



In and out of equilibrium I: Evolution of strategies in repeated games with discounting

Julián García^a, Matthijs van Veelen^{b,*}

^a Faculty of Information Technology, Monash University, Clayton, 3800 VIC, Australia

^b Department of Economics, Universiteit van Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands

Received 7 November 2014; final version received 28 October 2015; accepted 21 November 2015

Available online 2 December 2015

Abstract

In the repeated prisoner's dilemma there is no strategy that is evolutionarily stable, and a profusion of neutrally stable ones. But how stable is neutrally stable? We show that in repeated games with large enough continuation probabilities, where the stage game is characterized by a conflict between individual and collective interests, there is always a neutral mutant that can drift into a population that is playing an equilibrium, and create a selective advantage for a second mutant. The existence of stepping stone paths out of *any* equilibrium determines the dynamics in finite populations playing the repeated prisoner's dilemma. © 2015 Elsevier Inc. All rights reserved.

JEL classification: C73

Keywords: Repeated games; Evolutionary game theory

“Everything changes, nothing remains the same”

Buddha

1. Introduction

Repeated games typically have many equilibria. But how stable are these equilibria? And are some equilibria perhaps more stable than others? In this paper we use refinements from

* Corresponding author.

E-mail addresses: julian.garcia@monash.edu (J. García), c.m.vanveelen@uva.nl (M. van Veelen).

evolutionary game theory to determine how stable equilibria are, and to help understand the evolutionary dynamics in populations of individuals that are playing such games.

For the repeated prisoner's dilemma we know that there is no pure strategy that is evolutionarily stable (Selten and Hammerstein, 1984), and it is straightforward to extend that argument to non-trivial repeated games in general, and to mixed strategies with finite support. There are, on the other hand, very many neutrally stable strategies in the repeated prisoner's dilemma (Bendor and Swistak, 1995, 1997, 1998). These NSS'es range from fully defecting to fully cooperative. Neutral stability is a weaker version of evolutionary stability, that does not imply asymptotic stability in the replicator dynamics. That, however, does not rule out that there might be a *set* of NSS'es that is evolutionarily stable.

In order to investigate whether or not there are such evolutionarily stable sets, and, more in general, to determine how stable these NSS'es are, we will use the concept of robustness against indirect invasions (RAII, van Veelen, 2012). RAII is less strict than ESS, because it allows for neutral mutants. It is also more strict than NSS, because it does not allow for neutral mutants that serve as a stepping stone for other mutants that have an actual selective advantage, once the first mutant has gained enough of a foothold in the population, for instance through neutral drift. Robustness against indirect invasions preserves a tight link with the replicator dynamics for infinite populations, as well as with stochastic, finite population dynamics of which the replicator dynamics are the large population limit (see Traulsen et al., 2005, 2006). If a strategy is RAII, then it is an element of an ES set (van Veelen, 2012; Balkenborg and Schlag, 2001), which is, as a set, asymptotically stable in the replicator dynamics (Thomas, 1985). Vice versa, if there is an ES set, then all of its elements are RAII. Moreover, the way robustness against indirect invasions deals with neutral mutants implies that it matches the qualitative equilibrium analysis typically applied to stochastic, finite population dynamics, in which neutral mutants play a pivotal role (Nowak, 2006).

It turns out that for repeated games in which the stage game shows a conflict between individual and collective interests – like the prisoner's dilemma – no equilibrium is RAII, provided that the continuation probability is sufficiently high. In other words, any equilibrium can be upset by an at first harmless mutant, which serves as a stepping stone, or a springboard, for the invasion of a second mutant. Stepping stone paths with decreasing cooperation exist for all equilibria, unless there is no cooperation in equilibrium to begin with. Stepping stone paths with increasing cooperation exist for all equilibria that fall sufficiently short of full cooperation. What “sufficiently short” is, depends on the continuation probability.

Simulations show that not only do stepping stone paths out of any equilibrium exist for the repeated prisoner's dilemma, evolution also finds them. With a mutation procedure that is not biased, and that allows for all finite state automata to be reached as mutants, we find that indirect invasions are indeed the driver of the dynamics. The population finds itself in equilibrium most of the time, with regular transitions from equilibrium to equilibrium that do indeed follow these stepping stone paths, both with rising and with declining levels of cooperation. The implications for the dynamics are further illustrated by comparing the repeated prisoner's dilemma – which has no NSS'es that are RAII – to a repeated coordination game which does. We find that if the population size increases, the number of transitions out of equilibrium in the repeated coordination game quickly goes to zero, while in the repeated prisoner's dilemma the population keeps moving from equilibrium to equilibrium regularly. Whether or not there are equilibria that are RAII therefore makes a huge difference for the evolutionary dynamics in repeated games.

Under very reasonable dynamics, equilibria of the repeated games we tend to look at when we study cooperation in repeated interactions, are therefore relatively *unstable* when we compare

them to games for which we do have equilibria that are RAI or even ESS. But even though nothing is as stable as one would perhaps hope for, the typical transitions in and out of cooperation imply that reciprocity and cooperation will evolve, time and again, and to varying levels, and that they will also be undermined, over and over again. The non-existence of strategies that are RAI in the repeated prisoner's dilemma also implies that if we study a restricted strategy space, and find one or more strategies to be stable, we can always make that stability result disappear by enlarging the strategy space to include the right stepping stones.

1.1. Definitions and results in the literature

The literature on evolution in repeated games has used a variety of definitions of evolutionary stability. At first glance it therefore seems to contain contradicting results as well as duplications. [Axelrod and Hamilton \(1981\)](#) consider the repeated prisoner's dilemma, and claim that both Tit-for-tat and AllD are evolutionarily stable. [Bendor and Swistak \(1995\)](#) point out that they have only checked the first condition in [Definition 1](#) below. This implies that they have shown that these strategies are symmetric Nash equilibria, but not that they are evolutionarily stable according to the now generally accepted definition by [Maynard Smith and Price \(1973\)](#). We first give the pure strategy version of that definition, applied to repeated games – which is definition [2] in [Bendor and Swistak \(1995\)](#). Let \mathcal{S} be a space of all pure strategies for the repeated game, and let $\Pi : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}$ be the payoff function, where $\Pi(S, T)$ is the payoff of a player playing strategy S against a player playing strategy T . The game is symmetric, in that the payoff of the opponent T in this encounter is given by $\Pi(T, S)$.

Definition 1. (PURE ESS; MAYNARD SMITH & PRICE) A strategy $S \in \mathcal{S}$ is evolutionarily stable if both

$$\Pi(S, S) \geq \Pi(T, S) \text{ for all } T \text{ and}$$

$$\text{if } \Pi(S, S) = \Pi(T, S) \text{ then } \Pi(S, T) > \Pi(T, T) \text{ for all } T \neq S$$

A strategy that is evolutionarily stable is asymptotically stable in the replicator dynamics, if the space of pure strategies is finite ([Taylor and Jonker, 1978](#)).

Considering the repeated prisoner's dilemma, [Selten and Hammerstein \(1984\)](#) argue Tit-for-tat is not an ESS, because AllC is a neutral mutant; both strategies get the same payoff, both against Tit-for-tat, and against AllC. Their argument is easily extended to imply that every pure strategy in every non-trivial repeated game has neutral mutants (where a trivial game would be one in which the stage game has a singleton strategy set). The reason is that for every equilibrium strategy S playing against itself, there is always an off-equilibrium path. On the off-equilibrium path a strategy can be changed without consequences for payoffs. This creates a mutant strategy T for which $\Pi(T, S) = \Pi(S, S) = \Pi(T, T) = \Pi(S, T)$ and hence no strategy S can be ESS (see also Selten, 1983, for a discussion of evolutionary stability in extensive form games with finite trees).

[Boyd and Lorberbaum \(1987\)](#) also show that no pure ESS exists, but, as [Bendor and Swistak \(1995\)](#) point out, they use a different definition of an ESS.

Definition 2. (PURE ESS; BOYD & LORBERBAUM) A strategy $S \in \mathcal{S}$ is evolutionarily stable if for all T both

$\Pi(S, S) \geq \Pi(T, S)$ and

if $\Pi(S, S) = \Pi(T, S)$ then $\Pi(S, U) \geq \Pi(T, U)$ for all U

This definition is both less and more stringent than the pure strategy version of Maynard Smith & Price. It is less stringent, because the last inequality is non-strict. More importantly, it is also more stringent, because in the second half of the definition, S is required to do worse than T , not just against T , but against all possible strategies U . Bendor and Swistak (1995) give a beautifully simple example of a 3×3 matrix game that shows that some strategies that are ESS'es, according to the standard definition, are not ESS'es according to the definition of Boyd & Lorberbaum. The latter definition therefore rules out equilibria that are relevant for dynamics. The literature has therefore converged on using Maynard Smith & Price's definition.

Bendor and Swistak (1995) furthermore use a notion of stability – their definition [3] – that is now called neutral stability (see Maynard Smith, 1982; Weibull, 1995).

Definition 3. (PURE NSS) A strategy S is neutrally stable if both

$\Pi(S, S) \geq \Pi(T, S)$ for all T and

if $\Pi(T, T) = \Pi(S, T)$ then $\Pi(T, S) \geq \Pi(S, S)$

A strategy that is neutrally stable is Lyapunov stable in the replicator dynamics (Thomas, 1985; Bomze and Weibull, 1995), but not necessarily asymptotically stable.

Bendor and Swistak (1995) show that there are equilibria, ranging from fully defecting to fully cooperative, that are NSS.

Axelrod and Hamilton (1981), Selten and Hammerstein (1984), Boyd and Lorberbaum (1987) and Bendor and Swistak (1995) focus on pure strategies, as do many other papers about evolution in repeated games – see for instance Fudenberg and Maskin (1990), Binmore and Samuelson (1992),¹ and Cooper (1996). The standard definition of an ESS however also allows for mixed strategies. We would like to do the same. We therefore equate mixed strategies with probability distributions over \mathcal{S} . Section 4.1 shows how \mathcal{S} can be endowed with a metric to make it a separable metric space. Strategies P and Q will then be probability measures on $(\mathcal{S}, \mathcal{B})$ with Borel σ -field \mathcal{B} .

Definition 4. (MIXED ESS; MAYNARD SMITH & PRICE) A strategy P is evolutionarily stable if both

$\Pi(P, P) \geq \Pi(Q, P)$ for all Q and

if $\Pi(P, P) = \Pi(Q, P)$ then $\Pi(P, Q) > \Pi(Q, Q)$ for all $Q \neq P$

Below we state a straightforward generalization of Selten & Hammerstein's argument, not only to non-trivial games in general, but also to finite mixtures of strategies. A finite mixture only allows for a finite number of equilibrium paths and hence there is always an infinite number of off-equilibrium paths left on which behavior can be changed without affecting payoffs. In terms of Definition 4, this implies that no strategy P can be ESS if P is a probability distribution that puts probability 1 on a finite number of strategies.

¹ With the exception of Sections 8 and 9. They do however not use the standard mixed ESS, but define a *polymorphous MESS*, which is special version for lexicographic preferences, where complexity costs matter only if payoffs are equal.

Proposition 5. *In a non-trivial repeated game there is no finite mixture of strategies that is evolutionarily stable.*

Proof. See [Appendix A](#); this is a straightforward generalization of [Selten and Hammerstein \(1984\)](#). □

[Farrell and Ware \(1989\)](#) show that no finite mixture can be ESS in the repeated prisoner's dilemma. They, however, use the mixed equilibrium counterpart of [Boyd & Lorberbaum's](#) definition.

For infinitely large strategy sets, the link between evolutionary stability and asymptotic stability in replicator dynamics is a bit less straightforward than for finite sets of pure strategies ([van Veelen and Spreij, 2009](#)), but the standard definition is still more informative about the dynamical behavior. Our point of departure is therefore that with non-trivial repeated games, there is no finite mixture that is evolutionarily stable, and that for the repeated prisoner's dilemma there is a range of strategies that are neutrally stable, from fully defecting to fully cooperative, as shown in [Bendor and Swistak \(1995, 1997, 1998\)](#). The results in this paper will imply that none of these NSS'es are RAI, which indicates that they are relatively *unstable*, compared to NSS'es that are RAI.

[Bendor and Swistak \(1995, 1997, 1998\)](#) also find that the 'basins of non-repulsion' of more cooperative NSS'es are larger than those of less cooperative ones. This suggests that more cooperative NSS'es might also be more stable in reasonable evolutionary dynamics. Our finding that none of the NSS'es are RAI suggests that all NSS'es, cooperative and less cooperative ones, are, in a basic sense, equally unstable. In [Section 2.2](#) we show why the size of the basin of non-repulsion is less informative about the dynamics than the fact that all NSS'es are susceptible to indirect invasions, as the simulations also illustrate.

1.2. Other settings

Our setup allows for discounting, but has no complexity costs, no errors, and no population structure. These other elements are obviously important. Within classical game theory, complexity costs are taken into account in [Rubinstein \(1986\)](#) and [Abreu and Rubinstein \(1988\)](#) – see also [Kalai and Stanford \(1988\)](#). Players there have lexicographic preferences, where complexity costs only matter in the event of payoff ties. Going from Nash equilibria to NSS'es or ESS'es with lexicographic preferences can be done in different ways (see [Samuelson and Swinkels, 2003](#)). These different ways lead to different results. All strategies that satisfy the lexicographic stability concept in [Binmore and Samuelson \(1992\)](#) are efficient, while AIID is the only strategy that survives refinement in [Vohij \(2002\)](#). Rather than having lexicographic preferences, [Cooper \(1996\)](#) considers repeated games with (small) complexity costs, and proves that there are NSS'es with levels of cooperation that range from fully defecting to fully cooperative. In [van Veelen and García \(2012\)](#) we extend Cooper's result from the case without discounting to the case with discounting. Moreover, with small, but positive complexity costs, AIID is an ESS, and even has a uniform invasion barrier, while none of the cooperative equilibria are RAI. These cooperative equilibria therefore are less stable than AIID.

Without complexity costs, the results from this paper would apply, and we would have a range of equally unstable NSS'es. Choosing ever smaller complexity costs for a fixed population size brings the dynamics ever closer to what they would be without complexity costs. Choosing ever

larger population sizes for a fixed complexity cost brings the dynamics ever closer to what the prediction with complexity costs suggests (van Veelen and García, 2012, see also Schlag, 1993).

The importance of mistakes is highlighted by, among others, May (1987), Hirschleifer and Coll (1988), Fudenberg and Maskin (1990), Nowak and Sigmund (1990, 1992), Kim (1994), Miller (1996), and Axelrod (1997). The approach in Fudenberg and Maskin (1990) is also lexicographic; payoffs in the presence of n mistakes are of lower-order importance than payoffs in the presence of $n - 1$ mistakes. For the repeated prisoner's dilemma, their results imply that equilibria must be fully cooperative in order to satisfy their equilibrium requirement.

In van Veelen et al. (2012) we combine repetition with population structure.² There we consider repeated prisoner's dilemmas with "equal gains from switching" (which means that the game can be parametrized with parameters for costs and benefits only; see Nowak and Sigmund, 1990) and pure equilibria only. Here we consider pure equilibria in general repeated games (Theorems 7 and 8) and finite mixtures in general prisoner's dilemmas (Theorems 10 and 11), all in well-mixed populations, or, in other words, without population structure. The simulation program used in van Veelen et al. (2012) is the same as the program we use here, but then extended with population structure.

Evolutionary stability in the finitely repeated prisoner's dilemma is considered in Cressman (1996). The relevance of backward induction in the finitely repeated case implies that the complications there are rather different from the infinitely repeated case that we consider.

The proofs of Theorems 8 and 11 use an argument that is similar to the "secret handshake" argument in Robson (1990); see also Fudenberg and Maskin (1990).

2. Preliminaries

2.1. Robustness against indirect invasions

With no ESS and many NSS'es, a natural question to ask is exactly how stable or unstable those NSS'es are. As the only difference between the definitions of an ESS and an NSS is that the latter allows for invasions by neutral mutants, while the former does not, the question then becomes how much harm these neutral mutants can do.

If we think for example of the repeated prisoner's dilemma and the strategy Tit-for-tat, then it is clear that the worst enemy to cooperation is not AllD, but a succession of first AllC and then AllD. Tit-for-tat can easily resist an invasion of AllD, but not of AllC, which is a neutral mutant of Tit-for-tat. If AllC attains a high enough share of the population by random drift, then AllD gets a strict advantage and can invade the population (see Fig. 1).

In order to distinguish between neutral mutants that do and neutral mutants that do not serve as a springboard for other mutants, we use the concept of robustness against indirect invasions (van Veelen, 2012). For a strategy to be robust against indirect invasions (RAII) it must not only be an NSS, but there must also not be a sequence of neutral mutants that open the door for each other, one after another, until some mutant strategy has an actual selective advantage. We reproduce the definition below. As a preparation, three sets are defined for any strategy P : the set of (evolutionary) worse, equal and better performers against P .

$$S_W(P) = \{Q \mid \Pi(Q, P) < \Pi(P, P) \text{ or} \\ (\Pi(Q, P) = \Pi(P, P) \text{ and } \Pi(Q, Q) < \Pi(P, Q))\}$$

² van Veelen et al. (2012) is a follow-up on this paper, but published before it.

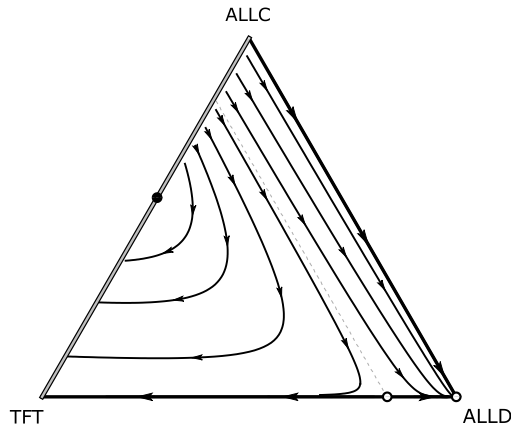


Fig. 1. An indirect invasion with decreasing cooperation. The simplex shows the replicator dynamics in the normal form game generated by *Allc*, *AllD*, and *Tft*, for $\delta = \frac{3}{4}$. *Allc* is a neutral mutant of *Tft*, and *AllD* can directly invade *Allc*.

$$S_E(P) = \{Q \mid \Pi(Q, P) = \Pi(P, P) \text{ and } \Pi(Q, Q) = \Pi(P, Q)\}$$

$$S_B(P) = \{Q \mid \Pi(Q, P) > \Pi(P, P) \text{ or } (\Pi(Q, P) = \Pi(P, P) \text{ and } \Pi(Q, Q) > \Pi(P, Q))\}$$

With the sets of worse, better and equal performers at hand, we can restate the definitions of an NSS and an ESS in a shorter way. A strategy P is an NSS if $S_B(P) = \emptyset$ and a strategy P is an ESS if $S_B(P) = \emptyset$ and $S_E(P) = \{P\}$. These sets also help define robustness against indirect invasions.

Definition 6. A strategy P is robust against indirect invasions (RAII) if

- (1) $S_B(P) = \emptyset$ and
- (2) $\nexists Q^1, \dots, Q^n, n \geq 2$, such that
$$\begin{cases} Q^1 \in S_E(P) \\ Q^i \in S_E(Q^{i-1}), \quad 2 \leq i \leq n-1 \\ Q^n \in S_B(Q^{n-1}) \end{cases}$$

If a strategy is RAI, then it gives us a set of strategies that is asymptotically stable in the replicator dynamics. [van Veelen \(2012\)](#) uses Theorem 3 in [Balkenborg and Schlag \(2001\)](#) to show that if P is RAI, then this strategy, together with its neutral mutants, and the neutral mutants of the neutral mutants, and so on, form a (minimal) ES-set. From [Thomas \(1985\)](#) we furthermore know that ES-sets are asymptotically stable in the replicator dynamics. The converse is also true; if X is an ES-set, and $P \in X$, then P is RAI.

The stepping stones that would make a strategy not RAI are not to be confused with the stepping stones that play a role in what is called step-by-step evolution in [Ellison \(2000\)](#). There step-by-step evolution can take the population from one ESS to the basin of attraction of another one through a sequence of population states, all of which might also be ESS'es. These steps will be relevant, if noise is relatively unlikely to get a population out of one ESS directly into the basin of attraction of another, compared to getting it to the same ESS through a sequence of intermediate ESS'es. Each step nonetheless may require a mutant that at low frequency is at an actual disadvantage. If we find that an equilibrium is not RAI, on the other hand, all that is required to leave it is a sequence of neutral mutants and a mutant with an actual advantage taking

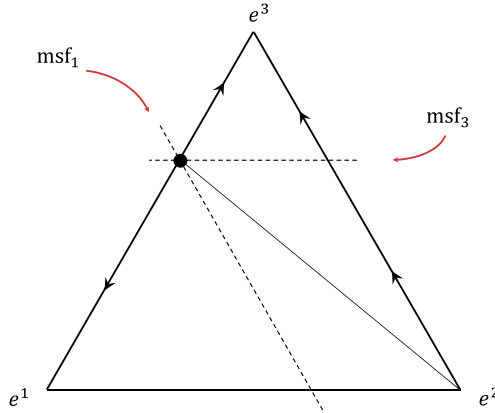


Fig. 2. The minimal stabilizing frequencies for pure strategies 1 and 3.

over the population. That implies that such an equilibrium is far more unstable than an ESS is, because leaving an ESS will always have to involve noise temporarily overwhelming the force of selection (see also [Kandori et al., 1993](#); [Young, 1998](#); [Sandholm and Staudigl, 2015](#)).

2.2. Minimal stabilizing frequencies and *p*-dominance

[Bendor and Swistak \(1995, 1997\)](#) showed that there is a range of NSS'es for the repeated prisoner's dilemma (Theorem 4). This will be our point of departure. They also show that fully cooperative, retaliating strategies have larger basins of non-repulsion than nasty ones, that never cooperate first (Theorems 5 and 6). These results concern “minimal stabilizing frequencies”, and how these depend on the continuation probability. The interpretation of these theorems is that, because of their larger basin of non-repulsion, cooperative equilibria are easier to obtain, and harder to disrupt than defecting ones. This interpretation rests on the assumption that the robustness of an equilibrium can be determined by looking at its minimal stabilizing frequency.

In order to illustrate that robustness in reasonable stochastic dynamics does not have to go hand in hand with a low minimal stabilizing frequency, we can look at a 3×3 matrix game example.

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ -1 & 1 & 2 \end{bmatrix}$$

For this game, the minimal stabilizing frequencies of strategies 1 and 3 are $1/3$ and $2/3$, respectively.³ If x_i denotes the share of pure strategy i , then $\pi(e^1, x) = \pi(e^2, x) = 1$ and $\pi(e^3, x) = 1 - 2x_1 + x_3$. That implies that $msf_1 = 1/3$ is both the lowest share for which the payoff of pure strategy 1 against x is at least as large as any other strategy present in the population if $x_1 > msf_1$, and the lowest share for which the payoff of pure strategy 1 against x is at least as large as the average, under the same condition. Similarly, $msf_3 = 2/3$ (see also [Fig. 2](#)).

³ The minimal stabilizing frequency is not formally defined in [Bendor and Swistak \(1995, 1997\)](#), but the description there leaves little room for ambiguity in this example.

Strategy 1 therefore has a lower minimal stabilizing frequency than strategy 3, and it has a larger basin of non-repulsion. The interpretation in [Bendor and Swistak \(1995, 1997\)](#) would then imply that strategy 1 is more robust than strategy 3. That, however, is not the case. Strategy 1 is not RAI, while strategy 3 is, and in many reasonable stochastic dynamics, it is strategy 1 that is left much more easily than strategy 3. This is caused by the availability of the indirect invasion, out of 1, through 2, and then to 3, while strategy 3 does not have a similar stepping stone path out.

Also with repeated prisoner’s dilemmas, we will see that the equilibria are left through indirect invasions, and not over the edges that determine how large the minimal stabilizing frequency is. Having a lower minimal stabilizing frequency therefore is no guarantee that a strategy is also more robust in the dynamics.

Although stabilizing frequencies are not formally defined in [Bendor and Swistak \(1995, 1997\)](#), its application in these papers would be consistent with a non-strict version of p -dominance ([Morris et al., 1995](#)). A pure strategy in a symmetric game with finitely many pure strategies is p -dominant if it is the unique best response when it is played by at least a proportion p of the population. The application of stabilizing frequencies in [Bendor and Swistak \(1995, 1997\)](#) suggests that the existence of *alternative* best responses at a given frequency does not prevent that frequency from being labelled stabilizing. In other words, a strategy does not have to do strictly better than other strategies at frequencies below a threshold, in order for that threshold to qualify as a stabilizing frequency, whereas it does have to do strictly better than all other strategies at frequency p , or larger, in order for this strategy to be called p -dominant.

Comparing basins of attraction by looking at p -dominance can be useful for describing dynamic behavior; see for instance [Ellison \(2000\)](#), [Sandholm \(2001\)](#), and [Oyama et al. \(2015\)](#). The example above illustrates why slightly weaker, but similar results could not be derived using a non-strict version of p -dominance.

2.3. Repeated games

The example above with AIIC, AIID and Tit-for-tat shows that in the repeated prisoner’s dilemma Tit-for-tat is not RAI. In Section 3 we will see that it is not just Tit-for-tat, and not just the repeated prisoner’s dilemma. [Theorem 7](#) implies that any equilibrium in which actions are played that are not also equilibria of the stage game can be undermined by a succession of two mutations. [Theorem 8](#) states that if there are possible gains from alternative courses of action, and the probability of continuation is sufficiently high, then also a stepping stone route into more cooperation exists. Together they imply that for instance in repeated prisoner’s dilemmas, no equilibrium is RAI, if the continuation probability is high enough. Both theorems come in a pure strategy version for general repeated games ([Theorems 7 and 8](#)) and a mixed strategy version for repeated prisoner’s dilemmas ([Theorems 10 and 11](#)).

We start with a few formal definitions. Consider a symmetric one-shot 2-player game g characterized by a set of players $I = \{1, 2\}$, an action space A , equal for both players, and a payoff function $\pi : A \times A \rightarrow \mathbb{R}$. Using a discount factor $\delta \in (0, 1)$, interpreted as a continuation probability, this one-shot game is turned into a repeated one, which will be called $\Gamma(\delta)$. A history at time t is a list of the actions played up to and including time $t - 1$, where an empty pair of brackets is used to denote the history ‘no history’. If $a_{t,i}$ is the action played by player i at time t , then these histories are:

$$h_1 = ()$$

$$h_t = ((a_{1,1}, a_{1,2}), \dots, (a_{t-1,1}, a_{t-1,2})), \quad t = 2, 3, \dots$$

Sometimes we will also write $(h_t, (a_{t,1}, a_{t,2}))$ for a history h_{t+1} . The set of possible histories at time t is:

$$H_1 = \{h_1\}$$

$$H_t = \prod_{i=1}^{t-1} (A \times A) \quad t = 2, 3, \dots$$

and the set of all possible histories is:

$$H = \bigcup_{t=1}^{\infty} H_t.$$

A pure strategy is a function that maps histories to the action space; $S : H \rightarrow A$. For two strategies, say S and T , the course of actions is determined by recursion; all actions at all stages are determined by the initiation

$$h_1^{S,T} = (), h_1^{T,S} = ()$$

and the recursion step for $t = 1, 2, \dots$

$$a_t^{S,T} = \left(S \left(h_t^{S,T} \right), T \left(h_t^{T,S} \right) \right), a_t^{T,S} = \left(T \left(h_t^{T,S} \right), S \left(h_t^{S,T} \right) \right)$$

$$h_{t+1}^{S,T} = \left(h_t^{S,T}, a_t^{S,T} \right), h_{t+1}^{T,S} = \left(h_t^{T,S}, a_t^{T,S} \right)$$

For $0 \leq \delta < 1$ the discounted, normalized payoffs to (a player that uses) strategy S against strategy T is given by:

$$\Pi(S, T) = (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} \pi \left(a_t^{S,T} \right)$$

3. Stepping stones in either direction

With these definitions, we can prove the first theorem. Note that strategies here are pure, and that we write that S is an equilibrium strategy, which is short for (S, S) being a symmetric Nash equilibrium of the game $\Gamma(\delta)$. We do not imply or require that the equilibrium is subgame perfect.

Theorem 7. *Let S be a strategy in the game $\Gamma(\delta)$ and let there be a time τ at which $a_{\tau}^{S,S}$ is not an equilibrium of the stage game. Then S is not RAI.*

Proof. Assume that S is an equilibrium (if it is not, it is trivially not RAI). Let T be the strategy that equals S for all histories, except for those that are elements of the set $\widehat{H} = \{h_t \mid t > \tau, a_{\tau,2} \in \arg \max_{a \in A} \pi(a, S(h_t^{S,S}))\}$. These histories only occur off the equilibrium path, since it is assumed that players playing S against each other do not play an equilibrium of the stage game at time τ . For those histories $h_t \in \widehat{H}$ we take $T(h_t) = S(h_t^{S,S})$. Obviously, the paths of T against S , T against T , S against S and S against T are all the same; $h_t^{T,S} = h_t^{T,T} = h_t^{S,S} = h_t^{S,T} \forall t$. Consequently the corresponding payoffs are also equal; $\Pi(T, S) = \Pi(S, S) = \Pi(T, T) = \Pi(S, T)$.

Now let U be the strategy that equals S , except after $h_\tau^{S,S}$, for which we take $U(h_\tau^{S,S}) \in \arg \max_{a \in A} \pi(a, S(h_\tau^{S,S}))$, and except after histories that are elements of the set $\tilde{H} = \{h_t \mid t > \tau, a_{\tau,1} \in \arg \max_{a \in A} \pi(a, S(h_\tau^{S,S}))\}$, for which we take $U(h_t) = S(h_t^{S,S}), h_t \in \tilde{H}$.

It is obvious that $\Pi(U, S) \leq \Pi(S, S)$, for S is an equilibrium, and it is also clear that $\Pi(U, T) > \Pi(T, T) = \Pi(S, T)$, because U improves itself at time τ without being punished by T . As $\Pi(U, T) > \Pi(T, T)$, while $\Pi(S, S) = \Pi(T, S)$ and $\Pi(S, T) = \Pi(T, T)$, S is not RAI.

Note that $\Pi(U, T) > \Pi(T, T) = \Pi(S, S) \geq \Pi(U, S)$, and therefore that $T \neq S$. In other words, if $T = S$, then U does strictly better against S than S itself and that contradicts S being an equilibrium. \square

This theorem indicates that as soon as there are equilibrium actions that must be upheld by the threat of punishment, then there are mutants that do not punish, and subsequently there are other mutants that take advantage of the first mutant not punishing.

The proof constructs only one way out of equilibrium. While this particular stepping stone path changes behavior for histories that are elements of rather moderate sets \hat{H} and \tilde{H} , other ways out of equilibrium may come with changes on larger, and maybe even more natural sets of histories, as for instance the example in Fig. 1 shows. But what the theorem shows is that if there is cooperation in equilibrium, at least the existence of an indirect way out is guaranteed.

While the reference points in Theorem 7 are the equilibria of the one-shot game, we will now focus on departures from the other extreme: the maximal feasible symmetric payoffs. Therefore we define $\pi_{\max} = \max_{a \in A} \pi(a, a)$ and $a_{\max} \in \arg \max_{a \in A} \pi(a, a)$. Note that a_{\max} is an action, while $a_\tau^{S,S} = (S(h_\tau^{S,S}), S(h_\tau^{S,S}))$ is an action profile. The following theorem states that if there is a point in the course of play of an equilibrium strategy at which the costs of unilaterally initiating cooperation could be offset by future gains from (increased) cooperation, then the strategy is not RAI. The proof uses a “secret handshake” argument, as introduced by Robson (1990). Because there are no separate signals here, the handshake will have to consist of a move in the repeated game itself (see also Fudenberg and Maskin, 1990). This implies that one single mutant, that both has the signal, and reacts to it by increased cooperation, can typically not invade equilibria. What is needed is a sequence of two mutations, where the first one develops the required reaction to the handshake, and the second one develops the handshake itself (see also Fig. 3).

Theorem 8. *Let S be a strategy in the game $\Gamma(\delta)$ and let there be a time τ , for which the following hold:*

1. $\pi(a_\tau^{S,S}) - \pi(a_{\max}, S(h_\tau^{S,S})) < \sum_{t=\tau+1}^{\infty} \delta^{t-\tau} (\pi_{\max} - \pi(a_\tau^{S,S}))$.
2. $a_{\max} \neq S(h_\tau^{S,S})$

Then S is not RAI.

Proof. Assume that S is an equilibrium (if it is not, it is trivially not RAI). Let T be the strategy that equals S for all histories, except for those that are elements of the set $\hat{H} = \{h_t \mid t > \tau, a_{u,2} = a_{\max}, u \geq \tau\}$. These histories only occur off the equilibrium path, as it is assumed that $a_{\max} \neq S(h_\tau^{S,S})$. For those histories $h_t \in \hat{H}$ we take $T(h_t) = a_{\max}$. Obviously, the paths of T against S , T against T , S against S and S against T are all the

same; $h_t^{T,S} = h_t^{T,T} = h_t^{S,S} = h_t^{S,T} \forall t$. Consequently the corresponding payoffs are also equal; $\Pi(T, S) = \Pi(S, S) = \Pi(T, T) = \Pi(S, T)$.

Now let U be the strategy that equals S , except after the history $h_\tau^{S,S}$, for which we choose $U(h_\tau^{S,S}) = a_{\max}$ and except after histories that are elements of the set $\tilde{H} = \{h_t \mid t > \tau, a_{\tau,1} = a_{\max} \text{ and } a_{u,2} = a_{\max}, u > \tau\}$, for which we take $U(h_t) = a_{\max}, h_t \in \tilde{H}$.

It is obvious that $\Pi(U, S) \leq \Pi(S, S)$, for S is an equilibrium, and it is also clear that $\Pi(U, T) > \Pi(T, T) = \Pi(S, T)$, because that follows directly from the first requirement of the theorem. As $\Pi(U, T) > \Pi(T, T)$, while $\Pi(S, S) = \Pi(T, S)$ and $\Pi(S, T) = \Pi(T, T)$, S is not RAI.

As in the proof of [Theorem 7](#), S being an equilibrium implies that $T \neq S$. \square

The requirements in this theorem are slightly less simple to check for than those in [Theorem 7](#), but when translated to prisoner’s dilemmas, it turns out to imply something that is relatively easy to handle. The next corollary states that for repeated prisoner’s dilemmas, all symmetric equilibria with payoffs less than $\pi(C, C) - (1 - \delta)[\pi(C, C) - \pi(C, D)]$ are not RAI. It therefore implies that for any strategy S with discounted, normalized payoffs $\Pi(S, S)$ smaller than the efficient symmetric payoff $\pi(C, C)$, there is a $\bar{\delta} \in (0, 1)$ such that S is indirectly invadable for all $\delta \in (\bar{\delta}, 1)$.

Corollary 9. *In a repeated prisoner’s dilemma, all strategies S with $\Pi(S, S) < \pi(C, C) - (1 - \delta)[\pi(C, C) - \pi(C, D)]$ are not RAI.*

Proof. First realize that S is $\pi(C, C) - \Pi(S, S)$ short from full, symmetric efficiency. Then choose as time τ in [Theorem 8](#) the first period that S plays defect. The second requirement of the same theorem is then automatically fulfilled.

The following can then be derived

$$\begin{aligned} \Pi(S, S) < \pi(C, C) - (1 - \delta)(\pi(C, C) - \pi(C, D)) &\Rightarrow \\ \Pi(S, S) < \pi(C, C) - (1 - \delta)\delta^\tau(\pi(C, C) - \pi(C, D)) &\Leftrightarrow \\ \pi(D, D) - \pi(C, D) < \frac{1}{(1 - \delta)\delta^\tau} [\pi(C, C) - \Pi(S, S)] - [\pi(C, C) - \pi(D, D)] &\Leftrightarrow \\ \pi(a_\tau^{S,S}) - \pi(a_{\max}, S(h_{\tau-1}^{S,S})) < \sum_{t=\tau}^\infty \delta^{t-\tau} (\pi_{\max} - \pi(a_t^{S,S})) - (\pi_{\max} - \pi(a_\tau^{S,S})) \end{aligned}$$

This satisfies the first requirement of [Theorem 8](#). \square

Again, the proof of [Theorem 8](#) only gives one stepping stone route out of equilibrium, but there may be lots of ways in which successive mutants can throw an equilibrium off balance and increase the level of cooperation.

Together, [Theorems 7 and 8](#) imply that in games with a conflict between individual and collective interests, like the prisoner’s dilemma, there is no strategy that is RAI, provided that δ is sufficiently high. For many equilibria there will both be stepping stone paths out with increasing and with decreasing cooperation. If the δ is not sufficiently high, and there is a strategy that is RAI, then [Theorem 7](#) still implies that this strategy cannot feature any cooperation. For the repeated prisoner’s dilemma that implies that if there is a strategy that is RAI, it can only be AIID.

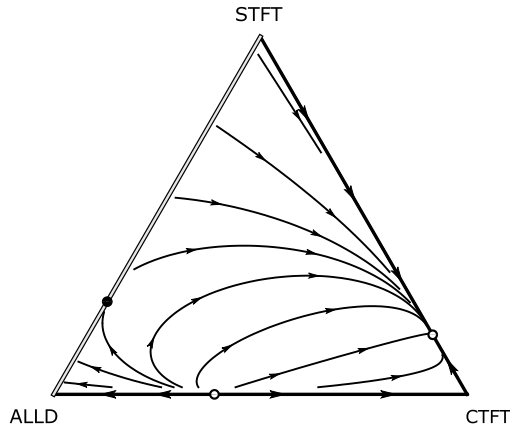


Fig. 3. An indirect invasion with increasing cooperation. The simplex shows the replicator dynamics in the normal form game generated by *AllD*, *Suspicious TitForTat (STFT)*, and *Cooperate TitForTat (CTFT)*, for $\delta = \frac{3}{4}$. *STFT* only differs from regular *TFT* in that its initial state is the defecting one, and it is a neutral mutant to *AllD*. *CTFT* cooperates on the first *two* moves and then imitates the opponent’s previous action, and it can directly invade *STFT*.

3.1. Mixed strategies

In evolutionary as well as in standard game theory, equilibrium concepts usually allow for mixed strategies. While the standard setting of symmetric 2-person bi-matrix games (see Weibull, 1995) naturally comes with definitions in terms of mixed strategies, the literature on repeated games is much more focused on pure equilibria (with exceptions such as for instance Section 8 and 9 in Binmore and Samuelson, 1992). It seems, however, that it would be natural to include mixed strategies here too, especially since the paths out of equilibrium at least at first lead away from pure strategies (or homogeneous populations) and into mixtures of strategies. While Theorem 5 shows that there is also no mixed ESS, Theorems 7 and 8 do not yet exclude the possibility that there is a mixture of strategies that is robust against indirect invasions. In this subsection we therefore give the equivalents of those theorems for finite mixtures. Here we will directly focus on repeated prisoner’s dilemmas rather than repeated games in general. This will keep notation simpler, it hopefully helps the intuition, and still captures the essentials. Also, $\Pi(D, D)$ will be used to denote $(1 - \delta) \sum_{t=0}^{\infty} \delta^{t-1} \pi(D, D) = \pi(D, D)$, which is the normalized discounted payoff of *AllD* against *AllD*.

Theorem 10. *Let P be a finite mixture of strategies in the repeated prisoner’s dilemma. If $\Pi(P, P) > \Pi(D, D)$ then P is not RAIL.*

Proof. See Appendix A. \square

As with the pure strategy version, the proof in the appendix just constructs one particular way out of equilibrium, while there may be many other stepping stone paths.

In order to formulate the mixed strategy counterpart for increasing cooperation, it will be helpful to have definitions that allow us to group together pure strategies that, up to a given time t , behave identically when interacting with a given pure strategy from the same mix. Let $P_1, \dots, P_n \in \mathcal{S}$ be the composing pure strategies of P , and let p_1, \dots, p_n , with $\sum_{i=1}^n p_i = 1$, be the

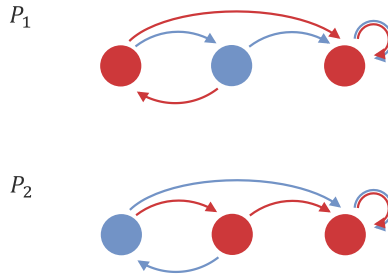


Fig. 4. An example of a mixed equilibrium. Blue stands for cooperation, red for defection. All automata start in the leftmost state.

probabilities with which they are played in P . We first define $E_i(j)_t = \{P_l \mid h_t^{P_i, P_l} = h_t^{P_i, P_j}\}$, which makes it the set of strategies against which the history of P_i at time t is the same as against P_j . Since we assume that P is a finite mixture, we know that $\lim_{t \rightarrow \infty} E_i(j)_t = E_i(j)$, where $E_i(j)$ is defined as $E_i(j) = \{P_l \mid a_t^{P_i, P_l} = a_t^{P_i, P_j} \forall t\}$ (see also the proof of Theorem 10).

For any defection by P_i that occurs along a path of interaction between P_i and P_j , we can discount P_i 's possible gains in the future, and compare them to the current period loss of switching from D to C , which could be used to initiate mutual full cooperation ever after. If strategy P_i defects at time t , then its gains and losses are to be averaged over the pure strategies that, up to time t , interact with P_i in the same way that P_j does, weighted with their frequencies. The proof of Theorem 11 will use a neutral mutant in which all pure strategies in $E_i(j)_t$ are replaced by versions that allow for a second mutant, in which P_i is replaced by a version that induces full mutual cooperation ever after with all of the new versions of the strategies in $E_i(j)_t$. This implies that possible gains for P_i in interactions with all pure strategies in $E_i(j)_t$ are relevant. If the possible gains for P_i of mutual full cooperation outweigh the cost of signaling, then P is not RAI.

In order to guarantee that the switch from D to C induces an off equilibrium path, the theorem moreover requires that all strategies that up to time t interact with P_j in the same way that P_i does, play D at time t .

Theorem 11. *Let P be a finite mixture of strategies in the repeated prisoner's dilemma, and let there be pure composing strategies P_i and P_j and a time t for which the following hold:*

1.
$$\sum_{P_l \in E_i(j)_t} p_l \left(\pi \left(a_t^{P_i, P_l} \right) - \pi \left(C, a_{t,2}^{P_i, P_l} \right) \right) < \sum_{P_l \in E_i(j)_t} p_l \sum_{u=t+1}^{\infty} \delta^{u-t} \left(\pi(C, C) - \pi \left(a_u^{P_i, P_l} \right) \right).$$
2. $a_{t,1}^{P_k, P_j} = D$ for all k for which $h_t^{P_k, P_j} = h_t^{P_i, P_j}$.

Then P is not RAI

Proof. See Appendix A. \square

An example might help illustrate different possible indirect invasions into one mixed strategy. Suppose we have a mix with equilibrium proportions of P_1 and P_2 from Fig. 4. In a prisoner's dilemma with $\pi(D, D) = 1$, $\pi(C, D) = 0$, $\pi(D, C) = 3$ and $\pi(C, C) = 2$, their normalized

payoffs against themselves and each other are $\Pi(P_1, P_1) = 1$, $\Pi(P_1, P_2) = \frac{3}{1+\delta}$, $\Pi(P_2, P_1) = \frac{3\delta}{1+\delta}$ and $\Pi(P_2, P_2) = 2 - \delta$. For $\delta > \frac{1}{2}$, the equilibrium frequency of P_1 is $\frac{1-\delta+\delta^2}{\delta+\delta^2}$. This is an equilibrium, because, conditional on having D as a first move, P_1 is the best a strategy can do against either pure strategy, and conditional on having C as a first move, P_2 is the best a strategy can do against either pure strategy.

The first possible indirect invasion goes with taking $i = 1$ and $j = 1$ in [Theorem 11](#). Suppose we replace all P_1 -players with players that play a mutated version of P_1 . The mutated version behaves just like P_1 , except after all histories that start with $k > 0$ rounds of (D, D) – like any history between two P_1 players – but then, from time $k + 1$ onwards, feature action profiles in which the other player plays C . After all of those histories, the mutated P_1 plays C . Since those histories do not occur between any two strategies in the mix, a new mix, in which P_1 is replaced with its mutated version, is a neutral mutant of the original mix.

Now a second version of P_1 arises, which behaves just like P_1 , except after all histories that start with k rounds of (D, D) . After those, the second mutated version of P_1 plays C . When matched with each other, the first and the second mutated versions of P_1 cooperate forever after round $k + 1$. The second mutated version still behaves the same against P_2 as P_1 did, and as its first mutated version did. When interacting with the first mutated version of P_1 , it gets higher discounted payoffs than P_1 or the first mutated version of P_1 did, provided that getting 2 instead of 1 forever after is worth getting 0 instead of 1 in round $k + 1$. The δ for which that starts being the case is $\frac{1}{2}$, and that implies that, considering P_1 interacting with P_1 , this equilibrium is not robust against these particular indirect invasions for all $t = k + 1 > 1$, and $\delta > \frac{1}{2}$.

Between P_2 and P_2 the situation is similar, with indirect invasions that are possible for all $t > 1$ as soon as $\delta > \frac{1}{2}$. Also between P_1 and P_2 increases in cooperation are possible. If we take $i = 1$ and $j = 2$, then a first mutant version of P_2 could behave just like P_2 , except after histories that start with $2k > 1$ rounds alternating between (C, D) and (D, C) – like any history between a P_1 and a P_2 player, from the perspective of the P_2 player – but then, from time $2k + 1$ onwards, feature the action profile (C, C) . For all of those histories, the mutated P_2 plays C . Since those histories do not occur between any two strategies in the mix, a new mix, in which P_2 is replaced with its mutated version, is a neutral mutant of the original mix.

Now a mutated version of P_1 arises, which behaves just like P_1 , except after all histories that start with $2k > 1$ rounds alternating between (D, C) and (C, D) . For those, the mutated version of P_1 plays C . When matched with each other, the mutated versions of P_1 and P_2 cooperate forever after round $2k$. The mutated version of P_1 still behaves the same against P_1 as P_1 did. When interacting with the mutated version of P_2 , it gets higher discounted payoffs than P_2 did, provided that getting 2 forever after – instead of getting 3 in odd rounds, and 0 in even ones – is worth getting 2 instead of 3 in round $2k + 1$. The δ for which that starts being the case is $\frac{1}{2}$, and that implies that, considering P_1 interacting with P_2 , this equilibrium is not robust against these particular indirect invasions for all odd $t > 2$, and $\delta > \frac{1}{2}$. For $i = 2$ and $j = 1$, similar indirect invasions are possible for all even $t > 1$, and $\delta > \frac{1}{2}$.

These indirect invasions are not exactly the same as the ones used in the proof of [Theorem 11](#); they are slightly simpler. They do however give the same threshold δ 's for which these indirect invasions are possible, because the subsequent two mutants induce full mutual cooperation. With more exploitable first mutants, which switch to unconditional cooperation after the handshake, one can also construct indirect invasions for lower δ 's. With such indirect invasions, the mutant P_i and P_j might also end up cooperating less with each other than the original P_i and P_j did, and

such indirect invasions can therefore not categorically be classified as increasing or decreasing cooperation.

3.1.1. Behavior strategies and finite mixtures

The literature on repeated games regularly works with the convenient and intuitive notion of behavior strategies. Our results can also be phrased in terms of behavior strategies, but there are a few reasons why in this context using mixed strategies might allow for more concise formulation of the results. One reason is that our results are about finite mixtures. That implies that if we were to rephrase the current results in terms of behavior strategies, we would have to restrict attention to those that only mix at a finite number of information sets. This would make the proofs a bit more roundabout than they currently are. To link up with stability results concerning RAI and ES-sets, we would moreover need to include auxiliary results, showing that a mixed strategy that is RAI corresponds to a behavior strategy that is RAI, and vice versa, and that a mixed strategy that is ESS corresponds with a behavior strategy that is ESS, but not vice versa.⁴

The reason why Selten (1983) uses behavioral strategies is that it helps get rid of spurious duplications in the description of strategic possibilities by mixed strategies, which create spurious neutral mutants. In our setting we do not have nature making chance moves, and we are less focused on stage games with mixed equilibria. That implies that it might require some ingenuity to construct examples where using behavior strategies actually does get rid of spurious mutants in our setting too. But even if such examples exist, it remains the case that our results are negative, and show that strategies are *not* RAI. While doing away with “spurious neutral mutants” can turn a non-ESS in an ESS – which is why Selten uses behavioral strategies – it cannot turn a strategy that is not RAI into a strategy that is RAI, as the neutral mutants that make it not RAI are clearly not spurious, because they open doors for other mutants (see also Selten, 1983, pp. 294–297).

Almost all of the literature focuses on pure strategies (see for instance Axelrod and Hamilton, 1981; Selten and Hammerstein, 1984; Boyd and Lorberbaum, 1987; Bendor and Swistak, 1995; Fudenberg and Maskin, 1990; Binmore and Samuelson, 1992; Cooper, 1996).⁵ Finite mixtures therefore are a step in the right direction. It would of course be interesting to know if similar results also hold for infinite mixtures, as the proofs for finite mixtures do not seem to allow for straightforward extension to infinite mixtures.

4. Indirect invasions in the simulations

In Section 3 it was already mentioned that the proofs only provide two stepping stone paths out of equilibrium; one with increasing and one with decreasing cooperation. This is enough to show that an equilibrium is not RAI. Still, since there is an uncountably infinite number of strategies,⁶ one could imagine that the existence of only one or two stepping stone paths out would not necessarily make an equilibrium very unstable in reasonable dynamics. We do however know that the paths constructed in the proofs are not the only paths out, and that there are in fact many

⁴ It might have been good had such results be included in van Veelen, 2012, because for some applications to finite extensive form games, there is a real difference.

⁵ Again, with the exception of Sections 8 and 9 in Binmore and Samuelson (1992).

⁶ Since a strategy $S: H \rightarrow A$ is a function that maps the set of histories H on the action space A , the set of strategies S is at least as large as the power set of H if the number of actions in A is larger than 1. Since H is countably infinite, we know from Cantor's Theorem that the power set of H is uncountably infinite.

quite similar ways out of equilibrium. Unfortunately, with an uncountably infinite strategy space, there is no way to determine how many paths out would be enough to be able to say with some confidence that a mutation process will actually find them. More precisely, even if the number of paths out of equilibrium would also be uncountably infinite for every equilibrium, a specific mutation process could still imply that when the population is at an equilibrium, the mutations needed for an indirect invasion occur with probability 0, while on the other hand, even if there would be only one path out of each equilibrium, a specific mutation process could imply that it occurs with positive probability. What matters therefore is the combination of a mutation process and the possible indirect invasions. (A similar point in a different context was made by [Bergin and Lipman, 1996](#).) In order to be able to say if these indirect invasions do indeed drive the evolutionary dynamics in relevant, interesting settings, we will therefore have to combine the game with a priori reasonable mutation processes and look at the dynamics. For our simulations we therefore start out with a natural choice for a mutation process, given that we do not want to exclude any part of the strategy space beforehand by construction of the mutation process.

4.1. Finite state automata

Existing simulations typically use a fixed, finite strategy space.⁷ That however is exactly what we want to avoid, given that it is the richness of the strategy space that allows for the existence of stepping stone paths out of every equilibrium. Therefore we chose to represent strategies as finite state automata,⁸ and combined it with a mutation mechanism that guarantees that every finite automaton can be reached by a finite sequence of mutations from any other automaton. In our setup, mutations add new strategies all the time, while selection may remove strategies from the population.

A finite state automaton is a list of states, and for every state it prescribes what the automaton plays when in that state, to which state it goes if the opponent plays cooperate, and to which state it goes if the opponent plays defect. More formally, a finite automaton, or a Moore machine, can then be represented by a tuple $\{\{1, \dots, N_S\}, \lambda_S, \mu_S\}$, where N_S is the number of states, $\lambda_S : \{1, \dots, N_S\} \rightarrow \{C, D\}$ gives the output in every state, and $\mu_S : \{1, \dots, N_S\} \times \{C, D\} \rightarrow \{1, \dots, N_S\}$ gives the transitions (see for instance [Rubinstein, 1986](#); [Abreu and Rubinstein, 1988](#); [Kalai and Stanford, 1988](#)).

The previous sections consider an unrestricted strategy space. Because not all strategies are finite state automata, it is worth pointing out that running simulations with finite state automata

⁷ To our knowledge, [Lindgren and Nordahl \(1994\)](#) is the only exception. Although their space is potentially infinite, what they are able to say based on theory is restricted to automata of size one and two. The (small) part of the paper that deals with a potentially infinite space is mostly speculative. Also they study spatial structure, and not the setting of a well-mixed population. The strategy spaces in [Ho \(1996\)](#) and [Miller \(1996\)](#) are finite, but extremely large. An interesting feature of their reproduction stage is that it is sexual, in the sense that the making of every offspring involves two parents. This allows for recombination. Therefore they also have a mutation procedure that implies that it depends on the current population which automata are likely to mutate into the population. [Linster \(1992, 1994\)](#) restricts attention to automata with two states. [Hirshleifer and Coll \(1988\)](#), [Young and Foster \(1991\)](#), and [Imhof et al. \(2005\)](#) consider three strategies; ALLD, ALLC and Tit-for-Tat.

⁸ The program also has the options to represent strategies with regular expressions, or to let Turing machines evolve. The set of regular expressions is equivalent to the set of finite automata, but because they are represented differently, the likelihoods of mutations also are different; a mutation that is a single step in one representation requires a series of steps in the other and vice versa. This is discussed in more detail on www.evolutionandgames.com, which also has the simulation program on it. The set of Turing machines is a richer set of strategies that embraces the set of finite automata.

still leaves us with a relatively rich set of strategies. We would like to do that by showing that when we define a relatively natural metric on the (unrestricted) strategy set \mathcal{S} , the set of finite state automata is dense in \mathcal{S} .

We would like to have a distance that reflects how differently, or similarly, two strategies play. We therefore will use a function that, for any given history, returns 0 if the two strategies play the same action, and 1 if they play a different one. Let $f : H \times \mathcal{S} \times \mathcal{S} \rightarrow \{0, 1\}$ be defined by

$$f(h_t, S, T) = \begin{cases} 0 & \text{if } S(h_t) = T(h_t) \\ 1 & \text{if } S(h_t) \neq T(h_t) \end{cases}$$

We assume that the action space A is finite, and that it has k elements, a_1, \dots, a_k . The number of possible histories in H_t – the set of all histories at time t – is k^{2t-2} . For any given t , we will give all histories equal weights. Because a history at time t is reached at all with probability δ^{t-1} , the set H_t as a whole is weighted with that probability. Therefore we define the distance between S and T , both $S, T \in \mathcal{S}$, as follows:

$$d(S, T) = (1 - \delta) \sum_{t=1}^{\infty} \rho^{t-1} \sum_{h_t \in H_t} |f(h_t, S, T)| \quad \text{with } \rho = \frac{\delta}{k^2} \text{ and } \delta \in (0, 1).$$

With this definition, $d(S, S) = 0$, and, if we take the example of the repeated prisoner’s dilemma, $d(AllC, AllD) = 1$, which implies that both are independent of δ . Distances between many pairs of strategies however will depend on δ .

If we take for $\mathcal{S}_t \subset \mathcal{S}$ the set of strategies in \mathcal{S} that all play a_1 for all histories h_u with $u > t$, then it is a finite set; it has $k^{(\sum_{v=1}^t k^{2v-2})} = k^{\binom{k^{2t}-1}{k^2-1}}$ elements. The set $\bigcup_{t=1}^{\infty} \mathcal{S}_t$ is therefore countable, but it is easy to see that it is dense in \mathcal{S} . If we now restrict attention to the repeated prisoner’s dilemma, and consider the set of all finite state automata, then this set is also countable, and dense in \mathcal{S} .

In the simulations, four kinds of mutations can occur. Mutations can add a state, delete a state, change the action that is played when in a state, or change which state to go to, depending on the action of the other player. If a state is added, one transition is randomly selected to then become a transition to this state. If a state is deleted, then all transitions towards that state are one by one randomly reassigned to another state, where each state is equally likely to be chosen. The probabilities of the four different types of mutations are of the same order, and chosen such that the average size of the automata does not explode, as it would if we choose the probability of adding a state to be too large compared to the probability of deleting a state. Besides the effect on the average size of automata, simulation results are not sensitive to even considerable changes in these probabilities.

The most important property of these four types of mutations together is that now every finite state automaton can be reached from every other finite state automaton in a finite sequence of mutations. That implies that we can always get arbitrarily close to any strategy through a finite sequence of mutations. This we think is an attractive property of a mutation scheme, given that we want to avoid excluding specific parts of the strategy space a priori.

In the simulations we will also include infinitely repeated games without discounting. Since the simulations are restricted to finite state automata, the simple limit of means is always well-defined there:

$$\Pi(S, T) = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \sum_{t=1}^{\tau} \pi(a_t^{S,T})$$

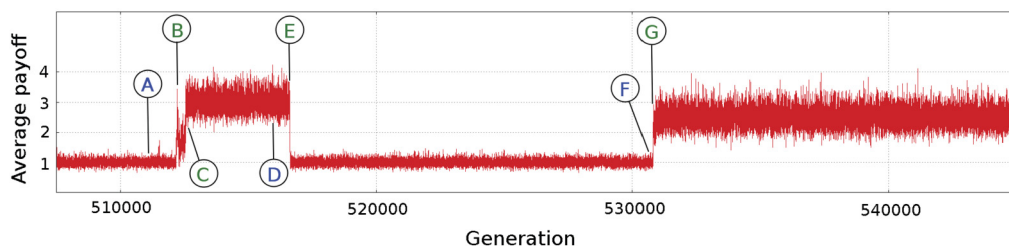


Fig. 5. Part of a typical run, with $\delta = 0.75$, and game payoffs $P = 1$, $R = 3$, $S = 0$ and $T = 4$. Blue letters indicate where relevant neutral mutants occur, green letters indicate advantageous mutants entering. The sequence of strategies is given in Fig. 6, their payoffs are given in Fig. 7.

4.2. The simulations

The basics of the simulations are simple. There are N individuals, and every generation they are randomly matched in pairs to play a repeated prisoner's dilemma. N is even, so every individual plays exactly once. With a probability of breakdown smaller than 1, the number of repetitions is a random variable. This, and the randomness of the matching, creates noise in the payoffs of individuals. These payoffs are used in the update step. In the update step, all individuals in the new generation are drawn one by one, and independently, from a distribution where the probability of being the offspring of individual j from the old generation is proportional to the payoff of j . After the new generation has been drawn, any individual mutates with a small probability. This completes the cycle for a generation. The cycle is repeated a large number of times.⁹

Fig. 5 depicts average payoffs in a short piece of a representative simulation run. The shifts in average payoffs suggest that the population might be walking from one equilibrium to the other. If we look at the strategies at different moments in time, this is exactly what we find. Fig. 6 depicts the strategies that enter the population at the points indicated by letters in Fig. 5. Fig. 7 gives the relevant payoffs in the payoff matrix between those strategies, in the order in which they appear in the simulation. Most of the payoffs between strategies that never live together in this part of the simulation run are left out, and the discount factor used for computations is the same as in the simulation.

In this part of the run, the population visits four equilibria. In the beginning, all individuals play AllD. The first indirect invasion (a neutral mutant followed by a mutant with a selective advantage) brings the population to a mixture of the two mutants. This mixture would be an equilibrium if the strategy space were to be restricted to these two strategies, but not for the unrestricted strategy space; the mixture is outperformed by a third mutant that appears at marker C. This mutant dominates the mutant that enters at marker B, and once that strategy has disappeared, it dominates the mutant that entered at A, and goes to fixation. This establishes full cooperation. After this we get an indirect invasion back to AllD, and finally we get an indirect invasion that establishes a strategy that, when played against itself, starts with defection, and then plays cooperate ever after.

⁹ A process in which the update step uses the payoffs to draw an entire new generation is called a Wright–Fisher process. The best known alternative is a Moran process, which only replaces one individual every cycle. Our program does have the option to run as a Moran process, but because the only difference lies in the speed of selection we report results that use the Wright–Fisher process. The online material at www.evolutionandgames.com contains a more detailed description.

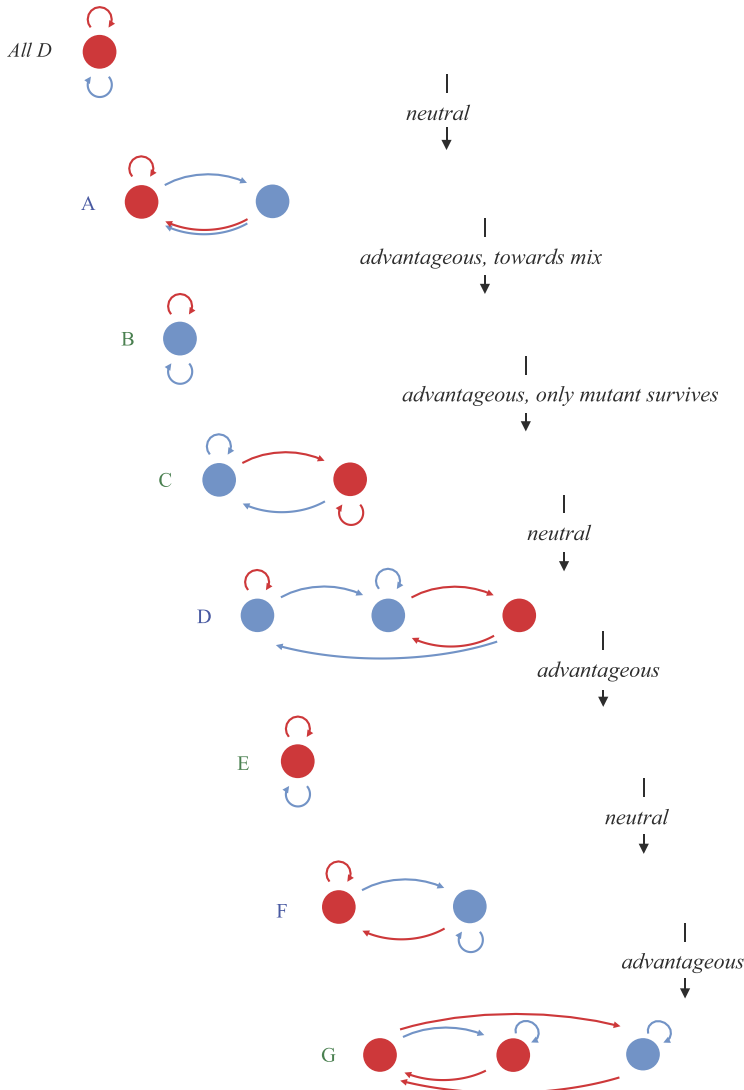


Fig. 6. The sequence of indirect invasions.

Besides looking at whether equilibria are indeed left through indirect invasions (which we will return to in the next subsection) we can also have a glimpse at the richness of the equilibria that the population visits. Fig. 8 only depicts the five strategies that are most frequently observed over the entire run, but the list is much longer (see also Fig 1. in van Veelen et al., 2012).

Since the same strategy can be produced by a variety of finite state automata, all equivalent finite state automata are lumped together with one equivalent strategy with a minimum number of states, which is depicted in the figure. For minimizing the size of automata we use the Hopcroft algorithm (Hopcroft, 1971). The strategy that is used most frequently, and by a large margin, is AllD, which can be encoded by any finite state automaton in which the output in any of its states is to defect. The simplest version of that has only one state.

	All D	A	B	C	D	E	F	G
All D	1	1	4					
A	1	1	$3\frac{4}{7}$	$2\frac{2}{7}$				
B	0	$1\frac{2}{7}$	3	3				
C		$1\frac{5}{7}$	3	3	3	$\frac{3}{4}$		
D				3	3	0		
E				$1\frac{3}{4}$	4	1	1	$1\frac{6}{7}$
F						1	1	$1\frac{109}{175}$
G						$\frac{4}{7}$	$1\frac{12}{35}$	$2\frac{1}{2}$

Fig. 7. The relevant payoffs between subsequent invaders at $\delta = 0.75$.

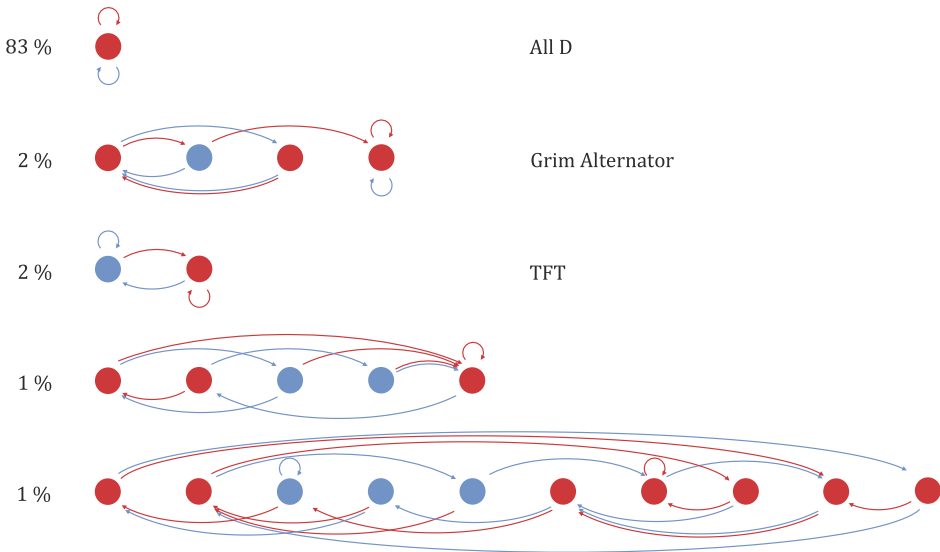


Fig. 8. The top 5 strategies for a run with $\delta = 0.75$, game payoffs $P = 1, R = 3, S = 0$ and $T = 4$, and a million generations.

Besides All D, there is a large variety of strategies that occur in considerable smaller frequencies, when averaged over the entire run. The second most frequent strategy is a strategy that, when it meets a copy of itself, plays cooperate every second round. If its opponent fails to also play cooperate in any of these odd rounds, it punishes with defecting forever. It would moreover maximally exploit All C.

The third most frequent strategy in this run is a well known one: Tit-for-Tat. The fourth most frequent strategy is part of a mixture, almost all of the time that it is there in the population. On its own, it is not an equilibrium strategy, but combined with the right counterpart, it constitutes an equilibrium mix. The counterparts that it is combined with during the run share the following

properties. When they play against a copy of themselves, they only play cooperate once, and then defect forever. When the two different automata in the mix meet, then the following path unfolds:

<i>D</i>	<i>C</i>	<i>D</i>	<i>D</i>	<i>C</i>	<i>D</i>	<i>D</i>	...
<i>C</i>	<i>D</i>	<i>C</i>	<i>C</i>	<i>D</i>	<i>C</i>	<i>C</i>	...

Both composing pure strategies do poorly against copies of themselves, but they evoke recurrent cooperation in each other, where one cooperates once every three rounds, and the other cooperates twice every block of three rounds.

The last strategy in the top 5 performs a handshake of 3 defections when it meets a copy of itself. After this handshake, it cooperates forever. A defection in round 4 or higher would make this strategy return to its initial state. In the initial state, the cheapest way for its opponent to get it back to a cooperative state is to sit through another full handshake.

4.3. Indirect invasions matter

The aim of the simulations is to find out if the presence, or absence, of indirect invasions indeed makes a substantial difference for evolutionary dynamics in repeated games. Therefore we ran simulations with different population sizes in order to see if, and how, equilibria are left. For the repeated prisoner's dilemma we first compared the number of times equilibria were left by indirect invasions to the number of times equilibria were left at all. The analysis of the output, however, is not that straightforward. The way in which the data are processed and analysed are described in some detail in [van Veelen and García \(2010\)](#) and in online material at www.evolutionandgames.com/repeatedgames.¹⁰ Here we will focus mostly on the results.

For a range of population sizes we generated simulation runs. In order to count indirect invasions as well as other ways in which equilibria are left, we need to choose which starting points to consider. As starting points of paths out of equilibrium, we took all equilibria of the repeated game that the simulation visited, and that were not themselves reached by a neutral invasion. The reason why we did not simply consider all equilibria visited by the simulation run is that it is very well possible that an equilibrium is first invaded by a neutral mutants that still is an equilibrium itself. In fact, equilibria typically are followed by a sequence of neutral mutants that have the same self-play – which, for as long as they are equilibria, is the equilibrium path. This implies that if we find a sequence of neutral mutants that themselves are equilibria, which is followed, first, by a neutral invasion to a state that is not an equilibrium, and then by an advantageous mutant, then it is reasonable to count the whole sequence as one single indirect invasion. It is for sure an indirect invasion starting from the last equilibrium, and also one when we start from the first one, and from all equilibria in between, but counting it as just one single indirect invasion is more than reasonable. All other sequences out of equilibrium – zero or more neutral mutants followed by the entry of a mutant with a disadvantage – are qualified as “other invasions”.

¹⁰ This includes a routine that finds the payoff of the best response against any pure strategy. If the payoff of S against itself equals this payoff, then S is a Nash equilibrium. This is a useful device, since the infinity of the strategy space does not allow us to determine whether or not S is a symmetric Nash equilibrium by simply comparing the payoff of S against itself to the payoff of all other strategies against S one after the other.

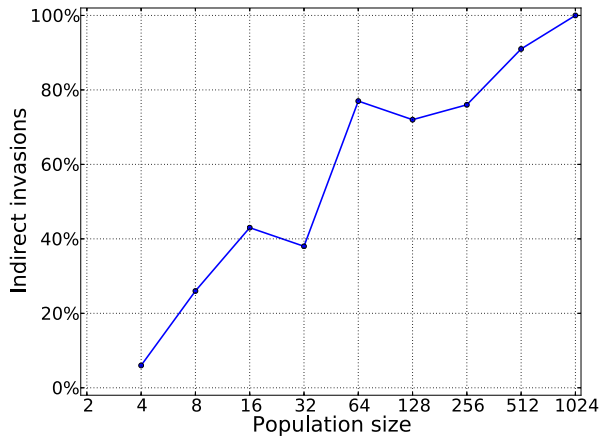


Fig. 9. The share of indirect invasions out of pure equilibria goes to 1 as the population size increases. The continuation probability is 0.75, and the game has payoffs $P = 1$, $R = 3$, $S = 0$ and $T = 4$.

If we restrict attention to pure equilibria, then the data indicate that the share of indirect invasions goes to 1 if the population size grows large. As a matter of fact, all equilibria are left through indirect invasions already for a population of 1024 (see Fig. 9).¹¹

We would also like to compare the dynamic behavior in a repeated prisoner’s dilemma, which does not have equilibria that are RAI, to a repeated game that does have equilibria that are RAI. Therefore we replace the prisoner’s dilemma as a stage game with a coordination game. If we do, then we preserve everything, including the richness of the strategy space. The only difference is that now we do get equilibria that are RAI (but not ESS). That means that what sets these repeated games apart is the existence or nonexistence of the possibility for indirect invasions. We therefore replace the prisoner’s dilemma

$$\begin{bmatrix} 3 & 0 \\ 4 & 1 \end{bmatrix}$$

with the following stage game:

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

In Fig. 10 we compare the number of transitions in simulations for the repeated prisoner’s dilemma to the number of transitions in the repeated coordination game. What counts as a transition, is if the population goes to a state where the self-play is different (all strategies with the same self-play are only neutral mutants of each other). While the number of transitions decreases for both games if the population increases, the number of transitions in the repeated coordination game divided by the number of transitions in the repeated prisoner’s dilemma goes to 0 rather rapidly. Transitions in the repeated coordination game are already not observed anymore for

¹¹ For combinations of two or more finite automata it is harder to get such clean data. Making an automated procedure to determine whether or not mixtures of automata are equilibria is far more complicated than constructing one for pure strategies and thereby well beyond the scope of this paper. For not too large population sizes, mixed population states are moreover left relatively easily by one of the composing strategies going extinct, leaving the population in a disequilibrium state, which then gets invaded, either by the strategy that went extinct just before, or by another strategy.

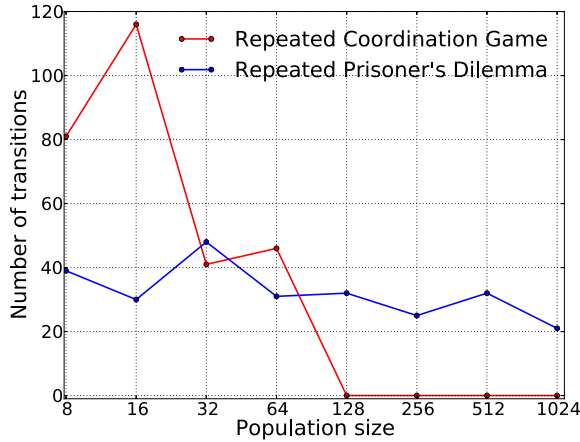


Fig. 10. The number of transitions leaving pure equilibria in a repeated prisoner's dilemma and in a repeated coordination game for different population sizes. Note that in the repeated coordination game, all RAI equilibria are pure, and all mixed equilibria are not NSS.

population size 128, while the number of indirect invasions is still quite sizeable at 1024 in the repeated prisoner's dilemma. One can therefore conclude that a population in the latter remains infinitely more mobile than the former.

The number of transitions away from pure equilibria in the repeated prisoner's dilemma is decreasing only very slightly. This fits what we expect. The fixation probability of neutral mutants is expected to determine the rate at which indirect invasions occur in large populations; once neutral mutants have opened the door for advantageous ones, invasions typically happen relatively quickly. The fixation probability of a neutral mutant is $\frac{1}{N}$, where N is the population size (see Theorem 1.1 in Durrett, 2008). At the same time the mutation probability *per individual* is constant, so the expected number of neutral mutants entering per generation is proportional to N . If we can assume that every neutral mutant has either becomes fixed or gone extinct before the other neutral mutant appears, then the expected number of transitions by neutral mutants should be constant, as the decrease in fixation probability is compensated by an increase in numbers of neutral mutants entering. But with an increase in population size, fixation times also increase, and the larger the population, the more neutral mutants enter in a population that has not yet gone to fixation. This interference implies that we will be observing slightly less fixation events for larger populations.

The traffic between equilibria for the repeated prisoner's dilemma, where no equilibrium is RAI, therefore decreases slightly, but remains of the same order of magnitude, if the population size increases. With the repeated coordination game the transitions between equilibria quickly become orders of magnitude less frequent. It is comparisons between transitions like the latter ones that are central to the concepts of stochastically stable states (Young, 1993), or long-run equilibria (Kandori et al., 1993). The dynamics for the repeated prisoner's dilemma therefore typically feature orders of magnitude more transitions than the dynamics studied in the literature on the evolution of conventions (see also Young, 1998; Ellison, 2000).

5. Discussion

Both the theorems and the simulations in this paper indicate that there is a fundamental instability in repeated games, provided that the stage game is characterized by a conflict between

individual and collective interests. The prime example is the repeated prisoner's dilemma. [Theorems 10 and 11](#) show that with sufficiently large continuation probability δ , there is no finite mixture of strategies that is robust against indirect invasions in the repeated prisoner's dilemma. In other words: every equilibrium can be upset, either by a mutant, if the strategy is not an NSS, or by a succession of mutants, if the strategy is an NSS. The simulations show that under very reasonable mutation schemes these stepping stone paths out of equilibrium not only exist, but evolution also finds them.

The richness of the strategy space therefore excludes that there is an equilibrium refinement, or a static stability concept, that, by only looking at the game itself, singles out an equilibrium that will be significantly more stable than all others in a population with random matching, mutation and selection. One important conclusion is that what we can expect to evolve will essentially depend – besides on δ – on the structure of the mutation probabilities, or, more precisely, on which mutations are relatively likely. The proofs of the results show that there are different possible stepping stone paths out of equilibrium; paths with increasing levels of cooperation, and paths with decreasing levels of cooperation. Whether we can expect cooperation to increase or decrease therefore also depends on the probabilities with which the mutations necessary for the different paths occur.

If we allow ourselves to restrict the strategy space, then that offers a possibility to get stability results. It is however important to see that restricting the strategy space to, say, a strict subset \mathcal{T} of \mathcal{S} , is in fact a special case of a combination of a starting point (somewhere within \mathcal{T}) and an assumption concerning mutation probabilities (they are zero for all mutations from elements of \mathcal{T} to elements of $\mathcal{S} \setminus \mathcal{T}$). This therefore falls within the message that the starting point and mutation probabilities are decisive. The results here however also imply that any stability result that is achieved by restricting the strategy space is not robust to relaxations of the restriction on the strategy space. If the strategies needed for the indirect invasions are apparently barred by exclusion from the strategy space, they nonetheless still exist, and adding them to the strategy space (i.e. allowing mutations to them) would render the strategy that was stable within the restricted strategy set \mathcal{T} unstable.

Simulations show that with a reasonable mutation process, a population that is not too small does indeed walk from equilibrium to equilibrium through indirect invasions. Since the infinite population model is meant to produce results that help us understand what happens in large, but still finite populations, the simulations thereby also emphasize the importance of neutral mutants and the need to have a concept that deals with indirect invasions in our theory for infinite populations.

Acknowledgments

We would like to thank Gerard van der Laan, Tore Ellingsen, Martin Nowak, Corina Tarnita, Dave Rand, Arthur Schram, Peter Spreij, Aljaz Ule, Drew Fudenberg, Jörgen Weibull, Jonathan Bendor, Piotr Swistak, Peyton Young, and participants of the Tinbergen Institute 2010 and 2014 conferences. We would also like to thank Larry Samuelson for suggesting to present the results concerning repeated games separate from the concept of robustness against indirect invasions. Christian Hellwig, two anonymous reviewers, and especially Bill Sandholm turned this into a much better paper. [Figs. 1 and 3](#) were made with Dynamo Version 1.4.1 by Bill Sandholm and Emin Dokumaci, which is a great service to users. Both authors gratefully acknowledge financial support by the Netherlands' Organization for Scientific Research (NWO grants 051-12-012, 051-14-046, and 451-04-073) and the Max Planck Society.

Appendix A. Proofs of theorems

Proof of Proposition 5. Assume that P is a finite mixture of strategies. Let P_1, \dots, P_n denote the composing pure strategies of P and let p_1, \dots, p_n with $\sum_{i=1}^n p_i = 1$ be the probabilities with which they are played in P . It is safe to assume that P is a Nash equilibrium, as being ESS implies being a Nash equilibrium.

There can be no more than n^2 paths that are followed by combinations of two pure strategies from this mixture. There is, however, a (countably) infinite number of possible paths; if k represents the number of possible actions of each player in the stage game, then there are k^2 possible action profiles per repetition, and there is an infinite number of repetitions. (Note that a game is non-trivial if $k > 1$.) For every finite mixture of strategies, we can create a new strategy that performs exactly as well as the other strategies in the mixture. Take one of the strategies present in the mixture, say strategy n , and mutate it into strategy $n + 1$ by only changing its behavior for a history that does not occur along any of the at most n^2 paths followed by duo's of strategies from this mixture interacting. Some such changes could turn it into one of the other $n - 1$ strategies, but there is a (countably) infinite number of possible histories to chose from (see also Section 3) and only a finite number of strategies in the mixture, so there always exists one such mutant that really is a new strategy. This new strategy does not cause any changes; when paired with any of the n strategies both strategies n and $n + 1$ follow the same paths and also the path of n with itself is the same as $n + 1$ with itself. Hence $n + 1$ receives exactly the same payoff as the other strategies from the mixture and we have a mutant that is not driven out. Therefore the finite mixture is not evolutionarily stable. □

Proof of Theorem 10. Assume that P is an equilibrium (if it is not, it is trivially not robust against indirect invasions). Let $P_1, \dots, P_n \in \mathcal{S}$ be the composing pure strategies of P and let $p_1, \dots, p_n, \sum_{i=1}^n p_i = 1$, be the probabilities with which they are played in P . If $\Pi(P, P) > \Pi(D, D)$, then obviously not all combinations of P_i and P_j , with $1 \leq i, j \leq n$, can always play D when they interact. So there must be at least one i and one j , with $1 \leq i, j \leq n$, and a time τ for which $a_\tau^{P_i, P_j} \neq (D, D)$. First it is clear that there cannot only be a finite number of times that C is played in the mixture. Suppose that were true, and there is a time τ and an i and a j , with $1 \leq i, j \leq n$, for which $a_\tau^{P_i, P_j} \neq (D, D)$ and $a_t^{P_i, P_j} = (D, D)$ for all i and j and $t > \tau$, then the mixture is not an equilibrium; without restricting generality we can assume that $P_i \left(h_\tau^{P_i, P_j} \right) = C$ and then a strategy that equals P_i for all histories at times $t < \tau$ and plays D for all histories at times $t \geq \tau$ earns a higher payoff than P_i and therefore also higher than all other composing strategies. Hence C must be played infinitely many times in the mixed population. Since there is only a finite number of combinations (P_i, P_j) , it also follows that there is at least one in which P_i plays C an infinite number of times.

Let (P_i, P_j) be a combination of strategies in which P_i plays C infinitely often. Let $E(i, j)$ be the set of combinations of strategies (P_k, P_l) for which $a_t^{P_k, P_l} = a_t^{P_i, P_j} \forall t$, that is, strategies P_k and P_l that follow the same path as when P_i interacts with P_j . Given that P is a finite mixture, there is a finite time τ' which is sufficiently large to determine whether or not $(P_k, P_l) \in E(i, j)$, that is, there is a τ' such that if $a_t^{P_k, P_l} = a_t^{P_i, P_j} \forall t \leq \tau'$ then $a_t^{P_k, P_l} = a_t^{P_i, P_j} \forall t$. Let $\tau'' > \tau'$ be the first time t after τ' at which $a_{t,1}^{P_i, P_j} = C$.

Let $E_i(j)$ be the set of strategies P_l such that $(P_i, P_l) \in E(i, j)$. For all $P_l \in E_i(j)$ one can define Q_l as the pure strategy that equals P_l for all histories, except for those in the set

$$\widehat{H} = \left\{ h_t \mid t > \tau'', a_u = a_u^{P_j, P_i} \text{ for } u \leq \tau' \text{ and } a_{\tau'', 2} = D \right\}.$$

These histories only occur off all equilibrium paths, since the history up to and including τ' implies that this history does not occur along an equilibrium path outside $E(i, j)$, as experienced by j , while the remainder implies that it does not occur along equilibrium paths in $E(i, j)$. For the histories $h_t \in \widehat{H}$ we take $Q_l(h_t) = a_{t,1}^{P_j, P_i} = P_l(h_t^{P_j, P_i})$. Obviously, the path of Q_l against P_m is the same as the path of P_l against P_m for all $m, 1 \leq m \leq n$ and all $P_l \in E_i(j)$. Define Q as the strategy that plays Q_l with probability p_l for $P_l \in E_i(j)$ and P_l with probability p_l for all $P_l \notin E_i(j)$. For this strategy we have $\Pi(Q, P) = \Pi(P, P) = \Pi(Q, Q) = \Pi(P, Q)$.

Let R be the strategy that equals P_i , except for $h_{\tau''}^{P_i, P_j}$, for which we take $R(h_{\tau''}^{P_i, P_j}) = D$ and except for histories that are elements of the set

$$\widetilde{H} = \left\{ h_t \mid t > \tau'', a_u = a_u^{P_i, P_j} \text{ for } u \leq \tau' \text{ and } a_{\tau'', 1} = D \right\},$$

for which we take $R(h_t) = a_{t,1}^{P_i, P_j} = P_i(h_t^{P_i, P_j}), h_t \in \widetilde{H}$.

Because P is an equilibrium, it must be that $\Pi(R, P) \leq \Pi(P, P)$. It is also clear that $\Pi(R, Q) > \Pi(Q, Q) = \Pi(P, Q)$, because R improves itself against strategies $Q_l \in E_i(j)$ at time τ'' without being punished and remains unchanged against strategies that are not in $E_i(j)$. As $\Pi(R, Q) > \Pi(Q, Q)$, while $\Pi(P, P) = \Pi(Q, P)$ and $\Pi(P, Q) = \Pi(Q, Q)$, P is not robust against indirect invasions.

Note that if $Q_l = P_l \forall P_l \in E_i(j)$, that would contradict P being an equilibrium, because if $P = Q$ then $\Pi(R, Q) > \Pi(Q, Q)$ would contradict that $\Pi(R, P) \leq \Pi(P, P)$. \square

Proof of Theorem 11. Assume that P is an equilibrium (if it is not, it is trivially not robust against indirect invasions). Take i, j and τ such that $\delta_{ij, \tau} = \min_{k,l,t} \delta_{kl,t}$. For all $P_l \in E_i(j)_\tau$ one can define Q_l as the pure strategy that equals P_l for all histories, except for those that are elements of the set

$$\widehat{H} = \left\{ h_t \mid t > \tau, a_u = a_u^{P_j, P_i} \text{ for } u < \tau, a_{u,2} = C, u \geq \tau \right\}.$$

These histories only occur off all equilibrium paths, if indeed $\delta_{ij, \tau} < 1$. For those histories $h_t \in \widehat{H}$ we take $Q_l(h_t) = C$. Obviously, the path of Q_l against P_m is the same as the path of P_l against P_m for all $m, 1 \leq m \leq n$ and all $P_l \in E_i(j)_\tau$. Define Q as the strategy that plays Q_l with probability p_l for $P_l \in E_i(j)_\tau$ and P_l with probability p_l for all $P_l \notin E_i(j)$. Consequently the corresponding payoffs are also equal; $\Pi(Q, P) = \Pi(P, P) = \Pi(Q, Q) = \Pi(P, Q)$.

Now let R be the strategy that equals P_i , except for the history $h_\tau^{P_i, P_j}$, for which we choose $R(h_\tau^{P_i, P_j}) = C$ and except for the histories that are elements of the set

$$\widetilde{H} = \left\{ h_t \mid t > \tau, a_u = a_u^{P_i, P_j} \text{ for } u < \tau, a_{\tau,1} = C \text{ and } a_{u,2} = C, u > \tau \right\},$$

for which we also take $R(h_t) = C, h_t \in \widetilde{H}$.

Because P is an equilibrium, it must be that $\Pi(R, P) \leq \Pi(P, P)$. It is also clear that $\Pi(R, Q) > \Pi(Q, Q) = \Pi(P, Q)$, because R improves itself against strategies $Q_l \in E_i(j)$ at time τ without being punished and remains unchanged against strategies that are not in $E_i(j)_\tau$. As $\Pi(R, Q) > \Pi(Q, Q)$, while $\Pi(P, P) = \Pi(Q, P)$ and $\Pi(P, Q) = \Pi(Q, Q)$, P is not robust against indirect invasions.

Note again that if $Q_l = P_l \forall P_l \in E_i(j)_\tau$, that would contradict P being an equilibrium, because if $P = Q$ then $\Pi(R, Q) > \Pi(Q, Q)$ would contradict that $\Pi(R, P) \leq \Pi(P, P)$. \square

References

- Abreu, D., Rubinstein, A., 1988. The structure of Nash equilibrium in repeated games with finite automata. *Econometrica* 56, 1259–1281.
- Axelrod, R., 1997. *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton University Press.
- Axelrod, R., Hamilton, W.D., 1981. The evolution of cooperation. *Science* 211, 1390–1396.
- Balkenborg, D., Schlag, K.H., 2001. Evolutionarily stable sets. *Int. J. Game Theory* 29, 571–595.
- Bendor, J., Swistak, P., 1995. Types of evolutionary stability and the problem of cooperation. *Proc. Natl. Acad. Sci. USA* 92, 3596–3600.
- Bendor, J., Swistak, P., 1997. The evolutionary stability of cooperation. *Am. Polit. Sci. Rev.* 91, 290–307.
- Bendor, J., Swistak, P., 1998. Evolutionary equilibria: characterization theorems and their implications. *Theory Decis.* 45, 99–159.
- Bergin, J., Lipman, B.L., 1996. Evolution with state-dependent mutations. *Econometrica* 64, 943–956.
- Binmore, K.G., Samuelson, L., 1992. Evolutionary stability in repeated games played by finite automata. *J. Econ. Theory* 57, 278–305.
- Bomze, I.M., Weibull, J.W., 1995. Does neutral stability imply Lyapunov stability? *Games Econ. Behav.* 11, 173–192.
- Boyd, R., Lorberbaum, J.P., 1987. No pure strategy is evolutionary stable in the iterated prisoner's dilemma game. *Nature* 327, 58–59.
- Cooper, D.J., 1996. Supergames played by finite automata with finite costs of complexity in an evolutionary setting. *J. Econ. Theory* 68, 266–275.
- Cressman, R., 1996. Evolutionary stability in the finitely repeated prisoner's dilemma. *J. Econ. Theory* 68, 234–248.
- Durrett, R., 2008. *Probability Models for DNA Sequence Evolution*. Probability and Its Applications, 2nd edition. Springer, New York.
- Ellison, G., 2000. Basins of attraction, long-run stochastic stability, and the speed of step-by-step evolution. *Rev. Econ. Stud.* 67 (1), 17–45.
- Farrell, J., Ware, R., 1989. Evolutionary stability in the repeated prisoner's dilemma. *Theor. Popul. Biol.* 36, 161–166.
- Fudenberg, D., Maskin, E., 1990. Evolution and cooperation in noisy repeated games. *Am. Econ. Rev.* 80, 274–279.
- Hirshleifer, J., Coll, J., 1988. What strategies can support the evolutionary emergence of cooperation? *J. Confl. Resolut.* 32, 367–398.
- Ho, T.H., 1996. Finite automata play repeated prisoner's dilemma with information processing costs. *J. Econ. Dyn. Control* 20, 173–207.
- Hopcroft, J., 1971. An $n \log n$ algorithm for minimizing states in a finite automaton. In: *Theory of Machines and Computations*, Proc. Internat. Sympos., Technion, Haifa, 1971. Academic Press, New York, pp. 189–196.
- Imhof, L.A., Fudenberg, D., Nowak, M.A., 2005. Evolutionary cycles of cooperation and defection. *Proc. Natl. Acad. Sci. USA* 102, 10797–10800.
- Kalai, E., Stanford, W., 1988. Finite rationality and interpersonal complexity in repeated games. *Econometrica* 56, 397–410.
- Kandori, M., Mailath, G.J., Rob, R., 1993. Learning, mutation, and long run equilibria in games. *Econometrica* 61, 29–56.
- Kim, Y.G., 1994. Evolutionarily stable strategies in the repeated prisoner's dilemma. *Math. Soc. Sci.* 28, 167–197.
- Lindgren, K., Nordahl, M.G., 1994. Evolutionary dynamics of spatial games. *Physica D* 75, 292–309.
- Linster, B.G., 1992. Evolutionary stability in the infinitely repeated prisoners' dilemma played by two-state Moore machines. *South. Econ. J.* 58, 880–903.
- Linster, B.G., 1994. Stochastic evolutionary dynamics in the repeated prisoners' dilemma. *Econ. Inq.* 32, 342–357.
- May, R.M., 1987. More evolution of cooperation. *Nature* 327, 15–17.
- Maynard Smith, J., 1982. *Evolution and the Theory of Games*. Cambridge University Press.
- Maynard Smith, J., Price, G.R., 1973. The logic of animal conflict. *Nature* 246, 15–18.
- Miller, J., 1996. The coevolution of automata in the repeated prisoner's dilemma. *J. Econ. Behav. Organ.* 29, 87–112.
- Morris, S., Rob, R., Shin, H.S., 1995. p -Dominance and belief potential. *Econometrica* 63 (1), 145–157.
- Nowak, M.A., 2006. *Evolutionary Dynamics*. Harvard University Press, Cambridge, MA.
- Nowak, M.A., Sigmund, K., 1990. The evolution of stochastic strategies in the prisoner's dilemma. *Acta Appl. Math.* 20, 247–265.

- Nowak, M.A., Sigmund, K., 1992. Tit for tat in heterogeneous populations. *Nature* 355, 250–253.
- Oyama, D., Sandholm, W.H., Tercieux, O., 2015. Sampling best response dynamics and deterministic equilibrium selection. *Theor. Econ.* 10 (1), 243–281.
- Robson, A.J., 1990. Efficiency in evolutionary games: Darwin, Nash and the secret handshake. *J. Theor. Biol.* 144, 379–396.
- Rubinstein, A., 1986. Finite automata play the repeated prisoner's dilemma. *J. Econ. Theory* 39, 83–96.
- Samuelson, L., Swinkels, J.M., 2003. Evolutionary stability and lexicographic preferences. *Games Econ. Behav.* 44, 332–342.
- Sandholm, W.H., 2001. Almost global convergence to p-dominant equilibria. *Int. J. Game Theory* 30, 107–116.
- Sandholm, W.H., Staudigl, M., 2015. Large deviations and stochastic stability in the small noise double limit. *Theor. Econ.*
- Schlag, K.H., 1993. Dynamic stability in the repeated prisoners' dilemma played by finite automata. Discussion paper B-243.
- Selten, R., 1983. Evolutionary stability in extensive two-person games. *Math. Soc. Sci.* 5, 269–363.
- Selten, R., Hammerstein, P., 1984. Gaps in Harley's argument on evolutionarily stable learning rules and in the logic of "tit for tat". *Behav. Brain Sci.* 7, 115–116.
- Taylor, P.D., Jonker, L., 1978. Evolutionary stable strategies and game dynamics. *Math. Biosci.* 40, 145–156.
- Thomas, B., 1985. On evolutionarily stable sets. *J. Math. Biol.* 22, 105–115.
- Traulsen, A., Claussen, J.C., Hauert, C., 2005. Coevolutionary dynamics: from finite to infinite populations. *Phys. Rev. Lett.* 95, 238701.
- Traulsen, A., Nowak, M.A., Pacheco, J.M., 2006. Stochastic dynamics of invasion and fixation. *Phys. Rev. E* 74, 011909.
- van Veelen, M., 2012. Robustness against indirect invasions. *Games Econ. Behav.* 74, 382–393.
- van Veelen, M., García, J., 2010. In and out of equilibrium: evolution of strategies in repeated games with discounting. TI working paper 10-037/1.
- van Veelen, M., García, J., 2012. In and out of equilibrium II: evolution in repeated games with discounting and complexity costs. TI working paper 12-089/1.
- van Veelen, M., García, J., Rand, D.G., Nowak, M.A., 2012. Direct reciprocity in structured populations. *Proc. Natl. Acad. Sci. USA* 109, 9929–9934.
- van Veelen, M., Spreij, P., 2009. Evolution in games with a continuous action space. *Econ. Theory* 39, 355–376.
- Volij, O., 2002. In defense of defect. *Games Econ. Behav.* 39, 309–321.
- Weibull, J.W., 1995. *Evolutionary Game Theory*. MIT Press, Cambridge.
- Young, H.P., 1993. The evolution of conventions. *Econometrica* 61, 57–84.
- Young, H.P., 1998. *Individual Strategy and Social Structure*. Princeton University Press, Princeton.
- Young, H.P., Foster, D., 1991. Cooperation in the short and in the long run. *Games Econ. Behav.* 3, 145–156.