TI 2024-022/I
Tinbergen Institute Discussion Paper

Repeated prisoner's dilemmas with errors: how much subgame-perfection, how much forgiveness, and how much cooperation?

Christopher Graser[1]
Matthijs van Veelen[2]

1 University of Amsterdam and Tinbergen Institute

2 University of Amsterdam and Tinbergen Institute

Repeated prisoner's dilemmas with errors: how much subgame-perfection, how much forgiveness, and how much cooperation?

Christopher Graser and Matthijs van Veelen[*]

April 3, 2024

Abstract

We consider the repeated prisoner's dilemma with implementation errors, and look at the resulting population dynamics, both analytically and with simulations. We show that with implementation errors, pure equilibrium strategies represented by finite state automata exhibit a structure that we call *self-mirroring*. Because selection easily spreads thinly on subgames that are reached after (multiple) errors, we find that in the simulations, strategies are often not best responding in all subgames. We also explore how forgiveness and cooperation respond to changes in the error rate and the continuation probability. Close to an error rate of 0, both show a hump-shaped pattern. We also explore how forgiveness and cooperation change with the error rate in models with a limited strategy set, and we have results for forgiveness at high error rates and/or low continuation probabilities.

---

[*] Department of Economics, Universiteit van Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, and Tinbergen Institute.

1 Introduction

In the repeated prisoner's dilemma, mutual cooperation between players can be stabilised, if players condition their behaviour on the past actions of their interaction partner. However, if players' actions are subject to implementation errors, and players that intend to play cooperate play defect with some positive probability, players cannot distinguish between intended and unintended defections of their opponent. Therefore, if strategies punish defections too harshly, and do not allow for a route back to mutual cooperation after an unintended defection, they will fail to sustain long-term cooperation. This creates a tension between deterring others from defecting intentionally, and forgiving unintentional defections.

This tension has long been recognised. Wu and Axelrod (1995) found that in a setting with implementation errors, Generous Tit-for-Tat and Contrite Tit-for-Tat would outperform other strategies, if added to the original set of strategies from Axelrod's tournaments (Axelrod and Hamilton, 1981). The latter strategy was first proposed by Sugden (1986) and shown to be evolutionarily stable by Boyd (1989). In a simulation that starts with the strategies from Axelrod's original tournament, plus four strategies including Generous Tit-for-Tat and Contrite Tit-for-Tat, Wu and Axelrod (1995) find that Contrite Tit-for-Tat ends up dominating the population.

In a setting that does not restrict the set of strategies to any specific subset, Fudenberg and Maskin (1990) and Dal Bó and Pujals (2020) derive results that suggest that if the error rate is low and the continuation probability is high, selection favours strategies that are highly cooperative as well as forgiving.[1] It is however not obvious whether or how these results, that

---

[1] Fudenberg and Maskin (1990) consider infinitely repeated games without discounting, and use a lexicographic evaluation of payoffs, where payoff differences in the presence of m mistakes only matter if the payoffs in the presence of i mistakes are equal for $i = 0, ..., m-1$. Dal Bó and Pujals (2020) derive results in which the error rate goes to 0, and the continuation probability goes to 1. In the introduction we do not go into (further) detail,

separate equilibria that do from equilibria that do not satisfy certain static stability properties, translate to properties of population dynamics. Because the value of static stability results lies in the degree to which they are informative about evolutionary dynamics, in this paper we explicitly consider dynamics too. In order to be able to compare the relative stability of different equilibria, we chose a model with stochastic selection dynamics, that allows for transitions between equilibria. The average amount of time that a population spends in an equilibrium will then reflect a combination of how easily it invades other equilibria, and how stable it is itself, once established. There are quite a few ingredients that can make a difference for what happens in the dynamics; the stage game, the error rate, the continuation probability, the population size, and the mutation kernel. By exploring different values for these parameters, and by combining the simulations with theoretical results, new and existing, we will try to establish to what extent what we see in the simulations generalises.

Our first result is that, in the presence of errors, for a strategy that can be represented by a finite state automaton, in order to be a Nash equilibrium, it must exhibit a structure that we label *self-mirroring.* Contrite Tit-for-Tat is an example of a strategy that is self-mirroring, and so are Grim Trigger and All D. Tit-for-Tat, on the other hand, and Win-stay-lose-shift (a.k.a. Pavlov) are not. We do however also note that selection inevitably has to be spread thin on subgames that are only reached after errors – in a way that we will make precise in Section 4. It is therefore not to be expected that we will only see Nash equilibria (where Nash equilibria with errors, roughly speaking, are subgame perfect equilibria without errors). The simulations confirm that; almost all of the time the population consists of a strategy that, although it is close to playing a best response against itself, is not a Nash equilibrium.

---

but both of these suggest that full efficiency and forgiveness have a selective advantage for specific combinations of small error rates and large continuation probabilities.

For combinations of error rates and continuation probabilities that make it hard to sustain cooperation at all (that is, for high error rates and/or low continuation probabilities), we find that different sets of strategies come with different, partially opposite implications for how forgiving they should be in order to be an equilibrium. Our Theorem 2 implies that for symmetric strategies, there is a subset of the parameter space where those that punish more harshly remain equilibria for higher error rates and lower continuation probabilities compared to strategies that are more forgiving (while there is also a subset of the parameter space where a symmetric strategy that cannot unambiguously qualified as forgiving or unforgiving remains an equilibrium). Asymmetric strategies like Contrite Tit-for-Tat, on the other hand, are forgiving, and their forgiveness is needed for them to remain equilibria at high error rates and low continuation probabilities.

A third result, relevant for small error rates, establishes continuity in a large class of dynamics. Applied to a setting with errors, it implies that for continuation probabilities under 1, the invariant distribution changes continuously with the error rate. Properties of those dynamics in the limit of the error rate going to 0 therefore cannot be different from what they are in the absence of errors. Without errors, we know that there is a large set of equilibria, with different levels of cooperation, and, for those that are at least cooperative to some degree, with different levels of forgiveness. There is also a large subset of equilibria, with levels of cooperation varying from fully cooperative to fully defecting, that are neutrally stable (they are NSSs, Bendor and Swistak, 1995). No equilibrium, however, is robust against indirect invasions (RAII, van Veelen, 2012; García and van Veelen, 2016). All NSSs therefore they are all similarly (in)stable (see García and van Veelen, 2016 and van Veelen et al., 2012). This is true for all continuation probabilities. The continuity result, together with the observation that in the absence of errors, cooperative equilibria are not more stable than defecting ones, implies that for continuation probabilities under 1, the average amount of cooperation for error rates close to 0 will be close to the average amount of cooperation in the absence of errors. This does however leave the door open for a discontinuity at a combination of an error rate of 0 and a continuation probability of 1, and it does allow for what one might take results from Fudenberg and Maskin (1990) and Dal Bó and Pujals (2020) to suggest, which is that there may be a sequence in which the error rate goes to 0, the continuation probability goes to 1, and the population size goes to infinity, in which cooperation rates in the dynamics approach 1. While we show within an example with a reduced strategy set that this is possible, the simulations with a larger, more general strategy set show that even for reasonably large continuation probabilities and population sizes, adding errors only creates a modest hump in cooperation; as the error rate increases from 0, the average level of cooperation first increases a bit, and rather quickly starts decreasing again. The same is true for forgiveness.

One model with a restricted strategy set allows us to calculate invariant distributions, and illustrates that the average amount of forgiveness is the result of a shifting balance between selection pressure against wasting too much on punishing errors harshly when playing with copies of oneself, and pressure against being too vulnerable to invasions by other, less cooperative strategies. This can create a hump in forgiveness, starting at an error rate of 0. Another model, also with a restricted strategy set, shows how errors can initially increase the amount of cooperation, by making paths out of defection easier relative to paths out of cooperation. These paths would be indirect invasions in the absence of errors, and these changes can create a hump in cooperation.

2 Preliminaries

2.1 Strategies and payoffs

We start with a few formal definitions. Consider a set of players $I = \{1, 2\}$, an action space $A = \{C, D\}$, equal for both players, and a payoff function $\pi : A \times A \to \mathbb{R}$. Using a discount factor $\delta \in (0, 1)$, interpreted as a continuation probability, and an error rate $\epsilon \in \left(0, \frac{1}{2}\right)$, this one-shot game is turned into a repeated one.

A history at time t is a list of 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:

$$\begin{aligned} &h_1 = () \\ &h_t = \left(\left(a_{1,1}, a_{1,2}\right), ..., \left(a_{t-1,1}, a_{t-1,2}\right)\right), \qquad t = 2, 3, ... \end{aligned}$$

Sometimes we will also write $(h_t, (a_{t,1}, a_{t,2}))$ for a history h_{t+1}. We will also write $h_t^{\leftarrow}$ for history h_t, as seen from the perspective of player 2.

The set of possible histories at time t is:

$$\begin{aligned} &H_1 = \{h_1\} \\ &H_t = \textstyle\prod_{i=1}^{t-1} (A \times A) \qquad t = 2, 3, ... \end{aligned}$$

and the set of all possible histories is:

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

A pure strategy is a function $S : H \to A$ that maps histories to the action space. We assume that this strategy is executed with errors. If $S(h) = C$, then action C is played with probability $1 - \epsilon$, and $\neg S(h_t) = D$ is played with probability ϵ, with the straightforward mirror image if $S(h) = D$.

For a combination of strategies S and T, one can determine how many errors have occurred in history h_t. Let $EC_{S,T}(h_t)$ count the number of actions in the history that differ from what the strategies would have prescribed. If h_i, $i = 1, ..., t$ are the truncated histories of length $0 \leq i \leq t$, that agree with h_t in all the periods that they share, then player 1 made an error in period i if $a_{i,1} \neq S(h_i)$, and player 2 made an error in period i if $a_{i,2} \neq T(h_i^{\leftarrow})$. The function $EC_{S,T}(h_t)$ adds the number of errors for every individual period over all periods $i = 1, ..., t$. Between strategies S and T, history h_t then occurs with probability

$$p_{S,T}(h_t) = (1-\epsilon)^{2(t-1)-EC_{S,T}(h_t)}\,\epsilon^{EC_{S,T}(h_t)}$$

For $0 \leq \delta < 1$ and $0 \leq \epsilon < \frac{1}{2}$ the discounted, normalised payoffs to (a player that uses) strategy S against strategy T is given by:

$$\Pi(S,T) = (1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}\sum_{h_t \in H_t} p_{S,T}(h_t)\begin{pmatrix} (1-\epsilon)^2\pi\left(S(h_t), T(h_t^{\leftarrow})\right) \\ +\,(1-\epsilon)\epsilon\pi\left(\neg S(h_t), T(h_t^{\leftarrow})\right) \\ +\,(1-\epsilon)\epsilon\pi\left(S(h_t), \neg T(h_t^{\leftarrow})\right) \\ +\,\epsilon^2\pi\left(\neg S(h_t), \neg T(h_t^{\leftarrow})\right) \end{pmatrix}$$

For $\delta = 1$ we consider the limit-of-means for the infinitely repeated game:

$$\Pi(S,T) = \lim_{\tau\to\infty}\frac{1}{\tau}\sum_{t=1}^{\tau}\sum_{h_t \in H_t} p_{S,T}(h_t)\begin{pmatrix} (1-\epsilon)^2\pi\left(S(h_t), T(h_t^{\leftarrow})\right) \\ +\,(1-\epsilon)\epsilon\pi\left(\neg S(h_t), T(h_t^{\leftarrow})\right) \\ +\,(1-\epsilon)\epsilon\pi\left(S(h_t), \neg T(h_t^{\leftarrow})\right) \\ +\,\epsilon^2\pi\left(\neg S(h_t), \neg T(h_t^{\leftarrow})\right) \end{pmatrix}$$

With strategies that can be represented by finite state automata, the limit-of-means always exists.

2.2 The stage game

We will consider prisoner's dilemmas with "equal gains from switching". This implies that the payoffs can be defined with only two parameters.

$$\begin{bmatrix} & C & D \\ C & b & 0 \\ D & b+c & c \end{bmatrix}$$

The b can be interpreted as the benefits conferred on the other player by cooperating (the left column equals the right column plus b), while c is the cost of cooperating (the top row equals the bottom row minus c)

When describing a prisoner's dilemma with equal gains from switching, the following matrix is also regularly used:

$$\begin{bmatrix} & C & D \\ C & b-c & -c \\ D & b & 0 \end{bmatrix}$$

The equilibrium analysis remains the same if a constant is subtracted from all entries, and therefore these two matrices are equivalent. In the simulations, probabilities are computed, which requires payoffs to be non-negative. In the simulations, we therefore use the first payoff matrix, and for consistency, we also use the first payoff matrix in the analysis. The key results that follow can also be formulated and proven for prisoner's dilemmas with unequal gains from switching, but since adding this further complexity would not lead too any deeper, or qualitatively different insights, we focus on the two parameter version of the game.

2.3 Finite state automata

For the simulations, and for a large part of the theory, we will restrict attention to strategies that can be represented by finite state automata (FSAs). With errors, a finite state automaton requires conditioning, not just on the action of the partner, but also on its own action. In this setting, an FSA, or a Moore machine M, is a tuple $\{\{1, ..., n_M\}, \lambda_M, \mu_M\}$, where n_M is the number of states, $\lambda_M : \{1, ..., n_M\} \rightarrow \{C, D\}$ gives the (intended) output in every state, and $\mu_M : \{1, ..., n_M\} \times \{C, D\}^2 \rightarrow \{1, ..., n_M\}$ gives the transitions as a function of the state, its own action, and the action of the other. Sometimes the formal definition of a machine also specifies in which state the machine starts, but because the states can always be renumbered so that the starting state is the first, we assume, without loss of generality, that the machine starts in state 1. Sometimes we will also order the remaining states, so that, if it plays against a copy of itself, then, in the absence of errors, it will transition from state i to state $i + 1$, until for the first time it goes back to a state it has already been in, or remains in the state it is currently in.

Not all strategies can be represented by FSAs. Different FSAs, on the other hand, can represent one and the same strategy. A straightforward example is the strategy All D which simply defects after all possible histories; any FSA for which the output in all states is D results in the strategy All D (see Figure 1). Sometimes it will be useful to consider a minimal FSA, that is, an FSA for which one cannot find another FSA with fewer states, that represents the same strategy. For a given strategy S that can be represented by an FSA, the minimal FSA is unique, up to the renumbering of states (Hopcroft, 1971). García and van Veelen (2016) moreover define a natural distance between strategies, with which one can show that the set of FSAs is dense in the set of all strategies.[2] This implies that restricting attention

---

[2] The set of FSAs in García and van Veelen (2016) is different from the set of FSAs here, in that there the transitions are only conditioning on the other player's action. This is all

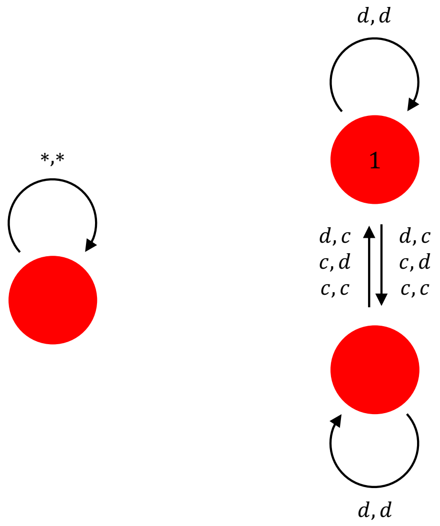


Figure 1: **Two Finite State Automata that both result in All D**. An asterisk indicates either C or D, so $*, *$ indicates all possible action profiles. Red means that the output in that state is D, blue means it is C. Where necessary, the initial state is indicated with a 1. The automaton on the left is the minimal FSA that represents All D. Both of them are self-mirroring (see Section 3).

to strategies that can be represented by an FSA still leaves us with a large and representative set of strategies. Repeated games with the strategy space

---

that is needed to define a strategy in the absence of errors. With errors, FSAs also need to condition transitions on their own realised action. On page 178, García and van Veelen (2016) state that it is *easy to see* that the set of FSAs is dense in the set of all strategies, and they repeat that claim on page 128 of van Veelen and García (2019). For the set of FSAs that we use here, that is indeed true, but for the set of FSAs used in García and van Veelen (2016) and van Veelen and García (2019), in a setting without errors, it is not easy to see, because it is not true. For FSAs that only condition their transitions on the action of the other player, one can construct a strategy that responds differently to two histories, where in one of them, all actions of the player itself are consistent with what the FSA prescribes, and in the other, not all of them are. Strategies that respond differently to those histories cannot be approximated arbitrarily closely by FSAs that only condition on the actions of the other player. Because the difference without errors is created by strategies that respond differently to histories that cannot occur in the absence of errors, this is inconsequential, but the claim there is nonetheless incorrect.

restricted to FSAs are sometimes also called machine games.

2.4 Stability concepts

Without errors, there are many Nash equilibria, and many NSSs (Bendor and Swistak, 1995). Using more restrictive stability concepts does not select between these many equilibria; there are no ESSs (Selten and Hammerstein, 1984), and no strategies that are RAII (García and van Veelen, 2016). Without errors, there is moreover no selection for subgame perfection; the set of NSSs contains both equilibrium strategies that are subgame perfect, such as Grim Trigger, and equilibrium strategies that are not, such as Tit-for-Tat. Simulations in García and van Veelen (2016) and van Veelen et al. (2012) confirm that reasonable dynamics will visit a range of NSSs, and that these are similarly (in)stable, regardless of their equilibrium level of cooperation, and regardless of whether they are subgame perfect or not.

With errors, the relation between the sets of strategies that satisfy the different equilibrium concepts changes. What is and what is not a Nash equilibrium of course depends on the discount rate δ and the error rate ϵ, but since with errors, all subgames are reached with positive probabilities, any Nash equilibrium is going to have to best respond in every subgame. If without errors, a strategy is a Nash equilibrium, but not subgame perfect, then there is a subgame in which it is not best responding against itself. Generically, if a strategy is a Nash equilibrium without errors, there is an open interval $(0, \bar{\epsilon})$ such that this strategy makes the right choices on the error-free path for error rates within this interval. Unless there is a subgame that is reached on the error-free path at which the player is indifferent between C and D at error rate 0, the fact that stage game payoffs and continuation payoffs are continuous in ϵ implies that there is an open interval within which in all subgames that are reached on the error-free path, what is optimal at $\epsilon = 0$ is still optimal within the open interval. (Here we use the fact that an FSA is finite, and there is only a finite number of states that can be reached when playing against a copy of itself). With positive error rates, all subgames are reached with positive probability, and playing the subgame-perfect action in the subgames that are reached *only* under positive ϵ yields strictly higher payoff than this strategy earns against itself.

While there are going to be fewer Nash equilibria for low error rates than there are in the absence of errors, the pure Nash equilibria that remain are typically going to be strict – besides knife edge cases that exist if one chooses error rates that make players indifferent between cooperating and defecting in certain subgames. That means that subgame perfect Nash equilibria without errors – which are not ESS, and not RAII – are typically going to be ESSs, and therefore also RAII, for small enough, but non-zero error rates. In this paper, in the presence of errors, we will therefore typically not differentiate between these stability concepts, and by default, we will consider strict Nash equilibria – where it is implied that these are also ESS.

2.5 Computing and simulating population dynamics

In order to get computational efficiency under the constraints that the question at hand imposes, there are different choices one can make for the simulations. García and van Veelen (2016, 2018) and van Veelen et al. (2012) want to allow for the possibility that the population visits mixed equilibria. In that setting, agent based simulations that follow a Wright-Fisher process are a reasonably efficient choice. One can however also assume that the population spends most of its time in pure equilibria, and that mutations happen so infrequently that the likelihood that a second mutant emerges before a first mutant has either gone to fixation or gone extinct is negligibly low. Although agent based simulations with the Moran process are less

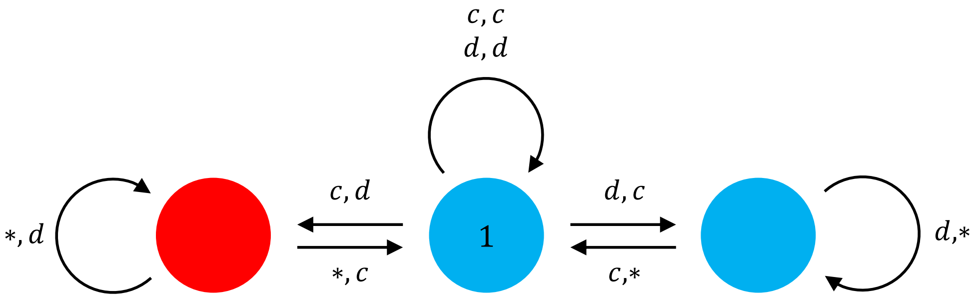


Figure 2: **Contrite Tit-for-tat**. This is a natural choice for an FSA that would represent the most basic version of Contrite Tit-for-Tat in a setting with errors (Sugden, 1986). Figure 3 gives a variation. This strategy is self-mirroring; when two players both play this strategy against each other, they are either both in the initial state, or one is in the left state, and its partner in the right one, or vice versa. In these figures, we will draw self-mirroring FSAs such that the strategy starts in the middle state at the top. To make sure that it is clear that this is the initial state, it is marked with a 1. This strategy can be an equilibrium for the right combination of sufficiently high δ and a sufficiently low ϵ (see Figure 7).

efficient than with the Wright-Fisher process, calculating fixation probabilities for the Moran process – under the assumption that we can use expected payoffs in the payoff matrix between all possible residents and mutants – is relatively easy. Rather than simulating every selection step explicitly, we can then numerically calculate the fixation probability for every new mutant and then determine its fate – fixation or extinction – by drawing a single random number. That is what is done in van Veelen and García (2019) in a setting with complexity costs, and that is also what we do here.

The assumption that we can use expected payoffs basically assumes that the population size is not too small. In the Moran process, for any given realisation of the payoffs, there is a (conditional) transition probability for that realisation. The (unconditional) transition probability from one state to another then is the expected value of those transition probabilities, where the expectation is taken with respect to the probabilities with which these payoffs occur. These probabilities reflect uncertainty with respect to the matching within the population, the duration of the game, and whether and when errors are made. For the calculation of the transition probabilities that we use here (which is more or less standard in the literature; see Nowak 2006), we instead first take the expectation of the realised payoffs, and then calculate just the one (unconditional) transition probability, as if there is no uncertainty about the realised payoffs, and the realised payoffs are equal to the expected payoffs. For small population sizes, there can be a difference between the two ways to compute the transition probabilities (for instance if the matching makes a difference for the total amount of payoffs), but for larger population sizes, these differences will vanish. We have also verified numerically that this approach matches agent based simulations, in which every selection step is explicitly simulated.

While the default simulations in this paper allow for mutants to explore all possible FSAs, we also have two models in which we consider restricted strategy sets. These strategy sets are small enough to allow us to explicitly calculate stationary distributions of the dynamics for a given mutation kernel. Those are used then to illustrate points and mechanisms relevant to the dynamics in the simulations that do allow for all FSAs.

3 Equilibrium strategies with errors

In this section we consider strategies that can be represented by FSAs, and we describe properties of equilibria. For constructing a best response, it will be useful to have a value function for the states of an FSA M, representing a strategy S, from the point of view of its opponent. Given a continuation probability, or discount factor $\delta \in [0, 1)$, and an error rate $\epsilon \in \left[0, \frac{1}{2}\right)$, the

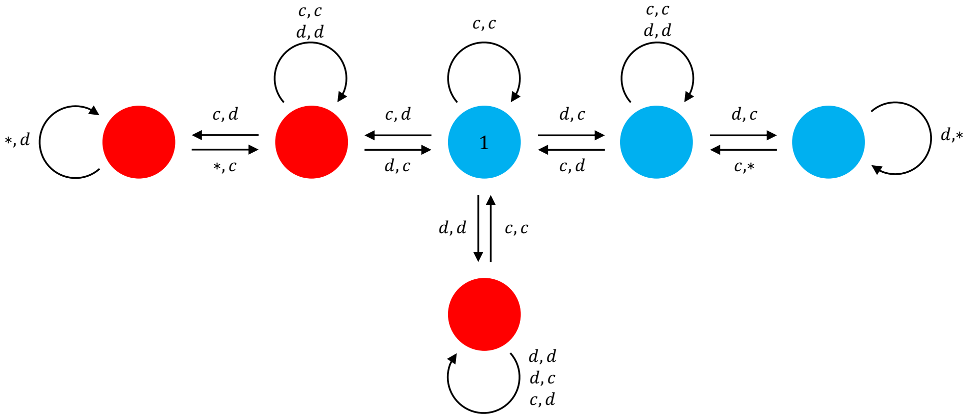


Figure 3: **A variant of Contrite Tit-for-tat**. This is a version of Contrite Tit-for-Tat that, when playing against itself, goes to a mutually defecting state if both make an error when in the initial state (at the center on top). Also the score keeping is different from the simpler version of Contrite Tit-for-tat, depicted in Figure 2, in that this one has a second level of contrition. The value to the opponent of this strategy being in different states goes from the lowest all the way to the left to the highest all the way to the right. This strategy is also self-mirroring, and it is an equilibrium for the right combination of a sufficiently high δ and a sufficiently low ϵ.

value to the opponent of a player using M of M being in state i is denoted by $V_M^*(i), i = 1, ..., n_M$, and it is the solution to the following system of equations.

$$V_M(i) = \max_{a \in \{C,D\}} \left\{ \begin{array}{l} (1-\epsilon)^2 (\pi(a, \lambda_M(i)) + \delta V_M(\mu_M(i, \lambda_M(i), a))) \\ + \epsilon(1-\epsilon)(\pi(\neg a, \lambda_M(i)) + \delta V_M(\mu_M(i, \lambda_M(i), \neg a))) \\ + \epsilon(1-\epsilon)(\pi(a, \neg\lambda_M(i)) + \delta V_M(\mu_M(i, \neg\lambda_M(i), a))) \\ + \epsilon^2 (\pi(\neg a, \neg\lambda_M(i)) + \delta V_M(\mu_M(i, \neg\lambda_M(i), \neg a))) \end{array} \right\} \qquad i = 1, ..., n_M$$

The maximum error rate we consider is $\frac{1}{2}$. At this error rate, any choice leads to a random coin flip in every subgame. Because every state is reached with positive probability, best responding against a player using M means playing an action that maximises the term between curly brackets, when M finds itself in state i. Creating a best response against the strategy S that M implies therefore can be done by making a finite state automaton M' that has the following properties.

1. M' has the same number of states as M does, that is, $n_M = n_{M'}$;
2. M' has a transition function that satisfies $\mu_{M'}(i, b, a) = \mu_M(i, a, b)$. This implies that those states are numbered in such a way that if M is in state i, then M' is also in its state i;
3. What M' plays when it is in state i (and M therefore is in its state i too) maximises the term between curly brackets in the definition of $V_M(i)$ above.

For M to represent a strategy that is a Nash equilibrium, its best response M' has to represent the same strategy that M represents. For M to represent a strategy that is a strict Nash equilibrium, the best response facing M has to be unique for every state of M.

Restricting attention to FSAs implies that we both reduce the set of possible equilibria, and the set of possible mutants, or alternative responses, we consider. Here, it may be good to recognise that FSAs that are equilibria within the set of FSAs, remain equilibria in the set of all strategies; if M' is the best response among all FSA, then it is also the best response among all strategies.

We will see below that a necessary condition for a *minimal* FSA M to be a Nash equilibrium is that it should be self-mirroring; it should be possible to make an equivalent FSA M' by renumbering the states in M, so that $\mu_{M'}(i, b, a) = \mu_M(i, a, b)$. The other condition is that what M' plays when in state i (and M therefore is in its state i too) maximises the term between

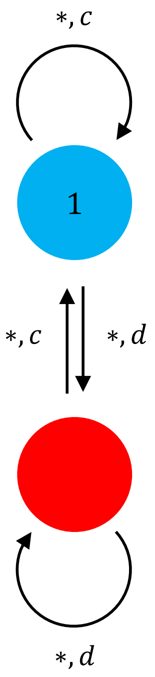


Figure 4: **(Normal) Tit-for-Tat**. This strategy is not self-mirroring. Starting in state 1 for both, if player 1 makes an error, but player 2 does not, then player 2 moves to the second state, and player 1 does not.

curly brackets in the definition of $V_M(i)$ above. Before we prove this, it may be useful to illustrate this central property of equilibria.

The FSAs in Figures 1, 2, and 3 are all self-mirroring. All D (Figure 1) is an equilibrium for all $\delta \in [0, 1]$ and $\epsilon \in \left[0, \frac{1}{2}\right)$. The different versions of Contrite Tit-for-Tat (Figs. 2 and 3) are equilibria for sufficiently high δ and sufficiently low ϵ. Tit-fot-Tat (Figure 4) is not self-mirroring, not subgame perfect without errors, and not a Nash equilibrium with errors. The FSA in Figure 5 is not self-mirroring, but it can be reduced to the FSA for Contrite Tit-for-Tat in Figure 2, which is self-mirroring.

The result below states that if a strategy is a strict Nash equilibrium, it can be represented by a minimal self-mirroring FSA. This is a very useful and

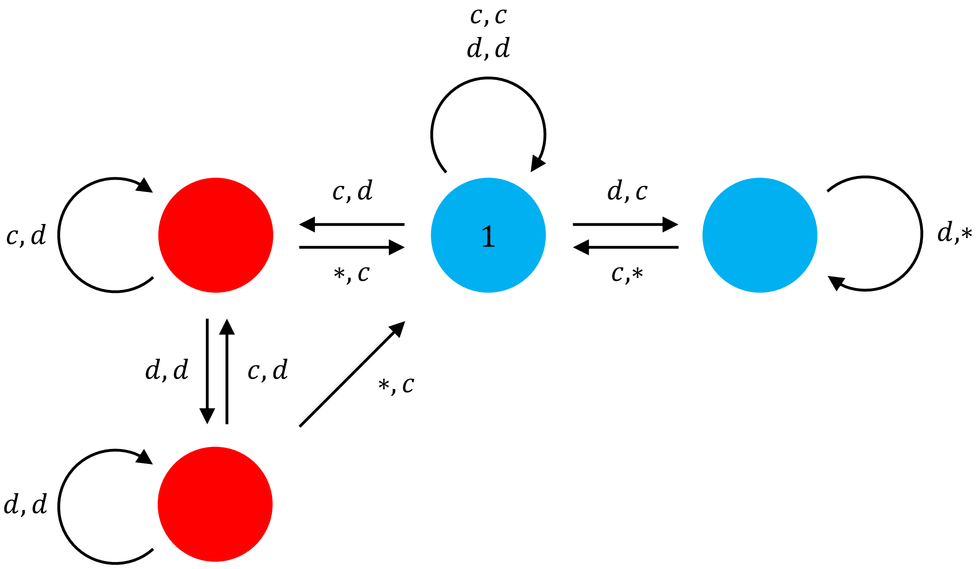


Figure 5: **An FSA that can be reduced to Contrite Tit-for-tat**. This FSA is not self-mirroring, but it can be reduced to the automaton that represents the strategy Contrite Tit-for-Tat from Figure 2, which is self-mirroring.

informative result, as it reduces the set of possible equilibria considerably.

Theorem 1. *If a finite state automaton M is minimal, and the strategy S it represents is a strict Nash equilibrium, then M is self-mirroring.*

Proof. For every state $i = 1, ..., n_M$, define the set of states $s(i)$ as follows. If two players meet, both use M, and one player is in state i, then $s(i)$ is the set of states that the other player can be in. Perhaps overly formally, $s(i)$ is the set of states j for which there is a history h_t that puts M in state i, while $h_t^{\leftarrow}$ puts M in state j. Now suppose that there is a state i for which $s(i)$ is not a singleton. That means that there are states j and k, $j \neq k$, with $j \in s(i)$ and $k \in s(i)$. Now starting in state j, a continuation history

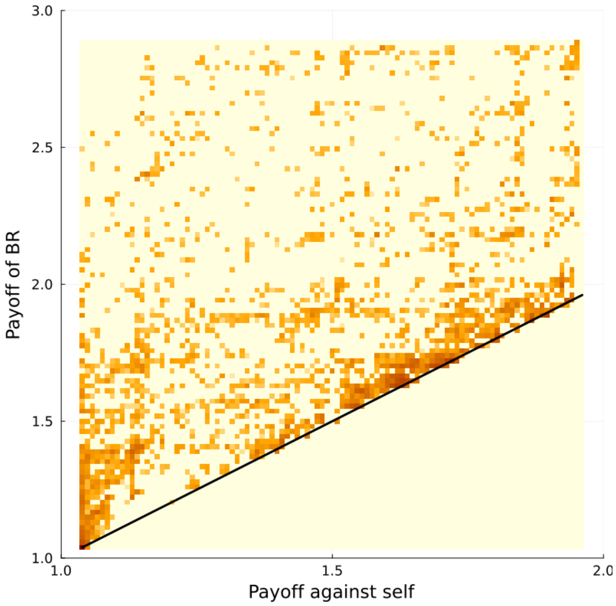
g_t – which is a sequence of $t-1$ action profiles, like normal histories, as defined in Section 2 are – would put M in state j', and, starting in state k, the same continuation history g_t would put M in state k'. If the output in those states j' and k' is the same for every continuation history g_t – that is, if $\lambda_M(j') = \lambda_M(k')$ for all g_t and some combination of $j, k \in s(i)$ – then M was not minimal, as the same strategy could be represented by an FSA M' that would not have state j, and redirect all arrows that point to state j in M so that they point to state k in M'. If on the other hand there is a continuation history g_t for which the output in states j' and k' differs, then M is not a strict equilibrium. To see that, let i' be the state that M transitions to if it starts at state i and observes the continuation history $g_t^{\leftarrow}$. Since M is a Nash equilibrium, the outputs in both j' or in k' need to maximise the term in brackets in the definition of V_M, applied to i', but as these outputs are not identical, there are multiple best responses against M in state i', and M does not represent a *strict* Nash equilibrium. Finally, since $s(i)$ is a singleton for all i, we can now renumber the states, and relabel state $s(i)$ as state i. This gives an M' for which $\mu_{M'}(i,b,a) = \mu_M(i,a,b)$. $\square$

4 Small imperfections

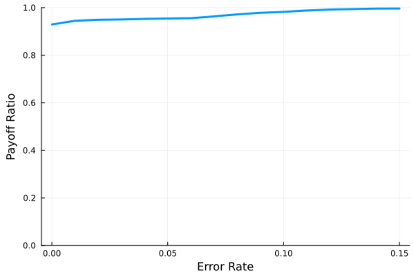
Before discussing how often we see equilibria in simulations, in which strategies are represented by FSAs, it is worth reflecting on selection off the error-free path in general. In the infinitely repeated prisoner's dilemma, there are infinitely many subgames. Therefore, if we were to represent a strategy as a (countably infinitely long) list with binary items that specify for every history whether it plays C or D, then it is clear that selection will have to be spread thin somewhere. Somewhat more precisely, in every population, there are subgames that are reached arbitrarily infrequently, making selection there also arbitrarily weak. In such a setting, with selection acting on choices for every history separately, the less often a history occurs (or the less often a subgame is reached), the less we expect to find the choice there to be optimal.

In the simulations, and in most of the theory in this paper, we represent strategies by FSAs. That means that by definition, for any FSA there are states that it arrives at through multiple histories, and there are even states that it arrives at through infinitely many histories. By changing the output in a state, a mutant can therefore change the strategy for a whole set of histories in one go. Similarly, other mutations (deleting or adding a state, or redirecting a transition) can also change the strategy it represents for whole sets of histories. By bundling histories together, a mutant could undo a number of suboptimalities, which implies that it it also aggregates the differences in expected payoffs. That can make selection against not best responding more effective. There can however still be states that are reached very infrequently, and therefore it is possible that a strategy is very close to being a best response against itself, while not being an equilibrium. This is the case if a strategy only chooses a suboptimal action for a set of histories that is very unlikely to occur.

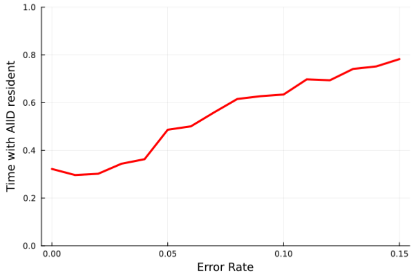
Figure 6 illustrates how this shows up in the simulations. We take the difference between the expected payoff of playing a best response to a strategy, and the expected payoff that this strategy earns against itself, to quantify how far away from being a Nash equilibrium a strategy is (Figure 6a). There is one strategy that deserves special attention, and that is All D. This strategy is subgame perfect without errors, and it is a Nash equilibrium with errors for all $\epsilon > 0$ and $\delta < 1$. The time that the population spends in All D varies as a function of the error rate (Figure 6c), and because this is by far the most frequent strategy in the simulations, we consider the changes in the share of All D separately. In Figure 6b we then consider the remaining strategies, that cooperate in at least one state, together. This set of strate-



(a) Distribution at $\epsilon = 0.02$



(b) Proximity to best response



(c) Time with All D resident

Figure 6: **Selection for subgame perfection.** In order to measure how far a strategy is from what would be subgame perfection without errors, and what being a Nash equilibrium is with errors, we compare, for every strategy that we observe in the simulation, what payoff best responding against it would result in, and what payoff it gets against itself. In panel (a) we see the density of combinations of those from a simulation run with $\frac{b}{c} = 2$, $\delta = 0.9$, $\epsilon = 0.02$, and $5 \cdot 10^6$ mutant arrivals. On the black diagonal line, the expected payoff that a strategy earns against itself is the same as the expected payoff that a best-responding strategy would earn. Strategies on this line, therefore, are Nash equilibria. In order to also make combinations with a low frequency visible, we first log_{10}-transformed the relative frequencies before applying the colour scheme. The bottom left pixel of this distribution corresponds to the strategy All D, which has by far the highest frequency. For panel (b), we aggregate these differences, weighted by the frequencies, and repeat that for a variety of ϵ's. Here, we do however exclude All D. Panel (c) plots the proportion of time spent with All D as a resident as a function of the error rate. All D is a Nash equilibrium with errors, for all error rates, and it is subgame perfect without errors.

gies contains strategies for which all cooperative states can only be reached after errors, and therefore it remains possible that the expected payoffs are close to the payoff of mutual defection.

If we fix a strategy that is not best responding in a state that is only reached after an error, then the difference between the payoff that a player would get by best responding and the payoff that this strategy gets against itself increases with ϵ. This, in turn, increases the selection pressure against such a strategy. The first effect increases the gap for an individual strategy, selection decreases the gap in the average distribution of strategies. In the simulations, the average gap decreases as the error rate increases, which suggests that the effect of selection is bigger (see Figure 6). The average number of states that automata have does not vary much with the error rate (not shown).

5 Forgiveness at high error rates or low continuation probabilities

At $\epsilon = \frac{1}{2}$, all choices are inconsequential, and every strategy is as good as any other. All strategies therefore are Nash equilibria, none of them are strict, and no strategy has a selective advantage over any other strategy. In simulations in general, we expect that the mutation probabilities always matter, but if the mutation kernel is not biased in favour of cooperation or defection, then at $\epsilon = \frac{1}{2}$ we should observe equal amounts of both C and D as intended actions (of course the actual actions at an error rate of $\frac{1}{2}$ are always balanced, no matter what strategies the population is composed of). Unbiased mutations would entail that mutating the output in a state from D to C is equally likely as mutating from C to D; that a newly added state is equally likely to have C or D as output; that mutations on the transitions make them point to all other states equally likely; and that the likelihood of a state being deleted is also independent of its output. The choice of a mutation kernel matters at all mutation rates, and certainly also at $\epsilon = \frac{1}{2}$, where all strategies behave the same and earn the same expected payoffs, but there is a large set of unbiased mutation kernels, that all lead to the same balanced average intended actions at $\epsilon = \frac{1}{2}$.

Generally, high error rates make it hard to stabilise cooperation. This is because cooperative strategies must, if their partner defects, also respond with some defection in future rounds so as to deter intentional defections. At higher error rates, such responses are triggered ever more often by unintentional defections when cooperative strategies play against copies of themselves. As the likelihood of remaining in a cooperative state, even if both players intend to cooperate, decreases, it becomes ever less attractive to play cooperate.

As a consequence, many cooperative strategies stop being Nash equilibria as the error rate increases. Our Theorem 2 below provides a simple condition for when symmetric strategies stop being equilibria. Symmetry is defined in Dal Bó and Pujals (2020) as follows: a strategy s is symmetric if $s\left(h_t\right) = s\left(h_t^{\leftarrow}\right)$ for any history h_t. For strategies represented by FSAs, this means that its minimal representation must have the property that, when the strategy plays against a copy of itself, both players always find themselves in the same state. Grim Trigger and All D, for instance, are symmetric, but Tit-for-Tat and Contrite Tit-for-Tat are not.

For this theorem, we, moreover need to introduce a strategy which we call Minimal Grim Trigger. This strategy, like Grim Trigger, has two states – one where it cooperates and one where it defects in all future rounds. However, unlike Grim Trigger, Minimal Grim Trigger only reaches its defection state if *both* players defect in the first state. After unilateral defections, Minimal Grim Trigger remains in its cooperate state.

Theorem 2. *A necessary and sufficient condition for Grim Trigger to be a Nash equilibrium is:*

$$\frac{b}{c} \geq \frac{1-\delta(1-\epsilon)\epsilon}{\delta(1-\epsilon)(1-2\epsilon)}$$

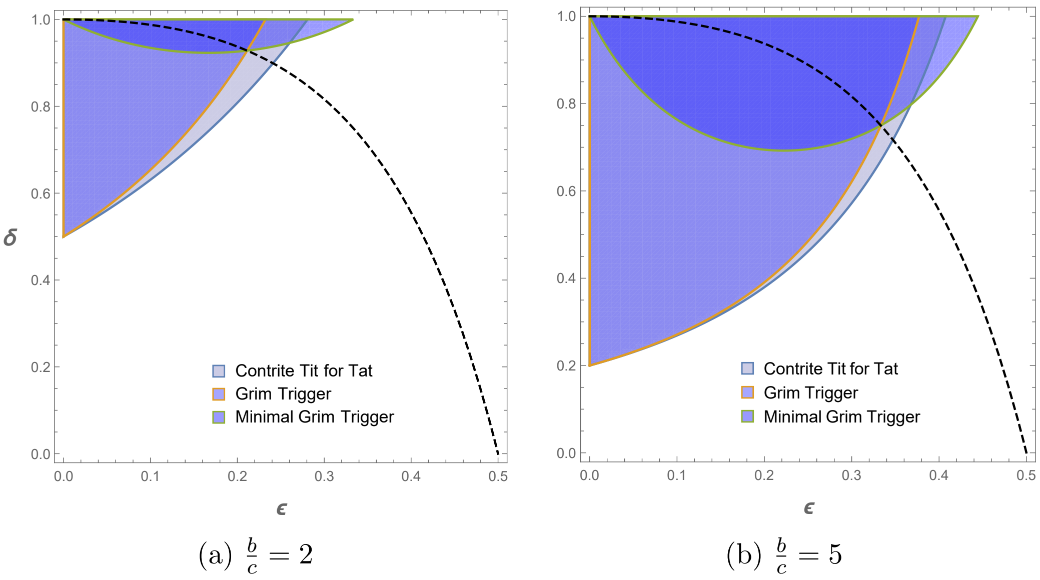
A necessary and sufficient condition for Minimal Grim Trigger to be a Nash equilibrium is:

$$\frac{b}{c} \geq \frac{1-\delta(1-\epsilon+\epsilon^2)}{\delta\epsilon(1-2\epsilon)}$$

If neither Grim Trigger nor Minimal Grim Trigger are Nash equilibria for a given combination of c, b, δ and ϵ, then there exists no other symmetric strategy that is a Nash equilibrium for this parameter combination.

Proof. See Appendix A. □

Theorem 2, and the reasoning in the proof of Theorem 2, suggest that within the set of symmetric strategies, the strategies that constitute Nash equilibria at high error rates, are those that once they reach their punishment state, have no route back to cooperation. The threat of long periods of mutual defection most effectively disincentivises unilateral defections in cooperative states, and Grim Trigger constitutes the extreme of indefinite defection after observing a defection in a cooperative state. Similarly, at higher error rates, where unilateral defections by one's interaction partner become very likely, Minimal Grim Trigger punishes mutual defection in the strongest possible way. However, the opposite pattern emerges when we look at asymmetric strategies. For instance Contrite Tit-for-Tat (Figure 2), which has a very quick route back to cooperation after a defection occurred, is a Nash equilibrium even under error rates that are too high for any symmetric strategy to be a Nash equilibria (see Figure 7). Intuitively, with Contrite Tit-for-Tat, the incentive to cooperate in state 1 is especially strong, because additionally to being punished for defecting, players are *compensated* in subsequent rounds if in the current round they end up cooperating while their



(a) $\frac{b}{c} = 2$ (b) $\frac{b}{c} = 5$

Figure 7: **Parameter ranges for which Minimal Grim Trigger, Grim Trigger and Contrite Tit-for-Tat are Nash equilibria**. As $\frac{b}{c}$ increases, all three strategies become stable on ever larger regions in the ϵ-δ square. The dashed line represents the boundary above which the incentive constraint for cooperation is satisfied by a larger margin for Minimal Grim Trigger than it is for Grim Trigger.

opponent defects. Therefore, even with a short punishment phase, Contrite Tit-for-Tat can stabilise cooperation in state 1. Moreover, there is a limit to how much these asymmetric punishment sequences, where the punishing player defects and the player that is being punished cooperates, can be exploited to stabilise cooperation: the player that is being punished only has an incentive to cooperate and thereby to comply with the punishment scheme, if compliance leads to a sufficiently quick return to the a state in which both players cooperate. Hence, it is precisely those asymmetric strategies which have short asymmetric punishment phases that remain equilibria at high error rates.

Summarizing, for parameter combinations for which sustaining cooperation is relatively hard, there is a difference between symmetric and asymmetric strategies. Symmetric strategies that do not punish harshly enough are not equilibria, and symmetric strategies that are equilibria therefore are not very forgiving. Asymmetric, but still self-mirroring strategies, like Contrite Tit-for-Tat, on the other hand, can be equilibria past the point where no symmetric equilibria exist. These strategies cannot punish too harshly in order to incentivise the individual that makes an error to repent by cooperating while the other one defects. Symmetric and asymmetric equilibria are very different, in that asymmetric strategies require some sort of penitence to go (back) to mutual cooperation, while symmetric strategies do not.

6 Cooperation and forgiveness at error rates close to 0

At $\epsilon = 0$, there is no selection for or against subgame perfection, nor is there selection for or against forgiveness, unless it affects what is optimal in the absence of errors. Strategies that would (sometimes) cooperate when they are matched with a copy of themselves, but that are so forgiving that it would be optimal to play defect against them, are not Nash equilibria. Those are obviously selected against, and show up only very shortly in simulations, typically as neutral mutants that are part of an indirect invasion. Between equilibrium strategies that all clear the bar of punishing sufficiently hard to incentivise cooperation, but that are not equally forgiving, the dynamics do not really differentiate in the absence of errors.[3] The only levels of forgiveness

---

[3] As we will see, there are dynamics for combinations of strategy sets and population sizes for which direct invasions between cooperative and uncooperative equilibria play a role (whereas for large strategy sets and large population sizes, all traffic between equilibria happens by indirect invasions). In such cases, as we will see, the punishment length may still make a difference in the absence of errors, although the difference for instance in the model of Section 8 is minimal (see Figure 22).

with a clear selective disadvantage at $\epsilon = 0$, therefore, are levels that are so forgiving that they do not incentivise cooperation.

As mentioned in the introduction, there are papers that describe properties of equilibria at vanishingly small error rates. Dal Bó and Pujals (2020) do that by comparing the sizes of the basins of attraction of equilibria with different levels of cooperation, on the error-free path, and with different levels of forgiveness when errors do occur. By considering a lexicographic evaluation of payoffs, Fudenberg and Maskin (1990) effectively use a lexicographic stability concept, where payoffs resulting from actions in histories with m errors only matter if the payoffs for histories with i errors, $i = 0, m - 1$, are all equal. Both of these papers suggest that ever smaller error rates, and ever higher discount rates, could allow dynamics to get arbitrarily close to full efficiency, and that forgiveness will be selected for.

Before discussing the interesting relation between their results and our dynamics, we will first focus on something more straightforward in the dynamics. If we take a fixed continuation probability $\delta < 1$, we can explore what the effect of adding errors is. As we will see in Figure 8, increasing the error rate from 0 comes with a little hump in cooperation and forgiveness. The simulations do however give no reason to believe that there is a discontinuity at $\epsilon = 0$; it seems that in the limit of the error rate going to 0, the average amount of cooperation and forgiveness converges to what they are at $\epsilon = 0$.

The observation that changes in the properties of the process when going from no errors to small error rates are continuous, is not an artifact of our setup, mutation procedure, or parameter choices. Rather, for any reasonable stochastic selection process with errors, we should expect the dynamics in the limit of small error rates to coincide with the dynamics in the absence of errors. This is formalised by the following theorem. For interpreting this result, it is helpful to realise that the dynamics are a Markov chain with a very large state space. This Markov chain may or may not have an invariant distribution, depending on the mutation kernel. If the mutation kernel is too likely to add states relative to how likely it is to delete states, then the average size may tend to keep growing, and no invariant distribution exists. If the mutation probabilities imply that adding a state is not too likely an event compared to deleting a state, then an invariant distribution does exist. What we do in the simulations is to estimate properties of these invariant distributions, such as the share of certain (types of) strategies, and the average amount of cooperation, forgiveness, and subgame perfection (or, more precisely, the average distance between the payoffs of the best response against a strategy and the payoffs a strategy gets against itself). The following theorem establishes conditions under which these invariant distributions change continuously as a function of the error rate. Because this result carries over to changes in Markov chains as a result of any parameter in general, including, within our setup, the continuation probability, or, in other settings, for instance complexity costs, the ϵ in the result will represent any parameter that has an effect on transition probabilities.

Theorem 3. *In a discrete time Markov process with countable state space, let the transition probabilities $P_\epsilon(i,j)$ between any states i and j be continuous, differentiable functions of some parameter ϵ. If for every ϵ this process has a unique stationary distribution, and if there exist an $m > 0$ and an M such that*

1. $\forall i: \ (\forall \epsilon' : P_\epsilon(i,i) = P_{\epsilon'}(i,i)) \vee (P_\epsilon(i,i) \leq 1 - m)$
2. $\forall i,j: \ |\frac{dP(i,j)}{d\epsilon}| \leq M$

then this stationary distribution is continuous and differentiable in ϵ, and the derivative of the weight $\sigma(i)$ of each state i in the stationary distribution is bounded by:

$$|\frac{d\sigma(i)}{d\epsilon}| \leq \frac{M}{2m}$$

Proof. The proof is in Appendix B.[4] □

The first condition in Theorem 3 is typically satisfied for a Moran process, a Wright-Fisher process, or any other reasonable stochastic finite population process. If the ϵ represents the error rate, with the continuation probability fixed, or if it represents the continuation probability, with the error rate fixed, then for all absorbed (monomorphic) states, it is automatically true that for all ϵ, the probability $P_\epsilon(i, i)$ of staying in the same state is equal; all transitions to other states are due to mutations, and the probabilities relating to those are independent of the error rate or the continuation probability. For states that are not absorbed, there is always the possibility that a replacement event changes the population state, and the probability of that happening is easily bounded away from zero across all mixed states. On top of that, for all states, absorbed or not, the presence of a positive mutation rate is going to put an upper bound on the probability of staying in the same state. Also, if we were to think of the ϵ in the theorem as a two-dimensional variable, representing both the error rate and the continuation probability, the first condition would still hold for all combinations of those in any reasonable stochastic finite population process, including the Moran Process with a positive mutation rate.

The second condition in Theorem 3 is also satisfied in the Moran process if the ϵ represents the error rate, but only for $\delta < 1$. The probability of going from state i to state j in the Moran process is the expected value of transition probabilities conditional on a realisation of the payoffs – where the expectation is taken over the probabilities with which these payoffs occur. The transition probabilities, conditional on a realisation of the payoffs, are independent of the error rate. The probability of each realisation depends

---

[4] Of course, when applying this theorem to argue that the stationary distribution does not change discontinuously at $\epsilon = 0$, we take $\frac{d}{d\epsilon}$ to mean the right-sided derivative at zero.

continuously on the error rate, and therefore the second condition is satisfied.[5] For details see Appendix F.

For infinitely repeated prisoner's dilemmas ($\delta = 1$), and with ϵ still representing the error rate, the second condition will however easily be violated. The payoffs of a strategy such as for instance Grim Trigger is discontinuous at $\epsilon = 0$; against itself earns a per period average of b at $\epsilon = 0$, and a per period average of $(1 - \epsilon) c + \epsilon b$ at all $\epsilon > 0$. That means that in the limit of vanishing error rates, the payoff of Grim Trigger against itself is $c \neq b$. In the next section, we will encounter a strategy called Jumpstarter, and this strategy goes the other way; against itself it earns a per period average of c at $\epsilon = 0$, but for all $\epsilon > 0$ it earns a period average of $\frac{1}{2}(b + c)$. This means that Condition 2 is not satisfied. In Section 7 we will moreover see that in a strategy set containing All C, All D, Grim, and Jumpstarter, we do indeed see a discontinuity at an error rate of 0 for a continuation probability of 1. This is visible both in panel (a) and in panel (b) of Figure 16.

Going back to discount rates below 1, and applying the result to error rates close to 0, we can then combine it with what we know to be true for repeated games in the absence of errors. From Bendor and Swistak (1995) we know that for every $\delta < 1$, in the absence of errors, there are NSSs varying from fully cooperative to fully uncooperative. From García and van Veelen (2016) we know that none of them are RAII. All NSSs therefore are similarly (un)stable. From van Veelen et al. (2012), we moreover know that for normal, unbiased mutation kernels, the average amount of cooperation for normal $b/c-$ratios remains well below 1 for all $\delta < 1$ (for $b/c = 2$, for instance, the average amount of cooperation stays well below 0.5). The theorem then implies that we should also not expect spectacular cooperation rates if we

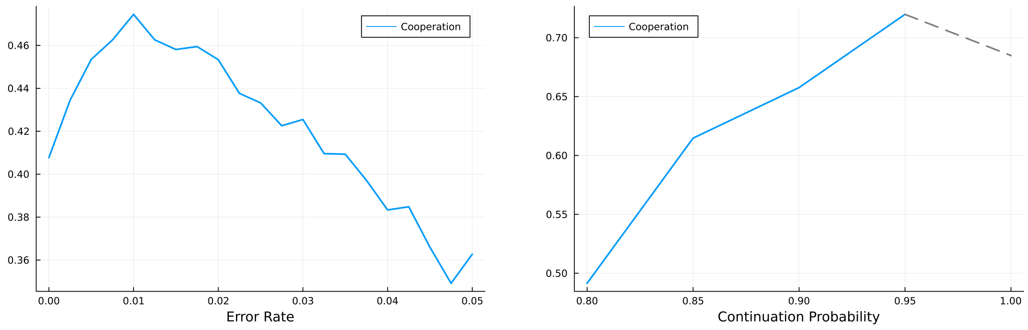
---

[5] In the simulations we do here, we assume that we are in the low mutation limit. This allows us to use a computationally much more efficient approach, in which we determine the fate of mutants arriving by just calculating the fixation probability numerically, and drawing from the resulting straightforward coin flip.

add small amounts of errors.

The second condition in Theorem 3 is also satisfied for the Moran process if the parameter in the theorem represents the continuation probability, and the error rate is fixed, as long as the continuation probability is not 1. To see why Condition 2 is not satisfied at a continuation probability of 1, with or without errors, we can consider strategies that one could call "All C with a handshake". We denote these with $d_T C$; when playing against copies of themselves, they first go through T defecting states before they end up in a cooperative state, which is never left. The left derivative at $\delta = 1$ keeps increasing with T, and is not bounded. The derivative of the transition probabilities at $\delta = 1$ in a population state that contains some individuals that play $d_T C$ and some that play another strategy then is not bounded. That does not mean that it is easy to construct a combination of a strategy set and a mutation kernel for which the the dynamics are in fact discontinuous at $\delta = 1$, but it does show that Theorem 3 cannot be used to show continuity for a model with a strategy set that contains all FSAs.

The following section does however contain an example that shows that it is not true that changes in the invariant distribution are continuous in changes in the parameter space, if we take the latter to be $[0, 1] \times \left[0, \frac{1}{2}\right]$ for (δ, ϵ). In this case, there may be sequences converging to $(\delta, \epsilon) = (1, 0)$ that get arbitrarily close to full cooperation (or more precisely, there may be sequences that, besides ϵ going to 0 and δ going to 1 also include the population size going to infinity) while the average amount of cooperation at an error rate of 0 and without discounting is not 100%. This implies that there is space for dynamic implications of results in Fudenberg and Maskin (1990) and Dal Bó and Pujals (2020). Both suggest that there are equilibria that become fully efficient in the limit of ϵ going to 0 and δ going to 1, that also become more stable than other equilibria in a sequence along which ϵ goes to 0, and δ goes to 1. In Section 9.1 we will return to the relation



(a) Cooperation as a function of ϵ

(b) Cooperation as a function of δ

Figure 8: **Cooperation in the simulations allowing for all FSAs**. Panel (a) shows the average cooperation rates as a function of the error rate for in simulation runs with $\delta = 0.8$, $N = 200$, $\frac{b}{c} = 3$, and $2 \cdot 10^8$ mutant arrivals. Panel (b) shows the average cooperation rates as a function of the continuation probability in simulation runs with $\epsilon = 0.1$, $N = 200$, $\frac{b}{c} = 3$, and 10^7 mutant arrivals. In the interval between $\delta = 0.95$ and $\delta = 1$ simulations become prohibitively expensive as games last very long in expectation, and payoff sums take longer to converge. At $\delta = 1$, payoffs can be computed in a different, not too expensive way. We therefore do not know how the curve plotted in panel (b) behaves between the points connected by the dashed line.

between the results in Dal Bó and Pujals (2020) and what happens in the dynamics.

Theorem 3 also applies to other ways in which the dynamics can be perturbed. Instead of execution error, one could for instance introduce perception error. The perturbation moreover also does not have to be an error; another possibility is that one can for instance introduce complexity costs to a varying degree (van Veelen and García, 2019). With complexity costs, one can then also study what happens when these becomes vanishingly small, and this setup then also satisfies the conditions of Theorem 3. This theorem then implies that the invariant distribution of the finite population dynamics for vanishingly small complexity costs will also be the same as it is in the absence of complexity costs. The simulation results in van Veelen and García (2019) confirm that. In the same way that results concerning static equilibrium concepts in Fudenberg and Maskin (1990) and Dal Bó and Pujals (2020) suggest that there are fully cooperative equilibria that are more stable than not fully cooperative ones in the limit of vanishingly small error rates, and continuation probabilities close to 1, results in Binmore and Samuelson (1992) suggest that the same might be true for vanishingly small complexity costs and continuation probabilities close to 1. This turns out not to be a property that is reflected in the dynamics (see also Cooper, 1996 and Volij, 2002).

What is different between adding complexity costs and adding errors, is that with complexity costs, the level of cooperation in the invariant distribution with complexity cost approaches the level of cooperation in the absence of complexity costs *from below*, when complexity costs vanish. That implies that adding complexity costs not only does not get the dynamics arbitrarily close to full cooperation, it is actually *bad* for cooperation to add complexity costs. Why complexity costs hinder rather than help the evolution of cooperation in this setting, is discussed in detail in van Veelen and García (2019). With execution errors, on the other hand, this is not the case. As we have seen, in the simulations, the average amount of cooperation initially increases as the error rate increases. That suggests that adding errors, in very modest quantities, has a positive effect on the average amount of cooperation that evolves. The following section investigates why that is, and in this section we also see that the fact that errors can have a positive effect on the level of cooperation also opens the door for the possibility that there is a sequence of Moran processes, with error rates going to 0, continuation probabilities going to 1, and population sizes to infinity, for which the cooperation rate goes to 1.

7 Ways out of All D

The purpose of this section is threefold. We first of all consider how errors affect what without errors would be an indirect invasion into All D. We focus on the first step, and see that it is possible that errors *increase* the fixation probability of a mutant that would be neutral in the absence of errors. The second objective is to recreate the hump in cooperation that we see in the simulations in a simple model. The third objective is to show that it is possible to have a sequence of choices for the parameters so that the average amount of cooperation in the dynamics converges to 1.

7.1 Jumpstarters

In the absence of errors, we know that the dynamics are dominated by indirect invasions (see García and van Veelen, 2016 and van Veelen et al., 2012). Stepping stone paths out of All D consist of neutral mutants that would reciprocate cooperation, but that do not initiate it; and then a second mutant that initiates cooperation. The second mutant has a selective advantage if the share of the first is sufficiently large. Our Theorem 3 therefore implies that in the limit of error rates going to 0, the dynamics will be similar, and the same sequences of mutants that drive the dynamics without errors will drive the dynamics with infrequent errors. As error rates increase, starting from 0, it is interesting to see how the probabilities of these transitions change, since average cooperation levels in the simulations initially increase when the error rate goes up (see Figure 8a).

With positive error rates, some strategies that are neutral mutants without errors do not need a second invader to start cooperating amongst themselves. Instead, they will automatically latch on to an error to initiate mutual cooperation. In this section, we therefore look at a class of strategies that we call *jumpstarters*. These jumpstarters defect after any history that consists

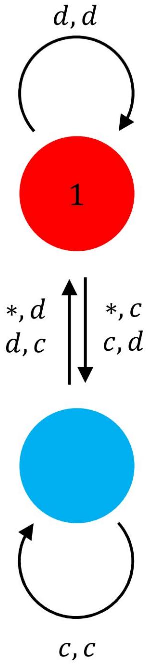


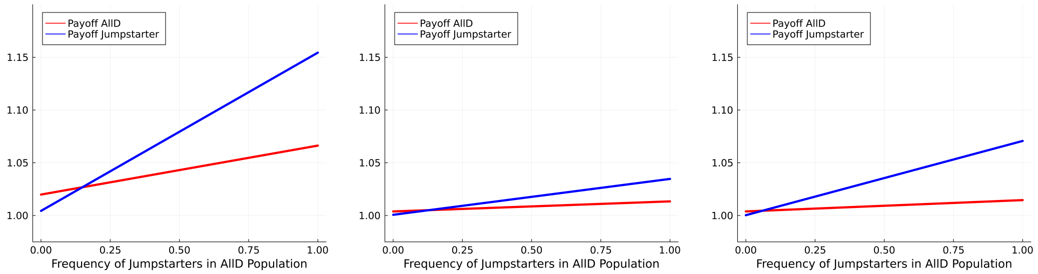
Figure 9: **The standard Jumpstarter.** This strategy goes to state 2 after every action profile that is not d, d. Other jumpstarters, that for instance go to state 2 only after mutual cooperation are also possible. Jumpstarters can also be constructed for residents other than All D, if they leave enough room for increased mutual cooperation. The standard jumpstarter depicted here is self-mirroring and symmetric.

of mutual defections only, and therefore they are neutral mutants of All D at $\epsilon = 0$. They do however use instances of unintended cooperation to switch to their cooperative state, and therefore, at positive error rates, they jumpstart cooperation among themselves. Figure 9 depicts an example of such a strategy. This one switches to the cooperative state after any deviation from mutual defection, and switches back after any deviation from mutual cooperation. In order to establish how the process of these strategies invading All D, and then being invaded themselves, changes as the error rate increases, we will have to carefully contrast static stability concepts and deterministic infinite population dynamics on the one hand – which one might be tempted to think suggests that these jumpstarters cannot invade All D at positive error rates.

For the dynamics, there are three variables that matter: the error rate ϵ, the discount rate δ, and the population size N. The first relevant observation is that for populations that consist of a resident All D and a mutant jumpstarter, the latter has a selective disadvantage at low frequencies. Their attempts to jumpstart cooperation after an error are not reciprocated by players that play All D, and therefore, if they only meet players that play All D, that implies a loss in payoffs that is not compensated for by anything. As a result of this, there is always going to be an interval of frequencies for which their payoff is below the payoff of All D – even though for low error rates, this difference is going to be rather small. The traditional criterion for whether or not mutants can invade that is for instance used in the definition of an ESS, would therefore discard this mutant, because at invasion, it has a payoff that is lower than All D.

If jumpstarters do meet other jumpstarters, however, errors help them to establish mutual cooperation. For a fixed error rate in combination with a sufficiently high discount rate, therefore, there will be a frequency of jumpstarters at which they start outperforming All D. The game between the two then becomes a coordination game. Increasing the δ for a fixed ϵ increases the probability that an error happens before the game ends, and therefore that mutual cooperation ensues between two jumpstarters. This leads to an increase of the mutual payoff between jumpstarters, and that moves the point that separates their basins of attraction to the left (see Figure 10b and 10c).

For low error rates, the gap between the payoff of All D and the payoff of jumpstarters at low frequencies of jumpstarters is small (see Figure 10a and 10b). A low ϵ in combination with a sufficiently high δ can therefore



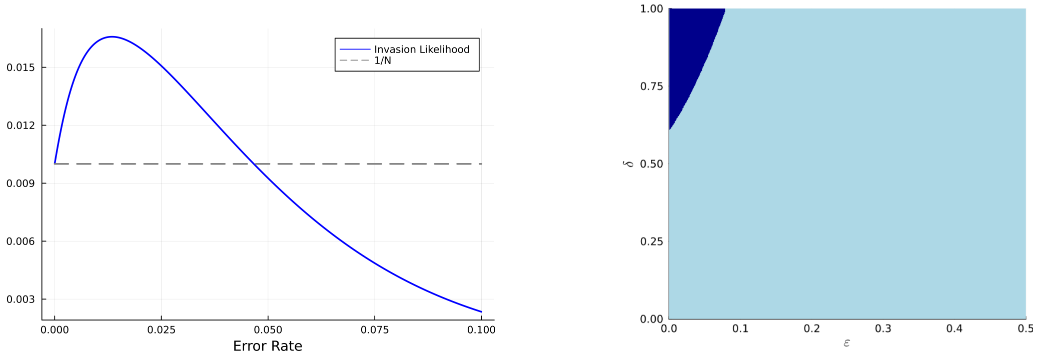
(a) $\delta = 0.8$, $\epsilon = 0.01$ (b) $\delta = 0.8$, $\epsilon = 0.002$ (c) $\delta = 0.9$, $\epsilon = 0.002$

Figure 10: **Payoffs of Jumpstarter and All D.** Lowering the error rate reduces the difference in payoffs for low frequencies of jumpstarters, as can be seen going from panel (a) to panel (b). Increasing the continuation probability exaggerates the difference in favour of the jumpstarters at higher frequencies, as can be seen going from panel (b) to (c). The jumpstarter used here is the standard one from Figure 9. The b/c-ratio is 3.

raise the fixation probabilities of jumpstarters above the fixation probability of a neutral mutant (see Figures 10 and 11). The fixation probability of a neutral mutant is $\frac{1}{N}$, which serves as a benchmark; mutants with a fixation probability higher than this are said to have a selective advantage (Nowak, 2006).

For a fixed discount rate δ, initially, an increase in the error rate only mildly contributes to the disadvantage of the jumpstarter at low frequencies, while it increases the advantage at higher frequencies much more. At higher error rates, further increases in the error rate do not contribute as much to bringing about mutual cooperation between jumpstarters, or to bringing it about earlier, while the increase in error rate does still contribute more or less equally to the disadvantage at low frequencies of jumpstarters. This combines to the hump-shaped probability of fixation we see in Figure 11.

Of course, as the population size grows large, stochastic selection dynamics resemble deterministic ones ever more closely. For a fixed error rate ϵ and a fixed discount factor δ, the probability that randomness gets the pop-



(a) Fixation Probability

(b) Below or over $\frac{1}{N}$

Figure 11: **Jumpstarters invading All D.** Panel (a) shows that the payoff disadvantage that jumpstarters have at low frequencies does not rule out fixation probabilities that are larger than $\frac{1}{N}$, as long as the error rate is small enough. Panel (b) divides the parameter space up in a region where the fixation probability is above and one where it is below $\frac{1}{N}$. The parameters are: $\frac{b}{c} = 3$, $N = 100$, and, in panel (a), $\delta = 0.8$. The jumpstarter here is still the one from Figure 9.

ulation out of the basin of attraction of All D (relative to the probability of neutral drift traveling the same distance) decreases with population size, and this decrease cannot be compensated for by higher certainty that, once in the basin of attraction of the jumpstarter, this jumpstarter goes to fixation. This implies that, for a fixed combination of ϵ and δ, we can alwasy find a large enough population size, so that the jumpstarter in Figure 9 would have a fixation probability below $\frac{1}{N}$. This is illustrated in Figure 12. What is also illustrated by this figure, however, is that for some combinations of a b/c-ratio and a continuation probability δ, however large one chooses the population size N, there is always an error rate ϵ for which the fixation probability of this jumpstarter is larger than $\frac{1}{N}$. This observation is made precise with Theorem 4.

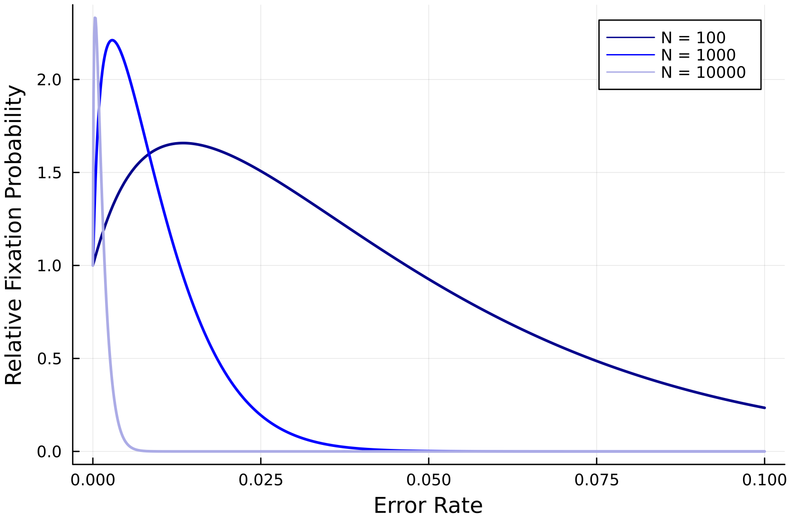


Figure 12: **Fixation probability of Jumpstarter, multiplied by population size N for different values of N.** Values above 1 indicate higher than neutral invasion likelihood. For a fixed error rate, there is always a population size, such that the fixation rate drops below $\frac{1}{N}$. For every population size N, on the other hand, there is an error rate ϵ such that the fixation probability is above $\frac{1}{N}$. The parameters are $\frac{b}{c} = 3$ and $\delta = 0.8$.

Theorem 4. *For any combination of population size N and any $\frac{b}{c} > 1$, if δ satisfies*

$$\delta > \frac{3b + (3N-1)\,c}{(N+1)\,b + (2N-1)\,c}$$

then there exists an $\bar{\epsilon} > 0$ so that for error rates in the open interval $\epsilon \in (0, \bar{\epsilon})$ the jumpstarter depicted in Figure 9 has a probability greater than $\frac{1}{N}$ of fixating in a population of All D.

Proof. See Appendix C □

This theorem implies that taking the limit of the population size going to infinity does not render this jumpstarter irrelevant for the dynamics for error rates that are not bounded away from 0. Roughly speaking, for $\delta > \frac{3c}{2c+b}$,

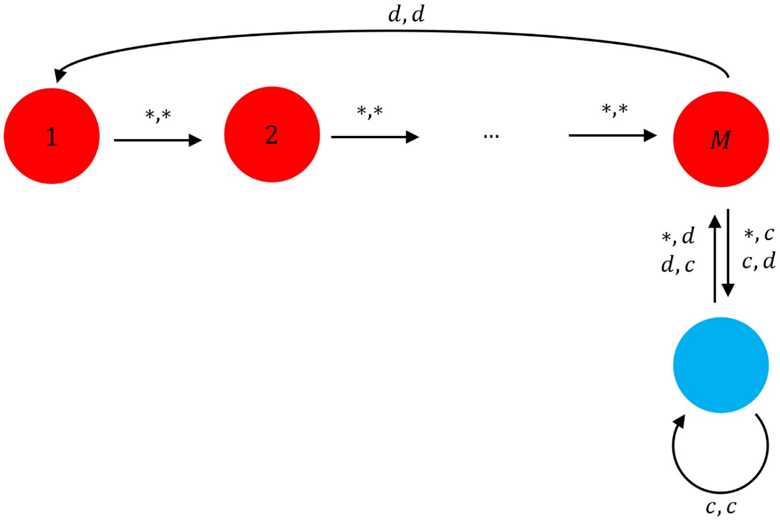


Figure 13: **Jumpstarter with a lower jumpstarting rate.** This strategy regulates the jumpstarting rate by only allowing for errors to make it go to the cooperative state once every M rounds.

every population size N comes with an error rate ϵ that is low enough to make the jumsptarter from Figure 9 have a higher than neutral fixation probability.

In Theorem 4 we fix the pair of strategies, and let the error rate ϵ shrink as the population size N increases. However, we can also keep the error rate fixed, and change which strategy we consider the fixation probability of. Rather than being tied to a jumpstarter that gets triggered into transitioning to a cooperative state by every error, and then relying on a decrease of the error rate for lowering the incidence of such transitions, one can also have jumpstarters that endogenously reduce the incidence, by ignoring some errors, and only transitioning after others. This way, one can show that also for error rates that are not low, there are jumstarters that have a fixation probability larger than $\frac{1}{N}$.

Theorem 5. *For $N \to \infty$, if δ satisfies*

$$\delta > \frac{3c}{(1-\epsilon)(1-2\epsilon)b + (1-\epsilon)(2-\epsilon)c}$$

there exists a jumpstarter that has a probability greater than $\frac{1}{N}$ of fixating in a population of All D.

Proof. Here we only give an outline; the complete proof is in Appendix D. We construct a jumpstarter that, in the absence of errors, cycles through $M+1$ states and defects in each of these states. In the first M states this jumpstarter moves from state i to state $i+1$ independently of the actions that are played. In state $M+1$ this new jumpstarter behaves akin to the jumpstarter considered in Theorem 3: if no error occurs, it returns to state 1, and if an error occurs, it moves to a cooperative state, and stays there until a further error occurs, which brings her back to state 1 (see Figure 13). This construction has the advantage that the jumpstarting rate can be made arbitrarily small, by choosing a sufficiently high number M. We can therefore use arguments similar to those in the proof of Theorem 4, where the jumpstarting rate was set by (arbitrarily small) ϵ.

□

We can illustrate this result by drawing the boundaries of the region within which jumpstarters exist that have fixation probabilities above $\frac{1}{N}$. These boundaries depend on the b/c-ratio, and are shown in Figure 14.

It is worth noting that these jumpstarters themselves are not necessarily equilibria. For low error rates, on average it takes long for an error to happen. A mutant that does not leave it up to chance, and just starts cooperating right away, might therefore have a selective advantage if the jumpstarter has gone to fixation. Such a sequence of mutants (first a jumpstarter, and then a strategy that initiates cooperation without waiting for errors to happen) then exactly matches what an indirect invasion would be without errors. We

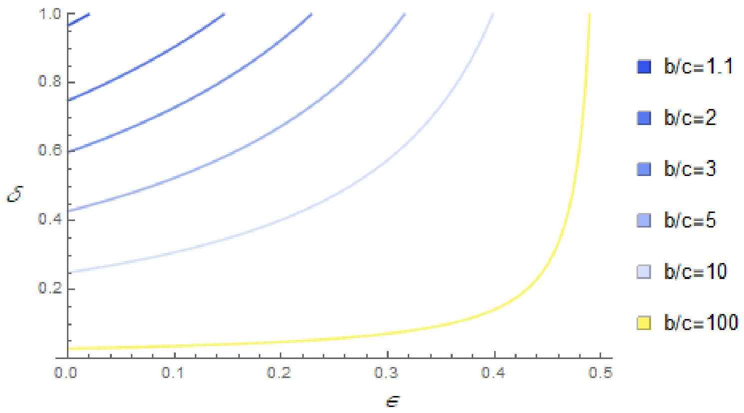


Figure 14: **Thresholds Theorem 5.** The curves correspond to the thresholds indicated in Theorem 5. To the top left of each curve, there exists a jumpstarter with sufficiently low jumpstarting rate (as sketched in Figure 13) which has a greater than neutral fixation probability in large populations ($N \to \infty$).

know from García and van Veelen (2016) and van Veelen (2012) that these indirect invasions drive the dynamics at an error rate of 0, and Theorem 3 tells us that the changes in the likelihoods of these transitions therefore must be relevant as we increase the error rate from 0.

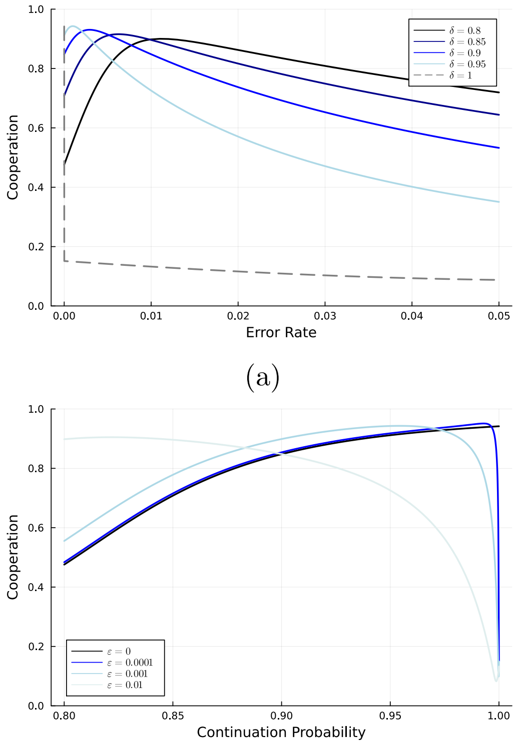
Another element to consider is that here, we take the error rate to be exogenous. Alternatively, we could allow the error rate to be an individual characteristic, or we could allow FSAs to randomise in order to determine the output in a state. In either case, the fidelity itself (of the strategy as a whole, or of the output in any given state) would then be subject to evolution. In such an alternative setup, sometimes higher error rates could be selected for, and sometimes lower error rates – just like sometimes cooperation is selected for, and sometimes defection. In a population of All D, for instance, lower error rates would be selected for. A population that consists of error-free All D might however be vulnerable to invasion by a jumpstarter. This would then require a sequence of mutations, in which they first acquire the capacity to be triggered into cooperation, and then the capacity to trigger others to cooperate, for instance by mutating into a strategy with a positive error rate.

7.2 A model with 4 strategies

If we consider a reduced strategy space that includes All C, All D, Grim Trigger, and the jumpstarter from Figure 9, then this contains a fully cooperative equilibrium (Grim Trigger) and a fully defecting one (All D). At error rate 0, Grim Trigger and All C are neutral mutants of each other. All C then is vulnerable to invasion by All D, so a sequence of first All C, and then either All D or the jumpstarter constitutes an indirect invasion into Grim. Also All D and the jumpstarter are neutral mutations of each other at error rate 0. The jumpstarter is vulnerable to invasion by All C or Grim, so these paths would constitute indirect invasions into All D.

For this model, we see a clear hump in the average amount of cooperation, as it changes with error rate ϵ (see Figure 15a). This hump moves to the left, and the size of the peak increases, as the continuation probability approaches 1. For a fixed population size, the amount of cooperation at the peak does however not approach 1. There maybe no reason to expect that it would, given that this model is restricted to only four strategies, but with results from Fudenberg and Maskin (1990) and Dal Bó and Pujals (2020) in mind, one might hope to see the peak approach 1. Making that happen might however also require the population size to be taken into account, and below we will see that this does not even require making sure we pick the right sequence of δ's.

For every $\epsilon > 0$, we also see a hump if we track the average amount of cooperation as a function of the continuation probability δ (see Figure 15b). There is a subtle detail, which is that for some error rates, as the continuation probability approaches 1, the cooperation rates first dips below what they



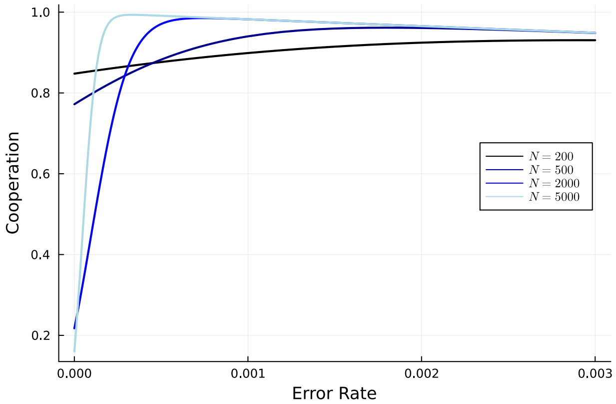
(a)

(b)

Figure 15: **Average amounts of cooperation in the four-strategy model.** In this model, there are four strategies: All C, All D, Grim Trigger and the Jumpstarter from Figure 9. We numerically compute the stationary distribution and resulting cooperation rates, assuming equal mutation rates between all four strategies. The parameters are: $N = 200$ and $b/c = 3$.

are at $\delta = 1$, and then approach the cooperation rates at $\delta = 1$ from below. Figures 15a and 15b moreover both show the discontinuity at $\epsilon = 0$ for the continuation probability fixed at $\delta = 1$. As noted in Section 6, this is caused by the discontinuity in payoffs of Grim Trigger and Jumpstarter at $\epsilon = 0$ for the continuation probability fixed at $\delta = 1$. The four-strategy model we consider here does not require the continuation probability to approach 1 for this to happen.

In this model, we can take a fixed δ, and if this continuation probability is



(a)

Figure 16: **Approaching full cooperation by increasing the population size.** In the same four-strategy model, for a combination of $\epsilon = 0.01$ $\delta = 0.9$, we find that increasing the population size gets the average amounts of cooperation ever closer to 1.

sufficiently high, there will be a sequence of ϵ's and population sizes, where the error rate goes to 0 and the population size to infinity, such that the cooperation rate approaches 1 (see Figure 16a). That is in a sense even more spectacular than what one would imagine results in Fudenberg and Maskin (1990) and Dal Bó and Pujals (2020) might imply for dynamics with strategy sets that are not similarly restricted. The results in those papers might be taken to suggest that there can be a sequence of combinations of ϵ, δ, and N, such that $\epsilon \to 0$, $\delta \to 1$ and $N \to \infty$, for which the cooperation rate in the associated dynamics approaches 1.

8 Forgiveness and punishment in a simple model with a restricted strategy set

In this section we consider a restricted set of strategies in order to explore selective pressures for or against strategies that differ in how long they punish. There are of course dimensions other than the punishment length along which punishment can differ. Contrite Tit-for-Tat, for instance, requires its partner to first "make things right" by cooperating, while Contrite still defects. Thereby it makes forgiveness contingent on what its partner does. As we have seen in Section 5, at high error rates, there are differences between on the one hand asymmetric strategies like Contrite Tit-for-tat, that is an equilibrium at high error rates, thanks to its forgiveness, and symmetric strategies, like Grim Trigger, that need to be rather unforgiving in order to be be equilibrium. In the set we consider here, however, strategies are symmetric, and forgiveness is not conditional. Within this set, at relatively modest error rates, there is an intermediate level of punishment that is the most frequent in the invariant distribution.

The set of strategies that we consider here consists of strategies S_n for $n = 0, ..., 20$, and is completed by All D. A strategy S_n cooperates as long as both players cooperate (see Figure 17). If any deviation from mutual cooperation is observed, while in the cooperative state, it subsequently defects for n rounds, and then resumes cooperation. Strategies with higher n, thus, punish defections more harshly. The strategy S_0 is also included, and this is of course the same strategy as All C.

For this set of strategies, Figure 20 reflects the sizes of the basins of attraction between any pair of strategies in this set. Besides the details of the geometric pattern, which we will get to below, there are two main ingredients that are visible in this figure. The first is that in pairs that consist of All D and S_n for some n, the higher n, the smaller the basin of attraction of All

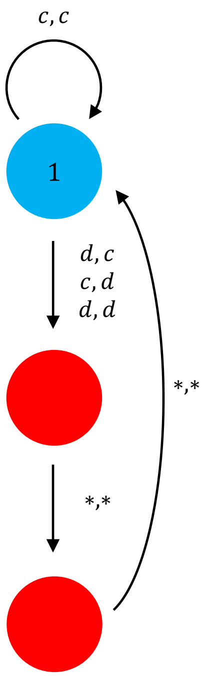
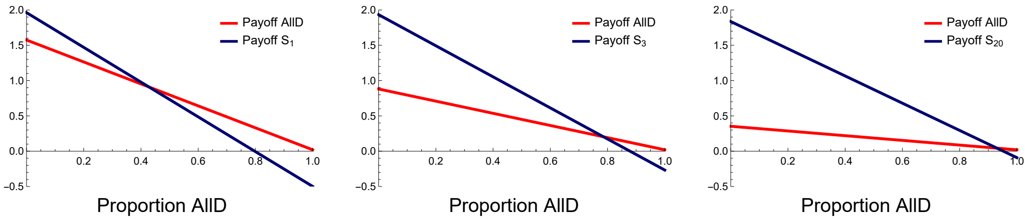


Figure 17: **A strategy with a fixed punishment duration: $\mathbf{S_2}$**. When this strategy finds itself in state 1, then after any deviation from mutual cooperation, it will (try to) play defect for two periods, and then return to the first state, regardless. The strategies S_n are self-mirroring, and therefore they can be Nash equilibria, depending on the b/c ratio, the δ and the ϵ.

D (see Figure 18). This is because for higher n, S_n punishes more harshly, and therefore spend less time cooperating against All D, which, obviously, never intentionally reciprocates. This is consistent with what we see at the top row and at the right column of Figure 20.

Between strategies S_{n_1} and S_{n_2}, on the other hand, the basin of attraction of strategies with the shorter punishment phase is larger than the basin of attraction of the strategy with the longer punishment phase (see Figures 19 and 20). When playing against a copy of oneself, being able to quickly resume mutual cooperation after an error, is good for payoffs. The difference in the

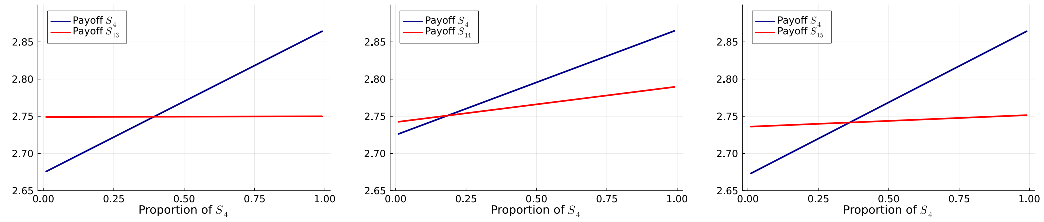


(a) S_1 vs. All D (b) S_3 vs. All D (c) S_{20} vs. All D

Figure 18: **Payoff of S_1, S_3, and S_{20} vs All D with infrequent errors.**. The plots show average payoffs per stage game for $\frac{b}{c} = 3$, $\delta = 0.9$, and $\epsilon = 0.01$. As residents, all three strategies have a payoff advantage. The point at which the payoffs intersect separates their basis of attraction. The size of basin of attraction of S_n increases with the punishment length.

payoffs that are attained against copies of themselves dominates the fact that in mixed matches, the strategy with the shorter punishment phase cooperates a bit more in expectation. The strategy with the shorter punishment length therefore has the larger basin of attraction.

The non-monotonicities in Figure 20 that create the geometric structure are caused by the fact that for some combinations of strategies S_{n_1} and S_{n_2}, the respective punishment lengths n_1 and n_2 satisfy $k_1 \cdot (n_1 + 1) + n_1 = k_2 \cdot (n_2 + 1) + n_2$ for relatively small integers k_1 and k_2. If relatively small k_1 and k_2 can be found for which this is satisfied, then after a defection, the strategies S_{n_1} and S_{n_2} return to cooperation relevantly soon. If, on the other hand, the equality is only satisfied for larger k_1 and k_2, then it is much more likely that a match between S_{n_1} and S_{n_2} ends before cooperation is resumed at all. All of this implies that for combinations of strategies that return to cooperation sooner, their mutual payoffs are higher compared to combinations of strategies that take (much) longer. The high mutual payoffs exaggerate the effect of the difference between how the two strategies do against copies of themselves (see Figure 19). Therefore, in the left upper half, the small basins of attraction of the residents become even smaller if



(a) S_4 vs. S_{13} (b) S_4 vs. S_{14} (c) S_4 vs. S_{15}

Figure 19: **Payoff of $\mathbf{S_4}$ vs. $\mathbf{S_{13}}$, $\mathbf{S_{14}}$, and $\mathbf{S_{15}}$.** In the absence of errors, the strategies S_4 and S_{13} return to cooperation after $69 = 13 \cdot (4+1) + 4 = 4 \cdot (13+1) + 13$ periods. S_4 and S_{14} return to cooperation after $29 = 5 \cdot (4+1) + 4 = 1 \cdot (14+1) + 14$ periods. S_4 and S_{15} return to cooperation after $79 = 15 \cdot (4+1) + 4 = 4 \cdot (15+1) + 15$ periods. While the intersection generally moves leftwards, as can be seen by comparing panel (a) and (c), a reduced punishment length in mixed pairs leads to higher payoffs within such pairs, exaggerating the deviations from 0.5. The parameters are: $\delta = 0.9$, $\frac{b}{c} = 3$ and $\epsilon = 0.01$

low numbers k_1 and k_2 exist, and in the right lower half, the large basins of attraction of the residents become even larger.

If we then assume a mutation process in which mutations from any strategy to any other strategy happen with the same probability, then the two main effects combined imply that within the set of strategies S_n, the tendency is for punishment lengths to get ever shorter. This also holds for pairs of strategies in which the shorter punishment length is too short to make S_n an equilibrium strategy. Obviously, All D can easily invade S_n if S_n is not an equilibrium (and therefore its basin of attraction has size 0), but also in general, the shorter the punishment length gets, the smaller its basin of attraction, and the more likely All D is to invade. All D itself, in turn, is left with a bias towards strategies with longer punishment phases, because, as we have seen, its invasion barrier against S_n decreases as n grows larger. These ingredients together results in invariant distributions in which strategies with very short or very long punishment phases have relatively low weights, while

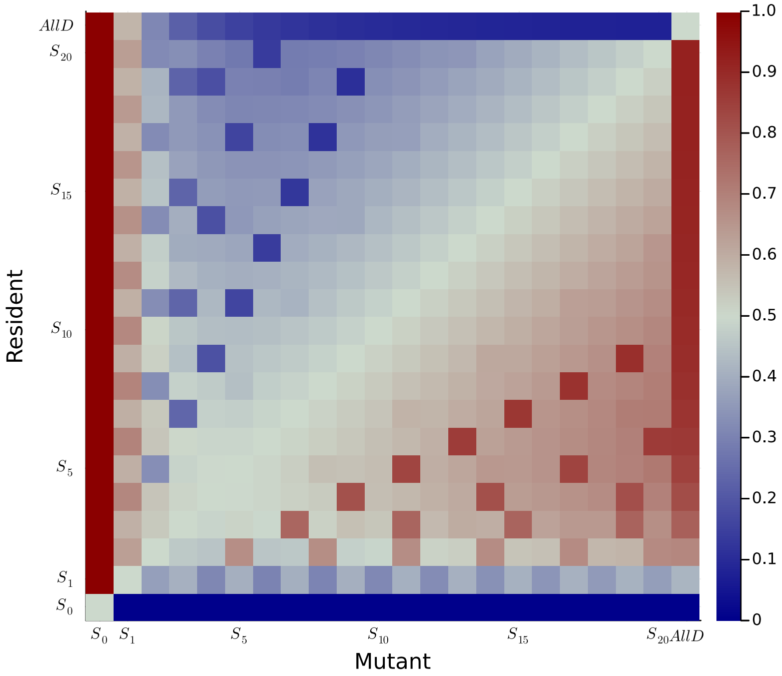


Figure 20: **Basins of Attraction.** Colors indicate the largest mutant-frequency at which the resident has a payoff advantage over the mutant, i.e. the boundary of the basin of attraction of the resident against the mutant. Parameter values: $\frac{b}{c} = 3$, $\delta = 0.9$ and $\epsilon = 0.01$.

strategies with intermediate punishment lengths have higher weights (Figure 21).

For higher $\frac{b}{c}$ ratios the peak of the invariant distribution shifts towards strategies that punish for fewer periods. In the limit of errors going to 0, all strategies S_n get identical payoffs against each other (or, in other words, they are all neutral mutants of each other) and the only difference between them is the basin of attraction when they are partnered with All D. Given that this is the only one ingredient that remains, and all strategies S_n are neutral against each other, the weights of S_n in the invariant distribution are increasing in n (see Figure 22).

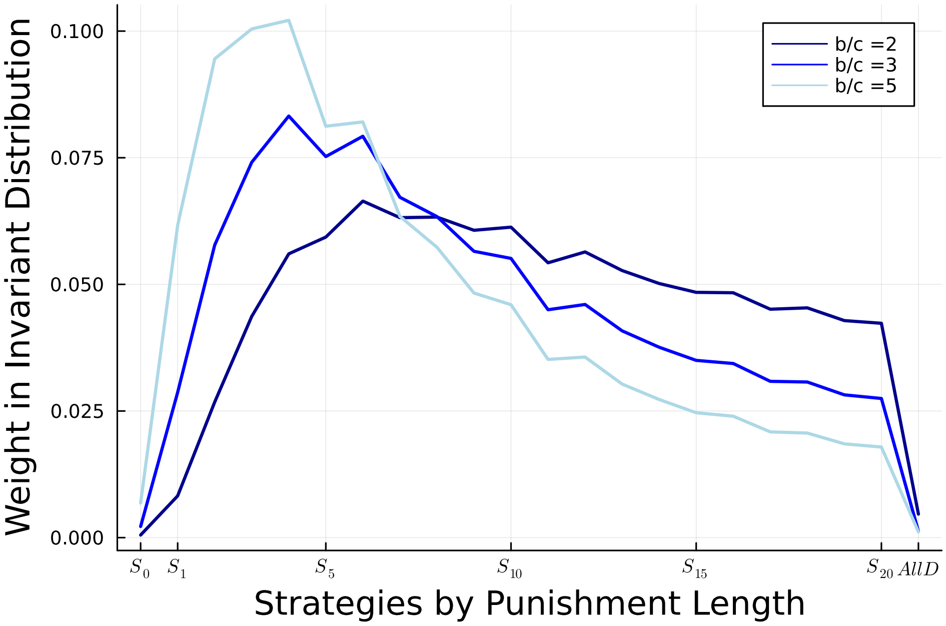


Figure 21: **The stationary distribution for different $b/c-$ratios**. Mutations happen u.a.r. between all 22 strategies. Parameters are: $N = 100$, $\delta = 0.9$, and $\epsilon = 0.01$. Because of the geometrical pattern in Figure 20, the stationary distribution looks a bit rugged. S_0 is All C, which is never an equilibrium, and for this ϵ and δ, S_1 is not an equilibrium for a b/c-ratio of 2, but it is for $b/c = 3$ and $b/c = 5$. All other strategies are equilibria for all b/c-ratios, but the basins of attraction of All D against strategies with high punishment length are small.

The shifting balance between the negative effect on the payoff when playing against a copy of oneself of punishment being long and harsh, and the negative effect on the payoff when playing against All D of punishment being (too) short and forgiving can lead to different patterns. For a $b/c-$ratio of 3, there is just a valley in the punishment length (or a hump in forgiveness); and for a $b/c-$ratio of 2, this valley is preceded by a little hump. For a $b/c-$ratio of 2, there is also a little hump in cooperativeness, which indicates that this does not need the presence of jumpstarters (see Figure 23).

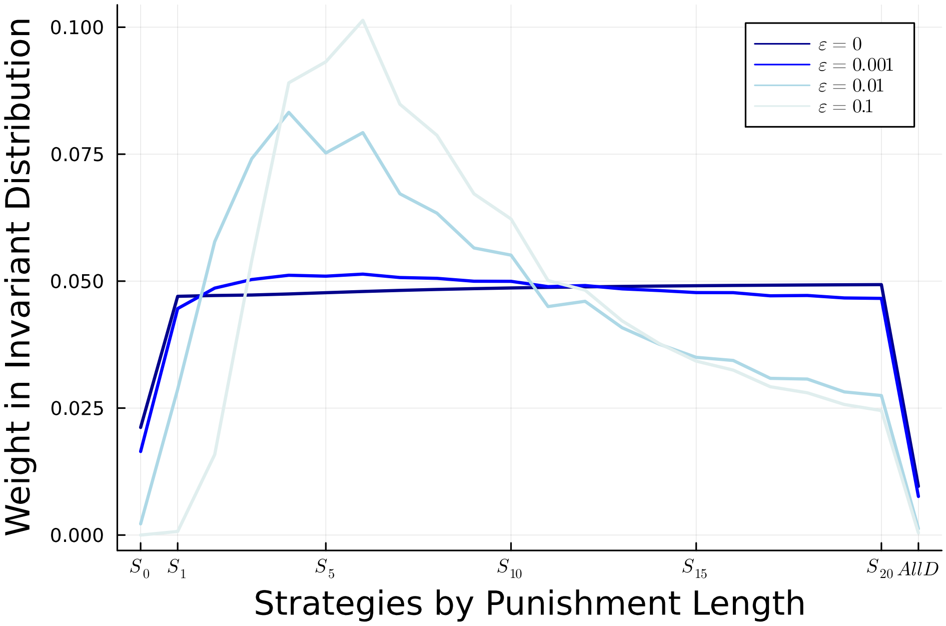
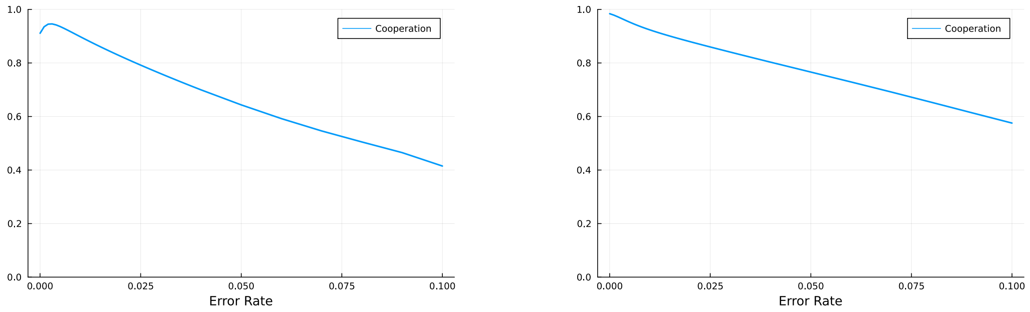


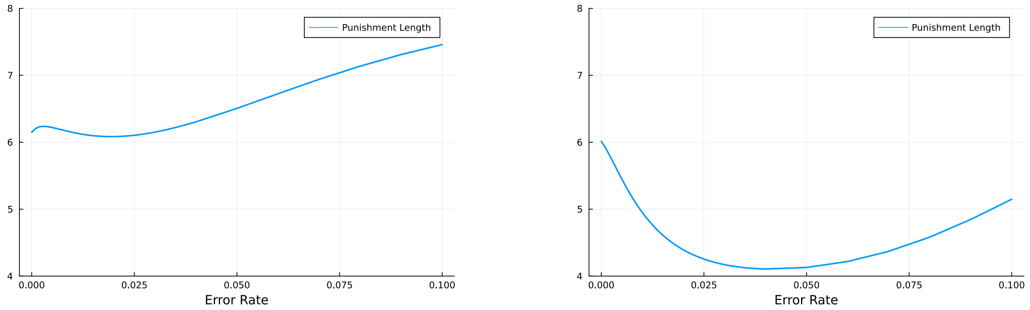
Figure 22: **The stationary distribution for different error rates**. Mutations happen u.a.r. between all 22 strategies. Parameters are: $N = 100$, $\delta = 0.9$, and $\frac{b}{c} = 3$. Because of the geometrical pattern in Figure 20, the stationary distribution looks a bit rugged. S_0 is All C, which is never an equilibrium. All other strategies are equilibria for all error rates, but the basins of attraction of All D against strategies with high punishment length are small.

9 Forgiveness and punishment as a function of the error rate in the simulations

If we want to explore how the amount of forgiveness in the simulations changes as the error rate increases, we need a way to measure forgiveness. One of the ways to do that would be to follow the definition in Dal Bó and Pujals (2020) that separates strategies into two categories; strategies that



(a) Cooperation at $\frac{b}{c} = 2$ (b) Cooperation at $\frac{b}{c} = 3$



(c) Punishment at $\frac{b}{c} = 2$ (d) Punishment at $\frac{b}{c} = 3$

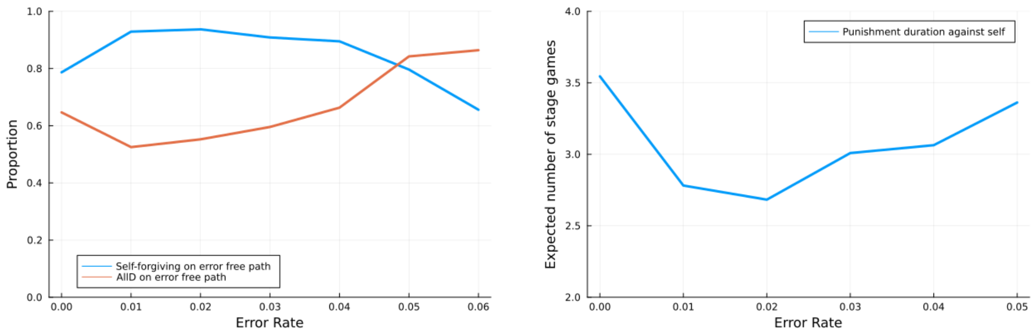
Figure 23: **Cooperation and punishment in the reduced model as a function of the error rate.** The level of cooperation in panels (a) and (b) is calculated as the weighted sum of the proportions at which the individual strategies, S_0 to S_{20} and All D, as residents play cooperate at a given ϵ. The weights reflect the proportion of time the population spends with a the respective strategy as a resident, and are calculated like in Figure 21 as the strategies' weights in the stationary distribution of a Moran process, where mutations happen with uniform probability between all strategies in the reduced strategy set, and where the mutation rate is infinitesimally small. Analogously, the punishment length in panels (c) and (d) is calculated as a weighted sum of expected (discounted) punishment lengths of the individual strategies. For strategy S_n the expected punishment length is computed as $\frac{1-\delta^n}{1-\delta}$. As in Figure 24, All D is not considered in this weighted sum, and the weights of all remaining strategies are normalised to add up to one. The continuation probability is $\delta = 0.9$.

are forgiving and strategies that are unforgiving. A strategy s then is called *unforgiving* if there exists a history h_t such that $s\left(h_t h_\tau\right) = D$ for any history $h_t h_\tau$ – where the latter is a notation that subdivides a history that starts with h_t into that initial sequence and the remainder. For FSAs, that means that a strategy is unforgiving, if it has a defecting state from which there is no sequence of transitions that brings it to a cooperative state. All strategies S_n discussed in the previous section are forgiving; they always return to the initial state from any other state, and they even do so regardless of what is being played. Contrite Tit-for-Tat, both the standard one depicted in Figure 2, and the version in Figure 3 also has a clear path back to the initial, cooperative state. On the error-free path, jumpstarter stays in the first, defecting state, but there are paths to the second, cooperative state. If a defection happens while in that state, then it returns to the first state, for which there is a path (back) to the cooperative state. This strategy therefore is also classified as forgiving. Tit-for-Tat is also forgiving, but Grim Trigger is not.

According to this definition, All D is also an unforgiving strategy. In their paper, that is conceptually and practically useful, as it is the inefficiency of continuation payoffs that is central to their argument, and this does not discern between the whole repeated game and a subgame that is reached after one or more errors. One could however also argue that, when All D plays against All D, it has nothing to be forgiven for. We could therefore also choose to measure forgiveness only within in the set of strategies excluding All D. Whether we include All D or not, what we find is that almost all of the strategies that are present in the population, for a variety of combinations of error rates and continuation probabilities, are either All D, or they are a strategy that is forgiving according to the definition from Dal Bó and Pujals (2020). This however is not the result of selection acting for or against strategies that are unforgiving; it is just a consequence of the fact that in our setting with errors, every state in an FSA has four outgoing arrows, which makes it really unlikely that, if an FSA has a cooperative state, there is a defecting state from which this cooperative state could never be reached – even if the sequence of errors and non-errors that this would require is very unlikely.

Another possibility for a definition of forgiveness is to consider what happens when an FSA plays against a copy of itself. For every cooperative state on the error-free path, we can then check if it has the property that if the opponent (and only the opponent) erroneously defects, while both are in that state, the automaton will eventually return to a cooperative state in the absence of further errors. If that is the case for all cooperative states on the error-free path, then we call this strategy *self-forgiving*. The strategies S_n all qualify as self-forgiving, and so does Contrite Tit-for-Tat. Grim Trigger is not self-forgiving, and neither is jumpstarter. Figure 24a shows that if we do use this binary measure of forgiveness, and we vary the error rate, there is a little hump in the amount of time the system spends in states with a self-forgiving resident.

This definition can however be a bit permissive, because strategies S_n, as discussed in Section 8, all qualify as self-forgiving, even if n is very large. For a non-binary measure for forgiveness that would reflect the difference between S_2 and S_{20}, we can also go over all cooperative states on the error-free path. For all of those, we can calculate the expected number of stage games that it takes after a unilateral defection, while in that state, until the automaton returns to a cooperate state, assuming no further error occurs. We can then take a weighted average of these times to return to a cooperative state over all the cooperate states on the error-free path, with weights proportional to the expected time spent in each of these cooperate states assuming that no error occurs. These per-automaton averages are then obviously themselves in turn weighted by the time spent with them as residents, where we exclude



(a) Self-forgiving residents

(b) Punishment length

Figure 24: **Forgiveness as a function of the error rate**. The parameters are: $\frac{b}{c} = 2$, $N = 200$ and $\delta = 0.9$. In panel (a), the red line indicates the proportion of time that the population spends with a resident that plays All D on the error-free path. For the blue lines in panel (a) and (b), only residents who do *not* play All D on the error-free path are considered. Both lines are constructed by looking at how the resident automata respond to single error-caused unilateral defections in those cooperate states which are reached on the error-free path when playing against copies of themselves. The blue line in panel (a) indicates the time that is spent with a resident that is *self-forgiving* (see the text for the definition) as a proportion of the time spent with all residents that do *not* play All D on the error-free path. The blue line in panel (b) indicates the average of the expected number of stage games that it takes until the automaton returns to a cooperate state after a unilateral defection (again, see the text for a more precise definition).

strategies that play All D on the error-free path. In Figure 24b, we see that for this measure, there is a little dip in the error-free punishment length, where a decrease in punishment length is of course seen as an increase in forgiveness.

As with cooperation, we therefore find that with either definition, forgiveness peaks at moderate, but non-zero error rates. In light of the results presented so far, this is no surprise. For one, selection on states that when meeting copies of oneself are only reached after an error is weak if errors are very infrequent. Similarly, as in section 7, mutants that jumpstart cooperation in a subgame in which the resident is unforgiving are more successful at moderate, but non-zero error rates.

Both of these arguments suggest that, how strongly the dynamics favour strategies which strike a good balance between forgiveness and punishment, varies with the error rate. However, what constitutes a good balance between forgiveness and punishment, is also a function of the error rate.

9.1 Relation to results in Dal Bó and Pujals (2020)

The results in Dal Bó and Pujals (2020) center around comparing the sizes of invasion barriers of different equilibria. In their paper, Lemma 1 states that All D does not have a uniformly large basin of attraction; Theorem 1 states that unforgiving strategies also do not have a uniformly large basin of attraction; and Theorem 2 states that if a strategy s has a uniformly large basin of attraction and is symmetric, then it is asymptotically efficient. A natural setting in which the size of the basin of attraction is the only ingredient one should focus on for understanding finite population dynamics, is, first of all, one in which there is a set of ESSs, or, in other words, in which there is a set of equilibria that all have positive invasion barriers against all mutants. Equilibria, moreover, would be left very rarely, and these rare transition would be caused by sets of agents making the same error simultaneously, thereby jumping from one equilibrium into the basin of attraction of the other. Once in the basin of attraction of the other equilibrium, selection then deterministically takes the population to this new equilibrium. These are the dynamics considered in Young (1993) and Kandori et al. (1993), and these dynamics justify the static stability concept of a *stochastically stable strategy* in Young (1993), or a *long-run equilibrium* in Kandori et al. (1993). In such a dynamics, obviously, the relative sizes of the basins of attraction are all that matters, as those determine how likely simultaneous errors move a population outside the basin of attraction of an equilibrium.

In the finite population dynamics that we consider, the size of the basin of attraction is not the only determinant of how easy or hard it is to leave an equilibrium. The Moran process is a stochastic, finite population version of the replicator dynamics. In this dynamic, there is noise, and getting out of a basin of attraction of one equilibrium is not happening as a result of a set of individuals that simultaneously make the same mistake, or that mutate into the same alternative strategy at the same time. Instead, one mutant with a payoff disadvantage may arise, and randomness in reproduction may result in the share of this mutant growing anyway. This would then have to happen over subsequent generations. The probability of making it out of the basin of attraction of the resident then not only depends on how far away the invasion barrier is, but also how shallow or deep the basin of attraction is (or, in other words, how strongly selection pushes back towards the equilibrium within the basin of attraction). The probability of making it out of a large, but shallow basin of attraction may be larger than the probability of making it out of a smaller, but deeper basin of attraction. Especially when we consider low error rates, for which payoff differences between strategies that share the same error-free path are very small, this makes a big difference.

The closer the continuation probability is to 1, however, the more the effect of adding a tiny bit of errors on payoff differences can be amplified. For a fixed δ, the differences between the payoffs of, for instance, 1) Grim Trigger against itself; 2) Grim Trigger against All C; 3) All C against Grim Trigger; and 4) All C against All C vanish in the limit of ϵ going to 0. For a fixed ϵ, on the other hand, these payoffs converge to $\epsilon b + (1-\epsilon)c$, $(1-\epsilon)(b+c)$, $\epsilon(b+c)$, and $(1-\epsilon)b + \epsilon c$, respectively, and those are as different as can be.

For a fixed δ, this implies that it becomes relevant what happens without errors. From García and van Veelen (2016) and van Veelen et al. (2012) we know that the dynamics without errors are characterised by indirect invasions. In those dynamics, all equilibria, in time, are undermined by a sequence of mutants, where first we get one or more neutral mutants that do not change the behaviour within the population. At some point a neutral mutant arises that would not punish deviations, or not sufficiently harshly, and then a mutant that takes advantage of the absence of threat of (sufficiently harsh) punishment invades, and changes what is being played in the population. A relevant observation here is that in the limit of error rates going to 0, the set of strategies that behave the same on the error-free path becomes the set of neutral mutants. Within this set, differences in how harshly they punish may give different strategies within this set different invasion barriers. How large these invasion barriers are all depends on *relative* payoff differences. In the limit of the error rate going to 0, all payoff differences within this set however vanish, and the fixation probability between any pair of neutral mutants becomes $\frac{1}{N}$, regardless of how their basins of attraction relate in the limit. This implies that differences in invasion barriers may be quite large in the limit of the error rate going to 0, while for the dynamics we consider, these mutants become indistinguishable.

Even though the size of the basin of attraction may not be the sole determining factor of stability in the dynamics, the discontinuity in payoffs at $(\delta, \epsilon) = (1, 0)$ allows for the possibility that there is a sequence of δ's and ϵ's, where δ goes to 1 and ϵ to 0, for which the difference with the invariant distribution at $(\delta, \epsilon) = (1, 0)$ becomes sizable. This would be consistent with Dal Bó and Pujals (2020), although it might take the population size going to infinity to really have the cooperation rate approach 1. This is however still speculative, as it is a computational challenge to let the population size become large.

10 Discussion

In this paper, we investigated what kind of strategies evolve in repeated prisoner's dilemmas with execution errors. We have combined theoretical results with simulations in order to make sure that our theoretical results are informative about population dynamics. The dynamics we use is the Moran process, which is a finite population, stochastic dynamics, that allows for transitions between equilibria. The relative stability of the different equilibria, combined with how easily they invade others, is then reflected in the amount of time the population spends in them. We are interested in what happens in the dynamics if the continuation probability is close to 1 and the error rate is close to 0 – which is a classical choice of a part of the parameter space to be interested in – but we are also interested in what happens when it is just the error rate that is close to 0, when it is just the continuation probability that is close to 1, or when neither of the two is close to 0 and 1, respectively. The reason to also be interested in other parts of the parameter space is that there are many ways in which results concerning specific ways to approach $(\delta, \epsilon) = (1, 0)$ may not be informative about realistic combinations of those that we might also be interested in.

10.1 How much subgame perfection?

Since the presence of execution errors implies that all subgames can be reached, one question is whether this leads to selection of equilibria that, in the absence of errors, would qualify as subgame perfect. A strategy that would not be subgame perfect without errors is not a Nash equilibrium with errors. Our first main result is that, in order to be a Nash equilibrium with errors, a finite state automaton (FSA) has to be self-mirroring. We also note that selection will have to be spread thinly on some subgames, or on some states that the FSAs can be in. For high error rates or low continuation probabilities, there are no cooperative equilibria, and since All D is subgame perfect, we find almost only subgame perfection there. For lower error rates and higher continuation probabilities, there are cooperative equilibria (although All D remains an equilibrium too), but the lower the error rate gets, the lower the probability that states off the error-free path are reached. This explains why in the simulations we find that most strategies (other than All D) are close to being a Nash equilibrium, but, when playing against a copy of themselves, they do respond suboptimally in some states, and therefore in some sets of subgames.

10.2 How much forgiveness?

In the simulations, which allow for all FSAs, we find that forgiveness and cooperation follow a hump-shaped pattern as a function of the error rate. Our Theorem 3 implies that the invariant distributions of the Markov chains defined by the Moran process vary continuously as a function of the error rate. Combined with the fact that at error rate 0, there is no reason to expect across the board selection for or against forgiveness, and that in these simulations, we find intermediate amounts of forgiveness at an error rate of 0, the continuity result implies that also for small error rates, we will find intermediate amounts of forgiveness. Increasing the error rate from 0 initially increase the amount of forgiveness, but not to spectacular degrees.

Within a restricted set of strategies, we find that there are two effects, the balance between which can shift as the error rate changes. With errors, strategies that punish more harshly get lower payoffs against themselves, which works against harsh punishment. Strategies that are (very) forgiving, on the other hand, are more vulnerable to invasion by defecting strategies. The shifting balance between these effects can create a hump in forgiveness (or the valley in punishment length), which replicates the hump in forgiveness in the simulations with unrestricted strategy set.

In the parameter region where equilibria get ever sparser, close to the boundary across which there are none (that is, with high error rates, or low continuation probabilities, or some mix of those) we do have theoretical results with respect to forgiveness, and they are mixed. Asymmetric strategies that are still equilibria in this parameter region (such as Contrite Tit-for-Tat) are very forgiving, and they have to be, in order to incentivise those that make mistakes to repent. This set of strategies is very different from the set of symmetric strategies, which do not require anything from the individual that made the error in order to go (back) to a cooperative state. For these symmetric strategies, the last strategies standing, as the error rate increases, or the discount rate decreases, are Grim Trigger and Minimal Grim Trigger (where the latter only goes to the punishment state after a mutual defection). One of our results implies that if neither of these is an equilibrium, then no other symmetric strategy is, and even though it takes more to trigger punishment in Minimal Grim Trigger, strategies that punish less severely if they do, compared to Grim Trigger and Minimal Grim Trigger, stop being equilibria before these harsh punishers do .

10.3 How much cooperation?

The average amount of cooperation for any given strategy depends on what it does on the error-free path, and on how harshly it punishes defections. For a fixed distribution of strategies, increasing the error rate increases the amount of punishment that is triggered, so without changes in the composition of the invariant distribution, and within the subset of equilibrium strategies that cooperate, the amount of cooperation would only go down. In the simulations, however, we find a hump as the error rate increases from 0. This means that the composition of the invariant distribution must be changing. Regarding the punishment of defections, we have just recapitulated that there is an initial shift towards shorter punishment lengths, or towards more forgiveness, which reduces the effect of the increased rate at which punishment is triggered. There must however also be a change in the average amount of cooperation on the error-free path. With a 4-strategy model, we have seen that there is scope for what would be indirect invasions out of defection in the absence of errors becoming more likely relative to what would be indirect invasions out of defection in the absence of errors. The errors therefore can shift the balance of indirect invasions out of defection and out of cooperation. The increase in cooperation that this brings does not reflect the ingredients that make full efficiency stable in Fudenberg and Maskin (1990) and Dal Bó and Pujals (2020).

What we do not investigate in this paper, is the possibility that mutants always being present can make the indirect invasion out of cooperative equilibria harder (Boyd and Lorberbaum, 1987). The idea there is that if there is for instance always a certain minimum fraction of All D around, then in a population that is otherwise composed of Grim Trigger, All C has a (slight) disadvantage compared to Grim Trigger, and stops being a neutral mutant. Because our model assumes the low mutation limit, this element is absent in our simulations. Errors do however play a somewhat similar role, as they also stop All C from being a neutral mutant of for instance Grim Trigger.

References

Axelrod, R., Hamilton, W.D., 1981. The evolution of cooperation. science 211, 1390–1396.

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

Binmore, K.G., Samuelson, L., 1992. Evolutionary stability in repeated games played by finite automata. Journal of economic theory 57, 278–305.

Boyd, R., 1989. Mistakes allow evolutionary stability in the repeated prisoner's dilemma game. Journal of theoretical Biology 136, 47–56.

Boyd, R., Lorberbaum, J.P., 1987. No pure strategy is evolutionarily stable in the repeated prisoner's dilemma game. Nature 327, 58–59.

Cooper, D.J., 1996. Supergames played by finite automata with finite costs of complexity in an evolutionary setting. Journal of Economic Theory 68, 266–275.

Dal Bó, P., Pujals, E.R., 2020. The evolutionary robustness of forgiveness and cooperation. arXiv preprint arXiv:1205.0958 – Updated version (2020) at http://www.econ.brown.edu/Faculty/Pedro_Dal_Bo/forgivenessandcooperation.pdf .

Fudenberg, D., Maskin, E., 1990. Evolution and cooperation in noisy repeated games. The American Economic Review 80, 274–279.

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

García, J., van Veelen, M., 2018. No strategy can win in the repeated prisoner's dilemma: linking game theory and computer simulations. Frontiers in Robotics and AI 5, 102.

Hopcroft, J., 1971. An n log n algorithm for minimizing states in a finite automaton, in: Theory of machines and computations. Elsevier, pp. 189–196.

Kandori, M., Mailath, G.J., Rob, R., 1993. Learning, mutation, and long run equilibria in games. Econometrica: Journal of the Econometric Society , 29–56.

Nowak, M.A., 2006. Evolutionary dynamics: exploring the equations of life. Harvard university press.

Nowak, M.A., Sasaki, A., Taylor, C., Fudenberg, D., 2004. Emergence of cooperation and evolutionary stability in finite populations. Nature 428, 646–650.

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

Sugden, R., 1986. The economics of welfare, rights and co-operation. Oxford: Basil Blackwell.

van Veelen, M., 2012. Robustness against indirect invasions. Games and Economic Behavior 74, 382–393.

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

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

Volij, O., 2002. In defense of defect. Games and Economic Behavior 39, 309–321.

Wu, J., Axelrod, R., 1995. How to cope with noise in the iterated prisoner's dilemma. Journal of Conflict resolution 39, 183–189.

Young, H.P., 1993. The evolution of conventions. Econometrica: Journal of the Econometric Society , 57–84.

A Proof Theorem 2

Proof. We start by reasoning through what a symmetric automaton that maximises incentives to cooperate for any given set of parameters, and that therefore is stable on a maximally large parameter region, looks like. This reasoning will lead us to strategies that in their cooperative state, behave identically either to Grim Trigger or to Minimal Grim Trigger. Finally, we derive the condition for these strategies to be Nash equilibria.

By definition, a cooperative FSA contains at least one state with the output *cooperate*. We consider an arbitrary cooperative state and label it state 1. If all states with the output *cooperate* in a cooperative FSA had the property that the transition associated with observing the action pair (C, D) in that state leads to a state with the output *cooperate*, then the FSA would not be a Nash equilibrium – just as All C is not an equilibrium. Hence, having a state with the output *cooperate* that transitions to a state with the output *defect* when observing a defection is a necessary condition for a cooperative strategy to be a Nash equilibrium, but it is clearly also a sufficient condition for the automaton to be cooperative. So, we can without loss of generality assume that after being defected on in state 1, the FSA immediately moves to a state with the output *defect*, which we label state 2.

The error-free path through the FSA after observing joint cooperation while being in state 1 has to include at least one state with the output *cooperate* – otherwise it would be optimal to already defect in state 1. Moreover, if a sequence of states on the error-free path after joint cooperation in state 1 starts with a state with the output *defect*, then we can increase the value of cooperating rather than defecting in state 1 by skipping this state and moving to the next state on the error-free path. Hence, we can assume for our purposes that after observing joint cooperation in state 1 the FSA moves to another state with the output *cooperate*. As this argument can be applied recursively to any cooperate state on the error-free path starting at state 1, it follows that all states on this path have the output *cooperate*. This makes the problem stationary. Hence, to analyze whether cooperation in an arbitrarily chosen cooperative state 1 is incentive-compatible as a function of the value of being in its punishment state 2, we can simply restrict attention to FSAs that remain in this state 1 after observing joint cooperation in state 1, and that consequently always move to the same punishment state 2 after observing a unilateral defection.

If we index the state that both players move to in case they both defect in state 1 by "3", and denote the value of being in the states 1, 2 and 3 by V_1, V_2 and V_3, then we can write the condition for cooperation in state 1 to be the payoff maximizing choice as follows:

$$\begin{aligned}&(1-\epsilon)(b-c)+\delta((1-\epsilon)^2V_1+2(1-\epsilon)\epsilon V_2+\epsilon^2V_3)\\ &\geq(1-\epsilon)b-\epsilon c+\delta((1-\epsilon)\epsilon V_1+((1-\epsilon)^2+\epsilon^2)V_2+\epsilon(1-\epsilon)V_3)\end{aligned}$$

This we can simplify a bit.

$$\begin{aligned}\delta\left((1-\epsilon)(1-2\epsilon)V_1-(1-4(1-\epsilon)\epsilon\right)V_2-\epsilon(1-2\epsilon)V_3)&\geq(1-2\epsilon)\,c\\ \delta\left((1-\epsilon)(1-2\epsilon)V_1-(1-2\epsilon)^2V_2-\epsilon(1-2\epsilon)V_3\right)&\geq(1-2\epsilon)\,c\\ \delta\left((1-\epsilon)V_1-(1-2\epsilon)V_2-\epsilon V_3\right)&\geq c\\ (1-\epsilon)V_1-(1-2\epsilon)V_2-\epsilon V_3&\geq\frac{c}{\delta}\end{aligned}$$

Under the assumption that this is an equilibrium, we can express V_1 in terms of V_1, V_2 and V_3:

$$V_1=(1-\epsilon)(b-c)+\delta\cdot\left((1-\epsilon)^2V_1+2\epsilon(1-\epsilon)V_2+\epsilon^2V_3\right)$$

Solving for V_1 yields:

$$V_1 = \frac{(1-\epsilon)(b-c) + \delta \cdot (2\epsilon(1-\epsilon)V_2 + \epsilon^2 V_3)}{1-\delta(1-\epsilon)^2}$$

We plug that back into the inequality above.

$$(1-\epsilon)\left(\frac{(1-\epsilon)(b-c) + \delta \cdot (2\epsilon(1-\epsilon)V_2 + \epsilon^2 V_3)}{1-\delta(1-\epsilon)^2}\right) - (1-2\epsilon)V_2 - \epsilon V_3 \geq \frac{c}{\delta}$$

Then we separate the factors involving V_2 and V_3.

$$\begin{aligned}&(1-\epsilon)\left(\frac{(1-\epsilon)(b-c)}{1-\delta(1-\epsilon)^2}\right) + (1-\epsilon)\left(\frac{(\delta \cdot (2\epsilon(1-\epsilon))}{1-\delta(1-\epsilon)^2}V_2\right) + (1-\epsilon)\left(\frac{\delta \cdot \epsilon^2}{1-\delta(1-\epsilon)^2}\right)V_3 \\ &- (1-2\epsilon)V_2 - \epsilon V_3 \geq \frac{c}{\delta}\end{aligned}$$

We multiply both sides with $1-\delta(1-\epsilon)^2$, and then we expand and simplify the factors that multiply V_2 and V_3. This takes a few steps.

$$\begin{aligned}&(1-\epsilon)(1-\epsilon)(b-c) \\ &+ \big(((1-\epsilon) \cdot \delta \cdot (2\epsilon(1-\epsilon)) - (1-2\epsilon) \cdot (1-\delta(1-\epsilon)^2)\big)\, V_2 \\ &+ \big((1-\epsilon) \cdot \delta \cdot \epsilon^2 \; - (1-\delta(1-\epsilon)^2)\epsilon\big)\, V_3 \\ &\geq \frac{c}{\delta}\end{aligned}$$

$$
\begin{aligned}
&(1-\epsilon)(1-\epsilon)(b-c) \\
&+\Big(((\delta-\delta\epsilon)\cdot\Big(2\epsilon-2\epsilon^2\Big)\Big)-(1-\delta(1-\epsilon)^2)+2\epsilon(1-\delta(1-\epsilon)^2)\Big)\,V_2 \\
&+\Big(\delta\epsilon^2-\delta\epsilon^3-\epsilon+\epsilon\delta(1-2\epsilon+\epsilon^2)\Big)\,V_3 \\
&\geq \frac{c}{\delta}
\end{aligned}
$$

$$
\begin{aligned}
&(1-\epsilon)(1-\epsilon)(b-c) \\
&+\Bigg(2\delta\epsilon-4\delta\epsilon^2+2\delta\epsilon^3 \\
&-(1-\delta(1-2\epsilon+\epsilon^2))+2\epsilon(1-\delta(1-2\epsilon+\epsilon^2))\Bigg)V_2 \\
&+\Big(\delta\epsilon^2-\delta\epsilon^3-\epsilon+\epsilon\delta-2\delta\epsilon^2+\delta\epsilon^3)\Big)\,V_3 \\
&\geq \frac{c}{\delta}
\end{aligned}
$$

$$
\begin{aligned}
&(1-\epsilon)(1-\epsilon)(b-c) \\
&+\Bigg(2\delta\epsilon-4\delta\epsilon^2+2\delta\epsilon^3 \\
&-1+\delta-2\delta\epsilon+\delta\epsilon^2+2\epsilon-2\epsilon(\delta-2\delta\epsilon+\delta\epsilon^2))\Bigg)V_2 \\
&+\Big(-\epsilon+\epsilon\delta-\delta\epsilon^2)\Big)\,V_3 \\
&\geq \frac{c}{\delta}
\end{aligned}
$$

$$
\begin{aligned}
&(1-\epsilon)(1-\epsilon)(b-c)\\
&+\Bigg(2\delta\epsilon-4\delta\epsilon^2+2\delta\epsilon^3\\
&-1+\delta-2\delta\epsilon+\delta\epsilon^2+2\epsilon-2\epsilon\delta+4\delta\epsilon^2-2\delta\epsilon^3))\Bigg)V_2\\
&-\big(1-\epsilon\delta+\delta\epsilon^2)\big)\,V_3\\
&\geq\frac{c}{\delta}
\end{aligned}
$$

$$
\begin{aligned}
&(1-\epsilon)(1-\epsilon)(b-c)\\
&+\Bigg(-1+\delta-2\delta\epsilon+\delta\epsilon^2+2\epsilon)\Bigg)V_2\\
&-\big(1-\epsilon\delta+\delta\epsilon^2)\big)\,V_3\\
&\geq\frac{c}{\delta}
\end{aligned}
$$

$$
\begin{aligned}
&(1-\epsilon)(1-\epsilon)(b-c)\\
&+\Bigg(-1+\delta(1-\epsilon)^2+2\epsilon)\Bigg)V_2\\
&-\big(1-\epsilon\delta+\delta\epsilon^2)\big)\,V_3\\
&\geq\frac{c}{\delta}
\end{aligned}
$$

$$
\begin{aligned}
&(1-\epsilon)(1-\epsilon)(b-c) \\
&-\left(1-\delta(1-\epsilon)^2-2\epsilon\right)V_2 \\
&-\left(1-\epsilon\delta+\delta\epsilon^2\right)V_3 \\
&\geq \frac{c}{\delta}
\end{aligned}
$$

The term $-(1-\epsilon\delta+\delta\epsilon^2))$ is always negative on $(\delta,\epsilon)\in(0,1)\times(0,\frac{1}{2})$. Our inequality is therefore most broadly satisfied if state 3 has the smallest possible value that can be stabilised in equilibrium – the output in this state has to be *defect*, and there can be no return to a cooperative state.

The term $-(1-\delta(1-\epsilon)^2-2\epsilon))$ is negative if $\delta<\frac{1-2\epsilon}{(1-\epsilon)^2}$. This expression corresponds to the dashed line in Figure 7. Depending on the sign of this term, the inequality above is therefore most broadly satisfied, for δ below this threshold, by the smallest possible value in state 2, or for δ above this threshold by the largest possible value in state 2.

Note that in the former case (δ below the threshold), we have therefore arrived at Grim Trigger. Regarding the latter case, the largest possible payoff is attained if V_2 corresponds to indefinite cooperation. However, for cooperate states to be incentive compatible they must punish some deviations from cooperation. State 2 therefore needs to satisfy the same incentive constraint as state 1, so that the behaviour that makes these states most broadly incentive compatible is identical, meaning that $V_2=V_1$. Note that for this case, we have arrived at Minimal Grim Trigger.

We start by deriving the condition for Grim Trigger to be an equilibrium, and then derive the condition for Minimal Grim Trigger to be an equilibrium. For Grim Trigger, we, thus, have $V_2=V_3=\frac{\epsilon(b-c)}{1-\delta}$. We, first rearrange using that the values in state 2 and 3 are equal. Recall our expression looked as follows.

$$\begin{aligned}
&\left(2\delta\epsilon(1-\epsilon)^2(1-2\epsilon) - 4\delta\epsilon(1-\epsilon)(1-\epsilon)^2 + \delta(1-\epsilon)^2 + 4(1-\epsilon)\epsilon - 1\right) V_2 \\
&\epsilon(1-2\epsilon)\left(\delta \cdot \epsilon(1-\epsilon) - \left(1-\delta(1-\epsilon)^2\right)\right) V_3 \\
&\geq \frac{1-2\epsilon}{\delta} c \left(1-\delta(1-\epsilon)^2\right) - (1-\epsilon)^2(1-2\epsilon)(b-c)
\end{aligned}$$

Substituting in $V_2 = V_3 = V$ allows us to simplify as follows.

$$\begin{aligned}
(1-\epsilon)\left(\frac{(1-\epsilon)(b-c) + \delta \cdot (2\epsilon(1-\epsilon)V + \epsilon^2 V)}{1-\delta(1-\epsilon)^2}\right) - (1-\epsilon)V &\geq \frac{c}{\delta} \\
(1-\epsilon)\left(\frac{(1-\epsilon)(b-c) + \delta \cdot (1-(1-\epsilon)^2)V)}{1-\delta(1-\epsilon)^2}\right) - (1-\epsilon)V &\geq \frac{c}{\delta} \\
\left(\frac{(1-\epsilon)(b-c) + \delta \cdot (1-(1-\epsilon)^2)V)}{1-\delta(1-\epsilon)^2}\right) - V &\geq \frac{c}{\delta(1-\epsilon)}
\end{aligned}$$

Now we plug in $V = \frac{\epsilon(b-c)}{1-\delta}$ and rearrange until we arrive at the inequality in the theorem.

$$\begin{aligned}
\left(\frac{(1-\epsilon)(b-c) + \delta \cdot (1-(1-\epsilon)^2)V)}{1-\delta(1-\epsilon)^2}\right) - V &\geq \frac{c}{\delta(1-\epsilon)} \\
\left(\frac{(\delta \cdot (1-(1-\epsilon)^2)) - (1-\delta(1-\epsilon)^2))}{1-\delta(1-\epsilon)^2}\right) V &\geq \frac{c}{\delta(1-\epsilon)} - \frac{(1-\epsilon)(b-c)}{1-\delta(1-\epsilon)^2} \\
\left(\frac{(\delta \cdot (1-(1-\epsilon)^2)) - (1-\delta(1-\epsilon)^2))\, \epsilon(b-c)}{1-\delta}\right) &\geq \frac{c(1-\delta(1-\epsilon)^2)}{\delta(1-\epsilon)} - (1-\epsilon)(b-c)
\end{aligned}$$

$$\begin{aligned}
\left(\frac{(\delta-1)\,\epsilon(b-c)}{1-\delta}\right) &\geq \frac{c(1-\delta(1-\epsilon)^2)}{\delta(1-\epsilon)} - (1-\epsilon)(b-c) \\
-\epsilon(b-c) &\geq \frac{c(1-\delta(1-\epsilon)^2)}{\delta(1-\epsilon)} - (1-\epsilon)(b-c) \\
(1-\epsilon-\epsilon)(b-c) &\geq \frac{c(1-\delta(1-\epsilon)^2)}{\delta(1-\epsilon)} \\
(1-2\epsilon)(b-c) &\geq \frac{c(1-\delta(1-\epsilon)^2)}{\delta(1-\epsilon)} \\
(b-c) &\geq \frac{c(1-\delta(1-\epsilon)^2)}{\delta(1-\epsilon)(1-2\epsilon)} \\
b-c\left(1+\frac{(1-\delta(1-\epsilon)^2)}{\delta(1-\epsilon)(1-2\epsilon)}\right) \geq 0 & \\
\frac{b}{c} &\geq 1+\frac{(1-\delta(1-\epsilon)^2)}{\delta(1-\epsilon)(1-2\epsilon)} \\
\frac{b}{c} &\geq \frac{\delta(1-\epsilon)(1-2\epsilon)+(1-\delta(1-\epsilon)^2)}{\delta(1-\epsilon)(1-2\epsilon)} \\
\frac{b}{c} &\geq \frac{1-\delta(1-\epsilon)\epsilon}{\delta(1-\epsilon)(1-2\epsilon)}
\end{aligned}$$

We now move on to derive the condition for Minimal Grim Trigger to be a Nash equilibrium. As before, we solve for V_1, now using that $V_1 = V_2$.

$$\begin{aligned}
V_1 &= (1-\epsilon)(b-c) + \delta\cdot\left((1-\epsilon^2)V_1 + \epsilon^2\frac{\epsilon(b-c)}{1-\delta}\right) \\
V_1\left(1-\delta(1-\epsilon^2)\right) &= (1-\epsilon)(b-c) + \delta\cdot\left(\epsilon^2\frac{\epsilon(b-c)}{1-\delta}\right) \\
V_1 &= \frac{(1-\epsilon)(b-c) + \left(\frac{\delta\epsilon^3(b-c)}{1-\delta}\right)}{1-\delta(1-\epsilon^2)}
\end{aligned}$$

Defecting rather than cooperating in state 1 yields the following payoff.

$$
\begin{aligned}
V_1' &= (1-\epsilon)b - \epsilon \cdot c + \delta \cdot \left((1-\epsilon\cdot(1-\epsilon))V_1' + (1-\epsilon)\cdot\epsilon\frac{\epsilon(b-c)}{1-\delta} \right) \\
V_1'(1-\delta(1-(1-\epsilon)\epsilon)) &= (1-\epsilon)b - \epsilon \cdot c + \delta \cdot \left((1-\epsilon)\cdot\epsilon\frac{\epsilon(b-c)}{1-\delta} \right) \\
V_1' &= \frac{(1-\epsilon)b - \epsilon \cdot c + \delta \cdot \left((1-\epsilon)\cdot\epsilon\frac{\epsilon(b-c)}{1-\delta} \right)}{1-\delta(1-(1-\epsilon)\epsilon}
\end{aligned}
$$

Now we compute the difference between these values $V_1 - V_1'$. Minimal Grim Trigger constitutes an equilibrium if this difference is positive.

$$
\begin{aligned}
V_1 - V_1' =& \frac{(1-\epsilon)(b-c) + \left(\frac{\delta\epsilon^3(b-c)}{1-\delta}\right)}{1-\delta(1-\epsilon^2)} - \frac{(1-\epsilon)b - \epsilon\cdot c + \delta\cdot\left((1-\epsilon)\cdot\epsilon\frac{\epsilon(b-c)}{1-\delta}\right)}{(1-\delta)(1-\delta(1-(1-\epsilon)\epsilon))} \\
\\
=& \frac{(1-\epsilon)(b-c)(1-\delta) + \delta\epsilon^3(b-c)}{(1-\delta)(1-\delta(1-\epsilon^2))} \\
& - \frac{(1-\delta)(1-\epsilon)b - (1-\delta)\epsilon\cdot c + \delta(1-\epsilon)\cdot\epsilon^2(b-c)}{1-\delta(1-(1-\epsilon)\epsilon}
\\
\\
=& \frac{(b-c)(1-\delta-\epsilon+\delta\epsilon+\delta\epsilon^3)}{(1-\delta)(1-\delta(1-\epsilon^2))} \\
& - \frac{b(1-\delta-\epsilon+\delta\epsilon+\delta\epsilon^2-\delta\epsilon^3)}{(1-\delta)(1-\delta(1-(1-\epsilon)\epsilon))} \\
& + \frac{(\epsilon-\delta\epsilon+\delta\epsilon^2-\delta\epsilon^3)c}{(1-\delta)(1-\delta(1-(1-\epsilon)\epsilon))}
\end{aligned}
$$

$$
\begin{aligned}
=&\frac{(1-\delta(1-(1-\epsilon)\epsilon))(b-c)(1-\delta-\epsilon+\delta\epsilon+\delta\epsilon^3)}{(1-\delta)(1-\delta(1-\epsilon^2))(1-\delta(1-(1-\epsilon)\epsilon))}\\
&-\frac{(1-\delta(1-\epsilon^2))b(1-\delta-\epsilon+\delta\epsilon+\delta\epsilon^2-\delta\epsilon^3)}{(1-\delta)(1-\delta(1-(1-\epsilon)\epsilon))(1-\delta(1-\epsilon^2))}\\
&+\frac{(1-\delta(1-\epsilon^2))(\epsilon-\delta\epsilon+\delta\epsilon^2-\delta\epsilon^3)c}{(1-\delta)(1-\delta(1-(1-\epsilon)\epsilon))(1-\delta(1-\epsilon^2))}
\end{aligned}
$$

The denominator is positive, so it will not affect the sign of the difference. We therefore drop it from the further calculations.

$$
\begin{aligned}
&(1-\delta(1-(1-\epsilon)\epsilon))(b-c)(1-\delta-\epsilon+\delta\epsilon+\delta\epsilon^3)\\
&-(1-\delta(1-\epsilon^2))b(1-\delta-\epsilon+\delta\epsilon+\delta\epsilon^2-\delta\epsilon^3)\\
&+(1-\delta(1-\epsilon^2))(\epsilon-\delta\epsilon+\delta\epsilon^2-\delta\epsilon^3)c
\end{aligned}
$$

$$
\begin{aligned}
=&(1-\delta+\delta\epsilon-\delta\epsilon^2)(b-c)(1-\delta-\epsilon+\delta\epsilon+\delta\epsilon^3)\\
&-(1-\delta+\delta\epsilon^2)b(1-\delta-\epsilon+\delta\epsilon+\delta\epsilon^2-\delta\epsilon^3)\\
&+(1-\delta+\delta\epsilon^2)(\epsilon-\delta\epsilon+\delta\epsilon^2-\delta\epsilon^3)c
\end{aligned}
$$

$$
\begin{aligned}
=&b\cdot\big((1-\delta+\delta\epsilon-\delta\epsilon^2)\cdot(1-\delta-\epsilon+\delta\epsilon+\delta\epsilon^3)-(1-\delta+\delta\epsilon^2)(1-\delta-\epsilon+\delta\epsilon+\delta\epsilon^2-\delta\epsilon^3)\big)\\
-&c((1-\delta+\delta\epsilon-\delta\epsilon^2)(1-\delta-\epsilon+\delta\epsilon+\delta\epsilon^3)-(1-\delta+\delta\epsilon^2)(\epsilon-\delta\epsilon+\delta\epsilon^2-\delta\epsilon^3))
\end{aligned}
$$

$$
\begin{aligned}
=&b \cdot \Bigg(1 - 2\delta + \delta^2 - \epsilon + 3\delta\epsilon - 2\delta^2\epsilon - 2\delta\epsilon^2 + 2\delta^2\epsilon^2 + 2\delta\epsilon^3 - 2\delta^2\epsilon^3 + \delta^2\epsilon^4 - \delta^2\epsilon^5 \\
&- (1 - 2\delta + \delta^2 - \epsilon + 2\delta\epsilon - \delta^2\epsilon + 2\delta\epsilon^2 - 2\delta^2\epsilon^2 - 2\delta\epsilon^3 + 2\delta^2\epsilon^3 + \delta^2\epsilon^4 - \delta^2\epsilon^5)\Bigg) \\
&- c\Bigg(1 - 2\delta + \delta^2 - \epsilon + 3\delta\epsilon - 2\delta^2\epsilon - 2\delta\epsilon^2 + 2\delta^2\epsilon^2 + 2\delta\epsilon^3 - 2\delta^2\epsilon^3 + \delta^2\epsilon^4 - \delta^2\epsilon^5 \\
&- (\epsilon - 2\delta\epsilon + \delta^2\epsilon + \delta\epsilon^2 - \delta^2\epsilon^2 + \delta^2\epsilon^4 - \delta^2\epsilon^5)\Bigg)
\end{aligned}
$$

$$
\begin{aligned}
=&b(\delta\epsilon - \delta^2\epsilon - 4\delta\epsilon^2 + 4\delta^2\epsilon^2 + 4\delta\epsilon 3 - 4\delta^2\epsilon^3) \\
&- c(1 - 2\delta + \delta^2 - 2\epsilon + 5\epsilon\delta - 3\delta^2\epsilon - 3\delta\epsilon^2 + 3\delta^2\epsilon^2 + 2\delta\epsilon^3 - 2\delta^2\epsilon^3)
\end{aligned}
$$

$$
\begin{aligned}
=&b(1 - \delta)\delta\epsilon(1 - 2\epsilon)^2 \\
&- c((1 - \delta)(1 - 2\epsilon)(1 - \delta + \delta\epsilon - \delta\epsilon^2))
\end{aligned}
$$

Both b and c are multiplied by factors that are positive in the considered parameter regions. The whole expression is therefore positive if the following holds.

$$
\begin{aligned}
\frac{b}{c} &\geq \frac{(1 - \delta)(1 - 2\epsilon)(1 - \delta + \delta\epsilon - \delta\epsilon^2)}{(1 - \delta)\delta\epsilon(1 - 2\epsilon)^2} \\
\frac{b}{c} &\geq \frac{(1 - \delta + \delta\epsilon - \delta\epsilon^2)}{\delta\epsilon(1 - 2\epsilon)} \\
\frac{b}{c} &\geq \frac{1 - \delta(1 - \epsilon + \epsilon^2)}{\delta\epsilon(1 - 2\epsilon)}
\end{aligned}
$$

$\square$

B Proof Theorem 3

Proof. Without loss of generality, we are considering differentiability with respect to ϵ at $\epsilon = 0$, which is indicated by the subscript 0.

If a stationary distribution σ exists, then all $\sigma(i)$ must satisfy $\sigma(i) = \frac{p_{j,i}}{p_{j,i}+P(i,i)}$, where generally, $P(i,j)$ is the per-iteration probability of moving from state i to state j, so that $P(i,i)$ is the per-iteration probability of remaining in state i, and $p_{j,i}$ is the aggregate per-iteration influx into state i, $p_{j,i} = \sum_{k \neq i} \sigma(k)P(k,i)$. (That follows from the fact that for a two-state process we have

$$\begin{aligned}
\sigma(i) &= \sigma(i)P(i,i) + \sigma(j)P(j,i) \\
&= \sigma(i)(1 - P(i,j)) + (1 - \sigma(i))P(j,i) \\
\implies \sigma(i) &= \frac{P(j,i)}{P(i,j) + P(j,i)}
\end{aligned}$$

and here we simply aggregate all states $\neq i$ into one state "j".)

Now, we proceed to show that the absolute value of the total differential

$$\frac{d\sigma_0(i)}{d\epsilon} = \frac{d\sigma_0(i)}{dP(i,j)}\frac{dP(i,j)}{d\epsilon} + \frac{d\sigma_0(i)}{dp_{j,i}}\frac{dp_{j,i}}{d\epsilon}$$

is bounded ($\max_i |\frac{d\sigma_0(i)}{d\epsilon}| < \infty$). This will imply that for any δ, we can find an ϵ such that $\max_i |\sigma_0(i) - s_\epsilon| < \delta$.

First, note that the absolute value of the first factors, $\frac{d\sigma_0(i)}{dP(i,j)}$ and $\frac{d\sigma_0(i)}{dp_{j,i}}$, is bounded by $\frac{1}{4m}$. For any given $P(i,j)$, the absolute value of the partial $\frac{d\sigma_0(i)}{dP(i,j)}$ is maximised by choosing $p_{j,i} = P(i,j)$.

$$|\frac{d\sigma(i)}{dP(i,j)}| = \frac{p_{j,i}}{(P(i,j)+p_{j,i})^2}$$
$$|\frac{d}{dp_{j,i}}\frac{d\sigma(i)}{dP(i,j)}| = |\frac{P(i,j)-p_{j,i}}{(P(i,j)+p_{j,i})^3}|$$

The resulting term $|\frac{d\sigma(i)}{dP(i,j)}| = \frac{P(i,j)}{(2P(i,j))^2}$ is maximal on $P(i,j) \in [m, 1]$ at $P(i,j) = m$. Hence $|\frac{ds(i)}{dP(i,j)}| \leq \frac{m}{(2m)^2} = \frac{1}{4m}$. (The reasoning for $|\frac{ds(i)}{dp_{j,i}}|$ is analogous.)

Furthermore, we have $|\frac{dP(i,j)}{d\epsilon}| \leq M$, so it follows that $|\frac{d\sigma_0(i)}{dP(i,j)}\frac{dP(i,j)}{d\epsilon}| \leq \frac{M}{4m}$.

Lastly, for $\frac{dp_{j,i}}{d\epsilon}$, we can use that $p_{j,i}$ can be approximated arbitrarily well by considering only finitely long sequences of states that the process could visit between leaving state i and returning to state i. Specifically, let $Pr_\epsilon(q)$ be the probability that the process, when leaving state i traverses a given sequence q of states $\neq i$ before returning to i. Then it holds that

$$p_{j,i} = \sum_{q \in Sequences} Pr_\epsilon(q)\frac{1}{length(q)}$$

Now if we choose some maximum considered sequence length L, we get an approximation $\tilde{p_{j,i}} = \sum_{length(q) \leq L} Pr_\epsilon(q)\frac{1}{length(q)}$. For this approximation it holds that

$$p_{j,i} - \tilde{p_{j,i}} \leq \frac{1}{L+1}$$

Let's choose L to be a function of ϵ, such as $L(\epsilon) = \lceil \frac{1}{\epsilon^2} \rceil$. Then,

$$\frac{dp_{j,i}}{d\epsilon} = \lim_{\epsilon \searrow 0} \frac{1}{\epsilon}\left(\frac{1}{L(\epsilon)+1} + \sum_{length(q) \leq L(\epsilon)} (Pr_\epsilon(q) - Pr_0(q))\frac{1}{length(q)}\right)$$

where Pr_ϵ and Pr_0 are the likelihoods of traversing q at 0 and at ϵ.

Now, $\sum_{length(q) \leq L(\epsilon)}(Pr_\epsilon(q) - Pr_0(q))\frac{1}{length(q)}$ is maximal if all changes in probability act on sequences of length one. This is because $Pr(q) = P(i, q_1) \cdot P(q_1, q_2) \cdots P(q_L, i)$, and the change in $Pr(q)$ is maximal if all $P(\cdot, \cdot)$ increase(decrease). Now the change in $P(\cdot, \cdot)$ is bounded by ϵM.

Further, $p^L - (p - \epsilon \cdot M)^L$ is maximised by $p = 1$ on $p \in [0, 1]$. Lastly, $\lim_{\epsilon \to 0} \frac{(1-(1-\epsilon)^L)}{L}$ is decreasing in L. In words, (A) the probability-change that a sequence of length L can undergo, if each individual transition rate can change only by $\epsilon \cdot M$, is maximal if the initial probability was 1, and (B) given that $(Pr(q) = 1)$, the probability change is maximised by choosing small L. Therefore, we are only considering sequences of one state, and the probability of such a sequence $Pr(q)$ can change by at most ϵM.

Hence, we arrive at the following approximation for $|\frac{d\sigma_0(i)}{dP(j,i)}||\frac{dP(j,i)}{d\epsilon}|$

$$\begin{aligned} |\frac{d\sigma_0(i)}{dP(j,i)}||\frac{dP(j,i)}{d\epsilon}| &\leq \frac{P(i,j)}{(P(i,j) + P(j,i))^2} \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left(\frac{1}{L(\epsilon) + 1} + \epsilon M \right) \\ &\leq \frac{1}{4m} M \end{aligned}$$

and overall, we arrive at

$$\frac{d\sigma(i)}{d\epsilon} \leq \frac{M}{2m}$$

□

C Proof Theorem 4

Proof. Nowak et al. (2004) show that in the limit of weak selection, a single mutant (in our case J for *jumpstarter*) has probability greater than $\frac{1}{N}$ of invading a population (where we now give the resident the subscript D for All D), if $(2N–1)\pi_{J,D} + (N–2)\pi_{J,J} > (2N–4)\pi_{D,D} + (N + 1)\pi_{D,J}$. We can use this limit result, because at an error rate of zero, jumpstarters are neutral mutants of All D, and because we are considering infinitesimally small ϵ. Since payoffs are differentiable functions of ϵ, this implies that payoff differences are also infinitesimally small.

We proceed by showing that the derivatives of the respective payoffs π w.r.t. ϵ at $\epsilon = 0$ behave in a way that ensures that the above inequality is satisfied in a neighbourhood of $\epsilon = 0$, which proves the claim.

Let's denote expected stage-game payoffs as functions of intended (error-free) actions by v_{dd}, v_{cd}, v_{dc}, and v_{cc}, and let's use the two-parameter payoff-matrix with non-negative payoffs introduced in section 2.2. Explicitly, the expected stage-game payoffs look as follows.

$$v_{dd} = c \cdot (1-\epsilon)^2 + (1-\epsilon) \cdot \epsilon \cdot (b+c) + \epsilon^2 \cdot b$$

$$v_{cc} = b \cdot (1-\epsilon)^2 + (1-\epsilon) \cdot \epsilon \cdot (b+c) + \epsilon^2 \cdot c$$

$$v_{cd} = (1-\epsilon) \cdot \epsilon \cdot (b+c) + \epsilon^2 \cdot (b+c)$$

$$v_{dc} = (b+c) \cdot (1-\epsilon)^2 + (1-\epsilon) \cdot \epsilon \cdot (b+c)$$

Moving on to payoffs, All D, when playing against copies of itself always earns

$$\pi_{D,D} = \frac{v_{dd}}{1-\delta}$$

.

Jumpstarters' payoffs against copies of themselves follow the following recursion

$$V_{J,J}(1) = v_{dd} + \delta \cdot \big((1-\epsilon)^2 \cdot V_{J,J}(1) + (1-(1-\epsilon)^2) \cdot V_{J,J}(2))\big)$$

$$V_{J,J}(2) = v_{cc} + \delta \cdot \big((1-\epsilon)^2 \cdot V_{J,J}(2) + (1-(1-\epsilon)^2) \cdot V_{J,J}(1)\big)$$

where $V_{J,J}(1)$ and $V_{J,J}(2)$ denote the values of J, playing against J being in state 1 and 2 respectively, and $\pi_{J,J} = V_{J,J}(1)$. Solving this for $V_{J,J}(1)$ and plugging in yields the following expression for $\pi_{J,J}$.

$$\pi_{J,J} = \frac{-2\delta\epsilon v_{cc} + \delta\epsilon^2 v_{cc} - v_{dd} + \delta v_{dd} - 2\delta\epsilon v_{dd} + \delta\epsilon^2 v_{dd}}{(1-\delta)(-1+\delta-4\delta\epsilon+2\delta\epsilon^2)}$$

Similarly, we can describe jumpstarters' payoffs against all D by this infinite recursion,

$$V_{J,D}(1) = v_{dd} + \delta \cdot \big((1-\epsilon)^2 \cdot V_{J,D}(1) + (1-(1-\epsilon)^2) \cdot V_J(2)\big)$$

$$V_{J,D}(2) = v_{cd} + \delta \cdot ((1-\epsilon)\epsilon \cdot V_{J,D}(2) + (1-(1-\epsilon)\epsilon) \cdot V_{J,D}(1))$$

which yields the following expression for $\pi_{J,D} = V_{J,D}(1)$

$$\pi_{J,D} = \frac{2\delta\epsilon v_{cd} - \delta\epsilon^2 v_{cd} + v_{dd} - \delta\epsilon v_{dd} + \delta\epsilon^2 v_{dd}}{(1-\delta)(1+\delta\epsilon)}$$

and, by symmetry, this expression for All D's payoffs.

$$\pi_{D,J} = \frac{2\delta\epsilon v_{dc} - \delta\epsilon^2 v_{dc} + v_{dd} - \delta\epsilon v_{dd} + \delta\epsilon^2 v_{dd}}{(1-\delta)(1+\delta\epsilon)}$$

As we are looking at the limit of $\epsilon \to 0$ we can ignore terms higher than first order in ϵ. Moreover, we leave out the common factor of $\frac{1}{1-\delta}$ to simplify payoff terms.

$$\pi_{D,D} = v_{dd}$$

$$\pi_{J,J} = \frac{-2\delta\epsilon v_{cc} - v_{dd} + \delta v_{dd} - 2\delta\epsilon v_{dd}}{(-1+\delta-4\delta\epsilon)}$$

$$\pi_{J,D} = \frac{2\delta\epsilon v_{cd} + v_{dd} - \delta\epsilon v_{dd}}{(1+\delta\epsilon)}$$

$$\pi_{D,J} = \frac{2\delta\epsilon v_{dc} + v_{dd} - \delta\epsilon v_{dd}}{(1+\delta\epsilon)}$$

Now we plug in the expressions for v_{dd}, v_{cc}, v_{cd}, and v_{dc} into the expressions for $\pi_{D,D}$, $\pi_{J,J}$, $\pi_{J,D}$, and $\pi_{D,J}$.

$$\begin{aligned}
\pi_{D,D} =& c(1-\epsilon)^2 + (b+c)(1-\epsilon)\epsilon \\
\pi_{J,J} =& \frac{1}{-1+\delta-4\delta\epsilon} \cdot \Big(-(c(1-\epsilon)^2) - (b+c)(1-\epsilon)\epsilon - 2\delta\epsilon(b(1-\epsilon)^2 \\
& + (b+c)(1-\epsilon)\epsilon) + \delta(c(1-\epsilon)^2 + (b+c)(1-\epsilon)\epsilon) \\
& - 2\delta\epsilon(c(1-\epsilon)^2 + (b+c)(1-\epsilon)\epsilon)\Big) \\
\pi_{J,D} =& \frac{1}{1+\delta\epsilon} \cdot \Big(c(1-\epsilon)^2 + (b+c)(1-\epsilon)\epsilon + 2(b+c)\delta(1-\epsilon)\epsilon^2 \\
& - \delta\epsilon(c(1-\epsilon)^2 + (b+c)(1-\epsilon)\epsilon)\Big) \\
\pi_{D,J} =& \frac{1}{1+\delta\epsilon} \cdot \Big(c(1-\epsilon)^2 + (b+c)(1-\epsilon)\epsilon - \delta\epsilon(c(1-\epsilon)^2 + (b+c)(1-\epsilon)\epsilon) \\
& + 2\delta\epsilon((b+c)(1-\epsilon)^2 + (b+c)(1-\epsilon)\epsilon)\Big)
\end{aligned}$$

Taking the derivative of each expression with respect to ϵ, we get the following expressions.

$$\frac{d\pi_{D,D}}{d\epsilon} = -c - b(-1+\epsilon) - b\epsilon$$

$$\begin{aligned}
\frac{d\pi_{J,J}}{d\epsilon} =& -\frac{(-1+\epsilon)(2b\delta\epsilon + c\delta(2+4\epsilon) + b(1+\delta+2\delta\epsilon))}{1+\delta(-1+4\epsilon)} \\
& + \frac{4\delta(-1+\epsilon)(c + b\epsilon(1+\delta+2\delta\epsilon) + c\delta(-1+2\epsilon+2\epsilon^2))}{(1+\delta(-1+4\epsilon))^2} \\
& - \frac{c + b\epsilon(1+\delta+2\delta\epsilon) + c\delta(-1+2\epsilon+2\epsilon^2)}{1+\delta(-1+4\epsilon)}
\end{aligned}$$

$$\begin{aligned}\frac{d\pi_{J,D}}{d\epsilon} &= -\frac{(-1+\epsilon)(b\delta\epsilon+2c\delta\epsilon+c\delta(-1+2\epsilon)+b(1+\delta\epsilon))}{1+\delta\epsilon}\\ &+\frac{\delta(-1+\epsilon)(c+c\delta\epsilon(-1+2\epsilon)+b\epsilon(1+\delta\epsilon))}{(1+\delta\epsilon)^2}\\ &-\frac{c+c\delta\epsilon(-1+2\epsilon)+b\epsilon(1+\delta\epsilon)}{1+\delta\epsilon}\end{aligned}$$

$$\begin{aligned}\frac{d\pi_{D,J}}{d\epsilon} &= -\frac{(-1+\epsilon)(c\delta+b(1-\delta(-2+\epsilon))-b\delta\epsilon)}{1+\delta\epsilon}\\ &+\frac{\delta(-1+\epsilon)(c+c\delta\epsilon+b(1-\delta(-2+\epsilon))\epsilon)}{(1+\delta\epsilon)^2}\\ &-\frac{c+c\delta\epsilon+b(1-\delta(-2+\epsilon))\epsilon}{1+\delta\epsilon}\end{aligned}$$

We evaluate these expressions at $\epsilon = 0$.

$$\left.\frac{d\pi_{D,D}}{d\epsilon}\right|_{\epsilon=0} = b - c$$

$$\left.\frac{d\pi_{J,J}}{d\epsilon}\right|_{\epsilon=0} = -\frac{c-c\delta}{1-\delta} - \frac{4\delta(c-c\delta)}{(1-\delta)^2} + \frac{2c\delta+b(1+\delta)}{1-\delta}$$

$$\left.\frac{d\pi_{J,D}}{d\epsilon}\right|_{\epsilon=0} = b - c - 2c\delta$$

$$\left.\frac{d\pi_{D,J}}{d\epsilon}\right|_{\epsilon=0} = -c + b(1+2\delta)$$

At $\epsilon = 0$ both sides of this inequality $(2N-1)\pi_{J,D}+(N-2)\pi_{J,J} > (2N-4)\pi_{D,D}+(N+1)\pi_{D,J}$ are equal. Hence, if we substitute these derivatives into the inequality from above, we arrive at a condition for the inequality to be satisfied in a neighbourhood of $\epsilon = 0$.

$$(-\frac{c-c\delta}{1-\delta}-\frac{4\delta(c-c\delta)}{(1-\delta)^2}+\frac{2c\delta+b(1+\delta)}{1-\delta})(-2+N)+(b-c-2c\delta)(-1+2N)$$
$$>(-c+b(1+2\delta))(1+N)+(b-c)(-4+2N)$$

Simplifying a bit – under the assumption that all variables are positive – we get the following.

$$(1-\delta)\delta(b(-3+\delta+\delta N)+c(3-3N+\delta(-1+2N)))>0$$

Finally, we use that $0<\delta<1$ and reformulate, to get the result in the theorem.

$$\begin{aligned}
&(b(-3+\delta+\delta N)+c(3-3N+\delta(-1+2N)))>0\\
&\quad -3b+\delta b+\delta Nb+c-3Nc+\delta(2N-1)c>0\\
&\qquad\qquad\implies \delta b+\delta Nb+\delta(2N-1)c>3b+3Nc-c\\
&\qquad\qquad\implies \delta(b+Nb+(2N-1)c)>3b+3Nc-c\\
&\qquad\qquad\qquad\qquad\implies \delta>\frac{3b+3Nc-c}{(b+Nb+(2N-1)c)}
\end{aligned}$$

□

D Proof Theorem 4

Here we assume that the jumpstater goes through M consecutive states as sketched in Figure 13. Then the Mth state is like the first state of the Figure 9 jumpstarter, and the $M+1$th state is like the second state in the Figure 9 jumpstarter.

Intuitively, what this achieves is that the jumpstarting rate can become arbitrarily low, even if ϵ is at a fixed non-zero level. For various steps in this proof we used the software Mathematica, Version 12.

Proof. As in the previous proof, we verify that the jumpstarter satisfies the inequality $(2N\text{–}1)\pi_{J,D} + (N\text{–}2)\pi_{J,J} > (2N\text{–}4)\pi_{D,D} + (N+1)\pi_{D,J}$. The subscript J now refers to the jumpstarter in Figure 13, and the subscript D still refers to All D. The payoff equations are a bit different. To simplify notation, let's just keep the terms v_{cc} etc. (everything is linear in payoffs, so it doesn't matter if the payoffs of unintended actions co-occur with the state-switch).

$$\pi_{DD} = \frac{v_{dd}}{1-\delta}$$

Let's call the number of non-jumpstarting states M. Then, all payoffs follow similar infinite recursions as before. Namely, to solve for $\pi_{J,J}$ we have this recursion.

$$V_{J,J}(1) = \frac{v_{dd}(1-\delta^{(M+1)})}{1-\delta} + \delta^{(M+1)} \cdot \big((1-\epsilon)^2 \cdot V_{J,J}(1) + (1-(1-\epsilon)^2) \cdot V_{J,J}(2)\big)$$

$$V_{J,J}(2) = v_{cc} + \delta \cdot \big((1-\epsilon)^2 \cdot V_{J,J}(2) + (1-(1-\epsilon)^2) \cdot V_{J,J}(1)\big)$$

and to solve for $\pi_{J,D}$ we have this recursion.

$$V_{J,D}(1) = \frac{v_{dd}(1-\delta^{M+1})}{1-\delta} + \delta^{(M+1)} \cdot \big((1-\epsilon)^2 \cdot V_{J,D}(1) + (1-(1-\epsilon)^2) \cdot V_{J,D}(2)\big)$$

$$V_{J,D}(2) = v_{cd} + \delta \cdot ((1-\epsilon)\epsilon \cdot V_{J,D}(2) + (1-(1-\epsilon)\epsilon) \cdot V_{J,D}(1))$$

We solve these recursions. For the jumpstarter-jumpstarter interaction we get

$$
\begin{aligned}
V_{J,J}(1) =& -\left(\frac{-\delta^{1+M}(1-(1-\epsilon)^2)v_{cc} - \big((1-\delta^{1+M})(1-\delta(1-\epsilon)^2)v_{dd}\big)}{-\delta^{2+M}(1-(1-\epsilon)^2)^2 + (1-\delta(1-\epsilon)^2)\left(1-\delta^{1+M}(1-\epsilon)^2\right)}\right) \\
V_{J,J}(2) =& \frac{1}{(1-\delta-\delta^{1+M}+\delta^{2+M}+2\delta\epsilon+2\delta^{1+M}\epsilon-4\delta^{2+M}\epsilon-\delta\epsilon^2-\delta^{1+M}\epsilon^2+2\delta^{2+M}\epsilon^2)} \\
& \cdot \frac{1}{(1-\delta)} \cdot \Bigg(v_{cc} - \delta v_{cc} - \delta^{1+M}v_{cc} + \delta^{2+M}v_{cc} + 2\delta^{1+M}\epsilon v_{cc} - 2\delta^{2+M}\epsilon v_{cc} \\
& - \delta^{1+M}\epsilon^2 v_{cc} + \delta^{2+M}\epsilon^2 v_{cc} + 2\delta\epsilon v_{dd} - 2\delta^{2+M}\epsilon v_{dd} - \delta\epsilon^2 v_{dd} + \delta^{2+M}\epsilon^2 v_{dd}\Bigg)
\end{aligned}
$$

and for the Jumpstarter versus All D interaction we get

$$
\begin{aligned}
V_{J,D}(1) =& -\left(\frac{-\delta^{1+M}(1-(1-\epsilon)^2)v_{cd} - \big((1-\delta^{1+M})(1-\delta(1-\epsilon)\epsilon)v_{dd}\big)}{(1-\delta)\left(1-\delta^{1+M}-\delta\epsilon+2\delta^{1+M}\epsilon-\delta^{2+M}\epsilon+\delta\epsilon^2-\delta^{1+M}\epsilon^2\right)}\right) \\
V_{J,D}(2) =& \frac{1}{(-1+\delta)\left(-1+\delta^{1+M}+\delta\epsilon-2\delta^{1+M}\epsilon+\delta^{2+M}\epsilon-\delta\epsilon^2+\delta^{1+M}\epsilon^2\right)} \\
& \cdot \Bigg(v_{cd} - \delta v_{cd} - \delta^{1+M}v_{cd} + \delta^{2+M}v_{cd} + 2\delta^{1+M}\epsilon v_{cd} - 2\delta^{2+M}\epsilon v_{cd} \\
& - \delta^{1+M}\epsilon^2 v_{cd} + \delta^{2+M}\epsilon^2 v_{cd} + \delta v_{dd} - \delta^{2+M}v_{dd} - \delta\epsilon v_{dd} + \delta^{2+M}\epsilon v_{dd} \\
& + \delta\epsilon^2 v_{dd} - \delta^{2+M}\epsilon^2 v_{dd}\Bigg)
\end{aligned}
$$

Again, by expression we get the same expression for $V_{D,J}(1)$, with v_{dc} rather than v_{cd}. Moreover, the expression for $\pi_{D,D}$ is as in the previous proof. We now plug these expressions into our inequality.

$$
\begin{aligned}
&\frac{-1}{-\delta^{2+M}(1-(1-\epsilon)^2)^2+(1-\delta(1-\epsilon)^2)(1-\delta^{1+M}(1-\epsilon)^2)}\cdot\Bigg(-\left(\frac{1}{1-\delta}\right)\\
&\cdot(1-\delta^{1+M})\cdot(1-\delta(1-\epsilon)^2)(c(1-\epsilon)^2+(b+c)(1-\epsilon)\epsilon+b\epsilon^2)\\
&-\delta^{1+M}(1-(1-\epsilon)^2)(b(1-\epsilon)^2+(b+c)(1-\epsilon)\epsilon+c\epsilon^2))(-2+N)\Bigg)\\
&+\frac{1}{1-\delta^{1+M}-\delta\epsilon+2\delta^{1+M}\epsilon-\delta^{2+M}\epsilon+\delta\epsilon^2-\delta^{1+M}\epsilon^2}\\
&\cdot\Bigg(-\delta^{1+M}(1-(1-\epsilon)^2)((b+c)(1-\epsilon)^2+(b+c)(1-\epsilon)\epsilon)\\
&-\left(\frac{1}{1-\delta}\right)(1-\delta^{1+M})(1-\delta(1-\epsilon)\epsilon)(c(1-\epsilon)^2\\
&+(b+c)(1-\epsilon)\epsilon+b\epsilon^2))(1+N)\Bigg)\\
&-\frac{(c(1-\epsilon)^2+(b+c)(1-\epsilon)\epsilon+b\epsilon^2)(-4+2N)}{1-\delta}\\
&-\frac{1}{1-\delta^{1+M}-\delta\epsilon+2\delta^{1+M}\epsilon-\delta^{2+M}\epsilon+\delta\epsilon^2-\delta^{1+M}\epsilon^2}\cdot\Bigg(-\left(\frac{1}{1-\delta}\right)\\
&\cdot(1-\delta^{1+M})(1-\delta(1-\epsilon)\epsilon)(c(1-\epsilon)^2+(b+c)(1-\epsilon)\epsilon+b\epsilon^2)\\
&-\delta^{1+M}(1-(1-\epsilon)^2)((b+c)(1-\epsilon)\epsilon+(b+c)\epsilon^2))(-1+2N)\Bigg)\\
&>0
\end{aligned}
$$

Of course, in the limit of M going to infinity, the all payoff differences between J and D vanish. We are interested in the limit of weak selection, and payoff differences are of order δ^M. Hence, we multiply the above inequality with δ^{-M}, and take $M \to \infty$. The first step, multiplying with δ^{-M}, and simplifying (done in Mathematica) yields.

$$\begin{aligned}
&\frac{1}{((\epsilon-1)^2\delta^{M+1}+\epsilon\delta^{M+2}-\delta(\epsilon-1)\epsilon-1)}\\
&\cdot\frac{1}{((2\epsilon^2-4\epsilon+1)\,\delta^{M+2}-(\epsilon-1)^2\delta^{M+1}-\delta(\epsilon-1)^2+1)}\\
&\cdot\Bigg(\delta\epsilon\left(2\epsilon^2-5\epsilon+2\right)b\Big(\delta^{M+2}\left(2\epsilon^2(N+1)-3\epsilon(N+2)+N+1\right)\\
&-3(\epsilon-1)^2\delta^{M+1}+\delta\left(\epsilon^2(1-2N)+3\epsilon N-N-1\right)+3\Big)\\
&+c\Big(\delta^{M+2}\left(\epsilon^2(4N-2)+\epsilon(6-9N)+2N-1\right)-3(\epsilon-1)^2(N-1)\delta^{M+1}\\
&-\delta(\epsilon-1)(\epsilon N+\epsilon-2N+1)+3(N-1)\Big)\Bigg)>0
\end{aligned}$$

Now, in the limit of $M \to \infty$, all the terms of order δ^M drop out, and we arrive at the following slightly shorter expression.

$$\begin{aligned}
&-\frac{\delta\epsilon\left(2\epsilon^2-5\epsilon+2\right)\left(b(\delta(\epsilon-1)(2\epsilon N-\epsilon-N-1)-3\right)}{\left(\delta(\epsilon-1)^2-1\right)\left(\delta(\epsilon-1)\epsilon+1\right)}\\
&-\frac{c(\delta(\epsilon-1)(\epsilon N+\epsilon-2N+1)-3N+3))}{\left(\delta(\epsilon-1)^2-1\right)\left(\delta(\epsilon-1)\epsilon+1\right)}\\
&>0
\end{aligned}$$

Lastly, we take the limit $N \to \infty$ since we are looking at large populations.

$$-\frac{\delta\epsilon\left(2\epsilon^2-5\epsilon+2\right)\left(b(\delta(\epsilon-1)(2\epsilon-1))+c(\delta(\epsilon-1)(\epsilon-2)-3)\right)}{\left(\delta(\epsilon-1)^2-1\right)\left(\delta(\epsilon-1)\epsilon+1\right)}>0$$

The denominator is negative, and $(2\epsilon^2-5\epsilon+2)$ is positive on $\epsilon\in[0,\frac{1}{2}]$ so everything simplifies further, and we can reformulate to arrive at the expression in our theorem.

$$\implies b(\delta(\epsilon-1)(2\epsilon-1)) + c(\delta(\epsilon-1)(\epsilon-2)-3) > 0$$
$$\implies b(\delta(\epsilon-1)(2\epsilon-1)) > c(3-\delta(\epsilon-1)(\epsilon-2))$$
$$\implies b(\delta(1-\epsilon)(1-2\epsilon)) > c(3-\delta(1-\epsilon)(2-\epsilon))$$
$$\implies \delta\left(b(1-\epsilon)(1-2\epsilon) + c(1-\epsilon)(2-\epsilon)\right) > 3c$$
$$\implies \delta > \frac{3c}{(b(1-\epsilon)(1-2\epsilon) + c(1-\epsilon)(2-\epsilon))}$$

□

E The basin of attraction of S_n vs All D is increasing in n

The payoff of S_n against itself follows this recursion.

$$V_{S_n,S_n}(1) = \pi_{cc} + \delta\cdot((1-\epsilon)\cdot V_{S_n,S_n}(1) + \epsilon\cdot V_{S_n,S_n}(2))$$
$$V_{S_n,S_n}(2) = \pi_{dd}\cdot\frac{(1-\delta^n)}{(1-\delta)} + \delta^n\cdot V_{S_n,S_n}(1)$$

Plugging in for $V_{S_n,S_n}(1)$ yields this expression.

$$V_{S_n,S_n}(1) = \pi_{cc} + \delta\cdot\left((1-\epsilon)\cdot V_{S_n,S_n}(1) + \epsilon\cdot\left(\pi_{dd}\cdot\frac{(1-\delta^n)}{(1-\delta)} + \delta^n\cdot V_{S_n,S_n}(1)\right)\right)$$
$$\implies V_{S_n,S_n}(1)(1-\delta(1-\epsilon)-\delta^{n+1}\epsilon) = \pi_{cc} + \delta\epsilon\pi_{dd}\frac{(1-\delta^n)}{(1-\delta)}$$
$$\implies V_{S_n,S_n}(1) = \frac{\pi_{cc} + \delta\epsilon\pi_{dd}\frac{(1-\delta^n)}{(1-\delta)}}{(1-\delta(1-\epsilon)-\delta^{n+1}\epsilon)}$$

For the interaction between S_n and All D (for brevity as above, ”D”), we have this recursion. The arguments of $V_{S_n,D}()$ and $V_{D,S_n}()$ refer to the state that S_n finds itself in – All D only has one state.

$$\begin{aligned}
V_{S_n,D}(1) &= \pi_{cd} + \delta \cdot ((1-\epsilon) \cdot V_{S_n,D}(2) + \epsilon \cdot V_{S_n,D}(1)) \\
V_{S_n,D}(2) &= \pi_{dd} \cdot \frac{(1-\delta^n)}{(1-\delta)} + \delta^n \cdot V_{S_n,D}(1)
\end{aligned}$$

As above, we plug in for the value in the second state of S_n.

$$\begin{aligned}
V_{S_n,D}(1) = \pi_{cd} + \delta \cdot ((1-\epsilon) \cdot \left(\pi_{dd} \cdot \frac{(1-\delta^n)}{(1-\delta)} + \delta^n \cdot V_1\right) + \epsilon \cdot V_{S_n,D}(1)) \\
\implies V_{S_n,D}(1) = (1-\delta^{n+1} \cdot (1-\epsilon) - \delta\epsilon) = \pi_{cd} + \delta \cdot (1-\epsilon) \cdot \pi_{dd} \cdot \frac{(1-\delta^n)}{(1-\delta)} \\
\implies V_{S_n,D}(1) = \frac{\pi_{cd} + \delta \cdot (1-\epsilon) \cdot \pi_{dd} \cdot \frac{(1-\delta^n)}{(1-\delta)}}{(1-\delta^{n+1} \cdot (1-\epsilon) - \delta\epsilon)}
\end{aligned}$$

Analogously, we get $V_{D,S_n}(1) = \frac{\pi_{dc} + \delta \cdot (1-\epsilon) \cdot \pi_{dd} \cdot \frac{(1-\delta^n)}{(1-\delta)}}{(1-\delta^{n+1} \cdot (1-\epsilon) - \delta\epsilon)}$, and All D simply earns a payoff of $\frac{\pi_{dd}}{1-\delta}$ when playing against copies of itself.

The boundary of the basin of attraction x_{S_n} of S_n is given by this expression,

$$x_{S_n} V_{S_n,S_n}(1) + (1-x_{S_n}) V_{S_n,D}(1) = x_{S_n} V_{D,S_n}(1) + (1-x_{S_n}) V_{D,D}(1)$$

which resolves to $x_{S_n} = \frac{V_{D,D}(1) - V_{S_n,D}(1)}{V_{D,D}(1) - V_{S_n,D}(1) + V_{S_n,S_n}(1) - V_{D,S_n}(1)}$.

We first consider the numerator, we plug in the expressions for the respective payoffs, and then those for π_{dd} and π_{cd}, and reformulate a few times along the way.

$$
\begin{aligned}
V_{D,D}(1) - V_{S_n,D}(1) &= \left(\frac{\pi_{dd}}{1-\delta}\right) - \left(\frac{\pi_{cd} + \delta \cdot (1-\epsilon) \cdot \pi_{dd} \cdot \frac{(1-\delta^n)}{(1-\delta)}}{(1 - \delta^{n+1} \cdot (1-\epsilon) - \delta\epsilon)}\right) \\
&= \frac{\pi_{dd}(1 - \delta^{n+1} \cdot (1-\epsilon) - \delta\epsilon) - (1-\delta)\pi_{cd} - \delta \cdot (1-\epsilon) \cdot \pi_{dd} \cdot (1-\delta^n)}{(1-\delta)(1 - \delta^{n+1} \cdot (1-\epsilon) - \delta\epsilon)} \\
&= \frac{\pi_{dd}(1 - \delta^{n+1} \cdot (1-\epsilon) - \delta\epsilon - \delta \cdot (1-\epsilon) \cdot (1-\delta^n)) - (1-\delta)\pi_{cd}}{(1-\delta)(1 - \delta^{n+1} \cdot (1-\epsilon) - \delta\epsilon)} \\
&= \frac{(1-\delta)(\pi_{dd} - \pi_{cd})}{(1-\delta)(1 - \delta^{n+1} \cdot (1-\epsilon) - \delta\epsilon)} \\
&= \frac{(-c\epsilon - (1-\epsilon)c)}{(1 - \delta^{n+1} \cdot (1-\epsilon) - \delta\epsilon)} \\
&= \frac{(c - 2c\epsilon)}{(1 - \delta^{n+1} \cdot (1-\epsilon) - \delta\epsilon)}
\end{aligned}
$$

For the denominator we proceed in the same way.

$$
\begin{aligned}
&V_{D,D}(1) - V_{S_n,D}(1) + V_{S_n,S_n}(1) - V_{D,S_n}(1)\\
=&\left(\frac{\pi_{dd}}{1-\delta}\right) - \left(\frac{\pi_{cd} + \delta \cdot (1-\epsilon) \cdot \pi_{dd} \cdot \frac{(1-\delta^n)}{(1-\delta)}}{(1-\delta^{n+1} \cdot (1-\epsilon) - \delta\epsilon)}\right) + \left(\frac{\pi_{cc} + \delta\epsilon\pi_{dd}\frac{(1-\delta^n)}{(1-\delta)}}{(1-\delta(1-\epsilon) - \delta^{n+1}\epsilon)}\right)\\
&-\left(\frac{\pi_{dc} + \delta \cdot (1-\epsilon) \cdot \pi_{dd} \cdot \frac{(1-\delta^n)}{(1-\delta)}}{(1-\delta^{n+1} \cdot (1-\epsilon) - \delta\epsilon)}\right)\\
=&\left(\frac{\pi_{dd}}{1-\delta}\right) + \left(\frac{\pi_{cc} + \delta\epsilon\pi_{dd}\frac{(1-\delta^n)}{(1-\delta)}}{(1-\delta(1-\epsilon) - \delta^{n+1}\epsilon)}\right) - \left(\frac{\pi_{dc} + \pi_{cd} + 2\delta \cdot (1-\epsilon) \cdot \pi_{dd} \cdot \frac{(1-\delta^n)}{(1-\delta)}}{(1-\delta^{n+1} \cdot (1-\epsilon) - \delta\epsilon)}\right)\\
=&\left(\frac{\epsilon(b-c)}{1-\delta}\right) + \left(\frac{(1-\epsilon)(b-c) + \delta\epsilon^2(b-c)\frac{(1-\delta^n)}{(1-\delta)}}{(1-\delta(1-\epsilon) - \delta^{n+1}\epsilon)}\right) - \left(\frac{(b-c) + 2\delta \cdot (1-\epsilon) \cdot \epsilon(b-c) \cdot \frac{(1-\delta^n)}{(1-\delta)}}{(1-\delta^{n+1} \cdot (1-\epsilon) - \delta\epsilon)}\right)\\
=&(b-c)\left(\left(\frac{\epsilon}{1-\delta}\right) + \left(\frac{(1-\epsilon) + \delta\epsilon^2\frac{(1-\delta^n)}{(1-\delta)}}{(1-\delta(1-\epsilon) - \delta^{n+1}\epsilon)}\right) - \left(\frac{1 + 2\delta \cdot (1-\epsilon) \cdot \epsilon \cdot \frac{(1-\delta^n)}{(1-\delta)}}{(1-\delta^{n+1} \cdot (1-\epsilon) - \delta\epsilon)}\right)\right)\\
=&\frac{(b-c)}{1-\delta}\left(\left(\frac{(1-\delta)\epsilon}{1-\delta}\right) + \left(\frac{(1-\delta)(1-\epsilon) + \delta\epsilon^2(1-\delta^n)}{(1-\delta(1-\epsilon) - \delta^{n+1}\epsilon)}\right) - \left(\frac{(1-\delta) + 2\delta \cdot (1-\epsilon) \cdot \epsilon \cdot (1-\delta^n)}{(1-\delta^{n+1} \cdot (1-\epsilon) - \delta\epsilon)}\right)\right)
\end{aligned}
$$

Bringing the term inside the outer brackets of this last expression to a common denominator and expanding terms, we get the following.

$$\begin{aligned}
&\frac{(b-c)}{(1-\delta)} \cdot \frac{1}{(1-\delta)(1-\delta(1-\epsilon)-\delta^{n+1}\epsilon)(1-\delta^{n+1}\cdot(1-\epsilon)-\delta\epsilon)} \cdot \Bigg[(1-\delta)\cdot \\
&\Bigg(\epsilon-\delta\epsilon-\delta^{1+n}\epsilon+\delta^{2+n}\epsilon+\delta^{2}\epsilon^{2}-2\delta^{2+n}\epsilon^{2}+\delta^{2+2n}\epsilon^{2}-\delta^{2}\epsilon^{3}+2\delta^{2+n}\epsilon^{3}-\delta^{2+2n}\epsilon^{3}\Bigg) \\
&+(1-\delta)\cdot\Bigg(1-\delta-\delta^{1+n}+\delta^{2+n}-\epsilon+\delta^{2}\epsilon+2\delta^{1+n}\epsilon-2\delta^{2+n}\epsilon+2\delta\epsilon^{2}-\delta^{2}\epsilon^{2} \\
&-2\delta^{1+n}\epsilon^{2}+\delta^{2+2n}\epsilon^{2}-\delta^{2}\epsilon^{3}+2\delta^{2+n}\epsilon^{3}-\delta^{2+2n}\epsilon^{3}\Bigg) \\
&-(1-\delta)\cdot\Bigg(1-2\delta+\delta^{2}+3\delta\epsilon-3\delta^{2}\epsilon-3\delta^{1+n}\epsilon+3\delta^{2+n}\epsilon-2\delta\epsilon^{2}+4\delta^{2}\epsilon^{2}+2\delta^{1+n}\epsilon^{2} \\
&-6\delta^{2+n}\epsilon^{2}+2\delta^{2+2n}\epsilon^{2}-2\delta^{2}\epsilon^{3}+4\delta^{2+n}\epsilon^{3}-2\delta^{2+2n}\epsilon^{3}\Bigg)\Bigg]
\end{aligned}$$

A factor of $(1-\delta)$ cancels, and we reduce the remaining expression in square brackets.

$$\frac{(b-c)}{(1-\delta)} \cdot \frac{\Bigg[(1-\delta)\delta(1-\delta^{n})(1-2\epsilon)^{2}\Bigg]}{(1-\delta(1-\epsilon)-\delta^{n+1}\epsilon)(1-\delta^{n+1}\cdot(1-\epsilon)-\delta\epsilon)}$$

Another factor of $(1-\delta)$ cancels.

$$\frac{(b-c)\delta(1-\delta^{n})(1-2\epsilon)^{2}}{(1-\delta(1-\epsilon)-\delta^{n+1}\epsilon)(1-\delta^{n+1}\cdot(1-\epsilon)-\delta\epsilon)}$$

With these separate expressions for numerator and denominator, we can summarise.

$$
\begin{aligned}
x_{S_n} &= \frac{\frac{(c-2c\epsilon)}{(1-\delta^{n+1}\cdot(1-\epsilon)-\delta\epsilon)}}{\frac{(b-c)\delta(1-\delta^{n})(1-2\epsilon)^2}{(1-\delta(1-\epsilon)-\delta^{n+1}\epsilon)(1-\delta^{n+1}\cdot(1-\epsilon)-\delta\epsilon)}} \\
&= \frac{(c-2c\epsilon)(1-\delta(1-\epsilon)-\delta^{n+1}\epsilon)}{(b-c)\delta(1-\delta^{n})(1-2\epsilon)^2} \\
&= \frac{c(1-\delta(1-\epsilon)-\delta^{n+1}\epsilon)}{(b-c)\delta(1-\delta^{n})(1-2\epsilon)} \\
&= \frac{c(1-\delta+\delta\epsilon-\delta^{n+1}\epsilon)}{(b-c)\delta(1-\delta^{n})(1-2\epsilon)}
\end{aligned}
$$

Now we can take a derivative of this expression w.r.t. n.

$$
\begin{aligned}
\frac{\partial x_{S_n}}{\partial n} &= \frac{-c\delta^{n+1}\epsilon\ln(\delta)\left((b-c)\delta(1-\delta^{n})(1-2\epsilon)\right)+\left((b-c)(1-2\epsilon)\delta^{n+1}\ln(\delta)c(1-\delta+\delta\epsilon-\delta^{n+1}\epsilon)\right)}{(b-c)^2\delta^2(1-\delta^{n})^2(1-2\epsilon)^2} \\
&= \frac{-c\delta^{n+1}\epsilon\ln(\delta)\left(\delta(1-\delta^{n})\right)+\left(\delta^{n+1}\ln(\delta)c(1-\delta+\delta\epsilon-\delta^{n+1}\epsilon)\right)}{(b-c)\delta^2(1-\delta^{n})^2(1-2\epsilon)}
\end{aligned}
$$

The numerator is positive for $\epsilon < \frac{1}{2}$, and the term $\ln(\delta)$ is negative for $\delta < 1$. The expression is therefore negative if the following inequality holds

$$
\begin{aligned}
&c\delta^{n+1}\epsilon\left(\delta(1-\delta^{n})\right)-\Big(\delta^{n+1}c(1-\delta+\delta\epsilon-\delta^{n+1}\epsilon)\Big)<0 \\
&c\delta^{n+1}\epsilon\left(\delta(1-\delta^{n})\right)<\Big(\delta^{n+1}c(1-\delta+\delta\epsilon-\delta^{n+1}\epsilon)\Big)<0 \\
&\qquad\qquad\iff \epsilon\delta-\epsilon\delta^{n+1}<1-\delta+\delta\epsilon-\delta^{n+1}\epsilon \\
&\qquad\qquad\iff 0<1-\delta
\end{aligned}
$$

F The process in our simulations satisfies condition 2 of Theorem 3 if the continuation probability is smaller than one

We start by showing that for continuation probabilities $\delta < 1$, the probabilities of state transitions $P(i, j)$ satisfy condition 2 of Theorem 3 if we take the parameter ϵ in Theorem 3 to denote the error rate – as the notation in the theorem is intended to suggest. In the end we remark that an analogous reasoning works if we take the parameter in Theorem 3 to denote the continuation probability δ.

In the Moran process, transition probabilities depend on the payoffs of all individuals. There are two sources of randomness affecting these payoffs. The first source is the matching. In a population of N individuals there are $\frac{N}{2}$ matched pairs. All possible matchings are equally likely, and their probability does not vary with the error rate or with the continuation probability. The second source of randomness, of course, is the stochasticity in what history occurs for a given matched pair.

We denote a profile of histories in a population as $v = (h_1, h_2, \dots, h_{\frac{N}{2}}) \in H^{\frac{N}{2}}$. Here, subindices denote the matched pairs, and not as in previous sections the duration of the match. The probability of a given transition, i.e. of a given selection event which leads to a change in composition of the population, can be calculated as the probability of this transition given a matching and a realised v, summed over all v that can occur for given matchings, and summed over all matchings. In each summation step, of course, the summands need to be weighted with the probabilities of the respective matchings and history profiles.

Since there are finitely many possible matchings, and since the probabilities of the different matchings do not vary in ϵ, it suffices to show that the probability of each transition conditional on a single arbitrary matching has a bounded derivative. If all summands of a finite sum have bounded derivatives, then the sum also has a bounded derivative.

Below, we show that $\sum_{v \in H^{\frac{N}{2}}} |\frac{dPr(v)}{d\epsilon}|$ is finite. This means that the *marginal* probability mass that is shifted when varying ϵ is bounded. Moreover, since transition probabilities lie between zero and one, shifting a probability mass of x between different history profiles can bring about at most a shift of x in the resulting transition probability for a given matching. Therefore, if $\sum_{v \in H^{\frac{N}{2}}} |\frac{dPr(v)}{d\epsilon}|$ has a finite derivative, so will the transition probability.

To show that $\sum_{v \in H^{\frac{N}{2}}} |\frac{dPr(v)}{d\epsilon}|$ is finite, we start by showing that for every pair of individuals, the sum $\sum_{h \in H} |\frac{dPr(h)}{d\epsilon}|$ is finite. It is convenient, here, to think of histories not in terms of the actions that are played by each player at each stage game, but in terms of whether those actions were or were not brought about by an error. For each given pair of strategies this, of course, uniquely identifies the actions that were played.

The probability that an interaction between two players results in a history h of k rounds in which n errors occur looks as follows.

$$Pr(h) = \epsilon^n \cdot (1-\epsilon)^{2k-n} \cdot \delta^{k-1} \cdot (1-\delta)$$

This expression has the following derivative with respect to ϵ.

$$\frac{dPr(h)}{d\epsilon} = \delta^{k-1} \cdot (1-\delta) \cdot \left[n \cdot \epsilon^{n-1} \cdot (1-\epsilon)^{2k-n} - (2k-n) \cdot (1-\epsilon)^{2k-n-1} \cdot \epsilon^n\right]$$

We distinguish two cases. If $\epsilon = 0$, the derivative of all histories with $n > 1$ vanishes. All probability mass is therefore shifted from histories with $n = 0$ to histories with $n = 1$. The magnitude of this shift looks as follows.

$$\begin{aligned}\frac{1}{2}\left(\sum_{h\in H}\left|\frac{dPr(h)}{d\epsilon}\right|\right)\Bigg|_{\epsilon=0} &= \sum_{k\in\mathbb{N}} \delta^{k-1}\cdot(1-\delta)\cdot\left[2k(1-\epsilon)^{2k-1}\right] \\ &= 2\cdot(1-\delta)\cdot\sum_{k\in\mathbb{N}}\delta^{k-1}\cdot k\end{aligned}$$

For this expression we simply plugged in $n = 0$ into the derivative from above, and used that there is only one possible history without errors for every duration k. The factor $\frac{1}{2}$ comes from the fact that these terms account for half of the shift – histories with $n = 1$ account for the other half.

This expression converges according to d'Alembert's criterion since the summand ratios approach a limit smaller than one: $\lim_{k\to\infty}\frac{\delta^k(k+1)}{\delta^{k-1}k} = \delta < 1$.

Alternatively, if $\epsilon \neq 0$ we can reformulate.

$$\begin{aligned}\left|\frac{dPr(h)}{d\epsilon}\right| &= \left|\delta^{k-1}\cdot(1-\delta)\cdot\left[n\cdot\epsilon^{n-1}\cdot(1-\epsilon)^{2k-n} - (2k-n)\cdot(1-\epsilon)^{2k-n-1}\cdot\epsilon^{n}\right]\right| \\ &= \left|Pr(h)\left[\frac{n}{\epsilon} - \frac{(2k-n)}{(1-\epsilon)}\right]\right| \\ &\leq Pr(h)\frac{1}{\min\{\epsilon, 1-\epsilon\}}2k\end{aligned}$$

Hence, the total shift in probability mass is bounded by the expression below, where now k_h denotes the length of history h. The equality in the expression below follows from the fact that the probabilities of all histories of length k sum to $\delta^{k-1}(1-\delta)$.

$$\sum_{h\in H}\left|\frac{dPr(h)}{d\epsilon}\right| \leq \frac{1}{\min\{\epsilon, 1-\epsilon\}}\sum_{h\in H} Pr(h)\cdot 2k_h = \frac{2\cdot(1-\delta)}{\min\{\epsilon,1-\epsilon\}}\sum_{k\in\mathbb{N}}\delta^{k-1}\cdot k$$

This expression converges as it is simply a scaled version of the sequence we had at $\epsilon = 0$.

Lastly, we show that it follows that $\sum_{v\in H^{\frac{N}{2}}} |\frac{dPr(v)}{d\epsilon}|$ is also finite. To do so, we simply show that the probability of the different v can be transformed to a similar functional form as the probabilities of the histories h. Firstly, we note that as outcomes across pairs are independent, $Pr(v)$ is simply the product of the likelihoods of individual histories.

$$Pr(v) = \prod_{h\in v} Pr(h)$$

$$\begin{aligned}
Pr(v) &= \prod_{h\in v}\epsilon^{n_h}\cdot(1-\epsilon)^{2k_h-n_h}\cdot\delta^{k_h-1}\cdot(1-\delta)\\
&= \epsilon^{\sum_{h\in v} n_h}\cdot(1-\epsilon)^{\sum_{h\in v} 2k_h-n_h}\cdot\delta^{\sum_{h\in v}(k_h-1)}\cdot(1-\delta)^{\frac{N}{2}}\\
&= \left(\frac{1-\delta}{\delta}\right)^{\frac{N}{2}-1}\left(\epsilon^{\sum_{h\in v} n_h}\cdot(1-\epsilon)^{\sum_{h\in v} 2k_h-n_h}\cdot\delta^{(\sum_{h\in v} k_h)-1}\cdot(1-\delta)\right)
\end{aligned}$$

Setting $n = \sum_{h\in v} n_h$ and $k = \sum_{h\in v} k_h$, we arrive back at a scaled version of our expression for the probability of a single history $P(h)$. The factor $\left(\frac{1-\delta}{\delta}\right)^{\frac{N}{2}-1}$ simply scales the sums in the derivations above, but does not affect their convergence. This proves the claim. Intuitively, we used here that we can simply view a history profile as one single concatenated history, and just need to correct for the likelihoods that the different histories end at their respective k_h.

As long as $\delta < 1$, we can also derive the analogue of condition 2 of Theorem 3 for δ instead of ϵ by following the steps above. The derivative $\frac{dPr(h)}{d\delta}$ then looks as follows.

$$\begin{aligned}\frac{dPr(h)}{d\delta} &= \left(\epsilon^n \cdot (1-\epsilon)^{2k-n}\right) \cdot \left[(k-1)\delta^{k-2} \cdot (1-\delta) - \delta^{k-1}\right] \\ &= Pr(h)\left[\frac{(k-1)}{\delta} - \frac{1}{(1-\delta)}\right]\end{aligned}$$

Phrased this in terms of history profiles, with $n = \sum_{h \in v} n_h$ and $k = \sum_{h \in v} k_h$, we then get the following.

$$\begin{aligned}\frac{dPr(v)}{d\delta} &= \left(\epsilon^n \cdot (1-\epsilon)^{2k-n}\right) \cdot \left[(k-1)\delta^{k-2} \cdot (1-\delta)^{\frac{N}{2}} - \frac{N}{2}(1-\delta)^{\frac{N}{2}-1}\delta^{k-1}\right] \\ &= Pr(v)\left[\frac{(k-1)}{\delta} - \frac{N}{2(1-\delta)}\right]\end{aligned}$$

All further arguments are identical to the arguments about changes in ϵ.