform population of doves be a NEGS. Intuitively, what happens when this “assortativity threshold” is crossed is that hawks become so likely to end up with other hawks that it is not worthwhile playing hawk even if the population of hawks is infinitely small. What all of this shows is that as predicted by theorem 6, the evolutionary optimum (doves only) will be a NEGS for some matching rule. But unless the environment is such that matching is sufficiently assortative, evolution will not lead to the evolutionary optimum. In a real-world situation where a specific matching rule and a specific payoff structure is in effect, this of course implies that evolution can easily produce a mixed population of hawks and doves. But as interesting is perhaps that evolution may in fact lead to the evolutionary optimum even without recourse to the extremities of either complete segregation or direct reciprocity. Furthermore, equation (18) tells us exactly which parameters account for the relationship between the level of assortativity in matching and evolutionary optimality. For example, less assortativity is needed if $A_2^2 - A_1^1$ is “small” which simply means that a Dove facing a Dove will gain relatively less from switching to Hawk ($A_1^1$ is the payoff to a Dove facing a Dove, and $A_2^2$ is the payoff to a Hawk facing a Dove).

Another illustration is provided by the Stag Hunt model of section 5.2. As we saw in that section, the Stag Hunt model has multiple NEGS for “low” levels of assortativity: Two uniform population states (everyone hunts for hare, everyone hunts for stag), and a mixed population state. For “high” levels of assortativity, only the state where everyone hunts for stag is a NEGS. The evolutionary optimum is for everyone to hunt for stag (and again this evolutionary optimum is supported by any matching rule since it is uniform). Thus the evolutionary optimum is a NEGS for all levels of assortativity which means that the prediction of theorem 6 bears out in a particularly strong way. But only if assortativity is sufficiently high will the evolutionary optimum be the unique NEGS, and so — just as in the Hawk-Doves model — evolution may not lead to the evolutionary optimum in a real-world situation with a specific matching rule and specific payoffs.27

---

25 Compare with direct reciprocity where the doves only outcome is always supported — along with any other payoff in the maximin set of the associated normal form game — as a subgame perfect Nash equilibrium in the infinitely repeated game (see e.g. Rubinstein (1979)).

26 In fact, this population state will be an ESSGS for such levels of assortativity, and so cannot be invaded by hawks even in the highly demanding sense of an ESSGS (in particular, it will be asymptotically stable for the replicator dynamics), see section 5.1 for details.

27 Concerning the situation with multiple NEGS, the standard way to “resolve” multiplicity in the evolutionary setting would be to think of this as a situation with path-dependence so that, depending on initial conditions, a society may end up either as

---

30
7 Conclusion

This paper had two main purposes. Firstly, to extend the existing machinery of evolutionary game theory to include models of group selection; and secondly, to use the new concepts developed to discuss the relationship between different kinds of selection and the fitness of populations. Two new equilibrium concepts were proposed, *Nash equilibrium with group selection* (NEGS) and *evolutionarily stable strategy with group selection* (ESSGS). These equilibrium concepts contain as special cases the standard ones; indeed when matching is random, the set of NEGS is just the symmetric Nash equilibria and the set of ESSGS is the evolutionarily stable strategies (theorem 4). We proceeded to show in our main theoretical result (theorem 5) that NEGS and ESSGS are for models with arbitrary matching rules what Nash equilibrium and ESS are for models with random matching. In particular, any stable steady state of the replicator dynamics is a NEGS and any ESSGS is an asymptotically stable steady state. As in the standard random matching setting, these results form the theoretical foundation upon which evolutionary game theory rests; hence our concepts extend the traditional game theoretic framework to models with group selection. As for the fitness of populations, our main result is the “second welfare theorem” of evolution (theorem 6) which states that any evolutionary optimum will be a NEGS under some matching rule.

From an applied point of view, the great advantage of the game theoretic approach is the additional structure it imposes compared to dynamic models of the replicator type. In particular, the analysis becomes simpler and the results become more powerful. Recall that all uniform population states (all individuals employing the same strategy) are steady states for the replicator dynamics. In fact, the set of steady states includes *everything* that is “evolutionary feasible” (and a good way to think of this set is in fact as evolutionary models’ parallel to the feasible set of an exchange economy). This of course makes stability analysis absolutely critical in the dynamic setting — the problem being that such stability analysis is *not* straight-forward in group selection models where the replicator dynamics forms a complex non-linear dynamical system.\(^28\) In contrast, we saw in section 5 that the set of NEGS and ESSGS can be computed with great ease in group selection games, and equally importantly, the game theoretic formulation allows for abstract analysis and the derivation of general results. An example of such a general result is theorem 4 which states that with random matching, the set of NEGS coincides one where everyone hunts for hares or everyone hunts for stags (as mentioned, the mixed NEGS is not an ESSGS). From our game theoretic perspective, it is however more natural to employ a suitable selection criterion (see the discussion in section 5.2, where we saw that a global games approach will actually favor the hares only outcome for “low” levels of assortativity because this is the *risk-dominant* outcome (Carlsson and Van Damme, 1993)).

\(^{28}\) Thus consider for example the discrete time replicator dynamics of equation (9). Unlike in models with random matching where the \(\pi\)'s are constants, these coefficients will in models with non-random matching depend on the population state through the matching rules in an often very complicated way. This of course makes even local stability analysis a daunting task.
with the symmetric Nash equilibria in the underlying normal form game which intuitively means that random matching precisely corresponds to “self-serving” behavior in general. Such a result would be impossible to establish within the traditional group selection framework of section 2. The “second welfare theorem” of evolution (theorem 6) is another example of this.

Often, matching is simply a geographical phenomenon (think of viruses or neighborhood imitation amongst humans), or could be a reflection of individuals’ limited ability to monitor other individuals (see the introduction for further details). But when matching rules correspond to institutions or conventions, not explaining how they come about misses half the story. A clear weakness of existing group selection models — including the results in this paper — is in this connection that the matching rules are taken as given. An obvious topic for future research would be to model the evolution of the matching rules (i.e., to endogenize them). Consider monitoring: If individuals gain an advantage by increasing their ability to monitor (by increasing their intelligence and memory), we can see how matching rules will over time evolve to be less and less random (typically more and more assortative). This then would be a true endogenous description of matching (institutions, conventions). The simplicity of the game theoretic framework presented in this paper should definitely put such a theory of matching rules within reach.

Appendices

A Finding all equilibria in $2 \times 2$ games

In this section we provide a tool that makes it easy for one to find and visualize NEGSs and ESSGSs in the $2 \times 2$ case. By use of our method, we can easily identify equilibria of such games by looking for intersections between two lines: one that depends on the payoffs (the equilibrium curve) and one that depends on the matching rule in effect (the matching rule curve). An example is shown in Figure 16; the equilibrium state is at the intersection of the two lines.

In what follows, we analyze games that have a payoff bimatrix of the general form presented in Table 4. Without loss of generality, we will assume that $A^1_1 \geq A^3_2$.

<table>
<thead>
<tr>
<th></th>
<th>Strategy 1</th>
<th>Strategy 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategy 1</td>
<td>$A^1_1, A^1_2$</td>
<td>$A^2_1, A^2_2$</td>
</tr>
<tr>
<td>Strategy 2</td>
<td>$A^2_2, A^2_1$</td>
<td>$A^3_2, A^3_1$</td>
</tr>
</tbody>
</table>

Table 4: The general form of a $2 \times 2$ game. $A^1_1 \geq A^3_2$. 

32
A.1 The Matching Rule Curve

A matching rule for the $2 \times 2$ case, will be of the form $f(x) = (f_1(x_1, x_2), f_2(x_1, x_2), f_3(x_1, x_2))$. Now notice that it can be easily described by only defining one of the three coordinates $f_i(x)$. This is because in order for $f$ to be a matching rule, it has to satisfy equations (3) (two linearly independent equations in our example of 2 strategies) so this leaves us with only one degree of freedom. We pick the value of $f_2(x)$ – that expresses the extent to which the two strategies get mixed with one another – to describe the matching rule. Of course, because there are only two strategies available, the state can be summarized by the proportion of individuals using Strategy 1 (the remaining individuals are clearly using Strategy 2). We will use $x$ to denote this proportion and thus to express the state. So any matching rule will be described by a function $\phi : [0, 1] \rightarrow [0, 1]$ and as the matching rule has to satisfy constraint (3), the three coordinates of $f$ can be calculated to be:

$$
\begin{align*}
    f_1(x) &= x - \frac{1}{2} \phi(x) \\
    f_2(x) &= \phi(x) \\
    f_3(x) &= 1 - x - \frac{1}{2} \phi(x).
\end{align*}
$$

(19)

More than that, the conditions $0 \leq f_1(x)$, $0 \leq f_2(x)$ and $0 \leq f_3(x)$ must be satisfied for all $x \in (0, 1)$. From these, we get that the values $\phi$ can take are restricted by:

$$
\begin{align*}
    0 \leq \phi(x) \leq 2x & \quad \text{for } x \in \left[0, \frac{1}{2}\right], \\
    0 \leq \phi(x) \leq 2(1-x) & \quad \text{for } x \in \left(\frac{1}{2}, 1\right].
\end{align*}
$$

(20)

So any matching rule in the case of 2-strategy, 2-person normal form games can be summarized by a function $\phi$ that satisfies (20).

It is now possible for us to draw diagrams that show what matching rules look like. Examples of graphs of matching rules are given in Figure 17. A matching rule is summarized by a line that begins at $(0,0)$, assumes values ‘within’ the triangle bounded by (20) and ends at $(1,0)$.

Equations (3) are in essence ‘balancing conditions’ similar to condition (2) in Alger and Weibull (2012). i.e. They ensure that the number of 1-strategists that are matched with 2-strategists is equal to the number of 2-strategists that are matched with 1-strategists.

Obviously, $x_1 = x$ and $x_2 = 1 - x$. 

Figure 16: Example of finding an equilibrium.
Under this formalization, the random matching rule will be given by
\[ \phi(x) = 2x - 2x^2 \]
whereas the complete segregation rule is simply
\[ \phi(x) = 0. \]

Another example would be the constant index of assortativity rule (Bergström, 2003) which can be summarized by
\[ \phi(x) = 2(1 - \alpha) x (1 - x) \]
where \( \alpha \in [0, 1] \) is the index of assortativity.

### A.2 The Equilibrium Curve

Under any matching rule, it is easy to show that an interior state \( x^* \) is an equilibrium iff:
\[ \pi_1(x^*) = \pi_2(x^*) \iff \]
\[ [(A_2^2 - A_2^3)x^* + (A_1^1 - A_1^2)(1 - x^*)]\phi(x^*) = 2(A_1^1 - A_2^3)x^*(1 - x^*) \]  
(21)

and, looking for boundary equilibria, if \( \phi \) is differentiable at 0 and at 1, for \( x = 0 \) to be an equilibrium, it must be the case that:
\[ \pi_1(0) \leq \pi_2(0) \quad \Rightarrow \quad (A_1^1 - A_2^3) \frac{\partial \phi}{\partial x}(0) \geq 2(A_1^1 - A_2^3) \]  
(22)

and for \( x = 1 \) to be an equilibrium, it must be the case that:
\[ \pi_1(1) \geq \pi_2(1) \quad \Rightarrow \quad (A_2^3 - A_2^2) \frac{\partial \phi}{\partial x}(1) \leq 2(A_1^1 - A_2^3). \]  
(23)

Now, provided that there is actually some strategic interaction occurring between the two players, \textit{i.e.} either \( A_2^2 \neq A_2^3 \) or \( A_1^1 \neq A_1^2 \) (or both), then from condition (21) we get two cases:
• If \( A_1 \neq A_3 \), then an interior state will be an equilibrium iff the value of \( \phi \) for that state is equal to the value of a function \( E \) for that given state. We will call this function the \textit{equilibrium curve} of the game and it is given by:

\[
E(x) = \frac{2(A_1 - A_3)x(1 - x)}{(A_2 - A_3)x + (A_1 - A_2)(1 - x)}. \tag{24}
\]

• In the case where \( A_1 = A_3 \), then the condition for an interior state to be an NEGS is:

\[
\begin{align*}
\phi(x) &= 0 \\
x &= \frac{A_1 - A_2}{A_2 - A_1 + A_1 - A_2} \quad \text{or} \quad x = \phi(x) = 0 \quad \text{or} \quad x = \frac{A_1 - A_2}{A_2 - A_1 + A_1 - A_2}.
\end{align*} \tag{25}
\]

Condition (25) says that any state for which the two strategies do not mix at all will be an equilibrium state (obviously, as no strategy gets an advantage over the other) and, more importantly, that the state \( \frac{A_1 - A_2}{A_2 - A_1 + A_1 - A_2} \) will be an equilibrium \textit{for all matching rules} (as long as this value is within the boundaries \((0,1)\).

\textbf{Stability and the Equilibrium Curve}  If we assume that the matching rule is ‘\( \mathcal{C}^1 \)’, then we can easily check that a state \( x \) will be an ESSGS iff

\[
\begin{align*}
\phi(x) &= E(x) \\
\frac{\partial \phi}{\partial x}(x) &> \frac{\partial E}{\partial x}(x).
\end{align*} \tag{26}
\]

Using the above analysis in conjunction with diagrams like the one in Figure 17 can help us spot NEGS and ESSGSs very easily. All one has to do is to plot the matching rule \( \phi \) and the equilibrium curve \( E \) on the same diagram. If the two lines meet at an interior state, then this state is a NEGS. If along with that the equilibrium curve is above the matching rule to the left of the state and below it to the right of the state, then the state is an ESSGS as well. Finally, for the states 0 and 1, one can say that in order for one of these states to be a NEGS (ESSGS), then it has to be that the slope of the matching rule is greater than (or equal to) the slope of the equilibrium curve at that state.

\textbf{A.3 Welfare in \( 2 \times 2 \) Games}  In the case of \( 2 \times 2 \) games, by using the formalization introduced above, we can make equilibrium welfare considerations. What we are interested in is to see how the different equilibria fare in terms of welfare. For a \( 2 \times 2 \) game, the welfare at state \( x \) when the value of the matching rule at \( x \) is \( \phi \) is given by:

\[
W(x, \phi) = A_2^3 + (A_1^1 - A_3^3)x + \frac{(A_1^2 + A_2^2 - A_1^1 - A_2^3)}{2} \phi \tag{27}
\]
And as long as $A_1^2 + A_2^2 \neq A_1^1 + A_2^3$, solving for $\phi$, we get:

$$\phi = \frac{2(W - A_3^3)}{A_1^2 + A_2^2 - A_1^1 - A_2^3} = \frac{2(A_1^1 - A_3^3)x}{A_1^2 + A_2^2 - A_1^1 - A_2^3}$$

(28)

For any value of $W$, the above equation gives the set of points on the $(x, \phi)$ plane that yield an average payoff of $W$ for the population. We will call such lines isogrowth lines as all points on each of these lines leads to the same growth rate of the population (which is the same as the average payoff). Drawing such lines can help us visualize what is really happening in terms of welfare under the various matching rules. More than that, by combining the isogrowth lines with the equilibrium curves of different games, we can see which matching rules can lead to some (utilitarian) optimality. An example of an isogrowth diagram is depicted in Figure 18. Finally, using the welfare function (27) along with the equilibrium curve (24) we can calculate the equilibrium welfare in the group selection game and then compare that to the expected payoff of a player in the normal form game. Such comparisons are carried out in Section 5 for three classes of $2 \times 2$ games.

**B Proof of Proposition 4**

Let us denote by $y^i \in S_m$ the (mixed) strategy used by player $i$ in the normal-form game $G$ and by $x^{-i} \in S_m$ the common strategy used in the normal-form game $G$ by all player $i$’s opponents. Let also $P_i(y^i|x^{-i})$ be the expected payoff of player $i$ in the normal-form game when he/she is using strategy $y^i$ and all of his opponents use strategy $x^{-i}$. Since $G$ is symmetric, we have $P_i(y^i|x^{-i}) = P_j(y^j|x^{-j})$ for all $i, j \in N$. So we can write $P(y|x)$ to express the expected payoff in the normal-form game of any player using strategy $y$ when all his opponents use the same strategy $x$.

A symmetric Nash equilibrium of game $G$ is a strategy $x^* \in S_m$ such that:

$$P(x^*|x^*) \geq P(y|x^*) \quad \text{for all } y \in S_m.$$
So, for \( x^* \) to be a symmetric Nash equilibrium, if every opponent of any given player \( i \) is using strategy \( x^* \), it must be a best response for player \( i \) to use the same strategy \( x^* \) as well.

On the other hand, a strategy \( x^* \) will be an equilibrium in game \( < I, G, f > \) iff:

\[
\Pi(x^*, x^*) \geq \Pi(y, x^*) \quad \text{for all } y \in S_m
\]

In order to prove the proposition, all we need to show is that

\[
\Pi(y, x) = P(y|x) \quad \text{for all } y \in S_m
\]  

(29)

under the random matching rule. If we let \( e_j \) be the probability vector that corresponds to pure strategy \( j \), then (29) boils down to

\[
\pi_j(x) = P(e_j|x) \quad \text{for all } j \in M.
\]  

(30)

Calculating \( \pi_j(x) \). Let us denote by \( M^i_{-j} \) the set of all strategies other than \( j \) represented in group \( i \) and by \( \Gamma^i_j \) the set of all groups that contain exactly one individual following strategy \( j \). Formally, \( M^i_{-j} = \{ k \in M \setminus \{ j \} | i \in \supp(j) \} \) and \( \Gamma^i_j = \{ i \in \Gamma_n, m | n_j = 1 \} \). Calculating \( \pi_j(x) \) under \( r^{n,m} \) yields:

\[
\pi_j(x) = \sum_{i \in \supp(j)} \frac{(n-1)!x_j^{n_j-1}n_i^i}{(n_j-1)!} \prod_{k \in M^i_{-j}} \frac{x_k}{n_k^i} A^i_j, x \in S_m \setminus bd_j(S_m)
\]  

(31)

\[
\pi_j(x) = \sum_{i \in \Gamma^i_j} (n-1)! \prod_{k \in M^i_{-j}} \frac{x_k}{n_k^i} A^i_j, x \in bd_j(S_m)
\]  

(32)

Calculating \( P(e_j|x) \). In general, all players use mixed strategies \( i.e. \) a randomization over the set of pure strategies \( M \). We will denote the pure strategy a player \( l \) ends up using after the randomization process has taken place – \( i.e. \) the realization of player \( l \)’s mixed strategy – as \( s^l \). The probability of a player ending up in a situation where his/her opponents follow (pure) strategies \( s^{-l} \in M^{n-1} \) with \( s^{-l} = (s^1, \ldots, s^{l-1}, s^{l+1}, \ldots, s^n) \) will be denoted by \( p(s^{-l}) \). When all player \( l \)’s opponents use the same strategy \( x \), those probabilities can be calculated to be:

\[
p(s^{-l}) = \prod_{k \in M} (x_k)v_k(s^{-l})
\]

where \( v_k(s^{-l}) \in \{0, 1, \ldots, n-1\} \) is the number of player \( l \)’s opponents using strategy \( k \) in the ordered set \( s_{-l} \).

Let us fix player \( l \)’s strategy (realization) to be \( s^l = e_j \). Since the game \( G \) is symmetric, the payoff of player \( l \) will not depend on the exact ordering in \( s^{-l} \) but on the vector \( v(s^{-l}) = (v_1(s^{-l}), \ldots, v_m(s^{-l})) \).

37
This means that different \( s^{-1} \)s with the same \( v(s^{-1}) \) will yield the same payoff for player \( l \). The number of the different \( v \) outcomes are elements is \( \gamma_{n-1,m} \). Let us use \( \kappa \in \Gamma_{n-1,m} \) to index the different \( v \). By abusing notation, we can calculate the probability of a specific \( v^\kappa \) to occur as

\[
p(v^\kappa) = \frac{(n-1)!}{\prod_{k \in M} v_k^{n_k} \prod_{k \in M} (x_k)^{v_k}}.
\] (33)

As player \( l \) is using strategy \( j \), if he ends up in a situation where his/her opponents’ realizations are \( \kappa \), it is as if he ends up in a group \( i \) where \( n_k^i = v_k^\kappa \) for \( k \neq j \) and \( n_j^i = v_j^\kappa + 1 \). This group will be in \( \text{supp}(j) \) and we will write \( i = j \triangleright \kappa \) and read: “\( i \) is the group that we get if we add an individual who uses strategy \( j \) to a set of opponents whose realizations are \( \kappa \).” Notice that the probabilities in (33) are independent of player \( l \)’s choice of strategy. So, the probability of player \( l \) ending up in situation \( i \) conditional on him using strategy \( j \) will be the same as the probability realization \( \kappa \) occurring. Using the \( i \)- rather than the \( \kappa \)-indexing, we can rewrite (33) (abusing the notation once again) as:

\[
p(i|j) = p(j \triangleright \kappa|j) = p(v^\kappa) = \frac{(n-1)!x_j^{n_j^i-1}}{(n_j^i - 1)!} \prod_{k \in M_{i-j}} \frac{(x_k)^{v_k^i}}{n_k^i!}.
\]

Now, in each of these cases \( i \), player \( l \) gets a payoff of \( A_j^i \) and his expected payoff is:

\[
P(e_j|x) = \sum_{i \in \text{supp}(j)} p(i|j)A_j^i = \sum_{i \in \text{supp}(j)} \frac{(n-1)!x_j^{n_j^i-1}}{(n_j^i - 1)!} \prod_{k \in M_{i-j}} \frac{(x_k)^{v_k^i}}{n_k^i!} A_j^i.
\] (34)

In the special case where \( x \in \text{bd}_j S_m \), player \( l \) can be sure that he is the only one using strategy \( j \) and thus, the only groups that get positive probability are the ones in \( \Gamma_j^1 \) which have \( n_j^1 = 1 \). So his/her expected payoff is:

\[
P(e_j|x) = \sum_{i \in \Gamma_j^1} p(i|j)A_j^i = \sum_{i \in \Gamma_j^1} (n-1)! \prod_{k \in M_{i-j}} \frac{(x_k)^{v_k^i}}{n_k^i!} A_j^i.
\] (35)

By comparing equation (31) to (34) and equation (32) to (35), we can see that

\[
\pi_j(x) = P(e_j|x)
\]

and as we showed that for an arbitrary \( j \), it holds for all \( j \in M \). \( \square \)

### C Proof of Theorem 1

**Best reply correspondence** The best reply correspondence (BRC) is a correspondence \( B : S_m \rightarrow S_m \) defined by:

\[
B(x) = \{ y \in S_m : (\forall y \in S_m) \Pi(y,x) \geq \Pi(y,x) \}
\]
and gives the mixed strategies an agent can follow so as to maximize his/her expected payoff given that the state is \( x \).

We also define the value function \( V: S_m \rightarrow \mathbb{R} \) that gives the maximum payoff an agent can achieve at any given state. Formally: \( V(x) = \max_{y \in S_m} \Pi(y, x) \).

### C.1 Equilibrium

We intend to show that under some assumptions on \( f \), an equilibrium state always exists. We will prove the existence result by using Kakutani’s fixed point theorem. In order to do that, we need to show that the BRC is convex-valued, nonempty-valued and upper hemicontinuous. These prerequisites are proven in Lemmata 7 and 8.

**Lemma 7** (Convex-valued BRC). For any group selection game under a matching rule \( \mathcal{G} = \langle I, G, f \rangle \) the best reply correspondence \( B \) is convex-valued.

**Proof.** We can identify three different cases for \( B(x) \):

- \( B(x) = \emptyset \) and thus \( B \) is convex-valued at \( x \).

- \( B(x) = \{ y^* \} \) i.e. the best reply correspondence contains only one element at \( x \) and thus \( B \) is convex-valued at \( x \).

- \( B(x) \) contains at least two elements at \( x \) i.e. there exist \( y_1^*, y_2^* \in S_m \) such that

  \[
  \Pi(y_1^*, x) \geq \Pi(y, x) \text{ for all } y \in S_m
  \]

  \[
  \Pi(y_2^*, x) \geq \Pi(y, x) \text{ for all } y \in S_m
  \]

  which is possible only if \( \Pi(y_1^*, x) = \Pi(y_2^*, x) = L \). Now, for all \( \lambda \in [0, 1] \) we have the following series of equalities:

  \[
  \Pi(\lambda y_1^* + (1 - \lambda)y_2^*, x) = (\lambda y_1^* + (1 - \lambda)y_2^*) \cdot \pi(x) =
  \]

  \[
  \lambda y_1^* \cdot \pi(x) + (1 - \lambda)y_2^* \cdot \pi(x) = \lambda \Pi(y_1^*, x) + (1 - \lambda)\Pi(y_2^*, x) =
  \]

  \[
  \Pi(y_1^*, x) = \Pi(y_2^*, x) = L
  \]

  So, for any \( y_1^*, y_2^* \in B(x) \) we get that \( \lambda y_1^* + (1 - \lambda)y_2^* \in B(x) \) for all \( \lambda \in [0, 1] \) and thus \( B \) is convex-valued at \( x \).

Since these are the only possible cases, we can conclude that \( B \) is convex-valued in \( S_m \).

**Lemma 8** (BRC: Nonempty-valued and upper hemicontinuous). For a group selection game under a matching rule \( \mathcal{G} = \langle I, G, f \rangle \), if
1. \( f \) is continuous on \( S_m \) and

2. the partial derivatives \( \partial_j f_i \) for all \( j \in M \) and all \( i \in \text{supp}(j) \) exist on \( \text{bd}_j(S_m) \)

then the best reply correspondence \( B \) is non-empty valued and upper hemicontinuous.

**Proof.** From assumption 2 of the lemma, we get that the limits \( \lim_{\tilde{x} \to x} f_i(\tilde{x}) = \partial_j f_i \) for all \( j \in M \) and all \( i \in \text{supp}(j) \) exist on \( \text{bd}_j(S_m) \) and from the definition of \( \pi_j (7) \), we get that

\[
\lim_{\tilde{x} \to x} \pi_j(\tilde{x}) = \pi_j(x) \text{ on } \text{bd}_j(S_m).
\]

So, \( \pi_j \) are continuous on \( \text{bd}_j(S_m) \) and since all \( f_i \) are continuous on \( S_m \), \( \pi_j \) are continuous on \( S_m \setminus \text{bd}_j(S_m) \) as sums of quotients of continuous functions. So, \( \pi \) is continuous on \( S_m \) and therefore, \( \Pi \) is continuous on \( S_m^2 \).

Now we can see that the conditions for Berge’s maximum theorem are satisfied: (i) \( S_m \) is compact and (ii) \( \Pi \) is continuous. So, using Berge’s theorem, we get that the value function \( V \) is continuous on \( S_m \) and that the best reply correspondence \( B \) is nonempty-valued, compact-valued, upper hemicontinuous and has a closed graph on \( S_m \).

The results needed are the nonempty-valuedness and upper hemicontinuity of \( B \). \( \square \)

Now we have all that is needed in prove the theorem. From the results of Lemmata 7 and 8, we know that \( B : S_m \to S_m \) is a nonempty-valued, convex-valued, upper hemicontinuous correspondence defined on the nonempty, compact and convex set \( S_m \). So, the conditions for the application of Kakutani’s fixed point theorem are satisfied. From Kakutani’s fixed point theorem, we get that there exists a \( x^* \in S_m \) such that \( x^* \in B(x^*) \) which means that there exists a \( x^* \in S_m \) such that

\[
\Pi(x^*, x^*) \geq \Pi(x, x^*) \text{ for all } x \in S_m.
\]

That is, \( \mathcal{G} \) has an equilibrium.

**D Proof of Theorem 6**

Let us define the following sets of group types:

\[
\mathcal{E}(x^*) = \{ i \in \Gamma_{n,m} : \text{supp}^{-1}(i) \subseteq I(x^*) \}
\]

\[
[M] = \{ i \in \Gamma_{n,m} : \text{supp}^{-1}(i) = \{ j \} \text{ for some } j \in M \}
\]

\( \mathcal{E}(x^*) \) consists of the group types that contain only individuals of types that are present in the population at \( x^* \). \( \mathcal{E}'(x^*) \) will denote its complement i.e. group types that contain at least one individual of one
of the types that are not present at $x^*$. $[M]$ consists of the groups types that contain only one type of individuals. We will denote the group type that contains only individuals of type $j$ by $[j]$. Now we can separate all group types in the following four categories:

- $SP(x^*) = \mathcal{E}(x^*) \cap [M]$ is the set of all group types that contain a single type of individuals that are present at $x^*$.
- $SA(x^*) = \mathcal{E}'(x^*) \cap [M]$ is the set of all group types that contain a single type of individuals that are absent at $x^*$.
- $MP(x^*) = \mathcal{E}(x^*) \setminus [M]$ is the set of all group types that contain more than one types of individuals that are present at $x^*$.
- $MA(x^*) = \mathcal{E}'(x^*) \setminus [M]$ is the set of all group types that contain more than one types of individuals and at least one of them is absent at $x^*$.

Let us define for any $x \in S_m$ the following:

$$\mu = \arg \min_{j \in I(x^*)} \frac{x_j}{x_j^*}$$

We construct $h$ as follows:

- For all $i \in MA(x^*)$ we define $h_i(x) = 0$.
- For all $i \in SA(x^*)$ we define $h_{[j]}(x) = x_j$.
- For all $i \in MP(x^*)$ we define $h_i(x) = \frac{x_{\mu}}{x_{\mu}^*} f_i^*(x^*)$.
- For all $i \in SP(x^*)$ we define $h_{[j]}(x) = \frac{x_{\mu}}{x_{\mu}^*} f_{[j]}^*(x^*) + x_j - \frac{x_{\mu}}{x_{\mu}^*} x_j$.

It is easy to check that $h$ is a matching rule as it satisfies definition 1. More than that it is also easy to see that $h(x^*) = f(x^*)$ and so $(x^*, h)$ is an evolutionary optimum. All we have to do is to show that $x^*$ is a NEGS under $h$.

Now let us define $A^* = \max_{[x_i] \in \pi} \hat{r}_i(x)$. As $(x^*, h)$ is an evolutionary optimum, it has to be that $x^*$ is a steady state of the replicator dynamics under $h$. So:

1. For all $j \in I(x^*)$ it has to be that $\pi_h(x^*) = A^*$ which is ensured by the fact that $h(x^*) = f(x^*)$ and
2. there is no restriction for all $j \in O(x^*)$.
For \( \mathbf{x}^* \) to be a NEGS it must hold that:

\[
\mathbf{x}^* \cdot \pi_h(\mathbf{x}^*) \geq \mathbf{y} \cdot \pi_h(\mathbf{x}^*) \quad \text{for all } \mathbf{y} \in S_m.
\]

Notice that from point 1. above, if \( \mathbf{x}^* \in \text{int} S_m \), it is a NEGS as \( \mathbf{y} \cdot \pi_h(\mathbf{x}^*) = A^* \) for all \( \mathbf{y} \in S_m \) and the proposition holds.

If \( \mathbf{x}^* \in \text{bd} S_m \), then all we need to do is show that \( \pi_{hj}(\mathbf{x}^*) \leq A^* \) for all \( j \in \mathcal{O}(\mathbf{x}^*) \). By definition,

\[
\pi_{hj}(\mathbf{x}^*) = \frac{\sum_{i \in \text{supp}(j)} n_i^j}{n} \partial^+ h_i(\mathbf{x}^*) A_i^j = \partial^+ h_{ij}(\mathbf{x}^*) A^j_{ij} + \sum_{i \in M \cap \text{supp}(j)} \frac{n_i^j}{n} \partial^+ h_i(\mathbf{x}^*) A_i^j = A^j_{ij}
\]

Finally, notice that under any matching rule the states \( e_j = (0, \ldots, 0, 1, 0, \ldots, 0) \) are steady states and the payoff of all individuals on these states is simply: \( \pi_h(e_j) = A^j_{ij} \). But as \( (\mathbf{x}^*, \mathbf{h}) \) is an evolutionary optimum, we know that \( A^* \leq A^* \) for all \( j \in M \). So, \( \pi_{hj}(\mathbf{x}^*) \leq A^* \) for all \( j \in M \).

\[
\square
\]

\section*{References}


