

van Veelen, Matthijs; Garcia, Julian

**Working Paper**

## In and Out of Equilibrium: Evolution of Strategies in Repeated Games with Discounting

Tinbergen Institute Discussion Paper, No. 10-037/1

**Provided in Cooperation with:**

Tinbergen Institute, Amsterdam and Rotterdam

*Suggested Citation:* van Veelen, Matthijs; Garcia, Julian (2010) : In and Out of Equilibrium: Evolution of Strategies in Repeated Games with Discounting, Tinbergen Institute Discussion Paper, No. 10-037/1, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:

<https://hdl.handle.net/10419/87038>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



TI 2010-037/1

Tinbergen Institute Discussion Paper

# In and out of Equilibrium: Evolution of Strategies in Repeated Games with Discounting

*Matthijs van Veelen<sup>1,3</sup>*

*Julián García<sup>2,3</sup>*

<sup>1</sup> CREED, Universiteit van Amsterdam;

<sup>2</sup> Vrije Universiteit Amsterdam;

<sup>3</sup> Tinbergen Institute.

### **Tinbergen Institute**

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

### **Tinbergen Institute Amsterdam**

Roetersstraat 31  
1018 WB Amsterdam  
The Netherlands  
Tel.: +31(0)20 551 3500  
Fax: +31(0)20 551 3555

### **Tinbergen Institute Rotterdam**

Burg. Oudlaan 50  
3062 PA Rotterdam  
The Netherlands  
Tel.: +31(0)10 408 8900  
Fax: +31(0)10 408 9031

Most TI discussion papers can be downloaded at  
<http://www.tinbergen.nl>.

# In and out of equilibrium: evolution of strategies in repeated games with discounting.

Matthijs van Veelen

CREED, Universiteit van Amsterdam,

Roetersstraat 11, 1018 WB Amsterdam, the Netherlands;

c.m.vanveelen@uva.nl

Julián García

Faculty of Economics and Business Administration, Vrije Universiteit,

De Boelelaan 1105, 1081 HV Amsterdam, the Netherlands;

j.garcia@feweb.vu.nl

April 9, 2010

## Abstract

Repeated games tend to have large sets of equilibria. We also know that in the repeated prisoners dilemma there is a profusion of neutrally stable strategies, but no strategy that is evolutionarily stable. This paper shows that for all of these neutrally stable strategies there is a stepping stone path out; there is always a neutral mutant that can enter a population and create an actual selective advantage for a second mutant. Such stepping stone paths out of equilibrium generally exist both in the direction of more and in the direction of less cooperation.

While the central theorems show that such paths out of equilibrium exist, they could still be rare compared to the size of the strategy space. Simulations however suggest that they are not too rare to be found by a reasonable mutation process, and that typical simulation paths take the population from equilibrium to equilibrium through a series of indirect invasions.

Instability does not mean we cannot draw qualitative conclusions though. The very nature of the indirect invasions implies that the population will on average be (somewhat) reciprocal and (reasonably) cooperative.

“Everything changes, nothing remains the same”

*Buddha*

## 1 Introduction: joy in repetition

Repeating a game generally opens up a variety of possibilities for equilibrium behaviour that the one-shot version does not possess. Repeated games therefore have been studied extensively; see for instance Friedman (1971), Aumann & Shapley (1976), Rubinstein (1979, 1980), Fudenberg & Maskin (1986), Abreu (1988) and van Damme (1989). The standard example under study is the prisoners dilemma, where the Nash equilibrium in the one-shot game is not Pareto-efficient and where repetition offers a possible escape from inefficiency. An interesting and natural follow up question is if evolution found an escape route too, and if it did, if it is the same escape route as the one that game theorists found. Again the literature is quite substantial, with for instance Axelrod (1984), Boyd & Lorberbaum (1987), Farrell & Ware (1989), Fudenberg & Maskin (1990), Binmore & Samuelson (1992, 1997), Bendor & Swistak (1995, 1997, 1998), Lorberbaum, Bohning, Shastri & Sine (2002) and Fudenberg, Imhof & Nowak (2005). The main problem these papers face is that in general there is no evolutionarily stable strategy in repeated games, while evolutionary stability is the main and usually also the most promising tool from the evolutionary game theory toolbox (see Weibull, 1995).

This paper examines how unfortunate that is. Helped by the careful distinctions between different definitions of stability from Bendor & Swistak (1995) and using arguments that are similar (but not identical) to those in Selten & Hammerstein (1984) and Farrell & Ware (1989) we begin with a general theorem concerning the non-existence of a finite mixture of strategies that is evolutionarily stable in the classical definition (Maynard Smith & Price, 1973, Maynard Smith, 1974). One way of dealing with such a negative result is to try out less demanding equilibrium refinements in order to overcome non-existence. Although positive results have been achieved with this approach (see Bendor & Swistak (1995, 1997, 1998) and, in slightly different settings, by Fudenberg & Maskin (1990) and Binmore & Samuelson (1992, 1997)), we will argue that there is a fundamental instability of all equilibria in interesting, non-trivial repeated games. We will do this by showing that no equilibrium is robust against indirect invasions (Van Veelen, 2010). In other words, we show that 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. It is shown that for repeated games such stepping stone paths out of equilibrium generally exist, both in the direction of more and in the direction of less cooperation - that is, if a higher resp.

lower level of cooperation is possible and, for increases in cooperation, if the probability of breakdown is small enough. This indicates that there is no population state that, once it is reached, cannot be overturned by a succession of mutants.

The proof of this central result is by construction of one such path out of equilibrium with increasing cooperation and one with decreasing cooperation. We do however also show that the whole, unrestricted strategy space for repeated games with discounting is very large; it is uncountably infinite. Therefore one could wonder whether the existence of one or two stepping stone paths out of equilibrium really is a problem one should worry about. We do however suspect that there are many similar ways out of equilibrium, and in order to show that these stepping stone paths are indeed important for the dynamics, we ran simulations. We find that 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.

There are a few conclusions to be drawn from these findings. The first is that we should be aware that for the repeated games we tend to look at when we study cooperation with repeated interactions, *instability* is just as important to focus on as *stability*. That does not imply that we can not say anything about what kind of strategies we can expect to find if we look at a repeated game being played in a population where strategies are subject to mutation and selection. Quite the contrary; we will see below that even though nothing is as stable as we would hope for, the transitions typically share some characteristics that imply that on average - the average also being taken over time - we can expect strategies to be somewhat nice as well as somewhat reciprocal. Exactly how nice and how reciprocal we can expect strategies to be on average depends on the assumptions concerning mutation probabilities. Along with the simulations we therefore developed measures for reciprocity and cooperativeness of strategies.

As restrictions on the strategy space are just a special case of a choice for mutation probabilities and a starting point, this embraces questions one could have about the development of the literature too. The central theorem also implies that if a restricted strategy space is used in the analysis of evolution of strategies in repeated games, and if we then do find an ESS, then one can always extend the strategy space so that this strategy is no longer an ESS and in fact not even weakly robust against indirect invasions. So rather than drawing conclusions from settings with a restricted strategy space, it is more important to think of assumptions concerning mutation probabilities as essential.

## 2 No ESS

The literature concerning evolutionary stability and repeated games can at first sight be a bit confusing. The reason, as Bendor & Swistak (1995) show, is that different authors have used different definitions of evolutionary stability. They also convincingly argue that

Maynard Smith's (1974) definition of an evolutionarily stable strategy (ESS), and perhaps a weaker version, that Maynard Smith (1982) calls a neutrally stable strategy (NSS), are dynamically the most interesting and meaningful ones. We will therefore adopt the more standard definition of an evolutionarily stable strategy. Here  $\mathcal{S}$  is a space of pure strategies for the repeated game and  $\Pi : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}$  is the payoff function, where  $\Pi(S, T)$  is the payoff of a player playing strategy  $S$  against a player playing strategy  $T$ . The payoff of the opponent  $T$  in this encounter is given by  $\Pi(T, S)$ , thereby assuming that the game is symmetric.

The pure strategy version of the definition - [2] in Bendor & Swistak (1995) - is as follows.

**Definition 1 (Pure ESS)** *A strategy  $S \in \mathcal{S}$  is evolutionarily stable if both*

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

*if  $\Pi(S, S) = \Pi(T, S)$  then  $\Pi(S, T) > \Pi(T, T)$  for all  $T \neq S$*

As the standard definition of an ESS also allows for mixed strategies, we would like to do the same here. We therefore equate mixed strategies with probability distributions over the pure strategy space. Section 3 and Appendix A show how  $\mathcal{S}$  can be endowed with a metric to make it a separable metric space. Strategies  $P$  and  $Q$  will then be a probability measures on  $(\mathcal{S}, \mathcal{B})$  with Borel  $\sigma$ -field  $\mathcal{B}$ .

**Definition 2 (Mixed ESS)** *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  $Q \neq P$*

Using Definition 1 - the pure strategy definition of an ESS - Selten & Hammerstein (1984) argue 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 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.

The following theorem states that finite mixtures of strategies can also not be evolutionarily stable. Phrased in terms of Definition 2, we claim that no strategy  $P$  can be ESS if  $P$  is a probability distribution that puts probability 1 on a finite number of strategies. The proof is a simple generalization of Selten & Hammerstein's argument; in a finite mixture there is only a finite number of equilibrium paths and hence there is always an infinite number of off-equilibrium paths left on which behaviour can be changed without affecting payoffs. Please note that Farrell & Ware (1989) make the same claim - and prove it - but they use a different definition of evolutionary stability. Furthermore we will focus on games with discounting, but the theorem below holds for undiscounted, infinitely repeated games too.

**Theorem 3** *In a non-trivial repeated game there is no finite mixture of strategies that is evolutionarily stable*

**Proof.** 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 behaviour 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. ■

One reaction to a non-existence result like this is to be less demanding. Bendor & Swistak (1995, 1997, 1998) did this and chose to look at strategies that satisfy a weaker condition - [3] in their paper. This condition equals Definition 1, but then with all inequalities non-strict. They chose to name strategies that satisfy this relaxed condition evolutionarily stable too, but clarity might be served with following Maynard Smith (1982) and Weibull (1995) in terming such strategies neutrally stable (NSS). In the current paper, the definition also includes mixed strategies, as opposed to Bendor & Swistak (1995, 1997, 1998)

**Definition 4 (Mixed NSS)** *A strategy  $P$  is neutrally stable if both*

$$\begin{aligned} &\Pi(P, P) \geq \Pi(Q, P) \text{ for all } Q \text{ and} \\ &\text{if } \Pi(P, P) = \Pi(Q, P) \text{ then } \Pi(P, Q) \geq \Pi(Q, Q) \end{aligned}$$

While there is no ESS, Bendor & Swistak (1995, 1997, 1998) do find a profusion of (pure) NSS'es. They also find that nice and retaliatory strategies have larger basins of 'non-repulsion'.<sup>1</sup>

---

<sup>1</sup>In a finite automata setting with complexity costs and lexicographic preferences Binmore & Samuelson



### 3 Stepping stones in either direction

A question one could ask is how much is lost if the demands are lowered from evolutionary to neutral stability. After all, the fact that there is no ESS does by itself not make neutral stability a more stable concept. It therefore seems worth trying to find out exactly how stable or unstable those NSS'es in the repeated prisoners dilemma are. As the only difference between the definitions of an ESS and an NSS is that the latter allows for invasions by neutral mutants, the question then becomes how much harm these neutral mutants can do.

If we think for example of the strategy Tit-for-tat, then it is clear that cooperation's worst enemy 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 Van Veelen (2010) the concept of robustness against indirect invasions is introduced. For a strategy to be robust against indirect invasions (RAII) it must not only be a NSS, but there must also not be a sequence of neutral mutants that opens the door for each other, one after another, until some mutant strategy has an actual selective advantage. The example above shows that Tit-for-tat is not RAI. Below we will see that it is not just Tit-for-tat, but that there is in fact no strategy in a non-trivial repeated game that is RAI; the proofs of the theorems below construct ways to leave any equilibrium in only two steps, both in the direction of higher and in the direction of lower levels of cooperation (if higher resp. lower levels are possible). More precisely, Theorem 5 shows that any positive level of cooperation can be undermined by a succession of two mutations, while Theorem 6 states that if there are possible gains from (increased) cooperation, and the probability of continuation is sufficiently high, then also a stepping stone route into more cooperation exists. Together they imply that no equilibrium in interesting repeated games with low enough probability of breakdown is RAI, and mostly there are ways out of equilibrium in the direction of in- as well as in the direction of decreasing cooperation. Both theorems come in a pure strategy version for expositional clarity and connection to the literature (5 and 6) and a mixed strategy version for generality (8 and 9).

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  $\pi : A \times A \rightarrow \mathbb{R}^2$ . Using a discount factor  $\delta$ , 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:

---

(1992, 1997) relax the requirement of an ESS to a MESS (see also Swinkels & Samuelson (2003) for a perfectly accurate and meaningful characterization of the different definitions). Their results are also in favour of efficiency.

$$h_1 = ()$$

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

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:

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

It will furthermore be useful to have a way of writing down a history with the roles of the players reversed. Given a history  $h_t$  as they are defined above, its mirror image  $h_t^{\leftarrow}$  is found by simply renumbering the players:

$$h_1^{\leftarrow} = ()$$

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

The reason why histories with roles reversed are needed, is that we assume that both players label themselves as player 1 and the other as player 2 and therefore face mirrored histories as they go along.

A 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

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

and the recursion step

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

$$h_{t+1}^{S,T} = \left( h_t, a_t^{S,T} \right), \quad t = 1, 2, \dots$$

The discounted normalised payoffs to (a player that uses) strategy  $S$  against strategy  $T$  is given by:

$$\Pi(S, T) = (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} \pi_1 \left( a_t^{S,T} \right)$$

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 equilibrium of the game  $\Gamma(\delta)$ .

**Theorem 5** *Let  $S$  be a strategy in the game  $\Gamma(\delta)$  and let there be a time  $\tau$  at which  $a_\tau^{S,S}$  is not an equilibrium of the stage game. Then  $S$  is not weakly robust against indirect invasions.*

**Proof.** Assume that  $S$  is an equilibrium (if it is not, it is trivially not robust against indirect invasions). Let  $T$  be the strategy that equals  $S$  for all histories, except for those that are elements of the set  $\hat{H} = \{h_t \mid t > \tau, a_{\tau,2} = \arg \max_{a \in A} \pi_2(S(h_\tau^{S,S}), a)\}$ . 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  $\tau$ . For those histories  $h_t \in \hat{H}$  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 for  $h_\tau^{S,S}$ , for which we take  $U(h_\tau^{S,S}) = \arg \max_{a \in A} \pi_1(a, S(h_\tau^{S,S}))$  and except for histories that are elements of the set  $\tilde{H} = \{h_t \mid t > \tau, a_{\tau,1} = \arg \max_{a \in A} \pi_1(a, S(h_\tau^{S,S}))\}$ , 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  $\tau$  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 weakly robust against indirect invasions.

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. ■

What this theorem indicates is that as soon as there are equilibrium actions that must be upheld by the threat of punishment, then there can be mutants that do not punish, and subsequently there can be other mutants that takes advantage of the first mutant not punishing. One thing worth noting is that the proof constructs only one way out of equilibrium. While this particular stepping stone path changes behaviour 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 Figure 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.

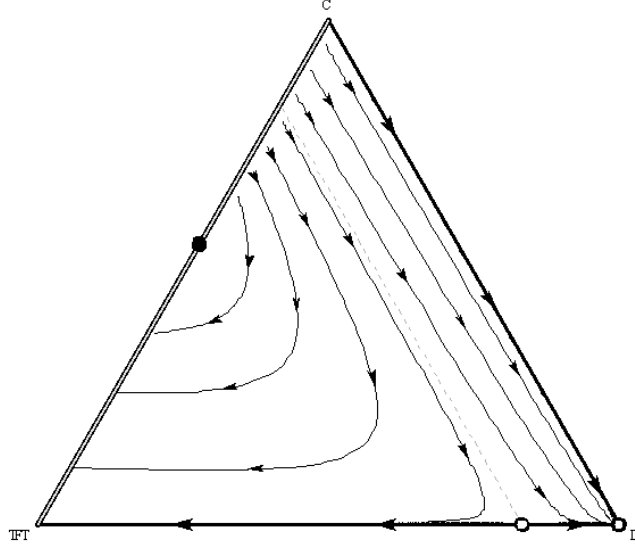


Figure 1. An example of an indirect invasion with decreasing cooperation; TFT, AllC and AllD. The dynamics are computed for  $\delta = \frac{3}{4}$ .

While the reference point in Theorem 5 is the equilibrium of the one-shot game, we will now focus on departures from what in non-trivial games is the other extreme: the maximally feasible symmetric payoffs. Therefore we define  $\pi_{\max} = \max_{a \in A} \pi_1(a, a)$  and  $a_{\max} = \arg \max_{a \in A} \pi_1(a, a)$ . 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 unilaterally initiating cooperation could be offset by future gains from (increased) cooperation, then the strategy is not robust against indirect invasions.

**Theorem 6** *Let  $S$  be a strategy in the game  $\Gamma(\delta)$  and let there be a time  $\tau$ , for which the following holds:*

1.  $\pi_1(a_{\tau}^{S,S}) - \pi_1(a_{\max}, S(h_{\tau}^{S,S})) < \sum_{t=\tau+1}^{\infty} \delta^{t-\tau} (\pi_{\max} - \pi_1(a_t^{S,S}))$ .
2.  $a_{\max} \neq S(h_{\tau}^{S,S})$

*Then  $S$  is not weakly robust against indirect invasions.*

**Proof.** Assume that  $S$  is an equilibrium (if it is not, it is trivially not robust against indirect invasions). 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  $h_t \in \widehat{H}$  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 for the history  $h_{\tau}^{S,S}$ , for which we choose  $U(h_{\tau}^{S,S}) = a_{\max}$  and except for the histories that are elements of the set  $\widetilde{H} = \{h_t \mid t > \tau, a_{\tau,1} = a_{\max} \text{ and } a_{u,2} = a_{\max}, u > \tau\}$ , for which we take  $U(h_t) = a_{\max}, h_t \in \widetilde{H}$ .

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 weakly robust against indirect invasions.

As in the proof of Theorem 5,  $S$  being an equilibrium implies that  $T \neq S$ . ■

The requirements in this theorem are slightly less simple to check for than those in Theorem 5, but when translated to prisoners dilemma's, it turns out to imply something that is relatively easy to handle. Before doing so, however, it is good to realize that discounted, normalised payoffs that belong to a combination of two strategies can vary with  $\delta$  and that they do so in different ways. If we look at symmetric equilibria, then it might be that two different strategies that, when played against themselves, both have the same discounted, normalised payoff for a given  $\delta$ , while a higher  $\delta$  increases them for one and decreases them for the other strategy.

The next theorem states that for repeated prisoners dilemma's, all symmetric equilibria with payoffs less than  $\pi_1(C, C) - (1 - \delta)[\pi_1(C, C) - \pi_1(C, D)]$  are not robust against indirect invasions. If we take more or less standard values, that is  $\pi_1(D, C) = 4, \pi_1(C, C) = 3, \pi_1(D, D) = 1, \pi_1(C, D) = 0$ , then this amounts to  $3\delta$ ; all strategies  $S$  with with payoffs  $\Pi(S, S)$  less than  $3\delta$  are indirectly invadable. There may be many other equilibria that are also not robust against indirect invasions, but Theorem 7 shows that at least all strategies with relatively low payoffs satisfy the criteria for Theorem 6. It also means that the closer  $\delta$  gets to 1, the more strategies are shown to be vulnerable to indirect invasions with increasing cooperation, and for any strategy  $S$  with payoff  $\Pi(S, S) < \pi_1(C, C) - (1 - \delta)[\pi_1(C, C) - \pi_1(C, D)]$  there is a  $\bar{\delta} \in (0, 1)$  such that  $S$  indirectly invadable for all  $\delta \in (\bar{\delta}, 1)$ . Together with Theorem 5 that implies that for sufficiently high  $\delta$  no symmetric equilibrium strategy is robust against indirect invasions.

**Theorem 7** *In a repeated prisoners dilemma, all strategies  $S$  with  $\Pi(S, S) < \pi_1(C, C) - (1 - \delta)[\pi_1(C, C) - \pi_1(C, D)]$  are not weakly robust against indirect invasions.*

**Proof.** First realize that  $S$  is  $\pi_1(C, C) - \Pi(S, S)$  short from full, symmetric efficiency. Then choose as time  $\tau$  in Theorem 6 the first period that  $S$  plays defect. The second requirement of the same theorem is then automatically fulfilled.

The following can then be derived

$$\Pi(S, S) < \pi_1(C, C) - (1 - \delta)(\pi_1(C, C) - \pi_1(C, D)) \Rightarrow$$

$$\Pi(S, S) < \pi_1(C, C) - (1 - \delta)\delta^\tau(\pi_1(C, C) - \pi_1(C, D)) \Leftrightarrow$$

$$\pi_1(D, D) - \pi_1(C, D) < \frac{1}{(1-\delta)\delta^\tau} [\pi_1(C, C) - \Pi(S, S)] - [\pi_1(C, C) - \pi_1(D, D)] \Leftrightarrow$$

$$\pi_1(a_\tau^{S,S}) - \pi_1(a_{\max, S}(h_{\tau-1}^{S,S})) < \sum_{t=\tau}^{\infty} \delta^{t-\tau} (\pi_{\max} - \pi_1(a_t^{S,S})) - (\pi_{\max} - \pi_1(a_\tau^{S,S}))$$

This satisfies the first requirement of Theorem 6. ■

Again, the proof of Theorem 6 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 with an increasing level of cooperation.

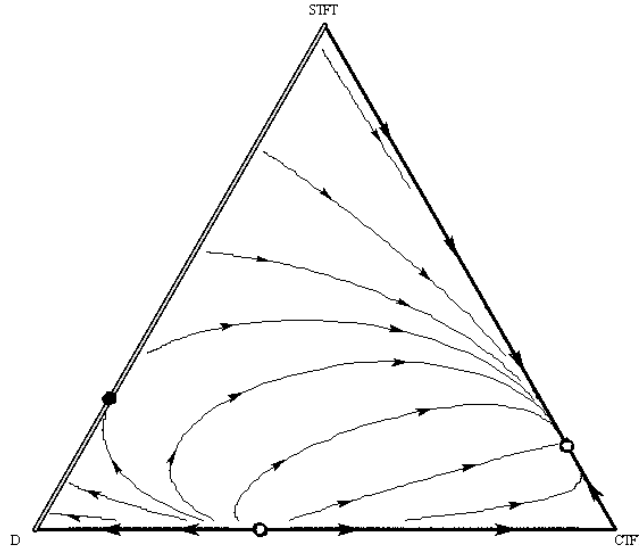


Figure 2. An example of an indirect invasion with increasing cooperation; ALLD, Suspicious TitForTat (STFT, also known as TatForTit) and Cooperate-TitForTat (CTFT, cooperates on the first two moves and then imitates the opponent). The dynamics are computed for  $\delta = \frac{3}{4}$ .

### 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 focussed on pure equilibria (with exceptions such as for instance Binmore & Samuelson, 1992, and Samuelson & Swinkels, 2003). It seems however no less 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 3 shows that there is also no mixed ESS, Theorems 5 and 6 do not yet exclude the possibility that there is a mixture of strategies that is RAI. In this subsection we therefore give the equivalents of those theorems for finite mixtures. Here we will directly focus on repeated prisoners dilemma's 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  $(1 - \delta) \sum_{t=0}^{\infty} \delta^{t-1} \pi_1(D, D) = \pi_1(D, D)$ , which is the normalised discounted payoff of *AllD* against *AllD*.

#### Theorem 8

*Let  $P$  be a finite mixture of strategies in  $\Gamma(\delta)$ .*

*If  $\Pi(P, P) > \Pi(D, D)$  then  $P$  is not weakly robust against indirect invasions.*

**Proof.** See Appendix B.1 ■

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, some of which can be considered to be more likely than others. But the theorem shows that indirect invasions are always possible for equilibria with cooperation.

In order to formulate the mixed strategy counterpart for increasing cooperation, it will be helpful to define the following. 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 probabilities with which they are played in  $P$ . For any defection that occurs along a path of interaction between any two strategies  $P_i$  and  $P_j$  from  $P$  we can discount the possible gains in the future and compare it to the current period loss of switching from  $D$  to  $C$  as an initiation of cooperation. Therefore we first define  $E_i(j)_t = \left\{ P_l \mid h_t^{P_i, P_l} = h_t^{P_i, P_j} \right\}$ , 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 (see also the proof of Theorem 8) as  $E_i(j) = \left\{ P_l \mid a_t^{P_i, P_l} = a_t^{P_i, P_j} \forall t \right\}$ . For any combination of strategies  $(P_i, P_j)$  and any time  $t$  we can compute  $\delta_{ij,t}$  as follows:

$$\delta_{ij,t} = \begin{cases} \delta \text{ such that} \\ \sum_{P_l \in E_i(j)_t} p_l \left( \pi_1 \left( a_t^{P_i, P_l} \right) - \pi_1 \left( C, a_{t,2}^{P_i, P_l} \right) \right) = \sum_{P_l \in E_i(j)_t} p_l \sum_{u=t+1}^{\infty} \delta^{u-t} \left( \pi_1 (C, C) - \pi_1 \left( a_u^{P_i, P_l} \right) \right) \\ \text{if } a_{t,1}^{P_i, P_j} = D \text{ and the equation has a solution } \delta \in (0, 1) \\ 1 \quad \text{otherwise} \end{cases}$$

This definition greatly simplifies the formulation of the next theorem. Note that the condition is very modest; the continuation probability  $\delta$  only has to be larger than the smallest  $\delta_{ij,t}$ .

**Theorem 9**

*Let  $P$  be a finite mixture of strategies in  $\Gamma(\delta)$ .*

*If  $\min_{i,j,t} \delta_{ij,t} < \delta < 1$  then  $P$  is not weakly robust against indirect invasions.*

**Proof.** See Appendix B.2 ■

## 4 Indirect invasions in the simulations.

In Section 3 it was already mentioned that the proofs only provide two stepping stones 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 (see Appendix A), one could say that the existence of only one or two stepping stone paths out would not necessarily make an equilibrium very unstable. We do however know that the paths constructed in the proofs are not the only paths out and we conjecture that in fact there will be very many 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 when 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. In order to be able to say if these indirect invasions indeed drive the evolutionary dynamics in relevant, interesting settings, we will therefore have to combine the game with a priori reasonable mutation processes. For our simulation approach we simply started out with what we think is the most natural choice for a set of mutation processes, if we do not want to exclude any part of the strategy space.



## 4.1 The simulations

The basics of the simulation are quite simple. There are  $N$  individuals and every generation they are randomly matched in pairs to play the repeated game. Because the game has a probability of breakdown that is 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 simulations reported here we use the Wright-Fisher process. In the Wright-Fisher process 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  $j$ . (The program also has the option to run it as a Moran process, but that is computationally very inefficient. The online material at [www.creedexperiment/evolution](http://www.creedexperiment/evolution) contains a more detailed description). After the new generation has been drawn, all individuals mutate with a small probability. This completes the cycle for one generation. The cycle is repeated a large number of times.

Strategies are programmed explicitly as finite automata.<sup>2</sup> A strategy 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. There are four types of mutations we allow for: mutations that add a state, mutations that delete a state, mutations that change the output when in a state, and mutations that change for a given state to which state this player goes given an action of the opponent. We chose mutation schemes where all of those four types of mutations are possible. If that is the case, then there are two things that are worth realizing. The first is that every finite automaton can be reached by a finite sequence of mutations from any other automaton. The other is that if we choose a very natural distance, then the set of all finite automata is dense in the set of all strategies (see Appendix A). Together, this implies that we can get arbitrarily close to any strategy through a finite sequence of mutations. This we think is a very attractive property of a mutation scheme.

These ingredients are all there is to the simulation program itself. Still this simple setup gives us a dynamic process with a few quite interesting features. As we will see below, this evolutionary process always finds a stepping stone path out of equilibrium in reasonable time. In fact, with increasing population size, indirect invasions come to dominate everything else as a driver of the dynamic process. That however does not mean that the notion of an equilibrium is not important; after leaving an equilibrium, the dynamics

---

<sup>2</sup>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 at the website that 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.

tend to take the population to a new equilibrium relatively fast. The population settles in the new equilibrium for a while, until another indirect invasion occurs. Under a reasonable set of mutation schemes, populations of a reasonable size therefore appear to walk from equilibrium to equilibrium through indirect invasions (see also Figure 3).

One feature that we find particularly appealing in these simulations, is that they nicely walk through a large strategy space in a way that shares quite some features with how we think of actual evolutionary processes. While there is typically only a few strategies present at any point in time, the process nonetheless takes the populations through many different parts of a vast strategy space. This means that locally we can very well describe what happens in the simulations with dynamics on simplices of relatively low dimensions. New mutations however allow the process as a whole to go from one simplex of low dimension to another, typically through shared facets. What is also realistic, is that mutations have to work with what is there; the probability of a certain mutant entering the population depends on how similar the mutant is to what is there at the moment in the current population. This matches with how we think many evolutionary processes take populations through what typically is a vast space of possibilities. Even though mutations are local - in the sense that they only alter existing strategies - the dynamics of the process as a whole can be quite rich, with a population that, although mostly in or close to equilibrium, still makes its way through a rich strategy space.

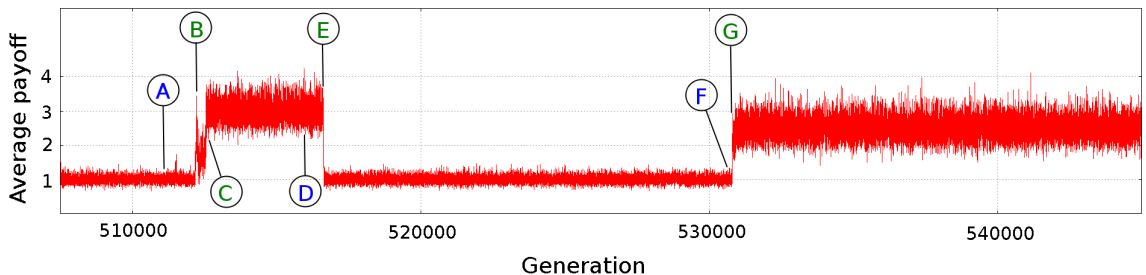


Figure 3. Part of a typical run. The population size is 128, the continuation probability is 0.75. Blue letters indicate where relevant neutral mutants occur, green letters indicate advantageous mutants entering. It moves from a fully defective equilibrium to a fully cooperative one, back to a fully defective one, and then to an equilibrium with partial cooperation. The actual sequence of strategies is given in Appendix D.

## 4.2 Capturing transitions

The aim of the simulations is to find out if the possibility of indirect invasions indeed makes a substantial difference for evolutionary dynamics in repeated games. Before being able to say if an equilibrium was left through an indirect invasion, it is important to first be able to say if it was left at all. While transitions are made possible by a mutation process that

constantly produces new strategies, that very same production of new strategies also creates noise in the population. This implies that if we think for instance of a pure equilibrium, we should not only classify a population as being at that equilibrium if the population consists of that one strategy only and nothing else. Given the frequent introduction of mutants, most of which enter only to be eliminated from the population before ever having attained a considerable share, we should also classify nearby population states as being at that equilibrium, and create a bandwidth which allows us to disregard the noise.

If the population at time  $t$  consists of strategy  $A$  only, and at time  $t + 100$  of strategy  $B$  only, then it is fair to say that at least one transition has occurred. If on the other hand the population at time  $t$ , and at time  $t + 100$ , and at all times in between, consists of between 90 and 100% strategy  $A$  and a remainder that is composed of an ever changing set of other strategies, then it seems reasonable to assume that a transition has not occurred, and that the little differences only reflect the regular influx and extinction of new mutations.

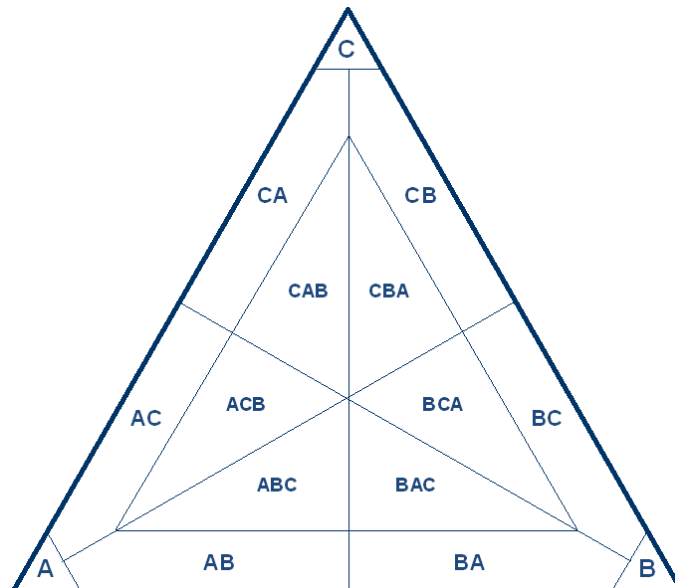


Figure 4. With only three strategies present, this depicts the classification of population states with a threshold of 90%. A population that consists, for example, of 65% strategy A, 30% strategy B and 5% strategy C is classified as a 2 dimensional mixture of A (most popular) and B (second most popular). With four strategies, we get a three dimensional simplex with a smaller simplex-shaped area in the middle that represents the ‘dark zone’.

We therefore begin the classification of a population state by ranking the composing strate-

gies from frequent to infrequent. Then we look at the minimum number of strategies that is needed to capture at least a fixed percentage of the population (below we choose 90% for the threshold). A population state is then characterized by its dimensionality (pure, 2 strategies, 3 strategies, more than 3 strategies) and the actual most popular strategy, resp. the 2 most popular and the 3 most popular strategies as they are ranked (see Fig. 4). The classification thereby never ignores more than 10% of the population. With the mutation rates and population sizes we chose in the simulations below, a population where the three most popular strategies made up less than 90% of the population was a rare exception.

This classification allows us, at least to some extent, to pick up three types of (possible) equilibria; pure ones, mixed ones with two strategies, and mixed ones with three strategies. If the population is at a pure equilibrium, we expect that it finds itself in a corner pocket (see Fig. 4), and that most of the mutants do not take the population outside this corner pocket. If a population is a mixed equilibrium with two strategies, it should find itself somewhere in between two edges, on the facet of the simplex. The construction of the pocket excludes that the areas that are meant to capture the mixed equilibria consist of the whole facet, because a small part of it is already contained in the pocket. But, again with a threshold of 90%, if the population finds itself in a mixed equilibrium where both strategies account for more than 10% of the equilibrium frequencies - and hence the equilibrium is not contained in a corner pocket - we expect that most mutations will not take the population outside the area it is in. On the facet itself we expect that on average the population will also be pushed in the direction of its equilibrium composition.

With this way to classify population states, we can follow the population as it travels from region to region. At any such transition, we can check if this transition can be associated with a neutral mutant entering or exiting the population, or with advantageous or disadvantageous mutants entering or exiting the population. This gives us the possibility to characterize a sequence of transitions as an indirect invasion or an invasion of a different kind. If strategy  $A$  is a pure equilibrium strategy, and  $B$  is a neutral mutant of  $A$ , and  $C$  has a strict advantage against  $B$ , then - with obvious abbreviations - going from the region  $A$  to  $AB$  to  $BA$  to  $B$  to  $BC$  will be classified as an indirect invasion. If  $A$  is a pure equilibrium, and  $D$  is a mutant with a selective disadvantage, then going from region  $A$  to region  $AD$  is classified as a different invasion. (On [www.creedexperiment/evolution](http://www.creedexperiment/evolution) we go into more detail on possible boundary crossings and possible sequences of boundary crossings).

This immediately points out the trade-off that we face for the choice of a threshold. If we choose a threshold that is larger than  $\frac{N-1}{N}$ , where  $N$  is the population size, then any mutant entering the population will take it outside the corner pocket. All mutants will therefore be recorded as transitions, and all disadvantageous mutants will be recorded as “different invasions”, even though they might be extinct the next generation already. This high threshold thereby leaves no room at all to observe what we are interested in, which is

the difference in how selection acts on different (sequences of) mutants. On the other hand, if we choose the threshold at 50%, then we leave no room to observe dynamics near mixed equilibria at all, as described above. So a lower threshold means more room to observe selection at work, but also more mixed equilibria that will go unnoticed, because they end up in corner pockets.

### 4.3 Capturing indirect invasions

Theorems 5 and 6 concern stepping stone paths out of pure equilibria. Since the theorems suggest that equilibria can be left through neutral mutants that open doors for other, advantageous mutants, it makes sense to first acknowledge that there will also be neutral mutants that themselves still are equilibria, and that therefore do not yet open such doors. As the starting point of a path out of a pure equilibrium, we therefore only choose those equilibria that were not themselves reached by a neutral invasion. Thereby we allow for neutral mutants that themselves are still equilibria just to be a part of the stepping stone path. As expected, all such sequences of neutral mutants turn out to 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, followed, first, by a neutral invasion to a state that is not an equilibrium, which in turn is followed by an advantageous mutant, then it is reasonable to count the whole sequence as one single indirect invasion. It is for sure a two-step indirect invasion when counted from the last equilibrium, and one can very well also see it as an indirect invasion into the first and into all equilibria in between, but counting it as just one single indirect invasion is more than reasonable.

If a pure equilibrium is left through a sequence of one or more boundary crossings that can be characterized as neutral, followed by one boundary crossing that can be characterized as the entry of an advantageous mutant, then it is qualified as an indirect invasion. All other sequences out of equilibrium (zero or more neutral mutants followed by a boundary crossing that can be characterized as the entry of disadvantageous mutant) are qualified as “other invasions”. Notice that theorems 5 and 6 concern leaving equilibria, and not what happens after an equilibrium is left. What we are after here is therefore first and foremost to capture paths away from the equilibrium that occur in the simulations.

Finding those starting points (pure equilibria that are themselves not reached by a neutral mutant) is facilitated by an automated procedure called the best responder. This procedure can, for a single finite automaton, find the best response to it, and thereby determine if it is a best response to itself and hence a pure equilibrium (see Appendix C).

We ran simulations for different population sizes, and first looked at how pure equilibria are left. The data indicate that for pure equilibria the share of indirect invasions goes to 1 with an increase in population size (see Fig. 5).

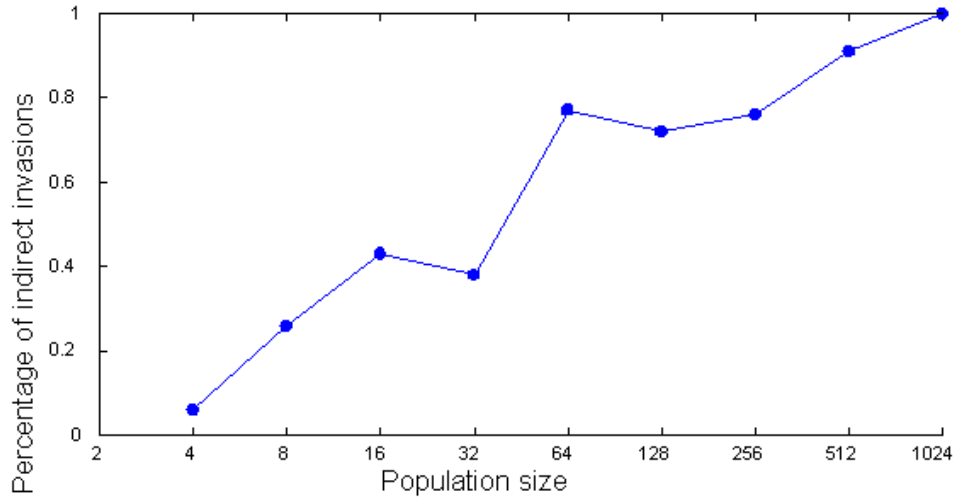


Figure 5. Transitions out of pure equilibria. As the population increases, the share of transitions that are indirect invasions goes to one.

For combinations of two or more finite automata it is harder to get such clean data. Constructing a best responder to a mixture of strategies is far more complicated than constructing one for pure strategies and thereby well beyond the scope of this paper. This implies that we do not have an automated procedure that determines if mixed states are equilibria. The best we can do without having this procedure at our disposal is to disregard as many obvious mixed disequilibrium states as possible, and see how the remainder of the mixed states - again, not themselves reached by a neutral invasion - is left. Here we of course again set apart those that are left by indirect invasions, but in the remainder there is an extra category. Since we expect that a mixed equilibrium might also be left relatively easily by one of the composing strategies going extinct, we also count separately how many are left by the population moving into a region of a lower dimension. We find that for every population size, the vast majority of mixed states consisting of 2 strategies are left by one of the two fixating. (The data and a more elaborate discussion are in the online material). The simulations do not render enough data for states consisting of 3 pure strategies, as the population spends most of its time in states with low dimensions.

To support the claim that the dynamics take a population from equilibrium to equilibrium through indirect invasions, it is worthwhile to look at what happens after one equilibrium is left, and how long it takes to get to another. Although not the main aim

of the simulations - which focus on leaving equilibria, and not on arriving at others - it is interesting to note that the percentage of time spent in equilibrium also increases with the population size. With increasing population size, it gets a bit harder to pinpoint how much time is spent in equilibrium; more time is spent outside of the corner pockets, for which we do not have a mixed state best responder procedure. We can however give upper and lower bounds, as Figure 6 does. Note that with small population sizes (the left side of Fig. 6) and fixed per individual mutation rates, only few mutations occur. This implies that by chance a population can get stuck in a disequilibrium state for a very long time, just because the right mutation takes forever to appear.

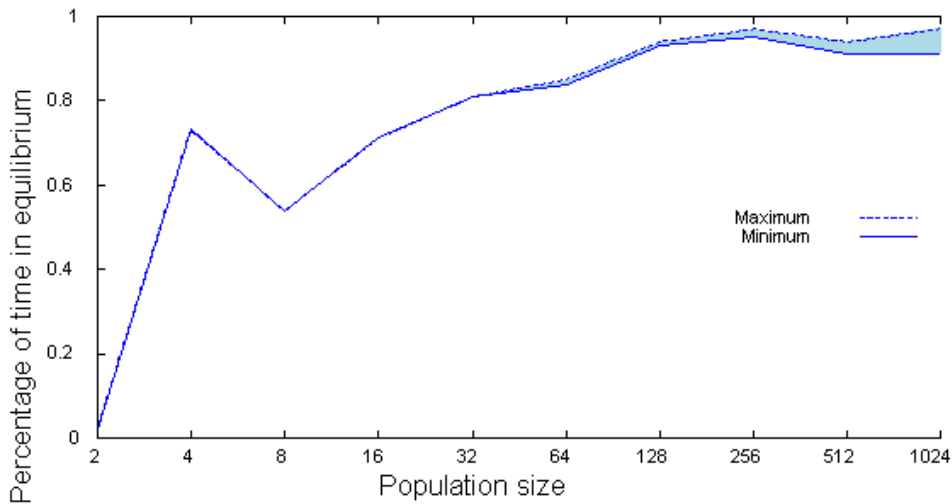


Figure 6. Time spent in equilibrium as a function of population size.

Together, these findings strongly suggest that populations playing repeated prisoners dilemmas do indeed walk from equilibrium to equilibrium through indirect invasions, already for populations that are not even that large. The leaving of pure equilibria is shown to become dominated completely by indirect invasions, while mixed states tend to be left by fixation events of the composing pure strategies more than through indirect invasions.

#### 4.4 Comparison between equilibria that are not RAI with equilibria that are

Our claim is that it matters that equilibria in the repeated prisoners dilemma, although neutrally stable (NSS), are not robust against indirect invasions (RAI). If we rank a few well

known different equilibrium concepts in evolutionary game theory according to decreasing levels of stability, then we get the following inclusions.

$$\Delta^{ESS} \subset \Delta^{NSS} \subset \Delta^{NE}$$

A Nash equilibrium that is not *NSS* is easily left; not being *NSS* means that there is a mutant for which the replicator dynamics pushes the population away from the equilibrium if this mutant enters the population. An *ESS* on the other hand is very stable; there is a neighbourhood such that the replicator dynamics take the population back to the ESS from any point in that neighbourhood. Still, stability remains a relative thing; although it is much harder to leave an *ESS* - it takes orders of magnitude more time in any reasonable stochastic dynamic process - than it is to leave a Nash equilibrium that is not *NSS*, it is not impossible. In fact, some of the most interesting papers in the literature on evolutionary game theory compare the stability of different *ESS*'es in one and the same game by comparing how many (simultaneous) mutations it takes to get from one *ESS* to the other (see for instance Kandori, Mailath & Rob, 1993, Kandori & Rob, 1995, Foster and Young, 1990, Young, 1993, 1998, Ellison, 2000). It may take very long, but the idea is that it could still be that one equilibrium is left more easily than the other. This then gives rise to a further refinement (stochastic stability in Young, 1993, or long run equilibria in Kandori, Mailath & Rob, 1993).

Our claim is that in order to properly compare stability, the categorization with *ESS*, *NSS* and *NE* is in some cases a bit too rough. Some *NSS*'es are, and some are not *RAII*. Those that are *RAII*, we suggest, are much more stable than those that are not. Since being *RAII* implies that this strategy is contained in a setwise generalisation of an *ESS* (see van Veelen, 2010) one could even say that in the sequence of inclusions below, the bigger gap is actually between *RAII* and *NSS*, and not between *ESS* and *RAII*.

$$\Delta^{ESS} \subset \Delta^{RAII} \subset \Delta^{NSS} \subset \Delta^{NE}$$

In order to indicate that it does indeed matter that *NSS*'es in a repeated prisoners dilemma are not *RAII*, we will compare it to a game that is rather similar in a lot of respects, but that has equilibria that are in fact *RAII*. If we replace the prisoners dilemma as a stage game with a coordination game, then we preserve everything, including the richness of the strategy space. The only difference is that now we do get equilibria that are *RAII* (but not *ESS*). That means that what sets these repeated games apart is the existence resp. nonexistence of the possibility for indirect invasions.

2	0
0	2



Below we compare the number of transitions in simulations for the repeated prisoners 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.

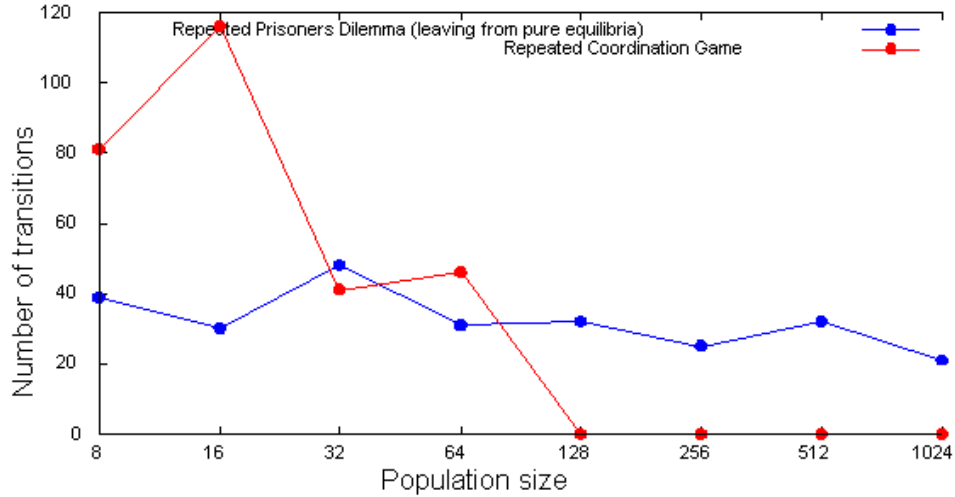


Figure 7. The number of transitions leaving pure equilibria in a repeated prisoners dilemma and in a repeated coordination game for different population sizes. Note that in the repeated coordination game, all RAI equilibria are pure, and all mixed equilibria are not NSS.

The number of transitions away from pure equilibria in the repeated prisoners dilemma is decreasing only very slightly. This fits what we expect; with example 6 in Van Veelen (2010) in mind, we expect that it is the fixation probability of neutral mutants that determines the speed at which indirect invasions occur in large populations. The fixation probability of a neutral mutant is  $\frac{1}{N}$ , where  $N$  is the population size, but since the mutation probability *per individual* is constant, the number of neutral mutants entering is proportional to  $N$ . If we can assume that every neutral mutant has either fixated 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 fixated. This interference implies that we will be seeing slightly

less fixation events for larger populations.

The number of transitions in the repeated coordination game on the other hand decreases much more drastically. The data suggest that the number of transitions out of pure equilibria in the repeated coordination game divided by the number of transitions in the repeated prisoners dilemma goes to 0 rather rapidly. One can therefore conclude that evolution in a population playing the repeated prisoners dilemma remains infinitely more mobile than evolution in a population playing a repeated coordination game, and that the possibility of indirect invasions makes all the difference.

## 4.5 Average reciprocity and niceness

The appendix also contains ways to measure reciprocity and cooperativeness for strategies. Not surprisingly, we find that strategies are on average somewhat reciprocal and moderately cooperative. This is rather natural; all equilibrium strategies are between uncooperative and fully cooperative and between not reciprocal and very reciprocal. A process that takes the population from equilibrium to equilibrium, with in- as well as decreasing levels of cooperation therefore must lead to an average that is somewhat reciprocal and moderately cooperative. Paths from one equilibrium to the other also typically first exhibit an increase (decrease) of the level of reciprocity, which is followed by an increase (decrease) of the level of cooperation. Note that, although the measure for cooperativeness can by construction not fall below 0 or be larger than 1, the measure for reciprocity can be negative.

There are however also equilibria where cooperation is preceded by a ‘negative handshake’. If we look at the following automaton, then it constitutes an equilibrium where cooperation that comes too soon is actually punished. We do sometimes see equilibria like this arising in the population (see also Appendix D, that contains the sequence of strategies that comes with the part of a run that is depicted in Fig. 3). Reciprocity here is not unambiguously positive; this strategy rewards cooperation when in state 4 and 5, but it does also punish cooperation when in state 1, 2 or 3 (see again the online tutorial for more detailed description of measures of reciprocity).

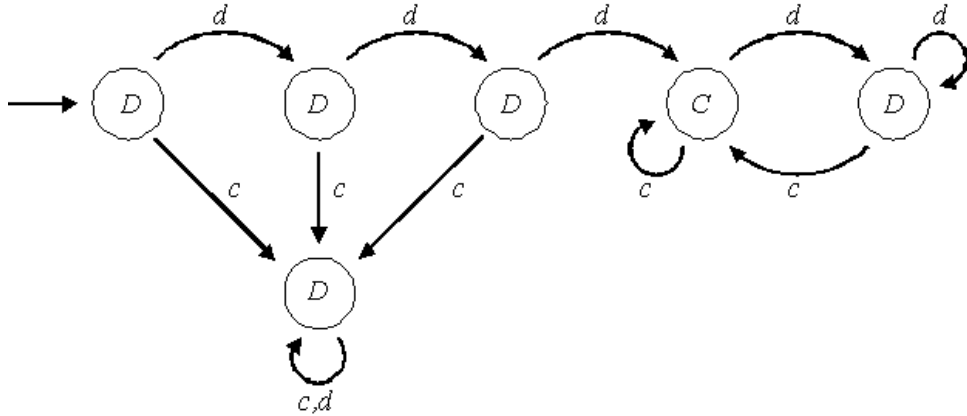


Figure 8.  $D^3TFT$ . With this strategy, three  $D$ 's is the code for cooperation to start. Once past the initial uncooperative phase, it plays just like ordinary  $TFT$ , but any deviation from the initial uncooperative phase implies eternal defection.

## 5 Conclusion and 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 of course the repeated prisoners dilemma. Theorems 8 and 9 show that with sufficiently large continuation probability  $\delta$ , there is no strategy in the repeated prisoners dilemma that is robust against indirect invasions. 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 actually 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 can predict what happens in a population with random matching, mutation and selection. One important thing that this tells us, is that what we can expect to evolve will essentially depend - besides on  $\delta$  - 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 stepping stone paths out of equilibrium, both with in- and with decreasing levels of cooperation. Whether we can expect cooperation to in- or decrease therefore depends on how many more of these paths there are, and, more importantly, on the probabilities with which the different mutations occur. Also the starting point might matter, although it seems that a natural starting point for evolution is the strategy to always defect.

If we allow ourselves to restrict the strategy space, then that offers a possibility to get stability results. It is however important too see that restricting the strategy space to, say, a 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 more or less falls within the message that 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  $\mathcal{T}$  instable.

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 of robustness against indirect invasions in our theory for infinite populations.

Playing with the simulations is also seriously fun, and can be done at [www.creedexperiment.nl/evolution](http://www.creedexperiment.nl/evolution).

## 6 Acknowledgements

Matthijs would like to thank Gerard van der Laan, Maus Sabelis, Martin Nowak, Corina Tarnita, Dave Rand, Arthur Schram, Peter Spreij and Aljaz Ule. 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. Figures 1 and 2 were made with Dynamo Version 1.4.1 by William H. 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).

## References

- [1] Abreu, D. (1988). "On the theory of infinitely repeated games with discounting," *Econometrica* **56**, 383-396.
- [2] Aumann, R., and L. Shapley (1976). "Long term competition: a game theoretic analysis," mimeo, Hebrew University.
- [3] Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic Books.
- [4] Bendor, J. and P. Swistak (1995). "Types of evolutionary stability and the problem of cooperation". *Proceedings of the National Academy of Sciences* **92**, 3596-360.
- [5] Bendor, J. and P. Swistak (1997). "The evolutionary stability of cooperation," *American Political Science Review* **91** (2), 290-307.
- [6] Bendor, J. and P. Swistak (1998). "Evolutionary equilibria: characterization theorems and their implications," *Theory and Decision* **45** (2), 99-159.
- [7] Binmore, K.G., and L. Samuelson (1992). "Evolutionary stability in repeated games played by finite automata," *Journal of Economic Theory* **57**, 278-305.
- [8] Binmore, K.G., and L. Samuelson (1997). "Muddling through: noisy equilibrium selection," *Journal of Economic Theory* **74**, 235-265.
- [9] Boyd, R. and J. P. Lorberbaum (1987). "No pure strategy is stable in the repeated prisoner's dilemma game, *Nature* **327**, 58-59.
- [10] van Damme, E. (1989). "Renegotiation-proof equilibria in repeated prisoner's dilemma," *Journal of Economic Theory* **47**, 206-217.
- [11] Dal Bó, P. 2005. "Cooperation under the Shadow of the Future: Experimental Evidence from Infinitely Repeated Games." *American Economic Review*, 95(5): 1591-1604.
- [12] Dal Bó, P., and G.R.. Fréchette (forthcoming) The evolution of cooperation in infinitely repeated games: experimental evidence. *American Economic Review*,
- [13] Ellison, G. (2000). "Basins of attraction, long-run stochastic stability, and the speed of step-by-step evolution." *Review of Economic Studies*, **67** (1), 17-45
- [14] Foster, D., and Young, H. P. (1990): "Stochastic Evolutionary Game Dynamics," *Theoretical Population Biology* **38**, 219-232.
- [15] Farrell, J. and E. Maskin (1989). "Renegotiation proof equilibria: reply," *Journal of Economic Theory* **49**, 376-378.
- [16] Farrell, J. and R. Ware (1989). "Evolutionary stability in the repeated prisoner's dilemma," *Theoretical Population Biology* **36**, 161-166.

- [17] Friedman, J. (1971). "A noncooperative equilibrium for supergames," *Review of Economic Studies* **38**, 1-12.
- [18] Fudenberg, D. and E. Maskin (1986). "The folk theorem in repeated games with discounting or with incomplete information," *Econometrica* **54**, 533-554.
- [19] Fudenberg, D., and E. Maskin (1990). "Evolution and cooperation in noisy repeated games," *American Economic Review* **80** (2), 274-279.
- [20] Fudenberg, D., L.A. Imhof, and M.A. Nowak (2005). "Evolutionary Cycles of Cooperation and Defection", *Proceedings of the National Academy of Sciences* **102** (31), 10797-10800
- [21] Imhof, L.A., and M.A. Nowak (2006). "Evolutionary game dynamics in a Wright-Fisher process," *Journal of Mathematical Biology* **52**, 667-681.
- [22] M. Kandori, G. J. Mailath and R. Rob (1993). "Learning, mutation, and the long run equilibria in games," *Econometrica* **61**, 29-56.
- [23] Lorberbaum, J.P. (1994). "No strategy is evolutionarily stable in the repeated prisoner's dilemma," *Journal of Theoretical Biology* **168**, 117-130.
- [24] Lorberbaum, J.P., D.E. Bohning, A. Shastri and L.E. Sine (2002). "Are there really no evolutionarily stable strategies in the iterated prisoner's dilemma?" *Journal of Theoretical Biology* **214**(2), 155-69.
- [25] Maynard Smith, J. and G.R. Price (1973). "The logic of animal conflict," *Nature* **246**, 15-18.
- [26] Maynard Smith, J (1982). *Evolution and the theory of games*. Cambridge: Cambridge University Press.
- [27] Moran, P.A.P. (1962). *The Statistical Processes of Evolutionary Theory*. Oxford: Clarendon Press.
- [28] Nowak, M.A. (2006) *Evolutionary dynamics: exploring the equations of life*. Harvard University Press, Cambridge, MA.
- [29] Rubinstein, A. (1979) "Equilibrium in supergames with the overtaking criterion," *Journal of Economic Theory* **21**, 1-9.
- [30] Rubinstein, A. (1980) "Strong perfect equilibrium in supergames," *International Journal of Game Theory* **9**, 1-12.
- [31] Samuelson, L. and J.M. Swinkels (2003). "Evolutionary stability and lexicographic preferences," *Games and Economic Behavior, Elsevier* **44** (2), 332-342.

- [32] Selten, R. and P. Hammerstein (1984). “Gaps in Harley’s argument on evolutionarily stable learning rules and in the logic of “tit for tat”,” *Behavioral and Brain Sciences* **7**, 115-116.
- [33] Thomas, B. (1985). “On evolutionarily stable sets,” *Journal of Mathematical Biology* **22**, 105-115.
- [34] Van Veelen, M. (2010). “But some neutrally stable strategies are more neutrally stable than others,” Tinbergen Institute discussion paper 10-033/1.
- [35] Weibull, J.W. (1995). *Evolutionary Game Theory*. Cambridge MA: MIT Press.
- [36] Young, H.P. (1993). “The evolution of conventions,” *Econometrica* **61**, 57–84.
- [37] Young, H.P. (1998). *Individual Strategy and Social Structure*, Princeton: Princeton University Press.

## A A metric for $\mathcal{S}$

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  $\mathcal{S}$  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.

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  $H_t$  therefore is  $k^{2t-2}$ .

Define the distance between  $S$  and  $T$ , both  $S, T \in \mathcal{S}$ , as follows:

$$d(S, T) = \sum_{t=1}^{\infty} \rho^t \sum_{h_t \in H_t} |f(h_t, S, T)|$$

with  $\rho = \frac{\delta}{k^2}$  and  $\delta \in (0, 1)$ .

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  $k^{(\sum_{v=1}^t k^{2v-2})} = k^{\left(\frac{k^{2t}-1}{k^2-1}\right)}$  elements. The set  $\bigcup_{t=1}^{\infty} \mathcal{S}_t$  is therefore countable, but it is easy to see that it is dense in  $\mathcal{S}$ .

## B Proofs of theorems

### B.1

**Proof of Theorem 8** 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  $\tau$  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  $\tau$  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  $P_i(h_{\tau}^{P_i, P_j}) = C$  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  $\tau'$  which is sufficiently large to determine whether or not  $(P_k, P_l) \in E(i, j)$ , that is, there is a  $\tau'$  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  $\tau'$  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  $\hat{H} = \left\{ h_t \mid t > \tau'', a_u = a_u^{P_j, P_i} \text{ for } u \leq \tau' \text{ and } a_{\tau'', 2} = D \right\}$ . These histories only occur off all equilibrium paths, since the history up to and including  $\tau'$  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  $h_t \in \hat{H}$  we take  $Q_l(h_t) = a_t^{P_j, P_i} = P_l(h_t^{P_j, P_i})$ . 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 that  $\Pi(Q, P) = \Pi(P, P) = \Pi(Q, Q) = \Pi(P, Q)$ .

Let  $R$  be the strategy that equals  $P_i$ , except for  $h_{\tau'', 1}^{P_i, P_j}$ , for which we take  $R(h_{\tau'', 1}^{P_i, P_j}) = D$  and except for histories that are elements of the set  $\tilde{H} = \left\{ h_t \mid t > \tau'', a_u = a_u^{P_i, P_j} \text{ for } u \leq \tau' \text{ and } a_{\tau'', 1} = D \right\}$ , for which we take  $R(h_t) = P_i(h_t^{P_i, P_j})$ ,  $h_t \in \tilde{H}$ .

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  $\tau''$  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 weakly 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  $\Pi(R, P) \leq \Pi(P, P)$ .

## B.2

**Proof of Theorem 9** Assume that  $P$  is an equilibrium (if it is not, it is trivially not robust against indirect invasions). Take  $i, j$  and  $\tau$  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  $\hat{H} = \{h_t \mid t > \tau, a_{u, 2} = C, u \geq \tau\}$ . These histories only occur off all equilibrium paths; the assumption implies that  $\delta_{ij, \tau} < 1$  and hence it is not possible that  $a_{\tau, 2}^{P_i, P_i} = C$ , for that would make  $\pi_1(a_\tau^{P_i, P_i}) - \pi_1(C, a_\tau^{P_i, P_i}) = 0 \forall P_l \in E_i(j)_\tau$ . For those histories  $h_t \in \hat{H}$  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  $\tilde{H} = \left\{ h_t \mid t > \tau, a_u = a_u^{P_i, P_j} \text{ for } u < \tau, a_{\tau,1} = C \text{ and } a_{u,2} = C, u > \tau \right\}$ , for which we also take  $R(h_t) = C, h_t \in \tilde{H}$ .

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  $\tau''$  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 weakly 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  $\Pi(R, P) \leq \Pi(P, P)$ .

## C The best responder

In order to be able to determine if a finite automaton - and hence a pure strategy - is a Nash equilibrium, we have constructed a small routine in the program called the best responder. This routine finds the payoff of the best response against strategy  $S$ , as well as a best response. 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 simply compare the payoff of  $S$  against itself to the payoff of all other strategies against  $S$  one after the other.

Suppose strategy  $S$  is an automaton with  $K$  states. Any state  $k$  is characterized by an action played by  $S$  when it finds itself in this state, and a list of transitions as a function of the action played by the opponent of  $S$ . With a slightly abusive notation -  $S$  is a function of histories elsewhere, while here it is easier to make it a function of states - we will write the first as  $S : \{1, \dots, K\} \rightarrow A$  and the latter as  $t_k : A \rightarrow \{1, \dots, K\}$ ,  $k = 1, \dots, K$ .

The value of being in state  $k$  is denoted by  $V(k)$ . We aim to find a solution to the following system:

$$V(k) = \max_{a \in A} \{ \pi_1(a, S(k)) + \delta V(t_k(a)) \} \quad k = 1, \dots, K$$

Let  $V^*(k), k = 1, \dots, K$  be the solution to this system. The discounted value in the initial state,  $(1 - \delta)V^*(1)$ , is the maximal discounted payoff to be earned against  $S$ , and  $a_k^* = \arg \max_{a \in A} \{ \pi_1(a, S(k)) + \delta V^*(t_k(a)) \}$  gives the optimal action when  $S$  is in state  $k$ .

The best responder does the following iteration.

Initialisation step:

$$V_1(k) = 0, \quad k = 1, \dots, K$$

Iteration step:

$$V_{n+1}(k) = \max_{a \in A} \{ \pi_1(a, S(k)) + \delta V_n(t_k(a)) \} \quad k = 1, \dots, K$$

It is quite straightforward that this iteration converges, as is shown in the following simple lemma. We will assume that the initialisation makes sure that we begin with values for all states that are below the solution of the system (whenever this procedure is invoked, we make sure that is in fact the case) but that is not actually necessary for convergence.

**Lemma 10** *If  $V_1(k) \leq V^*(k)$  for all  $k$  and if  $\delta \in [0, 1)$  then the above iteration converges to  $V^*(k)$ ,  $k = 1, \dots, K$ .*

**Proof.** First, if  $V_n(k) \leq V^*(k)$  for all  $k$ , then also

$$\begin{aligned} V_{n+1}(k) &= \max_{a \in A} \{ \pi_1(a, S(k)) + \delta V_n(t_k(a)) \} \\ &\leq \max_{a \in A} \{ \pi_1(a, S(k)) + \delta V^*(t_k(a)) \} = V^*(k) \text{ for all } k. \end{aligned}$$

Hence  $V^*(k) - V_n(k) \geq 0$  for all states  $k$  and all iterations  $n$ .

By definition we also have

$$V_{n+1}(k) \geq \pi_1(a_k^*, S(k)) + \delta V_n(t_k(a_k^*)) \text{ for all } k.$$

Therefore

$$0 \leq V^*(k) - V_{n+1}(k) \leq \delta (V^*(t_k(a_k^*)) - V_n(t_k(a_k^*))) \text{ for all } k.$$

This implies that

$$0 \leq \max_k (V^*(k) - V_{n+1}(k)) \leq \delta \max_k (V^*(k) - V_n(k))$$

and since  $\delta < 1$  we find that  $\lim_{n \rightarrow \infty} (V^*(k) - V_n(k)) = 0$  for all  $k$ . ■

The best responder gives us both the maximum payoff  $(1 - \delta) V^*(1)$  when playing against  $S$ , and an optimal strategy when playing against  $S$ , as  $a_k^*$  prescribes what to play when  $S$  is in state  $k$ . For numerical reasons, we actually use the latter. It is important for us to determine whether or not the payoff of  $S$  against itself is exactly equal to the maximum payoff when playing against  $S$ . The iteration with which the best responder finds the latter can have a numerical inaccuracy in it, and the evaluating the payoff of two given strategies

against each other (here:  $S$  against  $S$ ) can too. These are however different inaccuracies, so in order to have the *same* inaccuracies in both, we use the strategy that the best responder gives, first let it play that against  $S$ , then let  $S$  play against itself, and compare the two payoffs. If they are equal, then  $S$  is a best response to itself.

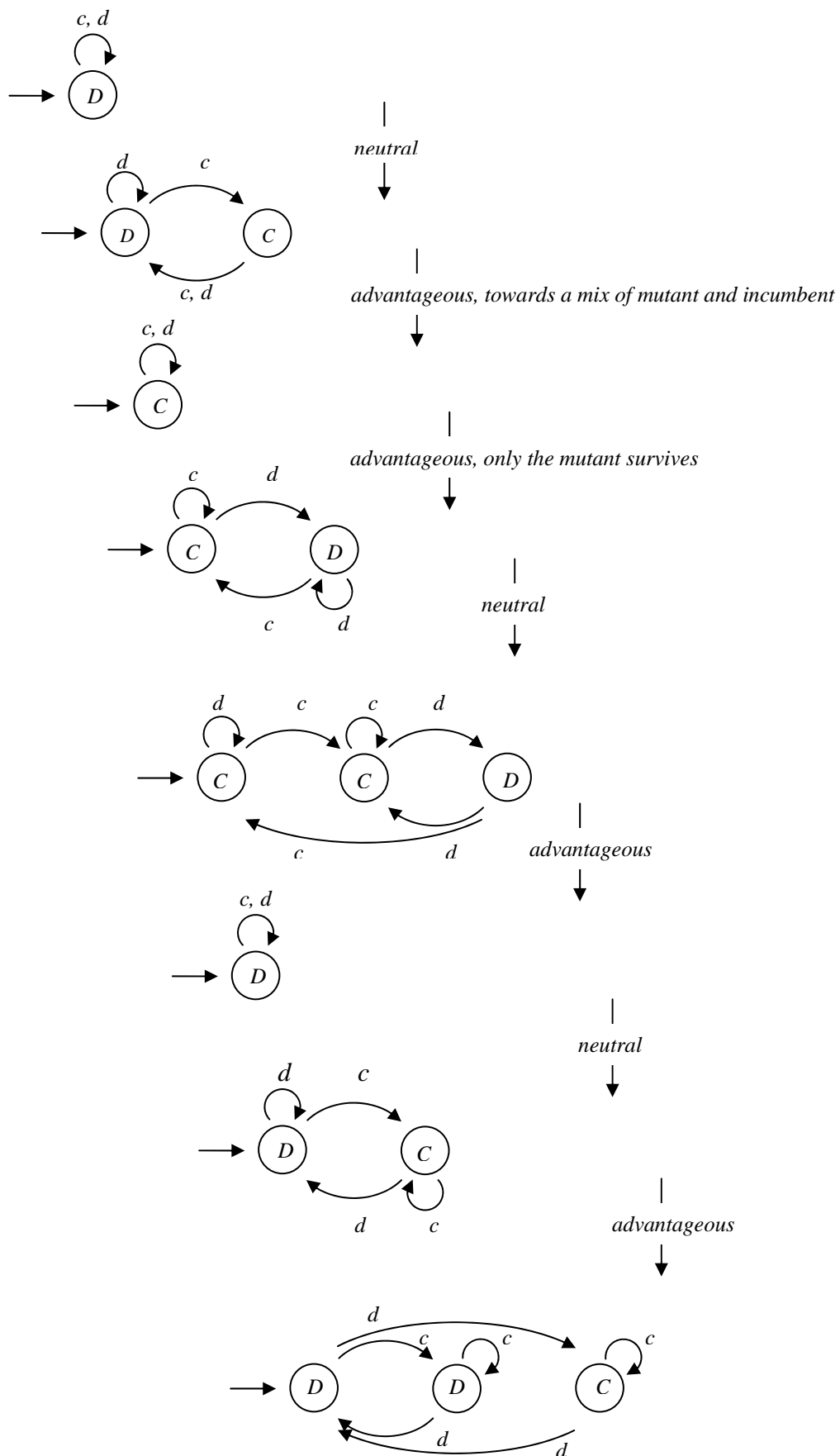
Note that the computer program works with phenotypes, not with genotypes, so two different ways to encode for instance the strategy  $AllD$  will be treated as one and the same strategy.

## D A typical sequence of strategies

The relevant payoffs for the sequence of strategies that go with the transitions indicated in Fig. 3 are given below. The population starts at All D. 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 is an equilibrium when the strategy is 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 #3 dominates mutant #2, and once #2 has disappeared, it dominates mutant #1, and goes to fixation. This establishes full cooperation. After this we get an indirect invasion back to All D, 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.

$$\begin{array}{ccccccc}
 1 & 1 & 4 & & & & \\
 1 & 1 & 3\frac{4}{7} & 2\frac{2}{7} & & & \\
 0 & 1\frac{2}{7} & 3 & 3 & & & \\
 & 1\frac{5}{7} & 3 & 3 & 3 & \frac{3}{4} & \\
 & & & 3 & 3 & 0 & \\
 & & & & 1\frac{3}{4} & 4 & 1 & 1 & 1\frac{6}{7} \\
 & & & & & & 1 & 1 & 1\frac{109}{175} \\
 & & & & & & \frac{4}{7} & 1\frac{12}{35} & 2\frac{1}{2}
 \end{array}$$

The actual automata are on the next page.



## E Measures of cooperativeness and reciprocity<sup>3</sup>

There are two reasons why we would like to have measures for cooperativeness and reciprocity. The first and most important reason is obvious; we would simply like to know how cooperative, and how reciprocal, strategies are. The second reason is that these measures may serve as an indication of indirect invasions. Typically an indirect invasion is characterized by a change in reciprocity followed by a change in cooperativeness.

Any measure of cooperativeness will have to weigh the different histories, and as we will see, every choice how to weigh them has appealing properties and drawbacks. In contrast to earlier definitions, here it is more natural to look at histories that only reflect what actions the *other* player has played. This captures all relevant histories for the measurement of cooperativeness, because what a strategy  $S$  itself has played is uniquely determined by the history of actions by the other player.

$$\begin{aligned}\bar{h}_1 &= () \\ \bar{h}_t &= (a_{1,2}, \dots, a_{t-1,2}), \quad t = 2, 3, \dots\end{aligned}$$

Again, we will sometimes also write  $(\bar{h}_t, a_{t,2})$  for a history  $\bar{h}_{t+1}$ , and we get the following sets of possible histories at time  $t$

$$\begin{aligned}\bar{H}_1 &= \{\bar{h}_1\} \\ \bar{H}_t &= \prod_{i=1}^{t-1} A \quad t = 2, 3, \dots\end{aligned}$$

With the repeated prisoners dilemma we have  $A = \{C, D\}$ , so in that case there are  $2^{t-1}$  histories  $\bar{h}_t \in \bar{H}_t$

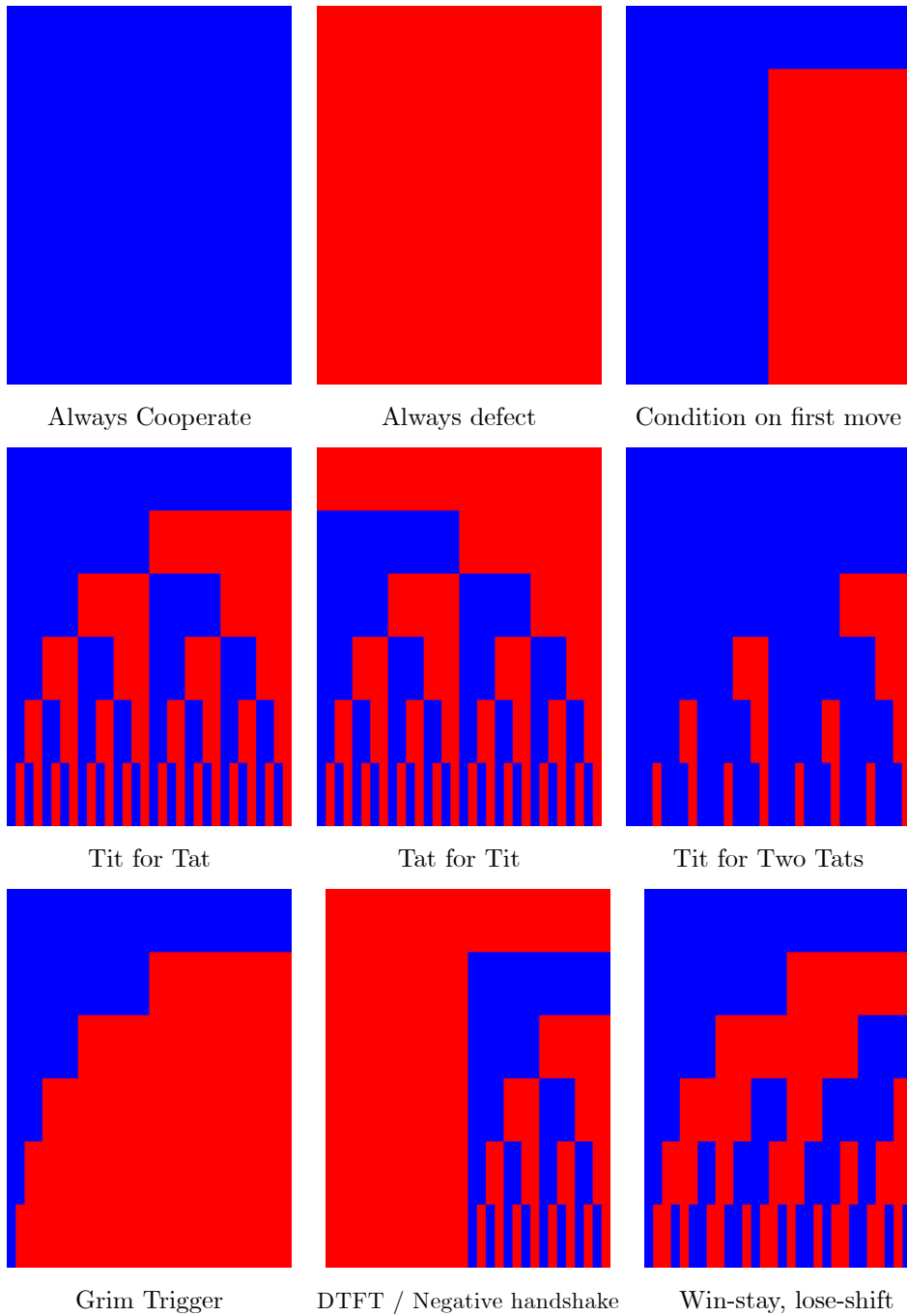
We begin with a measure that tells us how cooperative a strategy is, given that it is facing a history  $\bar{h}_t$ . If we weigh a history at time  $t + s$  with the probability that the game actually reaches round  $t + s - 1$ , given that it has already reached round  $t - 1$ , and if we also divide by the number of different histories of length  $t + s - 1$ , under the restriction that the first  $t - 1$  rounds of these histories are given by  $\bar{h}_t$ , we get the following. Note that this measure does not depend on the environment a strategy is in.

$$C(S, \bar{h}_t) = (1 - \delta) \sum_{s=0}^{\infty} \left(\frac{\delta}{2}\right)^s \left( \sum_{\substack{\bar{h} \in \bar{H}_{t+s} \\ \bar{h}^i = \bar{h}_t^i, i=1, \dots, t-1}} \mathbf{1}_{\{S(\bar{h})=C\}} \right)$$

The overall cooperativeness of a strategy  $S$  can then be defined as the cooperativeness at the beginning, where we have the empty history;  $C(S) = C(S, \bar{h}_1)$ .

---

<sup>3</sup>This is meant to only appear on [www.creedexperiment.nl/evolution](http://www.creedexperiment.nl/evolution), where one can run simulations and measure cooperativeness and reciprocity of the evolving population



An intuition for what this measure does can be gained from the above figure. The top bar represents the empty history. The second bar represents histories of length 1, and is split in two; the history where the other has cooperated, and the one where the other has defected. The third bar represents histories of length 2, and is split in four; the histories  $CC$ ,  $CD$ ,  $DC$  and  $DD$ . This continues indefinitely, but for the pictures we restrict ourselves to histories

of length 5 or less. If a part is blue, then that means that the strategy reacts to this history with cooperation, if it is red, then the strategy reacts with defection. Cooperativeness weighs the blueness of those pictures.

It is obvious that  $C(AllC) = 1$  and that  $C(AllD) = 0$ . A strategy that starts with cooperation, and further conditions play on the first move entirely has cooperativeness  $1 - \frac{\delta}{2}$ . This is sensible; if  $\delta = 0$ , then the first move is the only move, and since this strategy cooperates on the first move, it should have cooperativeness measure 1. On the other hand, except for the first move, this strategy cooperates in exactly half of the histories of length  $t$  for all  $t > 1$ . Therefore it makes sense that if  $\delta$  goes to 1, then cooperativeness goes to  $\frac{1}{2}$ .

More simple computations shows that  $C(TitForTat) = 1 - \frac{\delta}{2}$ ,  $C(TatForTit) = \frac{\delta}{2}$  and  $C(TitForTwoTats) = \frac{3}{4} + \frac{1-\delta^2}{4}$ . The last simple computation shows that  $C(GRIM) = \frac{2-2\delta}{2-\delta}$ , which is 1 at  $\delta = 0$  for similar reasons, and goes to 0 if  $\delta$  goes to 1.<sup>4</sup>

A measure for reciprocity can be constructed by comparing how much the cooperativeness of strategy  $S$  is changed if its opponent plays  $D$  rather than  $C$ . Again, histories of the same length are weighted equally here.

$$R(S) = \sum_{t=1}^{\infty} \sum_{\bar{h} \in \bar{H}_t} \left(\frac{\delta}{2}\right)^{t-1} [C(S, (\bar{h}, C)) - C(S, (\bar{h}, D))]$$

In the figure above, this is visualized as the difference in blueness below two neighbouring bits that share their history up to the one before last period.

Simple calculations now give that  $R(AllC) = R(AllD) = 0$  and  $R(TFT) = 1$ . Also, if we look at a strategy that only conditions on the first move and that defects forever if the first move of the other was  $D$ , and cooperates forever if the first move of the other was  $C$ , this strategy also has reciprocity 1. It is not hard to see that  $-1$  and  $+1$  are in fact the lower and upper bounds for reciprocity with this equal weighing of all strategies of the same length. Also  $R(GRIM) = \frac{4(1-\delta)}{(2-\delta)^2}$ . Note that here the reciprocity of Grim Trigger is lower than that of  $TFT$ , which is due to the fact that for many histories  $GRIM$  will only play a sequence of  $D$ 's either way.<sup>5</sup>

Alternatively we could measure the cooperativeness and reciprocity of a strategy given the population it is in. In that case, we should not weight all histories of a given length equally, but in the proportions in which they do occur given the actual population of

---


$$\begin{aligned}
{}^4 C(TFT) &= (1-\delta) \left(1 + \sum_{t=1}^{\infty} \frac{1}{2} \delta^t\right) = \frac{1}{2} (1-\delta) + \frac{1}{2} \\
C(GRIM) &= (1-\delta) \sum_{t=0}^{\infty} \left(\frac{1}{2} \delta\right)^t = (1-\delta) \frac{1}{1-\frac{\delta}{2}} = \frac{2-2\delta}{2-\delta} \\
{}^5 R(TFT) &= \sum_{t=1}^{\infty} \sum_{\bar{h} \in \bar{H}_t} \left(\frac{\delta}{2}\right)^{t-1} \left[\frac{1}{2} + \frac{1}{2} (1-\delta) - \left(\frac{1}{2} - \frac{1}{2} (1-\delta)\right)\right] 2^{t-1} = \sum_{t=1}^{\infty} \delta^{t-1} (1-\delta) = 1 \\
R(GRIM) &= \sum_{t=1}^{\infty} \left(\frac{\delta}{2}\right)^{t-1} \left[\frac{2-2\delta}{2-\delta} - 0\right] = \frac{4(1-\delta)}{(2-\delta)^2}
\end{aligned}$$



strategies. Hence the weight  $(\frac{1}{2})^{t-1}$ , which is one divided by the number of strategies in  $H_t$ , will then be replaced by their actual proportions. For instance, in a population that consists of any mixture of *AllC*, *TFT* and *GRIM*, the only history at time  $t$  that occurs, is a sequences of  $t-1$  consecutive  $C$ 's. The measure for cooperativeness then simply becomes the expected times a strategy plays  $C$  divided by the expected number of rounds.

Suppose the population is given by a vector of frequencies  $x = [x_1, \dots, x_N]$  where  $x_i$  is the frequency of strategy  $S_i$ . Then we define the population-dependent cooperativeness of a strategy  $S$  as follows:

$$C_x(S) = (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} \left( \sum_{i=1}^N x_i \mathbf{1}_{\{S(\bar{h}_t^{S,S_i})=C\}} \right)$$

In the population with only *AllC*, *TFT* and *GRIM*, cooperativeness of all these three strategies is 1. In an infinitely large population of  $\alpha$  *TFT* and  $1 - \alpha$  *AllD*, cooperativeness of *TFT* is  $1 - \delta + \delta\alpha$ .

A reasonable way of measuring reciprocity is to compare actual histories with histories that would have unfolded after one-step deviations. So let  $h_{t,a,s}^{S,T}$  be the history that for the first  $t-1$  steps unfolds recursively between strategy  $S$  and  $T$  - as above;

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

and the recursion step

$$\begin{aligned} a_i^{S,T} &= \left( S \left( h_i^{S,T} \right), T \left( h_i^{S,T \leftarrow} \right) \right) \\ h_{i+1}^{S,T} &= \left( h_i, a_i^{S,T} \right), \end{aligned} \quad i = 1, 2, \dots, t-1.$$

Only at time  $t$  the opponent plays  $a$ , while strategy  $S$  does not deviate and plays  $S \left( h_t^{S,T} \right)$ ;

$$\begin{aligned} a_t^{S,T} &= \left( S \left( h_t^{S,T} \right), a \right) \\ h_{t+1}^{S,T} &= \left( h_t, a_t^{S,T} \right). \end{aligned}$$

After that we go back to the normal recursion step

$$\begin{aligned} a_i^{S,T} &= \left( S \left( h_i^{S,T} \right), T \left( h_i^{S,T \leftarrow} \right) \right) \\ h_{i+1}^{S,T} &= \left( h_i, a_i^{S,T} \right), \end{aligned} \quad i = t+1, \dots, t+s$$

If  $T \left( h_i^{S,T \leftarrow} \right) = a$ , then this is  $h_{t,a,s}^{S,T}$  is just the actual history  $h_{t+1+s}^{S,T}$ , but if  $T \left( h_i^{S,T \leftarrow} \right) \neq a$ , it is a counterfactual history after a one-step deviation. The history  $\bar{h}_{t,a,s}^{S,T}$ , as above, just gives the actions of player  $T$  and ignores those of player  $S$ .

With this definition, we can make a measure for reciprocity as follows:

$$R_x(S) = (1 - \delta)^2 \sum_{t=1}^{\infty} \delta^{t-1} \sum_{i=1}^N x_i \sum_{s=0}^{\infty} \delta^s \left( \mathbf{1}_{\{S(\bar{h}_{t,C,s}^{S,S_i})=C\}} - \mathbf{1}_{\{S(\bar{h}_{t,D,s}^{S,S_i})=C\}} \right)$$

In a population of *GRIM* only, it is relatively easy to see that the reciprocity of *GRIM* is 1; the path of play between *GRIM* and itself is just a sequence of *C*'s, while after a deviation *GRIM* just plays a sequence of *D*'s. So at any time  $t$ , the discounted difference between these sequence from then on, normalised by multiplying by one of the  $(1 - \delta)$ 's, is 1. So discounting over  $t$  and normalizing by the other  $(1 - \delta)$  gives a reciprocity measure of 1.

In a population of *TFT* only, the reciprocity of *TFT* is  $\frac{1}{1+\delta}$ . Again, the path of play between *TFT* and itself is a sequence of *C*'s, but now, on the path after any deviation, *TFT* plays a sequence of alternating *D*'s and *C*'s. So at any time  $t$ , the discounted and normalized difference between them is  $(1 - \delta) \left( \frac{1}{1-\delta} - \frac{\delta}{1-\delta^2} \right) = \frac{1}{1+\delta}$ . So discounting over  $t$  and normalizing by the other  $(1 - \delta)$  gives the same number.

This comparison is in line with what we might want from a reciprocity measure that depends on the actual population; at any stage, the threat that grim trigger poses to itself is maximal, and larger than the threat *TFT* poses to itself.

Another comparison shows that this measure also picks up the fact that, if the punishment in Grim Trigger is in fact triggered on the path of play, then on the remainder of the path, *GRIM* is actually not reciprocal at all anymore. With *TFT* on the other hand, reciprocity actually remains the same, whether the punishment has been triggered in the past or not. This is reflected in the following example.

In a population of  $\alpha$  *TitForTat* and  $(1 - \alpha)$  *TatForTit*, the reciprocity of *TitForTat* is  $\alpha \frac{1}{1+\delta} + (1 - \alpha) \frac{1}{1+\delta} = \frac{1}{1+\delta}$ ; the reciprocity of *TitForTat* against itself was computed above, and for the computation of *TitForTat* against *TatForTit* it is enough to realize that the comparison between actual and counterfactual is between *DCDC...* versus *CCCC...* at odd  $t$ 's and between *CDCD...* versus *DDDD...* at even  $t$ 's.

In a population of  $\alpha$  *GRIM* and  $1 - \alpha$  *TatForTit*, the reciprocity of *GRIM* is  $\alpha + (1 - \alpha)(1 - \delta) = 1 - \delta + \alpha\delta$ ; in the latter interaction, *GRIM* only alters its behaviour in response to a change at  $t = 1$ . So for not too large  $\alpha$  - implying that there is enough *TatForTit* to have a noticeable effect of the punishment being triggered - and not too small  $\delta$ , *GRIM* now is actually less reciprocal than *TitForTat*.

Dependence on the population a strategy finds itself in can be seen as a good or a bad thing. For picking up indirect invasions, it seems to be a good thing; changes "far off" the current path of play between strategies do not change this reciprocity measure; it only changes if a strategy mutates into one that reacts differently to a one-step deviation. This implies that only becoming more or becoming less reciprocal in the way that is relevant for indirect invasions is picked up. On the other hand, it does not lend itself for a general,

environment-independent statement of how reciprocal or cooperative a strategy is.

Another possibility is to only look at the actions directly after a one-step deviation. Then we would get

$$R_x(S) = (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} \sum_{i=1}^N x_i \left( \mathbf{1}_{\{S(\bar{h}_t^{S, S_i}, C)=C\}} - \mathbf{1}_{\{S(\bar{h}_t^{S, S_i}, D)=C\}} \right)$$

Here the reciprocity of *AUC* is 0 again. The reciprocity of both *TFT* and *GRIM* in a population of *TFT*, *GRIM* and *AUC* is 1 here.