Repeated Games with Endogenous Separation

Segismundo S. Izquierdo, Luis R. Izquierdo and Matthijs van Veelen*

July 26, 2021

Abstract

We consider repeated games with endogenous separation – also known as voluntarily separable or voluntary partnership games – and their evolutionary dynamics. We formulate the replicator dynamics for games with endogenous separation, and provide a definition of neutral stability that guarantees Lyapunov stability in the replicator dynamics. We also provide existence results for monomorphic neutrally stable states in games with endogenous separation.

*Segismundo S. Izquierdo: BioEcoUva and Department of Industrial Organization, Universidad de Valladolid, 47011 Valladolid, Spain, segis@eii.uva.es. Luis R. Izquierdo: Department of Management Engineering, Universidad de Burgos, 09006 Burgos, Spain, lrizquierdo@ubu.es. Matthijs van Veelen: Department of Economics and Business, University of Amsterdam, 1012 WX Amsterdam, The Netherlands, and Tinbergen Institute, 1082 MS Amsterdam, The Netherlands, C.M.vanVeelen@uva.nl. If only it were still possible, we would have loved to thank William H. Sandholm for his very helpful comments. S. Izquierdo thanks the University of Wisconsin-Madison for hosting him during part of this research, and gratefully acknowledges sponsorship by the Fulbright Program and by the Spanish Ministry of Science, Innovation and Universities (PRX19/00113). We acknowledge financial support from the Spanish State Research Agency (PID2020-118906GB-I00/AEI/10.13039/501100011033).

1 Introduction

There is a large literature on repeated games where players are tied to their partners.[1] In this classic setup, reciprocity can sustain cooperation in games like the prisoners' dilemma. In an alternative setup, players also have the option to end the partnership, and go find a new match. Here players can discipline each other, not just by reciprocating, but also by leaving, for instance if their partner defects in the prisoners' dilemma. The option to leave, on the other hand, limits the scope for punishment by reciprocation.

Repeated games in which partners also have the option to leave have received different names in the literature. They are referred to as voluntary separable repeated games (Fujiwara-Greve and Okuno-Fujiwara, 2009; Fujiwara-Greve et al., 2015), voluntary partnership games (Vesely and Yang, 2010), conditional dissociation games (Izquierdo et al., 2010, 2014), games with the option to leave (Vesely and Yang, 2012), endogenously repeated games (Fujiwara-Greve et al., 2016),[2] partner switching games (Wubs et al., 2016), or games with endogenous match separation (Deb et al., 2020).[3] We will stay close to the last name, and call them games with endogenous separation.

Games with endogenous separation consist of a symmetric n-player stage game that is played repeatedly by the players in a population[4], who are grouped in n-player partnerships. In the body of the paper we will assume a two-player game[5], while appendix B provides a generalisation to n-player games. Players can condition their stage game action on the history of actions within their partnership, but not on past actions in previous partnerships, or, in other words, there is no information flow when changing partnership (Ghosh and Ray, 1996). After playing the stage game, players can unilaterally break their current partnership (hence the name *endogenous separation*), in which case both partners go single. Partnerships may also be broken by some exogenous factor, resulting in an exogenous partnership survival probability δ. At the beginning of every period, single players are randomly matched in new partnerships to play the stage game, along with the remaining pairs. Different combinations of strategies can have different expected duration. This creates a discrepancy between, on the one hand, the shares of strategies in the pool of singles – where strategies that tend to break up earlier return at a higher rate – and the population as a whole on the other. This feature complicates the analysis of such games.

In this paper we will focus on the stability of equilibria in games with endogenous separation,

[1] It is impossible to do justice to this whole literature with just a few citations. Some classic ones are Mailath and Samuelson (2006), Friedman (1971), Fudenberg and Maskin (1986), and the papers collected in Fudenberg and Levine (2008). For the state of the art, one can go to Deb et al. (2020) and references therein. In an evolutionary setting, classic papers are Hamilton and Axelrod (1981) and Bendor and Swistak (1995). Evolutionary dynamics in repeated games are considered in García and van Veelen (2016), van Veelen and García (2019) and van Veelen et al. (2012).

[2] This paper considers an extension of the model to two populations.

[3] Other pioneering studies of similar models, but without specific names for the game, are Schuessler (1989), Datta (1996), Kranton (1996), Carmichael and MacLeod (1997), and Watson (1999).

[4] The framework can be easily extended to non-symmetric n-player games played in n populations.

[5] This is in line with most of the literature, with a few exceptions such as Kurokawa (2019, 2021), who study some specific models with more than two players, and consider variations on how many players must choose to break a partnership for it to be broken.

and therefore also on out-of-equilibrium dynamics, in the evolutionary setting of population games generated by agents who are matched to play a normal form game (Sandholm, 2010b, pp. 4-5). The special case $\delta = 0$, where all partnerships are broken and randomly rematched after each stage game, can be considered to be the standard reference process that a population is assumed to undergo in the framework of population dynamics, for agents that are matched to play a normal form game with a finite strategy set (Friedman, 1998).

We provide several new results for games with endogenous separation. First, we formulate the replicator dynamics (Taylor and Jonker, 1978) for these games, with the standard replicator dynamics for normal form games being a special case ($\delta = 0$) in our framework. We illustrate the dynamics with some examples.

Second, we propose a definition of a neutrally stable state for games with endogenous separation (NSS^{ES}). For the two-player case, the pioneering papers by Carmichael and MacLeod (1997) and Fujiwara-Greve and Okuno-Fujiwara (2009) propose alternative definitions of neutral stability. However, as we show in appendix C, the conditions for neutral stability considered in those papers are not equivalent to any of the standard conditions for neutral stability in finite games (Bomze and Weibull, 1995). As a result, those definitions leave out states that one would naturally expect to qualify as neutrally stable. Besides, the definition by Carmichael and MacLeod (1997) is based on a non-explicit function, and the definition by Fujiwara-Greve and Okuno-Fujiwara (2009) can include states that, arguably, one would not naturally expect to qualify as neutrally stable, and which do not present Lyapunov stability in the replicator dynamics. In contrast, the condition for neutral stability that we propose here is explicit and is consistent with a standard definition of neutral stability for finite games.

Third, we study the relationship between our definition of neutral stability and the replicator dynamics. Our central result shows that being a neutrally stable state NSS^{ES} implies Lyapunov stability in the replicator dynamics for games with endogenous separation.

Last, we provide several results for monomorphic neutrally stable states in games with endogenous separation, or, equivalently, for *neutrally stable strategies:* in the framework of polymorphic populations made up by pure-strategists that we are considering, a monomorphic state is a state in which all players in the population use the same pure strategy; if a monomorphic state is neutrally stable, we say that the strategy being played at that state is a *neutrally stable strategy*. Focusing on the two-player case, our first result shows that if a strategy ever breaks up a partnership when playing against itself, it needs to satisfy very stringent conditions in order to be neutrally stable: it must always play an action corresponding to a symmetric pure Nash equilibrium of the stage game, and it must always attain the highest possible payoff among all symmetric action profiles of the stage game. The second result shows that a strategy that, when playing against itself, always plays a strict Nash equilibrium of the stage game, and never leaves, is neutrally stable. The third result is an existence theorem for neutrally stable strategies that shows that, for large enough δ,

any infinitely repeated sequence of symmetric profiles that provides players an average per-period payoff greater than the minmax payoff of the stage game can be supported by a neutrally stable strategy that starts any new partnership by playing a minmax action during a long enough initial phase. The "trust-building" strategies studied by Fujiwara-Greve and Okuno-Fujiwara (2009) for the prisoners' dilemma constitute a special case of the family of neutrally stable strategies that we present here. Their polymorphic equilibria, however, do not constitute neutrally stable states, for the same reasons discussed by Vesely and Yang (2012).

The rest of the paper is structured as follows. In section 2 we describe the endogenous separation model. Section 3 presents the payoff function for an incumbent population and for a (possibly polymorphic) group of potential invaders. In section 4 we formulate the replicator dynamics for games with endogenous separation and present some examples. Our definition of a neutrally stable state for games with endogenous separation (NSS^{ES}) is provided in section 5, with several properties of neutrally stable strategies being presented in section 6. Section 7 summarizes our results. The proof of a key result for the analysis of games with endogenous separation, relating the distribution of strategies in the pool of singles with the distribution of strategies in the whole population, is detailed in appendix A. The n-player case is discussed in appendix B.

2 The endogenous separation model

We consider a population of players that in each period find themselves in partnerships in order to play a symmetric two-player normal form game (fig. 1).

We refer to pure strategies in the stage game as *actions* and reserve the word *strategy* for decision rules in the associated game with endogenous separation (Mailath and Samuelson, 2006). As in the standard evolutionary setting for population games (Sandholm, 2010b, p. 4), we assume that players only use (pure) actions in the stage game, and pure strategies (as defined below) in the repeated game. The stage game is denoted by $G = \{A, u\}$, and is defined by a finite set of actions A and a payoff function $u\colon A^2 \to \mathbb{R}$.

Players in games with endogenous separation have the option to leave their partner after any stage game. The decision to leave their partner depends on the history of actions taken by each of the players within the current partnership. Formally, a strategy in a game with endogenous separation is defined as follows (Fujiwara-Greve and Okuno-Fujiwara, 2009). Let $t = 1, 2, ...$ indicate the periods within a partnership, i.e. the number of times the stage game has been played by the partnership. Let H_t be the set of all possible partnership histories, given that the partnership persists at time $t \geq 2$, which implies that $H_t = A^{2(t-1)}$. Let $H_1 = \{\emptyset\}$.

Definition 1. *A pure strategy i in a game with endogenous separation consists of $i = (a_t, b_t)_{t=1}^{\infty}$ where*

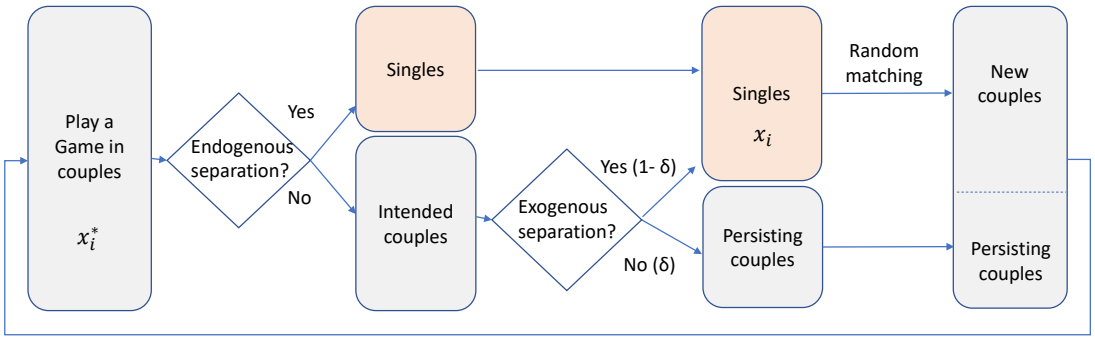


Figure 1: Sequence of events in games with endogenous separation. The proportion of players in the population using pure strategy i is x_i^*. The corresponding proportion of single players at the matching stage using pure strategy i is x_i.

$a_t : H_t \to A$ *specifies a stage-game action* $a_t(h_t) \in A$ *given the partnership history* $h_t \in H_t$, *and*

$b_t : H_t \times A^2 \to \{stay, leave\}$ *specifies whether to stay or leave the partner, depending on the partnership history* $h_t \in H_t$ *and the current-period action profile.*

The set of pure strategies in a game with endogenous separation is denoted by $\mathcal{S}$, and it is uncountably infinite (García and van Veelen, 2016).

At the beginning of each period, there will be single players, who find themselves in the "pool of singles", as well as players in surviving partnerships that were matched in a previous period. Single players are randomly matched in new partnerships, and all partnerships, new and surviving ones, play the stage game, choosing their actions according to their strategy. Subsequently, and depending on the history within the partnership, players choose to break up the partnership, or stay together, again according to their strategy. If any of the two players decides to leave, the partnership is broken, and both ex-partners are sent to the pool of singles, where all singles will be randomly matched again in new partnerships before the next stage game. Partnerships whose players choose to continue together manage to do so with probability $\delta \in [0, 1)$, and are otherwise broken by exogenous factors, sending both players to the pool of singles.

This endogenous separation framework[6] is represented in Figure 1.

A useful way to think of the dynamics in the short run, where there is no strategy updating, is given in fig. 2. This will also help link the frequencies in the pool of singles and the population as a whole.

[6]This is similar to the framework of Fujiwara-Greve and Okuno-Fujiwara (2009). The only difference is that in their model, individuals die with some probability, and when they die, they are replaced by a new single using the same strategy. An individual whose partner dies, also goes to the pool of singles. That amounts to a given exogenous probability with which a partnership survives to the next round (namely the square of the probability of survival of an individual).

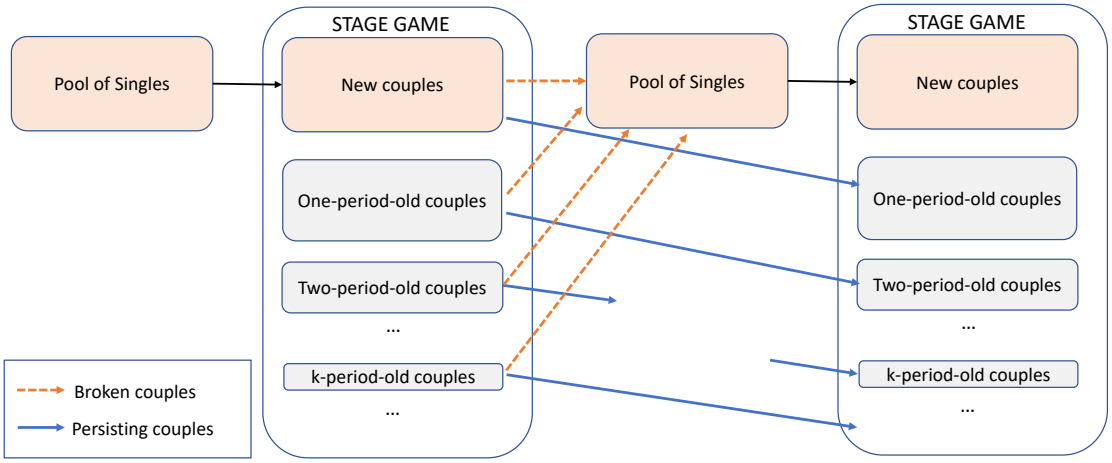


Figure 2: Sequence of events within two consecutive periods of the model.

The standard evolutionary model for a population of agents matched to play a symmetric normal form game $\{A, u\}$ can be considered to be a strategy-constrained game with endogenous separation in which $\delta = 0$ and where the strategy set $\mathcal{S}$ has been substituted by a finite set S_A which contains one strategy $i_a \in S_A$ for each action $a \in A$, with strategy i_a beginning a new partnership by playing action a. Note that the course of action followed by strategy i_a beyond its first action is irrelevant when $\delta = 0$.

3 Payoffs at steady states of the short-term dynamics

As in the standard evolutionary framework for large populations, we regard the proportion of individuals playing strategy i at the matching stage, x_i, as a continuous variable, and, likewise, we assume that the proportions of partnerships (i-strategists matched with j-strategists) resulting from the matching process equal the expected proportions. We will also assume that the total number of strategies being played in the population at any given time is finite. Therefore, the candidates for equilibria that we will consider have finite support. When considering stability, we allow this support to be any finite subset of $\mathcal{S}$, so no strategy is excluded as a possible ingredient of an equilibrium, or as a member of a group of potential invaders.

Let $\boldsymbol{x}^* = \{x_i^*\}_{i \in \mathcal{S}}$ be a strategy distribution in the population with finite support S, where x_i^* is the fraction of players in the population that use strategy i. This implies that $x_i^* \in [0, 1]$, $\sum_{i \in S} x_i^* = 1$ and $x_i^* = 0$ if $i \notin S$. The pool of singles is described by $\boldsymbol{x} = \{x_i\}_{i \in \mathcal{S}}$, where x_i is the fraction of players in the pool of singles that use strategy i, and for which the same restrictions apply. The composition of the pool of singles after one round of play depends on how the players in the population were matched, and for how long they were matched (see fig. 2). Therefore,

one population distribution $\boldsymbol{x}^*$ could produce a range of pool distributions $\boldsymbol{x}$. How the players will be matched in the next round, and for how long they will have been, also depends on how they were matched in the previous round, and for how long. For calculating the payoffs, we will assume that the population has reached a steady state of these short-term dynamics. In this steady state, the shares of the different matches are constant, and therefore also the composition of the pool of singles remains the same. Below we show how one can calculate the population strategy distribution $\boldsymbol{x}^*$ from the strategy distribution in the pool of singles $\boldsymbol{x}$, and in appendix A we show that there is one and only one steady-state pool distribution $\boldsymbol{x}$ that corresponds to each population strategy distribution $\boldsymbol{x}^*$. We will therefore refer to $\boldsymbol{x}^*$ as the population state corresponding to the pool state $\boldsymbol{x}$, and vice versa.

For every pair of pure strategies $i, j \in \mathcal{S}$, let their break-up period T_{ij} be the number of periods that an i-strategist and a j-strategist play together if their partnership is not broken by exogenous factors.[7] At a steady state $\boldsymbol{x}$, the share of individuals using strategy i in the whole population is proportional to $\bar{x}_i^*$ as defined below. Normalizing the mass of the pool of singles to 1, $\bar{x}_i^*$ includes a mass of $x_i x_j$ individuals playing i in newly formed $\{ij\}$ partnerships, plus $x_i x_j \delta$ individuals in one-period-old $\{ij\}$ partnerships if $T_{ij} \geq 2$, plus $x_i x_j \delta^2$ individuals in two-period-old $\{ij\}$ partnerships if $T_{ij} \geq 3$, and so on (see fig. 2); and it does so for all possible strategies j that a partner can have:

$$\bar{x}_i^* = \sum_{j \in S} x_i x_j \sum_{t=1}^{T_{ij}} \delta^{t-1} = \sum_{j \in S} x_i x_j \frac{1 - \delta^{T_{ij}}}{1 - \delta}.$$

We will call $\bar{x}_i^*$ the mass of i-players in the population. We furthermore denote the population-to-pool proportion of an $\{ij\}$ partnership by L_{ij}:

$$L_{ij} = \sum_{t=1}^{T_{ij}} \delta^{t-1} = \frac{1 - \delta^{T_{ij}}}{1 - \delta}.$$

$L_{ij} \geq 1$ is an expansion factor that indicates how many $\{ij\}$ partnerships we find in the population for each newly made (i.e., 0-period-old) $\{ij\}$ partnership formed after matching in the pool of singles. L_{ij} is also the expected length of an $\{ij\}$ partnership. In a steady state of the short-term dynamics, the relation between the fraction x_i^* of players in the whole population using strategy i and the fraction of players in the pool of singles using each strategy is then given by

$$x_i^* = \frac{\bar{x}_i^*}{\sum_k \bar{x}_k^*} = \frac{\sum_j x_i L_{ij} x_j}{\sum_k \sum_j x_k L_{kj} x_j} \equiv f_i(\boldsymbol{x}) \tag{1}$$

for $j, k \in S$.

For $t = 1, ..., T_{ij}$ and $i, j \in \mathcal{S}$, let a_t^{ij} be the action profile in period t in an $\{ij\}$ partnership, and let $u(a_t^{ij})$ be the associated payoff to the player using strategy i. Then, in any one period, the total

[7] If an i-strategist and a j-strategist never decide to break their partnership, then $T_{ij} = \infty$.

payoff to players using strategy i in partnerships with players using strategy j is

$$x_i x_j \sum_{t=1}^{T_{ij}} \delta^{t-1} u(a_t^{ij}).$$

Defining V_{ij}[8] as $V_{ij} = \sum_{t=1}^{T_{ij}} \delta^{t-1} u(a_t^{ij})$, the total payoff to the mass of players using strategy i in any one period is therefore

$$V_i(\boldsymbol{x}) = x_i \sum_{j \in S} V_{ij} x_j$$

Considering that the total mass of players using strategy i is $\bar{x}_i^* = x_i \sum_j L_{ij} x_j$, the average payoff to a player using pure strategy i, at a steady state with a pool strategy distribution $\boldsymbol{x}$, is

$$v_i(\boldsymbol{x}) = \frac{V_i(\boldsymbol{x})}{\bar{x}_i^*} = \frac{\sum_{j \in S} V_{ij} x_j}{\sum_{j \in S} L_{ij} x_j} \tag{2}$$

The last expression also allows us to extend the definition of $v_i(\boldsymbol{x})$ to strategies $i \in \mathcal{S}$ that are not in the support of $\boldsymbol{x}$. If strategy i is not in the support of $\boldsymbol{x}$, the payoff to strategy i can be interpreted as the payoff to a potential entrant using strategy i.[9]

We can then define the payoff to a (group of players with) strategy distribution $\boldsymbol{y}^*$ in a population with steady strategy distribution in the pool of singles $\boldsymbol{x}$ as:

$$v(\boldsymbol{y}^*, \boldsymbol{x}) = \sum_{i \in supp(\boldsymbol{y}^*)} y_i^* v_i(\boldsymbol{x}).$$

For monomorphic populations in which a single pure strategy $j \in \mathcal{S}$ is played, the payoff to strategy $i \in \mathcal{S}$ in a population of j-strategists is then

$$v_{ij} = \frac{V_{ij}}{L_{ij}} = (1-\delta) \frac{\sum_{t=1}^{T_{ij}} \delta^{t-1} u(a_t^{ij})}{1 - \delta^{T_{ij}}}$$

Equation (2) can be rewritten as $v_i(\boldsymbol{x}) = \sum_j \frac{L_{ij} x_j}{\sum_k L_{ik} x_k} v_{ij}$, for $j, k \in supp(\boldsymbol{x})$. This implies that if $\boldsymbol{x}$ is polymorphic, then $v_i(\boldsymbol{x})$ is a (strictly) convex combination of the payoffs v_{ij} for $j \in supp(\boldsymbol{x})$. This property will be useful when considering possible invasions of some strategy i by another strategy j. For instance, if $v_{ji} = v_{ii} = v_{ij} < v_{jj}$ and $supp(\boldsymbol{x}) = \{i, j\}$, then $v_i(\boldsymbol{x}) = v_{ii} < v_j(\boldsymbol{x})$.

The extension to the n-player case of the formulas presented in this section is indicated in appendix B.

[8] V_{ij} coincides with the expected per-partnership payoff to a player using strategy i in a match with a player that uses strategy j, which constitutes an alternative approach to calculate payoffs at steady states. That approach is followed by Fujiwara-Greve and Okuno-Fujiwara (2009).

[9] More precisely, this is the limit as $\epsilon \to 0$ of the payoff to a player using strategy i in a pool distribution such that a fraction of ϵ players use strategy i and a fraction of $(1-\epsilon)$ players have strategy distribution $\boldsymbol{x}$.

4 The replicator dynamics for games with endogenous separation

The general single-population replicator dynamics, as defined in Taylor and Jonker (1978), can also be formulated for games with endogenous separation. Suppose that individuals in a population are programmed to play some pure strategy i within a finite set S of s pure strategies. Let x_i^* (respectively, x_i) be the fraction of individuals in the population (respectively, in the pool of singles) that are programmed to play pure strategy $i \in S$. If we assume that the inflow of strategies (due to reproduction or strategy adoption) is proportional to their current share in the population as a whole x_i^*, and to their payoff $v_i(\boldsymbol{x})$, and that their outflow (from death or from strategy revision) is proportional to their share in the population,[10] we obtain the replicator dynamics for games with endogenous separation:

$$\dot{x}_i^* = \left[v_i(\boldsymbol{x}) - \sum_j x_j^* v_j(\boldsymbol{x}) \right] x_i^* \tag{3}$$

where

$$x_i^* = \frac{x_i \sum_j L_{ij} x_j}{\sum_k x_k \sum_j L_{kj} x_j} \tag{4}$$

A possible microfoundation for these dynamics is that individuals playing pure strategy i, whose prevalence in the population is x_i^*, reproduce at rate $v_i(\boldsymbol{x})$, while there is a uniform death rate in the population. An alternative microfoundation would be to assume that individuals, occasionally and independently, reconsider their strategy choice and, in order to choose a new strategy, use one of the revision protocols leading to the replicator dynamics.[11]

Numbering the finite set S of s strategies, we can –with a slight abuse of notation– represent a state $\boldsymbol{x}$ with support in S by its associated (column) vector $\boldsymbol{x}$, a point in the $(s-1)$-dimensional simplex $\Delta(S) \subset \mathbb{R}^s$. Using the symbol $\circ$ for the Hadamard product, and writing $\boldsymbol{L}$ for the matrix (L_{ij}) of expected lengths of each possible partnership of the strategies in S, the relationship eq. (4) between a population state and its corresponding pool state in vector notation is then given by the function $f_L : \Delta(S) \to \Delta(S)$, such that

$$\boldsymbol{x}^* = f_L(\boldsymbol{x}) = \frac{\boldsymbol{x} \circ (\boldsymbol{L}\boldsymbol{x})}{\|\boldsymbol{x} \circ (\boldsymbol{L}\boldsymbol{x})\|_1}. \tag{5}$$

If $\delta = 0$, then $\boldsymbol{x}^* = \boldsymbol{x}$, and the replicator dynamics (3) reduces to the standard replicator dynamics for pairwise interactions, which has a bilinear aggregate payoff function; a matrix $\boldsymbol{A}$

[10] If revision of strategies were to happen with a frequency for each strategy proportional to their presence in the pool of singles, we would obtain different dynamics: $\dot{x}_i^* = x_i^* v_i(\boldsymbol{x}) - x_i[\sum_j x_j^* v_j(\boldsymbol{x})]$. This could be a natural model if one assumes that players revise their strategy with some probability (only) every time their current partnership is broken. Note that in that case individuals whose partnerships are broken more frequently would revise their strategies more frequently.

[11] See Sandholm (2010b, chapter 10) for an overview of revision protocols, and Sandholm (2010a, example 1) or Izquierdo et al. (2019, examples A.1, A.2, and remark A.3) for the link with the replicator dynamics.

exists such that $v(\boldsymbol{y}, \boldsymbol{x}) = \boldsymbol{y}^T \boldsymbol{A}\boldsymbol{x}$. Here, A_{ij} is the payoff of the initial action of strategy i against the initial action of strategy j in the stage game.

$$\dot{x}_i = \left[v_i(\boldsymbol{x}) - \sum_j x_j v_j(\boldsymbol{x}) \right] x_i = \left[(\boldsymbol{A}\boldsymbol{x})_i - \boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} \right] x_i \tag{6}$$

Equation (3) defines a trajectory $\boldsymbol{x}^*(t; \boldsymbol{x}_0^*)$, starting from an initial population distribution $\boldsymbol{x}_0^*$ and its associated pool distribution $\boldsymbol{x}_0 = f_L^{-1}(\boldsymbol{x}_0^*)$. In order to show this, in appendix A we first prove the existence of a Lipschitz continuous inverse function $f_L^{-1} : \Delta(S) \to \Delta(S)$ that provides the pool distribution $\boldsymbol{x} = f_L^{-1}(\boldsymbol{x}^*)$ associated with a population distribution $\boldsymbol{x}^*$. We also show that, for any number of strategies $s > 3$, f_L^{-1} does not admit a general closed-form algebraic expression. After showing the existence of f_L^{-1}, we can consider the Lipschitz payoff function $F : \Delta(S) \to \mathbb{R}^s$ such that $F_i(\boldsymbol{x}^*) = v_i(f_L^{-1}(\boldsymbol{x}^*))$ and write eq. (3) as

$$\dot{x}_i^* = x_i^* \left[F_i(\boldsymbol{x}^*) - \sum_j x_j^* F_j(\boldsymbol{x}^*) \right] \tag{7}$$

which is the replicator dynamics as defined by Taylor and Jonker (1978) or Sandholm (2010b, p. 126), for a game with a Lipschitz continuous payoff function F, with the payoff function characterizing the population game. The replicator dynamics for games with endogenous separation (3) therefore is a special case of (7), corresponding to $F_i(x^*) = v_i(f_L^{-1}(\boldsymbol{x}^*))$, but in general it is different from the standard replicator dynamics for pairwise interactions (6), where the payoff functions are linear in $\boldsymbol{x}$: $F_i(\boldsymbol{x}) = (\boldsymbol{A}\boldsymbol{x})_i$.

We now present two examples.

Example 1. Let the stage game be a prisoners' dilemma, with actions C (for cooperate) and D (for defect), and stage game payoffs $u(DC) = 5, u(CC) = 4, u(DD) = 1$ and $u(CD) = 0$. Under the standard replicator dynamics (6), all interior solution trajectories converge to the state where all players defect and obtain a payoff of 1. Now consider one strategy that always plays C, and one that always plays D, while both strategies stay if their partner plays C, and leave if their partner plays D. We call these strategies CSL (Cooperate, Stay if your partner cooperates, Leave if your partner defects) and DSL (Schuessler, 1989). For $\delta > 0$, interactions can last longer than 1 period, and players can break up undesirable partnerships; and if δ is sufficiently high, the dynamics will change qualitatively.

Let x^*_{CSL} be the fraction of players using strategy CSL in the population. For $\delta = 0$, the payoff functions vary linearly with x^*_{CSL}, and $v_{\text{DSL}} = 1 + v_{\text{CSL}}$ for any population state. However, for $\delta > 0$, payoffs are not linear, and there can be states at which $v_{\text{DSL}} < v_{\text{CSL}}$, as shown in fig. 3. This figure also shows the rest points of the replicator dynamics, and the direction in which the system moves from any state, for different values of δ. For $\delta > 0.64$, there is a stable equilibrium with a

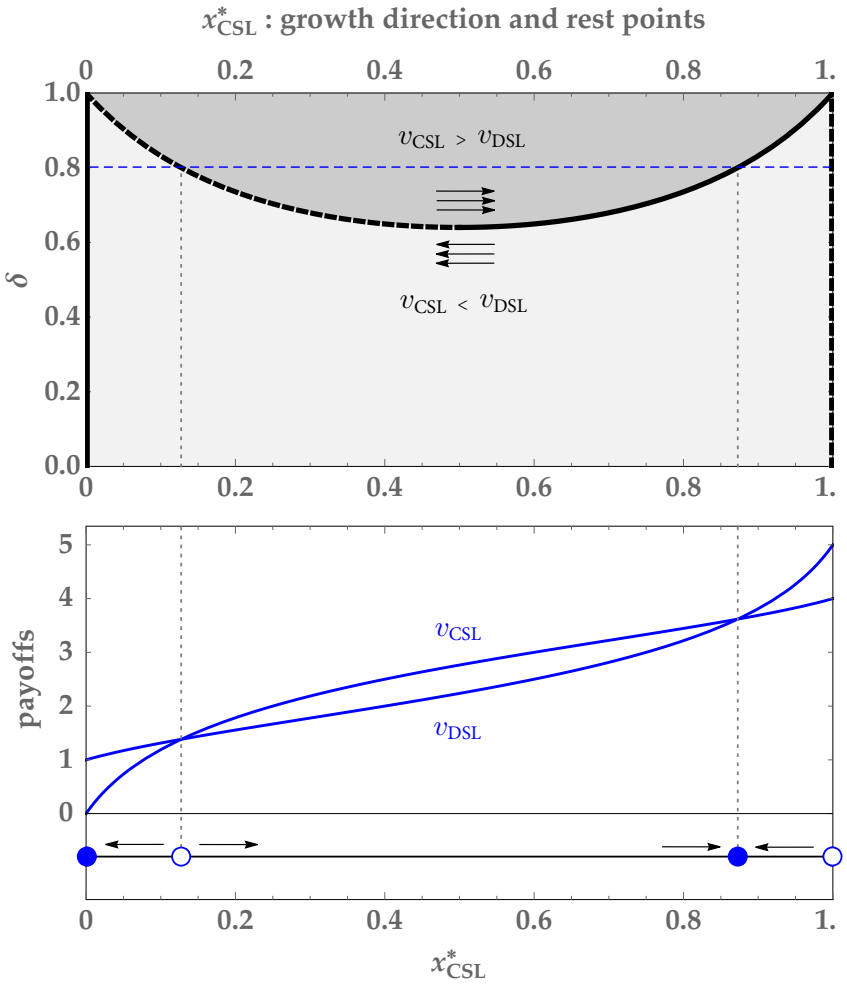
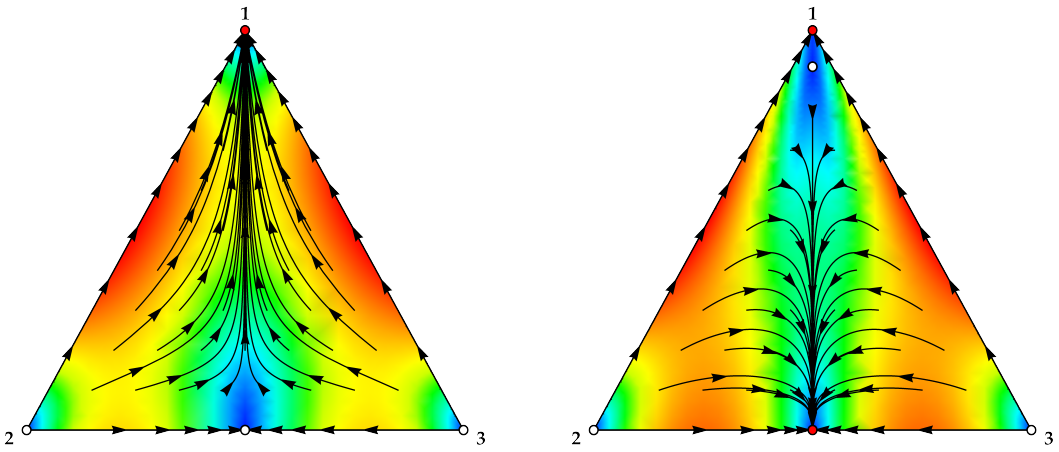


Figure 3: Dynamics of example 1. Top: Sets of values of x^*_{CSL} and δ at which $\dot{x}^*_{\text{CSL}}$ is positive (dark grey), negative (light grey) or zero (rest points, thick black lines; solid for stable and dashed for unstable). Below: payoffs to CSL (v_{CSL}) and to DSL (v_{DSL}) as a function of x^*_{CSL} for $\delta = 0.8$. Bottom: Direction of movement for x^*_{CSL}. Stable rest points are shown as solid dots; unstable rest points are shown as empty dots.



(a) $\delta = 0$ (b) $\delta = 0.9$

Figure 4: Phase portraits in the 2-dimensional simplex of the extended replicator dynamics for example 2. Subfigure (a) shows the dynamics for $\delta = 0$ (i.e. all partnerships are broken at the end of every period. This is the standard replicator dynamics for the stage game); subfigure (b) shows the dynamics for $\delta = 0.9$

majority of CSL players. The proportion of CSL players in the stable equilibrium approaches 1 as δ tends to 1. Moreover, the basin of attraction of this — mostly cooperative — equilibrium gets larger as δ increases, covering almost all of the state space when δ is close to 1. The lower panel in fig. 3 shows the dynamics for $\delta = 0.8$, which has two stable rest points (the monomorphic state where all play DSL, and a bimorphic state where $x^*_{\text{CSL}} = 0.87$) and one unstable rest point.

Example 2. Consider a stage game with three actions and payoff matrix

$$U = \begin{bmatrix} 3 & 3 & 3 \\ 0 & 0 & 5 \\ 0 & 5 & 0 \end{bmatrix}$$

Here we can consider the following three strategies: strategy $i \in \{1, 2, 3\}$ always plays action i, stays if it obtains its largest possible stage game payoff, given that it plays i (i.e. $\max_j u(a^{ij})$), and breaks the partnership otherwise. Figure 4a shows the dynamics for this game when $\delta = 0$, i.e. the setting where all partnerships are broken exogenously at the end of every period. This scenario corresponds to the replicator dynamics where the stage game is not repeated. In this setting, the monomorphic state $[1, 0, 0]$, where all players use strategy 1, is almost globally asymptotically stable (i.e. all solutions starting in the interior of the simplex converge to it). At this state, every player obtains a payoff of 3.

The dynamics can be very different if players have the opportunity to keep or break up their partnerships. Figure 4b shows the dynamics for $\delta = 0.9$, which means that the expected length of

partnerships that are not broken endogenously (strategy 1 with itself, and strategy 2 with strategy 3) is 10 periods. In this setting, the bimorphic state $[0, \frac{1}{2}, \frac{1}{2}]$, where half the population plays strategy 2 and the other half plays strategy 3, is asymptotically stable and has a very large basin of attraction (see fig. 4b). The payoff to the players at this state is 4.54, which is substantially higher than the payoffs in the case where interactions last for one period only. This is possible because the partnership with the highest payoff (strategy 2 with strategy 3) is now allowed to last longer, while endogenous separation allows for a quick breakup of partnerships with low payoffs.

The single-population replicator dynamics for two-player games with endogenous separation presented in this section is extended to n-player games in appendix B.

5 Nash equilibria and Neutrally Stable States with endogenous separation

5.1 *Nash equilibrium*

Let $\mathcal{F}(\mathcal{S})$ be the set of strategy distributions over $\mathcal{S}$ with finite support. In other words, if $\boldsymbol{x}^* \in \mathcal{F}(\mathcal{S})$, then there is a finite subset S of $\mathcal{S}$, for which $x_i^* \in [0,1]$ and $\sum_{i \in S} x_i^* = 1$, while $x_j^* = 0$ for any strategy $j \notin S$.

A definition of a Nash equilibrium in terms of population states $\boldsymbol{x}^*$ is impractical given that the payoff functions $F_i(x^*)$, as considered in eq. (7), do not admit a general closed-form algebraic expression when the support of x^* includes more than three strategies. When defining Nash equilibria and neutrally stable states, it can then be helpful to do it in terms of a population state and its associated pool state, $\{\boldsymbol{x}^*, \boldsymbol{x}\}$, where $\boldsymbol{x}^*$ is the population distribution corresponding to the steady-state pool distribution $\boldsymbol{x}$, i.e., $\boldsymbol{x}^* = f(\boldsymbol{x})$, according to eq. (1).

If the definition below is satisfied, we can refer to a Nash population state[12] $\boldsymbol{x}^*$, a Nash pool state $\boldsymbol{x}$, or a Nash population-pool state $\{\boldsymbol{x}^*, \boldsymbol{x}\}$.

Definition 2 (Nash Equilibrium)**.** *A population-pool state $\{\boldsymbol{x}^*, \boldsymbol{x}\}$ is a Nash equilibrium of a game with endogenous separation if for all $\boldsymbol{y}^* \in \mathcal{F}(\mathcal{S})$*

$$v(\boldsymbol{x}^*, \boldsymbol{x}) \geq v(\boldsymbol{y}^*, \boldsymbol{x})$$

On the left side of this inequality we have $v(\boldsymbol{x}^*, \boldsymbol{x}) = \sum_i x_i^* v_i(\boldsymbol{x})$, where $\boldsymbol{x}^* = f(\boldsymbol{x})$ and $i \in supp(\boldsymbol{x})$. This is the average payoff to incumbent players at pool state $\boldsymbol{x}$. On the right side we

[12] The game in Carmichael and MacLeod (1997) involves a phase where messages and gifts can be exchanged, so strategies there need to also specify a message and a gift. Other than that, their definition of a Nash population equilibrium is equivalent to definition 2. Also the definition of a Nash pool equilibrium in Fujiwara-Greve and Okuno-Fujiwara (2009) is equivalent.

see $v(\boldsymbol{y}^*, \boldsymbol{x}) = \sum_i y_i^* v_i(\boldsymbol{x})$, for $i \in supp(\boldsymbol{y}^*)$, which can be interpreted as the average payoff to an infinitesimally small group of players whose strategy distribution is $\boldsymbol{y}^*$, in a population with pool state $\boldsymbol{x}$.

It follows from Definition 2 that $\{\boldsymbol{x}^*, \boldsymbol{x}\}$ is a Nash equilibrium if and only if there is no pure strategy $i \in \mathcal{S}$ such that $v_i(\boldsymbol{x}) > v(\boldsymbol{x}^*, \boldsymbol{x})$. This will usually be the easiest way to assess whether a state $\{\boldsymbol{x}^*, \boldsymbol{x}\}$ is a Nash equilibrium.

Definition 2 implies that all pure strategies in the support of $\boldsymbol{x}^*$ (which is also the support of $\boldsymbol{x}$) must earn equal payoffs. Consequently, as in the standard framework, Nash equilibrium states are rest points of the replicator dynamics eq. (3), while not all rest points are Nash equilibria: all monomorphic states, for instance, are rest points of the replicator dynamics, but not necessarily Nash equilibria.

If a monomorphic distribution $\boldsymbol{i}$, corresponding to pure strategy $i \in \mathcal{S}$, is a Nash equilibrium state of a game with endogenous separation, we say that strategy i is a **Nash equilibrium strategy** of that game.

The following two results are relatively straightforward counterparts of observations that are routinely used for repeated games without the option to leave. The first provides a straightforward relation between symmetric pure Nash equilibria of a stage game and Nash equilibria of the corresponding game with endogenous separation.

Observation 1. *If the stage game of a game with endogenous separation has some symmetric Nash equilibrium in actions (a,a), any state composed of strategies that always play action a at every period (whatever their partner does) constitutes a Nash equilibrium state of the game with endogenous separation, regardless of the strategies' choices to leave or keep a partner.*

Proof. If the strategies in $\boldsymbol{x}$ always play action a, then no strategy can obtain a greater payoff than the stage payoff $u(a, a)$ at any period when matched with them, while $u(a, a)$ is the payoff that all the strategies in $\boldsymbol{x}$ obtain at every period when matched among themselves. □

As a result of observation 1, in a prisoners' dilemma with endogenous separation, any state composed of strategies that always defect in the stage game therefore constitutes a Nash equilibrium, regardless of the conditions under which the strategies break up or stay in a partnership; and any of those strategies, such as DSL in example 1, is a Nash equilibrium strategy.

Next, we provide a necessary condition for a strategy with a finite break-up period with itself to be a Nash equilibrium strategy.

Observation 2. *Consider a game with endogenous separation with $\delta > 0$. If strategy i has a finite break-up period T_{ii} and is a Nash equilibrium strategy, then it must play an action profile corresponding to a symmetric pure Nash equilibrium of the stage game at the last period T_{ii} of a partnership with itself.*

Proof. Suppose, to the contrary, that the action profile at T_{ii} when strategy i meets itself is not a Nash action profile. Now consider a strategy j that, in matches with i, plays like i itself up to period $T_{ii}-1$ but then, at period T_{ii}, plays a best response to the action adopted by i and leaves. Such strategy j, when playing with strategy i, would obtain payoff $v_{ji} > v_{ii}$. Therefore i is not a Nash equilibrium strategy. □

5.2 *Neutrally Stable States*

We begin this section with a note about nomenclature. In the evolutionary dynamics framework that we are considering of polymorphic populations and pure strategists (Sandholm, 2010b, p. 275), given that a population state can be identified by a point in the mixed strategy space, it is not uncommon to find the name *evolutionarily stable strategy* when referring to evolutionarily stable states (see Thomas (1984) or Bomze and Pötscher (1989, p. 15) for a discussion). To avoid the possible confusion, and following the approach adopted by many other authors (Sandholm, 2010b; Bomze and Pötscher, 1989; Hofbauer and Sigmund, 1998), here we will always use the term "*state*" when referring to a population (or pool) state, and will keep the term *strategy* for pure strategies of the game with endogenous separation, as defined before. We use *"Neutrally stable strategy"* to denote a pure strategy that, when adopted by all the individuals in a population, gives rise to a (monomorphic) neutrally stable state.

As in repeated games without the option to leave, there are no evolutionarily stable states in games with endogenous separation[13]. This is because for any equilibrium with finite support, there are always strategies that differ only off the equilibrium path, and those will obtain the same payoff as the equilibrium strategies. This rules out evolutionary stability (Selten, 1983; Selten and Hammerstein, 1984; García and van Veelen, 2016), and we consequently focus on the weaker stability condition of neutral stability. Given that the payoff functions F_i (see eq. (7)) in games with endogenous separation are non linear, we focus on definitions of neutral stability that guarantee stability in the replicator dynamics with non-linear payoff functions (Bomze and Weibull, 1995).

In the standard evolutionary framework, considering a finite set of s pure strategies S, a polyhedral population state space (here we assume that the state space is the whole simplex $\Delta(S) \in \mathbb{R}^s_{\geq 0}$ associated with the finite strategy set S, so it is a polyhedral space[14]), Lipschitz continuous payoff functions ϕ_i and an aggregate payoff function $u(\boldsymbol{x}, \boldsymbol{y}) = \sum_i x_i \phi_i(\boldsymbol{y})$, we say that a state $\boldsymbol{x} \in \Delta(S)$ is a neutrally stable state ($\boldsymbol{x} \in NSS$) if it satisfies any of the following two equivalent[15] conditions (Thomas, 1985; Bomze and Weibull, 1995):

[13] This is also the case for $\delta = 0$, unless we consider a strategy-constrained game.

[14] The state space is polyhedral if it is the intersection of finitely many closed affine half spaces (Bomze and Pötscher, 1989, p. 73).

[15] Thomas (1985) uses the terms *weakly evolutionarily stable* for a state satisfying condition 2, and shows that, in our considered framework (polyhedral state space) condition 2 implies condition 1. In a more general setting, Bomze and Weibull (1995) show that condition 1, which they call *strong unbeatability*, implies condition 2, and that consequently, if the state space is polyhedral, both conditions are equivalent.

Condition 1. *There is a neighborhood O of $\boldsymbol{x}$ in $\Delta(S)$ such that $u(\boldsymbol{x},\boldsymbol{y}) \geq u(\boldsymbol{y},\boldsymbol{y})$ for every $\boldsymbol{y} \in O$.*

Condition 2. *$\boldsymbol{x}$ is a Nash equilibrium and there is a neighborhood O of $\boldsymbol{x}$ in $\Delta(S)$ such that $u(\boldsymbol{x},\boldsymbol{y}) \geq u(\boldsymbol{y},\boldsymbol{y})$ for every $\boldsymbol{y} \in O$ satisfying $u(\boldsymbol{y},\boldsymbol{x}) = u(\boldsymbol{x},\boldsymbol{x})$.*

When studying the neutral stability of a Nash equilibrium (which is a prerequisite for neutral stability), the advantage of condition 2 over condition 1 is that the property $u(\boldsymbol{x},\boldsymbol{y}) \geq u(\boldsymbol{y},\boldsymbol{y})$ only needs to be checked for those states $\boldsymbol{y}$ that obtain the same payoff against $\boldsymbol{x}$ as $\boldsymbol{x}$ itself.

In the special case of linear payoff functions ϕ_i (and polyhedral state space), condition 1 and condition 2 are also equivalent to condition 3 below, used by Taylor and Jonker (1978) to define a neutrally stable state.

Condition 3. *For every $\boldsymbol{y} \in \Delta(S)$ there is some $\bar{\epsilon}_y \in (0,1)$ such that $u(\boldsymbol{x}, \epsilon\boldsymbol{y} + (1-\epsilon)\boldsymbol{x}) \geq u(\boldsymbol{y}, \epsilon\boldsymbol{y} + (1-\epsilon)\boldsymbol{x})$ for all $\epsilon \in (0, \bar{\epsilon}_y)$.*

The linear case corresponds to the standard evolutionary model for finite two-player normal form games with random matching. However, still considering a polyhedral space, if the linearity assumption is removed[16], condition 3 is weaker than condition 1 or condition 2, and it does not guarantee Lyapunov stability in the replicator dynamics (Bomze and Weibull, 1995).

For games with endogenous separation, we have an uncountably infinite strategy set $\mathcal{S}$ and the payoff functions are $F_i(\boldsymbol{x}^*) = v_i(f_L^{-1}(\boldsymbol{x}^*))$, which are non linear if $\delta > 0$. In this context, we say that a definition of a neutrally stable state for games with endogenous separation satisfies *finite-set neutral stability* if $\boldsymbol{x}^*$ being a neutrally stable state in the game with endogenous separation ($\boldsymbol{x}^* \in NSS^{ES}$) implies that for every finite set of strategies $S \in \mathcal{S}$ containing the support of $\boldsymbol{x}^*$, and taking $\Delta(S)$ as the (polyhedral) state space, $\boldsymbol{x}^* \in NSS$. Finite-set neutral stability can be formulated equivalently using either condition 1 or condition 2, and it implies that condition 3 is satisfied in every $\Delta(S)$. Taking condition 1, finite-set neutral stability means that for every finite set of strategies S such that $\boldsymbol{x}^* \in \Delta(S)$, there is a neighborhood O_S^* of $\boldsymbol{x}^*$ in $\Delta(S)$ such that $\sum_i x_i^* F_i(\boldsymbol{y}^*) \geq \sum_j y_j^* F_j(\boldsymbol{y}^*)$ for every $\boldsymbol{y}^* \in O_S^*$.

For two-player games with endogenous separation, Carmichael and MacLeod (1997) and Fujiwara-Greve and Okuno-Fujiwara (2009) provide pioneering definitions of neutral stability. However, as we show in appendix C, the first of those definitions is not practical to work with because it involves the non explicit function f_L^{-1}, while the second definition is explicit, but it does not satisfy finite-set neutral stability, and it does not guarantee Lyapunov stability in the replicator dynamics. Besides (see example C.1), both definitions leave out states that one would naturally expect to qualify as neutrally stable (and which are Lyapunov stable in the replicator dynamics).

[16]The payoff function is non linear in games with endogenous separation and $\delta > 0$, as well as in standard population games with more than two players.

Taking into account the infinite strategy space $\mathcal{S}$, the lack of an explicit formula for f_L^{-1}, and the advantage of condition 2 over condition 1 in polyhedral spaces, we provide the following two equivalent definitions of a neutrally stable state for a game with endogenous separation (NSS^{ES}).

Definition 3. *A population-pool state $\{\boldsymbol{x}^*, \boldsymbol{x}\}$ is a neutrally stable state of a game with endogenous separation (NSS^{ES}) if for any finite set of strategies $S \subset \mathcal{S}$ containing the support of $\boldsymbol{x}$ there is a neighborhood O_S of $\boldsymbol{x}$ in $\Delta(S)$ such that, for every $\boldsymbol{y} \in O_S$,*

$$v(\boldsymbol{x}^*, \boldsymbol{y}) \geq v(\boldsymbol{y}^*, \boldsymbol{y})$$

Definition 4. *A population-pool state $\{\boldsymbol{x}^*, \boldsymbol{x}\}$ is a neutrally stable state of a game with endogenous separation (NSS^{ES}) if it is a Nash equilibrium and for any finite set of strategies $S \subset \mathcal{S}$ with $\boldsymbol{x} \in \Delta(S)$ there is a neighborhood O_S of $\boldsymbol{x}$ in $\Delta(S)$ such that, for every $\boldsymbol{y} \in O_S$ satisfying $v(\boldsymbol{y}^*, \boldsymbol{x}) = v(\boldsymbol{x}^*, \boldsymbol{x})$,*

$$v(\boldsymbol{x}^*, \boldsymbol{y}) \geq v(\boldsymbol{y}^*, \boldsymbol{y})$$

These definitions are almost the same as condition 1 and condition 2, with only two differences. The first is that here we have infinitely many pure strategies, and we allow the neighbourhood to depend on the finite subset S of $\mathcal{S}$ we consider. This is enough for our central result. Of course one can also think of a stronger definition, where it is required that there is one single neighbourhood O of $\boldsymbol{x}$ in $\mathcal{F}(\mathcal{S})$ such that the conditions hold, rather than a neighbourhood for every subset S of pure strategies. This would obviously imply our definition; just choose $O_S = O \cap \Delta(S)$.[17]

The second difference is that here we consider a neighbourhood O_S of the pool state $\boldsymbol{x}$, while a direct translation to this setting would suggest that we consider a neighbourhood O_S^* of the population state $\boldsymbol{x}^*$. The reason for this choice is that f_L^{-1}, and, consequently, $v_i(f_L^{-1}(\boldsymbol{x}^*))$, do not admit a general closed-form algebraic expression. In appendix A, however, we show that if we consider distributions over a finite set S of s strategies, represented as points in a simplex $\Delta(S)$ in $\mathbb{R}^s$, then f_L is a bi-Lipschitz homeomorphism on $\Delta(S)$. This implies that f_L and f_L^{-1} preserve neighborhoods: if O is a neighborhood of $\boldsymbol{x}$ in $\Delta(S)$ and $\boldsymbol{x}^* = f_L(\boldsymbol{x})$, then $O^* = f_L(O)$ is a neighborhood of $\boldsymbol{x}^*$ in $\Delta(S)$, with the equivalent result for f_L^{-1}. Consequently, if there is a neighborhood O of $\boldsymbol{x}$ in which the associated population states satisfy some property, then there is also a neighborhood O^* of $\boldsymbol{x}^*$ in which the population states satisfy the property. This allows us to state the definition of a neutrally stable state in terms of $\boldsymbol{x}$ and $\boldsymbol{y}$, considering $\boldsymbol{x}^* = f_L(\boldsymbol{x})$ and $\boldsymbol{y}^* = f_L(\boldsymbol{y})$, which have an explicit formula (5). The equivalence of definition 3 and definition 4 follows from the preservation of neighborhoods by homeomorphisms, from the equivalence of condition 1 and condition 2 in polyhedral spaces with Lipschitz continuous payoff functions ϕ_i and aggregate

[17] To define a distance between strategy distributions $\boldsymbol{x} \in \mathcal{F}(\mathcal{S})$ and $\boldsymbol{y} \in \mathcal{F}(\mathcal{S})$, let $S = supp(\boldsymbol{x}) \cup supp(\boldsymbol{y})$, which we number from 1 to n. A possible distance now could be $d(\boldsymbol{x}, \boldsymbol{y}) = \sum_i^n |x_i - y_i|$. A neighborhood of $\boldsymbol{x} \in \mathcal{F}(\mathcal{S})$ is a subset of $\mathcal{F}(\mathcal{S})$ that includes all distributions $\boldsymbol{y} \in \mathcal{F}(\mathcal{S})$ with $d(\boldsymbol{x}, \boldsymbol{y}) < \epsilon$ for some $\epsilon > 0$. Because $\boldsymbol{x}$ and $\boldsymbol{y}$ correspond to points in R^n, any other distance based on an equivalent norm on the Euclidean space can be taken.

payoff function $u(\boldsymbol{x}, \boldsymbol{y}) = \sum_i x_i \phi_i(\boldsymbol{y})$ (Thomas, 1985; Bomze and Weibull, 1995), and from the fact that $\boldsymbol{x}$ is a Nash equilibrium (i.e., there is no strategy $i \in \mathcal{S}$ such that $v_i(\boldsymbol{x}) > v(\boldsymbol{x}^*, \boldsymbol{x})$) if and only if for any finite set of strategies S with $\boldsymbol{x} \in \Delta(S)$ there is no strategy $i \in S$ such that $v_i(\boldsymbol{x}) > v(\boldsymbol{x}^*, \boldsymbol{x})$.

The next result shows that a neutrally stable state $\boldsymbol{x}^* \in NSS^{ES}$ is Lyapunov stable in the replicator dynamics (3), for any finite set of strategies S that contains the support of $\boldsymbol{x}^*$.

Theorem 1. *A neutrally stable state $\hat{\boldsymbol{x}}^*$ of a game with endogenous separation is a Lyapunov stable state of the replicator dynamics (3), for any finite set of strategies S such that $\hat{\boldsymbol{x}}^* \in \Delta(S)$.*

Proof. Let $\hat{\boldsymbol{x}}^* \in \mathcal{S}$ be a neutrally stable population state. Consider the replicator dynamics, eq. (3), for a finite set S of s pure strategies such that $\hat{\boldsymbol{x}}^* \in \Delta(S)$. Numbering the pure strategies from 1 to s, strategy distributions $\hat{\boldsymbol{x}}^*$ and $\hat{\boldsymbol{x}}$ correspond to vectors $\hat{\boldsymbol{x}}^*$ and $\hat{\boldsymbol{x}} \in \Delta(S) \subset R^s_{\geq 0}$. In appendix A we show that the function $f_L : \Delta(S) \to \Delta(S)$ such that $\boldsymbol{x}^* = f_L(\boldsymbol{x})$ is a bi-Lipschitz homeomorphism in $\Delta(S)$. Now, consider as in eq. (7) the payoff function $F : \Delta(S) \to \mathbb{R}^s$ defined by $F_i(\boldsymbol{x}^*) = v_i(f_L^{-1}(\boldsymbol{x}^*))$, and let $E(\boldsymbol{y}^*, \boldsymbol{x}^*) = \sum_j y_j^* F_j(\boldsymbol{x}^*)$. Then the replicator dynamics (3) for games with endogenous separation can be written as

$$\dot{x}_i^* = x_i^* \left[F_i(\boldsymbol{x}^*) - E(\boldsymbol{x}^*, \boldsymbol{x}^*) \right] \tag{8}$$

which is the dynamics studied by Thomas (1985). That F is Lipschitz continuous follows easily from f_L^{-1} being Lipschitz continuous and from the definition of v_i. Considering that homeomorphisms preserve neighborhoods and that $E(\boldsymbol{x}^*, \boldsymbol{y}^*) = v(\boldsymbol{x}^*, \boldsymbol{y})$, $\hat{\boldsymbol{x}}^*$ satisfies the conditions for being weakly evolutionarily stable in Thomas (1985), with $\Delta(S)$ as the state space: $\hat{\boldsymbol{x}}^*$ is a Nash equilibrium and there is a neighborhood O^* of $\hat{\boldsymbol{x}}^*$ in $\Delta(S)$ such that

$$E(\hat{\boldsymbol{x}}^*, \boldsymbol{y}^*) \geq E(\boldsymbol{y}^*, \boldsymbol{y}^*)$$

for all $\boldsymbol{y}^* \in O^*$ with $E(\boldsymbol{y}^*, \hat{\boldsymbol{x}}^*) = E(\hat{\boldsymbol{x}}^*, \hat{\boldsymbol{x}}^*)$. Theorem 1 in Thomas (1985) then implies that $\hat{\boldsymbol{x}}^*$ is a Lyapunov stable state of the replicator dynamics (3) for the set of strategies S and state space $\Delta(S)$. As we made no assumptions on S other that it is finite and it contains the support of $\hat{\boldsymbol{x}}^*$, $\hat{\boldsymbol{x}}^*$ is Lyapunov stable in the replicator dynamics (3) for any finite set of strategies that contains the support of $\hat{\boldsymbol{x}}^*$. □

The definition of an NSS^{ES} and the Lyapunov stability result (theorem 1) in this section apply to n-player games as well, as indicated in appendix B. In the next section we present some results for monomorphic neutrally stable states in two-player games with endogenous separation, with the extension to n-player games presented in appendix B.

6 Neutrally stable strategies with endogenous separation

This section contains three results pertaining to monomorphic neutrally stable states in two-player games with endogenous separation. As indicated before, we refer to the one pure strategy that everyone plays in such a population state as a *neutrally stable strategy*.

Our first theorem provides necessary conditions for neutral stability for a strategy with a finite break-up period when playing against itself. These conditions turn out to be quite restrictive.

Theorem 2. *If a strategy i with finite break-up period T_{ii} is a neutrally stable strategy of a game with endogenous separation with survival rate $\delta > 0$, then: a) the maximum payoff attainable at a symmetric action profile of the stage game corresponds to a Nash equilibrium of the stage game, and b) the action profiles played in an $\{ii\}$ partnership belong to the set of efficient symmetric Nash equilibria in actions.*

Proof. Note first that the payoff to strategy i in a population of j-strategists can be written as

$$\begin{aligned} v_{ij} = \frac{V_{ij}}{L_{ij}} &= \frac{\sum_{t=1}^{T_{ij}} \delta^{t-1} u(a_t^{ij})}{\frac{1-\delta^{T_{ij}}}{1-\delta}} \\ &= \left[\sum_{t=1}^{T_{ij}} \delta^{t-1} u(a_t^{ij}) + \delta^{T_{ij}} \sum_{t=1}^{T_{ij}} \delta^{t-1} u(a_t^{ij}) + \delta^{2T_{ij}} \sum_{t=1}^{T_{ij}} \delta^{t-1} u(a_t^{ij}) + ... \right] (1-\delta) \end{aligned} \tag{9}$$

If we denote by $h_{ij} = \{a_1^{ij}, ..., a_{T_{ij}}^{ij}, a_1^{ij}, ..., a_{T_{ij}}^{ij}, ...\}$ the infinite sequence of outcomes or action profiles corresponding to an i-strategist when entering a population of j-strategists when no exogenous breakup events occur (so an $\{ij\}$ partnership keeps beginning again after each set of T_{ij} periods), formula 9 shows that the average per-period payoff to an i-strategist in a population of j-strategists coincides with the *normalized payoff* (Mailath and Samuelson, 2006) to i corresponding to the sequence of outcomes in h_{ij}. Note also that if $T_{ij} = 1$ then $v_{ij} = u(a_1^{ij})$, and otherwise v_{ij} is a convex combination of the set of payoffs $\{u(a_t^{ij})\}$ for $t \in \{1, ..., T_{ij}\}$. Let m be the maximum payoff obtainable at a symmetric action profile of the stage game, and let (b, b) be one of the symmetric action profiles (there may be more than one) that attain that maximum symmetric payoff. Suppose that $v_{ii} < m$, with T_{ii} being finite. Consider a strategy j that when playing with i behaves like i up to period T_{ii}, but at that period does not break the partnership and turns to playing action b forever, without breaking the partnership. That would make play between strategy i and strategy j unfold in the same way as it does between two players that play strategy i, and hence $v_{ji} = v_{ii} = v_{ij}$. Two players that play strategy j increase their average per round payoff after round i, and hence $v_{jj} > v_{ii}$. Pure strategy i therefore is not neutrally stable. This shows that, if i is a neutrally stable strategy and T_{ii} is finite, then $v_{ii} = m$, which implies that all the action profiles in h_{ii} must provide the maximum payoff m attainable at a symmetric action profile of the stage game.

Suppose that payoff m is obtained by some action profile (c, c) that is not a Nash equilibrium of the stage game, and strategy i plays action c at some period k of an $\{ii\}$ partnership. Then a strategy j that when playing with i chooses the same action as i up to period k (obtaining m at every period up to k), but at period k plays a best response action to c and breaks the partnership, obtains a payoff $v_{ji} > m = v_{ii}$, which cannot happen if i is neutrally stable. □

Example 3. Consider a 2-player pure coordination game as the stage game of a game with endogenous separation with $\delta > 0$. A necessary condition for a strategy i with finite break-up period T_{ii} to be neutrally stable is to always play the efficient action profile with itself.

Theorem 2 shows that one cannot construct a neutrally stable strategy of a game with endogenous separation with a finite break-up period by making it play symmetric strict Nash equilibria of the stage game, if there is a more efficient symmetric outcome and $\delta > 0$.

Corollary 2.1. *Consider a stage game such that the most efficient symmetric action profile is not a Nash equilibrium. Then, for any $\delta > 0$, no strategy i with a finite break-up period T_{ii} can be neutrally stable in the associated game with endogenous separation.*

Example 4. Consider a game with endogenous separation with the Prisoners' Dilemma or the Hawk-Dove game as stage game. As the most efficient symmetric action profile in those stage games is not a Nash equilibrium, no strategy with a finite break-up period with itself can be neutrally stable for any $\delta > 0$.

Our next lemma provides a sufficient condition for neutral stability for a strategy i that never leaves when playing against itself. The condition is that any strategy j that, when playing in a population of i-players, plays actions that are different from what i plays against itself, must earn a lower payoff than strategy i. Such a result may seem obvious, but what needs to be ruled out is that distributions $\boldsymbol{y}^*$ exist that do equally good against i ($v(\boldsymbol{y}^*, \boldsymbol{i}) = v_{ii}$), and better against themselves ($v(\boldsymbol{y}^*, \boldsymbol{y}) > v(\boldsymbol{i}, \boldsymbol{y})$). The proof does that by showing that any two pure strategies that obtain a payoff equal to v_{ii} when playing with i must play the same sequence of actions also when being matched between them, and therefore obtain that same (and not a higher) payoff.

Lemma 1. *If a strategy i that never leaves when playing against itself satisfies $v_{ii} > v_{ji}$ for any strategy j with $h_{ji} \neq h_{ii}$, then strategy i is a neutrally stable strategy.*

Proof. By hypothesis, any strategy j with $h_{ji} \neq h_{ii}$ obtains some payoff $v_{ji} < v_{ii}$, and, considering that h_{ii} is a series of symmetric outcomes, any strategy j with $h_{ji} = h_{ii}$ must satisfy $h_{ji} = h_{ii} = h_{ij} = h_{jj}$, and consequently $v_{ji} = v_{ii} = v_{ij} = v_{jj}$. Note that if an $\{ij\}$ partnership is broken without a deviation from the actions in h_{ii} happening, it is because strategy j chose to break the partnership, and it will do the same with any other strategy that generates the same path, including itself. Considering that $v(\boldsymbol{y}^*, \boldsymbol{i}) = \sum_{j \in supp(\boldsymbol{y}^*)} y_j^* v_{ji}$ is a strictly convex combination of the values

v_{ji} for $j \in supp(\boldsymbol{y}^*)$, if $\boldsymbol{y}^*$ is such that $v(\boldsymbol{y}^*, \boldsymbol{i}) = v_{ii}$ then any pair of strategies $j, k \in supp(\boldsymbol{y}^*)$ are such that $v_{ji} = v_{ki} = v_{ii}$, which implies $h_{ji} = h_{ki} = h_{ii}$ and then also the equality of all the prospective sequences $h_{jk} = h_{ii}$ for $j, k \in supp(\boldsymbol{y}^*)$. Given that $\boldsymbol{y}$ and $\boldsymbol{y}^*$ have the same support, for every population-pool state $\{\boldsymbol{y}^*, \boldsymbol{y}\}$ satisfying $v(\boldsymbol{y}^*, \boldsymbol{i}) = v_{ii}$, we have $v_i(\boldsymbol{y}) = v(\boldsymbol{y}^*, \boldsymbol{y}) = v_{ii}$ and consequently, $\boldsymbol{i}$ is a neutrally stable state. □

Corollary 2.2. *If the stage game of a game with endogenous separation has some strict Nash symmetric equilibrium in actions (a, a), then any strategy that always plays action a and never leaves a partner playing action a is neutrally stable.*

Example 5. If the stage game is a coordination game in which all symmetric action profiles are strict Nash equilibria, such as, e.g., a generic Stag Hunt, any strategy that always plays the same action and never leaves a partner using that same action is neutrally stable. If the stage game is a Prisoners' Dilemma, any strategy that always plays defect, and never leaves a partner that plays that action is neutrally stable.

Finally, we provide an existence theorem for neutrally stable strategies in symmetric games with endogenous separation. Let $w = \min_j \max_i u(a^{ij})$ be the minimax payoff of the stage game and let b be one of the minimax actions, i.e., an action such that $\max_i u(a^{ib}) = w$.

Theorem 3. *For large enough $\delta < 1$, any finite sequence or pattern of symmetric outcomes $\Phi_p = \{(a_1, a_1), (a_2, a_2), ..., (a_{k_p}, a_{k_p})\}$ with an average payoff $\bar{u} = k_p^{-1} \sum_{k=1}^{k_p} u(a_k, a_k)$ greater than the minimax payoff of the stage game can be sustained as an indefinitely repeated pattern by a neutrally stable strategy i such that, in the equilibrium path h_{ii}, the repeated play of the pattern Φ_p is preceded by a sufficiently long "deviation-deterring" phase in which a minimax action profile (b, b) is played.*

An example of the family of neutrally stable strategies included in theorem 3 are the "trust-building" strategies for the prisoners' dilemma game with endogenous separation studied by Fujiwara-Greve and Okuno-Fujiwara (2009), which, after a deviation-deterring phase playing (D, D), support the indefinitely repeated pattern $\{(C, C)\}$. In fact, theorem 3 shows that, for large enough δ, any indefinitely repeated pattern that includes some $\{(C, C)\}$ and some $\{(D, D)\}$ outcomes in the repeated pattern can be supported by a neutrally stable strategy, after a long enough deviation-deterring phase playing (D, D). In the setup of Carmichael and MacLeod (1997), the additional gift-giving stage allows players to start their match with a costly and inefficient exchange of gifts, which also has the effect of making it an unattractive prospect to be broken up with, and having to go through the costly initial phase with a new partner.

Proof. Consider a strategy i such that, when playing against itself, does not leave, and for which an $\{ii\}$ partnership starts with playing the minimax profile (b, b) for T periods, followed by infinitely many repetitions of the pattern Φ_p. If the other strategy deviates from this sequence, strategy i

breaks the partnership. Let $u_b = u(b,b)$ be the payoff corresponding to the action profile (b,b). As b is a minimax action, $u_b \leq w$. Then

$$v_{ii} = (1-\delta^T)u_b + \delta^T \frac{(1-\delta)}{1-\delta^{k_p}} \sum_{k=1}^{k_p} \delta^{k-1} u(a_k, a_k)$$

The key of the proof is that, given any infinitely repeated sequence of K payoffs $\{u_1, u_2, ..., u_K\}$, the infinite sum of the discounted payoffs multiplied by $(1-\delta)$ converges to the mean of the K payoffs as δ approaches 1, i.e.:[18]

$$\lim_{\delta \to 1} \frac{1-\delta}{1-\delta^K} \sum_{k=1}^{K} \delta^{k-1} u_k = \frac{\sum_{k=1}^{K} u_k}{K}$$

Using this result it is easy to see that, for any fixed T, $v_{ii} \to \bar{u}$ as $\delta \to 1$. Then, taking $\epsilon = \bar{u} - w$, for any fixed T there is a $\delta_1(T) < 1$ such that for $\delta > \delta_1(T)$ we have $v_{ii} > w + \frac{\epsilon}{2}$. Assume $\delta > \delta_1(T)$. A strategy j that deviates during the deviation-deterring phase would obtain $v_{ji} \leq w < v_{ii}$. During the pattern-playing phase, if a deviation is profitable at any point, then it must be profitable during the first occurrence of the pattern (see lemma 2 below for a formal proof). The payoff v_{ji} to a strategy j that deviates (either in action or by leaving) during the pattern-playing phase at the K-th action profile (a_K, a_K) of the pattern, with $K \leq k_p$, obtaining a payoff B at that stage, is:

$$v_{ji} = \frac{1-\delta}{1-\delta^{T+K}} \left(\frac{1-\delta^T}{1-\delta} u_b + \delta^T \left(\sum_{k=1}^{K-1} \delta^{k-1} u(a_k, a_k) + \delta^{K-1} B \right) \right).$$

And using L'Hopital rule we find:

$$\lim_{\delta \to 1} v_{ji} = \frac{u_b T + \sum_{k=1}^{K-1} u(a_k, a_k) + B}{T+K} = l_{ij}^{B,K,T}$$

Note that the term $l_{ij}^{B,K,T}$ converges to $u_b \leq w$ as $T \to \infty$, so we can find a value T_0 large enough to guarantee that $l_{ij}^{B,K,T_0} < w + \frac{\epsilon}{4}$ for any strategy j with $h_{ji} \neq h_{ii}$. Now, if we take a value of δ greater than $\delta_1(T_0)$ (so $v_{ii} > w + \frac{\epsilon}{2}$) and large enough to guarantee $v_{ji} < w + \frac{\epsilon}{2}$, we have $v_{ji} < v_{ii}$ for any strategy j with $h_{ji} \neq h_{ii}$, so, by lemma 1, strategy i is neutrally stable. □

The next lemma provides a formal proof of the claim that, during the pattern playing phase, if a deviation is profitable at any point, then it must be profitable during the first occurrence of the pattern.

Lemma 2. *Consider a strategy i such that $h_{ii} = \{\Phi_0, \Phi_p, \Phi_p, ...\}$ where $\Phi_0 = \{\Phi_{0,1}, ..., \Phi_{0,k_0}\}$ is a finite sequence of $k_0 \geq 1$ symmetric outcomes and where $\Phi_p = \{\Phi_{p,1}, ..., \Phi_{p,k_p}\}$ is a finite sequence*

[18] This can be shown using L'Hopital rule.

of $k_p \geq 1$ symmetric outcomes. Let Φ_1 be an arbitrary sequence of $k_1 \geq 1$ outcomes. A strategy j' with $h_{j'i} = \{(\Phi_0, \Phi_p, \Phi_1), (\Phi_0, \Phi_p, \Phi_1), ...\}$ satisfies $v_{j'i} > v_{ii}$, if and only if a strategy j with $h_{ji} = \{(\Phi_0, \Phi_1), (\Phi_0, \Phi_1), ...\}$ satisfies $v_{ji} > v_{ii}$.

Proof. Let the discounted values of the payoffs in each sequence be $\pi_0 = \sum_{k=1}^{k_0} \delta^{k-1} u(\Phi_{0,k})$, $\pi_p = \sum_{k=1}^{k_p} \delta^{k-1} u(\Phi_{p,k})$ and $\pi_1 = \sum_{k=1}^{k_1} \delta^{k-1} u(\Phi_{1,k})$. Then

$$v_{ii} = (1-\delta)(\pi_0 + \frac{\delta^{k_0}\pi_p}{1-\delta^{k_p}})$$

$$v_{j'i} = (1-\delta)\frac{\pi_0 + \delta^{k_0}\pi_p + \delta^{k_0+k_p}\pi_1}{1-\delta^{k_0+k_p+k_1}}$$

and

$$v_{ji} = (1-\delta)\frac{\pi_0 + \delta^{k_0}\pi_1}{1-\delta^{k_0+k_1}}$$

Any of the two conditions $v_{j'i} > v_{ii}$ or $v_{ji} > v_{ii}$, which require $\delta > 0$, can then be seen to be equivalent (rearranging and simplifying terms) to the condition

$$\frac{\pi_1 + \delta^{k_1}\pi_0}{1-\delta^{k_0+k_1}} > \frac{\pi_p}{1-\delta^{k_p}}$$

□

7 Conclusions

While some real-life situations tie partners together, others allow at least some freedom to cut ties and start over with a new partner. The setting in which partners cannot alter the duration of a partnership has become the standard, and is studied extensively. The setting where there is also the option to leave is considered only in a much smaller literature. In this paper we look at evolutionary dynamics for the latter setting. We formulate the replicator dynamics for games with endogenous separation, and, building on Carmichael and MacLeod (1997) and Fujiwara-Greve and Okuno-Fujiwara (2009), who study two-player games with endogenous separation, and on Thomas (1985) and Bomze and Weibull (1995), who study different neutral stability definitions and relate them to Lyapunov stability in the replicator dynamics, we give a definition of a neutrally stable state for n-player games with endogenous separation, and show that it guarantees Lyapunov stability in the replicator dynamics. A key ingredient for our results is the proof that the function relating the distribution of strategies in the population as a whole and the distribution in the pool of singles in a game with endogenous separation is a bi-Lipschitz homeomorphism (a diffeomorphism in the interior of the simplex).

We also provide several results for monomorphic neutrally stable states, where all individuals play a single pure strategy (a neutrally stable strategy). Our main result here is an existence

theorem that (in the two-player case) states that, for large enough δ, any infinitely repeated sequence of symmetric action profiles that provides players with an average payoff greater than the minmax payoff can be supported by a neutrally strategy, after a long enough initial phase in which a minmax action profile is played. This generalizes the idea of equilibria with a trust-building phase from Fujiwara-Greve and Okuno-Fujiwara (2009), where the stage game is a prisoners' dilemma. This trust-building phase can also be replaced by (wasteful) gift-giving, as in Carmichael and MacLeod (1997). We also provide an extension of this result to n-player games.

While most of the previous references on games with endogenous separation have focused on the prisoners' dilemma, here we present general results (the replicator dynamics and a definition of a neutrally stable state that guarantees Lyapunov stability in the replicator dynamics) for symmetric n-player games, as well as necessary conditions and constructive existence results for neutrally stable strategies in any symmetric game with endogenous separation.

A Existence and uniqueness of the solution trajectories of the replicator dynamics for games with endogenous separation

Consider a finite set S of s pure strategies $i \in \{1, ..., s\}$ of a game with endogenous separation, its associated simplex $\Delta(S) = \{\boldsymbol{x} \in \mathbb{R}^s_{\geq 0} : \sum_{i=1}^{s} x_i = 1\}$, and the replicator dynamics

$$\dot{x}_i^* = \left[v_i(\boldsymbol{x}) - \sum_j x_j^* v_j(\boldsymbol{x}) \right] x_i^* \tag{10}$$

where

$$\boldsymbol{x}^* = f_L(\boldsymbol{x}) = \frac{\boldsymbol{x} \circ (\boldsymbol{L}\boldsymbol{x})}{\|\boldsymbol{x} \circ (\boldsymbol{L}\boldsymbol{x})\|_1}$$

To show existence and uniqueness of the solution trajectories of the replicator dynamics (10) in $\Delta(S)$, we prove that the function f_L is a bi-Lipschitz homeomorphism in $\Delta(S)$, so there is an inverse function $f_L^{-1} : \Delta(S) \to \Delta(S)$, such that if $\boldsymbol{x}^* = f_L(\boldsymbol{x})$ then $\boldsymbol{x} = f_L^{-1}(\boldsymbol{x}^*)$, and f_L^{-1} is Lipschitz. We can then write the extended replicator dynamics using only the population prevalences $\boldsymbol{x}^*$:

$$\dot{x}_i^* = \left[F_i(\boldsymbol{x}^*) - \sum_j x_j^* F_j(\boldsymbol{x}^*) \right] x_i^* \tag{11}$$

where $F_i(\boldsymbol{x}^*) = v_i(f_L^{-1}(\boldsymbol{x}^*))$ is Lipschitz. This is then a standard replicator dynamics (with non-linear payoff function) (Sandholm, 2010b), and we can apply an extension of the the Picard-Lindelöf theorem for compact convex sets (see Sandholm (2010b, p. 135, Theorem 4.A.5) or Weibull (1995, p. 238, Proposition 6.1)) to show existence and uniqueness of a global solution through any

initial state $\boldsymbol{x}_0^* \in \Delta(S)$, as well as invariance of the simplex: the simplex $\Delta(S)$, its faces and its interior are invariant.

Note that the functions v_i are Lipschitz in $\Delta(S)$ because the denominator in (2) is always greater than or equal to 1. To show that F is a Lipschitz function we then need to show that there is a Lipschitz function f_L^{-1}. This will require several steps.

First, it is easy to check that f_L is continuous and preserves the faces of $\Delta(S)$, which implies that f_L is surjective (Idzik et al., 2014), i.e., for every population state $\boldsymbol{x}^* \in \Delta(S)$ there is at least one pool state $\boldsymbol{x} \in \Delta(S)$ such that $\boldsymbol{x}^* = f_L(\boldsymbol{x})$. Let us now consider the Jacobian of f_L. Working directly with f_L presents the problem that it is not defined outside $\Delta(S)$, so the partial derivatives are not defined. We can work with the extension of f_L to $\mathbb{R}^s$, but the Jacobian of this function at points in $\Delta(S)$ is null, because the division step in f_L projects points in $\mathbb{R}^s_{>0}$ onto $\Delta(S)$.

Consider the vector function $g : \mathbb{R}^s \to \mathbb{R}^s$ such that $g(\boldsymbol{x}) = \boldsymbol{x} \circ (\boldsymbol{L}\boldsymbol{x})$. By definition, for $\boldsymbol{x} \in \Delta(S)$, we have $f_L(\boldsymbol{x}) = \frac{g(\boldsymbol{x})}{\|g(\boldsymbol{x})\|_1}$. It is easy to see that g is continuously differentiable, being made up by polynomials (it is in fact smooth), and that for $\boldsymbol{x} \in \mathbb{R}^s_{\geq 0}$ it satisfies the following properties:

- (P1): $g_i(\boldsymbol{x}) = 0$ if and only if $x_i = 0$.
- (P2): For any scalar $k \geq 0$, $g(k\boldsymbol{x}) = k^2 g(\boldsymbol{x})$. This implies that the image of a line segment going from $\mathbf{0}$ to $\boldsymbol{x} \neq \mathbf{0}$ is the line segment that goes from $\mathbf{0}$ to $g(\boldsymbol{x})$, and the image of a ray from $\mathbf{0}$ through $\boldsymbol{x} \neq \mathbf{0}$ is a ray from $\mathbf{0}$ through $g(\boldsymbol{x})$.

Lemma A.1. *The Jacobian of g is non-singular in $\mathbb{R}^s_{\geq 0} \setminus \{\mathbf{0}\}$.*

Proof.

The elements of the Jacobian of g are $J_{ij}(\boldsymbol{x}) = x_i L_{ij} + \delta_{ij} \sum_k L_{ik} x_k$, where δ_{ij} is the Kronecker delta. Then, for $\boldsymbol{x} \in \mathbb{R}^s_{\geq 0} \setminus \{\mathbf{0}\}$:

- If $\boldsymbol{x} \in \mathbb{R}^s_{>0}$, the Jacobian $J(\boldsymbol{x})$ is a (column) strictly diagonally dominant matrix, and, consequently, $J(\boldsymbol{x})$ is non-singular (Serre, 2002, p. 73).
- If some $x_i = 0$ (and $\boldsymbol{x} \neq \mathbf{0}$), then $J_{ij}(\boldsymbol{x}) = \delta_{ij} \sum_k L_{ik} x_k$, so all the elements of the i^{th} row of the Jacobian are 0, but for the one at the diagonal, which is positive. The submatrix of the Jacobian corresponding to components of $\boldsymbol{x}$ that are positive is strictly diagonally dominant, so $J(\boldsymbol{x})$ is non-singular.

□

Lemma A.1 implies that g is locally invertible in $\mathbb{R}^s_{\geq 0} \setminus \{\mathbf{0}\}$, with a continuously differentiable local inverse at any point. Note also that Lemma A.1 implies that the Jacobian of g is bounded away from 0 in any compact set in $\mathbb{R}^s_{\geq 0} \setminus \{\mathbf{0}\}$.

Consider now $g|_{\mathbb{R}^s_{>0}}$, the restriction of g to $\mathbb{R}^s_{>0}$. Our next lemma shows global invertibility of $g|_{\mathbb{R}^s_{>0}}$.

Lemma A.2. *$g|_{\mathbb{R}^s_{>0}} : \mathbb{R}^s_{>0} \to \mathbb{R}^s_{>0}$ is a diffeomorphism from $\mathbb{R}^s_{>0}$ to $\mathbb{R}^s_{>0}$, i.e., it is a differentiable function with a differentiable inverse.*

Proof. We use theorem B in Gordon (1972): Let M_1 and M_2 be connected, oriented N-dimensional manifolds of class C^1, without boundary, and suppose that M_2 is simply connected. Then a C^1 map f from M_1 to M_2 is a diffeomorphism if and only if f is proper and the Jacobian of f never vanishes.

It is easy to check that $\mathbb{R}^s_{>0}$ with the standard orientation satisfies the conditions for the manifolds. Given that we have already proven that the Jacobian of $g|_{\mathbb{R}^s_{>0}}$ never vanishes (lemma A.1), it only rests to show that $g|_{\mathbb{R}^s_{>0}}$ is proper.

A map $f : X \to Y$ between topological spaces X and Y is proper if the inverse image of any compact set is compact (Lee, 2003, p. 45). In our context, being compact is equivalent to being closed and bounded. The proof that $g|_{\mathbb{R}^s_{>0}}$ is proper rests on the following two properties of g: i) For $\boldsymbol{x} \in \mathbb{R}^s_{>0}$, $||\boldsymbol{x}|| \to \infty \implies ||g(\boldsymbol{x})|| \to \infty$, and ii) $g(\partial\mathbb{R}^s_{>0}) \cap \mathbb{R}^s_{>0} = \emptyset$, where $\partial\mathbb{R}^s_{>0}$ is the boundary of $\mathbb{R}^s_{>0}$ in $\mathbb{R}^s$ (where some component $x_i = 0$; note that $g(\partial\mathbb{R}^s_{>0}) = \partial\mathbb{R}^s_{>0}$). Because of i), the preimage K^- of any compact set $K \subset \mathbb{R}^s_{>0}$ under $g|_{\mathbb{R}^s_{>0}}$ is bounded. Continuity of g guarantees that K^- is relatively closed in $\mathbb{R}^s_{>0}$, i.e., it is the intersection of a closed set with $\mathbb{R}^s_{>0}$ (Taylor, 2012, Th. 8.2.1). To show that K^- is closed we need to show that if $\boldsymbol{x}_k \to \boldsymbol{x}$ with $\boldsymbol{x}_k \in K^-$, then $\boldsymbol{x} \in \mathbb{R}^s_{>0}$. Assume $\boldsymbol{x}_k \to \boldsymbol{x}$ with $\boldsymbol{x}_k \in K^- \subset \mathbb{R}^s_{>0}$ (so $\boldsymbol{x} \in \mathbb{R}^s_{\geq 0}$). Then, by continuity, $g(\boldsymbol{x}_k) \to g(\boldsymbol{x})$, with $g(\boldsymbol{x}_k) \in K$. As K is compact, $g(\boldsymbol{x}) \in K \subset \mathbb{R}^s_{>0}$. Considering ii), this implies that $\boldsymbol{x} \in \mathbb{R}^s_{>0}$ ($\boldsymbol{x} \notin \partial\mathbb{R}^s_{>0}$, given that $g(\boldsymbol{x}) \in \mathbb{R}^s_{>0}$ and $g(\partial\mathbb{R}^s_{>0}) \cap \mathbb{R}^s_{>0} = \emptyset$) and, consequently, $\boldsymbol{x} \in K^-$.

□

Lemma A.3. *$g|_{\mathbb{R}^s_{\geq 0}} : \mathbb{R}^s_{\geq 0} \to \mathbb{R}^s_{\geq 0}$ is a homeomorphism from $\mathbb{R}^s_{\geq 0}$ to $\mathbb{R}^s_{\geq 0}$, i.e., it is a continuous function with a continuous inverse $g_+^{-1} : \mathbb{R}^s_{\geq 0} \to \mathbb{R}^s_{\geq 0}$.*

Proof. The surjectivity of f_L in $\Delta(S)$, combined with property (P2), imply that the image (codomain) of $\mathbb{R}^s_{\geq 0}$ under $g|_{\mathbb{R}^s_{\geq 0}}$ is $\mathbb{R}^s_{\geq 0}$, and the image of $\mathbb{R}^s_{>0}$ under $g|_{\mathbb{R}^s_{>0}}$ is $\mathbb{R}^s_{>0}$. Lemma A.2 implies that $g|_{\mathbb{R}^s_{>0}}$ is injective[19]. Furthermore, combined with property (P1), and considering that the boundary of $\mathbb{R}^s_{>0}$ is made up by $\mathbf{0}$ plus sets in which some components are 0 and the others belong to $\mathbb{R}^{s'}_{>0}$ for $1 \leq s' < s$, it follows that $g|_{\mathbb{R}^s_{\geq 0}}$ is injective. Now, given that the Jacobian of g is non-singular in $\mathbb{R}^s_{\geq 0} \setminus \{\mathbf{0}\}$ (lemma A.1), and considering again property (P1), this implies that $g|_{\mathbb{R}^s_{\geq 0}} : \mathbb{R}^s_{\geq 0} \to \mathbb{R}^s_{\geq 0}$ is a homeomorphism, i.e., a continuous function that has a continuous inverse function. □

Proposition A.1. *$f_L : \Delta(S) \to \Delta(S)$ is a bi-Lipschitz homeomorphism whose inverse function $f_L^{-1} : \Delta(S) \to \Delta(S)$ is $f_L^{-1}(\boldsymbol{x}^*) = \frac{g_+^{-1}(\boldsymbol{x}^*)}{\|g_+^{-1}(\boldsymbol{x}^*)\|_1}$.*

[19] An alternative proof can be made based on Theorem 4 in Gale and Nikaido (1965): considering that $J(\boldsymbol{x})$ in $\mathbb{R}^s_{>0}$ is a diagonally dominant matrix, $J(\boldsymbol{x})$ is a P-matrix.

To prove existence of the inverse function f_L^{-1}, we first show that the point $\alpha = \frac{g_+^{-1}(\boldsymbol{x}^*)}{\|g_+^{-1}(\boldsymbol{x}^*)\|_1} \in \Delta(S)$ satisfies $f_L(\alpha) = \boldsymbol{x}^*$, i.e., it belongs to the preimage of $\boldsymbol{x}^*$ under f_L; then we show that this preimage is a singleton, that there is no other point $\alpha' \in \Delta(S)$ satisfying $f_L(\alpha') = \boldsymbol{x}^*$.

Proof. i) α belongs to the preimage of $\boldsymbol{x}^*$ under f_L.

Let $\alpha = \frac{g_+^{-1}(\boldsymbol{x}^*)}{\|g_+^{-1}(\boldsymbol{x}^*)\|_1} \in \Delta(S)$. Note that, by definition, $f_L(\alpha) = \frac{g(\alpha)}{\|g(\alpha)\|_1}$, and remember (P2) that for any scalar $k \geq 0, g(k\boldsymbol{x}) = k^2 g(\boldsymbol{x})$. Then

$$f_L(\alpha) = \frac{g(\alpha)}{\|g(\alpha)\|_1} = \frac{g(\frac{g_+^{-1}(\boldsymbol{x}^*)}{\|g_+^{-1}(\boldsymbol{x}^*)\|_1})}{\|g(\frac{g_+^{-1}(\boldsymbol{x}^*)}{\|g_+^{-1}(\boldsymbol{x}^*)\|_1})\|_1} = \frac{\frac{g(g_+^{-1}(\boldsymbol{x}^*))}{\|g_+^{-1}(\boldsymbol{x}^*)\|_1^2}}{\|(\frac{g(g_+^{-1}(\boldsymbol{x}^*))}{\|g_+^{-1}(\boldsymbol{x}^*)\|_1^2})\|_1} = \frac{\boldsymbol{x}^*}{\|\boldsymbol{x}^*\|_1} = \boldsymbol{x}^*$$

ii) The preimage of x^* under f_L is the set $\{\alpha\}$.

Suppose now that there are $\alpha, \alpha' \in \Delta(S)$ satisfying $f_L(\alpha') = f_L(\alpha) = \boldsymbol{x}^*$ then

$$f_L(\alpha') = f_L(\alpha) \Rightarrow \frac{g(\alpha')}{\|g(\alpha')\|_1} = \frac{g(\alpha)}{\|g(\alpha)\|_1} \Rightarrow \frac{\|g(\alpha)\|_1}{\|g(\alpha')\|_1} g(\alpha') = g(\alpha) \Rightarrow g(k\alpha') = g(\alpha)$$

where $k = \sqrt{\frac{\|g(\alpha)\|_1}{\|g(\alpha')\|_1}}$. As g is injective in $\mathbb{R}^s_{\geq 0}$, $g(k\alpha') = g(\alpha)$ implies that $k\alpha' = \alpha$. But the fact that $\alpha, \alpha' \in \Delta(S)$, combined with $k\alpha' = \alpha$ imply that $\alpha' = \alpha$.

iii) $f_L : \Delta(S) \to \Delta(S)$ is bi-Lipschitz.

It is easy to see that f_L is Lipschitz, as it is a ratio of polynomials and, considering that $\sum_j L_{ij} x_j \geq 1$ for $\boldsymbol{x} \in \Delta(S)$, the denominator is $\sum_i x_i \sum_j L_{ij} x_j \geq 1$. For the inverse function, as $f_L^{-1}(\boldsymbol{x}^*) = \frac{g_+^{-1}(\boldsymbol{x}^*)}{\|g_+^{-1}(\boldsymbol{x}^*)\|_1}$, it only rests to show that, for $\boldsymbol{x}^* \in \Delta(S)$, $\|g_+^{-1}(\boldsymbol{x}^*)\|_1$ is bounded away from 0, as this implies both that the denominator in the defining ratio is bounded away from 0, and that $g_+^{-1}(\Delta(S))$ is contained in a compact set in $\mathbb{R}^s_{\geq 0} \backslash \{\mathbf{0}\}$, so, by the discussion in lemma A.1, $g_+^{-1}(\Delta(S))$ belongs to a region in which the Jacobian of g is bounded away from 0, and, consequently, the Jacobian of the inverse function g_+^{-1} at $\boldsymbol{x}^* \in int(\Delta(S))$ is bounded from above, and the Lipschitz condition holds considering the boundary of $\Delta(S)$ too.

To show that, for $\boldsymbol{x}^* \in \Delta(S)$, $\|g_+^{-1}(\boldsymbol{x}^*)\|_1$ is bounded away from 0, note first that, for $\boldsymbol{x} \in \mathbb{R}^s_{\geq 0}$, $\|g(\boldsymbol{x})\|_1 = \|\boldsymbol{x} \circ (\boldsymbol{L}\boldsymbol{x})\|_1 \geq \|\boldsymbol{x} \circ (\mathbf{1}\boldsymbol{x})\|_1 = \|\boldsymbol{x}\|_1^2$, which implies that, if $\boldsymbol{x}^* \in \Delta(S)$ and $\alpha = g_+^{-1}(\boldsymbol{x}^*)$ then $\|\alpha\|_1 \leq 1$. Besides, $g(\alpha) = \alpha \circ (\boldsymbol{L}\alpha) = \boldsymbol{x}^* \in \Delta(S)$, leading to $\sum_i \alpha_i (\sum_j L_{ij} \alpha_j) = 1$. Let $\bar{L} = Max_{i,j} L_{ij}$. Then we have $\bar{L}\|\alpha\|_1 = \sum_i \alpha_i \bar{L} \geq \sum_i \alpha_i (\sum_j L_{ij} \alpha_j) = 1$, which implies $\|\alpha\|_1 \geq \bar{L}^{-1}$.

□

We next show that, for more than three strategies, f_L^{-1} does not admit a general closed-form algebraic expression. The inverse functions in f_L^{-1} can be expressed as one of the solutions of the polynomial system (5), considering $\boldsymbol{x}^*$ and $\boldsymbol{L}$ as parameters and $\boldsymbol{x} = f_L^{-1}(\boldsymbol{x}^*)$ as the unknowns, and with the additional constraint $\sum x_i = 1$. For the two-strategy case, this system is:

$$x_1^* = \frac{x_1(L_{11}x_1 + L_{12}x_2)}{x_1(L_{11}x_1 + L_{12}x_2) + x_2(L_{21}x_1 + L_{22}x_2)}$$
$$x_2^* = \frac{x_2(L_{21}x_1 + L_{22}x_2)}{x_1(L_{11}x_1 + L_{12}x_2) + x_2(L_{21}x_1 + L_{22}x_2)}$$

Substituting $x_2 = 1 - x_1$ in the first equation, we obtain a second degree polynomial in x_1, with a standard general closed-form solution.

In the general case, using computational algebra techniques – Gröbner bases (Cox et al., 2015)– and considering specific values for the terms L_{ij} (equivalently, for δ and T_{ij}) and for $\boldsymbol{x}^*$, the solutions of the polynomial system (5) with $\sum x_i = 1$ can be obtained by generating first an auxiliary univariate polynomial in one of the variables x_i. An exploration of this auxiliary univariate polynomial of a Gröbner basis for different values of T_{ij}, taking rational values for δ and $\boldsymbol{x}^*$ (so that the coefficients of the univariate polynomial are rational) shows that, typically, this univariate polynomial is irreducible over the rationals and its degree is 4 for three strategies and 8 for four strategies. By the Abel-Ruffini theorem, for more than three strategies we can then expect that, in most cases, there will be no solution in radicals for $\boldsymbol{x} = f_L^{-1}(\boldsymbol{x}^*)$.

As an example, take as stage game any finite 2-player symmetric game and consider any four strategies (denoted 1, 2, 3 and 4) of the game with endogenous separation, with break-up periods $T_{ij} = \min(i, j)$. If we take, for instance, $\delta = \frac{1}{2}$ and the population state $\boldsymbol{x}^* = [\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}]$, the solution to (5) for x_4, considering $\sum_{i=1}^{4} x_i = 1$ and $0 \leq x_i \leq 1$, is the real root in $[0, 1]$ of the univariate (and irreducible over the rationals) polynomial

$$21\,x_4^8 - 2144\,x_4^6 - 2304\,x_4^5 + 49024\,x_4^4 + 61440\,x_4^3 - 251904\,x_4^2 - 344064\,x_4 + 86016$$

This root is an algebraic number (close to the rational 0.218) that does not admit an expression in radicals[20], i.e., it is not expressible in terms of addition, subtraction, multiplication, division, and root extraction (the elementary operations) on rational numbers.

B Extension to symmetric n-player games

The extension to n-player games of the definition of a strategy in a game with endogenous separation is straightforward. For symmetric n-player games in a single population, the formulas to calculate payoffs can be extended as follows.

As before, consider a steady-state pool state $\boldsymbol{x}$ with finite support S. Let $\Theta_n(S)$ be the set of $(n-1)$-tuples of the strategies in S and let $\boldsymbol{\theta} = (\theta_1, ..., \theta_{n-1}) \in \Theta_n(S)$ be one of such $(n-1)$-tuples of strategies. Let $\{i, \boldsymbol{\theta}\}$ represent a partnership in which one of the players plays strategy i and

[20] This can be checked using a Computational Algebra System such as Magma (Bosma et al., 1997), by calculating the Galois group of the polynomial with rational coefficients and checking that the Galois group is not solvable.

the other $n-1$ partners play the strategies in $\boldsymbol{\theta}$ (one each). Let the break-up period $T_{i,\boldsymbol{\theta}}$ be the number of periods that an $\{i,\boldsymbol{\theta}\}$ partnership stays together if the partnership is not broken by exogenous factors, and let the population-to-pool proportion of an $\{i,\boldsymbol{\theta}\}$ partnership (or the expected duration of the partnership) be

$$L_{i,\boldsymbol{\theta}} = \sum_{t=1}^{T_{i,\boldsymbol{\theta}}} \delta^{t-1} = \frac{1-\delta^{T_{i,\boldsymbol{\theta}}}}{1-\delta}.$$

In a steady state of the short-term dynamics, the relation between the fraction x_i^* of players in the whole population using strategy $i \in \mathcal{S}$ and the fraction of players in the pool of singles using each strategy is then given by

$$x_i^* = \frac{x_i \sum_{\boldsymbol{\theta}\in\Theta_n(S)} L_{i,\boldsymbol{\theta}} \prod_{k=1}^{n-1} x_{\theta_k}}{\sum_{j\in S} x_j \sum_{\boldsymbol{\theta}\in\Theta_n(S)} L_{j,\boldsymbol{\theta}} \prod_{k=1}^{n-1} x_{\theta_k}} \equiv f_i(\boldsymbol{x}) \tag{12}$$

The formula for the payoffs $V_i(\boldsymbol{x})$ is extended as

$$V_i(\boldsymbol{x}) = x_i \sum_{\boldsymbol{\theta}\in\Theta_n(S)} V_{i,\boldsymbol{\theta}} \prod_{k=1}^{n-1} x_{\theta_k}$$

where

$$V_{i,\boldsymbol{\theta}} = \sum_{t=1}^{T_{i,\boldsymbol{\theta}}} \delta^{t-1} u(a_t^{i,\boldsymbol{\theta}})$$

and where $u(a_t^{i,\boldsymbol{\theta}})$ is the payoff to an i-strategist at period t of a partnership in which the other $n-1$ players have $\boldsymbol{\theta}$ as their strategy profile. Finally, for any strategy $j \in \mathcal{S}$, we have that the average payoff to a player using pure strategy j, at a steady state with a pool strategy distribution $\boldsymbol{x}$ with finite support S, is

$$v_j(\boldsymbol{x}) = \frac{\sum_{\boldsymbol{\theta}\in\Theta_n(S)} V_{j,\boldsymbol{\theta}} \prod_{k=1}^{n-1} x_{\theta_k}}{\sum_{\boldsymbol{\theta}\in\Theta_n(S)} L_{j,\boldsymbol{\theta}} \prod_{k=1}^{n-1} x_{\theta_k}} \tag{13}$$

As before, the payoff to a (group of players with) strategy distribution $\boldsymbol{y}^*$ in a population with steady strategy distribution in the pool of singles $\boldsymbol{x}$ is

$$v(\boldsymbol{y}^*, \boldsymbol{x}) = \sum_{i\in supp(\boldsymbol{y}^*)} y_i^* v_i(\boldsymbol{x}).$$

For the replicator dynamics (3), where the relationship between $\boldsymbol{x}^*$ and $\boldsymbol{x}$ is given by eq. (12), considering a finite set S of s strategies and an initial population-pool state $\{\boldsymbol{x}^*, \boldsymbol{x}\}$ whose support is contained in S, states can be represented as vectors in $\mathbb{R}^s$, and existence and uniqueness of the solution trajectories can be shown as in appendix A. With some more detail, the functions g_i in

appendix A (i.e., the s components of the vector function $g : \mathbb{R}^s \to \mathbb{R}^s$) become

$$g_i(\boldsymbol{x}) = x_i \sum_{\boldsymbol{\theta} \in \Theta_n(S)} L_{i,\boldsymbol{\theta}} \prod_{k=1}^{n-1} x_{\theta_k}$$

It can then be checked that property (P1) holds, property (P2) generalizes as $g(k\boldsymbol{x}) = k^n g(\boldsymbol{x})$, the Jacobian $J(\boldsymbol{x})$ of g is a (column) strictly diagonally dominant matrix in $\mathbb{R}^s_{>0}$, and lemma A.1 holds, as well as the rest of lemmas and propositions in appendix A. The fact that the function $f_L : \Delta(S) \to \Delta(S)$, with $f_L(\boldsymbol{x}) = \frac{g(\boldsymbol{x})}{\|g(\boldsymbol{x})\|_1}$ (as in eq. (12), but considering $\boldsymbol{x}$ as a vector in $\mathbb{R}^s$, and $i \in \{1, ..., s\}$), is a bi-Lipschitz homeomorphism, can then be used as in the proof of theorem 1 to show that being an NSS^{ES} guarantees Lyapunov stability in the replicator dynamics.

For lemma 1, let h_{ji} be, as in the two-player case, the infinite prospective series of outcomes (n-player action profiles) that a j-strategist will find when entering a population of i-strategists if no exogenous breakup events occur. The extension of lemma 1 to n-player games is then

Lemma B.1. *If a strategy i that never leaves when playing in a partnership in which all player use strategy i, satisfies $v_i(\boldsymbol{i}) > v_j(\boldsymbol{i})$ for any strategy j with $h_{ji} \neq h_{ii}$, then strategy i is a neutrally stable strategy.*

Last, our existence theorem (theorem 3) needs to be adapted for the n-player case by considering the "same-action minimax payoff" w_h of the stage game, i.e., the minimum payoff that the other players can force on a player if they all use the same action: $w_h = \min_j \max_i u(a^{ij\cdots j})$. Here we also consider a same-action minimax profile, i.e., a homogeneous profile in which all players use a same action b such that $\max_i u(a^{ib\cdots b}) = w_h$. The adaptation of the proof is immediate. We state the extended theorem below for completeness.

Theorem 4. *For large enough $\delta < 1$, any finite sequence or pattern of k_p symmetric action profiles $\Phi_p = \{(a_1, ..., a_1), (a_2, ..., a_2), ..., (a_{k_p}, ..., a_{k_p})\}$ with an average payoff $\bar{u} = k_p^{-1} \sum_{k=1}^{k_p} u(a_k, ..., a_k)$ greater than the same-action minimax payoff of the stage game $w_h = min_j \; max_i \; u(a^{ij\cdots j})$ can be sustained as an indefinitely repeated pattern by a neutrally stable strategy i such that, in the equilibrium path h_{ii}, the repeated play of the pattern Φ_p is preceded by a sufficiently long "deviation-deterring" phase in which a same-action minimax action profile $(b, ..., b)$ is played.*

C Neutrally stable equilibrium and neutrally stable distribution

For two-player games, Carmichael and MacLeod (1997) assume that a function f^{-1} exists that gives the pool distribution $\boldsymbol{x}$ that corresponds to a given population distribution $\boldsymbol{x}^*$, and define a Neutrally Stable Equilibrium (NSE) in terms of the population state:

Definition C.1. *Carmichael and MacLeod (1997) A Nash equilibrium population state $\boldsymbol{x}^* \in \mathcal{F}(\mathcal{S})$ is a neutrally stable equilibrium if for every $\boldsymbol{y}^* \in \mathcal{F}(\mathcal{S})$ there exists an $\bar{\epsilon}_y \in (0,1)$ such that for any $\epsilon \in (0, \bar{\epsilon}_y)$,*

$$v_i(f^{-1}((1-\epsilon)\boldsymbol{x}^* + \epsilon \boldsymbol{y}^*)) \geq v_j(f^{-1}((1-\epsilon)\boldsymbol{x}^* + \epsilon \boldsymbol{y}^*))$$

for all $i \in supp(\boldsymbol{x}^*)$ and $j \in supp(\boldsymbol{y}^*)$.

Also for two-player games, Fujiwara-Greve and Okuno-Fujiwara (2009) give the following definition of a Neutrally Stable Distribution (*NSD*), in terms of the pool state:

Definition C.2. *Fujiwara-Greve and Okuno-Fujiwara (2009) A stationary distribution in the matching pool $\boldsymbol{x}$ is a neutrally stable distribution if for any $j \in \mathcal{S}$ there exists an $\bar{\epsilon}_j \in (0,1)$ such that for any $\epsilon \in (0, \bar{\epsilon}_j)$ and any $i \in supp(\boldsymbol{x})$,*

$$v_i((1-\epsilon)\,\boldsymbol{x} + \epsilon \boldsymbol{j}) \geq v_j((1-\epsilon)\,\boldsymbol{x} + \epsilon \boldsymbol{j})$$

where $\boldsymbol{j}$ is the strategy distribution consisting only of strategy j.

The examples below show that the definitions of an *NSE* and an *NSD* present several differences with the standard definition of an NSS. They also show that being an *NSD* does not guarantee Lyapunov stability in the replicator dynamics.

Example C.1. A neutrally stable state that, for $\delta = 0$, does not generate an NSE or an NSD. As a stage game, consider a *good Rock-Paper-Scissors* game (Sandholm, 2010b, p. 82) with payoff matrix

$$U = \begin{bmatrix} 0 & -1 & 3 \\ 3 & 0 & -1 \\ -1 & 3 & 0 \end{bmatrix}$$

In the standard evolutionary setting for the one-shot game, where the possible strategies are just the actions of the stage game, the Nash equilibrium state (in the three actions) $\boldsymbol{x}^{act} = \left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right]$ is a standard[21] NSS (Sandholm, 2010b, p. 90). For the game with endogenous separation, take $\delta = 0$. In this setting, any strategy of the game with endogenous separation is equivalent to an action in the stage game (the first action chosen by the strategy), and vice versa, given that all partnerships are broken after their first period together. Consequently, for a game with endogenous separation and $\delta = 0$ we can generate a NSS^{ES} in strategies ($\boldsymbol{x}^{str}$) from a NSS in actions ($\boldsymbol{x}^{act}$) of the stage game, by just substituting the (fraction of) players using action i in the NSS $\boldsymbol{x}^{act}$ by players using a strategy that begins by playing action i in the game with endogenous separation (and vice

[21] In this linear setting, conditions 1, 2 and 3 are equivalent, and any of them (or the other conditions for neutral stability considered by Bomze and Weibull (1995), all of which are equivalent in this setting) can then be considered the standard condition for neutral stability.

versa). In our example, considering for the repeated game one strategy (strategy 1) that starts playing Rock, one strategy (2) that starts playing Paper, and one strategy (3) that starts playing Scissors, the state (in the three strategies) $\boldsymbol{x}^{str} = \left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right]$ is an NSS^{ES} for $\delta = 0$.

On the other hand, $\boldsymbol{x}^{str} = \left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right]$ is neither an *NSE* nor an *NSD*, because an increase in the fraction of strategy 1 creates an advantage for strategy 1, which is the invader's strategy, over strategy 3, which is one of the strategies in the support of the incumbent population $\boldsymbol{x}^{str}$:

$$v_1((1-\epsilon)\,\boldsymbol{x}^{str} + \epsilon \boldsymbol{e}^1) = \frac{2}{3}(1-\epsilon) > \frac{2}{3}(1-\epsilon) - \epsilon = v_3((1-\epsilon)\,\boldsymbol{x}^{str} + \epsilon \boldsymbol{e}^1) \text{ for all } \epsilon \in (0,1)$$

where $\boldsymbol{e}^1 = [1, 0, 0]$.

The reason for this difference is that the standard NSS (or the NSS^{ES}) definitions require a robustness condition on the average payoff obtained by distribution $\boldsymbol{x}$, while the *NSE* and *NSD* definitions require (other) robustness conditions on the payoffs obtained by each of the pure strategies in the support of $\boldsymbol{x}$.

While example C.1 shows an NSS that, for $\delta = 0$, does not generate an *NSE* or an *NSD*, example C.2 below shows an *NSD* that does not satisfy finite-set neutral stability, or Lyapunov stability in the replicator dynamics.

Example C.2. An NSD that is not a standard NSS in a restricted finite strategy space. Consider a game with endogenous separation whose stage game has the following payoff matrix

$$U = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Take $\delta = 0$ and three strategies (1, 2 and 3) of the repeated game, each one playing the corresponding action of the stage game at the initial stage of a partnership. The Nash equilibrium state $\boldsymbol{x} = [1, 0, 0]$ is an *NSD*, because it is robust to independent invasions by any pure strategy of the repeated game, regardless of its initial action (which is all that matters when $\delta = 0$).

On the other hand, if we consider the finite set of strategies $S = \{1, 2, 3\}$, and we take $\Delta(S)$ as the state space, we have that $\boldsymbol{x}$ is not a standard[22] NSS, because any distribution in which strategies 2 and 3 are used would earn a payoff of 0 against $\boldsymbol{x}$, and a positive payoff against itself. The reason for this difference is that the *NSD* definition only requires robustness to invasions by monomorphic populations, while the NSS condition considers polymorphic invaders too[23]. This

[22] As in the previous example, in this linear setting, conditions 1, 2 and 3 are equivalent, and any of them can then be considered as the standard condition for neutral stability.

[23] Vesely and Yang (2010) and Vesely and Yang (2012) also use the definition of an *NSD* from Fujiwara-Greve and Okuno-Fujiwara (2009), with one important difference. While Fujiwara-Greve and Okuno-Fujiwara (2009) consider deterministic strategies, as we do, and as Carmichael and MacLeod (1997) do too, Vesely and Yang (2010, 2012) consider behavioral strategies, which allows the individuals that use these strategies to randomize. One important

example also shows that being an *NSD* does not guarantee Lyapunov stability in the replicator dynamics, as state $\boldsymbol{x} = [1, 0, 0]$ is an *NSD*, but it is dynamically unstable in the replicator dynamics for the three strategies.

For polymorphic equilibria, Vesely and Yang (2012) show that the polymorphic equilibria made up by combinations of trust-building strategies studied by Fujiwara-Greve and Okuno-Fujiwara (2009) are not robust to invasions by mixed strategists. For the same reasons, those polymorphic equilibria are not robust to invasions by a polymorphic population (as considered here), and consequently do not constitute an NSS^{ES}.

Bibliography

Bendor, J. and Swistak, P. (1995). Types of evolutionary stability and the problem of cooperation. *Proceedings of the National Academy of Sciences*, 92(8):3596–3600.

Bomze, I. M. and Pötscher, B. M. (1989). *Game theoretical foundations of evolutionary stability. Lecture notes in economics and mathematical systems*, volume 324. Springer Science & Business Media.

Bomze, I. M. and Weibull, J. W. (1995). Does neutral stability imply Lyapunov stability? *Games and Economic Behavior*, 11:173–192.

Bosma, W., Cannon, J., and Playoust, C. (1997). The Magma algebra system. I. The user language. *J. Symbolic Comput.*, 24(3-4):235–265. Computational algebra and number theory (London, 1993).

Carmichael, H. L. and MacLeod, W. B. (1997). Gift Giving and the Evolution of Cooperation. *International Economic Review*, 38(3):485.

Cox, D. A., Little, J., and O'Shea, D. (2015). *Ideals, Varieties, and Algorithms*. Undergraduate Texts in Mathematics. Springer International Publishing, Cham.

Datta, S. (1996). Building Trust. *STICERD - Theoretical Economics Paper Series*.

Deb, J., Sugaya, T., and Wolitzky, A. (2020). The Folk Theorem in Repeated Games With Anonymous Random Matching. *Econometrica*, 88(3):917–964.

Friedman, D. (1998). On economic applications of evolutionary game theory. *Journal of Evolutionary Economics*, 8(1):15–43.

consequence of that is that their mutants can mix. A population that consists of only strategy 1 therefore would not be an *NSD* in the version of Vesely and Yang (2010, 2012), because a mutant behavioral strategy that mixes between strategies 1 and 2 would be considered there.

Friedman, J. W. (1971). A non-cooperative equilibrium for supergames. *The Review of Economic Studies*, 38(1):1–12.

Fudenberg, D. and Levine, D. (2008). *A long-run collaboration on long-run games*. World Scientific.

Fudenberg, D. and Maskin, E. (1986). The folk theorem in repeated games with discounting or incomplete information. *Econometrica (1986-1998)*, 54(3):533.

Fujiwara-Greve, T., Greve, H. R., and Jonsson, S. (2016). Asymmetry of customer loss and recovery under endogenous partnerships: Theory and evidence. *International Economic Review*, 57(1):3–30.

Fujiwara-Greve, T. and Okuno-Fujiwara, M. (2009). Voluntarily separable repeated prisoner's dilemma. *Review of Economic Studies*, 76(3):993–1021.

Fujiwara-Greve, T., Okuno-Fujiwara, M., and Suzuki, N. (2015). Efficiency may improve when defectors exist. *Economic Theory*, 60(3):423–460.

Gale, D. and Nikaido, H. (1965). The Jacobian matrix and global univalence of mappings. *Math. Ann.*, 159(2):81–93.

García, J. and van Veelen, M. (2016). In and out of equilibrium I: Evolution of strategies in repeated games with discounting. *Journal of Economic Theory*, 161(1):161–189.

Ghosh, P. and Ray, D. (1996). Cooperation in Community Interaction without Information Flows. *The Review of Economic Studies*, 63(3):491.

Gordon, W. B. (1972). On the diffeomorphisms of euclidean space. *The American Mathematical Monthly*, 79(7):755–759.

Hamilton, W. D. and Axelrod, R. (1981). The evolution of cooperation. *Science*, 211(27):1390–1396.

Hofbauer, J. and Sigmund, K. (1998). *Evolutionary Games and Population Dynamics*. Cambridge University Press.

Idzik, A., Kulpa, W., and Maćkowiak, P. (2014). Equivalent forms of the Brouwer fixed point theorem I. *Topological Methods in Nonlinear Analysis*, 44(1):263–276.

Izquierdo, L. R., Izquierdo, S. S., and Sandholm, W. H. (2019). An introduction to ABED: Agent-based simulation of evolutionary game dynamics. *Games and Economic Behavior*, 118:434 – 462.

Izquierdo, L. R., Izquierdo, S. S., and Vega-Redondo, F. (2014). Leave and let leave: A sufficient condition to explain the evolutionary emergence of cooperation. *Journal of Economic Dynamics and Control*, 46:91–113.

Izquierdo, S. S., Izquierdo, L. R., and Vega-Redondo, F. (2010). The option to leave: Conditional dissociation in the evolution of cooperation. *Journal of Theoretical Biology*, 267(1):76–84.

Kranton, R. E. (1996). The Formation of Cooperative Relationships. *Journal of Law, Economics, and Organization*, 12(1):214–233.

Kurokawa, S. (2019). Three-player repeated games with an opt-out option. *Journal of Theoretical Biology*, 480:13–22.

Kurokawa, S. (2021). Effect of the group size on the evolution of cooperation when an exit option is present. *Journal of Theoretical Biology*, 521:110678.

Lee, J. M. (2003). *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer New York, New York, NY.

Mailath, G. J. and Samuelson, L. (2006). *Repeated Games and Reputations*. Oxford University Press.

Sandholm, W. H. (2010a). Pairwise comparison dynamics and evolutionary foundations for Nash equilibrium. *Games*, 1:3–17.

Sandholm, W. H. (2010b). *Population games and evolutionary dynamics*. The MIT Press.

Schuessler, R. (1989). Exit Threats and Cooperation under Anonymity. *Journal of Conflict Resolution*, 33(4):728–749.

Selten, R. (1983). Evolutionary stability in extensive two-person games. *Mathematical Social Sciences*, 5(3):269–363.

Selten, R. and Hammerstein, P. (1984). Gaps in harley's argument on evolutionarily stable learning rules and in the logic of "tit for tat". *Behavioral and Brain Sciences*, 7(1):115–116.

Serre, D. (2002). *Matrices*, volume 216 of *Graduate Texts in Mathematics*. Springer New York, New York, NY.

Taylor, J. L. (2012). *Foundations of analysis*. American Mathematical Society.

Taylor, P. D. and Jonker, L. B. (1978). Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, 40:145–156.

Thomas, B. (1984). Evolutionary stability: states and strategies. *Theoretical Population Biology*, 26(1):49–67.

Thomas, B. (1985). On evolutionarily stable sets. *Journal of Mathematical Biology*, 22:105–115.

van Veelen, M. and García, J. (2019). In and out of equilibrium II: Evolution in repeated games with discounting and complexity costs. *Games and Economic Behavior*, 115:113–130.

van Veelen, M., García, J., Rand, D. G., and Nowak, M. A. (2012). Direct reciprocity in structured populations. *Proceedings of the National Academy of Sciences USA*, 109(25):9929–9934.

Vesely, F. and Yang, C.-L. (2010). On optimal and neutrally stable population equilibrium in voluntary partnership prisoner's dilemma games. *SSRN*.

Vesely, F. and Yang, C.-L. (2012). Breakup, secret handshake and neutral stability in repeated prisoner's dilemma with option to leave: A note. *SSRN*.

Watson, J. (1999). Starting Small and Renegotiation. *Journal of Economic Theory*, 85(1):52–90.

Weibull, J. W. (1995). *Evolutionary Game Theory*. MIT Press, Cambridge.

Wubs, M., Bshary, R., and Lehmann, L. (2016). Coevolution between positive reciprocity, punishment, and partner switching in repeated interactions. *Proceedings of the Royal Society B: Biological Sciences*, 283(1832):20160488.