Supplementary Information:
Evolutionary processes that resolve cooperative dilemmas

Contents

1 General model **2**

1.1 The basic selection process 2

1.2 Memory-1 strategies of direct reciprocity 2

1.3 Stable states 5

1.4 Stable states in a simple invasion process 6

1.5 Parity and positivity 7

2 Linear decision functions **10**

2.1 Constrained parity violation 11

2.2 Larger parity violation 13

2.3 Comparison with M0 strategies 16

2.4 Equal gains from switching 17

2.5 Reduction 18

2.6 Sufficient conditions for linearity 21

3 Extensions **22**

3.1 Differentiable decision functions 22

3.2 Longer memory 24

1 General model

1.1 The basic selection process

We study two-player symmetric games which are associated with a set (or space) of strategies, $\mathcal{S}$. If player 1 uses strategy $A \in \mathcal{S}$ and player 2 uses strategy $B \in \mathcal{S}$, then player 1 is awarded payoff π_{AB}, and player 2 is awarded payoff π_{BA}.

We study evolutionary processes (or tournaments) on the strategy space $\mathcal{S}$. A process is indexed by discrete time steps. At each time step, there is a resident strategy A, which is the current state. A challenger strategy B is drawn from $\mathcal{S}$ at random. The payoff matrix between resident and challenger, $[(\pi_{AA}, \pi_{AB}), (\pi_{BA}, \pi_{BB})]$, can be written in shorthand as $[(a, b), (c, d)]$. That is,

$$a := \pi_{AA}, \;\; b := \pi_{AB}, \;\; c := \pi_{BA}, \;\; d := \pi_{BB} \tag{1}$$

Our selection process stipulates that A remains as the resident if and only if

$$U(a, b, c, d) > 0 \tag{2}$$

Otherwise, if $U(a, b, c, d) \leq 0$, then A is discarded and B becomes the new resident state. The procedure is repeated over many time steps. U is an arbitrary fixed function of four variables, which we assume to be continuous. We refer to U as the decision function of the selection process. In the following, we will blur the distinction between the process and the corresponding decision function.

We note that for generic processes, games and strategy spaces, perfect equality $U(a, b, c, d) = 0$ occurs with probability zero.

1.2 Memory-1 strategies of direct reciprocity

As an interesting application of the selection process described above, we take the space $\mathcal{S}$ to be a space of strategies from the setting of direct reciprocity. More specifically, we consider strategies for repeated 2×2 symmetric games. The payoff matrix for such a game is of the form

$$\begin{array}{c|cc} & \text{C} & \text{D} \\ \hline \text{C} & R & S \\ \text{D} & T & P \end{array} \tag{3}$$

Without loss of generality we label the actions 'C' and 'D' such that the payoff to two 'C' players is greater than the payoff to two 'D' players: $R \geq P$. In fact, we will assume $R > P$, unless we explicitly say otherwise. We sometimes study many different payoff matrices at once

— each is specified by a four-tuple $G = (R, S, T, P)$. We follow the literature on the evolution of cooperation, and refer to the action 'C' as 'cooperate' and 'D' as 'defect.'

In each round of the repeated game, the average of the two players' payoffs is either R (if both played C); P (if neither played C); or $(S + T)/2$ (if exactly one played C). Therefore, the maximum possible per-round average payoff to two players in the repeated game is

$$\pi_M := \max\{R, P, (S + T)/2\} = \max\{R, (S + T)/2\} \tag{4}$$

The minimum possible per-round average payoff to two players in the repeated game is

$$\pi_m := \min\{R, P, (S + T)/2\} = \min\{P, (S + T)/2\} \tag{5}$$

There are many strategy spaces of direct reciprocity which we could consider. Memory-n strategies are strategies which choose the next move based on the outcome of the previous n rounds. The outcome of each round is an element of {CC, CD, DC, DD}. This means there are 4^n possible outcomes for a sequence of n consecutive rounds. For each such outcome, a memory-n strategy specifies a probability for playing C. The space Mn of memory-n strategies is 4^n-dimensional. The simplest of these spaces is given by memory-0 (M0) strategies. An M0 strategy specifies a constant probability p for playing C in the next round.

In this paper, we focus on the space of memory-1 (M1) strategies (Nowak & Sigmund 1990, 1992). In the last section, we prove that almost all our results carry over to memory-n (Mn) strategies for $n > 1$. An M1 strategy $\mathbf{p} = (p_{\text{CC}}, p_{\text{CD}}, p_{\text{DC}}, p_{\text{DD}})$ is given by four probabilities p_{ij} — the probability to play C next if the focal player played i in the last round and the co-player played j. Usually we use the shorthand $\mathbf{p} = (p_1, p_2, p_3, p_4)$.

There is a good reason for choosing M1 as the strategy space when studying direct reciprocity. Regardless of the game parameters R, S, T, P, there is always an M1 strategy $\mathbf{p}$ such that if both players use $\mathbf{p}$, they each earn long-run average payoff π_M. For example, if $\pi_M = R$, then $\mathbf{p} = (1, 1, 1, 1)$ suffices. This is the strategy that always plays C. If $\pi_M = (S + T)/2$, then $\mathbf{p} = (1/2, 0, 1, 1/2)$ suffices. This is a strategy for which the two players jointly play CD and DC in alternating rounds, each player earning $\pi_M = (S + T)/2$ per-round on average.

By contrast, no M0 strategy can achieve CD and DC outcomes in alternating rounds. Consequently, if $\pi_M = (S + T)/2 > R$, there is no M0 strategy p for which the two players earn π_M per-round on average. This means that M1 is the smallest space Mn for which there is always a strategy which achieves the maximum possible self-payoff (average per-round payoff against itself in the repeated game). We are interested in finding evolutionary selection processes U for which the long term average payoff achieves the maximum value. This is a good reason to consider the space of M1 strategies.

If the focal player uses an M1 strategy $\mathbf{p}$ and the co-player uses an M1 strategy $\mathbf{q}$, then the outcome of the game may change randomly from round to round: indeed, there is an induced Markov chain, $M(\mathbf{p}, \mathbf{q})$, on the set of round outcomes {CC, CD, DC, DD}. A stationary distribution of the Markov chain describes the long-run average frequencies of the four outcomes. This means that a stationary distribution can be used to calculate the long-run average per-round payoff of $\mathbf{p}$ against $\mathbf{q}$, which we write as $\pi_{\mathbf{pq}}$. However, if the stationary distribution of the Markov chain is not unique, then the payoff $\pi_{\mathbf{pq}}$ is not uniquely defined. In that case, it depends on the starting moves of both players.

In an attempt to avoid some thorny technicalities in the future, we make the following definition:

Definition 1. *A memory-*1 *strategy* $\mathbf{p}$ *is called* ***eligible*** *if the Markov chain* $M(\mathbf{p}, \mathbf{p})$ *admits a unique stationary distribution.*

Proposition 2. *If* $\mathbf{p}$ *is eligible, then* $\pi_{\mathbf{pp}}$ *is uniquely defined for any* (R, S, T, P)*. Likewise,* $\pi_{\mathbf{pq}}$ *and* $\pi_{\mathbf{qp}}$ *are uniquely defined for any interior memory-*1 *strategy* $\mathbf{q}$*, and any* (R, S, T, P)*. These payoffs are given by the formula (6) of Press-Dyson.*

Proof. The first statement is true by the definition of $\pi_{\mathbf{pp}}$. The second statement is true for the following reason. Since $M(\mathbf{p}, \mathbf{p})$ has a unique stationary distribution, it has a unique closed irreducible set of states. (A closed irreducible set is a collection of states such that there is a positive probability of eventual transition between any two states in the set, and zero probability of eventual transition from a state inside the set to a state outside of the set.) Let $\mathbf{q}$ be an interior memory-1 strategy. Then every transition which occurs with positive probability in the Markov chain $M(\mathbf{p}, \mathbf{p})$, also occurs with positive probability in the Markov chain $M(\mathbf{p}, \mathbf{q})$. So $M(\mathbf{p}, \mathbf{q})$ also has a unique closed irreducible set of states, and hence a unique stationary distribution. □

Proposition 3. *Every memory-*1 *strategy* $\mathbf{p} = (p_1, p_2, p_3, p_4)$ *is an eligible strategy, except for the following five strategy types, which are not:*

1. *the square* $\mathbf{p} = (1, p_2, p_3, 0)$
2. *the square* $\mathbf{p} = (p_1, 1, 0, p_4)$
3. *the line* $\mathbf{p} = (1, 0, 1, p_4)$
4. *the line* $\mathbf{p} = (p_1, 0, 1, 0)$
5. *the point* $\mathbf{p} = (0, 0, 1, 1)$

Proof. The condition of being eligible says that the chain $\mathbf{M}(\mathbf{p},\mathbf{p})$ has a unique stationary distribution. It is a well-known fact that a finite Markov chain has a unique stationary distribution if and only if it has a unique closed irreducible set of states. The closed irreducible sets are determined by the signs of all of the state transitions (either positive or zero). In turn, these signs are determined by whether each p_i is $0, 1$, or somewhere in the open interval $(0, 1)$. In other words, there are three relevant possibilities for each of p_1, p_2, p_3, p_4. In total the number of strategy types needing to be considered is $3^4 = 81$. One can check by hand that the possibilities listed above are the ones which do not give rise to a unique closed irreducible set. □

From now on we generally assume that a resident M1 strategy for a selection process is an eligible strategy, unless specified otherwise. This usually guarantees that payoffs are well-defined.

For interior strategies $\mathbf{p}$ and $\mathbf{q}$, which are defined by $p_i, q_j \in (0, 1)$ for all i, j, the payoff $\pi_{\mathbf{pq}}$ is given by the formula (Press & Dyson 2012)

$$\pi_{\mathbf{pq}} = \frac{\det M(R, S, T, P)}{\det M(1, 1, 1, 1)}, \tag{6}$$

$$M(v_1, v_2, v_3, v_4) = \begin{pmatrix} -1 + p_1 q_1 & -1 + p_1 & -1 + q_1 & v_1 \\ p_2 q_3 & -1 + p_2 & q_3 & v_2 \\ p_3 q_2 & p_3 & -1 + q_2 & v_3 \\ p_4 q_4 & p_4 & q_4 & v_4 \end{pmatrix} \tag{7}$$

This formula also works in many cases where $\mathbf{p}$ or $\mathbf{q}$ are boundary strategies, which have p_i or q_j in $\{0, 1\}$ for some i or j. For instance, the formula is valid for any eligible strategy $\mathbf{p}$ and interior strategy $\mathbf{q}$.

The value of $\pi_{\mathbf{pp}}$ will simply be referred to as the payoff of $\mathbf{p}$. It is the average payoff received by each player when both use the strategy $\mathbf{p}$.

1.3 Stable states

The lifetime of a resident state in a selection process is the number of rounds before it is successfully replaced by a random challenger. The expected lifetime $\ell_U(A)$ of a resident state A can be calculated as follows:

$$\ell_U(A) := (1 - |\{B \,:\, U(a, b, c, d) > 0\}|)^{-1}$$

Here $|\;|$ represents the Lebesgue measure, and the set $\{B \,:\, U(a, b, c, d) > 0\}$ is the set of challenger strategies which cannot replace A.

Definition 4. *A resident state A is called a stable state for the process U if $\ell_U(A) = \infty$.*

A stable state A has an infinitely long expected lifetime. This occurs when the probability of a random challenger being able to replace A is zero. Notice that to determine whether a state is stable, it suffices to examine only interior strategies B (since boundary strategies have measure zero).

Our goal in defining stable states is that by acquiring analytical knowledge of these states we can come to understand the results of simulations of the selection process on the space M1.

Proposition 5. *Let A be a stable state for the process U. Then all states sufficiently close to A have a long expected lifetime.*

Proof. First, note that the set of all eligible states is open. So the expected lifetime ℓ_U is well-defined in a sufficiently small neighborhood of A.

Since A is stable, $\ell_U(A) = \infty$. Equivalently, the set $X = \{B : U(a,b,c,d) > 0\}$ of challengers which cannot replace A, is an open set of measure 1.

Let $L > 0$ be arbitrary. Pick a compact set $K \subseteq X$, with measure at least $1 - 1/L$. Since $K \subseteq X$ is compact and U is continuous, $\min_{B \in K} U(a,b,c,d) > 0$.

Now consider another resident strategy A', and define the shorthand

$$a' := \pi_{A'A'},\ b' := \pi_{A'B},\ c' := \pi_{BA'},\ d' := \pi_{BB}$$

If A' is sufficiently close to A, then $\min_{B \in K} U(a',b',c',d') > 0$. In that case

$$\ell_U(A') := (1 - |\{B : U(a',b',c',d') > 0\}|)^{-1} \geq (1 - |K|)^{-1} = L$$

So all strategies sufficiently close to A have lifetime at least L. Here L was arbitrary. □

1.4 Stable states in a simple invasion process

In this section we give an extended example. Consider the selection process $U(a,b,c,d) > a - c$. Here the resident, A, remains if the invasion fitness of the mutant, B, which is $c - a := \pi_{BA} - \pi_{AA}$, is negative. We call U the invasion process.

If A is a *strict Nash equilibrium*, then $\pi_{AA} > \pi_{BA}$ for all strategies B other than A. Thus, a strict Nash equilibrium is a stable state in the invasion process. No challenger strategy B can replace A. When asking for a strict NE in the space M1 we are asking for too much. These strategies do not typically exist, because one can find a strategy B with $\pi_{AA} = \pi_{BA}$. For instance, if A is the M1 strategy generous-tit-for-tat $(1, 1/4, 1, 1/4)$, in a standard donation game with costto-benefit ratio of $1/2$, then $\pi_{AA} \geq \pi_{BA}$ for all B where the payoffs are well-defined. But if B is ALLC $= (1,1,1,1)$ then B attains equality $\pi_{AA} = \pi_{BA}$.

If A is a ***Nash equilibrium***, then $\pi_{AA} \geq \pi_{BA}$ for all strategies B. When asking for a NE in the space M1 we are not asking for enough. For example, some NE strategies are equalizers[1,2]: $\pi_{AA} = \pi_{BA}$ for all B. If a resident strategy A is an equalizer, then we can say nothing about the expected lifetime of resident strategies which are very close to A.

Therefore, strict Nash equilibrium and Nash equilibrium strategies are not perfectly suited to understanding simulations of the invasion process on M1. However, the notion of a stable state is very similar to both of these concepts. It helps shed light on simulations by virtue of Proposition 5.

1.5 Parity and positivity

It is natural to consider selection processes $U(a,b,c,d)$ which implement a symmetric comparison between resident A and challenger B. The decision function U can be thought of as making a decision between the two strategies (based on the sign of $U(a,b,c,d)$) with a certain intensity (based on the magnitude of $U(a,b,c,d)$). A reasonable symmetry condition is that the decision and intensity is the same regardless of which strategy is in the role of resident and which is in the role of mutant. We call this condition *parity*:

$$U(d,c,b,a) = -U(a,b,c,d) \tag{8}$$

If we want to design a selection process which selects for high payoff, a second basic condition we should impose is that having a larger payoff $a = \pi_{AA}$ should contribute positively to the selection of the resident. We call this condition *positivity*: formally, it means that U is strictly monotonically increasing in the first variable.

Unexpectedly, the two conditions of parity and positivity are already enough to result in effective selection for high payoff. Our simulations have shown examples of this curious fact. It is also predicted by the following two propositions.

Proposition 6 (Parity and positivity)**.** *Let U be a selection process which satisfies both parity and positivity. If this process is applied to the space M1 for any game (R,S,T,P), then every stable state A of the process has maximum payoff π_M.*

Proof. By positivity, U is strictly monotonically increasing in the first variable. But by parity, that implies U is strictly monotonically decreasing in the fourth variable.

<u>Case 1:</u> $\pi_M = R$. Let A be an eligible memory-1 strategy and let

$$B(\epsilon) = (1 - \epsilon^2, \epsilon, 1 - \epsilon, \epsilon) \tag{9}$$

One can use the payoff formula (6) to compute π_{BB} and check that

$$\lim_{\epsilon \to 0} \pi_{BB} = R = \pi_M \tag{10}$$

In the same way it may be checked that the quantity $x = \lim_{\epsilon \to 0} \pi_{AB}$ exists and is equal to $\lim_{\epsilon \to 0} \pi_{BA}$. So, by continuity of U,

$$\lim_{\epsilon \to 0} U(a, b, c, d) = U(a, x, x, \pi_M)$$

Parity implies that $U(a, x, x, \pi_M) = -U(\pi_M, x, x, a)$. We have observed that U is increasing in the first variable and decreasing in the fourth. It follows that if $a < \pi_M$, then $U(a, x, x, \pi_M) < 0$. We can also say that $U(a, b, c, d) < 0$ for sufficiently small ϵ. So the selection process stipulates that resident A is replaced by challenger B, if ϵ is small enough. By continuity, there are also strategies nearby to $B(\epsilon)$ which can replace A. So A cannot be a stable state unless $\pi_{AA} =: a = \pi_M$

Case 2: $\pi_M = (S + T)/2$. The same argument works, except that we use $B(\epsilon) = (1 - \epsilon, \epsilon^2, 1 - \epsilon^2, \epsilon)$. □

Proposition 6 asserts that every stable state has maximum payoff. However, we would also like to know that such stable states exist at all.

Proposition 7. *Let U be a selection process which satisfies both parity and positivity. If this process is applied to the space M1 for any game (R, S, T, P), then the process has a stable state.*

Remark. Technically this proof will show that there is a family of states (with maximum payoff) which approach lifetime $+\infty$ in the limit. The limiting state itself is not eligible, because approaching it from different directions gives different payoff results. Nevertheless, it can be considered a kind of extended stable state, when approached from a specified direction.

Proof. To check that a state A is stable, it suffices to show that $U(a, b, c, d) > 0$ for all challengers B which are interior strategies.

Case 1: $\pi_M = R$. Let $A(\epsilon) = (1, \epsilon, 1, \epsilon)$. One can use the payoff formula (6) to compute π_{AA} and check that

$$\lim_{\epsilon \to 0} \pi_{AA} = R = \pi_M \tag{11}$$

Similarly, one can check that for every interior strategy B, the limit $x = \lim_{\epsilon\to 0} \pi_{AB}$ exists and is equal to $\lim_{\epsilon\to 0} \pi_{BA}$. It follows by continuity of U that

$$\lim_{\epsilon\to 0} U(a,b,c,d) = U(\pi_M, x, x, d)$$

Since B is an interior strategy, $\pi_{BB} =: d < \pi_M$. Now, by the logic in the proof of Proposition 6, we can conclude that $U(a,b,c,d) > 0$ for all sufficiently small ϵ. So we will say that $(1,\epsilon,1,\epsilon)$ is a stable state, where ϵ is infinitesimally small.

Case 2: $\pi_M = (S+T)/2$. The same argument carries over, except that we use $A(\epsilon) = (1-\epsilon, 0, 1, \epsilon)$. □

General selection processes exhibit various asymmetries, and usually do not satisfy parity. For instance, a decision function

$$U(a,b,c,d) = \exp(a + 2b - 4c + d) - 1, \tag{12}$$

does not.

One of our goals is to show that if the selection process U exhibits a certain degree of parity violation, then there must exist a game (R,S,T,P) for which the stable states A do not achieve payoff π_M (or else do not exist).

For simplicity we will only consider selection processes which satisfy the following:

Definition 8. *A selection process is centered if*

$$U(\pi,\, \pi,\, \pi,\, \pi) = 0, \tag{13}$$

for all $\pi \in \mathbb{R}$. In other words, the decision function reaches the point of indifference (0) if the variables a,b,c,d are all equal to each other.

Every function which satisfies parity is centered. The example (12) above is also centered.

2 Linear decision functions

A linear decision function U is one of the form

$$U_{\mathbf{x}}(a,b,c,d) := x_1 a + x_2 b + x_3 c + x_4 d \tag{14}$$

Here the vector $\mathbf{x} = (x_1, x_2, x_3, x_4)$ determines the decision function and the corresponding selection process. For example, the invasion process $U(a,b,c,d) = a - c$ which we have previously discussed is specified by $\mathbf{x} = (1, 0, -1, 0)$. This process replaces the resident whenever the challenger has a positive invasion fitness.

There is also a "risk-dominance process" which is defined by $\mathbf{x} = (1, 1, -1, -1)$. This process selects the strategy which is risk-dominant: that is, A is selected if and only if $a + b > c + d$. A third process is the so-called "nirvana" process which is defined by $\mathbf{x} = (1, 0, 0, -1)$. This process simply selects the strategy which has a higher payoff: A if $a > d$, and B otherwise.

A linear process $U_{\mathbf{x}}$ is centered according to Definition 8, if and only if it has $x_1 + x_2 + x_3 + x_4 = 0$. In this case, it has a further additive invariance property: for any constant K, we have

$$U_{\mathbf{x}}(a,b,c,d) = U_{\mathbf{x}}(a+K, b+K, c+K, d+K) \tag{15}$$

From now on, when we write $U_{\mathbf{x}}$, we assume that $x_1 + x_2 + x_3 + x_4 = 0$ unless specified otherwise.

The function $U_{\mathbf{x}}$ satisfies positivity if and only if $x_1 > 0$. In this case, we can normalize $\mathbf{x}$ so that $x_1 = 1$.

The resulting family of selection processes have linear decision function $U_{\mathbf{x}}$ where $\mathbf{x} = \mathbf{x}(\alpha, \beta)$ is the four-tuple defined by

$$\mathbf{x}(\alpha,\beta) := (1, \alpha, -\beta, -1-\alpha+\beta) \tag{16}$$

These decision functions are parameterized by α and β. We will sometimes refer to this as the $\alpha\beta$-plane of linear processes. We will devote substantial attention to understanding this subset of selection processes. First, we identify some important representatives:

1. The decision function $U_{(0,0)}(a,b,c,d) = a - d$ defines the "nirvana" process.
2. The decision function $U_{(1,1)}(a,b,c,d) = b - c$ defines a process in which the resident and challenger strive for a higher payoff against the other. We call this the 'direct competition' process.
3. The decision function $U_{(0,1)}(a,b,c,d) = a - c$ defines the invasion process.

4. The decision function $U_{(1,1)}(a,b,c,d) = a+b-c-d$ defines the risk-dominance process.

The linear selection processes which satisfy both parity and positivity are the ones of the form $U_{(\alpha,\alpha)}$. As a corollary of Propositions 6-7, we can say

Proposition 9. *Let $\alpha \in \mathbb{R}$. If the selection process $U_{(\alpha,\alpha)}$ is applied to the space M1, then:*

1. *Every stable state has maximum payoff π_M.*
2. *There is a stable state.*

In other words, when $\alpha = \beta$, the selection process $U_{(\alpha,\beta)}$ always favors maximum payoff. These processes lie on the diagonal of the $\alpha\beta$-plane.

2.1 Constrained parity violation

The parity condition for a process U can be framed as

$$U(a,b,c,d) + U(d,c,b,a) = 0 \tag{17}$$

For the process $U_{(\alpha,\beta)}$, this condition becomes

$$a+\alpha b-\beta c+(-1-\alpha+\beta)d+d+\alpha c-\beta b+(-1-\alpha+\beta)a = (\beta-\alpha)(a-b-c+d) \tag{18}$$

So the quantity $\beta - \alpha$ is a measure of the parity violation of a process $U_{(\alpha,\beta)}$.

From simulations, we have actually observed that there is a band $0 \le \beta - \alpha < 1$ of constrained parity violation, for which the selection process $U_{(\alpha,\beta)}$ always favors states with maximum payoff π_M. We refer to this as the L-band, since it appears to the left of (or above) the diagonal.

Every process in the L-band has an interpretation as follows:

1. The resident is assigned a linear "score" $S_A = a + g_1 b + g_2 c$, where $g_1, g_2 \in \mathbb{R}$. This score depends on a, b, c, and consequently depends on the challenger B.
2. The challenger is assigned a score $S_B = (1-2h)d + (g_1+h)c + (g_2+h)b$. In other words, the challenger's score has the same functional form as the resident's score, but subject to a small deformation parameterized by h. Here $0 \le h \le \frac{1}{2}$.

3. The decision function $U(a,b,c,d) = S_A - S_B$ is a comparison of scores. We have

$$\begin{aligned} U(a,b,c,d) &= S_A - S_B \\ &= a + g_1 b + g_2 c - (1-2h)d - (g_1+h)c - (g_2+h)b \\ &= a + (g_1 - g_2 - h)b + (g_2 - g_1 - h)c + (-1+2h)d \\ &= U_{(\alpha,\beta)}(a,b,c,d) \end{aligned}$$

where $\alpha = g_1 - g_2 - h$ and $\beta = g_1 - g_2 + h$. Note that $\beta - \alpha = 2h \in [0,1]$. So, by picking $g_1 - g_2$ and h appropriately, we can generate an arbitrary decision function $U_{(\alpha,\beta)}$ which is in the desired band.

The stochastic behavior of such processes is difficult to predict theoretically, but we have a result which seems to provide some explanation for simulation results.

Proposition 10. *Let $\alpha, \beta \in \mathbb{R}$ with $0 \le \beta - \alpha \le 1$. The selection process $U_{(\alpha,\beta)}$, applied to the space M1, admits a stable state with payoff π_M.*

Proof. Case 1: $\pi_M = R$. Let $A(\epsilon) = (1, \epsilon, 1, \epsilon)$ be a resident memory-1 strategy. Let B be a fixed interior challenger strategy. We mentioned in the proof of Proposition 6 that $\pi_{AA} = R = \pi_M$ and that $\lim_{\epsilon\to 0} \pi_{AB} = \lim_{\epsilon \to 0} \pi_{BA} =: x$. We claim that $x < \pi_M$. The reason is that $\pi_{AB} + \pi_{BA}$ is the sum of the two players' per-round average payoffs, which must be at most the maximum per-round total payoff $2R = 2\pi_M$. If $x = \pi_M$, then the sum of the two players' per-round average payoffs is equal to the maximum per-round total payoff $2R$ in the limit $\epsilon \to 0$. This is only possible if both players always play C in the limit $\epsilon \to 0$; that will not happen, since B is a fixed interior strategy. Now we have

$$\begin{aligned} \lim_{\epsilon \to 0} U_{(\alpha,\beta)}(a,b,c,d) &= \lim_{\epsilon \to 0} (a + \alpha b - \beta c + (-1 + \alpha - \beta)d) \\ &= (\pi_M - \pi_{BB}) + (\alpha - \beta)(x - \pi_{BB}) \end{aligned}$$

Now if $\beta - \alpha = 0$, then the limit is $\pi_M - \pi_{BB} > 0$. On the other hand if $\beta - \alpha = 1$, then the limit is $\pi_M - x > 0$. It follows by linearity that if $\beta - \alpha \in [0,1]$, the limit is positive. This shows that $A = (1, \epsilon, 1, \epsilon)$ is a stable state if ϵ is considered to be infinitesimal.
Case 2: $\pi_M = (S+T)/2$. The proof is nearly identical to the proof of Case 1, but with $A(\epsilon) = (1-\epsilon, 0, 1, \epsilon)$ instead. □

The proposition above establishes interesting behavior in L-band: the selection process can find stable states with maximum self-payoff π_M.

Above and below this band, we do not have the same behavior. The goal of the next section is to examine these regimes.

2.2 Larger parity violation

In the next two propositions, we first consider a fixed process with decision function $U_{(\alpha,-\alpha)}$. This process lies inside of the L-band if $-\frac{1}{2} \leq \alpha \leq 0$, and outside of the L-band if $\alpha < -\frac{1}{2}$ or $\alpha > 0$.

Proposition 11. *Suppose $\alpha > 0$. There exists a game (R, S, T, P) for which the selection process with decision function $U_{(\alpha,-\alpha)}$ has no stable state.*

Proof. Suppose the process has a stable state A. In that case, let B be an arbitrary interior memory-1 strategy. Since A is a stable state, we have

$$\begin{aligned} 0 < U_{(\alpha,-\alpha)}(a, b, c, d) &= a + \alpha(b + c) + (-1 - 2\alpha)d \\ &= (a - d) + \alpha(b + c - 2d) \end{aligned}$$

We claim that $a := \pi_{AA}$ is equal to π_M.
<u>Case 1: $\pi_M = R$</u>. Suppose $B(\epsilon) = (1 - \epsilon^2, \epsilon, 1 - \epsilon, \epsilon)$. As we mentioned in the proof of Proposition 6, $\lim_{\epsilon\to 0} \pi_{BB} = R = \pi_M$, and one can check that $\lim_{\epsilon\to 0} \pi_{BA} = \lim_{\epsilon\to 0} \pi_{AB} := x$ for any eligible memory-1 strategy A. Furthermore, we must have $x \leq \pi_M$ since π_M is the largest total payoff to both players which is achievable in a single round. So then,

$$0 \leq \lim_{\epsilon\to 0} U_{(\alpha,-\alpha)}(a, b, c, d) = (a - \pi_M) + 2\alpha(x - \pi_M) \leq a - \pi_M$$

It follows that $a = \pi_M$. We conclude that every stable state has payoff π_M.
<u>Case 2: $\pi_M = (S + T)/2$</u>. The conclusion is the same as for Case 1. The proof is the same except that we must use $B(\epsilon) = (1 - \epsilon, \epsilon^2, 1 - \epsilon^2, \epsilon)$.

In either case, every stable state achieves self-payoff π_M.

But now, consider the game $(R, S, T, P) = (1, 0, 4, 0)$, that is, the game with payoff matrix

	C	D
C	1	0
D	4	0

(19)

The maximum payoff is $\pi_M = 2$. The states which have this payoff are of the form $\mathbf{p} = (p_1, 0, 1, p_4)$. They are called self-alternators or just alternators[3], since they alternate between the outcomes CD and DC when playing against themselves. Let A be an eligible alternator strategy. Consider the challenger strategy

$$B(\epsilon) = \left(\frac{1}{2},\, 1-\epsilon,\, \epsilon,\, \frac{1}{2}\right) \tag{20}$$

It is easy to compute that

$$\lim_{\epsilon\to 0} U_{(\alpha,-\alpha)}(a,b,c,d) = \lim_{\epsilon\to 0}(a-d+\alpha(b+c-d)) = -2\alpha$$

Since $\alpha > 0$, the limit is negative. This shows that for some small $\epsilon > 0$, the challenger $B(\epsilon)$ can replace the resident A. So no such resident A can be a stable state. □

Proposition 12. *Let $\alpha < -\frac{1}{2}$. There exists a game (R, S, T, P) for which the process with decision function $U_{(\alpha,-\alpha)}$ has no stable state with payoff π_M.*

Proof. Consider the game (19) as above. The only strategies with payoff π_M are alternators $(p_1, 0, 1, p_4)$. Let A be such a strategy. Let $B(\epsilon) = (\epsilon^2, \epsilon, 1-\epsilon^2, \epsilon^2)$ be a challenger strategy which depends on $\epsilon > 0$. One can check, as in the proof of Proposition 11, that

$$\lim_{\epsilon\to 0} U_{(\alpha,-\alpha)}(a,b,c,d) = \lim_{\epsilon\to 0}(a-d+\alpha(b+c-d)) = 2+4\alpha$$

When $\alpha < -\frac{1}{2}$, it follows that there is some small $\epsilon > 0$ so that the challenger $B(\epsilon)$ can replace the resident A. No such resident A can be a stable state. □

We have shown that for processes of the form $U_{(\alpha,-\alpha)}$, with $\alpha > 0$ or $\alpha < -\frac{1}{2}$, we do not always have stable states with payoff π_M.

The final subsection, Reduction, describes a construction which relates some linear selection processes to others. This construction implies that any selection process $U_{(\alpha,\beta)}$ is equivalent, up to modifying (R, S, T, P), to the selection process $U_{(\alpha',-\alpha')}$ for $\alpha' = -(\beta-\alpha)/2$. Note that $U_{(\alpha,\beta)}$ is outside the band, if $\alpha' > 0$ or $\alpha' < -\frac{1}{2}$. We have already studied such process $U_{(\alpha',-\alpha')}$ in the preceding propositions. So we can conclude that every linear process $U_{(\alpha,\beta)}$ which is outside the band, has the following property: there is a game (R, S, T, P) for which there is no stable state with payoff π_M. In other words, we should not expect processes outside of the band to always select states with maximum payoff.

This means we can characterize the L-band $0 \le \beta - \alpha \le 1$ in the $\alpha\beta$-plane as follows:

Proposition 13. *The process with decision function $U_{(\alpha,\beta)}$ has the following property, if any only if $0 \le \beta - \alpha \le 1$: for any game (R, S, T, P), there exists a stable state with payoff π_M.*

The selection processes in the $\alpha\beta$-plane satisfy positivity, because the coefficient of a in $U_{(\alpha,\beta)}(a,b,c,d)$ is positive and in fact normalized to $+1$. For these processes, the payoff $a := \pi_{AA}$ contributes positively to the resident's chance of prevailing against a challenger.

There is another class of linear decision functions for which the coefficient of a is negative, and normalized to -1. We also wish to study these decision functions, especially if they exhibit the same behavior which characterizes the L-band.

First, consider a process $U_{\mathbf{x}}$ with $\mathbf{x} = (-1, x_2, x_3, 1 - x_2 - x_3)$. That is,

$$U_{\mathbf{x}}(a,b,c,d) = -a + x_2 b + x_3 c + (1 - x_2 - x_3)d \tag{21}$$

We claim that this process does not always have a stable state with payoff π_M.

Case 1: $x_2 + x_3 = 0$. Suppose, for an arbitrary game (R,S,T,P), there is a resident strategy A with payoff π_M, which is a stable state. Define a challenger strategy $B(\epsilon) = (1-\epsilon, \epsilon^2, 1-\epsilon, \epsilon^2)$. One can compute, using the payoff formula, that $\lim_{\epsilon\to 0} \pi_{BB} = P < \pi_M$. Furthermore, one can verify the existence and equality of limits $\lim_{\epsilon\to 0} \pi_{BA} = \lim_{\epsilon\to 0} \pi_{AB} =: x$. So now

$$\begin{aligned}
\lim_{\epsilon\to 0} U_{\mathbf{x}}(a,b,c,d) &= \lim_{\epsilon\to 0}(-a + x_2 b + x_3 c + (1 - x_2 - x_3)d) \\
&= -\pi_M + (x_2 + x_3)x + (1 - x_2 - x_3)P \\
&= -\pi_M + P < 0
\end{aligned}$$

So A is not a stable state, which is a contradiction.

Case 2: $x_2 + x_3 \neq 0$.

Suppose, for an arbitrary game (R,S,T,P), there is a resident strategy A with payoff π_M, which is a stable state. Stability means that for almost every challenger strategy B, we have

$$\begin{aligned}
0 < U_{\mathbf{x}}(a,b,c,d) &= -a + x_2 b + x_3 c + (1 - x_2 - x_3)d \\
&= -\pi_M + x_2 b + x_3 c + (1 - x_2 - x_3)d
\end{aligned}$$

By definition, $d := \pi_{BB} \leq \pi_M$. Let $\epsilon > 0$. By adding $(1+\epsilon)(\pi_M - d) \geq 0$ to the right hand side of the inequality, we have

$$\begin{aligned}
&0 < \epsilon\pi_M + x_2 b + x_3 c + (-\epsilon - x_2 - x_3)d \\
&0 < \pi_M + (x_2/\epsilon)b + (x_3/\epsilon)c + (-1 - (x_2/\epsilon) - (x_3/\epsilon))d
\end{aligned}$$

Equivalently, $0 < U_{(\alpha,\beta)}(a,b,c,d)$, where $\alpha = x_2/\epsilon$ and $\beta = -x_3/\epsilon$. So A is a stable state for $U_{(\alpha,\beta)}$. Note that by choosing ϵ small, we can assume $\beta - \alpha = -(x_2 + x_3)/\epsilon$ does not lie in the interval $[0, 1]$ for some ϵ. By Proposition 13, there is some game for which $U_{(\alpha,\beta)}$ has no stable state with self-payoff π_M. But we have shown that there is such a state A, contradiction.

This establishes the claim. Now we move on to processes with linear decision function of the form

$$U_{\mathbf{x}}(a, b, c, d) = x_2 b + x_3 c + (1 - x_2 - x_3)d \tag{22}$$

In case $x_2 + x_3 \neq 0$, the same approach as in Case 2 will work (using $\epsilon(\pi_M - d)$ rather than $(1 + \epsilon)(\pi_M - d)$). In case $x_2 + x_3 = 0$, there are three cases up to scaling: $\mathbf{x} = (0, 0, 0, 0)$, $\mathbf{x} = (0, 1, -1, 0)$, and $\mathbf{x} = (0, -1, 1, 0)$.

The process $U_{\mathbf{x}}$ with $\mathbf{x} = (0, 0, 0, 0)$ has no stable state. The two remaining possibilities — $\mathbf{x} = (0, 1, -1, 0)$ and $\mathbf{x} = (0, -1, 1, 0)$ — correspond, roughly speaking, to limits of processes $U_{(\alpha,\alpha)}$ in the two directions $\alpha \to \pm\infty$. Their behavior is somewhat singular and we do not characterize them directly here.

2.3 Comparison with M0 strategies

To contextualize the significance of the L-band, it is instructive to compare the space M1 of memory-1 strategies to the simpler space M0 of memory-0 strategies. Recall that M0 strategies give a single probability p for cooperating in each round.

When we apply a linear selection process to the space M0, it is impossible to guarantee the existence of a stable state with payoff π_M. In fact, as we have previously discussed, there are some games for which no M0 state at all has payoff π_M.

We could instead ask whether there is a stable state A which has the maximum feasible payoff *within the space* M0. For convenience, consider a game $(R, S, T, P) = (1, u, 1 + v, 0)$.

Proposition 14. *The maximum feasible payoff for an M*0 *strategy in the game* $(1, u, 1 + v, 0)$ *is*

$$\begin{cases} 1 & \text{if } u + v \leq 1 \\ \dfrac{(1 + u + v)^2}{4(u + v)} & \text{if } u + v \geq 1 \end{cases} \tag{23}$$

It turns out we have the following negative result.

Proposition 15. *Consider the nirvana selection process* $U_{(0,0)}(a, b, c, d) = a - d$ *on the space M*0 *of memory-0 strategies. This process satisfies the following property: for any game* (R, S, T, P) *with* $R > P$*, there is a stable state* A *such that the self-payoff* π_{AA} *is at a maximum in M*0*. There* *is no other decision function $U_{(\alpha,\beta)}$ which has the same property.*

Proof. It is clear by construction that for the nirvana process $U_{(0,0)}(a,b,c,d) = a - d$, a state A is stable if any only if the self-payoff π_{AA} is at a maximum in M0.

Consider the process $U_{\mathbf{x}}$ with $\mathbf{x}(\alpha, \beta)$. The space of strategies is $p \in [0,1]$. We focus our attention on the subspace of games $G = (1,\ u,\ 1+v,\ 0)$. With a small amount of algebra, we can establish the following:

- If $(u+v)(1+\alpha-\beta) \le 0$, then the algebraic condition for $p = 0$ to be the unique stable state is

$$\begin{aligned} \alpha u - \beta v + 1 - \beta &< 0; \\ \beta u - \alpha v + 1 - \beta &\le 0. \end{aligned} \tag{24}$$

- If $(u+v)(1+\alpha-\beta) > 0$, then the algebraic condition for $p = 0$ to be the unique stable state is

$$\begin{aligned} -(1-\beta)u - (1+\alpha)v + 1 - \beta &< 0; \\ (1+\alpha)u + (1-\beta)v + 1 - \beta &\le 0. \end{aligned} \tag{25}$$

Solutions (u, v) to equations (24) and (25) exist for $\alpha \neq -\beta$. Since $p = 0$ does not have the maximum payoff ($p = 1$ has higher payoff), we have shown that stable states with maximum payoff do not exist for $\alpha \neq -\beta$ for some games $G = (1,\ u,\ 1+v,\ 0)$.

We consider the case $\alpha = -\beta$ separately. If $\beta - \alpha > 2$ or $\alpha < -1$, then one can check that for any game $(1, u, 1+v, 0)$ for which $u + v = 0$, the unique stable state is $p = 0$. If $1 < \beta - \alpha \le 2$, then one can check that any game with $u + v > 1$ has no stable states. If $0 < \beta - \alpha \le 1$, then one can check that for any game with $u + v > 1$, there is some unique stable state $p \in (0, 1)$. Such an intermediate strategy does not maximize payoff. If $\beta - \alpha < 0$, then, for all games $(1, u, 1+v, 0)$ with $u + v$ sufficiently small, one can check that there is no stable state. When $\beta - \alpha = 0$ and $\alpha = -\beta$, we recover the nirvana process. □

2.4 Equal gains from switching

We return to selection processes on the space M1 of memory-1 strategies. Our simulations indicate that when $\alpha = -\beta > 0$, the process with linear decision function $U_{(\alpha,\beta)}$ tends toward states with maximum self-payoff π_M when the game satisfies *equal gains from switching*: that is, $R + P = S + T$[4]. We can in fact prove the following proposition.

Proposition 16. *Suppose* $R + P = S + T$*. Let* $\alpha > 0$*. Then the process* $U_{(\alpha,-\alpha)}$*, on the space M1, has a stable state* $\mathbf{p} = (1, 1, 1, 1)$ *with payoff* π_M*. Furthermore, every stable state has payoff* π_M*.*

Proof. Note the decision function

$$\begin{aligned} U_{(\alpha,-\alpha)}(a, b, c, d) &= a + \alpha b + \alpha c + (-1 - 2\alpha)d \\ &= (a - d) + \alpha(b + c - 2d) \end{aligned}$$

We have already shown, in the first part of the proof of Proposition 11, that every stable state has payoff π_M. (This also holds in the space Mn of memory-n strategies.)

Now we claim that $\mathbf{p} = (1, 1, 1, 1)$ is stable. First, note that $\pi_M = R$. The reason is that, by equal gains from switching, $(S+T)/2 = (R+P)/2 < R$, so $\pi_M := \max\{R, (S+T)/2\} = R$. So self-cooperators such as $\mathbf{p}$ have payoff π_M. For a resident $A := \mathbf{p}$ and an interior strategy B, we have

$$\begin{aligned} U_{(\alpha,-\alpha)}(a, b, c, d) &= (a - d) + \alpha(b + c - 2d) \\ &= \underbrace{\pi_M - d}_{>0} + \alpha(b + c - 2d) \end{aligned}$$

We claim that $b + c - 2d > 0$. Without loss of generality we can scale the game parameters R, S, T, P and add a single constant to all of them. Thus we can assume $(R, S, T, P) = (1, u, 1-u, 0)$ for some $u, v \in \mathbb{R}$. Moreover, one can check that the quantity $b+c-2d$ is only a function of three variables $R, S + T, P$. So again without loss of generality, we can assume $(R, S, T, P) = (1, 0, 1, 0)$. Now we simply check numerically that $b + c - 2d > 0$ for this game when A is the strategy $\mathbf{p} = (1, 1, 1, 1)$ and $B = (q_1, q_2, q_3, q_4)$ varies over $(0, 1)^4$. (Note that a randomly chosen challenger strategy B is an interior strategy, i.e. lies in $(0, 1)^4$, with probability 1.) □

2.5 Reduction

The space of centered linear processes $U_{\mathbf{x}}$ can be fully understood by studying a one dimensional family of processes $\mathbf{x}$ and four sporadic cases.

Until now we have assumed that the game (R, S, T, P) satisfies $R > P$. In this subsection, we allow R, S, T, P to be arbitrary real numbers.

Proposition 17. *Each process* $U_{\mathbf{x}}$ *for a given game* (R, S, T, P) *is equivalent — up to a simple* *mathematical transformation — to some process $U_{\mathbf{x}'}$ for a game (R', S', T', P'), where $\mathbf{x}'$ is chosen from the list:*

1. $(1, 0, z, -1 - 2z)$, $z \neq 0$
2. $(1, 1, -1, -1)$
3. $(0, 0, 0, 0)$
4. $(0, 1, -1, 0)$
5. $(0, 0, 1, -1)$

Remark. Case 1 above is the generic case.

Proof. Consider a process $U_{\mathbf{x}}$ and a game (R, S, T, P). We carry out the reduction in steps.

Step 1: Note that if $\lambda > 0$ is some constant, then replacing $\mathbf{x}$ with $\lambda\mathbf{x}$ does not affect the process. Similarly, replacing $\mathbf{x}$ with $-\mathbf{x}$, and replacing (R, S, T, P) with $(-R, -S, -T, -P)$ leads to an equivalent decision function or process.

By applying those two operations, it suffices to study the cases $\mathbf{x} = (0, 0, 0, 0), (0, 1, -1, 0), (0, x_2, 1-x_2, -1)$, and $(1, x_2, x_3, -1 - x_2 - x_3)$ for all games (R, S, T, P).

Step 2 (Optional): Define $\mathbf{p}' = (p'_1, p'_2, p'_3, p'_4) := (1-p_4, 1-p_3, 1-p_2, 1-p_1)$. It is impossible to distinguish between the transformed versions $\mathbf{p}'$ of the states $\mathbf{p}$ visited by the selection process $U_{\mathbf{x}}$ for game (P, T, S, R); and the actual states $\mathbf{p}$ visited by the selection process $U_{\mathbf{x}}$ for the game (R, S, T, P). The two lead to the same distribution over sequences of memory-1 strategies. This is easy to see by simply computing the decision function in both scenarios.

By applying this transformation if desired, we can assume that $R \geq P$.

Step 3: One of the cases we must consider, by virtue of Step 1, is $\mathbf{x} = (0, x_2, 1 - x_2, -1)$. However, if we replace (R, S, T, P) with (R, S', T', P) defined by

$$\begin{pmatrix} S' \\ T' \end{pmatrix} := \begin{pmatrix} 1 - x_2 & x_2 \\ x_2 & 1 - x_2 \end{pmatrix} \begin{pmatrix} S \\ T \end{pmatrix},$$

and we replace $\mathbf{x}$ with $(0, 0, 1, -1)$, then one can check the new decision function is the same as before.

Likewise, another case we must consider is $\mathbf{x} = (1, x_2, x_3, 1 - x_2 - x_3)$. There are two subcases. First, suppose $x_2 + x_3 = 0$. Then if we replace (R, S, T, P) with (R, S', T', P) where

$S' = x_2 S, T' = x_2 T$, and replace $\mathbf{x}$ with $(1, 1, -1, -1)$, then one can check the new decision function is the same as before.

Second, suppose $x_2 + x_3 \neq 0$. Then if we replace (R, S, T, P) with (R, S', T', P) where

$$\begin{pmatrix} S' \\ T' \end{pmatrix} := \frac{1}{x_2 + x_3} \begin{pmatrix} x_3 & x_2 \\ x_2 & x_3 \end{pmatrix} \begin{pmatrix} S \\ T \end{pmatrix},$$

and we replace $\mathbf{x}$ with $(1, 0, x_2 + x_3, -1 - x_2 - x_3)$, then one can check the new decision function is the same as before.

We have shown that every decision function, which depends on $\mathbf{x}, (R, S, T, P)$, leads to a selection process which can be understood as a simple mathematical transformation of a process where $R \geq P$ (if Step 2 is applied) and where $\mathbf{x}$ is one of the five possibilities listed in the proposition statement. □

We can summarize the outcome of the above proposition as follows: let $\mathbf{x} = (x_1, x_2, x_3, x_4)$ be a linear process and $G = (R, S, T, P)$ be a game.

1. If $x_1 \neq 0$ and $x_2 + x_3 \neq 0$ (the generic case), then an equivalent process and game is

$$\mathbf{x}' = \Big(1, 0, \frac{x_2 + x_3}{x_1}, \frac{x_4}{x_1}\Big),\ G' = \text{sgn}(x_1)\Big(R, \frac{x_2 T + x_3 S}{x_2 + x_3}, \frac{x_2 S + x_3 T}{x_2 + x_3}, P\Big)$$

2. If $x_1 \neq 0$ and $x_2 = -x_3$, then an equivalent process and game is

$$\mathbf{x}' = (1, 1, -1, -1),\ G' = \text{sgn}(x_1)\Big(R, \frac{x_2}{x_1} S, \frac{x_2}{x_1} T, P\Big)$$

3. If $x_1 = x_4 = x_2 = 0$, then the process and game are

$$\mathbf{x}' = \mathbf{x} = (0, 0, 0, 0),\ G' = G = (R, S, T, P)$$

4. If $x_1 = x_4 = 0$ and $x_2 \neq 0$, then an equivalent process and game are

$$\mathbf{x}' = (0, 1, -1, 0),\ G' = \text{sgn}(x_2)(R, S, T, P)$$

5. If $x_1 = 0$ and $x_4 \neq 0$, then an equivalent process and game is

$$\mathbf{x}' = (0, 0, 1, -1),\ G' = \text{sgn}(x_4)\Big(R, S - \frac{x_2}{x_4}(T - S), T - \frac{x_2}{x_4}(S - T), P\Big)$$

Next we present a proposition describing a relationship between the process $U_{\mathbf{x}}$ and the associated process $U_{-\mathbf{x}}$. In the main text, we primarily study games with payoff matrix $(R, S, T, P) = (1, u, 1+v, 0)$, where $u, v \in \mathbb{R}$. After a certain number t of rounds, we plot the average payoff π_t of the resident strategies visited. However, we typically normalize by transforming the payoff as

$$\pi'_t = \frac{\pi_t - \pi_m}{\pi_M - \pi_m} \tag{26}$$

The resulting value is a number between 0 and 1.

Proposition 18. *Let $\overline{\pi}_t$ be the expected value of (26) for the process $U_{\mathbf{x}}$ and game $(1, u, 1+v, 0)$. Then the expected value of (26) for the process $U_{-\mathbf{x}}$ and the game $(1, -v, 1-u, 0)$ is $1 - \overline{\pi}_t$.*

Proof. There is a simple relationship between the selection processes with decision functions $U_{\mathbf{x}}$ and $U_{-\mathbf{x}}$. For convenience, call these Process 1 and Process 2, respectively. The decision function for Process 1 in game $(1, u, 1+v, 0)$ is identical to the decision function for Process 2 in the negated game $(-1, -u, -1-v, 0)$.

However, the game $(-1, -u, -1-v, 0)$ is the same as the game $(0, -1-v, -u, -1)$ up to switching the labels of the actions C and D. By adding a constant 1 to all entries, we can further transform the game to $(1, -v, 1-u, 0)$. It is a simple matter to check that $\overline{\pi}_t$ transforms into $1 - \overline{\pi}_t$ under the operations above. □

2.6 Sufficient conditions for linearity

In this section, we show that any decision function $U(a, b, c, d)$ obeying three simple axioms is equivalent to a linear decision function. By "equivalent" we mean in the sense that the two decision functions have the same sign everywhere. Thus, the selection processes for the two functions are exactly the same.

A key feature of linear of linear decision functions is closure under addition of payoffs. Adding two sets of payoffs can be thought of as playing two games in parallel. If resident A defeats challenger B in game G and in game G' separately, then A also defeats B if G and G' are played at the same time and the payoffs summed. (And vice versa if A is defeated by B in game G and G'.)

Proposition 19 (Conditions for linearity)**.** *A continuous decision function $U(a, b, c, d)$ is equivalent to a centered linear decision function (14) if U satisfies the following two axioms:*

*Axiom 1 (**Addition of games**) For any pair of strategies A and B, if A is selected in game*

$G = (R, S, T, P)$ and in game $G' = (R', S', T', P')$, then A is selected in the game $G + G' = (R + R', S + S', T + T', P + P')$ (likewise if B is selected);

*Axiom 2 (**Weakly centered**) $U(\pi, \pi, \pi, \pi) = 0$ for **some** $\pi \neq 0$*

Proof. Let A be the strategy $(1, 1, 1, 1)$ and B be the strategy $(0, 0, 0, 0)$. For the game $G = (R, S, T, P)$, we have $(a, b, c, d) = (R, S, T, P)$. In other words, by choosing the game we can arrange for (a, b, c, d) to be an arbitrary given four-tuple. Then Axiom 1 implies that if $U(a, b, c, d) > 0$ and $U(a', b', c', d') > 0$, then $U(a + a', b + b', c + c', d + d') > 0$ (likewise for $U \leq 0$).

It can also be deduced from Axiom 1, that if $U(a, b, c, d) > 0$, then $U(\lambda a, \lambda b, \lambda c, \lambda d) > 0$ for λ positive and rational. (Likewise for ≤ 0.) By continuity of U, the set of payoffs on which the challenger is selected,

$$\Omega_c = \{(a, b, c, d) : U(a, b, c, d) \leq 0\}, \tag{27}$$

is closed. Thus Ω_c is closed under scaling by any nonnegative real λ. If Ω_c contains only the origin together with rays $(\lambda a, \lambda b, \lambda c, \lambda d)$, then its complement also consists only of rays. The complement of Ω_c is the set of payoffs $\Omega_r = \{(a, b, c, d) : U(a, b, c, d) > 0\}$ on which the resident is selected. Thus Ω_r is also closed under scaling by any positive and real λ. By axiom 1, both Ω_c and Ω_r are convex. The only way to partition $\mathbb{R}^n$ into two disjoint convex sets is via a hyperplane, by the Separating Hyperplane Theorem. Such a hyperplane must pass the origin $(0, 0, 0, 0)$ since both convex sets contains rays $(\lambda a, \lambda b, \lambda c, \lambda d)$. If the hyperplane is given by $x_1 a + x_2 b + x_3 c + x_4 d = 0$, then the decision function $U(a, b, c, d)$ is equivalent to the linear function $x_1 a + x_2 b + x_3 c + x_4 d$. Axiom 2 ensures that $x_1 + x_2 + x_3 + x_4 = 0$, so that the selection function is centered. □

In particular, Axiom 1 is implied by closure under addition of arguments: if $U(a, b, c, d) > 0$ and $U(a', b', c', d') > 0$, then $U(a + a', b + b', c + c', d + d') > 0$ (likewise for ≤ 0).

3 Extensions

3.1 Differentiable decision functions

We have already devoted some time to discussing symmetric decision functions as in Propositions 6-7. In this section, we examine other decision functions which are not necessarily linear.

Suppose $U(a, b, c, d)$ is a continuously differentiable decision function for a selection process which is centered in the sense of Definition 8. When the values a, b, c, d are small, $U(a, b, c, d)$ can be approximated by its first order Taylor expansion. More precisely, fix some π and (R, S, T, P)

and consider a game with payoff matrix

$$\begin{array}{c|cc} & \text{C} & \text{D} \\ \hline \text{C} & \pi+\delta R & \pi+\delta S \\ \text{D} & \pi+\delta T & \pi+\delta P, \end{array} \tag{28}$$

where $\delta > 0$ is small. This is a slight perturbation from a game of constant (strategy-independent) payoff π. The parameters a, b, c, d, which are convex combinations of the entries of the payoff matrix, have magnitudes $\pi + \mathcal{O}(\delta)$. Let U_1, U_2, U_3, U_4 be the derivatives of U in each argument, respectively, evaluated at the point (π, π, π, π). Since U is centered we have

$$0 = \frac{d}{d\pi} U(\pi, \pi, \pi, \pi) = U_1 + U_2 + U_3 + U_4 \tag{29}$$

for all π. The differentiability of U implies

$$\begin{aligned} U(a, b, c, d) &= U(\pi, \pi, \pi, \pi) + U_1(a - \pi) + U_2(b - \pi) + U_3(c - \pi) + U_4(d - \pi) + o(\delta) \\ &= U_1 a + U_2 b + U_3 c + U_4 d \; + \; o(\delta), \end{aligned} \tag{30}$$

where we have used equation (29) and $U(\pi, \pi, \pi, \pi)$ vanishing for all π. To leading order $\mathcal{O}(1)$ and ignoring terms of order $o(\delta)$, the decision function $U(a, b, c, d) \simeq U_1 a + U_2 b + U_3 c + U_4 d$ is linear. We refer to its linearization as $U'(a, b, c, d) = U_1 a + U_2 b + U_3 c + U_4 d$.

Suppose for some π, the derivatives U_1, U_2, U_3, U_4 taken at (π, π, π, π) has $U_1 < 0$, then the linear decision function U' is a process of the form (21). If $U_1 > 0$, but $U_2 + U_3 > 0$ or $U_2 + U_3 < -U_1$ then the decision function U' is a process beyond the L-band in the $\alpha\beta$-plane (16). In each case by the analysis of linear decision functions, there is some game (R, S, T, P) and thus the game,

$$\begin{array}{c|cc} & \text{C} & \text{D} \\ \hline \text{C} & \pi+\delta R & \pi+\delta S \\ \text{D} & \pi+\delta T & \pi+\delta P, \end{array} \tag{31}$$

for which the linear selection process U' has no stable state with payoff π_M. Linear selection processes are invariant under scaling by δ and adding π.

Thus for some π, if the decision function $U(a, b, c, d)$ has the derivatives U_1, U_2, U_3, U_4 taken at (π, π, π, π) obeying any of the following conditions

1. $U_1 < 0$;
2. $U_1 > 0$ and $U_2 + U_3 > 0$;

3. $U_1 > 0$ and $U_2 + U_3 < -U_1$,

then for the game (31), the process $U(a, b, c, d)$ has no stable state with maximal payoff π_M.

Remark. The functions described in Proposition 6 have $U_1 > 0$ and $U_2 + U_3 = 0$.

Now let $f(x)$ be a strictly increasing function with $f(0) = 0$. Consider the selection process with decision function

$$U(a, b, c, d) = f(a) + \alpha\, f(b) - \beta\, f(c) + (-1 - \alpha + \beta)\, f(d) \tag{32}$$

The process is easily seen to be centered. Furthermore, we can assume $f(0) = 0$ without loss of generality. A process with $f(0) \neq 0$ is equivalent to the process with $\tilde{f}(x) = f(x) - f(0)$, satisfying $\tilde{f}(0) = 0$.

Proposition 20. *Suppose* $0 \leq \beta - \alpha \leq 1$. *Then the selection process (32) has a stable state with payoff* π_M.

Proof. The proof is more or less identical to the proof of Proposition 10. □

Conversely, if it is *not* the case that $0 \leq \beta - \alpha \leq 1$, then we do not expect the process to have the property in Proposition 20.

The reason is that when R, S, T, P are small, then a, b, c, d are correspondingly small. So the decision function U behaves similarly to its first order expansion, which is a scaled version of $U'(a, b, c, d) = a + \alpha b - \beta c + (-1 - \alpha + \beta)d$. We have studied these processes above, and found that such a process always has a stable state with payoff π_M, if and only if $0 \leq \beta - \alpha \leq 1$.

3.2 Longer memory

Realistically, players may use strategies which are more complicated than the M1 strategies which we have studied above. It is interesting to ask which of our results also apply to selection processes on the space Mn of memory-n strategies, where $n > 1$.

We note that Propositions 6-7, as well as 9-10, also hold for the space Mn. The proofs are more or less the same. However, note that instead of computing the M1 payoff function 6 and taking limits by hand, one needs to argue differently to establish some of the equalities of limits which we claim in the proofs. (One method is to use the following tool: Press and Dyson show[1] that every memory-n strategy, when playing against a memory-1 strategy, achieves the same outcome as an equivalent memory-1 strategy.)

We also examine whether it is possible to generalize Propositions 11-12. A generalization of these propositions would claim that each process outside of the band does not always have a stable state with payoff π_M. In the next proposition, we show this is true for processes below the band, i.e. with decision function $U_{(\alpha,\beta)}$ and $\beta - \alpha < 0$.

Proposition 21. *Let $\alpha - \beta > 0$ and $n \geq 1$. There exists a game (R, S, T, P) for which the selection process with decision function $U_{(\alpha,\beta)}$, applied to the space Mn, has no stable state with payoff π_M.*

Proof. For games with $R > (S + T)/2$, we have $\pi_M = R$. The states with payoff π_M must cooperate whenever the previous n rounds consist solely of cooperation from both players. (Otherwise, such a strategy could not sustain full cooperation when playing against itself.)

The game played between two Mn strategies gives rise to a Markov chain on the set of possible outcomes for n consecutive rounds of the game. If A is eligible in the sense of Definition 1, then the Markov chain which arises when A plays against itself, has a unique stationary distribution. This stationary distribution is supported on the unique outcome of full cooperation, since the strategy cooperates against itself. If the last n rounds consist of full defection, then A cooperates with some strictly positive probability ϵ. (Otherwise, full defection would be another stationary distribution of the Markov chain.) Now consider the decision function $U_{(\alpha,-\alpha)}$. Suppose we have a game with the following payoff matrix:

$$\begin{array}{c|cc} & \text{C} & \text{D} \\ \hline \text{C} & 1 & -S \\ \text{D} & -S & 0, \end{array} \tag{33}$$

For A with given $\epsilon > 0$, we construct a memory-n challenger strategy B with the following three properties:

(i) When B plays against itself, the outcome is solely defection;

(ii) If only D is played in the last $n-1$ rounds, challenger B will play C with probability $1-\epsilon$.

(iii) For all other n-round outcomes, let p be the probability that A plays C. Then B plays D with probability p. This implies that the actions taken by A and B are opposites with probability

$$p^2 + (1-p)^2 \geq 1/2. \tag{34}$$

A and B are mutually defective with probability

$$p(1-p) < \frac{1}{4}. \tag{35}$$

Now suppose A and B play against each other in the repeated game. We study the frequency of mutual defection in a single round. Let A and B start the game by assuming a history of only defection. To escape from a memory-n history of defection, one player must play a C. But B can't play C by property (i), and A takes on average $1/\epsilon$ rounds to play a C. This means it takes on average $1/\epsilon$ rounds for A and B to escape from a history of n rounds of mutual defection. Now suppose the previous round contains at least one instance of C. Until n consecutive rounds consist of mutual defection, each new round will be mutual defection with probability at most $1/4$. It takes on average at least 4^{n-1} rounds for $n-1$ consecutive mutual defection to occur. But by property (ii) of the strategy B, it in fact takes on average at least $4^{n-1} \times 1/\epsilon$ rounds to achieve n consecutive rounds of mutual defection. So, starting at a memory-n history of n consecutive rounds of mutual defection, with probability $1-\epsilon$ the updated history is the same, and with probability ϵ, the updated history is different. The average number of rounds it takes to return to this memory-n history is at least $(1-\epsilon)\cdot 1 + \epsilon \cdot (4^{n-1}/\epsilon) \geq 4^{n-1}$. Since the average return time for this memory-n history is bounded below — uniformly in ϵ — it follows that the average frequency of this memory-n history is bounded above by a constant less than one — also uniformly in ϵ. But then the frequency of mutual defection DD in a single round is likewise bounded above by some constant less than one, uniformly in ϵ.

The combined frequencies of outcomes CD and DC have

$$v_{DC} + v_{CD} \geq 1 - v_{DD} > 0, \tag{36}$$

Now consider the selection process $U_{(\alpha,\beta)}$ with $\alpha - \beta > 0$. We have that $a := \pi_{AA} = 1$ and $d := \pi_{BB} = 0$. The decision function becomes

$$U_{(\alpha,\beta)}(a,b,c,d) = a + \alpha b - \beta c - (1+\alpha-\beta)d = 1 + \alpha b - \beta c. \tag{37}$$

We substitute in $b = c = v_{\mathrm{CC}} - Sv_{\mathrm{CD}} - Sv_{\mathrm{DC}}$. By choosing S appropriately, we can arrange that the challenger B replaces the resident A under the selection process $U_{(\alpha,\beta)}$. we have

$$U_{(\alpha,\beta)}(a,b,c,d) = (1 + (\alpha-\beta)v_{\mathrm{CC}}) - S(\alpha-\beta)(v_{\mathrm{CD}} + v_{\mathrm{DC}}) \tag{38}$$

Since $v_{\mathrm{CD}} + v_{\mathrm{DC}}$ is bounded below by some positive constant independent of ϵ, the quantity above is negative for large and positive S. Therefore, for the selection process $U_{(\alpha,-\alpha)}$, any state A with payoff π_M can be replaced by some challenger (and by continuity, at least a small locus of challengers). □

Linear processes above the band, i.e. with decision function $U_{(\alpha,\beta)}$ and $\beta - \alpha > 1$, remain to be studied. In the next proposition, we study those which have $\beta - \alpha > 2$. By Proposition 17, it suffices to study processes $U_{(\alpha,-\alpha)}$, with $\alpha < -1$.

Proposition 22. *Let $\alpha < -1$. The selection process with decision function $U_{(\alpha,-\alpha)}$, applied to the space Mn, has no stable state with payoff π_M.*

Proof sketch. Suppose A is a state with payoff π_M. For a challenger strategy $B = A + \epsilon$, for some small displacement vector ϵ, we have

$$\pi_{BB} = \pi_{AA} + \gamma + o(\epsilon) \tag{39}$$

Here γ is the first-order change in payoff as we move from A to B. Since π_M is the maximum payoff, we can assume that $\gamma < 0$.

Note that we also have $b := \pi_{AB} = \pi_{AA} + \lambda + o(\epsilon)$ and $c := \pi_{BA} = \pi_{AA} + \kappa + o(\epsilon)$, where $\lambda + \kappa = \gamma$.

Now we evaluate the process $U_{(\alpha,-\alpha)}$ with $\alpha < -1$ for resident A and challenger B:

$$\begin{aligned} U_{(\alpha,-\alpha)}(a,b,c,d) &= \pi_{AA} + \alpha\pi_{AB} + \alpha\pi_{BA} - (1+2\alpha)\pi_{BB} \\ &= \alpha(\lambda+\kappa) - (1+2\alpha)\gamma + o(\epsilon) && (40) \\ &= -(1+\alpha)\gamma + o(\epsilon) && (41) \end{aligned}$$

This is negative for small ϵ. We conclude that the resident A can be replaced by some nearby challenger strategy $B = A + \epsilon$. □

The proposition above implies (together with Proposition 17) that for processes with $\beta - \alpha > 2$, all states with maximum payoff π_M can be invaded by a local challenger with less than maximal payoff. We now prove a related and strong result: any process with $\beta > 1$ and $\alpha = 0$ has no stable state with maximum payoff, for a wide class of Prisoner's dilemma games. Applying Proposition 17 again, all processes with $\beta - \alpha > 1$ have no stable state with maximal payoff for a variety of games.

A game with the following payoff matrix is an example of a Prisoner's dilemma, in the sense that D is a strictly dominant action:

	C	D
C	1	$-S$
D	$S+2+2\Delta$	$0,$

(42)

for $S > 0$ and $0 < \Delta < \beta - 1$.

We have the following proposition.

Proposition 23. *Let S be positive and $0 < \Delta < \beta - 1$. Fix a payoff matrix $G = (1, -S, S + 2 + 2\Delta, 0)$. When $\alpha = 0$ and $\beta > 1$, the process $U_{(\alpha,\beta)}$, applied to the space Mn, has no stable state with payoff π_M.*

Proof. For game (42), the strategies A with maximum payoff $\pi_M = 1 + \Delta$ are alternators. Let ϵ_1 be the probability that A defects given n consecutive previous rounds of CC. Let ϵ_2 be the probability that A cooperates given n consecutive previous rounds of DD. Both ϵ_1 and ϵ_2 are strictly positive, if A is an eligible strategy. Consider a challenger strategy B which satisfies the following properties:

(i) When B plays itself, the only outcome is mutual defection.

(ii) For all other n-round outcomes, B defects with probability $\eta(\epsilon_1, \epsilon_2)$ as a function of ϵ_1, ϵ_2.

We explain, informally, how η will be chosen for given ϵ. First, pick η small enough such that B cooperates almost always when playing against A. One may worry that B and A may get stuck in mutual defection due to (i). But given a mutually defective history, A and B will get out of mutual defection since $\epsilon_2 > 0$. We pick η so small that once they get out of mutual defection, it takes them a very long time to get back to (n consecutive rounds of) mutual defection. In computing long-run average frequencies of the outcomes CC, CD, DC, DD, we can then essentially neglect the rounds of mutual defection that extend beyond n consecutive rounds, as they make only a small contribution. In the other rounds, B behaves as an M0 strategy that plays D with probability η, and the frequencies v_{CC} and v_{DC} of the outcomes CC and DC satisfy

$$v_{CC} + v_{DC} = 1 - \eta. \tag{43}$$

Second, we pick $\eta < \kappa(\epsilon_1)$ and $\eta < 1 - (1 + \Delta)/\beta$ for reasons that will be clear momentarily. Here $\kappa(\epsilon_1)$ is a function defined in the next paragraph. We have used the assumption $\Delta < \beta - 1$.

When A plays against B, we have argued that $v_{CC} + v_{DC} = 1 - \eta$ (ignoring a small contribution from some rounds of mutual defection).

We claim that the frequency v_{DC} is bounded below by some constant $\kappa(\epsilon_1)$ which is a function of ϵ_1. Divide the whole history of A playing against B into intervals of length about $n+1/\epsilon_1$. Since η is so small, we can ignore the rounds in which strategy B defects. Each interval on average has at least one round of DC: this is because if A has cooperated for n rounds against a fully cooperative opponent, then A is likely to defect once in the next $1/\epsilon_1$ rounds, by the definition of

ϵ_1. So the frequency v_{DC} of the outcome DC has a lower bound of approximately

$$\frac{1}{n+1/\epsilon_1} = \frac{\epsilon_1}{n\epsilon_1+1} := \kappa(\epsilon_1). \tag{44}$$

Additionally, the frequency of CD is bounded above by the probability of B defecting, i.e. η. Since η is chosen smaller than $\kappa(\epsilon_1)$, we have

$$v_{DC} - v_{CD} > 0. \tag{45}$$

Now we are ready to evaluate the process $U_{(\alpha,\beta)}$ with $\alpha = 0$ and $\beta > 1$ on resident A and challenger B. We have $a = 1 + \Delta$, $d = 0$, and $c = v_{CC} - Sv_{CD} + (S + 2 + 2\Delta)v_{DC}$ Evaluate the process $U_{(\alpha,\beta)}$ with $\alpha = 0$ and $\beta > 1$:

$$\begin{aligned} U_{(\alpha,\beta)}(a,b,c,d) &= 1 + \Delta - \beta(v_{CC} - Sv_{CD} + (S + 2 + 2\Delta)v_{DC}) \\ &< 1 + \Delta - \beta(v_{CC} + v_{DC}) - \beta S(v_{DC} - v_{CD}) \\ &< 1 + \Delta - \beta(1 - \eta) < 0, \end{aligned} \tag{46}$$

where we have plugged in equation (43) and (45), dropping the term involving S. This finishes the proof that A with maximal payoff $\pi_M = 1 + \Delta$ can be invaded by a defector B whose payoff is zero. □

We note that ALLD can neutrally replace the strategy B after B takes over A.

If a process $U_{(\alpha,\beta)}$ with $\alpha = 0$, $\beta = \beta_0 > 1$ does not have a stable state with maximum payoff for the game (42) for any particular $S > 0$ and $0 < \Delta < \beta_0 - 1$, then Proposition 17 implies that an arbitrary process $U_{(\alpha,\beta)}$ with $\beta - \alpha = \beta_0 > 1$ has the same problem for the game

	C	D
C	1	$-\frac{\beta+\alpha}{\beta_0}S - \frac{\alpha}{\beta_0}(2+2\Delta)$
D	$\frac{\beta+\alpha}{\beta_0}S + \frac{\beta}{\beta_0}(2+2\Delta)$	$0.$

(47)

We observe that all games of type (47) have $u + v > 1$. We call a game with $u + v > 1$ an alternation-optimal game. This is because the payoffs two two players who alternate between CD and DC outcomes earn higher payoffs than two players who mutually cooperate. Combining the results above together with Proposition 21, we have the following Proposition that characterizes the processes $U_{(\alpha,\beta)}$ for all α and β:

Proposition 24. *Suppose the decision function $U_{(\alpha,\beta)}$ is applied to the space Mn. For this process:*

(i) every stable state achieves maximum payoff π_M, for any game, if $\beta = \alpha$;

(ii) there is a stable state achieving maximum payoff π_M, for any game, if $0 \leq \beta - \alpha \leq 1$;

(iii) there is a family of Prisoner's dilemmas which are alternation-optimal, for which there is no stable state achieving maximum payoff, if $\beta - \alpha > 1$ and $\beta + \alpha > 0$;

(iv) there is a family of harmony games, which are alternation-optimal, for which there is no stable state achieving maximum payoff, if $\beta - \alpha > 1$ and $\beta + \alpha < 0$;

(v) there is a family of snowdrift games, which are alternation-optimal, for which there is no stable state achieving maximum payoff, if $\beta - \alpha > 1$, $\alpha < 0$ and $\beta > 0$;

(vi) there is a family of stag hunt games for which there is no stable state achieving maximum payoff, if $\beta - \alpha < 0$.

The reverse processes of the processes $U_{(\alpha,\beta)}$ have the form $U_{\mathbf{x}}$ with $\mathbf{x} = (-1, \alpha, -\beta, 1 - \alpha + \beta)$. The following proposition characterizes them for all α and β.

Proposition 25. *Suppose the decision function $U_{\mathbf{x}}$ with $\mathbf{x} = (-1, \alpha, -\beta, 1 - \alpha + \beta)$, is applied to the space Mn. For this process:*

(i) there is a family of Prisoner's dilemmas, which are alternation-optimal, for which there is no stable state achieving maximum payoff, if $\beta - \alpha > 0$ and $\beta + \alpha > 0$;

(ii) there is a family of harmony games, which are alternation-optimal, for which there is no stable state achieving maximum payoff, if $\beta - \alpha > 0$ and $\beta + \alpha < 0$;

(iii) there is a family of snowdrift games, which are alternation-optimal, for which there is no stable state achieving maximum payoff, if $\alpha < 0$ and $\beta > 0$;

(iv) there is a family of stag hunt games for which there is no stable state achieving maximum payoff, if $\beta - \alpha < 0$.

All of the statements in Propositions 24 and 25 contribute to the explanation of the results we observe in the simulations. For memory-1 and memory-2, we observe successful resident states with low payoff in prisoner dilemmas which are alternation-optimal, *precisely* for processes $\beta - \alpha > 1$ and $\beta + \alpha > 0$. We observe the same for harmony games which are alternation-optimal, *precisely* for $\beta - \alpha > 1$ and $\beta + \alpha < 0$.

References

[1] Press, W. H. & Dyson, F. J. Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent. *Proceedings of the National Academy of Sciences* **109**, 10409–10413 (2012).

[2] Boerlijst, M., Nowak, M. A. & Sigmund, K. Equal pay for all prisoners. *The American Mathematical Monthly* **104**, 303–305 (1997). URL https://doi.org/10.1080/00029890.1997.11990641. https://doi.org/10.1080/00029890.1997.11990641.

[3] Stewart, A. J. & Plotkin, J. B. Collapse of cooperation in evolving games. *Proceedings of the National Academy of Sciences* **111**, 17558–17563 (2014).

[4] Nowak, M. & Sigmund, K. The evolution of stochastic strategies in the prisoner's dilemma. *Acta Applicandae Mathematica* **20**, 247–265 (1990).