Three steps ahead

Heller, Yuval

University of Oxford, Department of Economics

13 June 2012

Three Steps Ahead (Preliminary Draft, June 13, 2012)

*Yuval Heller**

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

Abstract

Experimental evidence suggest that people only use 1-3 iterations of strategic reasoning, and that some people systematically use less iterations than others. In this paper, we present a novel evolutionary foundation for these stylized facts. In our model, agents interact in finitely repeated Prisoner's Dilemma, and each agent is characterized by the number of steps he thinks ahead. When two agents interact, each of them 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 unique stable outcome, in which the population includes a mixture of "naive" agents who think 1 step ahead, and "sophisticated" agents who think 2-3 steps ahead.

Keywords: Indirect evolution, cognitive hierarchy, bounded forward-looking, Prisoner's Dilemma, Cooperation. JEL Classification: C73, D03.

1 Introduction

Experimental evidence suggest that in new strategic interactions most people only use 1-3 iterations of strategic reasoning. This stylized fact is observed in different forms in various contexts. First, when playing long finite games, people only look a 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 finitely-repeated Prisoner's Dilemma, (Selten and Stoecker (1986) and the other references discussed at the of the introduction) and

*I would also like to express my deep gratitude to Itai Arieli, Vince Crawford, Peyton Young, and seminar participants at University of Birmingham, University of Oxford and University College London, for many useful comments, discussions and ideas.

"Centipede" games (McKelvey and Palfrey (1995); Nagel and Tang (1998)), and when interacting in sequential bargaining, players ignore future bargaining opportunities that are more than 1-2 steps ahead (Johnson, Camerer, Sen, and Rymon (2002)). Second, when facing serially dominated strategies, almost everyone make the first iteration (not playing a dominated action), many do the second iteration - assume that their opponent would not play dominated strategies, a few make the third iteration, and further iterations are rare (Beard and Beil Jr (1994); Rapoport and Amaldoss (2004)). Third, according to the models of cognitive hierarchy (or level-k), most players best respond to a belief that others use only at most two iterations of strategic reasoning (see, e.g., Stahl and Wilson (1994); Nagel (1995); Ho, Camerer, and Weigelt (1998); Bosch-Domenech, Montalvo, Nagel, and Satorra (2002); Crawford (2003); Camerer, Ho, and Chong (2004); Crawford and Iriberri (2007)).

A second stylized fact is the heterogeneity of the population: some people systemically use less iterations than others (see, e.g., Chong, Camerer, and Ho (2005) and Hyndman, Terracol, and Vaksmann (2012)). These stylized facts raise two related evolutionary puzzles. The first puzzle is why people only use 1-3 steps. Experimental evidence suggest that using more iterations is only unintuitive but not computationally complex (at-least in simple games): with appropriate guidance and feedback players can learn to use many iterations in a given game; however, when facing a new game people immediately return to use only 1-3 iterations (Johnson, Camerer, Sen, and Rymon (2002), Camerer (2003, Section 5.3.5)). In many games, being able to do one more step than the opponent gives a substantial advantage. As the cognitive cost of an additional level is moderate, it is puzzling why there hasn't been an "arms race" ("red queen effect", see Robson (2003)) that caused people to use more strategic iterations throughout the evolutionary process.

The second puzzle is how the "naive" people, who systematically use less iterations than the more "sophisticated" agents, survived the evolutionary process. At first glance, it seems that naive people would fare substantially less than sophisticated agents who enjoy the benefit of thinking one level ahead. *In this paper we present an evolutionary model that explains both puzzles and yields a unique sharp prediction: an heterogeneous population which only uses 1-3 strategic iterations.* Our model focuses on limited forward-looking in repeated Prisoner's Dilemma, but we believe that it can also shed light on other forms of bounded iterative reasoning.

Following the "indirect evolutionary approach" (Güth and Yaari (1992)) we present a reduced-form static analysis for a dynamic process that describes the evolution of types in a large population of agents.[1] Each agent has a type (level) in the set $\{L_1, ..., L_M\}$

[1] The indirect approach was mainly used to study evolution of preferences. Following, Stahl Dale (1993); Stennek (2000); Frenkel, Heller, and Teper (2012), we apply it to analyze evolution of cognitive biases.

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

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

that determines how many steps he looks ahead (as described in the next paragraph). At each generation the players are randomly matched and each couple plays M times (without rematching) the symmetric stage game of the Prisoner's Dilemma with the payoffs given in Table 1:[2] mutual cooperation (both players play C) yields both players $A > 2 + \sqrt{2}$, mutual defection (both players play D) gives 1, and if a single player defects, he obtains $A+1$ and his opponent gets 0. Observe that the parameter A is the ratio between what can be gained by mutual cooperation to the additional payoff that is obtained by defecting.[3] The total payoff from the repeated interaction is the undiscounted sum of payoffs. 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. Informally, this can be interpreted as an opportunity to observe your opponent's past behavior, or to observe a trait that is correlated with cognitive level (such as I.Q. level, see Gill and Prowse (2012)).

An agent of type L_k looks k steps ahead in his strategic reasoning. When the *horizon* (the number of remaining stages) is larger than k the agent must follow a simple heuristic - "grim": he cooperates if and only if is opponent never defected before.[4] When the horizon is equal to k, the agent begins to play strategically and he may choose any action. We interpret L_k's behavior to stem from bounded forward-looking: when the horizon is larger than k, he subjectively perceives it to be infinite, and he does not take into account the fact that the interaction has a well-defined final period, and that this final period has strategic implications. One can also consider our model as a reduced-form for an interaction with a random unknown long length, in which each type L_k gets a signal about the interaction's realized length k periods before the end (see Section 6). Observe that the set of strategies of type L_k is a strict subset of the set of strategies of type L_{k+1}, and that type L_M is fully-rational and has an unlimited set of strategies.

[2] All our results are independent of the value of M for every $M \geq 4$.

[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 (but the results remain qualitatively similar also without this assumption). Given this assumption we normalize, without loss of generality, the payoff of being a single cooperator to be 0, and the additional payoff of defecting to be 1.

[4] In Section 6 we discuss the extension of our model to a setup in which a player may choose his heuristic for long horizons, and the relation to the notion of analogy-based expectation equilibrium (Jehiel (2005)).

In common with much of the evolutionary literature, we use a static solution concept to tractably capture the stable points of a dynamic evolutionary process with the following three properties: (1) the frequency of types changes slowly according to their success: types that have yielded higher payoffs increase at the expense of those that yielded lower payoffs, and (2) given a distribution of types, players learn to best-respond to each other (i.e, they learn to play a Bayesian-Nash equilibrium, given their forward-looking constraints), and (3) rarely a few mutants with arbitrary type and behavior enter the population; if these mutants have a completely new type (which has not existed among the incumbents), then it is assumed that the incumbents learn to best reply to the mutants. This evolutionary process can be interpreted in two different ways: (1) biological process - types are genetically determined, and the payoff is the expected 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 another person's type, if the other person is more successful.

The *configuration* of the population is a pair consisting of a distribution of types and the strategy that each type L_k uses in the repeated game (which must be *grim* as long as the horizon is larger than k). A configuration is *stable* (a variant of Dekel, Ely, and Yilankaya (2007)'s definition) if it satisfies 3 conditions: (1) *balance* - each type in the population has the same expected payoff; (2) *equilibrium* - each type uses a best reply strategy; and (3) *resistance* - a small group of mutants that enters the population fares worse than the incumbents.[5]

Our main result shows that if p is distanced enough from both 0 and 1, then there exists a unique stable configuration which includes two kind of players: (1) *naive* agents of type L_1 who only begin defecting at the last stage, (2) *sophisticated* agents who look 2-3 steps ahead: usually they begin defecting two stages before the end, unless they observe that their opponent is sophisticated, and in this case, they begin defecting one stage earlier. The stability relies on the balance between the direct disadvantage of naive agents - they defect too late, and the indirect advantage - when nativity is being observed, it induces sophisticated opponents to postpone their defection, and this allows an additional round of mutual cooperation.[6]

[5] Our solution concept extends the notion of neutral stable strategy (NSS, Maynard Smith (1982)) from direct evolution, in which types completely determine the actions, to indirect evolution. In Appendix B we present a few alternative variants to the resistance condition (including Dekel, Ely, and Yilankaya (2007)'s definition), and we demonstrate that our results are robust to its exact properties.

[6] The proportion of sophisticated players is equal to $\frac{1}{p(A-1)}$. The types of the sophisticated agents are not uniquely determined. The sophisticated players may have any type of L_3 or higher (but all types play as if they were type L_3). If one adds to the model an arbitrarily small cost of having an higher cognitive level, then all the sophisticated players must have type L_3 (see Corollary 1).

It is interesting to note that stable configurations are very different when p is close to 0 or 1. In both cases, stable configurations must include fully-rational players who, when facing other fully-rational 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. The result of this "race" is that there must be some fully-rational agents in the population.

Existing evolutionary models that studied bounded strategic reasoning (Stahl Dale (1993); Stennek (2000)) focused on the case where types are unobservable ($p = 0$), and showed that: (1) the highest type always survives, and (2) other types may also survive if they do not play serially dominated strategies. Recently, Mohlin (2012) also dealt also with the case of fully-observable types ($p = 1$), and characterized conditions under which other types beside the highest may survive. This paper is the first to study partial observability in such a setup, which, perhaps surprisingly, leads to a much sharper prediction: a unique stable state in which everyone thinks 1-3 steps ahead.

Existing experimental results verify the plausibility of both our assumption of using "grim" strategy for large horizons, and of our main prediction. Selten and Stoecker (1986) study the behavior of players in iterated Prisoner Dilemma games of 10 rounds (similar results are presented in Andreoni and Miller (1993); Cooper, DeJong, Forsythe, and Ross (1996); Bruttel, Güth, and Kamecke (2012)). They show that: (1) if any player defected, then almost always both players defect at all remaining stages (a "grim"-like behavior), (2) usually there is mutual cooperation in the first 6 rounds, and (3) players begin defecting at the last 1-4 rounds.[7] Such behavior has two main explanations in the literature: (1) some players are altruistic, and (2) players have limited forward-looking.[8] Johnson, Camerer, Sen, and Rymon (2002) studied the relative importance of these explanations in a related sequential bargaining game, and their findings suggest the limited forward-looking is the main cause for this behavior.

The paper is structured as follows. Section 2 presents our model. Section 3 describes our solution concept (and in Appendix B we demonstrate the robustness of our results to a few plausible variants of this concept). In Section 4 we present our results, and it is followed by

[7] In Selten and Stoecker (1986)'s experiments players engaged in 25 sequences ("super-games") of iterated 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"). During these 13 sequences there is a slow drift in the behavior of players towards earlier defections. Nevertheless, defections before the last 4 rounds were infrequent also in the last couple of rounds.

[8] Heifetz and Pauzner (2005) explain this behavior with a different kind of cognitive limitations: at each node, each player has a small probability to be "confused" and choose a different action than the optimal one.

sketches of the proofs in Section 5 (formal proofs appear in Appendix A). We conclude in Section 6.

2 Model

We study a symmetric finitely-iterated Prisoner's Dilemma game that repeats M stages ($M \geq 4$), denoted by G. The payoff of each stage game are as described in Table 1 ($A > 2 + \sqrt{2}$). The payoff of the repeated game is the undiscounted sum of the stage payoffs. This payoff is interpreted, as standard in the evolutionary literature, as representing "success" or "fitness". Define the horizon of a stage as the number of remaining stages including the current stage. That is, the horizon at stage k is equal to $L - k + 1$. History h_k of length k is a sequence of k pairs, where the l-th pair describes the actions chosen by the players at stage l. Let H_k be the sets of histories of length k, and let $H = \cup_{1 \leq k < M} H_k$ be the set of all histories.

A pure strategy s is a function from H into $\{C, D\}$, and a behavioral strategy σ is a function from H into $\Delta(\{C, D\})$. With some abuse of notations we write $\sigma(h_k) = C$ when σ assigns probability 1 to playing C (and similarly for D). Let Σ be the set of behavioral strategies. Strategy σ is *k-grim* if whenever the horizon is larger than k: (1) σ assigns probability 1 to C if the opponent has never defected before, and (2) σ assigns probability 1 to D if the opponent has defected in the past. Let Σ_k be the set of k-grim behavioral strategies. Let $u(\sigma, \sigma')$ be the expected payoff of a player who plays behavioral strategy σ against an opponent who plays behavioral strategy σ'. Let $d_k \in \Sigma_k$ be the pure strategy that plays grim as long as the horizon is larger then k, and then defects at all following stages (when the horizon is at most k).

We imagine a large population randomly matched to play G. Different agents in the population differ in their cognitive ability, which is captured by their type. Let $\mathcal{L} = \{L_1, ..., L_M\}$ be the set of types (or levels).[9] An agent of type L_k looks only k steps ahead, and when the horizon is larger than k he ignores end-of-game strategic considerations and plays *grim* - defecting if and only if his opponent defected in the past. That is, an agent with type L_k can only play k-grim strategies. When the horizon is at most k, the agent is no longer limited in his play.

Following the model of partial observability of Dekel, Ely, and Yilankaya (2007), we assume that each player knows the type of his opponent with probability p (and get no information about his opponent's type with probability $1 - p$), independently of the event that his opponent knows his type. Let a *stranger* denote an opponent that his type was not

[9] We explicitly omit type of level 0 (who keep using use grim throughout the entire interaction). See Subsection 4 for a discussion of this assumption, and the influence of changing it.

observed.

3 Solution Concept

In this section we formally present a reduced-form stability concept (a variant of Dekel, Ely, and Yilankaya (2007)'s notion of stability), which extends neutral stable strategy (NSS, Maynard Smith (1982)) to indirect preferences. This concept is intended to capture the essential features of the following three components of the evolutionary process: (1) *mutations* - which introduces new types and behaviors in the population, (2) *optimization* - agents best respond to the behavior of the population within the limits of their bounded forward looking, and (3) *natural selection* - type composition is updated as successful types replicate. Mutations are modeled by considering an entry of a small group of players of any type who play any strategy. It is assumed that the population continues to play the pre-entry equilibrium, except when observing an opponent with a type that did not exist in the pre-entry distribution; in this latter case the incumbents are assumed to learn to play a best response to the mutants' strategy. Finally, natural selection is modeled by a static stability concept that identifies populations of types in which: (1) all types have the same fitness, (2) any small group of mutants that enter the population fare worse than the incumbents.

Suppose that the distribution of types in the population is given by $\mu \in \Delta(\mathcal{L})$. Let $C(\mu)$ be the support of μ. The interaction can be analyzed via the following two-player Bayesian game, $\Gamma_p(\mu)$. The types of the two players are drawn independently from μ, and each player with independent probability p observes the preferences of the other. With the complementary probability $1-p$, the player observes the uninformative signal $\emptyset$. A strategy for type L_k is a rule that specifies a behavioral *k-grim* strategy in the repeated game for each possible observation: $b_k : \mathcal{L} \cup \emptyset \to \Sigma_k$. We assume that aggregate play in the population corresponds to a symmetric Bayesian-Nash equilibrium of this game. That is, we assume that each individual of type L_k, when the horizon is at most k , plays a best reply to a correct belief about the distribution of his opponents' play. When type k is matched with type k' and plays strategy $\sigma_k \in \Sigma_k$, the expected payoff of k is:

$$pu(\sigma_k, b_{k'}(k)) + (1-p)\, u(\sigma_k, b_{k'}(\emptyset)).$$

The payoff is the average over two possibilities. With probability p, the opponent observes the type k', and with probability $1-p$ he observes $\emptyset$. An equilibrium b is thus characterized by two properties. First, type k chooses a k-grim optimal action $b_k(k')$ conditional on observing that the opponent's type is k':

$$b_k\left(k'\right) \in argmax_{\sigma_k \in \Sigma_k}\left(pu\left(\sigma_k, b_{k'}\left(k\right)\right) + \left(1-p\right)u\left(\sigma_k, b_{k'}\left(\emptyset\right)\right)\right)$$

for each $k' \in C\left(\mu\right)$. Second, type k chooses an optimal k-grim action conditional on observing nothing informative:

$$b_k\left(\emptyset\right) \in argmax_{\sigma_k \in \Sigma_k}\sum_{k' \in C(\mu)}\left(pu\left(\sigma_k, b_{k'}\left(k\right)\right) + \left(1-p\right)u\left(\sigma_k, b_{k'}\left(\emptyset\right)\right)\right)\mu\left(L_{k'}\right),$$

Let $B_p\left(\mu\right)$ denote the set of all Bayesian-Nash equilibria of the game $\Gamma_p\left(\mu\right)$. Given a population distribution $\mu$and an equilibrium $b \in B_p\left(\mu\right)$, the average fitness of type $L_k \in C\left(\mu\right)$ is denoted $\Pi_k\left(\mu|b\right)$ and is given by:

$$\sum_{k' \in C(\mu)}\left[p^2 u\left(b_k\left(k'\right), b_{k'}\left(k\right)\right) + p\left(1-p\right)u\left(b_k\left(k'\right), b_{k'}\left(\emptyset\right)\right)\right.$$

$$\left. + p\left(1-p\right)u\left(b_k\left(\emptyset\right), b_{k'}\left(k\right)\right) + \left(1-p\right)^2 u\left(b_k\left(\emptyset\right), b_{k'}\left(\emptyset\right)\right)\right]$$

This fitness, which depends on the equilibrium played, is the measure of evolutionary success for types. Hence, evolution depends both on the distribution of types and the equilibrium played given this distribution, and the stability definition applies to *configurations* - $\left(\mu, b\right)$, where $b \in B_p\left(\mu\right)$.

A configuration is stable if it is satisfies two conditions. First, it must be balanced: all types present must get the same fitness. If the configuration were not balanced, then some types have higher fitness than others and natural selection would alter the configuration as the former types multiply and the latter types recede. Formally:

Definition 1. Configuration $\left(\mu, b\right)$ is *balanced* if $u_k\left(\mu|b\right) = u_{k'}\left(\mu|b\right)$ for all $L_k, L_{k'} \in C\left(\mu\right)$.

Second, a stable configuration must resist entry by mutants who have some arbitrary type $L_{\tilde{k}}$ and play an arbitrary strategy $\sigma_{\tilde{k}}$. We assume that after the entry, the incumbents continue to play the same against strangers and incumbents, while they play a best response to the mutant's strategy when they observe a new type, which did not exist in the pre-entry population. Formally, the post-entry state of the population is described by a perturbed configuration:

Definition 2. Given configuration $\left(\mu, b\right)$, parameter $\epsilon > 0$, type $L_{\tilde{k}} \in \mathcal{L}$ and strategy $\sigma_{\tilde{k}} = \left(\sigma_{\tilde{k}}\left(k'\right) \in \Sigma_{\tilde{k}}\right)_{k' \in C(\mu) \cup \tilde{k} \cup \emptyset}$ for each $k' \in C\left(\mu\right) \cup \tilde{k} \cup \emptyset$, let the perturbed configuration $\left(\tilde{\mu}, \tilde{b}|\mu, b, \epsilon, L_{\tilde{k}}, \sigma_{\tilde{k}}\right)$ where $\tilde{\mu} \in \Delta\left(\mathcal{L}\right)$ and $\tilde{b}$ is a profile of strategies for the different types in $C\left(\tilde{\mu}\right)$ be as follows:

$$\tilde{\mu}(L_k) = \begin{cases} (1-\epsilon)\cdot\mu(L_k) & k \neq \tilde{k} \\ (1-\epsilon)\cdot\mu(L_{\tilde{k}}) + \epsilon & k = \tilde{k} \end{cases},$$

$$\tilde{b}_k(k') = \begin{cases} b_k(k') & k \neq \tilde{k},\ k' \in C(\mu)\cup\emptyset \\ k\text{-grim best response to } \sigma_{\tilde{k}}(k) & k \neq \tilde{k},\ k' = \tilde{k} \notin C(\mu) \\ (1-\epsilon)\cdot b_k(k') + \epsilon\cdot\sigma_{\tilde{k}}(k') & k = \tilde{k} \in C(\mu),\ ,k' \in C(\mu)\cup\emptyset \\ \sigma_{\tilde{k}}(k') & k = \tilde{k} \notin C(\mu),\ k' \in C(\mu)\cup\tilde{k}\cup\emptyset \end{cases}.$$

Given perturbed distribution $\left(\tilde{\mu},\tilde{b}\right)$, type $L_k \in C(\tilde{\mu})$ and k-grim strategy σ_k, define $\Pi_{k,\sigma_k}\left(\tilde{\mu}|\tilde{b}\right)$ as the expected payoff of an agent of type L_k who plays strategy σ_k against population $\tilde{\mu}$ who plays $\tilde{b}$. A configuration is stable if mutants fare worse than all incumbents in any perturbed configuration. Formally:

Definition 3. Configuration (μ,b) is *stable* if it is balanced and there exists $\epsilon' > 0$, such that for each parameter $0 < \epsilon \leq \epsilon'$, type $L_{\tilde{k}} \in \mathcal{L}$ and strategy $\sigma_{\tilde{k}} = \left(\sigma_{\tilde{k}}(k') \in \Sigma_{\tilde{k}}\right)_{k' \in C(\mu)\cup\tilde{k}\cup\emptyset}$, the mutants fare worse than all incumbents in any perturbed configuration $\left(\tilde{\mu},\tilde{b}\right) = \left(\tilde{\mu},\tilde{b}|\mu,b,\epsilon,L_{\tilde{k}},\sigma_{\tilde{k}}\right)$. That is:

1. If $L_{\tilde{k}} \notin C(\mu)$ then $\Pi_{\tilde{k}}\left(\tilde{\mu}|\tilde{b}\right) \leq \Pi_k\left(\tilde{\mu}|\tilde{b}\right)$ for each $k \in C(\mu)$.

2. If $L_{\tilde{k}} \in C(\mu)$ then $\Pi_{\tilde{k},\sigma_{\tilde{k}}}\left(\tilde{\mu}|\tilde{b}\right) \leq \Pi_{k,b_k}\left(\tilde{\mu}|\tilde{b}\right)$ for each $k \in C(\mu)$.

Our notion of stability extends the direct evolution notion of neutral stability strategy. Recall, that a mixed strategy σ in a normal-form game is *neutrally stable* (NSS, Maynard Smith (1982)) if for every strategy $\tilde{\sigma}$ there exists $\epsilon_0 > 0$ such that for every $0 < \epsilon \leq \epsilon_0$:

$$u(\tilde{\sigma}, \epsilon\sigma + (1-\epsilon)\tilde{\sigma}) \leq u(\sigma, \epsilon\sigma + (1-\epsilon)\tilde{\sigma}).$$

Observe, that when the set of types is a singleton and only includes fully-rational players, then our Definition 3 coincides with Maynard Smith (1982)'s neutral stability.

In Appendix B we present a few alternative notions of stability (including Dekel, Ely, and Yilankaya (2007)'s definition), which differ in the way the population reacts to the entry of the mutants, and in how much mutants are required to fare worse than the incumbents, and we demonstrate that our results are robust to the way in which stability is defined.

4 Results

Our main result gives a sharp prediction for the unique stable configuration in the interval $\frac{A}{(A-1)^2} < p < \frac{A-2}{A-1}$. In this configuration naive players (of type L_1) and sophisticated players (who play as if they were L_3, and start defecting 2-3 stages from then end) co-exist. Formally:

Theorem 1. *Let $\frac{A}{(A-1)^2} < p < \frac{A-2}{A-1}$. A configuration (μ, b) is stable if and only the following three conditions are satisfied:*

1. *Players of type L_1 (dubbed, "naive" players) have frequency $\mu(L_1) = 1 - \frac{1}{p(A-1)}$, and they all play d_1 (play "grim" until the last stage, and defect at the last stage).*
2. *There are no players of type L_2 ($\mu(L_1) = 0$).*
3. *All other players (types L_3 or more, dubbed "sophisticated" players) play the same strategy: d_3 against an observed sophisticated opponent (following "grim" until the last 3 stages, and defecting at the last 3 remaining stages), and d_2 in all other cases.*

Let $\mathcal{C}$ be the sets of configurations that satisfy conditions 1-3 above. The sketch of the proof is presented in the next section, and the formal proof is given in Appendix A.

All the stable configurations in $\mathcal{C}$ are "equivalent" in the sense that they differ only by the types of the sophisticated players (L_3 or higher) but all of them play as if they had type L_3. If one adds arbitrarily small costs for having higher cognitive levels, then there is a unique stable configuration in which all the sophisticated players have type L_3. Formally:

Corollary 1. *Let $f : \mathcal{L} \to \mathbb{R}^+$ be any strictly increasing function on the set of types ($f(L_k) > f(L_{k'}) \Leftrightarrow k > k'$) that represents cognitive costs. For each $\epsilon > 0$, let the ϵ perturbed model be the same as out basic model, except that the fitness of L_k is equal to his total payoff in the repeated Prisoner's Dilemma minus his cognitive cost, which is equal to $\epsilon \cdot f(L_k)$. Then:*

1. *For sufficiently small ϵ the game admits a unique stable configuration that only include naive players of type L_1 (that play d_1) and sophisticated players of type L_3 (that play d_3 against observed sophisticated opponents and d_2 otherwise).*
2. *In the limit when $\epsilon \to 0$, the share of naive players converge to $1 - \frac{1}{p(A-1)}$.*

The corollary follows from simple adaptations to the proof of Theorem 1 (proof is omitted).

Our next result, shows that in the benchmark cases when p is close to 0 and 1 the stable configurations are very different. In both cases, stable configurations must include fully-rational players who, when facing other fully-rational agents, defect at all stages. When p is close to 0, this occurs because the indirect advantage of lower types is too small and they can not exist in a stable configuration (because the probability of being identified 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 "race" is that there must some fully-rational agents in the population.

Formally:

Theorem 2.

1. *Let* $0 \leq p < \frac{1}{(M-2)\cdot(A-1)}$. *Then there exists a unique stable configuration where all players have type* L_M *and they defect at all stages.*

2. *Let* $1 \geq p > \frac{A-1}{A}$. *Then in any stable configuration there is a positive frequency of players of type* L_M*, and these players defect at all stages when observing an opponent of type* L_M*.*

5 Sketches of Proofs

In this section we present a few propositions that imply the results of the previous section, and sketch the intuition behind their proofs. The formal proof are presented in the appendix.

5.1 Stability of the Configurations in $\mathcal{C}$ (Theorem 1 - "if side")

The following proposition shows that the configurations in $\mathcal{C}$ are stable in the interval $\frac{A}{(A-1)^2} < p < \frac{A-1}{A}$.

Proposition 1. *Let* $\frac{A}{(A-1)^2} < p < \frac{A-1}{A}$. *Any configuration* $(\mu, b) \in \mathcal{C}$ *(see Theorem 1) is stable.*

The intuition of Proposition 1 is as follows. First we have to show that b only includes best-responses (given the bounded forward-looking). Naive players (L_1) play their unique dominating strategy - d_1 (as they must play grim when he horizon is larger than 1). Sophisticated players play d_3 against observed sophisticated players. This is optimal (and not d_4) for small enough p. They play d_2 otherwise. This is optimal (and not d_3) if $\mu(L_1)$ is large enough.

Next, we have to show that (μ, b) is balanced. In order to show it, we compare the fitness of naive and sophisticated agents as a function of their opponent. Naive agents succeed more only against an observing sophisticated opponent (who observed their type), because their observed naivety induces an additional round of mutual cooperation. Sophisticated agents fare better in the two other cases: against naive opponents and against an unobserving sophisticated opponent. This implies that there is a unique intermediate level of $\mu(L_1)$ that balance the payoff of the two kinds of players .

Finally, we have to show resistance to mutations. If ϵ more naive players join the populations, then due to the previous argument, naive agents fare worse and the excess frequency of naive players (the mutants) will be eliminated. The same holds for ϵ more sophisticated who join the population and play the same as the existing sophisticated players. Finally, one can show that ϵ sophisticated mutants (type 3 or more) who play different actions fare strictly worse then the incumbents.

5.2 Instability Outside $\mathcal{C}$ (Theorem 1 - "only if side")

The next proposition shows that any configuration outside $\mathcal{C}$ are unstable in the interval $\frac{1}{A-1} < p < \frac{A-2}{A-1}$.

Proposition 2. *Let* $\frac{1}{A-1} < p < \frac{A-2}{A-1}$ *and let* (μ, b) *be a configuration outside* C*. Then* (μ, b) *is not stable.*

The intuition behind the proof is as follows. First, observe that a configuration with a single type is not stable: 1) if the type is L_M, then the entire population defects all the time, and mutants of type L_1 would induce cooperation against them and invade the population; and 2) if the type is $L_k \neq L_M$, then mutants of type L_{k+1} can invade the population and get strictly higher payoff then the incumbents. Let L_k be the smallest type in the population. Then, type L_k must always defect when the horizon is k (as it is common knowledge that all players are rational at that stage), and all other types must defect at horizon $k+1$ (or sooner). The next step is to show that all other types in the population must play d_{k+1} against strangers. This is because if there is type $L_{k'}$ that plays d_l $(l > k+1)$ against strangers, then it implies (assuming that p is not too small) that type $L_{k'}$ fares worse against L_k, then L_k gets against itself (because members of $L_{k'}$ loose at least one round of mutual cooperation when facing unobserved L_1). This implies, that ϵ mutants of type L_k who enter the population and slightly increase L_k's frequency, would fare better than the incumbents of type $L_{k'}$ (before the entry, both types fared the same as the configuration was balanced; after the entry L_k's payoff becomes higher because there are more L_k agents). This contradicts the stability of the configuration.

Next we show that if p is not too close to 1, then all the sophisticated players play d_{k+2} when they observe that their opponent is sophisticated, and that type L_{k+1} cannot exist in the population (as its members would fare strictly worse then the more sophisticated types).

Finally, we show that if $L_k \neq L_1$, and p is not too small, then the population can be invaded by mutants of type L_1 as they would fare strictly better than the incumbents of type L_k (because they induce more mutual cooperation when their type is observed).

5.3 Stable Configuration for Low and High p-s (Theorem 2)

1. **Low p-s:** The configuration that everyone has type L_M (fully-rational) and begin defecting at the first stage is stable because the indirect advantage of naive mutants (with a lower type than L_M) is too small: they strictly lose when their naivety is unobserved, and their naivety is observed too rarely. Due to a similar argument, in any other configuration where different types co-exist, the lower type would fare strictly worse (and this implies the uniqueness).

2. **High p-s:** Assume to the contrary that no agent in the population ever defects at the first stage. Let L_k be the highest type in the population. Let $l < M$ be the horizon in which L_k players begin defecting when they observe an opponent of type L_K. If p is large enough, their opponent is likely to observe their signal as well and begin defecting at stage l as well. This implies (again for large enough p) that starting to defect one stage earlier is strictly better. This implies that either type $L_k$does not play a best response (if $L + 1 \leq k$) or that mutants with type L_{k+1} who play like type L_k except that they defect one stage earlier against L_k opponents would fare strictly better.

6 Concluding remarks

1. **Other heuristics for long horizons**: In our model we assumed that all players use a "grim" heuristic whenever the horizon is larger than their forward-looking ability. One could relax this assumption by allowing a player to choose his strategy for long horizons from some fixed set of heuristics. For example, the set of possible heuristics might be the strategies with "memory-1" (which depend only on the actions observed in the previous stage). Observe that these "memory-1" strategies include the "grim" heuristics: cooperate at stage 1, and cooperate at any later stage if and only if both players cooperated at the previous stage. A strategy of a player of type L_k in this setup specifies two strategic components for each possible signal about the opponent's type: (1) the heuristic he plays when the horizon is larger than k, and (2) the (unrestricted) strategy he plays when the horizon is at most k. It is immediate to apply our first result (Proposition 1) in this setup, and show that any configuration in $\mathcal{C}$ in which all players choose *grim* as their heuristic is stable. We conjuncture that there are only two sets of

stable configurations in this extended setup: (1) the efficient configurations in $\mathcal{C}$ in which all players use a "nice" (cooperate at the first stage) and "retaliating" heuristic (defect if your opponent defected at the previous stage), and (2) inefficient configurations in which all players defect at all stages (and use "always-defect" heuristics).

2. **Analogy-based expectation equilibrium:** Our model of bounded forward looking types could also be formulated using Jehiel (2005)'s Analogy-Based Expectation Equilibrium (ABEE). In this formulation a player of type L_k bundles all nodes with horizon of at least k into a single analogy class (while fully-differentiating among nodes with horizons smaller than k), and expects his opponent to play the same in all nodes of this class. The requirement that players play a Bayesian-Nash equilibrium in a configuration (restricted by k-grim consistency) is replaced with the requirement that players play an ABEE in a configuration: at each stage every player best-responses to his analogy-based expectations, and expectations correctly represent the average behavior in every class. As in the previous remark: (1) it is immediate to show that every configuration in $\mathcal{C}$ is stable in this formulation, and (2) we conjuncture that there are only two sets of stable configurations in this ABEE formulation: efficient $\mathcal{C}$-like configurations (in which everyone is "nice" and "retaliating" in his non-trivial analogy-class), and inefficient configurations in which all players defect at all stages.

3. **Random continuation probability**: Our model assumes that the repeated interaction has a deterministic constant length, and that players completely ignore this fact when the horizon is too large. These assumptions may seem unrealistic. However, one should note that the model may be a reduced-form for a more realistic interaction with a random length and incomplete information. Specifiably, let $\mathbf{T}$ be the random unknown length of each interaction. Assume that the interaction lasts at least M rounds ($Pr\left(\mathbf{T} \geq M\right) = 1$), and that the continuation probability at each stage ($Pr\left(\mathbf{T} > n | \mathbf{T} > n-1\right)$) is not too far from 1. Bounded forward-looking is modeled in this setup as the stage in which a player becomes aware to the timing of the final period: player of type L_k gets a signal about the final period of the interaction (i.e., about the realization of $\mathbf{T}$) k stages before the end. In this setup, players are not restricted in their strategies (each type may play any strategy at any horizon). As in the previous remarks: (1) it is immediate to see that any configuration in $\mathcal{C}$ is stable, and (2) we conjuncture that other stable configurations are only those in which everyone defects at all stages.

4. **Level 0**: In the main model we do not allow players to belong to "level-0" (L_0) who

follow *grim* strategy at all rounds of the interaction. Such "level-0" players play a strictly-dominated strategy (cooperating at the last stage), and we chose to omit them from the model as such extreme bounded forward-looking may seem implausible. We note that our results are qualitatively robust to the addition of type L_0 in the following sense. All of our results would remain shift a single step backwards: the naive players in the stable configurations in $\mathcal{C}$ would be of type L_0 instead of L_1, and the sophisticated players would look 1-2 steps ahead instead of 2-3 steps.

5. **Other games**: The formal analysis deals only with iterated Prisoner's Dilemma. However, we conjuncture that the results can be extended to other games in which iterated reasoning decreases payoffs. In particular, the extension of our results to "centipede"-like games (Rosenthal (1981)) is relatively straightforward. Such game can represent sequential interactions of gift exchanges. Such interactions were important in primitive hunter-gatherer populations (see, e.g., Haviland, Prins, and Walrath (2007), P. 440), which driven the biological evolution of human characteristics.

A Proofs

A.1 Proposition 1 - Stability of Configurations in $\mathcal{C}$

Proposition. ***1*** *Let* $\frac{A}{(A-1)^2} < p < \frac{A-1}{A}$. *Any configuration* $(\mu, b) \in \mathcal{C}$ *(see Theorem 1) is stable.*

Proof. We divide the proof into three lemmas: we first show that the configuration is balanced (Lemma 1), then we show that the configuration only includes strict best replies (Lemma 2), and finally we show stability against mutations (Lemma 3). □

Lemma 1. *Any configuration* $(\mu, b) \in \mathcal{C}$ *is balanced.*

Proof. Let $q = \mu(L_1)$ be the frequency of the naive players. A naive player gets $(L-1)A+1$ against a naive opponent, and $(L-2)A+1$ against a sophisticated player. A sophisticated player gets $(L-2)A+(A+1)+1 = (L-1)A+2$ against naive, and against a sophisticated opponent he gets : $(L-3)A+3$ if both players identify each other, $(L-3)A+(A+1)+2 = (L-2)A+3$ if only he identifies his opponent, $(L-3)A+0+2$ if only his opponent identifies him, and $(L-2)A+2$ if both players identify each other. The different types get the same payoff if:

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

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

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

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

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

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

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

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

Observe, that for each $p\geq\frac{1}{A-1}$ we get a valid value of $0\leq q\leq1$. □

Lemma 2. *In any configuration $(\mu,b)\in\mathcal{C}$ all types play a best response (thus, these configurations are well defined). Moreover, any deviation that induces a different play on-equilibrium-path, yields a strictly worse outcome.*

Proof. We have to show that all types play a best response (among the k-grim strategies). This is immediate for the naive player, as his only choice is between cooperating and defecting at the last stage, and the latter strictly dominates the former. We have to show that the sophisticated players play best responses. It is immediate that d_2 is a strict best response against naive opponents. 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 payoff of the last 3 rounds):

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

$$q\left(A-1\right)>\left(1-q\right)$$

$$q>\frac{1}{A}$$

Using (1) one obtains:

$$\frac{p\left(A-1\right)-1}{p\left(A-1\right)}>\frac{1}{A}$$

$$pA\left(A-1\right)-A>p\left(A-1\right)$$

$$pA^{2}-pA-A>pA-p$$

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

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

It is immediate that d_2 is also strictly better (against strangers) than any other strategy that induces a different play on-equilibrium-path. We are left with showing that it is strict better for a sophisticated player to play d_3 and not d_4 against a sophisticated opponent (and this immediately implies that d_3 is strictly better against identified sophisticated opponents than any other strategy that induces a different play on-equilibrium-path). This is true if the following inequality holds (focusing on the payoffs of the last 4 rounds, as all preceding payoffs are the same):

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

$$\left(1-p\right)\left(A-1\right)>p$$

$$A-1>Ap$$

$$p<\frac{A-1}{A}$$

□

Lemma 3. *Any configuration* $(\mu, b) \in \mathcal{C}$ *is stable.*

Proof. We have to show resistance against mutations (for all variants of stability presented in this paper). Observe first that naive players fare strictly worse than sophisticated agents against naive opponents (the sophisticated players obtain an additional fitness point by defecting when the horizon is equal to 2). As the configuration is balanced, it immediately implies that sophisticated players fare strictly worse than naive agents against sophisticated opponents. This implies that if ϵ naive (sophisticated) players join the population and play the same as the naive (sophisticated) incumbents, then naive (sophisticated) agents would fare worse (as their number has become larger and they fare worse against themselves). This implies that the mutants would fare worse than all the incumbents. Next, observe that any mutants of type L_2 would fare strictly worse than the sophisticated players: L_2 players would fare the same when they do not observe their opponent, and fare strictly worse when they observe a sophisticated opponent. Finally, if sophisticated mutants enter the population and play differently on-equilibrium-path, then they would earn strictly less due to Lemma (2). □

A.2 Proposition 2 - Instability of Configurations Outside $\mathcal{C}$

Proposition. *2 Let $\frac{1}{A-1} < p < \frac{A-2}{A-1}$ and let (μ, b) be a configuration outside C. Then (μ, b) is not stable.*

The proposition follows immediately from the following lemmas.

Lemma 4. *If any player ever defects in a configuration, then both players defect at all following stages.*

Proof. Assume to the contrary that player 1 defected at some stage of the game. After the deviation it is common knowledge, that player 1 plays rationally. Similarly, it is also common knowledge that player 2 either defects (if his horizon hasn't arrived yet) or play rationally. The fact that the game is dominance solvable then implies that both players must defect at all the following stages. □

The lemma immediately implies the following corollary.

Corollary 2. *It can be assumed without loss of generality that in any configuration, and given any signal about the opponent, all players play with positive probability only strategies d_l for some l.*

Lemma 5. *Let (μ, b) be a configuration. Let type $L_k \in C(\mu)$ be the smallest type in the population. Then: (1) L_k always defects with probability 1 at horizon k; (2) all other types in the population always defect with probability 1 at horizon $k+1$; and (3) if $\frac{1}{(M-2)\cdot(A-1)} < p$* and $\mu(L_k) = 1$ *then the configuration is not stable.*

Proof. It is common knowledge that all types are at least k. This implies that defecting when the horizon is equal to k is the unique strategy that survives iterations of eliminating dominated strategies, and thus all players must defect with probability 1 when the horizon is equal to k given any signal about the opponent (as the strategy profile in a configuration must be a Bayesian-Nash equilibrium). This, in turn, implies that all agent of type higher than k must defect with probability 1 at horizon $k+1$. To prove part (3), observe that if $k < M$, then ϵ mutants of type L_{k+1} would fare strictly better than the incumbents. If $k = M$, ϵ mutants of type L_1 would fare strictly better than the incumbents: with probability $1-p$ the mutant's type is unobservable and he would obtain 1 point less than the incumbents (when facing an incumbent). With probability p the mutant is identified and he obtains $(M-2)\cdot(A-1)-1$ points more than the incumbents. This implies the mutants would earn strictly more than the incumbents if and only if:

$$1 - p < p\left((M-2)\cdot(A-1) - 1\right) \Leftrightarrow$$

$$p > \frac{1}{(M-2)\cdot(A-1)}.$$

□

Lemma 6. *If Players of type L_k are indifferent between playing d_l and $d_{l'}$ against strangers with $k \geq l > l'$ in a stable configuration, then they can not play $d_{l'}$ with positive probability.*

Proof. Assume to the contrary that players of type L_k are indifferent between defecting at horizon l and horizon l' when playing against strangers, and that they play $d_{l'}$ with positive probability. Consider mutants of type L_k who defect at horizon l against strangers (and play the same as the incumbents in all other cases). Such mutants would fare strictly better than the incumbents of type L_k who happen to begin defecting against strangers only at the smaller horizon l': pre-entry both strategies yielded the same payoff, now as there a bit more early defectors, defecting at the larger horizon l is strictly better. This implies that the configuration cannot be stable or adjusting-stable. □

The lemma immediately implies the following corollary:

Corollary 3. *All players only use pure strategies when playing against strangers in a stable configuration.*

Lemma 7. *Let $p < \frac{A-2}{A-1}$. Let (μ, b) be a balanced configuration. Let type $L_{k_1} \in C(\mu)$ be the smallest type in the population. Assume that there exist type k_2 who begins defecting at* *horizon l_2 (with probability 1) against strangers and that $k_1 < l_2 - 1$. Then, the configuration is not stable.*

Proof. It is sufficient to show that type L_{k_1} gets an higher payoff when playing against L_{k_1} than L_{k_2} gets against L_{k_1} (because if this holds, then an entry of ϵ mutants of type L_{k_1} who play like the incumbents would violate stability and adjusting-stability). Due to Lemma 5, type k_1 always begin defecting at stage k_1, and thus gets against itself the following payoff: $(L - k_1)\,A + k_1$. Type k_2 gets 1 more if he observes k_1's type (probability p), but gets $(l_2 - k_1 - 1)\,(A-1) - 1$ less if he doesn't observe (probability $1-p$). Thus type k_1 gets a higher payoff against itself if:

$$((l_2 - k_1 - 1)\,(A-1) - 1)\,(1-p) > (A-2)\,(1-p) > p$$

This holds if:

$$p < \frac{A-2}{A-1}$$

□

Lemma 8. *Let $\frac{1}{A-1} < p < \frac{A-2}{A-1}$. Let (μ, b) be a balanced configuration. Let $L_{k_1} \in C(\mu)$ be the lowest type in the population. Then:*

1. *Type L_{k_1} always plays d_k (begin defecting at horizon k). All other types $L_k \neq L_{k_1}$ play d_{k_1+1} against strangers and against an observed L_{k_1}, and play d_{k_1+2} when they observe an opponent with any other type $L_k \neq L_{k_1}$.*

2. *If $\mu(L_{k_1}) \neq 1 - \frac{1}{p(A-1)}$ then the configuration is not stable.*

3. *If $k_1 > 1$ then the configuration is not stable.*

Proof.

1. Type L_{k_1} behavior is immediately implied from Lemma 5. By Lemma 8, all other types $L_k \neq L_{k_1}$ play d_{k_1+1} against strangers (and against observed L_{k_1}). If $p < \frac{A-1}{A}$, then by a similar argument to the one given in Lemma 2 all other types $L_k \neq L_{k_1}$ play d_{k_1+2} against observed types $L_k \neq L_{k_1}$.

2. Assume first that $\mu(L_{k_1}) = 1$. If $k_1 < N$ then ϵ mutants of type $k_1 + 1$ who always play d_{k_1+1} would earn strictly more than the incumbents. If $k_1 = N$, then we show that ϵ mutants of type L_1 would fare strictly more than the incumbents. The mutants would get 1 less point than the incumbents when their type is unobserved, and $(N-2)\cdot$

$(A-1)-1$ more points if their type is observed. Thus, they would earn strictly more than the incumbents if:

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

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

$$p>\frac{1}{(N-2)\cdot(A-1)},$$

which holds for every $p>\frac{1}{A-1}$ (as $N>2$). We are left with case that $\mu(L_{k_1})<1$. In this case, the balance between the payoff of type L_{k_1} and the higher types implies (in the same way as in the proof of Lemma 1) that $\mu(L_{k_1})=1-\frac{1}{p(A-1)}$.

3. Now assume $k_1>1$. Consider ϵ mutants of type 1 who enter the population. We show that these mutants earn strictly more than type k_1. This is true because type 1 and type k_1 get the same payoff whenever their types are unobserved and their opponent has a type larger than k_1. Type k_1 gets at most 1 more point when the opponent is of type k_1. Type 1 gets at least $A-1$ points more when the opponent identifies him and has type larger than k_1. Thus the mutants would get a strictly higher payoff if:

$(1-\mu(k_1))\,p\,(A-1)>\mu(k_1)$

$p\,(A-1)>\mu(k_1)\,(1+p\,(A-1))$

$\frac{p(A-1)}{1+p(A-1)}>\mu(k_1)=1-\frac{1}{p(A-1)}=\frac{p(A-1)-1}{p(A-1)}$

Which always holds. □

A.3 Theorem 2 - Stable Configurations Near 0 and 1

Theorem. *2*

1. *Let* $0\leq p<\frac{1}{(M-2)\cdot(A-1)}$. *Then there exists a unique stable configuration where all players have type* L_M *and they defect at all stages.*

2. *Let* $1\geq p>\frac{A-1}{A}$. *Then in any stable configuration there is a positive frequency of players of type* L_M, *and these players defect at all stages when observing an opponent of type* L_M.

Proof.

1. We begin by showing the stability of the configuration in which all players have type L_M and they defect at all stages. It is immediate that player best respond to each

other and that the balance property holds. By a similar argument to the proof of Lemma 5, ϵ mutants with type $k < L$ would fare worse than the incumbents if and only if $\frac{1}{(M-2)\cdot(A-1)} < p$. This implies that the configuration in which all players are fully-rational and defect at all stages is table if and only if $p < \frac{1}{(M-2)\cdot(A-1)}$.
Next we show that no other configuration is stable when $p < \frac{1}{A-1}$ (thus, if $\frac{1}{(M-2)\cdot(A-1)} < p < \frac{1}{A-1}$ then no stable configuration exist). By Lemmas 7 and 10, in any stable configuration, all players have wither type L_k and they always play d_k or they have higher types and they play against strangers d_{k+1}. If the entire population has type L_k and $k < M$, then mutants type L_{k+1} would earn strictly more. Otherwise, by similar arguments to those given in the proofs of Lemmas 1 and 2 if $p < \frac{1}{A-1}$ then either the configuration is unbalanced, or the higher types do not play a strictly best reply.

2. Assume first that no fully-rational players exist in the population: $\mu(L_N) = 0$. In this case, fully-rational mutants who defect one stage earlier then an identified opponent with the highest existing type, and imitate that highest existing type in all other cases, would earn strictly more then the incumbents. Now, assume that fully-rational players exist, and that they play d_k when observing a fully-rational opponent for some $k < N$. Consider fully-rational mutants who play d_{k+1} when observing a fully-rational opponent, and imitate incumbents otherwise. Such mutants would get 1 more point against a fully-rational opponent if both players identify each other. If only they identify their fully-rational opponent they would get at most $A-1$ less points. In all other cases they would fare the same as the fully-rational incumbents. This imply that the mutants would fare strictly better if:

$$p > (1-p)\cdot(A-1) \Leftrightarrow$$

$$pA > (A-1) \Leftrightarrow$$

$$p > \frac{A-1}{A}$$

□

B Robustness to Alternative Stability Notions

In this appendix we present a few alternative notions of stability (Subsection B.1), and we demonstrate the robustness of our results to the exact way in which stability is defined (subsection B.1.2). The presentation and the discussion of the alternative definitions may be of a separate interest for readers who are interested in applications of indirect evolutionary approach in other setups.

B.1 Alternative Definitions of Stability

In this subsection we present a few alternative ways to define stability, which differ in: (1) the way mutant's payoff is compared with the incumbent's payoff, (2) the way incumbents react to entry of mutants.

B.1.1 Weak Stability

The requirement in Definition 3 that the mutants must earn lower fitness than any incumbent may be too strong. When the entrants earn more than some incumbents, but less than the mean payoff of the incumbents, the mutants would be eliminated from he population. However, in this case it is not clear in general if the induced decline in the frequency of the incumbents who fare the worst would cause the aggregate behavior to move farther from the equilibrium. In some examples, the aggregate behavior would move back to the equilibrium. One such example, is a variant of the *Rock-Scissors-Paper* game with a victory yielding 1.5 and a loss giving -1. Assume that there are 3 types in the population $\{rock, scissors, paper\}$, and that each type limits its members to only use "its" action (i.e., types directly determine member's actions). Further assume that the pre-entry distribution of types is uniform over all these 3 types. Then an entry of ϵ mutants of type *rock* would violate our notion of stability (Definition 3), because the entrants would fare strictly better than incumbents of type *scissors*. Yet, it is well known that under the *replicator dynamics*, such mutants (that fare, on average, worse than the incumbents) would be eliminated without disturbing the equilibrium (see, e.g., Neeman, 1980).

In what follows we present a weaker notion, according to which entrants are only required to fare on average worse than the incumbents. We interpret this notion as a necessary requirement for evolutionary stability, and we use it to refine our uniqueness results in the next subsection. Formally:

Definition 4. A configuration (μ, b) is *weakly-stable* if it is balanced and there exists $\epsilon_0 > 0$, such that for each $0 < \epsilon \leq \epsilon_0$, $L_{\tilde{k}} \in \mathcal{L}$ and strategy profile $\sigma_{\tilde{k}} = \left(\sigma_{\tilde{k}}\left(k'\right) \in \Sigma_{\tilde{k}}\right)_{k' \in C(\mu) \cup \tilde{k} \cup \emptyset}$, the mutants fare worse on average than the incumbents in any perturbed configuration $\left(\tilde{\mu}, \tilde{b}\right) = \left(\tilde{\mu}, \tilde{b} | \mu, b, \epsilon, L_{\tilde{k}}, \sigma_{\tilde{k}}\right)$. That is:

1. If $L_{\tilde{k}} \notin C\left(\mu\right)$ then $\Pi_{\tilde{k}}\left(\tilde{\mu} | \tilde{b}\right) \leq \sum_{k \in C(\mu)} \mu\left(k\right) \cdot \Pi_{k}\left(\tilde{\mu} | \tilde{b}\right)$.

2. If $L_{\tilde{k}} \in C(\mu)$ then $\Pi_{\tilde{k},\sigma_{\tilde{k}}}\left(\tilde{\mu}|\tilde{b}\right) \leq \sum_{k\in C(\mu)} \mu(k) \cdot \Pi_{k,b_k}\left(\tilde{\mu}|\tilde{b}\right)$.

It is immediate to see any stable configuration is also weakly-stable.

B.1.2 DEY- Stability (Dekel, Ely, and Yilankaya (2007))

A different way one may criticize our definition of stability is our assumption that the incumbents continue playing the same strategy unless they observe a new mutant type. This implies that post-entry the incumbent may play an approximate best response, but not necessarily an exact one. This is because they do not take into account the presence of the ϵ mutants when playing against strangers. Dekel, Ely, and Yilankaya (2007)'s notion of stability takes the opposite assumption: post-entry the incumbents adapt their strategies such that the new strategy profile is again an exact equilibrium. They require that there exists post-entry equilibria in which the incumbents play is only slightly changed (dubbed, approximate *focal* equilibria), and that in all these equilibria the mutants would fare worse then all incumbents.

In order to formally present their notion, we shall introduce some notation. Let $N_{\epsilon_0}(\mu, L_{\tilde{k}})$ be the set of all distributions resulting from entry by no more than ϵ_0 mutants of type $L_{\tilde{k}}$ to population μ. Formally:

$$N_{\epsilon_0}(\mu, L_{\tilde{k}}) = \{\mu' : \mu' = (1-\epsilon)\mu + \epsilon L_{\tilde{k}},\, 0 \leq \epsilon < \epsilon_0\}$$

Let $\delta \geq 0$. Beginning with a configuration (μ, b), and following an entry by at most ϵ_0 mutants of type $L_{\tilde{k}}$ leading to $\tilde{\mu} \in N_{\epsilon}(\mu, L_{\tilde{k}})$, an equilibrium $\tilde{b}^E \in B_p(\tilde{\mu})$ is *δ-focal* if incumbent's behavior is changed by at most δ, that is, $\left|\tilde{b}_k^E(k') - b_k(k')\right| \leq \delta$ (whenever $p > 0$) and $\left|\tilde{b}_k^E(\emptyset) - b_k(\emptyset)\right| \leq \delta$ (whenever $p < 1$) for all $L_k, L_{k'} \in C(\mu)$.[10] An equilibrium is *focal* if it is 0-focal. Notice that when the entrants have a new type ($L_{\tilde{k}} \notin C(\mu)$), then a δ-focal equilibrium does not restrict the behavior of entrants, nor does it restrict the behavior of incumbents when they observe that they have been matched with mutants. Let $B_p^{\delta}(\tilde{\mu}|b)$ denote the set of all δ-focal equilibria relative to b if the post-entry distribution is $\tilde{\mu}$.

Dekel, Ely, and Yilankaya (2007) assume that any focal equilibrium can potentially arise following a mutation. Thus, *DEY-stability* requires that in all of them, entrants earn no higher fitness than any incumbent. However, not all post-entry populations will have focal equilibria. In that case, they require that approximate focal equilibria exist, and that in all of them entrants earn no higher fitness than any incumbent. Formally:

[10] We simplified the original more complex notion of almost-focal equilibria in Dekel, Ely, and Yilankaya (2007), using the fact that the set of possible types in our setup is finite.

Definition 5. A configuration (μ, b) is *DEY-stable* if it balanced and for every $\delta_0 > 0$ there exist $0 \le \delta \le \delta_0$ and $\epsilon > 0$ such that for every $L_{\tilde{k}} \in \mathcal{L}$ and $\tilde{\mu} \in N_\epsilon(\mu, L_{\tilde{k}})$, $u_k\left(\tilde{\mu}|\tilde{b}\right) \ge u_{\tilde{k}}\left(\tilde{\mu}|\tilde{b}\right)$ for all $\tilde{b} \in B_p^\delta(\tilde{\mu}|b)$ and $L_k \in C(\mu)$.

We think that Dekel, Ely, and Yilankaya (2007)'s assumption that the incumbents adjust their play to have an exact post-entry equilibrium in less plausible when the mutants have a type and play a strategy that was already existed before the entry. Such ϵ mutants can enter the population very quickly. In particular, they may not be "mutants" in the biological sense, they may simply be the result of a small difference between the realized number of offspring of some type and its expectation. Such a small change in the distribution of play, which does not introduce of a new type or a new strategy, is unlikely to be unnoticed, and the incumbents may not have an opportunity to adjust their play in this case.

Moreover, Dekel, Ely, and Yilankaya (2007)'s adjustment causes their notion to be inconsistent with Maynard Smith (1982)'s neutral stability, and to predict counter-intuitive result in the following example: a single-stage 2-strategy (T and B) symmetric coordination game with a single fully-rational type. One can see that the mixed symmetric equilibrium that gives equal weight to both actions is DEY-stable, while it is not stable according to our definition, nor it is neutral stable à la Maynard Smith (1982). Consider, an entry of ϵ mutants who play T. Such an entree is adjusted by the incumbents slightly decreasing the probability of choosing T from 0.5 to $0.5 - \epsilon$). Such a prediction is counter-intuitive: first, it is not clear how the incumbents become aware about the mutants, and second, the incumbents adjust by playing less often the better strategy (the one that before the adjustment yielded a strictly higher payoff).

B.1.3 Adjusting-Stability

When the mutants have a new type (or play a new strategy that was never played before), the adjustment to a new exact equilibrium seems more plausible, as the incumbents immediately observe things (types or actions) that were never existed before, thus the entree is quickly recognized. With this intuition we offer the following definition which is a mixture of our original definition (Definition 3) and DEY-stability. Specifically, when checking if a configuration is *adjusting-stable* we assume that post-entry behavior of incumbents is unadjusted (perturbed configuration as in Definition 2) if the mutants both have a type and play a strategy that existed in the pre-entry population, and it is adjusted into an approximate focal equilibrium if the type or the strategy are new (formal definition is straightforward and it is omitted for brevity).

B.1.4 Undominated-Stability

In our solution concept we assume that incumbents best reply to observed mutants (who belong to new types that have not existed in the population before). One can relax this assumption, and allow the incumbents to play against observed mutants any strategy that is not strictly dominated. With this relaxed assumption, one can define *undominated-stablity,* by requiring that mutants would fare worse than incumbents in any perturbed configuration and given any undominated strategy the incumbents play against observed mutants (not necessarily a best reply). It is immediate to see that undominated-stablity implies stability.

B.2 Robustness of The Results

In this subsection we show that our results are robust to other definitions of stability:

1. The stable configurations (the set $\mathcal{C}$) that we characterized in Theorem 1 are also stable according to all other definitions.
2. All other configurations (outside $\mathcal{C}$) are also unstable according to the adjusting-stability and undominated-stability variants (in the same interval of p-s).
3. The weak-stability and DEY-stability variants may induce other stable configurations, but all of them (in the interval $\max\left(\frac{1}{(M-2)\cdot(A-1)}, \frac{1}{(A-1)^2}\right) < p < 1 - \frac{2\cdot A-1}{A^2-A}$) would have similar properties to the configurations in $\mathcal{C}$: the population will include naive players of type L_1 and sophisticated players of higher types who start defecting only 2-4 stages before the end.

B.2.1 Adjusting-stability and Undominated-stability

It is immediate to see that all the proofs of our results (Appendix A) hold also for adjusting stability and undominated stability. In particular, the only place in the proofs that we relied on incumbents best-responding to observed mutants was in part (3) of Lemma 8, where entry of mutants of type L_1 was considered. In this case, the incumbents have a unique dominating strategy against identified mutants, and thus the proof remains unchanged.

B.2.2 Weak-Stability and DEY-Stability

Lemmas 1-5 hold for weak-stability and DEY-stability with minor adaptations. In particular, it is immediate that any configuration in $\mathcal{C}$ is also weakly-stable and DEY-stable. However, the proofs of Lemmas 6-7 are not valid for the setup of weak-stability and DEY-stability.

The following proposition shows that any weak-stable or DEY-stable configuration has key properties that are qualitatively similar to the configurations in $\mathcal{C}$: the population will include naive players of type L_1 and sophisticated players of higher types who start defecting 2-4 stages before the end.

Proposition 3. *Let max* $\left(\frac{1}{(M-2)\cdot(A-1)}, \frac{1}{(A-1)^2}\right) < p < 1-\frac{2\cdot A-1}{A^2-A}$ *and let* (μ, b) *be a weakly-stable (resp., DEY-stable) configuration. Then:*

1. $0 < \mu(L_1) < 1$. *Players of type* L_1 *always play* d_1.
2. *All other types* $L_k \neq L_1$ *only start defecting at the last four rounds.*

Proof. The proposition follows from the following three lemmas. □

Lemma 9. *Let* (μ, b) *be a weakly-stable (resp,. DEY-stable) configuration, let* $L_{k_1} \in C(\mu)$ *be the lowest type in the population and let* $p > \frac{1}{(M-2)\cdot(A-1)}$. *Then* $\mu(L_k) < 1$, *and the mean probability that a random player with type different than* L_{k_1} *plays* d_{k_1+1} *against strangers (denoted by* q*) is at-least:*

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

.

Proof. $\mu(L_{k_1}) < 1$ is implied by Lemma 5 (which is valid also for weak-stability and DEY-stability). Type L_{k_1} gets $(L-k_1)\cdot A+k_1$ 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 $(L-k_1)\cdot A+k_1+1$ when he observes his opponent's type, and an expected payoff of at most $q((L-k_1)A+k_1+1)+(1-q)\cdot((L-k_1-1)A+k_1+2)$. This implies that other types fare better when playing against L_{k_1} than the payoff that L_{k_1} gets against itself only if (subtracting the equal amount of $(L-k_1-1)\cdot A+k_1$ from each payoff):

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

$$A \leq 1+p\cdot A+(1-p)\cdot(q\cdot A+1-q)$$

$$A-\frac{1}{1-p} \leq q\cdot A+1-q$$

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

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

$$\frac{(A-1)\cdot(1-p)-1}{(A-1)\cdot(1-p)}\leq q \tag{2}$$

The configuration may be weakly-stable (resp., DEY-stable) only if $\frac{(A-1)\cdot(1-p)-1}{(A-1)\cdot(1-p)}\leq q$. □

Lemma 10. *Let (μ,b) be a weakly-stable (resp., DEY-stable) configuration, let $L_{k_1}\in C(\mu)$ be the lowest type in the population, and let $1-\frac{2\cdot A-1}{A^2-A}>p>\frac{1}{(M-2)\cdot(A-1)}$. Then:*

1. *No player defects with positive probability against strangers at horizon strictly larger than k_1+2.*

2. *No player defects with positive probability against any other player at horizon strictly larger than k_1+3.*

Proof.

1. Assume to the contrary that there is a type who defects with positive probability against strangers at horizon $l>k_1+2$. This implies that d_l yields a weekly better payoff against strangers than d_{k_1+2}. This can occur only if:

$$q\cdot(1-p)\cdot(A-1)\leq((1-q)+qp)\cdot 1$$

$$q\cdot(1-p)\cdot(A-1)\leq 1-q\cdot(1-p)$$

$$q\cdot(1-p)\cdot A\leq 1$$

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

Substituting (2) yields:

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

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

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

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

$$1-p\leq\frac{2\cdot A-1}{A\cdot(A-1)}$$

$$p \geq 1 - \frac{2 \cdot A - 1}{A^2 - A},$$

and we get a contradiction to $p < 1 - \frac{2 \cdot A - 1}{A^2 - A}$.

2. Strategy d_l ($l > k_1 + 3$) may yield a better payoff than d_{k_1+3} only if:

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

$$A - 1 \leq p \cdot A$$
$$\frac{A - 1}{A} \leq p.$$

The latter inequality cannot hold because $1 - \frac{2 \cdot A - 1}{A^2 - A} < \frac{A-1}{A}$ for every $A > 1$.

□

Lemma 11. *Let (μ, b) be a weakly-stable (resp. DEY-stable) configuration, let $L_{k_1} \in C(\mu)$ be the lowest type in the population, and let $1 - \frac{2 \cdot A - 1}{A^2 - A} > p > max\left(\frac{1}{(M-2) \cdot (A-1)}, \frac{1}{(A-1)^2}\right)$. Then:*

1. *If every type different than L_{k_1} always plays d_{k_1+1} against strangers, then the configuration must be in $\mathcal{C}$.*

2. *If there are types who play with positive probability $d_{k_1} + 2$ against strangers than $\mu(L_{k_1}) \leq \frac{1}{A}$.*

3. *k_1 must be equal to 1.*

Proof.

1. If everyone besides type L_{k_1} always plays d_{k_1+1} against strangers, then we can immediately apply the arguments in the proof of Lemma 8, and conclude that the configuration must be in $\mathcal{C}$.

2. The fact that there are types who play with positive probability $d_{k_1}+2$ against strangers implies that $d_{k_1} + 2$ yields a weakly-better payoff than d_{k_1+1} against strangers. This implies that:

$$\mu(L_{k_1}) \cdot (A - 1) \leq (1 - \mu(L_{k_1})) \cdot 1$$
$$\mu(L_{k_1}) \leq \frac{1}{A}. \tag{3}$$

3. Assume to the contrary that $k_1 > 1$. Observe that ϵ mutants of type L_1 would fare strictly better than the incumbents of type L_{k_1}(and thus would fare strictly better than

all incumbents in the post-entry perturbed-configuration / approximate focal equilibrium) if:

$$p \cdot (A-1) \cdot (1-\mu(L_{k_1})) > \mu(L_{k_1}) \cdot 1$$

$$p \cdot (A-1) > \mu(L_{k_1}) \cdot (1+p \cdot (A-1))$$

$$\frac{p \cdot (A-1)}{1+p \cdot (A-1)} > \mu(L_{k_1}).$$

Substituting (3) yields:

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

$$p \cdot A \cdot (A-1) > 1+p \cdot (A-1)$$

$$p > \frac{1}{(A-1)^2}$$

□

References

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

Beard, T., and R. Beil Jr (1994): "Do people rely on the self-interested maximization of others? An experimental test," *Management Science*, pp. 252–262.

Bosch-Domenech, A., J. Montalvo, R. Nagel, and A. Satorra (2002): "One, two,(three), infinity,...: Newspaper and lab beauty-contest experiments," *The American Economic Review*, 92(5), 1687–1701.

Bruttel, L., W. Güth, and U. Kamecke (2012): "Finitely repeated prisoners dilemma experiments without a commonly known end," *International Journal of Game Theory*, pp. 1–25.

Camerer, C. (2003): *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

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

Chong, J., C. Camerer, and T. Ho (2005): "Cognitive hierarchy: A limited thinking theory in games," *Experimental Business Research*, pp. 203–228.

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, 187–218.

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

Crawford, V., and N. Iriberri (2007): “Level-k Auctions: Can a Nonequilibrium Model of Strategic Thinking Explain the Winner’s Curse and Overbidding in Private-Value Auctions?,” *Econometrica*, 75(6), 1721–1770.

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,” .

Gill, D., and V. Prowse (2012): “Cognitive ability and learning to play equilibrium: A level-k analysis,” .

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, pp. 23–34. The University of Michigan Press, Ann Arbor.

Haviland, W., H. Prins, and D. Walrath (2007): *Cultural anthropology: the human challenge*. Wadsworth Pub Co.

Heifetz, A., and A. Pauzner (2005): “Backward induction with players who doubt others’ faultlessness,” *Mathematical Social Sciences*, 50(3), 252–267.

Ho, T., C. Camerer, and K. Weigelt (1998): “Iterated dominance and iterated best response in experimental" p-beauty contests",” *The American Economic Review*, 88(4), 947–969.

Hyndman, K., A. Terracol, and J. Vaksmann (2012): “Beliefs and (In) Stability in Normal-Form Games,” .

Jehiel, P. (2005): “Analogy-based expectation equilibrium,” *Journal of Economic theory*, 123(2), 81–104.

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. (1982): "Evolution and the theory of games," .

McKelvey, R., and T. Palfrey (1995): "Quantal response equilibria for normal form games," *Games and Economic Behavior*, 10(1), 6–38.

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

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.

Rapoport, A., and W. Amaldoss (2004): "Mixed strategies and iterative elimination of strongly dominated strategies: An experimental investigation of states of knowledge," *Journal of Economic Behavior & Organization*, 42(4), 483–521.

Robson, A. (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., and P. Wilson (1994): "Experimental evidence on players' models of other players," *Journal of economic behavior & organization*, 25(3), 309–327.

Stahl Dale, O. (1993): "Evolution of Smartn Players," *Games and Economic Behavior*, 5(4), 604–617.

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