Three steps ahead

Heller, Yuval

University of Oxford, Department of Economics

13 June 2012

Three Steps Ahead

Yuval Heller[*] *(January 10, 2013)*

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

Abstract

We present an evolutionary model of a population interacting in repeated Prisoner's Dilemma. Each type is characterized by the number of steps he looks ahead, and each agent has an independent probability to observe the opponent's type. We show that if this probability is not too close to 0 or 1, then the evolutionary process admits a stable outcome, in which the population includes a mixture of "naive" agents who look 1 step ahead, and "moderately sophisticated" agents who look 3 steps ahead. Moreover, this outcome is unique under the additional assumption that agents present reciprocity at early stages of the interaction.

KEYWORDS: bounded forward-looking, evolutionary stability, Prisoner's Dilemma. JEL Classification: C73, D03.

1 Introduction

Experimental evidence suggests that people look only few stages ahead and use backward induction reasoning to a limited extent. 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 "Centipede" game (McKelvey and Palfrey (1995); Nagel and Tang (1998)), and they ignore future opportunities that are more than 1-2 steps ahead when interacting in sequential bargaining (Neelin, Sonnenschein, and Spiegel (1988)). A second stylized fact is the heterogeneity of the population: some people systematically look fewer

[*] I would like to express my deep gratitude to Itai Arieli, Vince Crawford, Eddie Dekel, Eric Mohlin, Ariel Rubinstein, Peyton Young, seminar participants at Birmingham, Oxford, and University College London, and workshop participants in fifth Transatlantic Theory Workshop at Northwestern University, for many useful comments, discussions, and ideas.

	C	D
C	A,A	0,A+1
D	A+1,0	1,1

Tab. 1: Payoff at the symmetric stage game Prisoner's Dilemma ($A > 3.15$).

steps than others (see, e.g., Johnson, Camerer, Sen, and Rymon (2002)).[1]
These observations raise two related evolutionary puzzles. The first puzzle is why people only look few steps ahead. In many games, the ability to look one more step than your opponent gives 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 more steps ahead throughout the evolutionary process (the so called "red queen effect"; Robson (2003)). The second puzzle is how the "naive" people, who systematically look fewer steps ahead, survive.

In this paper we present an evolutionary model where agents, who differ in their forward looking ability, play repeated Prisoner's Dilemma. We characterize a stable heterogeneous population of naive agents (who look 1 step ahead) and moderately sophisticated agents (who look 3 steps ahead). Moreover, we show that under an additional assumption of reciprocal behavior at early stages of the interaction, this is the *unique* stable population.

In each generation agents from a large population are randomly matched and each couple plays (without rematching) repeated Prisoner's Dilemma. The stage payoffs are described in Table 1: mutual cooperation (both players play C) yields both players A, mutual defection (both players play D) gives 1, and if a single player defects, he obtains $A+1$ and his opponent gets 0. The length of the interactions at each generation, $\mathbf{T}$, has a geometric distribution with parameter λ. In what follows, we assume that mutual cooperation is sufficiently efficient ($A > 3$), and that the interaction is long enough (λ close to 1).[2]

Each agent in our model has a type in the set $\{L_1, L_2, ..., L_M, L_\infty\}$ that determines how many steps he looks ahead (results are independent of M, given that $M \geq 3$). Agents of type L_∞ are informed about the realized length of the interaction before the game begins. An agent of type L_k is informed about the realized length k periods before the end (after playing at stage $\mathbf{T} - k$). We interpret this information structure to stem from bounded forward-lookingness: type L_k becomes aware of the realized final period and its strategic

[1] Similar stylized facts are also observed with respect to the number of strategic iterations that players use in static games, as suggested by the cognitive hierarchy (or level-k) models (see, e.g., Stahl and Wilson (1994); Nagel (1995); Costa-Gomes, Crawford, and Broseta (2001); Camerer, Ho, and Chong (2004)).

[2] In order to simplify the presentation of the results, we assumed that: (1) defection yields the same additional payoff (relative to cooperation) regardless of the opponent's strategy, and (2) all interactions at each generation has the same length. The results remain qualitatively similar also without these assumptions.

"backward-induction" implications only k rounds before the end.

We assume that types are partially observable in the following way (similar to Dekel, Ely, and Yilankaya (2007)): before the interaction begins, each agent has an independent probability p to observe his opponent's type.[3] Informally, this can be interpreted as an opportunity to observe opponent's past behavior, or to observe a trait that is correlated with the forward-looking ability. The total payoff of an agent of type L_k is the undiscounted sum of payoffs in the repeated Prisoner's Dilemma minus an arbitrarily small cost that is increasing in k (a marginal cost for having a better forward-looking ability).

We capture the stable points of the dynamic evolutionary process by adapting the notion of evolutionarily stable strategy (ESS, Maynard-Smith (1974)) to a setup with different types. In such a setup, the state of the population is described by a *configuration* - a pair consisting of a distribution of types and the (possibly mixed) strategy that each type uses in the game. A configuration is *evolutionarily stable* if any sufficiently small group of mutants who invades the population is outperformed by the incumbents in the post-entry population.[4]

Our first result shows that if p is not too close to 0 and 1 (and this interval is increasing in A), then there exists an evolutionarily stable configuration, which includes two kinds of players: (1) *naive* agents of type L_1 who only begin defecting at the last stage (their proportion increases in both p and A), (2) *moderately sophisticated* agents of type L_3: usually they defect two stages before the end, unless they observe that their opponent is sophisticated, and, in this case, they begin defecting one stage earlier. Stability relies on the balance between the direct disadvantage of naive agents (defecting too late), and the indirect advantage - when naivety is observed by a moderately sophisticated opponent, it serves as a commitment device that allows an additional round of mutual cooperation. Higher types (L_4 or more) cannot invade the population because L_3's behavior remains the best-reply also when being informed earlier about the realized length of the interaction.

As common in evolutionary models, this environment admits other stable configurations. Our second result characterizes an additional assumption - *early-reciprocity*, under which, this is the *unique* stable configuration.[5] Specifically, we assume that as long as an agent is uninformed about the realized length of the interaction he must: (1) be "nice" - never defect before his opponent, and (2) "retaliate" - defect if the opponent defected in the previous stage. Two examples for such heuristics are "tit-for-tat" (defect if the opponent defected in

[3] Results remain the same if agents were able to observe only lower opponents' type (see remark 2).

[4] The "mutants" achieve the same payoff if they are *equivalent* to the incumbents: have the same distribution of types and play the same on-equilibrium path. If they are not equivalent, we require the mutants to obtain a strictly lower payoff.

[5] In addition, we refine evolutionary-stability by requiring stable configurations to satisfy "properness" (Myerson (1978)). This restricts the incumbents to use "reasonable" strategies against external mutants.

the previous stage) and "grim" (defect if the opponent ever defected in the past).

The plausibility of this assumption relies on Axelrod (1984)'s findings that reciprocal behavior is very successful in tournaments of infinitely repeated Prisoner's Dilemma. In addition, a support for this assumption and for the predicted behavior in our model, is given in experiments that study behavior in finitely-repeated Prisoner's Dilemma. Selten and Stoecker (1986) study games with 10 rounds and show: (1) if any player defected, then almost always both players defect at all remaining stages, (2) usually there is mutual cooperation in the first 6 rounds, and (3) players begin defecting in the last 1-4 rounds.[6] Johnson, Camerer, Sen, and Rymon (2002)'s findings suggest that bounded forward-lookingness is the main cause for this behavior.

It is interesting to note that stable configurations are very different when p is close to 0 or 1. In both cases (assuming early-reciprocity), stable configurations must include L_∞ agents who, when facing other L_∞ agents, defect at all stages. When p is close to 0, types are too rarely observed, and the indirect advantage of naive agents is too weak. When p is close to 1, there is an "arms race" between sophisticated agents who observe each other: each such agent wishes to defect one stage before his opponent.

Our formal analysis deals only with repeated Prisoner's Dilemma. It is straightforward to extend the results to other games in which looking far ahead decreases efficiency. One example for such games is "centipede" (Rosenthal (1981)), which can represent sequential gift exchange. Such interactions are important both in primitive *hunter-gatherer* societies (see, e.g., Haviland, Prins, and Walrath (2007), p. 440), as well as in modern societies.

We conclude by briefly surveying the related literature. 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 (1993); Stennek (2000) and Mohlin (2012) present evolutionary models of bounded strategic reasoning ("level-k"), which are related to our model when $p = 0$ or $p = 1$. This paper is novel in introducing partial observability in this setup, and showing that it yields qualitative different results. Crawford (2003) studies zero-sum games with "cheap talk" and show that naive and sophisticated agents may co-exist and obtain the same payoff. Finally, Mengel (2012) assumes bounded forward looking, and demonstrates in an interaction of repeated Prisoner's Dilemma that it can induce cooperative behavior while both myopic play and unlimited forward looking behavior only induce defections.

[6] In the experiment subjects engaged in 25 sequences ("super-games") of repeated Prisoner's Dilemma. The above results describe the behavior of subjects in the last 13 sequences (after the initial 12 sequences in which players are inexperienced and their actions are "noisier"). See similar results in Andreoni and Miller (1993); Cooper, DeJong, Forsythe, and Ross (1996); Bruttel, Güth, and Kamecke (2012).

The paper is structured as follows. Section 2 presents our model. In Section 3 we show the stability of the configuration in which L_1 and L_3 co-exist. In Section 4 we prove uniqueness under the assumption of early-reciprocity, and characterize the stable configurations for low and high p-s. Appendix A compares our notion of stability with Dekel, Ely, and Yilankaya (2007), and show that our results are similar also with their notion, and Appendix B includes the formal proofs.

2 Model

2.1 Payoffs, Strategies and Types

We consider a large population in which at each generation agents are randomly matched, and each pair of agents play the repeated Prisoner's Dilemma with a geometric random length: $\mathbf{T}$. The stage game includes two actions: $\{C, D\}$. The stage-payoffs are described in Table 1. As is standard in the evolutionary literature, this payoff is interpreted as representing "success" or "fitness." To simplify the presentation of the results we assume that: (1) the interaction lasts at least 3 rounds - $\mathbf{T}-2\sim\text{Geo}\,(\lambda)$; and (2) at each generation all interactions have the same length (while lengths in different generations are independent). The results remain qualitatively the same without these assumptions.

Agents in the population differ in their forward-looking ability, which is captured by their type. Fix an arbitrary integer $M \geq 4$, and let $\mathcal{L} = \{L_1, ..., L_M, L_\infty\}$ be the set of types.[7] An agent of type L_k is informed about the realized length of the interaction after playing at round $\mathbf{T} - k$, or at the beginning of the interaction if $\mathbf{T} < k$. We dub agents as *uninformed* before they receive the signal about the realized length, and as *informed* afterwords.

Remark 1. We set L_1 to be the minimal type in the model. However, our results are robust to this choice. Specifically, if the minimal type in $\mathcal{L}$ were L_k (for some $k \in \mathbb{N}$ including $k = 0$) instead of L_1, then all our results would hold for a "shifted" unique stable configuration, (μ^*, b^*), where L_1 is replaced by L_k and L_3 is replaced by L_{k+2}.

Let $c : \mathcal{L} \to \mathbb{R}^+$ be a strictly increasing function satisfying $c\,(L_1) = 0$, and let $\delta > 0$. Agents of type L_k bear a *cognitive cost* of $\delta \cdot c\,(L_k)$. The payoff of the repeated game is the undiscounted sum of the stage payoffs minus the cognitive cost. In what follows we focus on the case of: (1) arbitrarily low cognitive costs - sufficiently small δ, (2) long interaction - λ is close enough to 1, and (3) efficient mutual cooperation - A is large enough ($A > 3.2$ for stability, and $A > 4.6$ for uniqueness).

[7] The results also hold for $M = 3$ but this make the notations of the proof more cumbersome.

Following Dekel, Ely, and Yilankaya (2007), we assume that before the interaction begins each player observes the type of his opponent with probability p (and gets no information about his opponent's type with probability $1-p$), independently of the event that his opponent observes his type.[8] We use the term *stranger* to describe an opponent whose type is not observed.

Remark 2. In some environments it might be plausible to assume that agents only identify lower opponent's types (see, Mohlin (2012)). All of our results remain the same with asymmetric type observability, where the informative signal (obtained with probability p) is:

1. The opponent's exact type, if it is strictly lower than the agent's type.

2. If the opponent's type is weakly higher, then the agent only observes this fact.

A history of the game of length $t \geq 0$ is a triple $\left(L_{k'}, l, (a_i, a_{-i})^t\right)$where: (1) $L_{k'} \in \mathcal{L} \cup \phi$ describes the signal about the opponent's type (ϕ denotes a stranger); (2) $l \in \{1, ..., M, \infty\}$; $l = \infty$ describes the case of an uninformed agent - the number of remaining stages (dubbed, the *horizon*) is unknown, and $l < \infty$ describes the length of a known horizon; and (3) $(a_i, a_{-i})^t \in \{C \times D\}^t$ describe the t action-profiles that were observed so far in the game. Let H_t denote the set of all histories of length t, and let $H = \cup_{t \in \mathbb{N}} H_t$ be the set of all histories. A *pure strategy* (resp., *behavioral strategy*) is a function $b : H \to \{C, D\}$ (resp., $\beta : H \to \Delta(\{C, D\})$ from the set of histories to the set of pure (resp., mixed) actions.

2.2 Configurations

Following the "indirect evolutionary approach" (Güth and Yaari (1992)) we present a reduced-form static analysis of a dynamic process that describes the evolution of types.[9] This process can be interpreted as either: (1) biological process - types are genetically determined, and payoff is the number of offspring; and (2) learning and imitation process - an agent's type describes the way he perceives strategic interactions; once in a while an agent may decide to change his strategic framework and imitate the type of a more successful agent.

Given a distribution of types $\mu \in \Delta(\mathcal{L})$ let $C(\mu) \subseteq \mathcal{L}$ denote the support of μ (types with positive frequency). Types in $C(\mu)$ are called *incumbents*, and types outside $C(\mu)$ are called *external mutants*. The state of the population is described by a *configuration* - a pair consisting of a distribution of types and their strategy profile. Formally:

[8] The results remain qualitatively similar if the signal structure is slightly altered by: (1) a small positive correlation with the opponent's signal, or (2) a small probability that the informative signal is incorrect.

[9] The indirect approach was mainly used to study evolution of preferences. See Frenkel, Heller, and Teper (2012) for a previous adaptation of this approach to study evolution of cognitive biases.

Definition 1. *Configuration* (μ, β) is a pair where $\mu \in \Delta(\mathcal{L})$ is a distribution of types, and $\beta = (\beta_k)_{k \in C(\mu)}$ is the profile of behavioral strategies of the incumbents.

Remark 3. Note that a configuration also determines the behavior against external mutants. In Appendix A we show that our results remain qualitatively similar also with Dekel, Ely, and Yilankaya (2007)'s alternative notion, in which h the state of the population determines only the strategies that are used against incumbents.

Next, we define the mixture of two configurations as follows:

Definition 2. Let (μ, β) and (μ', β') be configurations, and let $0 < \epsilon < 1$. The *mixture configuration* $\left(\tilde{\mu}, \tilde{\beta}\right) = (1-\epsilon) \cdot (\mu, \beta) + \epsilon \cdot (\mu', \beta')$ is: $\tilde{\mu} = (1-\epsilon) \cdot \mu + \epsilon \cdot \mu'$, and:

$$\forall k \in C(\tilde{\mu}), \ \tilde{\beta}_k = \frac{(1-\epsilon) \cdot \mu(L_k) \cdot \beta_k + \epsilon \cdot \mu'(L_k) \cdot \beta'_k}{(1-\epsilon) \cdot \mu(L_k) + \epsilon \cdot \mu'(L_k)}.$$

When ϵ is small we interpret $(1-\epsilon) \cdot (\mu, \beta) + \epsilon \cdot (\mu', \beta')$ as a *post-entry configuration* after incumbents in state (μ, β) are invaded by ϵ mutants with configuration (μ', β'). Finally, we define two configurations as equivalent if they have the same distribution and they induce the same observed play. Formally:

Definition 3. Configurations (μ, β) and (μ', β') are equivalent $((\mu, \beta) \approx (\mu', \beta'))$ if: (1) $\mu = \mu'$, and (2) for each pair of incumbents $L_k, L_{k'} \in C(\mu)$, the observed play when type L_k plays against type $L_{k'}$ is the same in both configurations.

Note that that following the invasion of ϵ mutants, the incumbents in each of two equivalent configurations may act differently when facing these mutants.

3 Evolutionary Stability

3.1 Definition

In a model without types, the state of the population is described by a strategy. A strategy is neutrally (resp., evolutionarily) stable if any sufficiently small group of mutants who invades the population and plays an arbitrary strategy would achieve a weakly (strictly) lower payoff than the incumbents. Formally:

Definition 4. (Maynard-Smith (1974); Maynard Smith (1982)) Strategy $\sigma \in \Sigma$ is neutrally (resp., evolutionarily) stable if for any strategy σ' (resp., $\sigma' \neq \sigma$) there exists $\epsilon_{\sigma'} \in (0,1)$ such that for every $0 < \epsilon < \epsilon_{\sigma'}$: $u(\sigma, \epsilon\sigma' + (1-\epsilon)\sigma) \geq u(\sigma', \epsilon\sigma' + (1-\epsilon)\sigma)$. (resp., $u(\sigma, \epsilon\sigma' + (1-\epsilon)\sigma) > u(\sigma', \epsilon\sigma' + (1-\epsilon)\sigma)$).

In what follows we extend the notion of evolutionary stability from strategies to configurations. Given two configurations (μ,β) and (μ',β') define $u\left((\mu,\beta),(\mu',\beta')\right)$ as the expected payoff of a player from population (μ,β) who plays against an opponent from population (μ',β') (and the type of each player is observed with independent probability p). A configuration is *neutrally (evolutionarily) stable* if any sufficiently small group of mutants who invades the population would obtain a weakly (strictly) lower payoff than the incumbents in the post-entry population. Formally:

Definition 5. Configuration (μ,β) is *neutrally* (resp., *evolutionarily*) *stable* if for any configuration (μ',β') (resp., any $(\mu',\beta')\not\approx(\mu,\beta)$) there exists $\epsilon_{\sigma'}\in(0,1)$ such that $\forall\, 0<\epsilon<\epsilon_{\sigma'}$:

$$u\left((\mu,\beta),\epsilon\left(\mu',\beta'\right)+(1-\epsilon)\left(\mu,\beta\right)\right)\geq u\left((\mu',\beta'),\epsilon\left(\mu',\beta'\right)+(1-\epsilon)\left(\mu,\beta\right)\right)$$

$$(\text{resp., } u\left((\mu,\beta),\epsilon\left(\mu',\beta'\right)+(1-\epsilon)\left(\mu,\beta\right)\right)>u\left((\mu',\beta'),\epsilon\left(\mu',\beta'\right)+(1-\epsilon)\left(\mu,\beta\right)\right)).$$

Remark 4. Note that:

1. Evolutionarily stable configurations are only weakly stable against invasions of equivalent mutants.
2. Definition 5 is closely related to Maynard Smith (1982)'s Definition 4 in two ways:
   (a) When the set of types is a singleton, then Definition 5 and Definition 4 coincide.
   (b) Consider a two-player "meta-game" in which each player chooses a type and a strategy for that type. Note that a mixed "meta-strategy" in this game is a configuration. A symmetric strategy profile in this "meta-game" is a neutrally stable strategy if and only if it is a neutrally stable configuration.[10]
3. Similar to the standard setup without types (see, Taylor and Jonker (1978)), neutral stability implies (Lyapunov) dynamic stability: no small change in the population can take it away from a neutral stable configuration in any payoff-monotonic dynamics.

It is well known that any neutrally stable strategy is a Nash equilibrium. Similarly (see Proposition 2 in Appendix B.1) strategy profile β in a neutrally stable configuration (μ,β) is: (1) *balanced* - all incumbents obtain the same payoff, and (2) a Bayes–Nash equilibrium in the Bayesian game with distribution μ.

[10] An evolutionarily stable configuration may be only a neutrally stable strategy in the "meta-game" as "meta-strategies" that only deviate off-equilibrium path yield the same payoff as the incumbents in the post-entry population.

3.2 Stability Result

Our first result characterizes an evolutionarily stable configuration, (μ^*, b^*), in which *naive* players (type L_1) and *moderately sophisticated* players (type L_3) co-exist. Let the configuration (μ^*, b^*) be defined as follows:

1. The population includes only types L_1 and L_3 with the following frequencies:

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

2. Uninformed agents play *grim*: defect if and only if the opponent has ever defected.

3. Informed L_1 agents defect at the last stage. Informed L_3 players:

   (a) Against an observed type different from L_1: defect at the last three stages.

   (b) Against strangers and observed L_1: Follow "grim" at horizon=3, and defect at the last two stages.

Our first results shows that (μ^*, b^*) is stable if p is not too close to 0 or 1.

Theorem 1. *Let $\frac{A}{(A-1)^2} < p < \frac{A-1}{A}$, let $\delta > 0$ be sufficiently small and let $\lambda < 1$ be sufficiently large. Then (μ^*, b^*) is evolutionarily stable.*[11]

The formal proof appears in Appendix B.2. In what follows we briefly sketch its outline. First, we show that b^* is a Bayes–Nash equilibrium (given μ^*), and that following *grim* till the last three rounds is also a best reply for informed mutants of higher types (if p is not too close to 1). Next, we show that (μ^*, b^*) is balanced. In order to show this, we compare the fitness of L_1 and L_3 agents against different opponents. L_1 agents succeed more against an observing L_3 opponent (who observed their type), because their observed naivety induces an additional round of mutual cooperation. L_3 agents achieve a better payoff in the two other cases: against naive opponents and against an unobserving sophisticated opponent. This implies that there is a unique level of $\mu(L_1)$ that balances the payoff of the two kinds of players (if p is not too close to 0).

Finally, we use these two properties to show resistance to mutations. If ϵ more players of type L_1 (L_3) join the populations, then due to the previous arguments, they would have a strictly lower payoff than the incumbents (on average). Mutants of type L_2 are outperformed due to their inability to defect one stage earlier against an observed type L_3. Mutants of types L_4 or more are outperformed due to their higher cognitive costs.

[11] Note that the interval $\frac{A}{(A-1)^2} < p < \frac{A-1}{A}$ is increasing in A, it converges to the entire interval $[0,1]$ in the limit $A \to \infty$,and it is non-empty for each $A > 3.2$.

4 Uniqueness

Similar to other models of repeated interactions our environment admits additional stable configurations. In this section we show that if we further assume that uninformed agents present reciprocal behavior, then (μ^*, b^*) is the *unique* stable configuration.

4.1 Reciprocal Behavior

In an early-reciprocal strategy uninformed agents: (1) never defect first ("nice"), and (2) defect if the opponent has defected in the last stage ("retaliate"). Formally:

Definition 6. Behavioral strategy β is an *early-reciprocal strategy* if :

1. $\beta\left(\cdot,\infty,(a_i,a_{-i})^t\right)=C$ if $a_{-i}^{\tau}=C$ for each $\tau\leq t$ (*nice*); and
2. $\beta\left(\cdot,\infty,(a_i,a_{-i})^t\right)=D$ if $a_{-i}^{t}=D$ (*retaliate*).

As discussed in the introduction, reciprocal behavior includes a family of heuristics (such as, "tit-for-tat" and "grim"), which are very successful in experimental and simulated plays of infinitely repeated Prisoner's Dilemma (e.g., Axelrod (1984)). In what follows we assume that all agents (both incumbents and mutants) are limited to playing only early-reciprocal strategies (dubbed, *early-reciprocity assumption*).

4.2 Properness Refinement

Even with the early-reciprocity assumption, the interaction admits additional evolutionarily stable configurations. One such configuration is described in the following example.

Example. Consider the configuration that assigns mass 1 to L_∞ agents who defect at all stages against any observed opponent's type. One can see that this configuration is evolutionarily stable. However, the stability relies on the incumbents defecting at all stages against naive mutants (L_1). Such a strategy is strictly dominated by an alternative strategy that cooperates for the first $M-2$ stages against naive opponents (due to the early-reciprocity assumption). Thus, in the long run, as a response to recurrent entrees of naive mutants, incumbents are expected to evolve into cooperating at the first stages of the game when facing naive opponents, and the stability of the configuration will be lost.

Motivated by this example, we refine evolutionary stability by requiring properness (Myerson (1978)). We begin by formally defining properness in this setup. A configuration is interior if every combination of type and pure strategy has a positive probability. With some abuse of notation: (1) we denote by L_k also the distribution that assigns mass 1 to type L_k; and (2) we consider the behavioral strategy β as a mixed strategy, and denote by $\beta(b)$ the probability of the pure strategy b.

Definition 7. Configuration (μ, β) is *interior* if for each types $L_k \in \mathcal{L}$ and for each pure early-reciprocal strategy b: (1) $\mu(L_k) > 0$, and (2) $\beta(b) > 0$.

Given $\epsilon > 0$, an interior strategy configuration (μ, β) is *ϵ-proper* if for every type $L_k L_{k'} \in \mathcal{L}$ and every pure early-reciprocal strategy $b_k, b_{k'}$:

$$u\left(\left(L_k, b_k\right), (\mu, \beta)\right) < u\left(\left(L_{k'}, b_{k'}\right), (\mu, \beta)\right) \Rightarrow \mu\left(L_k\right) \cdot \beta\left(b_k\right) \leq \epsilon \cdot \mu\left(L_{k'}\right) \cdot \beta\left(b_{k'}\right).$$

A configuration is proper if it is the limit of some ϵ-proper equilibria when $\epsilon \rightarrow 0$.

Definition 8. Configuration (μ, β) is proper, if there exists a sequence $\left(\epsilon_n > 0\right)_{n \geq 1} \rightarrow 0$ and a sequence of ϵ_n-proper interior configurations (μ^n, β^n) such that $\mu^n \rightarrow_{n \rightarrow \infty} \mu$, and if $\mu(L_k) > 0$ then $\beta_k^n \rightarrow_{n \rightarrow \infty} \beta_k$.

Remark 5. Note that:

1. The configuration in the example above is not proper because in any interior configuration, cooperating in the first $M - 2$ stages strictly dominates early defection against observed L_1. This implies that always defecting against type L_1 cannot be played with positive probability in a proper configuration.

2. It is immediate to see that in every proper configuration (μ, β) the profile β is a balanced Bayes–Nash equilibrium.

3. Definition 8 is closely related to Myerson's (1978) definition of proper equilibrium:

   (a) If there is a single type, then Definition 5 and Myerson's definition coincide.

   (b) Consider again the two-player "meta-game" in which each player chooses a type and an early reciprocal strategy. A symmetric (mixed) strategy profile in this meta-game is a proper equilibrium if and only if it is a proper configuration.

4. In the proof of our uniqueness result (Theorem 2) we only use a weaker property that is implied by properness: the requirement that the strategy that an incumbent plays against an external mutant must be a best-reply to some strategy of the external mutant, which is best-reply to the incumbent configuration.

A configuration is a *proper naturally (evolutionarily) stable* if it is both proper and naturally (evolutionarily) stable. van Damme (1987) showed for normal-form games without types that evolutionary stability implies properness. In our setup, this does not hold, because we weakened evolutionary stability by allowing equivalent mutants to fare the same as the incumbents. Thus, we have to slightly enhance evolutionary stability by explicitly require properness.

4.3 Uniqueness Result

It is straightforward to show that (μ^*, b^*) is proper (Prop. 3, proved in Appendix B.3). Our next result shows that with the early-reciprocity assumption any proper neutrally stable configuration is equivalent to (μ^*, b^*).[12]

Theorem 2. *Let $A > 4.6$, $\frac{A}{(A-1)^2} < p < \frac{A-1}{A}$, $\delta > 0$ be sufficiently small, $\lambda < 1$ be sufficiently large, and assume early-reciprocity. Then if (μ, β) is a proper neutrally stable configuration, then it is equivalent to (μ^*, b^*).*[13]

The sketch of the proof is as follows (see Appendix B.4 for the formal proof). First, observe that a configuration with a single type is not stable: (1) if the type is L_∞, then the entire population defects all the time, and mutants of type L_1 induce cooperation against them and, in doing so, outperform the incumbents; and (2) if the type is $L_k \neq L_\infty$, then mutants of type L_∞ can invade the population. Let L_{k_1} be the smallest ("naive") type in the population. Then, it is immediate to see that type L_{k_1} must always defect when the horizon is at most k_1, and all other types must defect when the horizon is at most $k_1 + 1$.

The next step is to show that a large fraction of the non-naive population must cooperate at all horizons larger than $k_1 + 1$ when facing strangers. Otherwise, a small increase in the frequency of the naive players (type L_{k_1}) would improve their fitness relative to the non-naive agents (as many non-naive agents defect too early against unobserved naive opponents), and this implies instability. The fact that this fraction is so large implies that if there are non-naive players who defect at earlier horizons than $k_1 + 1$ against strangers, then: (1) the large fraction who defects at horizon $k_1 + 1$ against strangers must belong to type L_{k_1+1}, and (2) all the remaining non-naive players must defect at horizon $k_1 + 2$ against strangers. This characterization allows us to find the unique distribution of types that satisfies the balance

[12] The assumption $A > 4.6$ is required to have uniqueness in the entire interval $\frac{A}{(A-1)^2} < p < \frac{A-1}{A}$. For lower A-s the uniqueness may hold only in a sub-interval (as discussed in the proof in Appendix B.4).

[13] The uniqueness result holds also for $A < 4.6$ but in a smaller interval as detailed Corollary 1 in Appendix B.4.

of payoffs among the different types, but it turns out that this distribution is not stable against small perturbations in the frequency of the incumbents.

Finally, if all non-naive players defect at horizon k_1+1 against strangers, then it implies that they all defect at horizon k_1+2 against observed non-naive opponents, and the balance between the payoffs of the different types implies that the frequency of naive and non-naive players is the same as in μ^*. Finally, we show that if $k_1>1$, then the configuration can be invaded by mutants of type L_1, who would outperform the incumbents by inducing more mutual cooperation when being observed by the opponents.

4.4 Low and High p-s

Our main results (Theorems 1-2) characterized the unique stable configuration in the interval $\frac{A}{(A-1)^2}<p<\frac{A-1}{A}$. In this section we deal with the remaining intervals: low p-s (below $\frac{A}{(A-1)^2}$) and high p-s (above $\frac{A-1}{A}$), and show that the stable configurations are qualitatively different at these intervals. In both cases, stable configurations (if they exist) must include L_∞ players who, when facing L_∞ opponents, defect at all stages.

When p is close to 0, this occurs because the indirect advantage of lower types is too small and they cannot exist in a stable configuration (because the probability of being observed by the opponent is too low). When p is close to 1, there is an "arms race" between sophisticated agents who observe each other: each such agent wishes to defect one stage before his opponent. The result of this "arms race" is that in any stable configuration there must be L_∞ agents in the population, and these L_∞ players defect at the first stage when they observe a L_∞ opponent. Table 2 summarizes this result, which is formalized as follows (see proof in Appendix B.5):

Theorem 3. *Let $A>4.6$, $\delta>0$ be sufficiently small, $\lambda<1$ be sufficiently large, and assume early-reciprocity. Let (μ,β) be a proper neutrally stable strategy.*

1. *Let $p<\frac{A}{(A-1)^2}$. Then $\mu\left(L_\infty\right)=1$ and everyone defects at all stages.*
2. *Let $\frac{A}{(A-1)^2}<p<\frac{A-1}{A}$. Then $(\mu,\beta)\approx(\mu^*,b^*)$.*
3. *Let $\frac{A-1}{A}<p\leq 1$. Then $\mu\left(L_\infty\right)>0$, and L_∞ agents always defect against observed L_∞ opponents.*

A Comparison With Dekel, Ely, and Yilankaya (2007)'s Stability

In our notion of stability, the state of the population specifies the behavior of the incumbents also against external mutants. Dekel, Ely, and Yilankaya (2007) present an alternative

Tab. 2: Characterization of Proper Evolutionarily stable Configurations

Interval	Example ($A = 10$)	Characterization of Proper Evolutionarily stable Configurations
$0 < p < \frac{A}{(A-1)^2}$	$1\% < p < 12\%$	Necessary condition: only $\mu(L_\infty) = 1$.
$\frac{A}{(A-1)^2} < p < \frac{A-1}{A}$	$12\% < p < 90\%$	(μ^*, b^*) - Naive (L_1) and moderately sophisticated agents (L_3) co-exist.
$\frac{A-1}{A} < p < 1$	$90\% < p \leq 1$	Necessary condition: $\mu(L_\infty) > 0$.

notion according to which the state of the population only specifies the behavior of players against incumbents, and the behavior against external mutants is determined by a post-entry adaptation process.[14] In this appendix we describe Dekel, Ely, and Yilankaya (2007)'s notion, and show that our results remain qualitatively similar with this notion.

Dekel, Ely, and Yilankaya (2007) assume that the adaptation process according to which agents choose their strategies is much faster than the evolutionary process according to which the frequency of the types evolves. Thus, they assume that the post-entry population adjusts their play to an exact Bayes–Nash equilibrium immediately after mutants enter the population. Let a *compact configuration* be a pair consisting of a distribution of types and the strategy that each type uses against strangers and other incumbents (but behavior against external mutants is unspecified). A compact configuration (μ, β) is *(strictly) DEY-stable* if:

1. Strategy profile β is:
   (a) A Bayes–Nash equilibrium in the Bayesian game with the distribution of types μ.
   (b) Balanced - it induces the same payoff to all types in $C(\mu)$.
2. For each type $L_k \in \mathcal{L}$, there exists a sufficiently small ϵ_0 such that for each $\epsilon < \epsilon_0$, after ϵ mutants of type L_k invade the population:
   (a) There exist post-entry Bayes–Nash equilibria in which the incumbents' play is only slightly changed relative to the pre-entry play.
   (b) In all these equilibria the mutants are (strictly) outperformed by the incumbents.

[14] A similar approach is used in the notions of *mental equilibrium* (Winter, Garcia-Jurado, and Mendez-Naya (2010)) and *evolutionarily stable types* (Alger and Weibull (2012)). Both notions apply only to homogeneous populations that include a single type, and thus are not appropriate to deal with stability of heterogeneous populations.

With simple adaptations, Lemmas 1-5 apply also for DEY-stability. This immediately implies that (μ^*, b^*) is strictly DEY-stable, and that it is "qualitative unique" under the early reciprocity assumption. That is, any other DEY-stable configuration satisfies similar qualitative properties: (1) naive agents (type L_1) and moderately sophisticated agents (types in the set $\{L_2, L_3, L_4\}$) co-exist, and (2) higher types (L_5 and above) do not exist. Formally:

Proposition 1. *Let* $\max\left(\frac{1}{A-2}, \frac{2\cdot A-1}{A\cdot(A-1)}\right) < p < \frac{A-1}{A}$*, let* $\delta > 0$ *be sufficiently small, let* $\lambda < 1$ *be sufficiently large, and assume early-reciprocity. Then the compact configuration* (μ^*, b^*) *is strictly DEY-stable. Moreover, any other DEY-stable configuration* (μ, b) *satisfies:* $\sum_{k\leq 4} \mu(L_i) = 1,$ *and* $0 < \mu(L_1) < 1$.[15]

In this setup we only have the weaker "qualitative" uniqueness because Lemmas 6-7, which are required for "full" uniqueness, do not hold. The lemmas fail because DEY-stability does not consider what happens as a result of a small perturbation to the:

- strategies played by the incumbents (part 1 of both lemmas), as DEY-stability implicitly assumes that the incumbents immediately adjust back to their previous play (which remain Bayes–Nash equilibrium).

- frequencies of the different incumbent types (part 4 of Lemma 7), as DEY-stability only allows mutants to have a single type, while such perturbations can only be represented by entry of heterogeneous mutants.

Finally, we note that if one adapts DEY-stability by: (1) allowing non-external mutants to have several types, and (2) assuming that the adjustment to a new exact equilibrium takes place only after the entry of external mutants, then all of our results, including the "full" uniqueness would hold.

B Proofs

Throughout the proofs we use d_k to denote the following behavior of an informed agent: play *grim* if the horizon$>k$, and defect if horizon$\leq k$. Note the for an informed agent of type L_k, all behaviors d_m for $m \geq k$ are equivalent, as such an agent never encounter an horizon larger than k. In all the results we assume that $\delta > 0$ is sufficiently small, and that $\lambda < 1$ is sufficiently large.

[15] Note that the interval $\max\left(\frac{1}{A-2}, \frac{2\cdot A-1}{A\cdot(A-1)}\right) < p < \frac{A-1}{A}$ is increasing in A, it converges to the entire interval $[0, 1]$ in the limit $A \to \infty$, and it is non-empty for each $A > 3.4$.

B.1 Neutral Stability Implies Balanced Bayes–Nash Equilibrium

Proposition 2. *Let (μ, β) be a neutrally stable configuration. Then, the strategy profile β: (1) induces the same payoff for each type in the support of μ, and (2) is a Bayes–Nash equilibrium in the Bayesian game with distribution of types μ.*

Proof.

1. Assume to the contrary that β induces different payoffs to different types. Let $L_k \in C(\mu)$ be the type with the highest payoff. Then $u\left(\left(L_k, \beta_k\right), (\mu, \beta)\right) > u\left((\mu, \beta), (\mu, \beta)\right)$. This implies that for sufficiently small $\epsilon > 0$, mutants of type L_k who play β_k achieve a strictly higher payoff than the incumbents and this contradicts the stability.

2. Assume to the contrary that β is not a Bayes–Nash equilibrium. Let $L_k \in C(\mu)$ be the type who does not play a best response against (μ, β). This implies that there exists strategy $\beta_k^{'}$ such that $u\left(\left(L_k, \beta_k'\right), (\mu, \beta)\right) > u\left(\left(L_k, \beta_k\right), (\mu, \beta)\right)$. By the first part of the proposition, $u\left(\left(L_k, \beta_k\right), (\mu, \beta)\right) = u\left((\mu, \beta), (\mu, \beta)\right)$. This implies that for sufficiently small $\epsilon > 0$, mutants of type L_k who play β_k' obtain a strictly higher than the incumbents and this contradicts the stability of (μ, β).

□

B.2 Stability of (μ^*, b^*)

Theorem 1. *Let $\frac{A}{(A-1)^2} < p < \frac{A-1}{A}$. Then configuration (μ^*, b^*) is evolutionarily stable.*

Proof. In order to prove that evolutionary stability, we first show two auxiliary results: (μ^*, b^*) is balanced (Lemma 1), and b^* is a Bayes–Nash equilibrium (given μ^*) that is strict with respect to on-equilibrium path deviations (Lemma 2).

Lemma 1. *Configuration (μ^*, b^*) is balanced.*

Proof. Let $q = \mu(L_1)$. Agent of type L_1 gets $(\mathbf{T}-1) \cdot A + 1$ against L_1 opponent and $(\mathbf{T}-2) \cdot A + 1$ against L_3 opponent. Agent of type L_1 obtains $(\mathbf{T}-2) \cdot A + (A+1) + 1 = (\mathbf{T}-1) \cdot A + 2$ against L_1, and against L_3 opponent he gets: $(\mathbf{T}-3) A + 3$ if both players identify each other, $(\mathbf{T}-3) \cdot A + (A+1) + 2 = (\mathbf{T}-2) \cdot A + 3$ if only he identifies his opponent, $(\mathbf{T}-3) \cdot A + 0 + 2$ if only his opponent identifies him, and $(\mathbf{T}-2) \cdot A + 2$ if both players identify each other. Denote by $\delta_3 = \delta \cdot c(L_3)$ the cognitive cost of type L_3. The different types get the same payoff if:

$$q\cdot((\mathbf{T}-1)\cdot A+1)+(1-q)\cdot((\mathbf{T}-2)\cdot A+1)+\delta_3=q\cdot((\mathbf{T}-1)\cdot A+2)+(1-q)\cdot$$
$$\left(p^2\left((\mathbf{T}-3)A+3\right)+p\left(1-p\right)\left(((\mathbf{T}-2)\cdot A+3)+((\mathbf{T}-3)\cdot A+2)\right)+\left(1-p\right)^2\left((\mathbf{T}-2)\cdot A+2\right)\right)$$

$$\left(1-q\right)\left((\mathbf{T}-2)A+1-\left((\mathbf{T}-3)A+1+2p^2+p\left(1-p\right)\left(A+2+1\right)+\left(1-p\right)^2\left(A+1\right)\right)\right)+\delta_3=q$$

$$q=\left(1-q\right)\cdot\left(A-\left(2p^2+p\left(1-p\right)\left(A+3\right)+\left(1-p\right)^2\left(A+1\right)\right)\right)+\delta_3$$

$$q=\left(1-q\right)\left(A-\left(p^2\left(2-A-3+A+1\right)+p\left(A+3-2A-2\right)+\left(A+1\right)\right)\right)+\delta_3$$

$$q=\left(1-q\right)\left(A-\left(p\left(1-A\right)+\left(A+1\right)\right)\right)+\delta_3$$

$$q=\left(1-q\right)\left(-p\left(1-A\right)-1\right)+\delta_3=\left(1-q\right)\left(p\left(A-1\right)-1\right)+\delta_3$$

$$q\left(p\left(A-1\right)-1+1\right)=p\left(A-1\right)-1+\delta_3$$

$$q=\frac{p\left(A-1\right)-1+\delta_3}{p\left(A-1\right)}. \tag{B.1}$$

Note that for each $p>\frac{1}{A-1}$ we get a valid value of $0<q<1$. □

Lemma 2. *Strategy profile b^* is a Bayes–Nash equilibrium given the distribution μ^*. Moreover: (1) playing* grim *until the last three stages is also the best reply of informed mutants of higher types, and (2) any deviation that induces a different play on-equilibrium path yields a strictly worse payoff.*

Proof. For large enough $\lambda<1$, it is immediate that playing *grim* is a best reply for an uninformed agent. In what follows we focus on informed agents. We have to show that (1) Playing d_2 against observed L_1 opponents and strangers, and playing d_3 against observed non-naive opponents is a best reply of all informed agent (including informed mutants with types L_k with $k>3$); and (2) deviations on-equilibrium path are strictly worse. It is immediate that d_2 is a best reply against an observed naive opponent, and strictly better than all on-equilibrium path deviations. Next, we show that playing d_2 against a stranger is strictly better than playing d_3. This is true if the following inequality holds (looking at the payoffs of the last 3 rounds, as all preceding payoffs are the same):

$$q\cdot(2A+2)+(1-q)\cdot(2p+(1-p)\cdot(A+2))>q\cdot(A+3)+(1-q)\cdot(3p+(1-p)\cdot(A+3))$$

$$q\cdot(A-1)>(1-q)\;\Leftrightarrow\;q>\frac{1}{A}.$$

Using (B.1) one obtains:

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

$$p\cdot A^2-p\cdot A-A>p\cdot A-p\ \Leftrightarrow\ p\cdot\left(A^2-2A+1\right)>A\ \Leftrightarrow\ p>\frac{A}{(A-1)^2}.$$

It is then immediate that d_2 is also strictly better (against strangers) than any other deviation on-equilibrium path (also when a mutant agent is informed earlier about the realized length). We are left with showing that playing d_3 is strictly better than playing d_4 against an observed non-naive opponent (and this immediately implies that d_3 is also strictly better than any deviation on-equilibrium path also of a mutant which is informed earlier about the realized length). This is true if the following inequality holds (focusing on the payoffs of the last 4 rounds, as preceding payoffs are equal):

$$p\cdot(A+3)+(1-p)\cdot(2A+3)>p\cdot(A+4)+(1-p)\cdot(A+4)$$

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

We now use the lemmas to prove that (μ^*,b^*) is evolutionarily stable. That is, we have to show that after an invasion of ϵ mutants with configuration (μ,β) $((\mu,\beta)\not\approx(\mu^*,b^*))$, the incumbents obtain a strictly higher payoff than the mutants in the post-entry population (for sufficiently small $\epsilon>0$). □

First, consider mutants of types L_1 or L_3. If these mutants play differently against incumbents (strangers, L_1 or L_3) than do their incumbent counterparts on-equilibrium path, then they are strictly worse off by the previous lemmas. Note that when the proportion of L_1 agents becomes larger (smaller) relative to its proportion in μ^*, then the L_1 agents achieve a lower (higher) payoff than the L_3 agents. This is because L_1 agents obtain a strictly lower payoff than L_3 agents when facing L_1 opponents (L_3 players obtain an additional fitness point by defecting when the horizon is equal to 2). This implies that mutants of types L_1 or L_3 who play the same as their incumbent counterparts on-equilibrium path, obtain a strictly lower payoff than the incumbents (unless these mutants have the same distribution of types as the incumbents, and, in this case, they obtain the same payoff).

Next, consider mutants of different types (L_2 or L_4 or higher). Mutants of type L_2 achieve a strictly lower payoff against incumbents: they have the same payoff as L_3 in most cases, but they obtain a strictly lower payoff when they observe an opponent of type L_3 due to their inability to defect at horizon 3. Mutants of higher types (L_4 or more) obtain at most the incumbents' payoff when facing incumbents (as discussed before, L_3's strategy is a best-reply also when being informed earlier about the realized length), while they have a strictly larger cognitive cost $(\delta \cdot c(L_4))$. Thus these mutants achieve a strictly lower payoff than the incumbents. Finally, mutants may gain an advantage from a *secret handshake*-like behavior (Robson (1990)) - playing the same against incumbent types and strangers, while cooperating with each other when observing a mutant type (different from L_1 and L_3). However, for sufficiently small ϵ, such an advantage cannot compensate for the strict losses mentioned above, and this implies that any configuration of mutants would be outperformed by the incumbents. □

B.3 Properness of (μ^*, b^*)

We first show that (μ^*, b^*) is a proper configuration.

Proposition 3. Let $\frac{A}{(A-1)^2} < p < \frac{A-1}{A}$. Then configuration (μ^*, b^*) is proper.

Proof. Let the sequence of interior ϵ_n-proper configurations (μ^n, β^n) that converge to (μ^*, b^*) be defined as follows (for brevity we only sketch the main details of the construction). Each external mutant $L_k \notin \{L_1, L_3\}$ has $\mu^n(L_2) = O(\epsilon_n)$, while $\mu^n(L_1) = \mu(L_1) - O(\epsilon_n)$ and $\mu^n(L_3) = \mu(L_3) - O(\epsilon_n)$. With probability of $1 - O(\epsilon_n)$ all types play grim when being uninformed, and when being informed they play d_2 against strangers and L_1 opponents, and play d_3 against other observed opponents. Agents play all other pure strategies with probabilities of $O(\epsilon_n)$ or smaller magnitudes in a way that is consistent with ϵ_n-properness. Observe that such a configuration is ϵ_n-proper, and this implies the properness of (μ^*, b^*). □

B.4 Uniqueness of (μ^*, b^*)

Theorem 2. Let $A > 4.6$, $\frac{A}{(A-1)^2} < p < \frac{A-1}{A}$, and assume early-reciprocity. Then if (μ, β) is a proper neutrally stable configuration, then it is equivalent to (μ^*, b^*).[16]

Theorem **2** follows immediately from Lemmas 3-7. First, Lemma 3 shows that proper neutrally stable configuration must include more than one type in their support, and the lowest type must be at most $M-1$. Formally:

Lemma 3. *Let (μ, β) be a configuration such that β is a Bayes–Nash equilibrium given μ. Let $0 < p$. Let type $L_{k_1} \in C(\mu)$ be the smallest type in the population. Then:*

1. *Everyone defects (with probability 1) at any horizon weakly smaller than k_1.*

[16] If $A < 4.6$, Theorem **2** holds if p satisfies: $\max\left(\frac{1}{A-2}, \frac{2\cdot A-1}{A\cdot(A-1)}\right) < p < \frac{A-1}{A}$ or $p < 1 - \frac{2\cdot A-1}{A^2-A}$.

2. *Any type $L_k \neq L_{k_1}$ in the population defects (with probability 1) at horizon $k_1 + 1$.*

3. *If $\mu(L_{k_1}) = 1$ then the configuration* is *not neutrally stable.*

4. *If $k_1 = M$ then the configuration* is *not neutrally stable.*

5. *$\mu(L_{k_1}) < 1$ and $k_1 \leq M - 1$.*

Proof.

1. It is common knowledge that all types are at least k_1. This implies that defecting when the horizon is at most k_1 is the unique strictly dominance-solvable strategy. Thus, all players must defect with probability 1 when the horizon is at most k_1 given any signal about the opponent.

2. Part (1) implies that defecting is strictly better than cooperating at horizon $k_1 + 1$.

3. Observe that if $k_1 < \infty$, then ϵ mutants of type L_∞ who enter the population and play d_{k_1+1} strictly outperform the incumbents. If $k_1 = \infty$, then for sufficiently large λ, mutants of type L_1 will strictly outperform the incumbents, because they induce at least $\mathbf{T} - 2$ rounds of mutual cooperation when their nativity is being observed (due to the properness requirement).

4. If the lowest type in the population is L_M, then L_∞ agents strictly outperform agents with type L_{M-1} and the configuration cannot be neutrally stable.

5. It is immediately implied by the previous parts.

□

Remark 6. Note that if $\lambda$is not close enough to 1 (and this threshold is decreasing in p and converges to 1 as $p \to 0$), then the configuration in which all agents have type L_∞ and they always defect can be a proper evolutionary stable.

Given a configuration with more than one incumbent, we call the lowest incumbent type "*naive*," and all other incumbents are dubbed *non-naive types*. Let a *cooperative opponent* be an opponent who has not defected so far in the game. The following lemma shows that if everyone cooperates at all horizons strictly larger than $k_1 + 1$ in a proper neutrally stable configuration, then this configuration must be equivalent to (μ^*, b^*).

Lemma 4. *Let $p < \frac{A-1}{A}$, let (μ, β) be a proper neutrally stable strategy, and let type $L_{k_1} \in C(\mu)$ be the smallest type in the population. Assume that $\mu(L_{k_1}) < 1$, $k_1 \leq M-1$, and all types in the population cooperate at all horizons strictly larger than $k_1 + 1$ when facing cooperative strangers. Then:*

1. *No one defects at a horizon strictly larger than $k_1 + 2$ against any incumbent.*

2. *$\mu(L_{k_1+1}) = 0$ and all non-naive incumbents play d_{k_1+1} against strangers or observed types L_{k_1}, and they play d_{k_1+2} against any non-naive observed incumbents (up to off-equilibrium path deviations).*

3. *No incumbent has a type strictly larger than L_{k_1+2} (if $k_1 + 1 = M$, let $k_1 + 2 := \infty$).*

4. *The population only includes types $\{L_{k_1}, L_{k_1+2}\}$, and:*
$$\mu(L_{k_1}) = \frac{p(A-1) - 1 + \delta \cdot (c(L_{k_1+2}) - c(L_{k_1}))}{p(A-1)}$$
for any $p > \frac{A}{(A-1)^2}$, and no neutrally stable configuration exists if $p < \frac{A}{(A-1)^2}$.

5. *If $p > \frac{A}{(A-1)^2}$, then (μ, b) and (μ^*, b^*) are equivalent configurations.*

Proof.

1. We have to show that playing d_{k_1+2} is strictly better than an earlier defection against an observed non-naive incumbent. This is because defecting at horizon k_1+3 (defecting at a horizon strictly larger than k_1+3) yields $A-1$ (at least $2 \cdot (A-1)$) fewer points than d_{k_1+2} against an unobserving opponent and at most 1 (2) more points than d_{k_1+2} against an observing opponent. Thus d_{k_1+2} is strictly better than defecting at a horizon of at least k_1+3 if:
$$(1-p) \cdot (A-1) > p \;\Leftrightarrow\; (A-1) > A \cdot p \;\Leftrightarrow\; \frac{A-1}{A} > p.$$

2. By part (2) of the previous lemma all non-naive incumbents play d_{k_1+1} when facing strangers or observed L_{k_1}. It is immediate that d_{k_1+2} is strictly better than defecting at a horizon of at most k_1+1 when facing an observed non-naive incumbent. By the previous part, any incumbent with a type strictly larger than L_{k_1+1} plays d_{k_1+2} against observed non-naive incumbents. In order to complete the proof we have to show that all non-naive incumbents have type different than L_{k_1+1}. Assume to the contrary that: (I) all non-naive incumbents have type L_{k_1+1}; this implies that mutants of type

L_∞ who play d_{k_1+2} against non-naive incumbents and d_{k_1+1} against strangers or naive incumbents outperform the incumbents; or (II) some of the non-naive incumbents have type L_{k_1+1} while other incumbents have higher types; then for sufficiently small $\delta > 0$, the latter group outperforms the former.

3. Assume to the contrary that there are players of a type strictly higher than L_{k_1+2}. If there are also incumbents of type L_{k_1+2} then the previous part shows that both groups play the same on-equilibrium path, and thus the agents with the strictly higher types must obtain strictly lower payoffs due to the cognitive costs. Otherwise, any best-reply mutant type L_{k_1+1} must play d_{k_1+1} against strangers and naive incumbents (or an equivalent strategy that only differs off-equilibrium path), and this implies that in any proper configuration, non-naive incumbents cannot defect at horizons strictly higher than k_1+2 when facing an observed mutant type L_{k_1+1}. This implies that such mutants outperform the incumbents due to the cognitive costs.

4. $C(\mu) = \{L_{k_1}, L_{k_1+2}\}$ is immediate from the previous two parts. In any balanced configuration the naive and the non-naive incumbents must have the same payoff. By repeating Lemma 1's calculations, this implies that $\mu(L_{k_1}) = \frac{p(A-1)-1+\delta\cdot\left(c\left(L_{k_1+2}\right)-c\left(L_{k_1}\right)\right)}{p(A-1)}$. By repeating Lemma 2's calculations, this configuration cannot be stable if $p < \frac{A}{(A-1)^2}$.

5. If $L_{k_1} = L_1$ then the previous parts imply that $(\mu, \beta) \approx (\mu^*, b^*)$ are equivalent configurations. Assume to the contrary that $k_1 > 1$. We now show that ϵ mutants of type L_1 who invade the population outperform L_{k_1} incumbents (and this immediately implies that the mutants also outperform the incumbents of type L_{k_1+2}, as the post-entry difference in the payoffs between the incumbents is $O(\epsilon)$). When facing an opponent of type L_{k_1}, the mutants obtain one less point. When facing an unobserving opponent of type L_{k_1+2}, both types L_1 and L_{k_1}fare the same. When facing an observing opponent of type L_{k_1+2}, the mutants obtain at least $A-1$ more fitness points (by inducing a sophisticated opponent to postpone his defection). Thus the mutants outperform if:

$$\mu(L_{k_1}) < p\cdot(A-1)\cdot\mu(L_{k_1+2}) = p\cdot(A-1)\cdot(1-\mu(L_{k_1})) \Leftrightarrow \mu(L_{k_1}) < \frac{p\cdot(A-1)}{1+p\cdot(A-1)}.$$

By the previous part:

$$\mu(L_{k_1}) = \frac{p(A-1)-1+\delta\cdot(c(L_{k_1+2})-c(L_{k_1}))}{p(A-1)} < \frac{p\cdot(A-1)}{1+p\cdot(A-1)},$$

where the last inequality holds for a sufficiently small $\delta > 0$.

□

We now have to deal with the remaining case in which only a fraction of the non-naive players cooperate at all horizons strictly larger than k_1+1 when facing a cooperative stranger. First, Lemma 5 shows that the frequency of the naive players is small, and that these naive players must have type L_1, and that if p is not too small, then the population must include also types $L_2 - L_4$ but no higher types.

Lemma 5. *Let (μ, β) be a proper neutrally stable strategy, and let type $L_{k_1} \in C(\mu)$ be the smallest type in the population. Assume that $\mu(L_{k_1}) < 1$, $k_1 \leq M-1$, and that there are agents who defect at horizons larger than k_1+1 when facing cooperative strangers. Then:*

1. *$\mu(L_{k_1}) \leq \frac{1}{A}$.*
2. *If $p > \frac{1}{(A-1)^2}$, then $L_{k_1} = L_1$.*
3. *If $p > \frac{2 \cdot A-1}{A \cdot (A-1)}$, then $\mu(L_2) > 0$.*
4. *If $\mu(L_2) > 0$ and $\frac{A-1}{A} > p > \frac{1}{A-2}$ then:*
   - *(a) No incumbent defects at horizons > 3 when facing cooperative strangers.*
   - *(b) No incumbent defects at horizons > 4 when facing any cooperative incumbent.*
   - *(c) No incumbent has a type strictly higher than L_4.*

Proof.

1. The fact that there are incumbents who defect with positive probability at horizons strictly larger than k_1+1 against strangers implies that early defection (at horizon strictly larger than k_1+1) yields a weakly better payoff than d_{k_1+1} against cooperative strangers. Early defection at horizon k_1+2 ($>k_1+2$) yields at least $A-1$ ($2\cdot(A-1)$) fewer fitness points against naive agents, and at most 1 (2) more points against non-naive opponents. This can hold only if:
$$\mu(L_{k_1}) \cdot (A-1) \leq (1-\mu(L_{k_1})) \cdot 1 \Leftrightarrow \mu(L_{k_1}) \leq \frac{1}{A}. \tag{B.2}$$
2. Assume to the contrary that $k_1 > 1$. Observe that ϵ mutants of type L_1 outperform the incumbents of type L_{k_1} (and thus outperform all the incumbents in the post-entry configuration) if: $p \cdot (A-1) \cdot (1-\mu(L_{k_1})) > \mu(L_{k_1}) \cdot 1$. This is because the mutants

of type L_1 earn at least $A-1$ more points when their type is observed by a non-naive incumbent, they earn the same when their type is not observed by a non-naive incumbent, and they earn at most 1 less point when playing against a naive incumbent (type L_{k_1}). Thus the mutants achieve a strictly higher payoff if:

$$p \cdot (A-1) > \mu(L_2) \cdot (1 + p \cdot (A-1)) \Leftrightarrow \frac{p \cdot (A-1)}{1 + p \cdot (A-1)} > \mu(L_2).$$

Substituting (B.2) yields:

$$\frac{p \cdot (A-1)}{1 + p \cdot (A-1)} > \frac{1}{A} \Leftrightarrow p \cdot A \cdot (A-1) > 1 + p \cdot (A-1) \Leftrightarrow p > \frac{1}{(A-1)^2}.$$

3. Assume to the contrary that $\mu(L_2) = 0$. Balance implies that the naive players (L_1) must have the same payoff as the non-naive players. This can hold only if:

$$p \cdot (A-2) \cdot (1 - \mu(L_1)) < (1-p) \cdot (1 - \mu(L_1)) \cdot 2 + \mu(L_1).$$

This is because naive players obtain (on average) at least $A-2$ more fitness points when their type is observed by a non-naive opponent (as they induce their opponent to cooperate at least one more round), and non-naive agents get at most 1 more point against a naive opponent and at most 2 more points against a non-observing sophisticated opponent. Thus:

$$(p \cdot (A-2) - 2 \cdot (1-p)) \cdot (1 - \mu(L_1)) < \mu(L_1) \Leftrightarrow (p \cdot A - 2) \cdot (1 - \mu(L_1)) < \mu(L_1)$$

$$(p \cdot A - 2) < \mu(L_1) \cdot (p \cdot A - 1) \Leftrightarrow \mu(L_1) > \frac{p \cdot A - 2}{p \cdot A - 1}.$$

Substituting (B.2) yields:

$$\frac{1}{A} > \frac{p \cdot A - 2}{p \cdot A - 1} \Leftrightarrow p \cdot A - 1 > p \cdot A^2 - 2 \cdot A \Leftrightarrow p \cdot A^2 - (2+p) \cdot A + 1 < 0.$$

The last inequality holds if and only if $p > \frac{2 \cdot A - 1}{A \cdot (A-1)}$, a contradiction.

4. Let $\mu(L_{3+}) = 1 - \mu(L_1) - \mu(L_2)$.

   (a) Balance implies that types L_1 and L_2 obtain the same payoff. This can hold only if:

$$p \cdot \mu(L_{3+}) \cdot (A-1) < \mu(L_1) + \mu(L_2) + (1-p) \cdot \mu(L_{3+}).$$

This is because type L_1 obtains $A-1$ more fitness points against an observing opponent of type L_3 or higher, while type L_2 obtains 1 more point against types L_1 and L_2 and at most 1 more point against an unobserving type L_3 or higher. Thus:

$$p \cdot \mu\left(L_{3+}\right) \cdot (A-1) < 1 - \mu\left(L_{3+}\right) + (1-p) \cdot \mu\left(L_{3+}\right) = 1 - p \cdot \mu\left(L_{3+}\right)$$

$$p \cdot \mu\left(L_{3+}\right) \cdot A < 1 \Leftrightarrow \mu\left(L_{3+}\right) < \frac{1}{A \cdot p}.$$

This implies, together with (B.2):

$$\mu\left(L_2\right) = 1 - \mu\left(L_1\right) - \mu\left(L_{3+}\right) > 1 - \frac{1}{A} - \frac{1}{A \cdot p} = \frac{A \cdot p - p - 1}{A \cdot p}.$$

From the same argument as in part (1) of this lemma, if $\mu\left(L_2\right) > \frac{1}{A}$ then no incumbent defects at a horizon strictly larger than 3 when facing cooperative strangers. Substituting this inequality yields:

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

(b) The proof repeats the argument of part (2) of Lemma 4.

(c) The proof repeats the argument of part (3) of Lemma 4.

□

Lemma 6 shows that: (1) a large fraction of non-naive players cooperate at all horizons strictly larger than k_1+1 when facing cooperative strangers, (2) if $p < 1 - \frac{2A-1}{A^2-A}$ then no incumbent defects at a horizon strictly larger than k_1+2 when facing cooperative strangers, and (3) no incumbent has a type higher than L_{k_1+3}.

Lemma 6. *Let $p < 1 - \frac{2A-1}{A^2-A}$, let (μ, β) be a proper neutrally stable strategy, let type $L_{k_1} \in C(\mu)$ be the smallest type in the population, and assume the that $\mu\left(L_{k_1}\right) < 1$ and $k_1 \leq M-1$. Let η be the mean probability that a non-naive incumbent cooperates at all horizons strictly larger than k_1+1 when facing a cooperative stranger. Assume that $\eta < 1$. Then:*

1.

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

2. No player defects at horizon $>$ k_1+2 when facing cooperative strangers.

3. *No player defects at horizon* $> k_1 + 3$ *when facing cooperative incumbents.*

4. *No player in the population has a type strictly larger than* L_{k_1+3}.

Proof.

1. Type L_{k_1} gets $(\mathbf{T}-1)\cdot A+1$ points when playing against itself. A random player with a type different than L_{k_1} who plays against L_{k_1} gets at most $(\mathbf{T}-1)\cdot A+1+1$ when he observes his opponent's type, and an expected payoff of at most $\eta\cdot((\mathbf{T}-1)\,A+2)+(1-\eta)\cdot((\mathbf{T}-2)\cdot A+3)$. This implies that a necessary condition for other types to achieve a higher payoff (on average) when playing against L_1 than the payoff that L_1 gets against itself is (subtracting the equal amount of $(\mathbf{T}-2)\cdot A+1$ from each payoff):

$$A < p\cdot(A+1)+(1-p)\cdot(\eta\cdot(A+1)+2\cdot(1-\eta))$$

$$A < 1+p\cdot A+(1-p)\cdot(\eta\cdot A+1-\eta) \Leftrightarrow A-\frac{1}{1-p} < \eta\cdot A+1-\eta$$

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

$$\frac{(A-1)\cdot(1-p)-1}{(A-1)\cdot(1-p)} < \eta. \tag{B.3}$$

If (B.3) does not hold, then the configuration cannot be naturally stable, because a sufficiently small group of mutants with type L_1 who invade the population and play d_1 would outperform the incumbents.

2. We show that when facing strangers, all types cooperate with probability 1 at all horizons strictly larger than k_1+2. Assume to the contrary that there is a type who defects with positive probability against cooperative strangers at horizon $l > k_1+2$. This implies that defecting at horizon l yields a weakly better payoff against strangers than d_{k_1+2}. This can occur only if:

$$\eta\cdot(1-p)\cdot(A-1) \leq (1-\eta)+\eta\cdot p.$$

This is because if $l = k_1+3$ ($l > k_1+3$), d_{k_1+2} yields $A-1$ (at least $2\cdot(A-1)$) more points against non-observing opponents who cooperate at all horizons larger than k_1+1, and it yields at most 1 (2) fewer points against any other opponents. Thus:

$$\eta\cdot(1-p)\cdot(A-1) \leq 1-\eta\cdot(1-p) \Leftrightarrow \eta\cdot(1-p)\cdot A \leq 1\,(1-p) \Leftrightarrow \eta \leq \frac{1}{(1-p)\cdot A}.$$

Substituting (B.3) yields:

$$\frac{(A-1)\cdot(1-p)-1}{(A-1)\cdot(1-p)}\leq\frac{1}{(1-p)\cdot A}\Leftrightarrow A\cdot((A-1)\cdot(1-p)-1)\leq A-1$$

$$A\cdot(A-1)\cdot(1-p)-A\leq A-1\Leftrightarrow A\cdot(A-1)\cdot(1-p)\leq 2\cdot A-1$$

$$1-p\leq\frac{2\cdot A-1}{A\cdot(A-1)}\Leftrightarrow p\geq 1-\frac{2\cdot A-1}{A^2-A},$$

and we get a contradiction to $p<1-\frac{2\cdot A-1}{A^2-A}$. By part (2) of Lemma 3, all non-naive incumbents defect with probability 1 at any horizon of at most k_1+1. This implies that η of the non-naive incumbents play d_{k_1+1} against cooperative strangers and the remaining fraction plays $d_{k_{1+2}}$.

3. The proof repeats the argument of part (2) of Lemma 4.

4. The proof repeats the argument of part (3) of Lemma 4.

□

The following corollary is immediately implied by Lemmas 3-6:

Corollary 1. *Let*

$$\max\left(\frac{1}{A-2},\frac{2\cdot A-1}{A\cdot(A-1)}\right)<p<\frac{A-1}{A}\text{ or }0<p<1-\frac{2\cdot A-1}{A^2-A},$$

let $(\mu,\beta)\not\approx(\mu^*,b^*)$ be a stable proper neutrally stable configuration, and let $L_{k_1}\in C(\mu)$ be the lowest incumbent type ("naive"). Then:

1. *$\mu(L_{k_1})<1$.*

2. *All non-naive incumbents either play d_{k_1+1} or d_{k_1+2} against cooperative strangers.*

3. *$\mu(L_k)=0$ for every $k>k_1+3$.*

4. *If $p>\frac{1}{(A-1)^2}$ then $k_1=1$.*

Remark 7. Note that if $A>4.6$ then

$$\max\left(\frac{1}{A-2},\frac{2\cdot A-1}{A\cdot(A-1)}\right)<1-\frac{2\cdot A-1}{A^2-A},$$

which implies that Corollary 1 is valid in this case for each $p<\frac{A-1}{A}$.

Finally, Lemma 7 shows that the configurations characterized by Corollary 1 cannot be proper neutrally stable (unless it is equivalent to (μ^*, b^*)). To simplify notation, the lemma describes the case in which $L_{k_1} = L_1$ but it works the same (only with more cumbersome notations) for $L_{k_1} > L_1$ (which is possible when $p < \frac{1}{(A-1)^2}$).

Lemma 7. *Let $0 < p < \frac{A-1}{A}$, let (μ, β) be a configuration satisfying: (1) $0 < \mu(L_1) < 1$, (2) $\mu(L_k) = 0 \; \forall k > 4$, and (3) a positive fraction of non-naive incumbents play d_3 against cooperative strangers, and the remaining non-naive players play d_2 against cooperative strangers. Then (μ, β) cannot be proper neutrally stable.*

Proof. Assume to the contrary that (μ, β) is a proper neutrally stable configuration.

1. *All players who play d_2 against cooperative strangers have type L_2.*
   Assume to the contrary that there is a type $L_{\tilde{k}}$ ($\tilde{k} > 2$) that plays d_2 with positive probability against strangers (and by the previous lemma it plays d_3 with the remaining probability). Consider the following configuration of mutants ,(μ', b'), which as the same distribution of types as the incumbents, and play like the incumbents except when mutant of type $\tilde{k}$ faces a cooperative stranger: (1) $\mu' = \mu$, (2) for each $k \neq \tilde{k}$, $b'_k = b_k$, (3) for each $L_k \in \mathcal{L}$, $b'_{\tilde{k}} = b_{\tilde{k}}$ except that the mutants of type $\tilde{k}$ play d_3 against cooperative strangers. Observe that such mutants strictly outperform the incumbents: mutants of a type different than $L_{\tilde{k}}$ obtain the same payoff as their incumbent counterparts, while mutants of type $L_{\tilde{k}}$ achieve a strictly higher payoff when facing an unobserved opponent of type $L_{\tilde{k}}$ (pre-entry, both d_2 and d_3 yielded the same payoff; post-entry, there are a few more early defectors and thus d_3 yields a strictly higher payoff), and obtain the same payoff in all other cases. This implies that the configuration cannot be neutrally stable.

2. *β is characterized as follows: Uninformed agents follow any early-reciprocal behavior. Informed agents play as follows: L_1 agents play d_1; L_2 agents play d_2; L_3 agents play d_2 against observed L_1 and d_3 in all other cases; and L_4 agents play d_2 against observed L_1, d_3 against strangers and observed L_2, and d_4 against observed L_4 or L_3 (all strategies are determined up to off-equilibrium path deviations that do not change the observable play).*
   The strategies used against strangers are determined by the previous part and by Lemma 3. The strategies used against observed incumbents are best replies if $p < \frac{A-1}{A}$ by the same argument as in part (1) of Lemma 4.

3. $\mu(L_3) = 0$.

By a similar argument to part (2) of Lemma 4, agents of type L_4 outperform agents of type L_3 due to their unique ability to play d_4 against an observed type L_3 or L_4.

4. To simplify notation we characterize the frequency of each type in the case where the cognitive costs converge to 0 ($\delta \to o$). The arguments work very similarly (but the notation is more cumbersome) for small enough $\delta > 0$. Then:

$$\mu(L_1) = \frac{1}{A + p \cdot (1-p) \cdot (A-1)^2},\ \mu(L_2) = 1 - \frac{1 + A - p \cdot (A-1)}{A + p \cdot (1-p) \cdot (A-1)^2},$$

$$\mu(L_4) = \frac{1}{p \cdot (A-1) + 1}.$$

Let $\mu_k = \mu(L_k)$. The fact that (μ, b) is a balanced configuration implies that types L_1 and L_2 should have the same payoff. Type L_2 obtains 1 more fitness point against types L_1 and L_2, the same payoff against an unobserving type L_4, and $A-1$ fewer points against an observing type L_4. The balance between the payoffs implies:

$$(1 - \mu_4) = \mu_4 \cdot p \cdot (A-1) \Leftrightarrow \mu_4 = \frac{1}{p \cdot (A-1) + 1}. \tag{B.4}$$

Similarly, L_2 and L_4 should have the same payoff. Type L_2 obtains 1 less point against type L_2, the same number of points against observed type L_1, $A-1$ more points against unobserved type L_1, and the comparison against an opponent of type L_4 depends on observability: $A-2$ more points when both types are observed, 1 less point when both types are unobserved, 2 fewer points when only the opponent is observed, and $A-1$ more points when only the opponent is observing. Thus, the balance between the payoffs implies:

$$(1-p) \cdot \mu_1 \cdot (A-1) + \mu_4 \cdot \left(p^2 \cdot (A-2) - (1-p)^2 + p \cdot (1-p) \cdot (A-1-2)\right) = \mu_2$$

$$(1-p) \cdot \mu_1 \cdot (A-1) + \mu_4 \cdot \left(p^2 \cdot (A-3) - 1 + 2p + \left(p - p^2\right) \cdot (A-3)\right) = \mu_2$$

$$(1-p) \cdot \mu_1 \cdot (A-1) + \mu_4 \cdot (p \cdot (A-3) - 1 + 2p) = \mu_2$$

$$(1-p) \cdot \mu_1 \cdot (A-1) + \mu_4 \cdot (p \cdot (A-1) - 1) = \mu_2 = 1 - \mu_1 - \mu_4$$

$$\mu_4 \cdot p \cdot (A-1) = 1 - \mu_1 \cdot (1 + (1-p) \cdot (A-1)) \Leftrightarrow \mu_4 \cdot p \cdot (A-1) = 1 - \mu_1 \cdot (A - p \cdot (A-1))$$

$$\mu_1 \cdot (A - p \cdot (A-1)) = 1 - \mu_4 \cdot p \cdot (A-1) \Leftrightarrow \mu_1 = \frac{1 - \mu_4 \cdot p \cdot (A-1)}{A - p \cdot (A-1)}.$$

Substituting (B.4) yields:

$$\mu_1 = \frac{1-\frac{p\cdot(A-1)}{p\cdot(A-1)+1}}{A-p\cdot(A-1)} = \frac{\frac{1}{p\cdot(A-1)+1}}{A-p\cdot(A-1)}$$

$$\mu_1 = \frac{1}{(p\cdot(A-1)+1)\cdot(A-p\cdot(A-1))} = \frac{1}{A+p\cdot(1-p)\cdot(A-1)^2}.$$

This implies that:

$$\mu_2 = 1-\mu_1-\mu_4 = 1-\frac{1}{(p\cdot(A-1)+1)\cdot(A-p\cdot(A-1))} - \frac{1}{p\cdot(A-1)+1}$$

$$\mu_2 = 1-\frac{1+A-p\cdot(A-1)}{(p\cdot(A-1)+1)\cdot(A-p\cdot(A-1))} = 1-\frac{1+A-p\cdot(A-1)}{(p\cdot(A-1)+1)\cdot(A-p\cdot(A-1))}$$

$$\mu_2 = 1-\frac{1+A-p\cdot(A-1)}{A+p\cdot(1-p)\cdot(A-1)^2}.$$

If any of the μ_i-s is not between 0 and 1 then no neutrally stable configuration exists.

5. *The configuration is not neutrally stable.*

   A direct algebraic calculation reveals that for sufficiently small $\epsilon, \epsilon' > 0$:

   (a) If $p < 0.5$ then ϵ "imitating" mutants (who play the same strategies as the incumbents) with configuration (μ', b') with $\mu'(L_1) = 1-\mu(L_4)+\epsilon'$, $\mu'(L_2) = 0$, $\mu'(L_4) = \mu(L_4)-\epsilon'$, and $b' = b$ (play the same as the incumbents) outperform the incumbents in the post-entry population.

   (b) If $p > 0.5$ ϵ "imitating" mutants with a configuration (μ', b') with $\mu'(L_1) = 0$, $\mu'(L_2) = 1-\mu(L_4)+\epsilon'$, $\mu'(L_4) = \mu(L_4)-\epsilon'$, and $b' = b$ outperform the incumbents in the post-entry population for sufficiently small ϵ.

□

B.5 Stable Configurations Near 0 and 1 - Theorem 3

Parts 1 of Theorem **3** is immediate from Remarks 6 and 7, and from Corollary 1. Part 2 is immediate from Theorems 1-4. We have to prove part 3:

Theorem. *3. (part 3): Let $\frac{A-1}{A} < p \leq 1$. Then in any proper neutrally stable configuration, (μ, β), $\mu(L_\infty) > 0$, and L_∞ players defect at all stages when observing an L_∞ opponent.*

Proof. Let L_k be the highest type in the population. Let l be the largest horizon in which L_k agents begin defecting with positive probability against an observed cooperative opponent of the same type. If this probability is strictly less than 1, then by a similar argument to part (1) of Lemma 7, the configuration is not neutrally stable (ϵ "imitating" mutants who differ only in that the L_k mutants play d_l with probability 1 against observed L_k opponents, would outperform the incumbents). Now, if $l < k$, then ϵ mutants of type L_k who play d_{l+1} (start defecting one stage earlier) against observed L_k, and play the same as the incumbents in all other cases, outperform the incumbents of type L_k (and this implies they outperform all incumbents) if:

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

(because the mutants obtain 1 more point when their L_k opponent observes their type, and they get at most $A-1$ fewer points when he does not observe their type; they obtain the same payoff against strangers and other observed opponents). For similar reasons, if $l = k < \infty$, then ϵ mutants of type L_∞ who play d_{k+1} against observed L_k, and play the same as the incumbents of type L_k in all other cases, outperform incumbents of type L_k (and this implies they outperform all incumbents) in any proper neutrally stable configuration. □

References

Alger, I., and J. Weibull (2012): "Homo Moralis – Preference evolution under incomplete information and assortative matching," Discussion paper, Toulouse School of Economics.

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

Axelrod, R. (1984): *The Evolution of Cooperation.* Basic Books.

Bruttel, L., W. Güth, and U. Kamecke (2012): "Finitely repeated prisoners' dilemma experiments without a commonly known end," *International Journal of Game Theory*, 41(1), 23–47.

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

Camerer, C., T. Ho, and J. Chong (2004): "A cognitive hierarchy model of games," *The Quarterly Journal of Economics*, 119(3), 861–898.

COOPER, R., D. DEJONG, R. FORSYTHE, AND T. ROSS (1996): “Cooperation without reputation: Experimental evidence from prisoner’s dilemma games,” *Games and Economic Behavior*, 12(2), 187–218.

COSTA-GOMES, M., V. CRAWFORD, AND B. BROSETA (2001): “Cognition and behavior in normal-form games: An experimental study,” *Econometrica*, 69(5), 1193–1235.

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

DEKEL, E., J. C. ELY, AND O. YILANKAYA (2007): “Evolution of preferences,” *Review of Economic Studies*, 74(3), 685–704.

FRENKEL, S., Y. HELLER, AND R. TEPER (2012): “Endowment as a blessing,” *mimeo.*

GEANAKOPLOS, J., AND L. GRAY (1991): “When seeing further is not seeing better,” *Bulletin of the Santa Fe Institute*, 6(2).

GÜTH, W., AND M. YAARI (1992): “Explaining reciprocal behavior in simple strategic games: An evolutionary approach,” in *Explaining Process and Change: Approaches to Evolutionary Economics*, ed. by U. Witt. University of Michigan Press, Ann Arbor.

HAVILAND, W., H. PRINS, AND D. WALRATH (2007): *Cultural Anthropology: The Human Challenge*. Wadsworth Pub. Co.

JOHNSON, E., C. CAMERER, S. SEN, AND T. RYMON (2002): “Detecting failures of backward induction: Monitoring information search in sequential bargaining,” *Journal of Economic Theory*, 104(1), 16–47.

MAYNARD-SMITH, J. (1974): “The theory of games and the evolution of animal conflicts,” *Journal of Theoretical Biology*, 47(1), 209–221.

MAYNARD SMITH, J. (1982): *Evolution and the theory of games.* Cambridge University Press, Cambridge.

MCKELVEY, R., AND T. PALFREY (1995): “Quantal response equilibria for normal form games,” *Games and Economic Behavior*, 10(1), 6–38.

MENGEL, F. (2012): “Learning by (limited) forward looking players,” *mimeo.*

MOHLIN, E. (2012): “Evolution of theories of mind,” *Games and Economic Behavior*, 75(1), 299–318.

Myerson, R. (1978): "Refinements of the Nash equilibrium concept," *International Journal of Game Theory*, 7(2), 73–80.

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

Nagel, R., and F. Tang (1998): "Experimental results on the centipede game in normal form: An investigation on learning," *Journal of Mathematical Psychology*, 42(2), 356–384.

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

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

——— (2003): "The evolution of rationality and the Red Queen," *Journal of Economic Theory*, 111(1), 1–22.

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

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

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

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

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

Taylor, P., and L. Jonker (1978): "Evolutionary stable strategies and game dynamics," *Mathematical Biosciences*, 40(1), 145–156.

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

Winter, E., I. Garcia-Jurado, and L. Mendez-Naya (2010): "Mental equilibrium and rational emotions," *mimeo.*