Three steps ahead

Heller, Yuval

University of Oxford, Department of Economics

25 June 2013

Three Steps Ahead

*Yuval Heller**(October 4, 2013)*

Address: Nuffield College and Department of Economics, New Road, Oxford, OX1 1NF, United Kingdom. Email: yuval26@gmail.com or yuval.heller@economics.ox.ac.uk.

Abstract

We study a variant of the repeated Prisoner's Dilemma with uncertain horizon, in which each player chooses his foresight ability: that is, the timing in which he is informed about the realized length of the interaction. In addition, each player has an independent probability to observe the opponent's foresight ability. We show that if this probability is not too close to zero or one, then the game admits an evolutionarily stable strategy, in which agents who look one step ahead and agents who look three steps ahead co-exist. Moreover, this is the unique evolutionarily stable strategy in which players play efficiently at early stages of the interaction. We interpret our results as a novel evolutionary foundation for limited foresight, and as a new mechanism to induce cooperation in the repeated Prisoner's Dilemma.

KEYWORDS: Limited foresight, Prisoner's Dilemma, limit ESS. JEL Classification: C73, D03.

1 Introduction

Experimental evidence suggests that people look ahead by only few rounds. For example, players usually defect only at the last couple of stages when playing a finitely repeated Prisoner's Dilemma game (see, e.g., Selten & Stoecker, 1986), and they ignore future opportunities that are more than a few steps ahead when interacting in sequential bargaining (Neelin *et al.*, 1988). A second stylized fact is the heterogeneity of the population: some people systematically look fewer steps ahead than others (see, e.g., Johnson *et al.*, 2002).[1]

These observations raise two related evolutionary puzzles. In many games, the ability to look

---

[1] Similar stylized facts are also observed with respect to the number of strategic iterations in static games, as suggested by the "level-k" models (e.g., Stahl & Wilson, 1994; Nagel, 1995; Costa-Gomes *et al.*, 2001.

ahead by one more step than your opponent can give a substantial advantage. As the cognitive cost of an additional step is moderate in relatively simple games (see, e.g., Camerer, 2003, Section 5.3.5), it is puzzling why there has not been an "arms race" in which people learn to look many steps ahead throughout the evolutionary process (the so called "red queen effect"; Robson, 2003). The second puzzle is how "naive" people in a heterogeneous population, who systematically look fewer steps ahead, survive.

In this paper we present a reduced form static analysis of a dynamic evolutionary process of cultural learning in a large population of agents who play the repeated Prisoner's Dilemma.[2] Each agent is endowed by a type which determines his foresight ability and his behavior in the game. Most of the time agents follow the foresight ability and strategy that they have inherited. Every so often, a few agents experiment with a different type. The frequency of types evolves according to a *payoff-monotonic selection dynamic*: more successful types become more frequent. *Our main results characterize a stable heterogeneous population in which some agents look one step ahead and the remaining agents look three steps ahead, and show that this is the unique stable population in which players cooperate at early stages of the interaction.*

It is well known (Nachbar, 1990) that any dynamically stable state can be represented as a symmetric Nash equilibrium in a 2-player game, in which the players are a metaphor for the evolutionary process. This game includes an initial round in which players choose their foresight ability and $\mathbf{T}$ rounds of repeated Prisoner's Dilemma, where $\mathbf{T}$ is geometrically distributed with a continuation probability close to one (i.e., a high enough expected length).

At stage 0, each player chooses a *foresight ability* (abbreviated, *ability*) from the set $\{L_1, L_2, ..., L_k, ...\}$. A player with ability L_k obtains a private signal at round $\mathbf{T}-k$ about the realization of $\mathbf{T}$. We interpret k as the *horizon* (i.e., number of remaining steps) in which a player with ability L_k becomes aware of the strategic implications of the final period. We discuss this interpretation in Section 8.1. In addition, choosing ability L_k bears a cognitive cost of $c(L_k)$, which is weakly increasing in k (non-monotonic costs are discussed in Section 8.2).

Each player obtains a private signal about his opponent's ability (à la Dekel *et al.*, 2007): (1) the signal reveals the opponent's ability with probability p, and (2) it is non-informative otherwise (independently of the signal that is observed by the opponent). Our interpretation is that each player may observe his opponent's behavior in the past or a trait that is correlated with foresight ability, and he uses such observations to assess his opponent's ability.[3]

The payoffs and actions at stages $1 \leq t \leq \mathbf{T}$ are described in Table 1: Mutual cooperation yields $A > 1$, mutual defection gives 1, and if a single player defects, he obtains $A + 1$ and his

---

[2] This approach was pioneered by Maynard-Smith & Price (1973), and adapted in Guth & Yaari (1992) to study the evolution of preferences. See also Frenkel *et al.* (2012)'s application to cognitive biases.

[3] In Section 8.2 we relax the assumption that p is exogenous, and allow players to influence the probability of observing the opponent's ability.

	C	D
C	A A	0 $A+1$
D	$A+1$ 0	1 1

Tab. 1: Payoff at the symmetric stage game Prisoner's Dilemma ($A > 1$).

opponent gets 0.[4] The total payoff of the game is the undiscounted sum of payoffs.

We begin by characterizing a specific symmetric Nash equilibrium, σ^*, for every p that is not too close to zero and one (and the width of this interval is increasing in A, see Figure 1). The support of σ^* includes two abilities (dubbed, *incumbents*): L_1 and L_3, where $\mu(L_1)$ is increasing in A and p. Strategy σ^* induces a simple deterministic play: when the horizon is still uncertain players follow a "perfect" variant of "tit-for-tat" (dubbed, *pavlov*): defect if and only if players have played different actions in the previous round.[5] Everyone defects at the last round. A player with ability L_3 defects also at the penultimate round, and his behavior at the previous round (i.e., stage $\mathbf{T}-2$) depends on the signal about the opponent's ability: he follows *pavlov* if it is either L_1 or unknown, and defects otherwise.

Intuitively, the equilibrium relies on two observations: (1) if p is not too low, there is a unique frequency that induces a balance between the direct disadvantage of having ability L_1 (loosing one point by cooperating at horizon 2), and its indirect "commitment" advantage (when an L_3 opponent observes ability L_1, it induces him to cooperate an additional round); and (2) if p is not too high, then it is optimal to follow *pavlov* at stage $\mathbf{T}-3$ also when a player has higher ability than L_3.

Nash equilibria may be dynamically unstable. Maynard-Smith & Price (1973) refine it as follows: Nash equilibrium σ is an *evolutionarily stable strategy* (abbreviated, *ESS*) if it is a better reply against any other best-reply strategy σ' ($u(\sigma', \sigma) = u(\sigma, \sigma) \Rightarrow u(\sigma, \sigma') > u(\sigma', \sigma')$). The motivation is that an ESS, if adopted by a population of players, cannot be invaded by any alternative strategy that is initially rare. Cressman (1997) shows that an ESS is dynamically stable for a large set of payoff-monotonic selection dynamics. Repeated games rarely admit an ESS due to "equivalent" strategies that differ only in off-equilibrium paths. In particular, the repeated Prisoner's Dilemma does not admit any ESS (Lorberbaum, 1994).

---

[4] We assume that defection yields the same additional payoff (relative to cooperation) regardless of the opponent's strategy to simplify the presentation of the result and their proofs. The results remain qualitatively similar also without this assumption. Besides this assumption, the table represents a general Prisoner's Dilemma game (up to an affine normalization).

[5] The name "Pavlov" (Kraines & Kraines, 1989; Nowak & Sigmund, 1993) stems from the fact that it embodies an almost reflex-like response to the payoff of the previous round: it repeats its former move if it was rewarded by a high payoff (A or $A+1$), and it switches if it was punished by receiving a low payoff (0 or 1).

Selten (1983) adapts the notion of ESS to extensive-form games as follows. A *perturbation* is a function that assigns a minimal probability to play each action at each information set. Strategy σ is a *limit ESS* if it is the limit of ESS of a sequence of perturbed games when the perturbations converge to 0.[6] Observe that any ESS is a limit ESS, and that any limit ESS is a symmetric perfect equilibrium (Selten, 1975, see Corollary 1). Our first main result (Theorem 1) shows that σ^* is a limit ESS.[7]

Similar to other repeated games, the interaction admits many stable strategies. In Section 5, we present a "folk theorem" result: for any k, m and n, there exists a limit ESS in which everyone has ability L_k, and as long as the horizon is uncertain, players repeat cycles in which they cooperate m times and defect n times.

Thus, uniqueness is possible only when focusing on a subset of stable strategies. We shall say that a strategy is *early-nice* if players cooperate when: (1) the horizon is sufficiently large, and (2) no one has ever defected in the past. Equivalently, a strategy is early-nice if it induces efficient play (mutual cooperation) at early stages of the interaction, also when one of the players "trembles" and chooses a different ability. Our second main result (Theorem 2) shows that, if $A > 3$, then any early-nice limit ESS is equivalent to σ^*: it induces the same distribution of abilities and the same play on the equilibrium path. In Section 7 we extend the uniqueness result to weaker solution concepts: a neutrally stable strategy and a perfect equilibrium. Figure 1 graphically summarizes our main results for different values of A and p. Observe that no early-nice stable strategies exist if p is close to either zero or one.

The intuition of Theorem 2 is as follows. Let $L_{\underline{k}}$ be the lowest incumbent ability. Observe that everyone must defect during the last $\underline{k}$ rounds because the event of reaching the $\underline{k}^{\text{th}}$ to last round is common knowledge among the players. If $L_{\underline{k}}$ is the unique incumbent ability, then "mutants" with ability $L_{\underline{k}+1}$ outperform the incumbents by defecting one stage earlier. If there are two consecutive abilities, then the lower ability is outperformed by the higher one. If there is a gap of more than two steps between the lowest and the highest ability in the population and A is sufficiently large, then it turns out that the strategy is unstable to small perturbations in the frequencies of the different abilities in between. Thus the support must be $\left\{ L_{\underline{k}}, L_{\underline{k}+2} \right\}$ for some $\underline{k}$. If $L_{\underline{k}} > L_1$ then "mutants" with ability L_1 can induce additional rounds of mutual cooperation and outperform the incumbents. Finally, stability fails if p is too close to 0 because the indirect advantage of having a low ability is too small, and it fails if p is close to one because players are "trapped" in an "arms race" towards earlier defections and higher abilities.

Various arguments motivate us to focus on early-nice strategies: (1) Robson (1990) demonstrates that "secret-handshake mutants" in a rich enough environment can take the population

---

[6] A few examples for applications of limit ESS are: Samuelson (1991); Kim (1993, 1994); Bolton (1997); Leimar (1997).

[7] Moreover, we show that σ^* is the limit of ESS of *every* sequence of perturbed games (*strict limit ESS*).

Fig. 1: Summary of Main Results
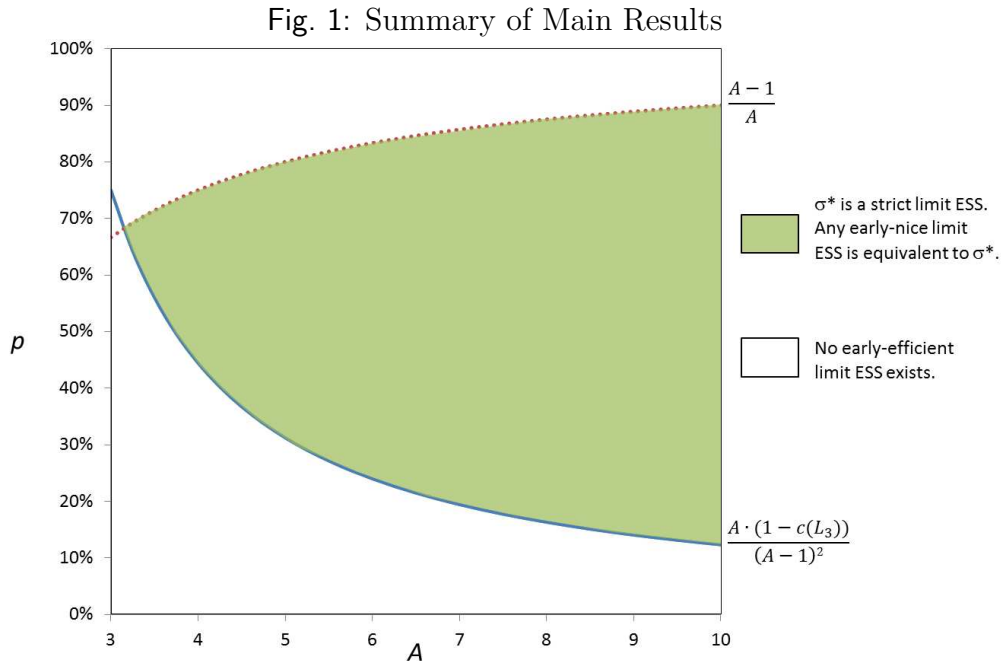
away from inefficient states; (2) Fundenberg & Maskin (1990) and Binmore & Samuelson (1992) study the undiscounted infinitely repeated Prisoner's Dilemma and show that mutual cooperation at almost all rounds is the unique evolutionarily stable outcome given a small amount of "noise" or small costs for using complex strategies; (3) Kreps *et al.* (1982) study the long finitely repeated Prisoner's Dilemma, and show that small uncertainty about the preferences of the opponent imply mutual cooperation at early stages of the game; (4) the tournaments of Axelrod (1984) and Wu & Axelrod (1995) suggest that "niceness" (not being the first to defect) might be a necessary requirement for evolutionary success; and finally, (5) Selten & Stoecker (1986) experimentally demonstrate that most subjects satisfy early-niceness when playing the repeated Prisoner's Dilemma in the lab.[8]

Our formal analysis deals only with the repeated Prisoner's Dilemma. It is relatively simple to extend the results to other games in which looking far ahead decreases efficiency, such as "centipede" (Rosenthal, 1981). Such interactions are important both in primitive *hunter-gatherer* societies (representing sequential gift exchange, see, e.g., Haviland *et al.*, 2007, p. 440), as well as in modern societies. In Section A.2 we sketch how to extend our results to a setup in which players also play other games.

We conclude by briefly surveying the related literature. The closest paper to ours is Jehiel

---

[8] Selten & Stoecker (1986) experimentally study how people play a repeated Prisoner's Dilemma with 10 rounds (see similar results in Andreoni & Miller (1993); Cooper *et al.* (1996); Bruttel *et al.* (2012)). They show that: (1) there is usually mutual cooperation in the first six rounds, (2) players begin defecting only during the last four rounds, and (3) if any player defected, then almost always both players defect at all remaining stages. Johnson *et al.* (2002)'s findings suggest that limited foresight is the main cause for this behavior.

(2001), which assumes a fixed level of limited foresight in the infinitely repeated Prisoner's Dilemma, and shows that in all equilibria, players cooperate at all stages except the first few rounds.[9] The key difference between our paper and Jehiel (2001) is that we obtain limited foresight as a result, rather than assuming it. That is, in our model players can acquire long foresight abilities with low costs (or without costs at all), and yet there is a stable state (unique under the additional assumption of early-niceness) in which everyone choose to look only a few steps ahead. In addition, the current paper presents a novel notion of limited foresight (see Section 8.1), which may be of independent interest.

Geanakoplos & Gray (1991) study complex sequential decision problems and describe circumstances under which looking too far ahead in a decision tree leads to poor choices. Stahl (1993), Stennek (2000) and Mohlin (2012) present evolutionary models of bounded strategic reasoning ("level-k"), which are related to our model when p is equal to zero or one. This paper is novel in introducing partial observability in this setup, and showing that it yields qualitatively different results. Crawford (2003) studies zero-sum games with "cheap talk" and shows that naive and sophisticated agents may co-exist and obtain the same payoff.

Finally, our paper is related to the literature that studies the stability of cooperation in the repeated Prisoner's Dilemma (various such papers are cited elsewhere in the introduction). One may gain additional insights by reinterpreting the abilities as voluntary commitment devices that are available to the players, with the property (which is implied by early-niceness) that a committed player cannot punish an opponent who has chosen not to commit. Our results demonstrate that such simple commitment devices can induce mutual cooperation at all rounds except the last few, provided that the probability of observing the opponent's commitment is not too close to either zero or one.

The paper is structured as follows. Section 2 presents the model. In Section 3, we characterize the symmetric Nash equilibrium σ^*. Section 4 shows that strategy σ^* is a limit ESS. Section 5 presents a "folk-theorem" result. Section 6 shows that σ^* is essentially the unique early-nice limit ESS, and Section 7 extends it to weaker solution concepts. In Section 8 we discuss the interpretation of limited foresight and sketch a few extensions and variants, which are formalized in Appendix A. Finally, Appendix B includes the formal proofs.

---

[9] Recently, Mengel (2012) obtains a similar result for the finitely repeated Prisoner's Dilemma while using stochastic stability as the solution concept. Two other related papers are Samuelson (1987) and Neyman (1999), which show that if the (exogenous) information structure slightly departs from common knowledge about the final period, then there is an equilibrium in which players almost always cooperate in the finitely repeated Prisoner's Dilemma.

2   Model

We study a symmetric 2-player extensive-form game. As discussed in the introduction, the players should be interpreted as a metaphor for the evolutionary process.[10]

2.1   Abilities and Signals

The interaction includes an initial round in which players choose their foresight ability and $\mathbf{T}$ rounds of repeated Prisoner's Dilemma. Random variable $\mathbf{T} - 2$ is geometrically distributed with parameter $1 - \delta$, where $0 < \delta < 1$ describes the continuation probability at each stage: $\delta = \Pr(\mathbf{T} > k | \mathbf{T} = k)$ (for each $k > 2$).[11] We focus on the case of δ close enough to 1.

At stage 0 each player $i \in \{1, 2\}$ chooses his *ability* from the set $\mathcal{L} = \{L_1, L_2. L_3, ..., L_k, ...\}$.[12] We shall say that L_k is larger (resp., weakly larger, smaller) than $L_{k'}$ if $k > k'$ (resp., $k \geq k'$ $k < k'$). Let $L_{\geq k}$ denote the set of abilities weakly larger than L_k. Intuitively, the ability of a player determines when he will become aware of the realized length of the interaction and its strategic implications. Formally, a player with ability L_k privately observes at round $\max(\mathbf{T} - k, 0)$ the realization of $\mathbf{T}$. In Section 8.1 we discuss the interpretation of the abilities and the uncertain length.

Players partially observe the ability of the opponent as follows (à la Dekel *et al.*, 2007). At the end of stage 0, each player privately observes his opponent's ability with probability p, and he obtains no information otherwise (independently of the signal that is observed by his opponent.)[13] We shall say that a player is *uninformed* as long as he has not yet received the signal about the realized length, and is *informed* afterwords. We shall use the term *stranger* to describe an opponent whose foresight ability is not observed, and we shall use the term *observing* (*non-observing*) to describe a player who has observed (not observed) his opponent's ability.

Let $c : \mathcal{L} \to \mathbb{R}^+$ be an arbitrary weakly increasing function, which describes the *cognitive cost* of each foresight ability.[14] That is, a player who chooses ability L_k obtains a negative payoff of $-c(L_k)$. Without loss of generality, we normalize: $c(L_1) = 0$. At each stage $1 \leq t \leq \mathbf{T}$ the players play the Prisoner's Dilemma as described in Table 1 with two pure actions -$\{C, D\}$.

---

[10] The model also fits two additional interpretations: (1) a biological evolutionary process in which the type is determined by the gene; and (2) a non-evolutionary strategic interaction in which a player first chooses how much effort to spend on detecting earlier signs that the interaction is going to end soon (his foresight ability), and the players play the repeated Prisoner's Dilemma.

[11] To simplify the presentation of the results, we assume that $\mathbf{T} - 2$ rather than $\mathbf{T}$ has a geometric distribution. The results remain qualitatively the same without this assumption.

[12] Results are robust to having either a maximal ability or a minimal ability different than L_1 (see Sec. 8.2).

[13] In Section 8.2 we show that the results are robust to the timing in which a player may observe his opponent's ability, and we demonstrate how to extend to model to allow p to be determined endogenously by the players.

[14] We relax the assumption of weakly increasing cognitive costs in Section 8.2.

2.2 Strategies and Payoffs

Given $i \in \{1, 2\}$, let $-i$ denote the other player. An *information set* of length $n > 0$ of player $i \in \{1, 2\}$ is a tuple $I = \left(L, l, s, \left(a^i, a^{-i}\right)^n\right)$ where: (1) $L \in \mathcal{L}$ is the player's ability (as chosen at stage 0); (2) $l \in \{1, ..., L\} \cup \infty$ is the number of remaining periods (dubbed, the *horizon*), with $l = \infty$ describing an uninformed agent; (3) $s \in \{\mathcal{L} \cup \phi\}$ is the signal about the opponent's ability, with $s = \phi$ describing a non-informative signal (i.e., facing a stranger); and (4) $\left(a^i, a^{-i}\right)^n \in \left(\{C, D\} \times \{C, D\}\right)^n$ describes the actions that were publicly observed so far in the game. Let $\mathcal{I}_n$ denote the set of all information sets of length n, and let $\mathcal{I} = \cup_{n \geq 1} \mathcal{I}_n$ be the set of all information sets.

A *behavior strategy* (abbreviated, *strategy*) is a pair $\sigma = (\mu, \beta)$, where $\mu \in \Delta(\mathcal{L})$ is a distribution over the abilities, and $\beta : \mathcal{I} \to \Delta(\{C, D\})$ is a function that assigns a mixed action for each information set (dubbed, *playing-rule*). The abilities in $supp(\mu)$ shall be called the *incumbents*. Let $\Sigma(B)$ denote the set of all strategies (playing-rules). With slight abuse of notation, we can identify a pure distribution with a single ability in its support. A pure playing-rule, which induces a deterministic play at all information sets, is described by the function $b : \mathcal{I} \to \{C, D\}$.

The total payoff of the game is the undiscounted sum of the stage payoffs (including the cognitive cost at stage 0). This is formalized as follows. A history of play (abbreviated, *history*) of length n is a tuple $\left(\left(L^1, L^2\right), \left(a^1, a^2\right)^n\right)$ where (L^1, L^2) describes the abilities chosen at stage 0, and $\left(a^1, a^2\right)^n$ describes the n actions taken at stages $1, ..., n$. Let H_n be the set of histories of length n. For each history $h_n \in H_n$, let the payoff of player 1 be defined as follows:

$$u(h_n) = u\left(\left(L^1, L^2\right), \left(a^1, a^2\right)^n\right) = \sum_{k \leq n} u\left(a_k^1, a_k^2\right) - c\left(L^1\right),$$

where $u(a^1, a^2)$ is the Prisoner's Dilemma stage payoff as given by Table 1.

For each game length T, history $h_T \in H_T$, and pair of strategies σ, σ', let $\Pr_{\sigma, \sigma'}(h_T | \mathbf{T} = T)$ be the probability of reaching history h_T when player 1 plays strategy σ, player 2 plays strategy σ', conditional on the random length of the game being equal to T. The expected payoff of a player who plays strategy σ and faces an opponent who plays strategy σ' is defined as follows:

$$u(\sigma, \sigma') = \sum_{T \in \mathbb{N}} \Pr(\mathbf{T} = T) \cdot \sum_{h_T \in H_T} \Pr_{\sigma, \sigma'}(h_T | \mathbf{T} = T) \cdot u(h_T).$$

Remark 1. Some readers may wonder why we study a cognitive bias (limited foresight) but allow agents to use complex strategies with perfect memory. We consider this aspect of the model an advantage rather than a weakness. The model allows agents to use complex strategies with long memories and long foresight abilities, and yet, it implies a unique early-nice stable

outcome in which all players choose to have a small foresight ability, and to use simple strategies that depend only on the realized actions in the previous stage. We note that all our results remain the same if one adds a restriction to the model either to how many rounds of play the agents can remember; or to the complexity of strategies that the agents may use.

3 Characterization of a Nash Equilibrium

As mentioned in the Introduction, we use a reduced form static analysis to study the long-run stable outcomes of a payoff-monotonic dynamic in which more successful types become more frequent. A state of the population is *Lyapunov stable* if no small change in the population composition can lead it away. Nachbar (1990) shows that any Lyapunov stable state is a Nash equilibrium as well. Motivated by this observation, we characterize in this section a specific Nash equilibrium, σ^*. We emphasis that this equilibrium behavior can be achieved by agents who passively follow their types, rather then actively maximize their payoffs.[15]

A strategy is a symmetric Nash equilibrium if it is a best-reply to itself.

Definition 1. Strategy $\sigma \in \Sigma$ is a *symmetric Nash equilibrium* if $u(\sigma, \sigma) \geq u(\sigma', \sigma) \; \forall \sigma' \in \Sigma$.

Strategy σ^* assigns positive probabilities to two abilities: L_1 and L_3, and it induces deterministic simple playing-rule: following *pavlov* (defect iff players played differently at the previous round) at long horizons; and *defect* at short horizons. The horizon in which the behavior changes from *pavlov* to *defect* depends on the signal about the opponent: it happens at horizon 2 against strangers and observed L_1 opponents, and at horizon 3 otherwise.

Definition 2. For every $p > 0$, $A > 1$ and $c(L_3) < 1$, let $\sigma^* = (\mu^*, b^*)$ be as follows:

$$\mu^*(L_1) = 1 - \frac{1 - c(L_3)}{p \cdot (A - 1)}, \;\; \mu^*(L_3) = \frac{1 - c(L_3)}{p \cdot (A - 1)}, \;\; \forall k \notin \{1, 3\} \; \mu^*(L_k) = 0;$$

$$b^*\left(L, l, s, \left(a^i, a^{-i}\right)^t\right) = \begin{cases} C & (l \geq 4 \text{ or } (l = 3 \text{ and } s \in \{L_1, \phi\})) \text{ and } \left(t = 0 \text{ or } a_t^i = a_t^{-i}\right) \\ D & \text{otherwise} \end{cases}.$$

It might be helpful to interpret strategy σ^* as describing a heterogeneous population of agents, which includes two fractions with different abilities. The L_1 agents play *pavlov* until the last round of the interaction (in which they defect). The L_3 agents play *pavlov* until they are being informed about the final period. Their behavior at this stage (horizon 3) depends on the signal about the opponent: they play *pavlov* if they face a stranger or L_1, and they defect if they obtain a different signal. In the last two stages the L_3 agents defect.

---

[15] Though the result also hold in the presence of sophisticated agents who explicitly maximize their payoffs.

Our first result shows that (μ^*, b^*) is a Nash equilibrium if p is not too close to zero or one, the cognitive cost of L_3 is not too high, and the continuation probability δ is close enough to one.

Proposition 1. *Assume that: (1) $\frac{A \cdot (1 - c(L_3))}{(A-1)^2} < p < \frac{A-1}{A}$, (2) $c(L_3) < \frac{1}{A}$, and (3) δ is sufficiently close to 1. Then σ^* is a symmetric Nash equilibrium.*

Proposition 1 is implied by Theorem 1 (which is proved in Appendix B, as with the other results in the paper). The sketch of the proof is as follows. The two fractions of agents in the population play a Hawk-Dove reduced game: when two L_1s ("doves") meet, they obtain a relatively high payoff due to cooperating at all rounds except the last one; an L_3 agent achieves one more utility point against L_1 by defecting at horizon 2; and when two L_3s meet, they obtain a relatively low payoff (if p is not too low) because when one of them observes the other, he defects at horizon 3, and this early defection implies an inefficient outcome. Thus, each fraction becomes less successful (relative to the other fraction) if its frequency becomes larger, and a unique frequency (μ^*) balances their payoffs.

Next, we show that b^* is a best reply to σ^* for all abilities (including abilities in $L_{\geq 4}$). Observe that being cooperative is optimal for uninformed agents (when δ is close enough to one). This is because the gain from defection (one point) is outweighed by the loss of $A - 1$ points at the next round, in which the opponent defects instead of cooperating. Informed L_1 agents play a dominant action at the last round (*defect*). Observe that the best reply against an opponent who plays *pavlov* until horizon k and defects at shorter horizons is to defect one round earlier at horizon $k + 1$. This implies that the behavior of informed L_3 agents against L_1 is optimal. The behavior of L_3 agents against strangers (start defecting at horizon 2) is optimal, as long as the fraction of L_1 agents ($\mu^*(L_1)$) is not too low, and this holds if p is not too small. The behavior of L_3 agents when observing an L_3 opponent (start defecting at horizon 3) is optimal (even if they had a larger foresight ability) as long as p is not too high and the opponent is likely enough to be unobserving. This implies that choosing any ability in $L_{\geq 4}$ cannot improve the payoff. Finally, choosing ability L_2 yields a strictly lower payoff because: (1) ability L_2 induces the same opponent's behavior as L_3; (2) L_2 agents cannot defect at horizon 3 when observing an L_3 opponent, and this yields them one less utility point.

Remark 2. Proposition 1 holds also if *pavlov* is replaced with a different reciprocal behavior that induces cooperative behavior on the equilibrium path, such as "tit-for-tat" (defect iff the opponent defected in the previous round), or "perfect-grim-trigger" (defect iff any player defected before). We present the results with *pavlov* because it satisfies three appealing properties: (1) it satisfies the refinement of evolutionary stability introduced in the next section; contrary to this, "tit-for-tat" implies non-optimal play off the equilibrium path: following a defection

of the opponent, it is strictly better to cooperate rather than defect; (2) it is a very simple strategy that depends only on the actions of the last round; and (3) it implies efficiency (mutual cooperation most of the time) also when there are small error probabilities (see, Nowak & Sigmund, 1993). Contrary to this, the "perfect-grim-trigger" induces inefficient play in "noisy" environments.

4 Evolutionary Stability

A Nash equilibrium may be dynamically unstable. Maynard-Smith & Price (1973) refined Nash equilibrium, and presented the notion of evolutionary stability. A symmetric Nash equilibrium σ is evolutionarily (neutrally) stable if it achieves a strictly (weakly) better payoff against any other best-reply strategy σ'. Formally:

Definition 3. (Maynard-Smith & Price (1973), as reformulated for behavior strategies in Selten (1983)) Strategy $\sigma \in \Sigma$ is *an evolutionarily (neutrally) stable* strategy (abbreviated, respectively, ESS, NSS) if: (1) it is a symmetric Nash equilibrium, and (2) $\forall \sigma' \neq \sigma$, if $u(\sigma', \sigma) = u(\sigma, \sigma)$ then $u(\sigma, \sigma') > u(\sigma', \sigma')$ ($u(\sigma, \sigma') \geq u(\sigma', \sigma')$).

The motivation for Definition 3 is that an ESS, if adopted by a population of players in a given environment, cannot be invaded by any alternative strategy that is initially rare. Cressman (1997) and Sandholm (2010) show that an ESS is an asymptotic stable state under a large variety of payoff-monotonic dynamic processes; that is, the population converges to the ESS from any close enough initial state.[16]

Repeated games rarely admit an ESS due to the existence of "equivalent" strategies that differ only off the equilibrium path. In particular, our model admits no ESS.[17] Selten (1983) slightly weakens this notion by requiring evolutionary stability in a converging sequence of perturbed games (but not necessarily in the unperturbed game). Formally:

Definition 4. (Selten (1983, 1988)) A perturbation (ubiquitous perturbation) ζ is a function that assigns a non-negative (positive) number for:

1. each ability at stage 0 such that $\sum_{L_k \in \mathcal{L}} \zeta(L_k) < 1$; and

2. each action (C or D) after each information set $I \in \mathcal{I}$, such that $\zeta(C)(I) + \zeta(D)(I) < 1$.

---

[16] They show this dynamic stability for a slightly stronger notion, regular evolutionarily stable strategy, which also satisfies strictness with respect to strategies outside its support. Strategy σ^* (Def. 2) satisfies regularity in any ubiquitous perturbed game if $c(L_3) < c(L_4)$ (see Selten (1988) for a formal definition of regularity in this setup.)

[17] See Lorberbaum (1994) for a proof that the repeated Prisoner's Dilemma with uncertain horizon does not admit any evolutionarily stable strategy. Similarly, one can adapt the proof, and show that it does not admit an evolutionarily stable set (Thomas, 1985) or an equilibrium evolutionarily stable set (Swinkels, 1992).

Let $\boldsymbol{\Gamma}(\zeta)$ be a (symmetric) perturbed version of the game described in Section 2 (*perturbed game*), in which each player is limited to choose strategy $\sigma = (\mu, \beta)$ that satisfies: (1) $\mu(L_k) \geq \zeta(L_k)$ for each $L_k \in \mathcal{L}$, and (2) $\zeta(I)(C) \leq \beta(I)(C) \leq 1 - \zeta(I)(D)$ for each $I \in \mathcal{I}$. Let $\Sigma(\zeta)$ (resp., $\Delta_\zeta(\mathcal{L})$, $B(\zeta)$) the set of all strategies (resp., distributions, playing-rules) that satisfy these two properties (resp., the first property, the second property). Let $M(\zeta)$ denote the *maximal tremble* of ζ: $M(\zeta) = \max\left(\sup_{L_k \in \mathcal{L}} \zeta(L_k), \sup_{I \in \mathcal{I}, a \in \{C,D\}} \zeta(I)(a)\right)$.

Definition 5. (Selten (1983)) Strategy $\sigma \in \Sigma$ is a *limit ESS* if there exists a sequence of perturbations $(\zeta_n)_{n \in \mathbb{N}}$ satisfying $\lim_{n \to \infty} M(\zeta_n) = 0$, and for each $n \in \mathbb{N}$, there exists an ESS σ_n of the perturbed game $\Gamma(\zeta_n)$, such that $\lim_{n \to \infty} \sigma_n = \sigma$ is satisfied.

Observe that any ESS is a limit-ESS, and that any limit ESS is a symmetric perfect equilibrium (Selten, 1975).[18]

In order to strengthen our stability result, we present a stronger notion than Definition 5 by requiring a *strict* limit ESS to be the limit of ESS of *every* sequence of strict perturbed games (rather than a specific sequence). The motivation (similar to Okada, 1981's notion of strict-perfection), is that a strong notion of stability should be robust to the specific structure of the perturbations. Formally:

Definition 6. Strategy $\sigma \in \Sigma$ is a *strict limit ESS* (*strict limit NSS*) if, for every sequence of ubiquitous perturbations $(\zeta_n)_{n \in \mathbb{N}}$ satisfying $\lim_{n \to \infty} M(\zeta_n) = 0$, and for every $n \in \mathbb{N}$, there exists an ESS (NSS) of $\Gamma(\zeta_n)$, such that $\lim_{n \to \infty} \sigma_n = \sigma$ is satisfied.

Our first main result strengthens Proposition 1 and shows that σ^* is a strict limit ESS.

Theorem 1. *Assume that: (1)* $\frac{A \cdot (1 - c(L_3))}{(A-1)^2} < p < \frac{A-1}{A}$*, (2)* $c(L_3) < \frac{1}{A}$*,*[19] *and (3)* δ *is sufficiently close to 1. Then* σ^* *is a strict limit NSS, and if* $c(L_4) > c(L_3)$ *then* σ^* *is a strict limit ESS.*

The sketch of the proof is as follows. Let ζ be any sufficiently small ubiquitous perturbation, and let $\sigma_\zeta^* = \left(\mu_\zeta^*, b_\zeta^*\right)$ be the closest strategy to σ^* in the perturbed game $G(\zeta)$ that satisfies $u\left(\left(L_1, b_\zeta^*\right), \sigma_\zeta^*\right) = u\left(\left(L_3, b_\zeta^*\right), \sigma_\zeta^*\right)$. Lorberbaum et al. (2002) proved that the perturbed-*pavlov* is a strict-best reply to itself when playing slightly-perturbed *standard* repeated Prisoner's Dilemma (in which players remain uninformed throughout the game). Together with the arguments from the sketch of proof of Proposition 1, this implies that (1) playing-rule b_ζ^* is a strict best-reply to σ_ζ^* (for all abilities), (2) ability L_2 achieves a strictly lower payoff than L_3, and (3) any ability $L_{>3}$ can achieve, at most, the same payoffs as L_3. The properties of the Hawk-Dove "meta-game" between abilities L_1 and L_3 (discussed in Section 3) imply that any strategy with

---

[18] See Corollary 1, which slightly strengthens the result for general extensive-form games of van Damme (1987, Corollary 9.8.6) that any limit ESS is a sequential equilibrium (Kreps & Wilson (1982))

[19] Assumption 2 can be slightly weakened as follows $c(L_3) < \min\left(1, \frac{1}{A} + \left(1 - \frac{1}{A}\right) \cdot c(L_2)\right)$.

a different frequency of L_1-s and L_3-s yields a strictly lower payoff. This shows that σ_ζ^* is an NSS of $\Gamma(\zeta)$ and an ESS if $c(L_4) > c(L_3)$.

Remark 3. Minor adaptions to the proof imply a slightly stronger result when $c(L_4) = c(L_3)$. Let $L_{as\_3} = \{L_k | k \geq 3, \ c(L_k) = c(L_3)\}$ be the abilities with the same costs as L_3. Then:

$$\Sigma^* = \{(\mu, \beta^*) | \mu(L_1) = \mu^*(L_1) \text{ and } \forall k \notin L_{as\_3} \cup \{L_1\} \ \ \mu(L_k) = 0\}$$

is a "strict limit evolutionarily stable set": it is the limit of evolutionarily stable sets (Thomas, 1985) of any sequence of converging ubiquitous perturbed games.

5 All Abilities Can be Stable

Strategy σ^* is efficient in the sense that players always cooperate on the equilibrium path, except for the last few rounds. The following proposition shows that the game also admits an inefficient stable strategy in which all players have ability L_1 and always defect.

Proposition 2. *Let $\sigma_{def} = (L_1, b_{def})$ with $b_{def} \equiv D$ (always defect). Then σ_{def} is a strict limit NSS. Moreover, if $c(L_2) > c(L_1)$ then σ_{def} is a strict limit ESS.*

The proof adapts Lorberbaum *et al.* (2002)'s result that defection is a strict best reply to itself in the slightly-perturbed repeated Prisoner's Dilemma.

The following proposition shows a "folk theorem" result: for any ability L_k and for any finite sequence of actions, there exists a strict limit ESS in which all players have ability L_k and they keep playing cycles of the sequence as long as they are uninformed. Formally:

Proposition 3. *Let $L_k \in \mathcal{L}$, $M \in \mathbb{N}$, and $S \in (\{C, D\})^M \neq (D, ..., D)$. Assume that $0 < p$, and that $\delta = \delta(p)$ is sufficiently close to 1. There exists a strict limit ESS $\sigma = (L_k, \beta)$ in which on the equilibrium path uninformed players repeat playing cycles of the sequence S.*

Kim (1994) studies the standard repeated Prisoner's Dilemma, and shows that any finite sequence of actions can be implemented as a strict limit ESS for δ sufficiently close to one by using "perfect-grim-trigger punishments" off-equilibrium path. Our proof extends Kim's result to the setup with abilities as follows. On the equilibrium path, players with ability L_k repeat playing cycles of the sequence S as long as they are uninformed, and they defect at the last k stages. If an L_k player observes an $L_{k'} \neq L_k$ opponent, he plays a cycle of an asymmetric sequence of action-profiles W', which yields the L_k ($L_{k'}$) player a higher (lower) payoff relative to sequence S. If any player deviate from this pattern, both players always defect.

6 Early-Niceness and Uniqueness

A strategy is early-nice if players who follow its playing-rule cooperate when: (1) the horizon is large enough, and (2) no one has ever defected before. Formally:

Definition 7. Strategy $\sigma = (\mu, \beta) \in \Sigma$ is *early*-nice if there exists $M_\sigma \in \mathbb{N}$ such that $\beta\left(L, l, s, \left(a^i, a^{-i}\right)^n\right)(C) = 1$ if: (1) $l > M_\sigma$, and (2) $\left(a^i, a^{-i}\right)^n = (C, C)^n$ (dubbed, *cooperative information set*).

Early-niceness implies efficient play (mutual cooperation) at early stages of the interaction on the equilibrium path. Proposition 3 shows that this implication is not enough to restrict the set of stable abilities: any ability L_k can be the unique incumbent in a limit ESS which induces early inefficient play only against non-incumbent abilities. Early-niceness also requires efficient play in cases in which one of the players (or both) has "trembled" and chosen an ability outside the support of μ. Note that early-niceness does not restrict the play of a "mutant" player who follows a different playing-rule.

Two strategies are equivalent if they induce the same distribution of abilities and the same observable play; they can only differ in their off-equilibrium behavior. Formally:

Definition 8. Strategies $\sigma, \sigma' \in \Sigma$ are equivalent if for each possible game length T and for each history $h_T \in H_T$: $\Pr_{\sigma,\sigma}\left(h_T | \mathbf{T} = T\right) = \Pr_{\sigma',\sigma'}\left(h_T | \mathbf{T} = T\right)$.

Our second main result shows that any early-nice limit ESS is equivalent to σ^* (assuming $a > 3$), and that there is no early-nice limit ESS for values of p that are close to either zero or one.

Theorem 2. *Assume that δ is close enough to 1, $A > 3$, $\forall k \in \mathbb{N}\ c\left(L_{k+1}\right) - c\left(L_k\right) < 1$, $c\left(L_4\right) < \frac{1}{A}$, and $0 \le p < 1$. If strategy $\sigma = (\mu, \beta)$ is a an early-nice limit ESS, then $\sigma \approx \sigma^*$. Moreover, if $p < \frac{A \cdot (1 - c(L_3))}{(A-1)^2}$ or $\frac{A-1}{A} < p$ then no early-nice limit ESS exists.*

The sketch of the proof is as follows. Let $L_{\underline{k}}$ be the lowest incumbent ability in $supp\left(\mu\right)$, and assume that all incumbents cooperate with probability one at horizons larger than M_σ. The inequality $c\left(L_{\underline{k}+1}\right) - c\left(L_{\underline{k}}\right) < 1$ implies that $\mu\left(L_{\underline{k}}\right) < 1$ (otherwise $L_{\underline{k}+1}$ incumbents could outperform the incumbents). Observe that on the equilibrium path, everyone defects at the last $\underline{k}$ rounds (because, when the horizon is equal to $\underline{k}$, this event becomes common knowledge among the players), and, as a result, all incumbents $L_{\ge \underline{k}+1}$ defect at horizon $\underline{k} + 1$. Next, we note that early-efficiency implies that if any player defects on the equilibrium path, then both players defect in all the remaining stages (as it becomes common knowledge that the horizon is at most M_σ).

We finish the proof by dealing with three separate cases:

1. $p < \frac{A-1}{A}$ *and all incumbents cooperate on the equilibrium path when facing a stranger at a horizon larger than $\underline{k} + 1$.* The assumption that $p < \frac{A-1}{A}$ implies that all incumbents cooperate on the equilibrium path when the horizon is larger than $\underline{k} + 2$ (because the opponent is likely to be unobserving and to cooperate until horizon $\underline{k} + 1$). This implies that σ must be equivalent to a shifted variant of σ^*, in which abilities $L_{\underline{k}}$ and $L_{\underline{k}+2}$ co-exist and p cannot be too low. Finally, if $L_{\underline{k}} \geq L_2$, then perfection and early-niceness imply that mutants with ability L_1 outperform the $L_{\underline{k}}$ incumbents by inducing additional rounds of cooperation when their ability is observed.

2. $p < \frac{A-1}{A}$ *and some incumbents defect on the equilibrium path when facing a stranger with a horizon larger than $\underline{k} + 1$.* First, we show that all incumbents $L_{\geq \underline{k}+2}$ must defect with probability one at horizon $\underline{k} + 2$ (otherwise the strategy is not stable to a perturbation that slightly increases the probability of defection at horizon $\underline{k} + 2$). Next, we show that if $A > 3$ then it implies that $\mu\left(L_{\underline{k}+1}\right) > 0$ (otherwise σ is not stable to a perturbation that slightly increases $\mu\left(L_{\underline{k}}\right)$). Finally, we compare the payoffs of $L_{\underline{k}}$ and $L_{\underline{k}+1}$: ability $L_{\underline{k}+1}$ yields an additional utility point against $\left\{L_{\underline{k}}, L_{\underline{k}+1}\right\}$ and an additional fixed loss against higher abilities. This implies that abilities $L_{\underline{k}}$ and $L_{\underline{k}+1}$ obtain the same payoff iff $\mu\left(L_{\geq \underline{k}+2}\right)$ is equal to a specific value, but then strategy σ is not stable to a perturbation that changes the frequency of abilities $\left\{L_{\underline{k}}, L_{\underline{k}+1}\right\}$ while keeping $\mu\left(L_{\geq \underline{k}+2}\right)$ fixed.

3. $p > \frac{A-1}{A}$. Let $k > M_\sigma$, and let m be the largest horizon in which a player with ability L_k, who observes an opponent with the same ability, defects with a positive probability. Neutral stability implies that the defection will be with probability one (otherwise the strategy is not stable to a perturbation that slightly increases the frequency of players that defect at horizon m). The inequality $p > \frac{A-1}{A}$ implies that that $m = k$ (otherwise it would be strictly better to defect at horizon $m+1$), and this contradicts the early-niceness.

We conclude with a few remarks on Theorem 2:

1. Replacing "pavlov" with "perfect-grim-trigger" (defect iff any player defected before) at long horizons yields a strict limit ESS that is equivalent to σ^* (but not identical, as they differ in the off-equilibrium behavior.)

2. In principal, one could adapt the mechanisms that lead to early-niceness in either Fundenberg & Maskin (1990), Binmore & Samuelson (1992) or Kreps et al. (1982), incorporate them in our model, and obtain early-niceness as part of the uniqueness result (rather than as an assumption). We choose not to do this because it involves technical difficulties that would make the model substantially less tractable and less transparent.

3. Theorem 2 holds for any $p < 1$. If $p = 1$, then the game may admit a limit ESS with large abilities in its support. Specifically, for each k with sufficiently small $c(L_k)$, one can show that if a limit ESS exists, it must assign a positive frequency for abilities $L_{\geq k}$ (see a related analysis in Mohlin, 2012).

4. If one omits the condition $c(L_4) < \frac{1}{A}$, then the uniqueness result still essentially holds, except that a limit ESS may also be equivalent to σ_2^* - a shifted variant of σ^* that includes abilities L_2 and L_4 (see Definition 2).

5. If the condition $\forall k \in \mathbb{N}\ c(L_{k+1}) - c(L_k) < 1$ does not hold, then for sufficiently low p there are additional "single-ability" limit ESS. Specifically, if $c(L_{k+1}) - c(L_k) > 1$ and $p < \frac{1}{(A-1)\cdot(k-2)}$, then a strategy that includes only ability L_k is also a limit ESS.

7   Uniqueness with Weaker Solution Concepts

Theorem 2 shows that σ^* is essentially the unique early-nice limit ESS. In this section we study which aspects of the uniqueness hold for weaker solution concepts.

A strategy is a perfect NSS (symmetric perfect equilibrium) if it is the limit of NSS (symmetric Nash equilibria) of a converging sequence of ubiquitous perturbed games.

Definition 9. $\sigma \in \Sigma$ is a *perfect NSS* (*symmetric perfect equilibrium*) if there exists a sequence of ubiquitous perturbations $(\zeta_n)_{n \in \mathbb{N}}$ satisfying $\lim_{n \to \infty} M(\zeta_n) = 0$, and for each $n \in \mathbb{N}$, there exists an NSS (symmetric Nash equilibrium) σ_n of $\Gamma(\zeta_n)$, such that $\lim_{n \to \infty} \sigma_n = \sigma$.

Observe that: (1) any limit ESS is a perfect NSS (by Lemma 1); and (2) any perfect NSS is a symmetric perfect equilibrium.

The following two definitions are useful to present the results of this section. Strategy $\sigma_{\underline{k}}^*$ is a $\underline{k}$-shifted variant of σ^*, in which ability $L_{\underline{k}}$ replaces L_1 and ability $L_{\underline{k}+2}$ replaces L_2. Formally:

Definition 10. For each $\underline{k}$, let strategy $\sigma_{\underline{k}}^* = \left(\mu_{\underline{k}}^*, b_{\underline{k}}^*\right)$ be as follows:

$$\mu_{\underline{k}}^*\left(L_{\underline{k}}\right) = 1 - \frac{1 - \left(c\left(L_{\underline{k}+2}\right) - c\left(L_{\underline{k}}\right)\right)}{p \cdot (A-1)}, \quad \mu^*\left(L_{\underline{k}+2}\right) = 1 - \mu_{\underline{k}}^*\left(L_{\underline{k}}\right), \quad \forall k \notin \{\underline{k}, \underline{k}+2\}\ \mu^*\left(L_k\right) = 0.$$

$$b_{\underline{k}}^*\left(L, l.s, \left(a^i, a^{-i}\right)^t\right) = \begin{cases} C & \left[l \geq \underline{k} + 3 \ \text{ or } \ \left(l = \underline{k} + 2 \ \text{ and } s \in \left\{L_{\underline{k}}, \phi\right\}\right)\right] \\ & \text{and } \left(t = 0 \ \text{ or } a_t^i = a_t^{-i}\right) \\ D & \text{otherwise} \end{cases}.$$

The set $\Sigma_{\underline{k}}^* \subseteq \Sigma$ includes all the strategies that differ from $\sigma_{\underline{k}}^*$ only by "redistributing" frequency $\mu_{\underline{k}}^* \left(L_{\underline{k}+2} \right)$ among other abilities L_k that have the same cost and play the same as $L_{\underline{k}+2}$ (given playing rule b^*). Formally:

Definition 11. For each $\underline{k}$, let

$$\Sigma_{\underline{k}}^* = \left\{ (\mu, b^*) \, | \mu \left(L_{\underline{k}} \right) = \mu_{\underline{k}}^* \left(L_{\underline{k}} \right), \; \forall k \neq \underline{k} \; \mu \left(L_k \right) > 0 \, \text{only if} \, \left(L_k \geq L_{\underline{k}+2} \, \text{and} \, c \left(L_k \right) = c \left(L_{\underline{k}+2} \right) \right) \right\}.$$

The following result shows which aspects of the uniqueness results hold with the weaker solution concepts. Part (1) shows that, in any symmetric perfect equilibrium, the minimal incumbent ability is L_1 and the maximal ability is either L_3 or L_4.[20] Part (2) shows that any early-efficient NSS is similar to a k-shifted variant of σ^*.[21] Part (3) shows that the uniqueness result essentially holds for early-nice perfect NSS.

Proposition 4. *Assume that δ is close enough to 1, $\forall k \in \mathbb{N} \; c \left(L_{k+1} \right) - c \left(L_k \right) < 1$, $A > 3$, and $c \left(L_4 \right) < \frac{1}{A}$.*

1. *Assume $\frac{1}{(A-1)^2} < p$. If $\sigma = (\mu, \beta)$ is a an early-nice symmetric perfect equilibrium, then $0 < \mu \left(L_1 \right) < 1$. Moreover, if $\frac{A+1}{(A-1) \cdot (A-2)} < p < \frac{A-1}{A}$ and $c \left(L_5 \right) > c \left(L_4 \right)$ then $\mu \left(L_{\geq 5} \right) = 0$.*

2. *Assume $p \neq 0.5$ and $p < \frac{A-1}{A}$. If σ is a an early-nice NSS, then it is equivalent to a strategy in $\cup_k \Sigma_k^*$. Moreover, if $p < \frac{A \cdot (1 - c(L_3))}{(A-1)^2}$, then no early-nice NSS exist.*

3. *Assume $p \notin \left\{ \frac{1}{2}, 1 \right\}$. If σ is an early-nice perfect NSS, then $\sigma \approx \sigma'$ for some $\sigma' \in \Sigma_1^*$. Moreover, if $p < \frac{A \cdot (1 - c(L_3))}{(A-1)^2}$ or $\frac{A-1}{A} < p$, then no early-nice perfect NSS exist.*

8 Discussion

8.1 Limited foresight and Uncertain Length

In this section we deal with three related questions: (1) Why do we model the interaction as having uncertain length? (2) Could similar results be obtained in a model with a fixed length? and (3) Why do we interpret abilities in our model as representing limited foresight?

As argued by Osborne & Rubinstein (1994, chapter 8.2):

---

[20] The same result holds for the weaker notion of sequential equilibrium (Kreps & Wilson (1982)) and to a "0-perfect" equilibrium, in which the perturbations must assign minimal positive probabilities only at stage 0.

[21] The result is stated for $p \notin 0.5$ and $p < \frac{A-1}{A}$. See Part (3-f) and footnote 26 for an additional strategy that may be an early-nice *perfect NSS when $p = 0.5$. When $p > \frac{A-1}{A}$ we can show that: (1) for each M, if $c \left(L_M \right)$ is sufficiently small, then any early-nice NSS includes ability L_M in its support, and (2) if M_0 is sufficiently large and $c \left(L_{M_0} \right)$ is sufficiently small, then no early-nice NSS exists.

"A model should attempt to capture the features of reality that the players perceive. ... In a situation that is objectively finite, a key criterion that determines whether we should use a model with a finite or an infinite horizon is whether the last period enters explicitly into the players' strategic considerations."

Following this argument, we present a "hybrid" model in which the horizon is infinite and uncertain, until close to the end, in which the final period reaches the agent's foresight ability.

Next we show that similar results can be obtained if the game has a fixed length. Consider a repeated Prisoner's Dilemma with a fixed length L. Agents with limited foresight in this setup must be unable to "count" how many rounds remain in the game. This can be formalized by restricting agents to strategies that depend only on the actions observed at the last m rounds, or to strategies that can be implemented by automota with a limited number of states. With such a restriction, one can adapt our main results (Theorems 1-2) to this setup.

Finally, we discuss the interpretation of limited foresight in our model, and compare it with an alternative "myopic" notion that is used in the existing literature (e.g., Jehiel, 2001). In our model, an agent with ability L_k perceives repeated interactions as infinite (if the horizon is larger than k). The alternative notion implies that such an agent perceives long repeated interactions as having, at most, k rounds. The comparison between the two notions can be facilitated by considering a long 2-player zero-sum game such as Chess. In this setup, agents with limited foresight (such as computer programs) base their play on a bounded minimax algorithm that looks a limited number of steps ahead and uses a heuristic evaluation function to assign values to the states at the horizon.

When moving from chess-like games to non-zero-sum repeated games, the key question is how L_k agents evaluate states that occur k steps ahead. The "myopic" notion assigns a constant value to all non-final states. Contrary to this, our notion bases the evaluation of non-final states on the history of play and its influence on the future play of his opponent. In particular, consider an L_1 agent who plays against an opponent who follows a "perfect-grim-trigger". The "myopic" notion implies the counter-intuitive prediction that the L_1 agent always defects. Our notion implies that he cooperates until the last stage. As described in Footnote 8, the experimental evidence from finitely repeated Prisoner's Dilemma games suggests that subject behave in a way which is consistent with our notion of limited foresight.

8.2   Extensions and Variants

We conclude by presenting a few extensions and variants of our model.

In the basic model we followed two common assumptions in the "indirect evolutionary literature" (see, e.g., Dekel *et al.*, 2007): (1) an agent can observe his opponent's ability with a

fixed exogenous probability; and (2) an agent cannot send a false signal about his ability. These assumptions may seem too restrictive. Completely relaxing them, by allowing each player to choose at stage 0 both an unobservable true ability and a "fake" ability that is observed by the opponent (a "cheap-talk" model), induces a unique behavior in any Nash equilibrium: everyone defects at all stages.[22] In Appendix A.1 we sketch a variant of the model, which partially relaxes these assumptions: each player chooses at stage 0 a true ability, a fake ability, and an effort level, and the probability in which a player observes the true ability (rather than the fake ability) of the opponent is increasing in the player's effort and decreasing in the opponent's effort. We show that a σ^*-like strategy remains stable in this setup.

Our basic model deals only with the repeated Prisoner's Dilemma, and assume that the cognitive cost function is increasing. In Appendix A.2, we sketch how to extend the results to other games (assuming that the repeated Prisoner's Dilemma is played with sufficiently high probability), and to non-monotonic cost functions (which may represent the advantages of having higher abilities in other games).

Next we show that our results are robust to various changes in the set of abilities. First, we consider the case in which the minimal ability in $\mathcal{L}$ is not L_1, but any other arbitrary ability $L_{\tilde{k}}$ (including ability "L_0", which is never informed about the realized length of the interaction). It is straightforward to see that all of our results hold in this setup except that σ^* is replaced with its shifted variant $\sigma_{\tilde{k}}^*$ (Definition 10), in which, ability $L_{\tilde{k}}$ ($L_{\tilde{k}+2}$) replaces ability L_1 (L_3). Next, we observe that our results hold also if the set of abilities $\mathcal{L}$ is extended to include ability L_∞, which is informed about the final period at the end of round 0.

The next variant introduces a maximal ability by restricting the set of abilities to be $\{L_1, ..., L_M\}$. Assuming that $M \geq 3$, Theorem 1 holds in this setup. Theorem 2 holds for ps which are not too close to either zero or one . Assuming that $c(L_M)$ is sufficiently low, one can complete the characterization for all values of p: (1) for low ps: if a limit ESS exists, then the only ability in its support is L_M (because the indirect "commitment" advantage of lower abilities is too small); and (2) for high ps ($p > \frac{A-1}{A}$): if a limit ESS exists, then its support includes ability L_M (as a result of the "arm race" for earlier defections and higher abilities).

Finally, we note that the main results (Theorem 1-2) hold for each of the following changes to the observation of the opponent's ability:

1. "Late observability": Players observe the opponent's ability later in the game (and not at

---

[22] To simplify the argument assume that $c(L_k) \equiv 0$ (the argument can be extended to positive and sufficiently small cognitive costs). Assume to the contrary that players cooperate with positive probability on the equilibrium path. Let m be the smallest horizon in which they cooperate with positive probability. Then the following strategy is a strictly-better reply: choosing an arbitrary large enough true ability, signaling one of the fake abilities of the incumbents, playing like the incumbents at horizons larger than m, and defecting during the last m stages.

the end of stage zero). For example, the results hold if a player with ability L_k obtains the signal about his opponent's ability at horizon k (when he becomes aware to the timing of the final period), or at horizon $\min(k, k' + 1)$ (i.e., a player only observes if his opponent is going to be informed about the final period at the next round).

2. Asymmetric observability (à la Mohlin, 2012): the informative signal (obtained with probability p) is the opponent's exact ability, only if it is strictly lower than the agent's ability; if the opponent's ability is weakly higher, then the agent only observes this fact.

3. Perturbed signals: having a weak correlation between signals of the two players.

A   Variants and Extensions - Formal Details

In this appendix we formally state the definitions and the results of two extensions of our model. The formal proofs, which are relatively simple adaptations to the proofs of the results in the basic model, are omitted for brevity.

A.1   False Signals and Endogenous Observability

In this section, we sketch a variant of the model in which players can influence the probability of observing the opponent's ability. A comprehensive analysis of this variant (with a general underlying game) is left for a separate paper (Heller & Mohlin, 2013).

At stage 0, each player i makes 3 choices: (1) *true ability* - $L^i \in \mathcal{L}$, (2) *fake ability* - $s^i \in \mathcal{L}$, and (3) effort level $e^i \in \mathbb{R}^+$, which costs e^i utility points.[23] The model also specifies an observation function $p : \mathbb{R}^+ \times \mathbb{R}^+ \to [0, 1]$. When a player who invests effort e^1 faces an opponent who invests effort e^2, he privately observes his opponent's true ability with probability $p(e^1, e^2)$ and observes the fake ability otherwise. We assume that $p(e^1, e^2)$ is (1) increasing and concave in the first parameter, (2) decreasing and convex in the second parameter, and (3) sub-modular: $\frac{\partial^2 p(e^1, e^2)}{\partial e^1 \partial e^2} < 0$ (i.e., the efforts of the two players are strategic substitutes).

A strategy in this setup is a pair $\sigma = (\mu, \beta)$ where $\mu \in \Delta(\mathcal{L} \times \mathcal{L} \times \mathbb{R}^+)$ is a distribution over the pure choices at stage 0 (true ability, fake ability and effort level).

Theorem 1 is extended to this setup as follows:

Proposition 5. *Assume that: (1)* $\exists e_0 < \frac{1}{A} - c(L_3)$ *s.t.* $\forall e \leq e_0$, $\frac{A \cdot (1 - c(L_3))}{(A-1)^2} < p(e, e) < \frac{A-1}{A}$

$$and \;\; \left. \frac{\partial p(e^1, e^2)}{\partial e^1} \right|_{(e^1, e^2) = (e_0, e_0)} - A \cdot \left. \frac{\partial p(e^1, e^2)}{\partial e^2} \right|_{(e^1, e^2) = (e_0, e_0)} < 1,$$

---

[23] Similar results also hold if each player chooses two different efforts: one for lying, and one for detecting lies.

(2) $c(L_4) > c(L_3) < \frac{1}{A}$, and (3) δ is sufficiently close to 1. Then there exists $0 < e^* < e_0$ such that $\sigma^*(e^*) = (\mu^*(e^*), b^*)$ is a limit ESS, where b^* is as Def. 2 and $\mu^*(e^*)$ is as follows:

$$supp(\mu^*) = \{(L_1, L_1, 0), (L_3, L_1, e^*)\}, \text{ and } \mu^*((L_3, L_1, e^*)) = \frac{1 - c(L_3) - e^*}{p \cdot (A - 1)}.$$

The stable strategy $\sigma^*(e^*)$ has two types in its support: (1) agents with ability L_1 who do not expend any effort; and (2) agents with ability L_3 who expend effort e^*(which is determined by the observability function) and try to deceive their opponent into thinking that they have ability L_1. Agents behave in the same way as in the basic model. In what follows, we briefly explain the first assumption (the remaining two assumptions are identical to Theorem 1), and the intuition as to why it implies the stability of $\sigma^*(e^*)$. Assumption (1) requires the existence of an effort level e_0 that satisfies three requirements. (I) e_0 is not too large. Observe that if $e < \frac{1}{A} - c(L_3)$, then the total cost of an agent with ability L_3 who invests effort e is smaller than $\frac{1}{A}$, and is outweighed by its gain from defecting one stage earlier in a population that includes a large enough fraction of agents with ability L_1. The next requirement, (II), is that $p(e, e)$ is not too close to zero or one (the same bounds as in Theorem 1) for any $e \leq e_0$: this implies that the induced observation probability when two L_3 agents meet each other (and each spends effort level e^* on the equilibrium path) is far enough from zero and one, which is required for stability from the same reasons as in the basic model. And the third requirement, (III), is that the marginal contribution of effort at e_0 (which is the sum of the marginal contributions induced by increasing the probability to observe the opponent's ability and by decreasing the probability that the opponent observes the agent's own type) is smaller then its marginal cost (=1): this condition implies (by convexity and sub-modularity) that there exists a stable effort level $e^* < e_0$.

We conjecture that one could also adapt Theorem 2 to this setup.

A.2 Other Games and Non-Monotonic Costs

Consider the following environment with other games: At the end of stage 0 (after both players have chosen their abilities), Nature randomly chooses whether to play the repeated Prisoner's Dilemma (with probability q) or an arbitrary extensive-form game (with probability $1 - q$)[24] and both players observe the realized game. A strategy in this setup is a triple (μ, β, γ), where the new component, γ, describes the playing rule in the other game. For each (μ, γ) define $v_{\mu,\gamma} : \mathcal{L} \to \mathbb{R}$ as the expected payoff of ability L_k conditional on playing the other game, when

---

[24] There is no loss of generality in assuming that what is played with probability $1 - q$ is a single game.

the opponent follows (μ, γ). Define the *net cost function* $c_{\mu,\gamma,q} : \mathcal{L} \to \mathbb{R}$ as follows:

$$c_{\mu,\gamma,q}(L_k) = (c(L_k) - c(L_1)) - (1-q) \cdot (v_{\mu,\gamma}(L_k) - v_{\mu,\gamma}(L_1)).$$

In words, $c_{\mu,\gamma,q}$ is the total external cost of choosing ability L_k (relative to L_1): the cognitive cost minus the additional gain that is achieved in the other game. We can extend the basic model by replacing the fixed cognitive cost function c with the family of net cost functions $c_{\mu,\gamma,q}$. This reduced form model captures other games without a loss of generality: any arbitrary environment with additional games is equivalent to a family of net cost functions.

We shall say that (μ, β, γ) is early efficient if (μ, β) is early efficient. Define $\Delta^*_{\gamma,q} \in \Delta(\mathcal{L})$ as the set of distributions μ^*_γ that satisfy the following conditions:

$$\mu^*_{\gamma,q}(L_1) = 1 - \frac{1 - c_{\mu^*_{\gamma,q}}(L_3)}{p \cdot (A-1)}, \quad \mu^*(L_3) = 1 - \mu^*_{\gamma,q}(L_1), \quad \forall k \notin \{1,3\} \ \mu^*_{\gamma,q}(L_k) = 0.$$

That is, $\mu^*_{\gamma,q}$ is similar to μ^* of the standard model, except that the cognitive cost of L_3 is replaced with its net cost. Theorem 2 is extended to this setup as follows:

Proposition 6. *Assume that $A > 3$, $c(L_3) < c(L_4) < \frac{1}{A}$, $p < 1$, $\forall k \in \mathbb{N} \ c(L_{k+1}) - c(L_k) < 1$, and that δ is sufficiently close to 1. Then there exists $q_0 < 1$, such that for each $q_0 \le q \le 1$ and each playing-rule γ, if strategy (μ, β, γ) is a an early-nice limit ESS, then $\mu \in \Delta^*_{\gamma,q}$ and $\beta \approx \beta^*$.*

Finally, we consider a simpler variant that keeps the structure of a single cost function $c : \mathcal{L} \to \mathbb{R}$, but relaxes the assumption that it is weakly increasing ($c(L_1)$ is still normalized to be 0). This non-monotonicity is interpreted as representing external advantages of higher abilities in other games. To simplify the presentation of the results assume that there exists a unique ability that minimizes $c(L_k)$ among all abilities in $L_{\ge 3}$, and let it be denoted by $L_{\hat{k}} = \operatorname{argmin}_{L_k \ge L_3}(c(L_k))$. Theorem 1 is extended to this setup with ability $L_{\hat{k}}$ replacing L_3. Formally:

Definition 12. For each $1 - p \cdot (A-1) < c(L_{\hat{k}}) < 1$, let $\hat{\sigma} = (\hat{\mu}, \beta^*)$ where $\hat{\mu}$ is as follows:

$$\hat{\mu}(L_{\hat{k}}) = \frac{1 - c(L_{\hat{k}})}{p \cdot (A-1)}, \quad \hat{\mu}(L_1) = 1 - \hat{\mu}(L_{\hat{k}}), \quad \forall k \notin \{1, \hat{k}\} \ \hat{\mu}(L_k) = 0;$$

Proposition 7. *Assume that: $\frac{A \cdot (1 - c(L_{\hat{k}}))}{(A-1)^2} < p < \frac{A-1}{A}$, $c(L_{\hat{k}}) < \frac{1}{A}$, $-c(L_{\hat{k}}) < p \cdot (A-1) - 1$, and δ is sufficiently close to 1. Then $\hat{\sigma}$ is a strict limit ESS.*

Similarly, Theorem 2 can be extended to this setup in a similar way (if the cognitive costs are sufficiently small).

B Proofs

B.1 Limit ESS and Ubiquitous Perturbations

The following lemma shows that if σ_n is an ESS of a perturbed game of the repeated Prisoner's Dilemma, then it is also an ESS of a nearby ubiquitous perturbed game.

Lemma 1. *Let ζ be a perturbation. Let $\sigma \in \Sigma$ be an ESS of the perturbed game $\Gamma(\zeta)$. Then for every $\epsilon > 0$, there exists an ubiquitous perturbation ζ' such that: (1) $|\zeta - \zeta'| < \epsilon$; (2) $\sigma' \in \Sigma$ is an ESS of the perturbed game $\Gamma(\zeta')$; and (3) $|\sigma' - \sigma| < \epsilon$.*

Proof. The fact that σ is an ESS implies that it must assign a positive probability to each information set (otherwise, an equivalent strategy σ' that differs only in information sets that are reached with zero probability would get the same payoff as σ: $u(\sigma, \sigma) = u(\sigma', \sigma)$ and $u(\sigma, \sigma') = u(\sigma', \sigma')$). This implies that σ must assign a positive probability for each ability and for each action at each information set in which the horizon is larger than 1. When the horizon is equal to 1, defection is a dominant action. Let $\epsilon > 0$ be sufficiently small. Define a ubiquitous perturbation ζ' as follows: (1) if $\zeta(I)(a) > 0$, let $\zeta'(I)(a) = \zeta(I)(a)$; (2) if $\zeta(I)(a) = 0$ and the horizon is larger than 1, let $\zeta'(I)(a) = \min(\epsilon, \sigma(a))$ (which is a positive number due to the previous argument); and (3) when the horizon is equal to one let $\zeta(I)(a) = \epsilon$. Let σ' be equal to σ except at horizon 1, in which it defects with probability $1 - \epsilon$. The above arguments imply that σ' is an ESS in $\Gamma(\zeta')$. $\qquad\square$

An immediate corollary of Lemma 1 is that every limit ESS is the limit of ESS of a sequence of ubiquitous perturbed games.

Corollary 1. *Let $\sigma \in \Sigma$ be a limit ESS. There exists a sequence of ubiquitous perturbations $(\zeta_n)_{n \in \mathbb{N}}$ satisfying $\lim_{n \to \infty} M(\zeta_n) = 0$, and for each $n \in \mathbb{N}$, there exists an ESS σ_n of the perturbed game $\Gamma(\zeta_n)$, such that $\lim_{n \to \infty} \sigma_n = \sigma$ is satisfied.*

Proof. The fact that σ is a limit ESS implies that there exists a sequence of perturbations $(\zeta_n)_{n \in \mathbb{N}}$ satisfying: $\lim_{n \to \infty} M(\zeta_n) = 0$, and for each $n \in \mathbb{N}$, there exists a strategy $\sigma_n \in \Sigma(\zeta_n)$, which is an ESS of $\Gamma(\zeta_n)$, and that $\lim_{n \to \infty} \sigma_n = \sigma$ is satisfied. Lemma 1 implies that there exists a sequence of ubiquitous perturbations $(\zeta'_n)_{n \in \mathbb{N}}$ with the same properties. $\qquad\square$

Remark 4. The corollary immediately implies that every limit ESS is a perfect NSS (Def. 9) and a symmetric perfect equilibrium (Selten, 1975). The proof of Lemma 1 relies on the property of the repeated Prisoner's Dilemma that each player has a dominant action at the last stage. Slightly weaker results are known for general extensive-form games: any limit ESS is a symmetric sequential equilibrium (van Damme, 1987, Corollary 9.8.6).

Tab. 2: Reduced Game (Players Choose abilities, and must follow Playing-Rule b^*)

	L_1	$L_{k\,(k\geq 3)}$	L_2
L_1	$2 \cdot A$	A	A
$L_{k\,(k\geq 3)}$	$2 \cdot A + 1 - c\left(L_k\right)$	$A + 1 - p \cdot A + p - c\left(L_k\right)$	$A + 1 + p - c\left(L_k\right)$
L_2	$2 \cdot A + 1 - c\left(L_2\right)$	$A + 1 - p \cdot A - c\left(L_2\right)$	$A + 1 - c\left(L_2\right)$

B.2 Theorem 1 - σ^* is a Strict Limit NSS / ESS

Proof. The proof includes several parts:

1. *Abilities L_1 and L_3 are best-replies given playing-rule b^*: $u\left(\left(L_1, b^*\right), \sigma^*\right) = u\left(\left(L_3, b^*\right), \sigma^*\right)$ $\geq u\left(\left(L_k, b^*\right), \sigma^*\right)$ for each $k \notin \{1, 3\}$ with strict inequality if: $c\left(L_4\right) > c\left(L_3\right)$.*

   (a) *Reduced game given b^*.* Playing-rule b^* induces a reduced normal form game in which each player chooses ability at stage 0, and then players follow b^* at the remaining rounds. Note that the choice of ability only influences the payoffs at stages 0 (cognitive cost), $\mathbf{T} - 1$ (=horizon 2) and $\mathbf{T} - 2$ (=horizon 3), as all abilities play the same at all other stages (they all play *pavlov* until stage $\mathbf{T} - \mathbf{3}$ and defect at stage $\mathbf{T}$). Henceforth, we focus only on the payoffs of these 3 stages, and present in Table 2 the symmetric payoff matrix of this reduced game. The payoffs of Table 2 are calculated as follows: Two players with ability L_1 who face each other cooperate at horizons 2 and 3, and obtain $2 \cdot A$ utility points. A player with ability L_k ($k \geq 2$), who faces L_1, defects at horizon 2 and obtains $2 \cdot A + 1$ points (and induces a cognitive cost), while the L_1 opponent obtains only A points. When two L_2-s face each other they both cooperate at horizon 3 and both defect at horizon 2, and they obtain $A + 1$ points. When an L_3 player faces an L_2, the outcome depends on whether or not L_3 is observing. With probability p, the L_3 player is observing and he obtains $A + 2$ points (by defecting at both horizons) and the L_2 opponent obtains one point; with probability $1 - p$, L_3 is not observing and both players obtain $A + 1$ points (both cooperate at horizon 3 and defect at horizon 2). Finally, when two L_3s face each other, the outcome depends on both observations. If both players are observing (probability p^2), they defect at both horizons and obtain 2. If both are unobserving (probability $(1 - p)^2$), they defect only at horizon 2 and obtain $A + 1$. If exactly one of them is observing, the observing player defects at horizon 3, and he obtains $A + 2$ and his opponent obtains one. Aggregating these possible outcomes yields the

following expected payoff at horizons 2 and 3:

$$p^2 \cdot 2 + (1 - p)^2 \cdot (A + 1) + p \cdot (1 - p) \cdot (A + 2) + (1 - p) \cdot p \cdot 1 = A + 1 - p \cdot A + p.$$

(b) Abilities $L_{>3}$ are weakly dominated by ability L_3 and strictly dominated if $c(L_4) > c(L_3)$: they obtain the same stage-payoffs but they bear higher cognitive costs.

(c) *Ability L_2 obtains a strictly lower payoff then ability L_1.*
    We have to show that the payoff of ability L_2 $((2 \cdot A + 1) \cdot \mu^*(L_1) + (A + 1 - p \cdot A) \cdot \mu^*(L_3) - c(L_2))$ is strictly smaller than the payoff of ability L_3 $((2 \cdot A + 1) \cdot \mu^*(L_1) + (A + 1 - p \cdot A + p) \cdot \mu^*(L_3) - c(L_3))$. This holds iff:

$$(A + 1 - p \cdot A) \cdot \mu^*(L_3) - c(L_2) \overset{?}{<} (A + 1 - p \cdot A + p) \cdot \mu^*(L_3) - c(L_3) \Leftrightarrow$$

$$c(L_3) - c(L_2) \overset{?}{<} p \cdot \mu^*(L_3) = \frac{1 - c(L_3)}{A - 1} \Leftrightarrow$$

$$c(L_3) \overset{?}{<} \frac{c(L_2) \cdot (A - 1) + 1}{A} = \frac{1}{A} + \left(1 - \frac{1}{A}\right) \cdot c(L_2),$$

and the latter inequality is implied by $c(L_3) < \frac{1}{A}$.

(d) *Frequency μ^* balances the payoffs between abilities L_1 and L_3.*
    Observe that if $A + 1 - p \cdot A + p - c(L_k) < A$, then the reduced game between these two abilities is of the Hawk-Dove variety: each ability is a strict best-reply to the other ability. This inequality holds iff:

$$A + 1 - p \cdot A + p - c(L_3) < A \Leftrightarrow \frac{1 - c(L_3)}{A - 1} < p.$$

The latter inequality holds due to the assumption that $\frac{A \cdot (1 - c(L_3))}{(A - 1)^2} < p$. It is well known that a Hawk-Dove game admits a unique mixed equilibrium. We now show that μ^* is the unique equilibrium. The payoff of L_1 is $2 \cdot A \cdot \mu^*(L_1) + A \cdot \mu^*(L_3)$, and the payoff of L_3 is $(2 \cdot A + 1) \cdot \mu^*(L_1) + (A + 1 - p \cdot A + p) \cdot \mu^*(L_3) - c(L_3)$. These payoffs are equal if:

$$2 \cdot A \cdot \mu^*(L_1) + A \cdot \mu^*(L_3) \overset{?}{=} (2 \cdot A + 1) \cdot \mu^*(L_1) + (A + 1 - p \cdot A + p) \cdot \mu^*(L_3) - c(L_3) \Leftrightarrow$$

$$(p \cdot A - p - 1) \cdot \mu^*(L_3) + c(L_3) \overset{?}{=} \mu^*(L_1) = 1 - \mu^*(L_3) \Leftrightarrow$$

$$p \cdot (A - 1) \cdot \mu^*(L_3) + c(L_3) \overset{?}{=} 1 \Leftrightarrow \mu^*(L_3) = \frac{1 - c(L_3)}{p \cdot (A - 1)}.$$

2. *Stability against other distributions*: If $(\mu, b^*) \neq \sigma^*$ is a best-reply to σ^* then $u(\sigma^*, (\mu, b^*)) \geq$

$u\left(\left(\mu, b^*\right),\left(\mu, b^*\right)\right)$ *with a strict equality if* $c\left(L_4\right) > c\left(L_3\right)$.

Part (1) implies that $\left(\mu, b^*\right)$ is a best-reply to σ^* iff $supp\left(\mu\right) \subseteq \{L_1, L_3\}$. The result is an immediate corollary of the well-known result (see, e.g., Weibull, 1995, Section 2.1.2) that the unique equilibrium in a Hawk-Dove game is an ESS. The intuition for this result is the observation that, if the frequency of ability L_1 becomes larger (smaller), they become relatively less (more) successful than ability L_3.

3. *The perturbed game* $\Gamma\left(\zeta\right)$: Let ζ be any ubiquitous perturbation with sufficiently small maximal tremble $M\left(\zeta\right)$. Let $\sigma_\zeta^* = \left(\mu_\zeta^*, \beta_\zeta^*\right)$ be defined as follows:

$$\beta_\zeta^*\left(I\right)\left(C\right) = \begin{cases} \zeta\left(I\right)\left(C\right) & b^*\left(I\right) = D \\ 1 - \zeta\left(I\right)\left(D\right) & b^*\left(I\right) = C \end{cases}, \quad \mu_\zeta^*\left(L_k\right) = \begin{cases} \mu^*\left(L_1\right) + \eta & k = 1 \\ \zeta\left(L_k\right) & k \neq 1, 3 \\ 1 - \sum_{k \neq 3} \mu_\zeta^*\left(L_k\right) & k = 3 \end{cases},$$

where η is chosen such that $u\left(\left(L_1, \beta_\zeta^*\right), \sigma_\zeta^*\right) = u\left(\left(L_3, \beta_\zeta^*\right), \sigma_\zeta^*\right)$. Observe that:

   (a) For sufficiently small $M\left(\zeta\right)$, such η exists and its order of magnitude is $O\left(M\left(\zeta\right)\right)$ (due to continuity and the properties of the reduced game described above). This implies that $\left|\sigma_\zeta^* - \sigma^*\right| \leq O\left(M\left(\zeta\right)\right)$.

   (b) β_ζ^* is the closest playing rule to b^* in $\Sigma\left(\zeta\right)$.

   (c) The perturbed reduced game between abilities, in which the fixed playing-rule is β_ζ^*, is still a game of Hawk-Dove variety (again, by continuity). Thus the results of parts (1)-(2) still hold in this setup.

   (d) All information sets are reached with positive probability in $\Gamma\left(\zeta\right)$.

4. *For every* $\mu \in \Delta_\zeta\left(\mathcal{L}\right)$ *and* $\beta \in B\left(\zeta\right)$, $u\left(\left(\mu, \beta_\zeta^*\right), \sigma_\zeta^*\right) > u\left(\left(\mu, \beta\right), \sigma_\zeta^*\right)$ *(i.e.,* β_ζ^* *is a strictly optimal playing-rule against strategy* σ_ζ^* *in* $\Gamma\left(\zeta\right)$ *for all abilities).*

   (a) β_ζ^* *is strictly optimal for uninformed agents.* Recall that as long as players are uninformed, playing-rule b^* is equal to *pavlov*, and that β_ζ^* is the closest strategy to b^* in $B\left(\zeta\right)$. Lorberbaum et al. (2002) study the standard repeated Prisoner's Dilemma, in which players remain uninformed throughout the game. They analyze a perturbation that assigns minimal probability $\epsilon > 0$ for each action at each information set. They show that the ϵ-*perturbed pavlov* (the strategy that defects with probability $1 - \epsilon$ if the players played different actions at the previous round, and cooperates with probability $1 - \epsilon$ otherwise) is a symmetric strict equilibrium (and hence, also an ESS) in the ϵ-*perturbed game*. Minor adaptations to their proof (omitted for brevity)

extend the result (for δ sufficiently close to 1): (1) for any ubiquitous perturbation; and (2) for the current setup, in which players are informed in the last few rounds.

(b) *β_ζ^* is strictly optimal for horizons 1 and 2.* Defection is a dominant action for horizon 1. The fact that σ_ζ^* induces a very high probability of defection at horizon 1 (regardless of the history), implies that defecting at horizon 2 is a strict best-reply.

(c) *Horizon 3 against L_3.* Defection at horizon 3 yields one more point immediately (relative to cooperation), while it does not affect future payoffs (because, with high probability, the opponent defects during the last two rounds regardless of the history).

(d) *Horizon 3 against L_1 and strangers.* If players played different actions in the previous round, then defection yields both a higher payoff in the current stage, and a higher expected payoff in the future (as the opponent is likely to defect in the current stage, and only mutual defection may lead to mutual cooperation at the next round). This argument works also in larger horizons, and in steps (e)-(f) below we focus on showing that β_ζ^* is optimal only after a previous round in which both players played the same.

If the players played the same action in the previous round, and the opponent is L_1, then cooperation yields (with high probability)[25] payoff vector of $A, A+1, 1$: A at horizon 3, $A+1$ at horizon 2 (as the L_1 opponent cooperates), and one at horizon 1. Defection at horizon 3 yields a vector payoff of at most $A+1, 1, 1$ (as the L_1 opponent defects at horizon 2). Thus cooperation yields $A-1$ more utility points.

We are left with showing that β_ζ^* yields a strictly better payoff against strangers. If the stranger has ability L_3, then defection yields one more utility point than cooperation at horizon 3, and payoffs during the last two rounds remain the same. Thus, defection yields a higher expected payoff against a stranger iff the frequency of L_3 opponents is sufficiently low: $\mu^*(L_3) \cdot 1 < (1 - \mu^*(L_3)) \cdot (A-1) \Leftrightarrow$

$$\mu^*(L_3) < \frac{A-1}{A} \Leftrightarrow \frac{1 - c(L_3)}{p \cdot (A-1)} < \frac{A-1}{A} \Leftrightarrow \frac{A \cdot (1 - c(L_3))}{(A-1)^2} < p.$$

(e) *Horizon 4 against L_3.* Cooperation at horizon 4 (assuming both players played the same at the previous stage) yields a payoff vector of $A, 1, 1, 1$ $(A, A+1, 1, 1)$ during the last four rounds when facing an observing (unobserving) opponent. Defection at horizon 4 yields a payoff of $A+1, 1, 1, 1$ in both cases. Thus, cooperation is a strict

---

[25] Henceforth in the analysis we present strict inequalities by using the payoffs that are induced by the unperturbed strategy σ^*, which approximates the payoffs that are induced by σ_ζ^*. For sufficiently small $M(\zeta)$, the inequalities hold also for the slightly-perturbed σ_ζ^*. For brevity, we also omit the phrasing "with high probability" in the remaining text.

best reply iff: $1 \cdot p < (1-p) \cdot (A-1) \Leftrightarrow p < \frac{A-1}{A}$.

(f) *Horizon 4 against strangers and L_1, and horizons larger than 4 against all opponents.*
Cooperation is a strict best reply (assuming both players played the same at the previous stage) because it yields one less utility point in the current stage (relative to defection), and $A-1$ more points in the next round.

5. Combining the above arguments implies σ_ζ^* is an NSS in $\Gamma(\zeta)$ and an ESS if $c(L_4) > c(L_3)$. This implies that σ^* is a strict limit NSS, and a strict limit ESS if $c(L_4) > c(L_3)$.

$\square$

B.3 Theorem 2: Uniqueness Result

Proof. We begin with a few notations. Let $\sigma = (\mu, \beta)$ be an early-nice limit ESS. Let $M_\sigma \in \mathbb{N}$ be a large enough integer such that with probability one everyone cooperates at any horizon larger than M_σ. We shall say that a player faces an incumbent (at a given information set) if he has observed the opponent to have an incumbent ability or if he faces a stranger (as with probability one strangers have incumbent abilities). Let $L_{\underline{k}} \in supp(\mu)$ be the lowest incumbent ability. Recall that an information set $I \in \mathcal{I}$ *cooperative* if both players have cooperated at all previous stages. We shall say "ability L_k does X" as an abbreviation for "playing rule β induces a player with ability L_k to do X". The proof includes the following parts:

1. Preliminary observations about strategy σ:

   (a) *On the equilibrium path, everyone defects in the last $\underline{k}$ rounds.*
   Intuitively, this is because it is common knowledge among the players whether or not the horizon is at most $\underline{k}$. The formal argument is as follows: Assume to the contrary that players cooperate with positive probability in the last $\underline{k}$ rounds on the equilibrium path. Let $m \le \underline{k}$ be the smallest horizon in which a player cooperates with positive probability on the equilibrium path. Consider a strategy σ' that coincides with σ, except that players defect at horizon m with probability one. Observe that $u(\sigma', \sigma) > u(\sigma, \sigma)$ as both strategies induce the same play and yield the same payoff against σ at all rounds except at horizon m, in which strategy σ' defects with probability one and yields a higher payoff.

   (b) With probability one, players with ability $L_{\ge \underline{k}+1}$ defect at horizon $\underline{k}+1$ when facing an incumbent. This is because defection at horizon $\underline{k}+1$ yields one more utility point without affecting the opponent's future play (due to the previous step). Similarly, this implies that players with ability $L_{\ge \underline{k}+2}$ defect with probability one at horizon $\underline{k}+2$ when facing an observed incumbent ability $L_{\ge \underline{k}+1}$.

(c) Early-niceness implies that uninformed players cooperate with probability one at cooperative information sets (because the unknown horizon has a positive probability to be larger than M_σ). This is also true if the player has a non-incumbent ability.

(d) *If any incumbent ability defects with positive probability when facing an incumbent at a cooperative information set, and the defection is realized in the game, then both players defect at all the remaining periods.*
The claim is implied by the observation that after such a defection it becomes common knowledge that the maximal horizon is M_σ. The proof is analogous to step (a), and it is omitted for brevity.

(e) $\mu\left(L_{\underline{k}}\right) < 1$. The assumption that $c\left(L_{\underline{k}+1}\right) - c\left(L_k\right) < 1$ implies that if $\mu\left(L_{\underline{k}}\right) = 1$, then any strategy σ' that assigns mass one to $L_{\underline{k}+1}$, cooperates when being uninformed and defects at the last $\underline{k} + 1$ stages, is a strictly better reply against σ.

2. Case I: Assume that $p < \frac{A-1}{A}$ and all incumbents cooperate when: (1) the opponent is a stranger, (2) the information set is cooperative, and (3) the horizon is strictly larger than $\underline{k} + 1$. Then:

(a) *All incumbents cooperate when: (1) the opponent is an incumbent, (2) the information set is cooperative, and (3) the horizon is strictly larger than $\underline{k} + 2$.*
The previous part implies that defection at horizon $\underline{k} + 2$ $(> \underline{k} + 2)$ yields at least $A - 1$ $(2 \cdot (A - 1))$ less points than cooperation against an unobserving opponent (probability $1 - p$). If the opponent is observing (probability p) the maximal gain from defection is one point (two points), which is obtained if the opponent were planning to defect at horizon $\underline{k} + 2$ (at the next round also after mutual cooperation at the current stage). Defection yields a strictly lower payoff if:

$$(1 - p) \cdot (A - 1) > p \cdot 1 \iff (A - 1) > A \cdot p \iff \frac{A-1}{A} > p.$$

(b) The previous step implies that all incumbents obtain the same payoff at all horizons except $\underline{k} + 1$ and $\underline{k} + 2$, and that the reduced game between the abilities at these horizons is analogous to Table 2 (where $L_{\underline{k}}$ replaces L_1). As a result: (1) $\mu\left(L_{>\underline{k}+1}\right) > 0$ (otherwise, $u\left(\left(L_{\underline{k}}, \beta\right), \sigma\right) < u\left(\left(L_{\underline{k}+1}, \beta\right), \sigma\right)$ because $c\left(L_{\underline{k}+1}\right) - c\left(L_{\underline{k}}\right) < 1$); and (2) $\forall k > \underline{k} + 2$, $\mu\left(L_{>\underline{k}+2}\right) > 0$ only if $c\left(L_k\right) = c\left(L_{\underline{k}+2}\right)$ (otherwise $u\left(\left(L_k, \beta\right), \sigma\right) < u\left(\left(L_{\underline{k}+2}, \beta\right), \sigma\right)$, and σ cannot be an equilibrium).

(c) $\mu\left(L_{\underline{k}+1}\right) = 0$. Assume to the contrary that $\mu\left(L_{\underline{k}+1}\right) > 0$. The fact that σ is an equilibrium implies that $u\left(\left(L_{\underline{k}}, \beta\right), \sigma\right) = u\left(\left(L_{\underline{k}+1}, \beta\right), \sigma\right) = u\left(\left(L_{\underline{k}+2}, \beta\right), \sigma\right)$.

Analogous calculations to part (1-c-d) of Theorem 1's proof imply that $L_{\underline{k}}$ and $L_{\underline{k}+1}$ obtain the same payoff only if:

$$c\left(L_{\underline{k}+1}\right) - c\left(L_{\underline{k}}\right) + \mu\left(L_{\geq \underline{k}+2}\right) \cdot p \cdot A = 1 \Leftrightarrow \mu\left(L_{\geq \underline{k}+2}\right) = \frac{1 - \left(c\left(L_{\underline{k}+1}\right) - c\left(L_{\underline{k}}\right)\right)}{p \cdot A}.$$

Let μ' be defined as follows: $\mu'\left(L_{\underline{k}}\right) = 0$, $\mu'\left(L_{\underline{k}+1}\right) = \mu\left(L_{\underline{k}}\right) + \mu\left(L_{\underline{k}+1}\right)$, and $\mu'\left(L_k\right) = \mu\left(L_k\right)$ for each $k \geq \underline{k} + 2$. The fact that $supp\left(\mu'\right) \subseteq supp\left(\mu\right)$ implies that $u\left(\left(\mu', \beta\right), \sigma\right) = u\left(\sigma, \sigma\right)$ and the equality $\mu\left(L_{\geq \underline{k}+2}\right) = \mu'\left(L_{\geq \underline{k}+2}\right)$ implies $u\left(\sigma, \left(\mu', \beta\right)\right) = u\left(\left(\mu', \beta\right), \left(\mu', \beta\right)\right)$ (because μ and μ' only differ in the frequency of $L_{\underline{k}}$ and $L_{\underline{k}+1}$, and these two abilities yield the same payoff).[26]

(d) If $c\left(L_{\underline{k}+2}\right) = c\left(L_{\underline{k}+3}\right)$ then σ is not a limit ESS. By the previous steps $u\left(\left(L_{\underline{k}+2}, \beta\right), \sigma\right) = u\left(\left(L_{\underline{k}+3}, \beta\right), \sigma\right)$ (because these two strategies play the same on the equilibrium path), and this implies that strategy $\sigma' = \left(\mu', \beta\right)$ which differs from $\sigma = \left(\mu, \beta\right)$ by an internal shift in the frequencies of abilities $L_{\underline{k}+2}$ and higher abilities with the same cognitive costs satisfy: $u\left(\sigma', \sigma\right) = u\left(\sigma, \sigma\right)$ and $u\left(\sigma', \sigma'\right) = u\left(\sigma, \sigma'\right)$. An analogous property would hold in any sufficiently close perturbed game, and thus σ cannot be a limit ESS.

(e) If $p < \frac{1 - \left(c\left(L_{\underline{k}+2}\right) - c\left(L_{\underline{k}}\right)\right)}{A - 1}$ or $c\left(L_{\underline{k}+2}\right) - c\left(L_{\underline{k}}\right) \geq 1$, then σ is not an equilibrium. Otherwise,

$$\mu\left(L_{\underline{k}}\right) = 1 - \frac{1 - \left(c\left(L_{\underline{k}+2}\right) - c\left(L_{\underline{k}}\right)\right)}{p \cdot (A - 1)}. \tag{B.1}$$

The argument is analogous to part (1-d) of the proof of Theorem 1.

(f) $L_{\underline{k}} = L_1$ (which implies by the previous steps that $\sigma \approx \sigma^*$). Assume to the contrary that $L_{\underline{k}} > L_1$:

    i. *If there is an incumbent ability which defects with positive probability against an observed L_1 opponent, then both players defect at all the renaming rounds.* Intuitively, this is because after such a defection is realized, it becomes common knowledge that the horizon is at most $\bar{k}$. The formal argument is as follows.[27] Assume to the contrary, that there is an incumbent ability who defects with positive probability when facing an observed L_1 opponent at a cooperative information set. Let $l \leq \bar{k}$ be the highest horizon in which an incumbent defects against an observed L_1. Assume to the contrary that either player cooperates

---

[26] One can show that slightly perturbing μ' to satisfy $\mu'\left(L_{\underline{k}+2}\right) = \mu\left(L_{\underline{k}+2}\right) + \epsilon$, would imply that $u\left(\sigma, \left(\mu', \beta\right)\right) < u\left(\left(\mu', \beta\right), \left(\mu', \beta\right)\right)$. That is, σ is not an NSS.

[27] Note that the argument is slightly more complex than part (1-a), as it deals with an information set off the equilibrium path.

with positive probability at any later stage. Let $m \leq l$ be the farthest round since the first defection, in which at least one of the players cooperates with positive probability. Consider strategy σ' that coincides with σ at all information sets, except that it defects (with probability 1) m rounds after the initial defection. Observe that strategy σ' yields a strictly higher payoff conditional on playing against L_1 opponents. Consider any ubiquitous perturbed game $\Gamma(\zeta)$ with sufficiently small $M(\zeta)$. By continuity, any strategy $\sigma'_\zeta \in \Sigma(\zeta)$ sufficiently close to σ' yields a strictly better payoff against any strategy $\sigma_\zeta \in \Sigma(\zeta)$ sufficiently close to σ (relative to the payoff that σ_ζ yields against itself). This contradicts the assumption that σ is a perfect equilibrium.

ii. *An incumbent ability which faces an observed L_1 opponent at a cooperative information set: (1) cooperates if the horizon is larger than 2; and (2) defects if the horizon is at most 2.*

   Defection at any horizon larger than 2 yields a strictly lower payoff due to the previous step. Cooperating at horizon 2 yields strictly lower payoff, because it immediately yields one less point, without changing the future play of the opponent (who always defects at the last stage, as it is a dominant action).

iii. If $L_{\underline{k}} > L_2$ then $u((L_1, \beta), \sigma) > u((L_{\underline{k}}, \beta), \sigma)$.

   By the previous parts, (L_1, β) achieves at most one less utility point (relative to $(L_{\underline{k}}, \beta)$) when facing an unobserving $L_{\underline{k}}$ opponent, and it achieves at least $A - 1$ $(A - 2)$ more points against an observing $L_{>\underline{k}}$ $(L_{\underline{k}})$ opponent. Thus, $u((L_1, \beta), \sigma) > u((L_{\underline{k}}, \beta), \sigma)$ if:

$$(1 - p) \cdot \mu\left(L_{\underline{k}}\right) \overset{?}{<} p \cdot \left(A - 1 - \mu\left(L_{\underline{k}}\right)\right) + c\left(L_{\underline{k}}\right) \Leftrightarrow \mu\left(L_{\underline{k}}\right) \overset{?}{<} p \cdot (A - 1) + c\left(L_{\underline{k}}\right).$$

   Substituting $\mu\left(L_{\underline{k}}\right)$ yields and defining $0 \leq x \equiv c\left(L_{\underline{k}+2}\right) - c\left(L_{\underline{k}}\right) < 1$:

$$\frac{p \cdot (A - 1) - 1 + x}{p \cdot (A - 1)} \overset{?}{<} p \cdot (A - 1) + c\left(L_{\underline{k}}\right) \Leftrightarrow x \overset{?}{<} p \cdot (A - 1) \cdot \left(p \cdot (A - 1) + c\left(L_{\underline{k}}\right) - 1\right) + 1$$

   Substituting $p \cdot (A - 1) \geq 1 - x$ yields:

$$x \overset{?}{<} (1 - x) \cdot \left(1 - x + c\left(L_{\underline{k}}\right) - 1\right) + 1 \Leftrightarrow x \overset{?}{<} (1 - x) \cdot \left(c\left(L_{\underline{k}}\right) - x\right) + 1$$

$$\Leftarrow x \overset{?}{<} (1 - x) \cdot (-x) + 1 \Leftrightarrow 1 - 2x + x^2 > 0 \Leftrightarrow (1 - x)^2 > 0 \Leftarrow 0 \leq x < 1.$$

iv. *If $L_{\underline{k}} = L_2$ and $c(L_4) < \frac{1}{A}$ then $u((L_1, \beta), \sigma) > u((L_2, \beta), \sigma)$.*

   By the previous parts, (L_1, b_1) achieves at most one less utility point (relative

to $(L_2, \beta))$ when facing an L_2 opponent, the same payoff when facing an unobserving $L_{>2}$ opponent, and at least $A - 1$ more points against an observing $L_{>2}$ opponent. Thus, $u((L_1, \beta), \sigma) > u((L_2, \beta), \sigma)$ if:

$$\mu(L_2) \overset{?}{<} p \cdot (A - 1) \cdot (1 - \mu(L_2)) + c(L_2) \Leftrightarrow \mu(L_2) \overset{?}{<} \frac{p \cdot (A - 1) + c(L_2)}{p \cdot (A - 1) + 1}.$$

Substituting $\mu(L_2)$ from (B.1), implies:

$$\frac{p \cdot (A - 1) - 1 + (c(L_4) - c(L_2))}{p \cdot (A - 1)} \overset{?}{<} \frac{p \cdot (A - 1) + c(L_2)}{p \cdot (A - 1) + 1}$$

$$\Leftrightarrow -1 + (p \cdot (A - 1) + 1) \cdot (c(L_4) - c(L_2)) \overset{?}{<} p \cdot (A - 1) \cdot c(L_2)$$

$$\Leftrightarrow c(L_4) - c(L_2) \overset{?}{<} \frac{1 + p \cdot (A - 1) \cdot c(L_2)}{1 + p \cdot (A - 1)},$$

and the last inequality is immediately implied by $c(L_4) < \frac{1}{A}$.

3. Case II: Assume that $p < \frac{A-1}{A}$ and there are incumbents who defect with positive probability when: (1) the opponent is a stranger, (2) the information set is cooperative, and (3) the horizon is strictly larger than $\underline{k} + 1$. Then:

   (a) $\mu\left(L_{\underline{k}}\right) \leq \frac{1}{A}$.
   Due to part (1-e), defection at horizon $\underline{k}+2$ $(> \underline{k}+2)$ yields $A-1$ (at least $2 \cdot (A-1)$) less utility points relative to cooperating until horizon $\underline{k}+1$ if the opponent has ability $L_{\underline{k}}$, and one (at most two) more points against any other opponent. Such a defection can yield a weakly better payoff only if:

   $$\mu\left(L_{\underline{k}}\right) \cdot (A - 1) \leq \left(1 - \mu\left(L_{\underline{k}}\right)\right) \Leftrightarrow \mu\left(L_{\underline{k}}\right) \leq \frac{1}{A}. \tag{B.2}$$

   (b) *All incumbent abilities $L_{\geq \underline{k}+2}$ defect with probability one when facing a stranger at a cooperative information set with horizon $\underline{k} + 2$.*
   Assume to the contrary that there is an incumbent ability $L_{\tilde{k}}$ $(\tilde{k} \geq \underline{k} + 2)$ that cooperates with positive probability against strangers at a cooperative information set with horizon $\underline{k} + 2$. Define strategy σ' to coincide with σ, except that σ' defects with probability one when ability $L_{\tilde{k}}$ faces a stranger at a cooperative information set with horizon $\underline{k} + 2$. The assumption of case II and part (1-e) imply that $u(\sigma', \sigma) \geq u(\sigma, \sigma)$, and that $u(\sigma', \sigma') > u(\sigma, \sigma')$, and we get a contradiction to neutral stability.

   (c) $\mu\left(L_{\underline{k}+1}\right) > 0$. Assume to the contrary that $\mu\left(L_{\underline{k}+1}\right) = 0$.

i. Assume that $p < \frac{A-2}{A-1}$: We compare the payoff of ability $L_{\underline{k}}$ and the mean payoff of any incumbent ability $L_{\geq \underline{k}+2}$ when facing an $L_{\underline{k}}$ opponent: $L_{\underline{k}}$ obtains one less utility point when the opponent's ability is observed, and at least $A-2$ more utility points when the opponent is a stranger. This implies that $u\left(\left(L_{\underline{k}}, \beta\right), \left(L_{\underline{k}}, \beta\right)\right) > u\left(\sigma, \left(L_{\underline{k}}, \beta\right)\right)$ (which contradicts neutral stability) if:

$$p < (A-2) \cdot (1-p) \Leftrightarrow p < \frac{A-2}{A-1}.$$

ii. Due to analogous arguments to parts (a-b), $\mu\left(L_{\underline{k}}\right) \leq \frac{1}{A}$ implies that all incumbent abilities $L_{\geq \underline{k}+3}$ defect with probability one when facing a stranger (or an incumbent ability in $L_{\geq \underline{k}+2}$) at a cooperative information set with horizon $\underline{k}+3$.

iii. Assume that $p \geq \frac{A-2}{A-1}$: To simplify notation let: $\alpha = \mu\left(L_{\geq \underline{k}+3}\right) / \mu\left(L_{\geq \underline{k}+2}\right)$, and $\mu = \mu\left(L_{\underline{k}}\right)$. We compare the payoff of ability $L_{\underline{k}}$ and the average payoff of abilities $L_{\geq \underline{k}+2}$. Ability $L_{\underline{k}}$ yields: (1) at least $A-2+\alpha \cdot (A-1)$ more points when facing an unobserved $L_{\underline{k}}$ opponent (probability $(1-p) \cdot \mu$), (2) one less point when facing an observed $L_{\underline{k}}$ opponent (probability $p \cdot \mu$), (3) at least $A-2+\alpha \cdot (A-1)$ more points when facing an observing $L_{\geq \underline{k}+2}$ opponent (probability $p \cdot (1-\mu)$), (4) at most one less point when facing an unobserved and unobserving $L_{\geq \underline{k}+2}$ opponent $((1-p)^2 \cdot (1-\mu))$, and (5) at most $1+\alpha$ less points when facing an observed and unobserving $L_{\geq \underline{k}+2}$ opponent (probability $(1-p) \cdot p \cdot (1-\mu)$). This implies that $u\left(\left(L_{\underline{k}}, \beta\right), \sigma\right) > u(\sigma, \sigma)$ (which contradicts σ being a Nash equilibrium) if:

$$(A-2+\alpha \cdot (A-1))(p \cdot (1-\mu) + (1-p) \cdot \mu) \stackrel{?}{>} p \cdot \mu + (1-p) \cdot (1-\mu) \cdot (1+p \cdot \alpha)$$

Substituting $A > 3$ yields:

$$\Leftarrow (1+2 \cdot \alpha)(\cdot p \cdot (1-\mu) + (1-p) \cdot \mu) \stackrel{?}{>} p \cdot \mu + (1-p) \cdot (1-\mu) \cdot (1+p \cdot \alpha)$$

$$\Leftrightarrow (2 \cdot p - 1) \cdot (1 - 2 \cdot \mu) + \alpha\left(2 \cdot (p \cdot (1-\mu) + (1-p) \cdot \mu) - (1-p) \cdot (1-\mu) \cdot p\right) \stackrel{?}{>} 0$$

$$\Leftrightarrow (2 \cdot p - 1) \cdot (1 - 2 \cdot \mu) + \alpha\left(p \cdot (1-\mu) \cdot (1+p) + 2 \cdot (1-p) \cdot \mu\right) \stackrel{?}{>} 0.$$

Substituting $p > \frac{A-2}{A-1} > \frac{1}{2}$ and $\mu < \frac{1}{A} < \frac{1}{3}$ implies the inequality.

(d)

$$\mu\left(L_{\geq \underline{k}+2}\right) = \frac{1 - c\left(L_{\underline{k}+1}\right)}{1 + (A-1) \cdot p}. \tag{B.3}$$

We compare the payoffs of ability $L_{\underline{k}}$ and ability $L_{\underline{k}+1}$. Ability $L_{\underline{k}}$ obtains: one less point when facing an $L_{\underline{k}}$ or $L_{\underline{k}+1}$ opponent, the same payoff when facing an unobserving $L_{\geq \underline{k}+2}$ opponent, and $A-1$ more points when facing an observing $L_{\geq \underline{k}+2}$ opponent. This implies that $u\left(\left(L_{\underline{k}},\beta\right),\sigma\right) = u\left(\left(L_{\underline{k}+1},\beta\right),\sigma\right)$ (which is implied by σ being an equilibrium) if and only if:

$$(A-1)\cdot p\cdot \mu\left(L_{\geq\underline{k}+2}\right) = 1 - \mu\left(L_{\geq\underline{k}+2}\right) - \left(c\left(L_{\underline{k}+1}\right) - c\left(L_{\underline{k}}\right)\right)$$

$$\Leftrightarrow \mu\left(L_{\geq\underline{k}+2}\right) = \frac{1 - \left(c\left(L_{\underline{k}+1}\right) - c\left(L_{\underline{k}}\right)\right)}{1 + (A-1)\cdot p}.$$

(e) *Strategy σ is not a limit ESS.*

Let μ'' be defined as follows: $\mu''\left(L_{\underline{k}}\right) = 0$, $\mu''\left(L_{\underline{k}+1}\right) = \mu\left(L_{\underline{k}}\right) + \mu\left(L_{\underline{k}+1}\right)$, and $\mu''(L_k) = \mu(L_k)$ for each $k > \underline{k}+1$. $supp\left(\mu''\right) \subseteq supp\left(\mu\right)$ implies that $u\left(\left(\mu'',\beta\right),\sigma\right) = u\left(\sigma,\sigma\right)$ and the previous part implies that $u\left(\sigma,\left(\mu'',\beta\right)\right) = u\left(\left(\mu'',\beta\right),\left(\mu'',\beta\right)\right)$ (because $\mu\left(L_{\geq\underline{k}+2}\right) = \mu''\left(L_{\geq\underline{k}+2}\right)$). An analogous property is satisfied in any sufficiently close perturbed game, and thus σ cannot be a limit ESS.[28]

4. Case III - Assume that $p \geq \frac{A-1}{A}$. Let $\tilde{k} > M_\sigma$. Let $m \leq M_\sigma$ be the highest horizon in which $L_{\tilde{k}}$ ability defects with a positive probability when facing an observed $L_{\tilde{k}}$ opponent at a cooperative information set (m cannot be higher than M_σ due to the assumption of early-niceness). An analogous argument to part (3b), implies that ability $L_{\tilde{k}}$ defects with probability one at horizon m when facing an observed $L_{\tilde{k}}$ opponent. Finally, an analogous argument to part (4e) of Theorem 1's proof.) shows that $p \geq \frac{A-1}{A}$ implies a contradiction to the assumption that σ is a limit ESS because defection yields a higher payoff than cooperation when facing an observed $L_{\tilde{k}}$ opponent at a cooperative information set with horizon $m+1$.[29]

□

---

[28] One can show that perturbing μ'' to satisfy either $\mu''\left(L_{\underline{k}+2}\right) = \mu\left(L_{\underline{k}+2}\right) + \epsilon$ or $\mu''\left(L_{\underline{k}+2}\right) = \mu\left(L_{\underline{k}+2}\right) - \epsilon$, would imply that $u\left(\sigma,\left(\mu'',\beta\right)\right) < u\left(\left(\mu',\beta\right),\left(\mu',\beta\right)\right)$ for any $p \neq 0.5$. That is, σ is not an NSS for any $p \neq 0.5$.

[29] If $p = \frac{A-1}{A}$, then defection and cooperation yields the same payoff, and one has to rely also on an analogous argument to part (3-3b) to imply the contradiction.

B.4   Other Results:

Proof of Part (1) of Proposition 4 (early-nice symmetric perfect equilibrium)

1. We begin by showing that $0 < \mu(L_1) < 1$. The preliminary observations and Case I of Theorem 2's proof hold with minor adaptations also for a symmetric perfect equilibrium. We are left with case II, in which there are incumbents who defect with positive probability when facing a stranger at a cooperative information set when the horizon is larger than $\underline{k} + 1$ (where $L_{\underline{k}}$ is the smallest incumbent). Part (3a) holds also in this setup and shows that $\mu\left(L_{\underline{k}}\right) \leq \frac{1}{A}$. Assume to the contrary that $L_{\underline{k}} \neq L_1$. We compare the payoff of abilities L_1 and $L_{\underline{k}}$ against σ. Ability L_1 obtains at most one less point when facing an $L_{\underline{k}}$ opponent (probability $\mu\left(L_{\underline{k}}\right)$), the same payoff when facing an unobserving $L_{>\underline{k}}$ opponent, and at least $A - 1$ more points when facing an observing $L_{>\underline{k}}$ opponent (probability $p \cdot \left(1 - \mu\left(L_{\underline{k}}\right)\right)$). Thus, $u\left((L_1, \beta), \sigma\right) > u\left(\left(L_{\underline{k}}, \beta\right), \sigma\right) = u(\sigma, \sigma)$ (and this contradicts σ being an equilibrium) if:

$$\mu\left(L_{\underline{k}}\right) < p \cdot \left(1 - \mu\left(L_{\underline{k}}\right)\right) \cdot (A - 1) \Leftrightarrow p > \frac{\mu\left(L_{\underline{k}}\right)}{\left(1 - \mu\left(L_{\underline{k}}\right)\right) \cdot (A - 1)} > \frac{\frac{1}{A}}{\frac{(A-1)}{A} \cdot (A - 1)} = \frac{1}{(A - 1)^2}.$$

2. We now show that $\mu(L_{\geq 5}) = 0$.

   (a) Assume first that $\mu(L_{\leq 2}) \leq \frac{1}{A}$. We compare the payoff of (L_1, β) and average payoff of $(L_{\geq 3}, \beta)$ against σ. Ability L_1 achieves at least $A - 2$ more points when facing an observing $L_{\geq 3}$ opponent (probability $p \cdot (1 - \mu(L_{\leq 2}))$), at most two less points when facing an $L_{\leq 2}$ opponent (probability $\mu(L_{\leq 2})$), and at most $1 + p$ less points when facing an unobserving $L_{\geq 3}$ opponent (probability $(1 - p) \cdot (1 - \mu(L_{\leq 2}))$). Thus $u((L_1, \beta), \sigma) > u((L_{\geq 3}, \beta), \sigma)$ (and this contradicts σ being an equilibrium) if:

$$p \cdot (1 - \mu(L_{\leq 2})) \cdot (A - 2) \stackrel{?}{>} 2 \cdot \mu(L_{\leq 2}) + (1 + p) \cdot (1 - p) \cdot (1 - \mu(L_{\leq 2}))$$

$$\Leftarrow p \cdot (1 - \mu(L_{\leq 2})) \cdot (A - 2) \stackrel{?}{>} 1 + \mu(L_{\leq 2}) \Leftrightarrow p \stackrel{?}{>} \frac{1 + \mu(L_{\leq 2})}{(1 - \mu(L_{\leq 2})) \cdot (A - 2)}$$

$$\Leftarrow p \stackrel{?}{>} \frac{\frac{A+1}{A}}{\frac{(A-1)}{A} \cdot (A - 2)} = \frac{A + 1}{(A - 1) \cdot (A - 2)}.$$

   (b) Assume that $\mu(L_{\leq 2}) > \frac{1}{A}$. By an analogous argument to part (4e) of Theorem 1's proof, it implies that it is strictly better to cooperate at any horizon larger than three when facing a stranger at a cooperative information set. The assumption that $p < \frac{A-1}{A}$ implies by an analogous argument to part (4-e) of Theorem 1's proof

that it is strictly better to cooperate at any horizon larger than four when facing an incumbent at a cooperative information set. Thus, on the equilibrium path all incumbents cooperate at all horizons larger than four. This implies that if $c(L_5) > c(L_4)$ then all incumbents have ability of at most L_4.

The proofs of parts (2-3) of Proposition 4 are very similar to the analogous parts of the proof of Theorem 2 (omitted for brevity).

Proof of Proposition 2 $\left((L_1, b_{defect})\right.$ is a strict limit ESS):

Proof. Lorberbaum et al. (2002) study a perturbed variant of the standard repeated Prisoner's Dilemma, in which there is a fixed minimal probability $\epsilon > 0$ for each action at each information set. They show that the *ϵ-perturbed defect* (the strategy that defects with probability $1 - \epsilon$ at all information sets) is a symmetric strict equilibrium (and hence, also an ESS) in the *ϵ-perturbed game.* Minor adaptations to their proof (omitted for brevity) allow us to extend the result (for δ sufficiently close to one): (1) for any ubiquitous perturbation; and (2) for the current setup, in which players may become informed earlier about the realized length of the game.   □

Proof of Proposition 3 (each L_k can be the unique incumbent in a strict limit ESS):

Proof. The proof includes the following parts:

1. *Notation and Preliminary definitions:*

   (a) Given a finite action profile $W_t = (W^1, W^2) \in (\{C, D\}, \{C, D\})^t$, let $\bar{u}(W)$ be the average stage payoff of player 1 who repeats playing cycles of W^1 and faces an opponent who repeats playing W^2. Let S_j denote the jth action in the sequence S. To simplify notation, assume without loss of generality that $S_1 = C$.

   (b) Let $M' \in \mathbb{N}$ and let $\dot{W}, \ddot{W}, \dddot{W} \in (\{C, D\}, \{C, D\})^{M'}$ be sequences of action profiles that satisfy the following properties: (1) the sequence $\ddot{W}$ is the "reflection" of $\dot{W}$, in which the roles of players 1 and 2 are exchanged: $\forall 1 \leq j \leq M, i \in \{1, 2\} \; \dot{W}_j^i = \ddot{W}_j^{-i}$; (2) the sequence $\dot{W}$ begins with defection: $\dot{W}_1 = D$. (3) $\dddot{W}$ is a symmetric action profile that begins with mutual defection ($\dddot{W}_1 = (D, D)$); and (4) the average stage payoffs are ordered as follows: $\bar{u}\left(\dot{W}\right) > \bar{u}\left((S, S)\right), \bar{u}\left(\dddot{W}\right) > \bar{u}\left(\ddot{W}\right) > 1$.

   (c) Let $W_t \in (\{C, D\}, \{C, D\})^t$ be a symmetric action profile of length t, in which both players repeat playing cycles of S: $\forall 1 \leq j \leq t \; W_{t,j} = (S_{j \bmod M}, S_{j \bmod M})$. Similarly, let $\dot{W}_t$ (resp., $\ddot{W}_t$, $\dddot{W}_t$) be an action profile of length t, in which both players repeat playing cycles of $\dot{W}$ (resp., $\ddot{W}$, $\dddot{W}$): $\forall 1 \leq j \leq t \; \dot{W}_{t,j} = \left(\dot{W}_{j \bmod M'}\right)$ (resp., $\ddot{W}_{t,j} = \left(\ddot{W}_{j \bmod M'}\right)$, $\dddot{W}_{t,j} = \left(\dddot{W}_{j \bmod M'}\right)$).

2. *Definition of the deterministic playing rule $b_{W,k}$:* At stage 1 $b_{W,k}(L_{k'}, l.s, \emptyset) = C$ iff $\left(s \in \{L_k, \phi\} \text{ and } \left(\exists k < l' < l \text{ s.t. } W^i_{t+1+l-l'} = C\right)\right)$. That is, at stage 1 each player cooperates only if he observes his opponent to have the incumbent ability (or a stranger), and in addition, the horizon is long enough such that his opponent is likely to cooperate in the future at least once. To simplify the notation below, we slightly abuse it, and write $s = L_k$ instead of $s = \phi$ when the opponent is a stranger (and has probability 1 to have the incumbent ability L_k). At the remaining stages $(t \geq 1)$ $b_{W,k}\left(L_{k'}, l.s = L_{k''}, (a^i, a^{-i})^t\right) = C$ iff either of the following conditions hold:

$$\left((a^i, a^{-i})^t = W_t \text{ and } W^i_{t+1} = C \text{ and}\right.$$

$$\left(\exists l_1, l_2 \text{ s.t. } ((k < l_1 < l_2 < l) \text{ or } (k'' < l_1 < l)) \text{ and } W^{-i}_{t+1+l-l_1} = W^{-i}_{t+1+l-l_2} = C\right);$$

$$\text{or } \left((a^i, a^{-i})^t = \dot{W}_t \text{ and } \dot{W}^i_{t+1} = C \text{ and } L_{k''} \neq L_k \text{ and}\right.$$

$$\left(\exists l_1, l_2 \text{ s.t. } ((k < l_1 < l_2 < l) \text{ or } (k'' < l_1 < l)) \text{ and } \ddot{W}^{-i}_{t+1+l-l_1} = \ddot{W}^{-i}_{t+1+l-l_2} = C\right);$$

$$\text{or } \left((a^i, a^{-i})^t = \ddot{W}_t \text{ and } \ddot{W}^i_{t+1} = C \text{ and } L_{k'} \neq L_k \text{ and } \left(\exists k'' < l_1 < l \text{ s.t. } \ddot{W}^{-i}_{t+1+l-l_1} = C\right)\right)$$

$$\text{or } \left((a^i, a^{-i})^t = \dddot{W}_t \text{ and } \dddot{W}^i_{t+1} = C \text{ and } L_{k'}, L_{k''} \neq L_k \text{ and } \left(\exists k'' < l_1 < l \text{ s.t. } \dddot{W}^{-i}_{t+1+l-l_1} = C\right)\right).$$

That is, the first action-profile determines which sequence the players should follow: W if it was (C, C), $\dot{W}$ if it was (D, C), $\ddot{W}$ if it was (C, D), and $\dddot{W}$ if it was (D, D). The players follow this cycle until either of the following occurs:

(a) It becomes common knowledge that either player has deviated in the past - in this case, both players defect at all remaining stages.

(b) A player knows that his opponent is not going to cooperate in the future (because the horizon is too short) - in this case he defects.

3. Fix an arbitrary ubiquitous perturbed game $\Gamma(\zeta)$ with sufficiently small maximal tremble $M(\zeta)$. Let $\sigma_{W,k,\zeta} = (\mu_{W,k,\zeta}, \beta_{W,k,\zeta}) \in \Sigma(\zeta)$ be the closest strategy to $(L_k, b_{W,k})$ in $\Sigma(\zeta)$, and let $\sigma = (\mu, \beta) \neq \sigma_{W,k,\zeta} \in \Sigma(\zeta)$ any other strategy. We now show that $u(\sigma, \sigma_{W,k,\zeta}) < u(\sigma_{W,k,\zeta}, \sigma_{W,k,\zeta})$ (i.e, $\sigma_{W,k,\zeta}$ is a symmetric strict equilibrium in $\Sigma(\zeta)$), which implies that $(L_k, b_{W,k})$ is a strict limit ESS. The argument is a simple adaptation to *Kim (1994)*'s folk theorem result and is briefly sketched as follows:

(a) $u((\mu, \beta), \sigma_{W,k,\zeta}) \leq u((\mu, \beta_{W,k,\zeta}), \sigma_{W,k,\zeta})$. This is because any deviation from playing rule $\beta_{W,k}$, which is observed by the opponent, leads the players to defect at all

remaining stages, and for δ sufficiently close to 1, the future loss outweighs the gain.

(b) $u\left(\left(\mu, \beta_{W,k,\zeta}\right), \sigma_{W,k,\zeta}\right) < u\left(\sigma_{W,k,\zeta}, \sigma_{W,k,\zeta}\right)$ *if* $\mu \neq \mu_{W,k}$. This is because playing-rule $\beta_{W,k,\zeta}$ induces a strictly higher payoff to L_k and any distribution $\mu \neq \mu_{W,k}$ assigns a smaller frequency to L_k and higher frequencies to all other abilities.

$\square$

References

Andreoni, J., & Miller, J.H. 1993. Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *The Economic Journal*, **103**(418), 570–585.

Axelrod, R.M. 1984. *The Evolution of Cooperation*. Basic Books.

Binmore, K.G., & Samuelson, L. 1992. Evolutionary stability in repeated games played by finite automata. *Journal of economic theory*, **57**(2), 278–305.

Bolton, Gary E. 1997. The rationality of splitting equally. *Journal of Economic Behavior and Organization*, **32**(3), 365–381.

Bruttel, L.V., Güth, W., & Kamecke, U. 2012. Finitely repeated prisoners' dilemma experiments without a commonly known end. *Int. J. of Game Theory*, **41(1)**, 23–47.

Camerer, C. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press, Princeton.

Cooper, R., DeJong, D.V., Forsythe, R., & Ross, T.W. 1996. Cooperation without reputation: Experimental evidence from prisoner's dilemma games. *Games and Economic Behavior*, **12(2)**, 187–218.

Costa-Gomes, M., Crawford, V.P., & Broseta, B. 2001. Cognition and behavior in normal-form games: An experimental study. *Econometrica*, **69**(5), 1193–1235.

Crawford, V.P. 2003. Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *The American Economic Review*, **93**(1), 133–149.

Cressman, Ross. 1997. Local stability of smooth selection dynamics for normal form games. *Mathematical Social Sciences*, **34**(1), 1–19.

Dekel, Eddie, Ely, Jeffrey C., & Yilankaya, Okan. 2007. Evolution of preferences. *Review of Economic Studies*, **74**(3), 685–704.

Frenkel, S., Heller, Y., & Teper, R. 2012. Endowment as a blessing. *mimeo.*

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

Geanakoplos, J., & Gray, L. 1991. When seeing further is not seeing better. *Bulletin of the Santa Fe Institute*, **6**(2).

Guth, Werner, & Yaari, Menahem. 1992. Explaining reciprocal behavior in simple strategic games: An evolutionary approach. *In:* Witt, Ulrich (ed), *Explaining Process and Change: Approaches to Evolutionary Economics.* University of Michigan Press, Ann Arbor.

Haviland, W.A., Prins, H.E.L., & Walrath, D. 2007. *Cultural Anthropology: The Human Challenge.* Wadsworth Pub. Co.

Heller, Y., & Mohlin, E. 2013. Read my lips. *mimeo.*

Jehiel, Philippe. 2001. Limited foresight may force cooperation. *The Review of Economic Studies*, **68**(2), 369–391.

Johnson, E.J., Camerer, C., Sen, S., & Rymon, T. 2002. Detecting failures of backward induction: Monitoring information search in sequential bargaining. *Journal of Economic Theory*, **104**(1), 16–47.

Kim, Yong-Gwan. 1993. Evolutionary stability in the asymmetric war of attrition. *Journal of theoretical biology*, **161**(1), 13.

Kim, Yong-Gwan. 1994. Evolutionarily stable strategies in the repeated prisoner's dilemma. *Mathematical Social Sciences*, **28**(3), 167–197.

Kraines, David, & Kraines, Vivian. 1989. Pavlov and the prisoner's dilemma. *Theory and decision*, **26**(1), 47–79.

Kreps, David M, & Wilson, Robert. 1982. Sequential equilibria. *Econometrica*, **50(4)**, 863–894.

Kreps, David M, Milgrom, Paul, Roberts, John, & Wilson, Robert. 1982. Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic theory*, **27**(2), 245–252.

Leimar, Olof. 1997. Repeated games: a state space approach. *Journal of Theoretical Biology*, **184**(4), 471–498.

Lorberbaum, Jeffrey. 1994. No strategy is evolutionarily stable in the repeated prisoner's dilemma. *Journal of Theoretical Biology*, **168**(2), 117.

Lorberbaum, Jeffrey P, Bohning, Daryl E, Shastri, Ananda, & Sine, Lauren E. 2002. Are There Really No Evolutionarily Stable Strategies in the Iterated Prisoners Dilemma? *Journal of Theoretical Biology*, **214**(2), 155–169.

Maynard-Smith, J., & Price, G.R. 1973. The Logic of Animal Conflict. *Nature*, **246**, 15.

Mengel, F. 2012. Learning by (limited) forward looking players. *mimeo.*

Mohlin, Erik. 2012. Evolution of theories of mind. *Games and Econ. Behav.*, **75**(1), 299–318.

Nachbar, John H. 1990. Evolutionary selection dynamics in games: convergence and limit properties. *International journal of game theory*, **19**(1), 59–89.

Nagel, R. 1995. Unraveling in guessing games: An experimental study. *The American Economic Review*, **85**(5), 1313–1326.

Neelin, J., Sonnenschein, H., & Spiegel, M. 1988. A further test of noncooperative bargaining theory: Comment. *The American Economic Review*, **78**(4), 824–836.

Neyman, Abraham. 1999. Cooperation in repeated games when the number of stages is not commonly known. *Econometrica*, **67**(1), 45–64.

Nowak, Martin, & Sigmund, Karl. 1993. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature*, **364**, 56–58.

Okada, Mr A. 1981. On stability of perfect equilibrium points. *International Journal of Game Theory*, **10**(2), 67–73.

Osborne, Martin J, & Rubinstein, Ariel. 1994. *Course in game theory.* The MIT press.

Robson, A.J. 1990. Efficiency in evolutionary games: Darwin, Nash, and the secret handshake. *Journal of Theoretical Biology*, **144**(3), 379–396.

Robson, A.J. 2003. The evolution of rationality and the Red Queen. *Journal of Economic Theory*, **111**, 1–22.

Rosenthal, R.W. 1981. Games of perfect information, predatory pricing and the chain-store paradox. *Journal of Economic Theory*, **25**(1), 92–100.

Samuelson, Larry. 1987. A note on uncertainty and cooperation in a finitely repeated prisoner's dilemma. *International Journal of Game Theory*, **16**(3), 187–195.

Samuelson, Larry. 1991. Limit evolutionarily stable strategies in two-player, normal form games. *Games and Economic Behavior*, **3**(1), 110–128.

Sandholm, William H. 2010. Local stability under evolutionary game dynamics. *Theoretical Economics*, **5**(1), 27–50.

Selten, R., & Stoecker, R. 1986. End behavior in sequences of finite Prisoner's Dilemma supergames: A learning theory approach. *Journal of Economic Behavior & Organization*, **7**(1), 47–70.

Selten, Reinhard. 1975. Reexamination of the perfectness concept for equilibrium points in extensive games. *International journal of game theory*, **4**(1), 25–55.

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

Selten, Reinhard. 1988. Evolutionary stability in extensive two-person games-correction and further development. *Mathematical Social Sciences*, **16**(3), 223–266.

Stahl, D.O. 1993. Evolution of smart-n players. *Games and Econ. Behavior*, **5**(4), 604–617.

Stahl, D.O., & Wilson, P.W. 1994. Experimental evidence on players' models of other players. *Journal of Economic Behavior and Organization*, **25**(3), 309–327.

Stennek, J. 2000. The survival value of assuming others to be rational. *International Journal of Game Theory*, **29**(2), 147–163.

Swinkels, J.M. 1992. Evolutionary stability with equilibrium entrants. *Journal of Economic Theory*, **57**(2), 306–332.

Thomas, Bernhard. 1985. On evolutionarily stable sets. *J. of Math. Biology*, **22**(1), 105–115.

van Damme, E. 1987. *Stability and Perfection of Nash Equilibria*. Springer, Berlin.

Weibull, Jörgen W. 1995. *Evolutionary game theory*. The MIT press.

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