Three steps ahead

Heller, Yuval

University of Oxford, Department of Economics

1 January 2015

Three steps ahead

Yuval Heller
Department of Economics, University of Oxford

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 0 or 1, then the game admits an evolutionarily stable strategy, in which agents who look one step ahead and agents who look three steps ahead coexist. 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 have limited foresight. 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 and 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 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. Each agent is endowed by a type that determines his foresight ability

---

[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 II and Wilson 1994, Nagel 1995, Costa-Gomes et al. 2001).

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

TABLE 1. Payoff at the symmetric stage game prisoner's dilemma ($A > 1$).

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.*

Our static analysis focuses on a symmetric two-player game, in which the set of actions of each player is the set of feasible types in the population. A mixed equilibrium in this auxiliary game describes a distribution of types in the population. It is well known (see, e.g., Nachbar 1990) that a distribution of types is dynamically stable only if its corresponding mixed strategy is a symmetric Nash equilibrium.

The auxiliary game includes an initial round in which players choose their foresight ability and **T** rounds of repeated prisoner's dilemma, where **T** is geometrically distributed with a continuation probability close to 1 (i.e., a high enough expected length). At stage 0, each player chooses a *foresight ability* (abbreviated, *ability*) from the set $\{L_1, L_2, \ldots, L_k, \ldots\}$. A player with ability L_k is privately informed at round $\mathbf{T} - k$ about the realization of **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 (nonmonotonic costs are discussed in Section 8.2).

Each player obtains a private signal about his opponent's ability (à la Dekel et al. 2007): the signal reveals the opponent's ability with probability p and it is noninformative 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.[2]

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 opponent gets 0.[3] The total payoff of the game is the undiscounted sum of payoffs.

---

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

[3]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).

We begin by characterizing a specific symmetric Nash equilibrium, σ^*, for every p that is not too close to 0 and 1 (and the width of this interval is increasing in A). 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.[4] 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, $\mu(L_1)$, 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); (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 and 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. 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).

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.[5] 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 2) shows that σ^* is a limit ESS.[6]

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 the horizon is sufficiently large and no one has ever defected in the past. Empirical evidence suggests that focusing on early-nice strategies is plausible: Selten and Stoecker (1986) experimentally

---

[4]The name "pavlov" (Kraines and Kraines 1989, Nowak and 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).

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

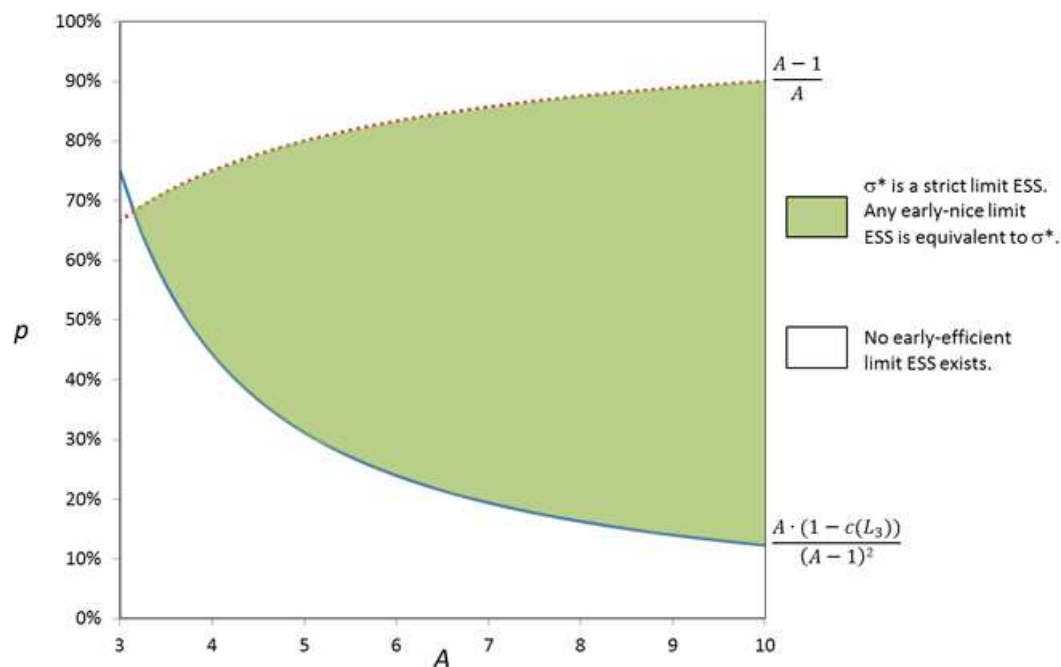[6]Moreover, we show that σ^* is the limit of ESS of *every* sequence of perturbed games (*strict limit ESS*).

FIGURE 1. Summary of main results.

demonstrate that most subjects satisfy early-niceness when playing the repeated pris-oner's dilemma in the lab,[7] and the tournaments of Axelrod (1984) and Wu and Axelrod (1995) suggest that "niceness" (not being the first to defect) might be a necessary re-quirement for evolutionary success. In Section 6, we adapt the results of Fudenberg and Maskin (1990) and show that all the non-early-nice strategies of the above folk-theorem result become unstable when the continuation probability converges to 1 (while the early-nice strategy σ^* remains stable).

Our second main result (Theorem 5) 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 solu-tion 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 0 or 1.

The intuition of Theorem 5 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 reach-ing the $\underline{k}$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 de-fecting one stage earlier. If there are two consecutive abilities, then the lower ability is

---

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

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 $\{L_{\underline{k}}, L_{\underline{k}+2}\}$ 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 1 because players are "trapped" in an arms race toward earlier defections and higher abilities.

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 in primitive *hunter–gatherer* societies (representing sequential gift exchange; see, e.g., Haviland et al. 2007, p. 440) as well as in modern societies.

We conclude by briefly surveying the related literature. Our paper is related to the literature that studies the stability of cooperation in the repeated prisoner's dilemma (e.g., the seminal work of Axelrod 1984, and the recent work of van Veelen and García 2010). Several aspects of our proofs rely on ideas from Kim (1994), Lorberbaum (1994), and Lorberbaum et al. (2002), which have been extended to the current setup with foresight abilities. Another related paper is the seminal work of Kreps et al. (1982), which shows that if one of the players may be committed to tit-for-tat behavior (and the commitment is unobservable by the opponent), then players mutually cooperate until the last few rounds in any equilibrium. One can interpret ability L_1 in our model as a similar commitment device. A key difference between the two models is that in Kreps et al. (1982), a committed player achieves a strictly lower payoff relative to a noncommitted player, while in our model, L_1 players achieve the maximal payoff.

A closely related paper is Jehiel (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.[8] 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 chooses to look only a few steps ahead. In addition, the current paper presents a novel notion of limited foresight, which may be of independent interest (see Section 8.1).

Geanakoplos and 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 II (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

---

[8]Recently, Mengel (2014) 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.

to 0 or 1. 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 coexist and obtain the same payoff.

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. Finally, Appendix B includes the formal proofs.

2. Model

As mentioned in the Introduction, we are interested in characterizing dynamically stable states in a large population of agents, where each agent is endowed with a type that determines his ability and his behavior. We do so by studying an auxiliary static symmetric two-player game in which the set of actions of each player is the set of feasible types of the agents.

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 $k > 2$).[9] We focus on the case of δ close to 1.

At stage 0, each player $i \in \{1, 2\}$ chooses his *ability* from the set $\mathcal{L} = \{L_1, L_2, L_3, \ldots, L_k, \ldots\}$.[10] 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).[11] 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* afterward. We shall use the term *stranger* to describe an opponent whose foresight ability is not observed, and

---

[9]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.

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

[11]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 the model to allow p to be determined endogenously by the players.

we shall use the term *observing* (*nonobserving*) 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.[12] 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: $\{C, D\}$.

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 $I = (L, l, s, (a^i, a^{-i})^n)$, where $L \in \mathcal{L}$ is the player's ability (as chosen at stage 0), $l \in \{1, \dots, L\} \cup \infty$ is the number of remaining periods (dubbed the horizon), with $l < \infty$ ($l = \infty$) describing an informed (uninformed) agent, $s \in \{\mathcal{L} \cup \phi\}$ is the signal about the opponent's ability, with $s = \phi$ describing a noninformative signal (i.e., facing a stranger), and $(a^i, a^{-i})^n \in (\{C, D\} \times \{C, D\})^n$ 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 $\mathcal{I} = \bigcup_{n \geq 1} \mathcal{I}_n$ 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(μ) shall be called the *incumbents*. Let Σ (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 $((L^1, L^2), (a^1, a^2)^n)$, where (L^1, L^2) describes the abilities chosen at stage 0 and $(a^1, a^2)^n$ describes the n actions taken at stages $1, \dots, n$. 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

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

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 and 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

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

---

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

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

We study the long-run stable outcomes of payoff-monotonic dynamics in which more successful types become more frequent. We interpret these dynamics to be the result of a process of cultural learning.[13] 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 symmetric Nash equilibrium of the auxiliary game. Motivated by this observation, we characterize in this section a specific Nash equilibrium, σ^*. We emphasize that this equilibrium behavior can be achieved by agents who passively follow their types, rather than actively maximize their payoffs.[14]

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 a deterministic simple playing rule. Players follow pavlov (defect if and only if players played differently in the previous round) when they are uninformed about the number of remaining rounds. Players with ability L_1 defect at the last stage. Finally, an L_3 player who is facing an L_1 opponent or does not know the opponent's ability starts defecting with two rounds to go; otherwise he starts defecting with three rounds to go.

DEFINITION 2. For every $p > 0$, $A > 1$, and $c(L_3) < 1$, let $\sigma^* = (\mu^*, b^*)$ be[15]

$$\mu^*(L_1) = 1 - \frac{1 - c(L_3)}{p \cdot (A - 1)}, \qquad \mu^*(L_3) = \frac{1 - c(L_3)}{p \cdot (A - 1)}$$

$$\forall k \notin \{1, 3\} \quad \mu^*(L_k) = 0$$

---

[13]The dynamics also fit a biological evolutionary process in which the type is determined by the gene.

[14]The results also hold in the presence of sophisticated agents who explicitly maximize their payoffs. Thus, the model can also fit a nonevolutionary strategic setup in which players explicitly choose how much effort to spend on detecting early signs that the interaction is going to end soon (foresight ability), and then they play the repeated prisoner's dilemma (with partial observability of the opponent's effort).

[15]Definition 2 describes the behavior of players at all information sets, including the behavior off the equilibrium path (e.g., after observing that there are five more rounds to go), which is important for the equilibrium refinements presented in the next section. Observe that L_1 players only observe either $l = \infty$ or $l = 1$, and thus they stop playing pavlov only in the last round.

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

Our first result shows that (μ^*, b^*) is a Nash equilibrium if p is not too close to 0 or 1, the cognitive cost of L_3 is not too high, and the continuation probability δ is close enough to 1.

THEOREM 1. *There exists $\bar{\delta} < 1$ such that for all $\delta \in (\bar{\delta}, 1)$, if $A \cdot (1 - c(L_3))/(A - 1)^2 < p < (A - 1)/A$ and $c(L_3) < 1/A$, then σ^* is a symmetric Nash equilibrium.*

Theorem 1 is implied by Theorem 2 (which is proved in Appendix B, along with the other results in the paper). The sketch of the proof is as follows. If we suppose that only abilities L_1 and L_3 are chosen, and that the players follow b^*, then the result is a hawk–dove game between these abilities: an L_3 player fares better against an L_1 opponent by defecting at horizon 2, while an L_1 player fares better against an L_3 opponent due to its indirect "commitment" advantage; that is, when the opponent observes ability L_1 (which happens with probability p), it induces him to cooperate for an additional round. Thus, each ability becomes less successful (relative to the other ability) if its frequency becomes larger. As a result, a unique frequency of L_1 players balances the payoffs of the two abilities, and this frequency is increasing in p. If p is not too small, then the frequency of L_1 players is sufficiently large, such that it is optimal for an L_3 player who does not know the opponent's ability to start defecting only in the penultimate round. If p is not too large, then an $L_{>3}$ player who observes an L_3 opponent will still wait to defect until there are three rounds to go, in the hope that the opponent has not observed his ability. Finally, an L_2 player is outperformed because he does not have the commitment advantage of the L_1 players, and, in addition, unlike the L_3 players, he is unable to defect three rounds before the end.

REMARK 2. Theorem 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 if and only if the opponent defected in the previous round) or perfect grim trigger (defect if and only if 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 nonoptimal 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 and Sigmund 1993), which allows its stability to remain robust when $\delta \to 1$ (see the invasion barrier analysis in Section 6). 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 and 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'. The formal definition follows.

DEFINITION 3 (Maynard Smith and 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 (i) it is a symmetric Nash equilibrium and (ii) $\forall \sigma' \neq \sigma$, 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.

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.[16] Selten (1983) slightly weakens this notion by requiring evolutionary stability in a converging sequence of perturbed games in which players rarely "tremble" and play "wrong" actions (but not necessarily in the unperturbed game). The formal definition follows.

DEFINITION 4 (Selten 1983, 1988). A *(full support) perturbation* ζ is a function that assigns a nonnegative (positive) number for

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

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

Let $\Gamma(\zeta)$ denote the (full support) *perturbed game* that results from perturbing the game described in Section 2 by (full support) perturbation ζ. In game $\Gamma(\zeta)$, each player is limited to choose strategy $\sigma = (\mu, \beta)$ that satisfies $\mu(L_k) \geq \zeta(L_k)$ for each $L_k \in \mathcal{L}$ and $\zeta(I)(C) \leq \beta(I)(C) \leq 1 - \zeta(I)(D)$ for each $I \in \mathcal{I}$. Let $\Sigma(\zeta)$ (resp., $\Delta_\zeta(\mathcal{L})$, $B(\zeta)$) be 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: $M(\zeta) = \max(\sup_{L_k \in \mathcal{L}} \zeta(L_k), \sup_{I \in \mathcal{I}, a \in \{C, D\}} \zeta(I)(a))$.

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).[17]

---

[16]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).

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

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's 1981 notion of strict perfection) is that a strong notion of stability should be robust to the specific structure of the perturbations. This is formally stated as follows.

DEFINITION 6. Strategy $\sigma \in \Sigma$ is a *strict limit ESS* (*strict limit NSS*) if, for every sequence of full support 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 Theorem 1 and shows that σ^* is a strict limit ESS.

THEOREM 2. *There exists $\bar{\delta} < 1$ such that for all $\delta \in (\bar{\delta}, 1)$, if $A \cdot (1 - c(L_3))/(A-1)^2 < p < (A-1)/A$ and $c(L_3) < 1/A$,[18] 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 full support perturbation, and let $\sigma_\zeta^* = (\mu_\zeta^*, b_\zeta^*)$ be the closest strategy to σ^* in the perturbed game $G(\zeta)$ that satisfies $u((L_1, b_\zeta^*), \sigma_\zeta^*) = u((L_3, b_\zeta^*), \sigma_\zeta^*)$. Lorberbaum et al. (2002) proved that the perturbed pavlov is a strict best-reply to itself when playing a 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 Theorem 1, this implies that playing rule b_ζ^* is a strict best-reply to σ_ζ^* (for all abilities), ability L_2 achieves a strictly lower payoff than L_3, and any ability $L_{>3}$ can achieve, at most, the same payoffs as L_3. The properties of the hawk–dove "metagame" between abilities L_1 and L_3 (discussed in Section 3) imply that any strategy with 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. We conclude this section with a few comments about the stability of σ^*:

1. *Stability without cognitive costs.* Minor adaptions to the proof imply a slightly stronger result when $c(L_4) = c(L_3)$. Let $L_{\text{as\_3}} = \{L_k \mid k \geq 3, c(L_k) = c(L_3)\}$ be the abilities with the same costs as L_3. Then

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

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

2. *Uniform limit ESS.* In Heller (2014), I show that the notion of "limit ESS" is too weak: it does not imply neutral stability and it may be dynamically unstable in the sense that almost any small perturbation takes the population away. These two issues are caused by the implicit assumption of the notion of limit ESS that

---

[18]This assumption can be slightly weakened as $c(L_3) < \min(1, 1/A + (1 - 1/A) \cdot c(L_2))$.

mutants are rarer than "trembling" incumbents. I solve these two issues by defining a slightly stronger notion, *uniform limit ESS*, which requires mutants to be strictly outperformed also but without this implicit assumption. Minor adaptations to the proof imply that σ^* is a uniform limit ESS.

3. *Dynamic stability of σ^*.* All of our results remain qualitatively the same if one restricts players to choose a foresight ability of at most L_M ($M \geq 3$) and a playing rule that depends only on the last $N \geq 3$ rounds. With such restrictions, each player has a finite set of strategies, and existing results imply that σ^* is dynamically stable:[19]

   (a) The results of Thomas (1985) imply that σ^* is Lyapunov stable in the unperturbed game under the replicator dynamics.

   (b) The results of Cressman (1997) and Sandholm (2010) imply that σ^* is asymptotically stable (i.e., populations starting close enough to σ^* eventually converge to it) under a large variety of payoff-monotonic dynamics and in any full support game $\Gamma(\zeta)$ with a sufficiently small $M(\zeta)$.[20]

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 theorem shows that the game also admits an inefficient stable strategy in which all players have ability L_1 and always defect.

THEOREM 3. *Let $\sigma_{\text{def}} = (L_1, b_{\text{def}})$ with $b_{\text{def}} \equiv D$ (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.'s (2002) result that defection is a strict best-reply to itself in the slightly perturbed repeated prisoner's dilemma.

The following theorem formally 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.

THEOREM 4. *Let $L_k \in \mathcal{L}$, $M \in \mathbb{N}$, and $S \in (\{C, D\})^M \neq (D, \ldots, D)$. Assume that $0 < p$. Then there exists $\bar{\delta} < 1$ (which depends on A, p, L_k, and S) such that for all $\delta \in (\bar{\delta}, 1)$, there exists a strict limit ESS $\sigma_{S,k} = (L_k, \beta_{S,k})$ in which, on the equilibrium path, uninformed players repeat playing cycles of the sequence S.*

---

[19]Note that the dynamic stability in the unperturbed game is relatively weak. Strategy σ^* is vulnerable to a sequence of two consecutive invasions: a neutral mutant who always cooperates, which creates a selective advantage for a second mutant who start defecting earlier in the game (as shown by van Veelen and García 2010, any strategy in the unperturbed repeated prisoner's dilemma has a similar vulnerability). However, as soon as any small perturbation (with full support) is introduced, then σ^* satisfies the strong notion of asymptotic stability and it is no longer vulnerable to a sequence of invasions.

[20]The results of Cressman (1997) and Sandholm (2010) require an additional mild requirement of regularity, namely, strictness with respect to strategies outside its support. Minor adaptations to the proof of Theorem 2 show that strategy σ^* is a regular ESS in $\Gamma(\zeta)$.

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 1 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 deviates from this pattern, both players always defect.

6. Early-niceness and uniqueness

6.1 *Early-niceness*

A strategy is early-nice if players who follow its playing rule cooperate when the horizon is large enough and no one has ever defected before. This is formalized as follows.

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

Early-niceness implies efficient play (mutual cooperation) at early stages of the interaction on the equilibrium path. Theorem 4 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 that induces early inefficient play only against nonincumbent 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. That is, it rules out the "discrimination" against mutants (playing a different cycle when observing a mutant ability), which is necessary for the stability of the various strategies of Theorem 4. Note that early-niceness does not restrict the play of a mutant player who follows a different playing rule.

In the Introduction, we presented an empirical motivation for early-niceness. We now present a theoretical justification when the continuation probability δ is close to 1. The argument adapts to the current setup the results of Fudenberg and Maskin (1990) for undiscounted infinite repeated games (see also related ideas in Robson 1990 and Bendor and Swistak 1997). We say that strategy σ has a (uniform) *invasion barrier* of $0 < \bar{\epsilon} < 1$ if for every mutant strategy σ', the incumbents strictly outperform the mutants in any post-entry population in which the frequency of the mutants is at most $\bar{\epsilon}$ (i.e., for each $0 < \epsilon < \bar{\epsilon}$ and each σ', the inequality $u(\sigma, (1 - \epsilon) \cdot \sigma + \epsilon \cdot \sigma') > u(\sigma', (1 - \epsilon) \cdot \sigma + \epsilon \cdot \sigma')$ holds). Note that (a) for finite games, a strategy is an ESS if and only if it has a positive invasion barrier (Weibull 1995, Proposition 2.5); (b) for games with infinite strategy spaces, the stronger notion of having a positive invasion barrier is required for implying dynamic stability (Oechssler and Riedel 2001); and (c) having a smaller invasion barrier implies a less robust stability. In what follows, we show that the invasion barriers of all the non-early-nice limit ESSs of the previous section converge to 0 as δ converges to 1.

This is in contrast to the early-nice strategy σ^* that has an invasion barrier bounded away from zero for all values of $\bar{\delta} < \delta < 1$.

Fix an arbitrary full support perturbed game $\Gamma(\zeta)$ with a sufficiently small $M(\zeta)$. We first deal with the invasion barrier of $\sigma_{\text{def}} = (L_1, \textit{defect})$ of Theorem 3. Let $\sigma' = (L_1, \textit{pavlov})$. Observe that σ' yields one less point against σ, but it obtains an expected gain of $(A-1)/(1-\delta)$ against σ' (compared with the payoffs of strategy σ against these opponents). Thus, the mutant strategy σ' outperforms the incumbent strategy σ in any post-entry population in which the mutants' share is at least $(1-\delta)/(A-1)$. This implies that the invasion barrier of σ_{def} converges to zero as $\delta \to 1$.

Next, consider one of the limit ESSs $\sigma_{S,k} = (L_k, \beta_{S,k})$ of Theorem 4. Let $\sigma'_{S,k}$ be a strategy that coincides with $\sigma_{S,k}$ as long as the players do not use the perfect-grim-trigger punishments. If the players reach a history in which they have to use a "punishment" (defect at all the remaining stages according to $\sigma_{S,k}$), then strategy $\sigma'_{S,k}$ induces them to play pavlov. By the same argument as in the case of σ_{def} above, the mutant strategy $\sigma'_{S,k}$ outperforms the incumbent strategy $\sigma_{S,k}$ in any post-entry population in which the mutants' share is at least $(1 - \delta)/(A - 1)$. We conjecture that the above argument can be extended to show that the invasion barrier of any non-early-nice strategy of finite complexity converges to zero as $\delta \to 1$.[21]

6.2 *Uniqueness result*

Two strategies are realization equivalent if they induce the same distribution over outcome paths; they can only differ in their off-equilibrium behavior. This is formally defined as follows.

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

Our second main result shows that any early-nice limit ESS is realization equivalent to σ^* (assuming $A > 3$) and that there is no early-nice limit ESS for values of p that are close to either 0 or 1.

THEOREM 5. *There exists $\bar{\delta} < 1$ such that for all $\delta \in (\bar{\delta}, 1)$, if $A > 3$ $\forall k \in \mathbb{N}$ $c(L_{k+1}) - c(L_k) < 1$, $c(L_4) < 1/A$, and $p < 1$, then strategy $\sigma = (\mu, \beta)$ is a an early-nice limit ESS only if $\sigma \approx \sigma^*$. Moreover, if $p < A \cdot (1 - c(L_3))/(A - 1)^2$ or $(A - 1)/A < p$, then no early-nice limit ESS exists.*

---

[21]The next step in the analogous result of Fudenberg and Maskin (1990) is to observe that if σ is a strategy with finite complexity, then there exists a history h^* that yields the lowest expected sum of payoffs in the remaining stages. Let σ'_{h^*} be a strategy that differs from σ only by playing after history h^* a different action from the one induced by σ, and mutually cooperating at all remaining stages if the opponent had done the same. A similar argument as above shows that the invasion barrier of σ against σ'_{h^*} converges to zero as $\delta \to 1$. In our setup, unlike in Fudenberg and Maskin (1990), players have private information, and this raises technical difficulties in the identification of history h^* and the characterization of σ'_{h^*}. Due to these technical difficulties, we leave the proof of the conjecture for future research.

The sketch of the proof is as follows. Let $L_{\underline{k}}$ be the lowest incumbent ability in supp(μ) and assume that all incumbents cooperate with probability 1 at horizons larger than M_σ. The inequality $c(L_{\underline{k}+1}) - c(L_{\underline{k}}) < 1$ implies that $\mu(L_{\underline{k}}) < 1$ (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 $L_{\geq \underline{k}+1}$ defect at horizon $\underline{k} + 1$. Next, we note that early-niceness 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. *We have $p < (A-1)/A$ and all incumbents cooperate on the equilibrium path when facing a stranger at a horizon larger than $\underline{k}+1$.* The assumption that $p < (A-1)/A$ 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 co-operate 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 $L_{\underline{k}+2}$ coexist 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. *We have $p < (A - 1)/A$ 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 1 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 $\mu(L_{\underline{k}+1}) > 0$ (otherwise σ is not stable to a perturbation that slightly increases $\mu(L_{\underline{k}})$). 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 $\{L_{\underline{k}}, L_{\underline{k}+1}\}$ and an additional fixed loss against higher abilities. This implies that abilities $L_{\underline{k}}$ and $L_{\underline{k}+1}$ obtain the same payoff if and only if $\mu(L_{\geq \underline{k}+2})$ is equal to a specific value, but then strategy σ is not stable to a perturbation that changes the frequency of abilities $\{L_{\underline{k}}, L_{\underline{k}+1}\}$ while keeping $\mu(L_{\geq \underline{k}+2})$ fixed.

3. *We have $p > (A - 1)/A$.* 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 1 (otherwise the strategy is not stable to a perturbation that slightly increases the frequency of players that defect at horizon m). The inequality $p > (A - 1)/A$ implies 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 5:

1. Replacing pavlov with perfect grim trigger (defect if and only if 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 principle, one could adapt the mechanisms that lead to early-niceness in either Fudenberg and Maskin (1990), Binmore and 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 5 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 $c(L_4) < 1/A$, 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 $p < 1/((A-1) \cdot (k-2))$, then a strategy that includes only ability L_k is also a limit ESS.

7. Uniqueness with weaker solution concepts

Theorem 5 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 full support perturbed games.

Definition 9. The strategy $\sigma \in \Sigma$ is a *perfect NSS* (*symmetric perfect equilibrium*) if there exists a sequence of full support 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 any limit ESS is a perfect NSS (by Lemma 1); and any perfect NSS is a symmetric perfect equilibrium.

The following two formal 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.

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

$$\mu_{\underline{k}}^*(L_{\underline{k}}) = 1 - \frac{1 - (c(L_{\underline{k}+2}) - c(L_{\underline{k}}))}{p \cdot (A-1)}, \qquad \mu^*(L_{\underline{k}+2}) = 1 - \mu_{\underline{k}}^*(L_{\underline{k}})$$

$$\forall k \notin \{\underline{k}, \underline{k}+2\} \quad \mu^*(L_k) = 0$$

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

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

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

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

$$\left. (L_k \geq L_{\underline{k}+2} \text{ and } c(L_k) = c(L_{\underline{k}+2})) \right\}.$$

The following theorem 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.[22] Part (2) shows that any early-nice NSS is similar to a k-shifted variant of σ^*.[23] Part (3) shows that the uniqueness result essentially holds for early-nice perfect NSS.

THEOREM 6. *There exists $\bar{\delta} < 1$ such that for all $\delta \in (\bar{\delta}, 1)$, if $\forall k \in \mathbb{N}$, $c(L_{k+1}) - c(L_k) < 1$, $A > 3$, and $c(L_4) < 1/A$, the following statements hold:*

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

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

3. *If $p \notin \{\frac{1}{2}, 1\}$ and σ is an early-nice perfect NSS, then $\sigma \approx \sigma'$ for some $\sigma' \in \Sigma_1^*$. Moreover, if $p < A \cdot (1 - c(L_3))/(A-1)^2$ or $(A-1)/A < p$, 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? (3) Why do we interpret abilities in our model as representing limited foresight?

As argued by Osborne and Rubinstein (1994, Chapter 8.2):

---

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

[23]The result is stated for $p \notin 0.5$ and $p < (A-1)/A$. See footnote 29 for an additional strategy that may be an early-nice *perfect NSS when $p = 0.5$. When $p > (A-1)/A$ we can show that for each M, if $c(L_M)$ is sufficiently small, then any early-nice NSS includes ability L_M in its support, and if M_0 is sufficiently large and $c(L_{M_0})$ 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, and 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 automata with a limited number of states. With such a restriction, one can adapt our main results (Theorems 2–5) to this setup.

Finally, we discuss the interpretation of limited foresight in our model and compare it with the alternative notion of Jehiel (2001). The comparison between the two notions can be facilitated by considering a long two-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 nonfinal positions k steps ahead. When moving from chess-like games to non-zero-sum repeated games, the "position" is the history of play (due to its influence on the future behavior of the opponent). Jehiel's (2001) notion assigns a history-independent random value to all nonfinal states. In contrast, our notion bases the evaluation of nonfinal states on the history of play by using an "infinite-horizon benchmark": assuming that there is probability δ to end the game in any future round. In particular, consider an L_1 agent who plays against an opponent who follows pavlov. Jehiel's notion implies the counterintuitive prediction that the L_1 agent usually defects (and always defects if the randomness in the evaluation function is sufficiently small), while our notion implies that he cooperates until the last stage. As described in footnote 7, the experimental evidence from finitely repeated prisoner's dilemma games suggests that subjects behave in a way that 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 evolutionary literature (see, e.g., Dekel et al. 2007): an agent can observe his opponent's ability with a fixed exogenous probability and 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.[24] In Appendix A, 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 with 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 assumes that the cognitive cost function is increasing. It is relatively simple to extend the results to an environment in which players may play other games as long as the probability of playing games in which looking far ahead decreases efficiency (such as in the repeated prisoner's dilemma) is sufficiently high. The results can also be extended to deal with non-monotonic cost functions, which may represent the advantages of having higher abilities in other games. Specifically, if one assumes that the cognitive costs are not too high, then the game admits a strict limit ESS similar to σ^* except that L_3 is replaced with the ability that minimizes the cognitive cost in $L_{\geq 3}$ (and the playing rule remains the same as in σ^*).

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 $\{L_1, \ldots, L_M\}$. Assuming that $M \geq 3$, Theorem 2 holds in this setup. Theorem 5 holds for p's that are not too close to either 0 or 1. Assuming that $c(L_M)$ is sufficiently low, one can complete the characterization for all values of p: for low p's, 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 for high p's ($p > (A-1)/A$), 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 (Theorems 2–5) 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 the end of stage 0). 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

---

[24]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.

aware of 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: There is a weak correlation between signals of the two players.

Appendix A: 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 future research.

At stage 0, each player i makes three 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.[25] 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 increasing and concave in the first parameter, decreasing and convex in the second parameter, and submodular: $\partial^2 p(e^1, e^2)/(\partial e^1 \partial e^2) < 0$ (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 2 is extended to this setup as follows.

THEOREM 7. *Assume that* $\exists e_0 < 1/A - c(L_3)$ *such that* $\forall e \le e_0$, $A \cdot (1 - c(L_3))/(A-1)^2 < p(e, e) < (A - 1)/A$ *and*

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

Additionally assume that $c(L_4) > c(L_3) < 1/A$ *and* 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 Definition 2 and* $\mu^*(e^*)$ *is*

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

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

---

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

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 additional two assumptions are identical to Theorem 2) and the intuition as to why it implies the stability of $\sigma^*(e^*)$. The first assumption requires the existence of an effort level e_0 that satisfies three requirements. (I) Effort level e_0 is not too large. Observe that if $e < 1/A - c(L_3)$, then the total cost of an agent with ability L_3 who invests effort e is smaller than $1/A$ 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. (II) The probability $p(e, e)$ is not too close to 0 or 1 (the same bounds as in Theorem 2) 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 0 and 1, which is required for stability from the same reasons as in the basic model. (III) 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 than its marginal cost ($= 1$). This condition implies (by convexity and submodularity) that there exists a stable effort level $e^* < e_0$.

We conjecture that one could also adapt Theorem 5 to this setup.

Appendix B: Proofs

B.1 *Limit ESS and full support 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 full support 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 a full support perturbation ζ' such that $|\zeta - \zeta'| < \epsilon$, $\sigma' \in \Sigma$ is an ESS of the perturbed game $\Gamma(\zeta')$, and $|\sigma' - \sigma| < \epsilon$.*

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 full support perturbation ζ' as follows: if $\zeta(I)(a) > 0$, let $\zeta'(I)(a) = \zeta(I)(a)$; 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 when the horizon is equal to 1, 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 $\Gamma(\zeta')$. □

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

COROLLARY 1. *Let $\sigma \in \Sigma$ be a limit ESS. There exists a sequence of full support 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 full support perturbations $(\zeta'_n)_{n\in\mathbb{N}}$ with the same properties.   □

REMARK 4. The corollary immediately implies that every limit ESS is a perfect NSS (Definition 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).

B.2 *Theorem 2: σ^* is a strict limit NSS/ESS*

PROOF OF THEOREM 2. The proof includes several parts.

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

   (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 $\mathbf{T} - 3$ and defect at stage $\mathbf{T}$). Henceforth, we focus only on the payoffs of these three stages. In Table 2 we present 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 1 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_3's 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 points. If

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

TABLE 2. Reduced game (players choose abilities and must follow playing rule b^*).

both are unobserving (probability $(1 - p)^2$), they defect only at horizon 2 and obtain $A + 1$ points. If exactly one of them is observing, the observing player defects at horizon 3, and he obtains $A + 2$ points and his opponent obtains 1 point. Aggregating these possible outcomes yields the following expected pay-off 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 $c(L_4) > c(L_3)$.* The players obtain the same stage payoffs but they bear higher cognitive costs.

(c) *Ability L_2 obtains a strictly lower payoff than 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 if and only if

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

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

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

and the latter inequality is implied by $c(L_3) < 1/A$.

(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 if and only if

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

The latter inequality holds due to the assumption that $A \cdot (1 - c(L_3))/(A - 1)^2 < p$. 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

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

2. *Stability against other distributions.* If $(\mu, b^*) \neq \sigma^*$ is a best-reply to σ^*, then $u(\sigma^*, (\mu, b^*)) \geq u((\mu, b^*), (\mu, b^*))$ with a strict equality if $c(L_4) > c(L_3)$. Part 1 implies that (μ, b^*) is a best-reply to σ^* if and only if $\mathrm{supp}(\mu) \subseteq \{L_1, L_3\}$. 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* $\Gamma(\zeta)$. Let ζ be any full support perturbation with sufficiently small maximal tremble $M(\zeta)$. Let $\sigma_\zeta^* = (\mu_\zeta^*, \beta_\zeta^*)$ be defined as

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

where η is chosen such that $u((L_1, \beta_\zeta^*), \sigma_\zeta^*) = u((L_3, \beta_\zeta^*), \sigma_\zeta^*)$. Make the following observations:

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

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

   (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 and 2 still hold in this setup.

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

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

(a) *Playing rule 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 $B(\zeta)$. 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 that 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) for any full support perturbation and for the current setup in which players are informed in the last few rounds.

(b) *The playing rule 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 in the next round). This argument works also in larger horizons, and in steps (e) and (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)[26] a payoff vector of $A, A+1, 1$: A at horizon 3, $A+1$ at horizon 2 (as the L_1 opponent cooperates), and 1 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

---

[26]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 phrase "with high probability" in the remaining text.

stranger if and only if the frequency of L_3 opponents is sufficiently low:

$$\mu^*(L_3) \cdot 1 < (1 - \mu^*(L_3)) \cdot (A - 1) \quad \Leftrightarrow \quad \mu^*(L_3) < \frac{A - 1}{A}$$

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

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

(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 best reply if and only if $1 \cdot p < (1 - p) \cdot (A - 1) \Leftrightarrow p < (A - 1)/A$.

(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 that 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 $c(L_4) > c(L_3)$. □

B.3 *Theorem 5: Uniqueness result*

PROOF OF THEOREM 5. We begin with a some notation. 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 1, 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 1, strangers have incumbent abilities). Let $L_{\underline{k}} \in \text{supp}(\mu)$ be the lowest incumbent ability. Recall that an information set $I \in \mathcal{I}$ is *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 $m \leq \underline{k}$ 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 1. 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 1 and yields a higher payoff.

(b) With probability 1, players with ability $L_{\geq \underline{k}+1}$ 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 $L_{\geq \underline{k}+2}$ defect with probability 1 at horizon $\underline{k}+2$ when facing an observed incumbent ability $L_{\geq \underline{k}+1}$.

(c) Early-niceness implies that uninformed players cooperate with probability 1 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 nonincumbent ability.

(d) *If any incumbent ability defects with positive probability when facing an incumbent at a cooperative information set and if 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) We have $\mu(L_{\underline{k}}) < 1$. The assumption that $c(L_{\underline{k}+1}) - c(L_{\underline{k}}) < 1$ implies that if $\mu(L_{\underline{k}}) = 1$, then any strategy σ' that assigns mass 1 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 $p < (A-1)/A$ and all incumbents cooperate when the opponent is a stranger, the information set is cooperative, and the horizon is strictly larger than $\underline{k}+1$. Then the following scenarios occur:

(a) *All incumbents cooperate when the opponent is an incumbent, the information set is cooperative, and 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

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

(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, $\mu(L_{>\underline{k}+1}) > 0$ (otherwise, $u((L_{\underline{k}}, \beta), \sigma) < u((L_{\underline{k}+1}, \beta), \sigma)$ because $c(L_{\underline{k}+1}) - c(L_{\underline{k}}) < 1$) and $\forall k > \underline{k}+2$, $\mu(L_{>\underline{k}+2}) > 0$ only if $c(L_{\underline{k}}) = c(L_{\underline{k}+2})$ (otherwise $u((L_k, \beta), \sigma) < u((L_{\underline{k}+2}, \beta), \sigma)$ and σ cannot be an equilibrium).

(c) We have $\mu(L_{\underline{k}+1}) = 0$. Assume to the contrary that $\mu(L_{\underline{k}+1}) > 0$. The fact that σ is an equilibrium implies that $u((L_{\underline{k}}, \beta), \sigma) = u((L_{\underline{k}+1}, \beta), \sigma) = u((L_{\underline{k}+2}, \beta), \sigma)$. Analogous calculations to part 1(c) and (d) of Theorem 2's proof imply that $L_{\underline{k}}$ and $L_{\underline{k}+1}$ obtain the same payoff only if

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

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

Let μ' be defined as $\mu'(L_{\underline{k}}) = 0$, $\mu'(L_{\underline{k}+1}) = \mu(L_{\underline{k}}) + \mu(L_{\underline{k}+1})$ and let $\mu'(L_k) = \mu(L_k)$ for each $k \geq \underline{k} + 2$. The fact that $\text{supp}(\mu') \subseteq \text{supp}(\mu)$ implies that $u((\mu', \beta), \sigma) = u(\sigma, \sigma)$ and the equality $\mu(L_{\geq \underline{k}+2}) = \mu'(L_{\geq \underline{k}+2})$ implies $u(\sigma, (\mu', \beta)) = u((\mu', \beta), (\mu', \beta))$ (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).[27]

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

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

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

The argument is analogous to part 1(d) of the proof of Theorem 2.

(f) We have $L_{\underline{k}} = L_1$ (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 that 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 $\overline{k}$. The formal argument is as follows.[28] 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 $l \leq \overline{k}$ be the highest horizon in which an incumbent defects against an observed L_1. Assume to the contrary that either player cooperates with positive probability at any later

---

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

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

stage. Let $m \le l$ 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 full support 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 that faces an observed L_1 opponent at a cooperative information set cooperates if the horizon is larger than 2 and 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 a 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

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

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

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

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

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

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

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

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

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

$$\Leftrightarrow \quad (1 - x)^2 > 0 \quad \Leftarrow \quad 0 \le x < 1.$$

(iv) *If $L_{\underline{k}} = L_2$ and $c(L_4) < 1/A$, 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 un-observing $L_{>2}$ 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

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

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

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

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

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

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

and the last inequality is immediately implied by $c(L_4) < 1/A$.

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

(a) *We have $\mu(L_{\underline{k}}) \leq 1/A$.* 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

$$\mu(L_{\underline{k}}) \cdot (A - 1) \leq (1 - \mu(L_{\underline{k}})) \quad \Leftrightarrow \quad \mu(L_{\underline{k}}) \leq \frac{1}{A}.$$

(b) *All incumbent abilities $L_{\geq \underline{k} + 2}$ defect with probability* 1 *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 1 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) *We have $\mu(L_{\underline{k} + 1}) > 0$.* Assume to the contrary that $\mu(L_{\underline{k} + 1}) = 0$.

   (i) Assume that $p < (A - 2)/(A - 1)$. 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 obtains at least $A - 2$ more utility points when the opponent

is a stranger. This implies that $u((L_{\underline{k}}, \beta), (L_{\underline{k}}, \beta)) > u(\sigma, (L_{\underline{k}}, \beta))$ (which contradicts neutral stability) if

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

(ii) Due to analogous arguments to parts (a) and (b), $\mu(L_{\underline{k}}) \leq 1/A$ implies that all incumbent abilities $L_{\geq \underline{k}+3}$ defect with probability 1 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 $p \geq (A-2)/(A-1)$. To simplify notation, let $\alpha = \mu(L_{\geq \underline{k}+3})/\mu(L_{\geq \underline{k}+2})$ and $\mu = \mu(L_{\underline{k}})$. 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 at least $A - 2 + \alpha \cdot (A-1)$ more points when facing an unobserved $L_{\underline{k}}$ opponent (probability $(1-p) \cdot \mu$), one less point when facing an observed $L_{\underline{k}}$ opponent (probability $p \cdot \mu$), at least $A - 2 + \alpha \cdot (A-1)$ more points when facing an observing $L_{\geq \underline{k}+2}$ opponent (probability $p \cdot (1-\mu)$), at most one less point when facing an unobserved and unobserving $L_{\geq \underline{k}+2}$ opponent $((1-p)^2 \cdot (1-\mu))$, and 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 $u((L_{\underline{k}}, \beta), \sigma) > u(\sigma, \sigma)$ (which contradicts σ being a Nash equilibrium) if

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

Substituting $A > 3$ yields

$$\Leftarrow \quad (1 + 2 \cdot \alpha) \cdot (p \cdot (1-\mu) + (1-p) \cdot \mu)$$
$$\overset{?}{>} p \cdot \mu + (1-p) \cdot (1-\mu) \cdot (1 + p \cdot \alpha)$$
$$\Leftrightarrow \quad (2 \cdot p - 1) \cdot (1 - 2 \cdot \mu)$$
$$+ \alpha\big(2 \cdot (p \cdot (1-\mu) + (1-p) \cdot \mu) - (1-p) \cdot (1-\mu) \cdot p\big) \overset{?}{>} 0$$
$$\Leftrightarrow \quad (2 \cdot p - 1) \cdot (1 - 2 \cdot \mu) + \alpha(p \cdot (1-\mu) \cdot (1+p) + 2 \cdot (1-p) \cdot \mu) \overset{?}{>} 0.$$

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

(d) *We have*

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

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 $u((L_{\underline{k}}, \beta), \sigma) = u((L_{\underline{k}+1}, \beta), \sigma)$ (which is implied by σ being an equilibrium) if and only if

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

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

(e) *Strategy σ is not a limit ESS.* Let μ'' be defined as $\mu''(L_{\underline{k}}) = 0$, $\mu''(L_{\underline{k}+1}) = \mu(L_{\underline{k}}) + \mu(L_{\underline{k}+1})$, and $\mu''(L_k) = \mu(L_k)$ for each $k > \underline{k} + 1$. The inclusion $\text{supp}(\mu'') \subseteq \text{supp}(\mu)$ implies that $u((\mu'', \beta), \sigma) = u(\sigma, \sigma)$ and the previous part implies that $u(\sigma, (\mu'', \beta)) = u((\mu'', \beta), (\mu'', \beta))$ (because $\mu(L_{\geq \underline{k}+2}) = \mu''(L_{\geq \underline{k}+2})$). An analogous property is satisfied in any sufficiently close perturbed game, and thus σ cannot be a limit ESS.[29]

4. *Case III.* Assume that $p \geq (A-1)/A$. 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 3(b), implies that ability $L_{\tilde{k}}$ defects with probability 1 at horizon m when facing an observed $L_{\tilde{k}}$ opponent. Finally, an analogous argument to part 4(e) of the proof of Theorem 2 shows that $p \geq (A-1)/A$ 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 $m + 1$.[30]

□

B.4 *Other results*

PROOF OF PART 1 OF THEOREM 6 (Early-nice symmetric perfect equilibrium).

1. We begin by showing that $0 < \mu(L_1) < 1$. The preliminary observations and Case I of Theorem 5's proof also hold with minor adaptations 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 3(a) holds also in this setup and shows that $\mu(L_{\underline{k}}) \leq 1/A$. 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 $\mu(L_{\underline{k}})$), the same payoff when facing an unobserving $L_{>\underline{k}}$ opponent, and at least $A - 1$

---

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

[30]If $p = (A-1)/A$, then defection and cooperation yield the same payoff, and one has to rely also on an analogous argument to part 3(b) to imply the contradiction.

more points when facing an observing $L_{>\underline{k}}$ opponent (probability $p \cdot (1 - \mu(L_{\underline{k}}))$). Thus, $u((L_1, \beta), \sigma) > u((L_{\underline{k}}, \beta), \sigma) = u(\sigma, \sigma)$ (and this contradicts σ being an equilibrium) if

$$\mu(L_{\underline{k}}) < p \cdot (1 - \mu(L_{\underline{k}})) \cdot (A - 1)$$

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

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

   (a) Assume first that $\mu(L_{\leq 2}) \leq 1/A$. 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

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

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

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

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

   (b) Assume that $\mu(L_{\leq 2}) > 1/A$. By an analogous argument to part 4(e) of Theorem 2's proof, it implies that it is strictly better to cooperate at any horizon larger than 3 when facing a stranger at a cooperative information set. The assumption that $p < (A - 1)/A$ implies by an analogous argument to part 4(e) of Theorem 2's proof that it is strictly better to cooperate at any horizon larger than 4 when facing an incumbent at a cooperative information set. Thus, on the equilibrium path, all incumbents cooperate at all horizons larger than 4. This implies that if $c(L_5) > c(L_4)$, then all incumbents have ability of at most L_4. $\qquad\square$

The proofs of parts 2 and 3 of Theorem 6 are very similar to the analogous parts of the proof of Theorem 5 (omitted for brevity).

PROOF OF THEOREM 3 ((L_1, b_{defect}) is a strict limit ESS). 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 1) for any full support perturbation and for the current setup, in which players may become informed earlier about the realized length of the game.   □

PROOF OF THEOREM 4 (Each L_k can be the unique incumbent in a strict limit ESS). 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: The sequence $\ddot{W}$ is the "reflection" of $\dot{W}$, in which the roles of players 1 and 2 are exchanged, $\forall 1 \le j \le M$, $i \in \{1, 2\}$ $\dot{W}_j^i = \ddot{W}_j^{-i}$; the sequence $\dot{W}$ begins with defection, $\dot{W}_1 = D$; the sequence $\dddot{W}$ is a symmetric action profile that begins with mutual defection ($\dddot{W}_1 = (D, D)$); and the average stage payoffs are ordered as $\bar{u}(\dot{W}) > \bar{u}((S, S))$, $\bar{u}(\dot{W}) > \bar{u}(\dddot{W}) > 1$.

   (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: $\forall 1 \le j \le t$, $W_{t,j} = (S_{j \bmod M}, S_{j \bmod M})$. 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., $\ddot{W}$, $\dddot{W}$): $\forall 1 \le j \le t$, $\dot{W}_{t,j} = (\dot{W}_{j \bmod M'})$ (resp., $\ddot{W}_{t,j} = (\ddot{W}_{j \bmod M'})$, $\dddot{W}_{t,j} = (\dddot{W}_{j \bmod M'})$).

2. *Definition of the deterministic playing rule $b_{W,k}$*: At stage 1, $b_{W,k}(L_{k'}, l, s, \varnothing) = C$ if and only if ($s \in \{L_k, \phi\}$ and ($\exists k < l' < l$ s.t. $W_{t+1+l-l'}^i = C$)). 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 ($t \ge 1$), $b_{W,k}(L_{k'}, l, s = L_{k''}, (a^i, a^{-i})^t) = C$ if and only if any one of the following conditions holds:

$$\left((a^i, a^{-i})^t = W_t \text{ and } W_{t+1}^i = C \text{ and} \right.$$
$$(\exists l_1, l_2 \text{ s.t. } ((k < l_1 < l_2 < l) \text{ or } (k'' < l_1 < l)) \text{ and}$$
$$\left. W_{t+1+l-l_1}^{-i} = W_{t+1+l-l_2}^{-i} = C)\right) \quad \text{or}$$
$$\left((a^i, a^{-i})^t = \dot{W}_t \text{ and } \dot{W}_{t+1}^i = C \text{ and } L_{k''} \ne L_k \text{ and} \right.$$
$$(\exists l_1, l_2 \text{ s.t. } ((k < l_1 < l_2 < l) \text{ or } (k'' < l_1 < l)) \text{ and}$$
$$\left. \ddot{W}_{t+1+l-l_1}^{-i} = \ddot{W}_{t+1+l-l_2}^{-i} = C)\right) \quad \text{or}$$

$$\big((a^i, a^{-i})^t = \ddot{W}_t \text{ and } \ddot{W}^i_{t+1} = C \text{ and } L_{k'} \neq L_k \text{ and}$$

$$(\exists k'' < l_1 < l \text{ s.t. } \ddot{W}^{-i}_{t+1+l-l_1} = C)\big) \quad \text{or}$$

$$\big((a^i, a^{-i})^t = \ddot{W}_t \text{ and } \ddot{W}^i_{t+1} = C \text{ and } L_{k'}, L_{k''} \neq L_k \text{ and}$$

$$(\exists k'' < l_1 < l \text{ s.t. } \ddot{W}^{-i}_{t+1+l-l_1} = C)\big).$$

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 events 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 full support 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* $\sigma = (\mu, \beta) \neq \sigma_{W,k,\zeta} \in \Sigma(\zeta)$ *be 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 of Kim's (1994) folk theorem result and is briefly sketched as follows:

(a) *We have* $u((\mu, \beta), \sigma_{W,k,\zeta}) \leq u((\mu, \beta_{W,k,\zeta}), \sigma_{W,k,\zeta})$. 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) *We have* $u((\mu, \beta_{W,k,\zeta}), \sigma_{W,k,\zeta}) < u(\sigma_{W,k,\zeta}, \sigma_{W,k,\zeta})$ *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.   □

References

Andreoni, James and John H. Miller (1993), "Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence." *Economic Journal*, 103, 570–585. [206]

Axelrod, Robert M. (1984), *The Evolution of Cooperation*. Basic Books, New York. [206, 207]

Bendor, Jonathan and Piotr Swistak (1997), "The evolutionary stability of cooperation." *American Political Science Review*, 91, 290–307. [215]

Binmore, Kenneth G. and Larry Samuelson (1992), "Evolutionary stability in repeated games played by finite automata." *Journal of Economic Theory*, 57, 278–305. [218]

Bolton, Gary E. (1997), "The rationality of splitting equally." *Journal of Economic Behavior & Organization*, 32, 365–381. [205]

Bruttel, Lisa V., Werner Güth, and Ulrich Kamecke (2012), "Finitely repeated prisoners' dilemma experiments without a commonly known end." *International Journal of Game Theory*, 41, 23–47. [206]

Camerer, Colin F. (2003), *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton University Press, Princeton. [203]

Cooper, Russel, Douglas V. DeJong, Robert Forsythe, and Thomas W. Ross (1996), "Cooperation without reputation: Experimental evidence from prisoner's dilemma games." *Games and Economic Behavior*, 12, 187–218. [206]

Costa-Gomes, Miguel, Vincent P. Crawford, and Bruno Broseta (2001), "Cognition and behavior in normal-form games: An experimental study." *Econometrica*, 69, 1193–1235. [203]

Crawford, Vincent P. (2003), "Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions." *American Economic Review*, 93, 133–149. [208]

Cressman, Ross (1997), "Local stability of smooth selection dynamics for normal form games." *Mathematical Social Sciences*, 34, 1–19. [214]

Dekel, Eddie, Jeffrey C. Ely, and Okan Yilankaya (2007), "Evolution of preferences." *Review of Economic Studies*, 74, 685–704. [204, 208, 220]

Fudenberg, Drew and Eric Maskin (1990), "Evolution and cooperation in noisy repeated games." *American Economic Review: Papers and Proceedings*, 80, 274–279. [206, 215, 216, 218]

Geanakoplos, John and Larry Gray (1991), "When seeing further is not seeing better." *Bulletin of the Santa Fe Institute*, 6, 4–9. [207]

Haviland, William A., Harald Prins, Dana Walrath, and Bunny McBride (2007), *Cultural Anthropology: The Human Challenge.* Wadsworth, Belmont. [207]

Heller, Yuval (2014), "Stability and trembles in extensive-form games." *Games and Economic Behavior*, 84, 132–136. [213]

Jehiel, Philippe (2001), "Limited foresight may force cooperation." *Review of Economic Studies*, 68, 369–391. [207, 220]

Johnson, Eric J., Colin Camerer, Sankar Sen, and Talia Rymon (2002), "Detecting failures of backward induction: Monitoring information search in sequential bargaining." *Journal of Economic Theory*, 104, 16–47. [203, 206]

Kim, Yong-Gwan (1993), "Evolutionary stability in the asymmetric war of attrition." *Journal of Theoretical Biology*, 161, 13–21. [205]

Kim, Yong-Gwan (1994), "Evolutionarily stable strategies in the repeated prisoner's dilemma." *Mathematical Social Sciences*, 28, 167–197. [205, 207, 215, 237]

Kraines, David and Vivian Kraines (1989), "Pavlov and the prisoner's dilemma." *Theory and Decision*, 26, 47–79. [205]

Kreps, David M. and Robert Wilson (1982), "Sequential equilibria." *Econometrica*, 50, 863–894. [212, 219]

Kreps, David M., Paul Milgrom, John Roberts, and Robert Wilson (1982), "Rational cooperation in the finitely repeated prisoners' dilemma." *Journal of Economic Theory*, 27, 245–252. [207, 218]

Leimar, Olof (1997), "Repeated games: A state space approach." *Journal of Theoretical Biology*, 184, 471–498. [205]

Lorberbaum, Jeffrey (1994), "No strategy is evolutionarily stable in the repeated prisoner's dilemma." *Journal of Theoretical Biology*, 168, 117–130. [205, 207, 212]

Lorberbaum, Jeffrey P., Daryl E. Bohning, Ananda Shastri, and Lauren E. Sine (2002), "Are there really no evolutionarily stable strategies in the iterated prisoner's dilemma?" *Journal of Theoretical Biology*, 214, 155–169. [207, 213, 214, 227, 235]

Maynard Smith, John and George R. Price (1973), "The logic of animal conflict." *Nature*, 246, 15–18. [205, 211, 212]

Mengel, Friederike (2014), "Learning by (limited) forward looking players." *Journal of Economic Behavior & Organization*, 108, 59–77. [207]

Mohlin, Erik (2012), "Evolution of theories of mind." *Games and Economic Behavior*, 75, 299–318. [207, 218, 222]

Nachbar, John H. (1990), "'Evolutionary' selection dynamics in games: Convergence and limit properties." *International Journal of Game Theory*, 19, 59–89. [204, 210]

Nagel, Rosemarie C. (1995), "Unraveling in guessing games: An experimental study." *American Economic Review*, 85, 1313–1326. [203]

Neelin, Janet, Hugo Sonnenschein, and Matthew Spiegel (1988), "A further test of noncooperative bargaining theory: Comment." *American Economic Review*, 78, 824–836. [203]

Neyman, Abraham (1999), "Cooperation in repeated games when the number of stages is not commonly known." *Econometrica*, 67, 45–64. [207]

Nowak, Martin and Karl Sigmund (1993), "A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game." *Nature*, 364, 56–58. [205, 211]

Oechssler, Jörg and Frank Riedel (2001), "Evolutionary dynamics on infinite strategy spaces." *Economic Theory*, 17, 141–162. [215]

Okada, Akira (1981), "On stability of perfect equilibrium points." *International Journal of Game Theory*, 10, 67–73. [213]

Osborne, Martin J. and Ariel Rubinstein (1994), *A Course in Game Theory*. MIT Press, Cambridge, Massachusetts. [219]

Robson, Arthur J. (1990), "Efficiency in evolutionary games: Darwin, Nash, and the secret handshake." *Journal of Theoretical Biology*, 144, 379–396. [215]

Robson, Arthur J. (2003), "The evolution of rationality and the Red Queen." *Journal of Economic Theory*, 111, 1–22. [203]

Rosenthal, Robert W. (1981), "Games of perfect information, predatory pricing and the chain-store paradox." *Journal of Economic Theory*, 25, 92–100. [207]

Samuelson, Larry (1987), "A note on uncertainty and cooperation in a finitely repeated prisoner's dilemma." *International Journal of Game Theory*, 16, 187–195. [207]

Samuelson, Larry (1991), "Limit evolutionarily stable strategies in two-player, normal form games." *Games and Economic Behavior*, 3, 110–128. [205]

Sandholm, William H. (2010), "Local stability under evolutionary game dynamics." *Theoretical Economics*, 5, 27–50. [214]

Selten, Reinhard (1975), "Reexamination of the perfectness concept for equilibrium points in extensive games." *International Journal of Game Theory*, 4, 25–55. [205, 212, 224]

Selten, Reinhard (1983), "Evolutionary stability in extensive two-person games." *Mathematical Social Sciences*, 5, 269–363. [205, 212]

Selten, Reinhard (1988), "Evolutionary stability in extensive two-person games—Correction and further development." *Mathematical Social Sciences*, 16, 223–266. [212]

Selten, Reinhard and Rolf Stoecker (1986), "End behavior in sequences of finite prisoner's dilemma supergames: A learning theory approach." *Journal of Economic Behavior & Organization*, 7, 47–70. [203, 205, 206]

Stahl II, Dale O. (1993), "Evolution of smart-n players." *Games and Economic Behavior*, 5, 604–617. [207]

Stahl II, Dale O. and Paul W. Wilson (1994), "Experimental evidence on players' models of other players." *Journal of Economic Behavior & Organization*, 25, 309–327. [203]

Stennek, Johan (2000), "The survival value of assuming others to be rational." *International Journal of Game Theory*, 29, 147–163. [207]

Swinkels, Jeroen M. (1992), "Evolutionary stability with equilibrium entrants." *Journal of Economic Theory*, 57, 306–332. [212]

Thomas, Bernhard (1985), "On evolutionarily stable sets." *Journal of Mathematical Biology*, 22, 105–115. [212, 213, 214]

van Damme, Eric (1987), *Stability and Perfection of Nash Equilibria*. Springer-Verlag, Berlin. [212, 224]

van Veelen, Matthijs and Julián García (2010), "In and out of equilibrium: Evolution of strategies in repeated games with discounting." Discussion Paper 10-037/1, Tinbergen Institute. [207, 214]

Weibull, Jörgen W. (1995), *Evolutionary Game Theory*. MIT Press, Cambridge, Massachusetts. [215, 226]

Wu, Jianzhong and Robert Axelrod (1995), "How to cope with noise in the iterated prisoner's dilemma." *Journal of Conflict Resolution*, 39, 183–189. [206]