Three steps ahead

Heller, Yuval

University of Oxford, Department of Economics

13 June 2012

Three Steps Ahead

Yuval Heller[*] *(August 31, 2012)*

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

Experimental evidence suggests that most people use only a few iterations of strategic reasoning, and that some people systematically use fewer iterations than others. In this paper, we present a novel evolutionary foundation for these stylized facts. In our model, agents interact in a 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 "moderately sophisticated" agents who think up to 3 steps ahead.

KEYWORDS: Indirect evolution, evolutionary stability, proper equilibrium, cognitive hierarchy, bounded forward-looking, Prisoner's Dilemma, Cooperation. JEL Classification: C73, D03.

1 Introduction

Experimental evidence suggests that in new strategic interactions most people use only a few iterations of strategic reasoning. This can be observed in different forms in different contexts. First, when playing long finite games, people look only 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 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 bargaining opportunities that are more than 1-2 steps

---

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

ahead when interacting in sequential bargaining (Neelin, Sonnenschein, and Spiegel (1988); Johnson, Camerer, Sen, and Rymon (2002)). Secondly, when facing iteratively dominated strategies, almost everyone completes the first iteration (not playing a dominated action), many complete the second iteration (assuming that their opponent does not play dominated strategies), a few complete the third iteration, while further iterations are rare (Costa-Gomes, Crawford, and Broseta (2001); Rapoport and Amaldoss (2004); Costa-Gomes and Crawford (2006)). Third, according to cognitive hierarchy (or level-k) models, most players best reply to a belief that others use at most two iterations of strategic reasoning (Stahl and Wilson (1994); Nagel (1995); Ho, Camerer, and Weigelt (1998); Bosch-Domenech, Montalvo, Nagel, and Satorra (2002); Camerer, Ho, and Chong (2004); Crawford and Iriberri (2007)).

A second stylized fact is the heterogeneity of the population: some people systematically use fewer iterations than others (Chong, Camerer, and Ho (2005); Costa-Gomes and Crawford (2006)). These observations raise two related evolutionary puzzles. The first puzzle is why people use only a few steps. Several experiments (e.g., Crawford (2008); Camerer (2003, Section 5.3.5)) 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 more iterations. In many games, the ability to complete one more step than one's opponent gives a substantial advantage. As the cognitive cost of an additional level is moderate, it is puzzling why there was not an "arms race" in which people learned to use more strategic iterations throughout the evolutionary process (the so called "red queen effect"; Robson (2003)).

The second puzzle is how the "naive" people, who systematically use fewer iterations than the more "sophisticated" agents, survived the evolutionary process. At first glance, it seems that sophisticated agents would outperform naive agents due to 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: a heterogeneous population of naive agents and moderately sophisticated agents, in which everyone completes only 1-3 strategic iterations.* Our model focuses on bounded forward-lookingness in the repeated Prisoner's Dilemma. 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 of a dynamic process that describes the evolution of types in a large population of agents.[1] This process can be interpreted in two different ways: (1) biological process - types are genetically determined, and the payoff is the expected number

---

[1] The indirect approach was mainly used to study evolution of preferences, and it is related to the literature on strategic delegation (e.g., Fershtman, Judd, and Kalai (1991)). Following, Stahl (1993), Stennek (2000), and Frenkel, Heller, and Teper (2012), we apply it to analyze the 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 ($A > 3.15$).

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.

In each generation the agents in the population 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: mutual cooperation (both players play C) yields both players $A > 3.15$, mutual defection (both players play D) gives 1, and if a single player defects, he obtains $A + 1$ and his opponent gets 0. Note that the parameter A is the ratio between what can be gained by mutual cooperation and the additional payoff that is obtained by defecting.[2] [3]

Each agent in our model has a type (level) in the set $\{L_1, ..., L_M\}$ that determines how many steps he looks ahead. 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. We assume that this heuristic satisfies two properties: (1) "nice" (never be the first player to defect), and (2) "retaliating" - defect if the opponent defected in the previous stage. Two examples for such heuristics are "grim" and "tit-for-tat."[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-lookingness: 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 of a random unknown long length, in which, k periods before the end, each type L_k gets a signal about the interaction's realized length (as discussed in Section 7). Note that the set

---

[2] All our results are independent of the value of M (given that $M \geq 4$). The assumption $A > 3.15$ is required for the solution we characterize below to be evolutionarily stable in a nonempty interval of p-s.

[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. 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] *Grim* heuristic defects if and only if the opponent defected in the past, and *Tit-for-tat* heuristic defects if and only if the opponent defected in the previous stage. In Section 7 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)).

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 is not limited in his choice of strategy.

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.[5] Informally, this can be interpreted as an opportunity to observe your opponent's past behavior, or to observe a trait that is correlated with the cognitive level (such as I.Q. level; see Gill and Prowse (2012)). 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.[6]

Evolutionary stability can be sustained by playing very badly when facing types outside the support of the distribution. However, this is unlikely to be stable in the long run, as the strategies played against non-existing types should slowly evolve into better strategies, as a response to recurrent entries of mutants. Thus, we refine evolutionary stability by also requiring *properness* (Myerson (1978)). In Section 6 we show that our results are robust to various plausible changes in the definition of stability in this setup.

Our main result shows that if p is not too close to 0 and 1 (and this interval is increasing in A), then there exists a unique proper evolutionarily stable configuration, which includes two kind of players: (1) *naive* agents of type L_1 who only begin defecting at the last stage, (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. The stability relies on the balance between the direct disadvantage of naive agents (they defect too late), and the indirect "commitment" advantage (when naivety is observed by a moderately sophisticated opponent, it serves as a commitment device that allows an additional round of mutual cooperation). The proportion of naive players is increasing in both p and A.

---

[5] The results remain the same if agents were able to observe only lower opponents' type (see Section 6).

[6] 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.

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. The characterization of stable configurations in the entire interval $p \in [0,1]$ is presented in Table 2 in Section 5.

Existing evolutionary models that studied bounded strategic reasoning (Stahl (1993); Stennek (2000)) focused on the case where types are unobservable ($p = 0$) , and showed that in various games: (1) the most sophisticated type always survives, and (2) lower (more naive) types can also survive if they do not play serially dominated strategies. Recently, Mohlin (2012) showed that there may be evolutionarily stable configurations in which the highest type does not survive, and he also studied the case in which higher types can perfectly observe lower types (a case similar to $p = 1$; see Section 6).[7] This paper is the first to introduce partial observability in this setup. Partial observability yields to a sharp and qualitative different prediction: only naive and moderately sophisticated agents survive.

Existing experimental results verify the plausibility of both our assumption of using "nice" and "retaliating" heuristics for large horizons, and of our main prediction. Selten and Stoecker (1986) study the behavior of players in iterated Prisoner's Dilemma games of 10 rounds.[8] They show that: (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.[9] Such behavior has two main explanations in the literature: (1) some players are altruistic, and (2) players have limited forward-looking ability.[10] Johnson, Camerer, Sen, and Rymon (2002) studied the relative importance of these explanations in a related sequential bargaining game, and their findings suggest that limited forward-lookingness is the main cause for this behavior.

A recent qualitative support for our result is given in Hyndman, Terracol, and Vaksmann

---

[7] See also Crawford (2003) for a strategic (non-evolutionary) model of zero-sum games with "cheap talk" in which naive and sophisticated agents may co-exist and obtain the same payoff.

[8] Similar results are presented in Andreoni and Miller (1993); Cooper, DeJong, Forsythe, and Ross (1996); Bruttel, Güth, and Kamecke (2012).

[9] In Selten and Stoecker's (1986) experiments players 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"). 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.

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

(2012), which experimentally studied the strategic behavior of people across different games. They showed that a fraction of the players consistently assign a low level of reasoning to their opponent, while the remaining players alternate between different assessments of their opponent's cognitive skill. The former fraction corresponds to "naive" agents in our model who best reply to a belief that the opponent does not have any forward-looking ability. The remaining players correspond to the "moderately sophisticated" agents who, depending on the signal they obtain, best reply to different beliefs about the opponent's ability.

The paper is structured as follows. Section 2 presents our model. In Section 3 we show that the configuration described above, in which levels L_1 and L_3 co-exist, is evolutionarily stable. In Section 4 we prove uniqueness. Section 5 characterizes the stable configurations for low and high p-s. Section 6 demonstrates that our results are robust to various changes in the model. We conclude in Section 7.

2 Model

2.1 Strategies and Types

We study a symmetric finitely repeated Prisoner's Dilemma game that repeats M stages ($M \geq 4$). The payoffs of each stage game are described in Table 1. As is standard in the evolutionary literature, this payoff is interpreted 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 m is equal to $M - m + 1$. History h_m of length m is a sequence of m pairs, where the l-th pair describes the actions chosen by the players at stage l. Let H_m be the sets of histories of length m, and let $H = \cup_{1 \leq m < M} H_m$ be the set of all nonterminal histories.

A pure strategy s is a function from H into $\{C, D\}$, and a behavioral strategy σ is a function from H into $\Delta(\{C, D\})$. Let S (Σ) be the set of pure (behavioral) strategies. With some abuse of notation, we also treat behavioral strategy $\sigma \in \Sigma$ as a mixed strategy, and denote by $\sigma(s)$ the probability that strategy σ assigns to the pure strategy s. Behavioral strategy (henceforth, strategy) σ is *k-nice-retaliating* 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 previous stage. Let Σ_k (S_k) be the set of *k-nice-retaliating* behavioral (pure) strategies. Let $d_k \in S_k$ be the pure strategy that plays *grim* as long as the horizon is strictly larger than k (he defects if and only if his opponent has defected in the past), and always defects at smaller horizons. Let $\mathcal{D} = \{d_k\}_{0 \leq k \leq M}$ be the set of all such "grim-then-defect" strategies. Let $u(\sigma, \sigma')$ be the expected payoff of a player who plays strategy σ against an opponent who plays σ'.

We imagine a large population in which in each generation agents are randomly matched, and each pair of agents play the repeated Prisoner's Dilemma. Different agents in the population differ in their cognitive abilities, which is captured by their type. Let $\mathcal{L} = \{L_1, ..., L_M\}$ be the set of types (or levels).[11] 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 a nice and retaliating heuristic . That is, an agent with type L_k can play only *k-nice-retaliating* strategies. Note that when the horizon is at most k, the agent is no longer limited in his play.

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 $\delta \cdot c(L_k)$. In what follows, we focus on the case where δ is sufficiently small (cognitive costs are arbitrarily low). The payoff of the repeated game is the undiscounted sum of the stage payoffs minus the cognitive cost.

Following the model of partial observability of Dekel, Ely, and Yilankaya (2007), we assume that 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. In Section 6 we demonstrate that our results remain the same also if agents can observe only lower types. Finally, we use the term *stranger* to describe an opponent whose type was not observed.

2.2 Configurations

The state of the population is described by a *configuration* - a pair consisting of a distribution of types and the strategy that each type uses in the game. Formally (where $C(\mu)$ denotes the support of μ):

Definition 1. *Configuration (or population)* (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 signal-dependent behavioral strategies that is played by each type. That is, for each type $L_k \in C(\mu)$, $\beta_k : \mathcal{L} \cup \emptyset \to \Sigma_k$ is a signal-dependent strategy that specifies a behavioral *k-nice-retaliating* strategy for each possible observation about the opponent's type (including observations outside $C(\mu)$).

Given a configuration (μ, b), we call the types in $C(\mu)$ *existing types* or *incumbents*, and types outside $C(\mu)$ are called *non-existing* types or *mutant* types.

Remark 1. We note two points regarding Definition 1:

---

[11] We explicitly omit level 0 (L_0, who uses a nice and retaliating heuristic throughout the entire interaction). The results are qualitatively the same if L_0 is included (see Section 6.1.1).

- A configuration also determines the strategies that are used against non-existing types. In Section 6 we propose an alternative stability notion, according to which, the state of the population determines only the strategies that are used against existing types (similar to the definition of a configuration in Dekel, Ely, and Yilankaya (2007)).
- Agents of type L_k can use a behavioral (non-pure) strategy. As usual in such models, this can be interpreted as either: (1) each agent randomly chooses his actions, or (2) different fractions of type L_k play different pure strategies.

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:

- $\tilde{\mu} = (1-\epsilon) \cdot \mu + \epsilon \cdot \mu'$.
- For each $k \in C(\tilde{\mu})$:

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

When ϵ is small we interpret $(1-\epsilon) \cdot (\mu, \beta) + \epsilon \cdot (\mu', \beta')$ as a *post-entry configuration* after the population of incumbents in state (μ, β) is 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. $\mu = \mu'$.
2. For each pair of types $L_k, L_{k'} \in C(\mu)$, 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 very 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 "mutant" strategy σ' (resp., $\sigma' \neq \sigma$) there exists some $\epsilon_{\sigma'} \in (0,1)$ such that for every $0 < \epsilon < \epsilon_{\sigma'}$:

$$u\left(\sigma, \epsilon\sigma' + (1-\epsilon)\sigma\right) \geq u\left(\sigma', \epsilon\sigma' + (1-\epsilon)\sigma\right).$$

(resp., $u\left(\sigma, \epsilon\sigma' + (1-\epsilon)\sigma\right) > u\left(\sigma', \epsilon\sigma' + (1-\epsilon)\sigma\right)$).

In what follows we extend the notion of evolutionary stability from strategies to configurations. Given two configurations (μ, b) and (μ', b') define $u\left((\mu, b), (\mu', b')\right)$ as the expected payoff of a player from population (μ, b) who plays against an opponent from population (μ', b') (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 "mutant" configuration (μ', β') (resp., any $(\mu', \beta') \not\approx (\mu, \beta)$) there exists some $\epsilon_{\sigma'} \in (0,1)$ such that for every $0 < \epsilon < \epsilon_{\sigma'}$:

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

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

Remark 2. Note that:

1. Any evolutionarily stable configuration is also neutrally stable.
2. Evolutionarily stable configurations are only weakly stable against invasions of mutants who are equivalent to the incumbents (have the same distribution of types and play the same on-equilibrium path).
3. 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 type L_k and a signal-dependent *k-nice-retaliating* strategy. Note that a mixed 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.[12]

With some abuse of notation we denote by L_k also the distribution that assigns mass 1 to type L_k. It is well known that any neutrally stable strategy is a Nash equilibrium. Proposition 4 (in Appendix A.1) shows that the strategy profile b in neutrally stable configuration (μ, b) is: (1) *balanced* - all incumbents obtain the same payoff, and (2) a Bayes–Nash equilibrium in the Bayesian game with a distribution of types μ.

3.2 Result (Stability)

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:

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

2. The "naive" agents of type L_1 play d_1 : play "grim" until the last stage, and defect at the last stage.

3. The "moderately sophisticated" agents of type L_3 play:

    (a) d_2 against strangers and observed L_1 (follow "grim" until the last 2 stages, and defect at the last 2 remaining stages).

    (b) d_3 against any observed type different from L_1.

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

The formal proof appears in Appendix A.2. In what follows we briefly sketch the outline of the proof. First, we show that b^* is a Bayes–Nash equilibrium (given μ^*), and that type L_3's strategy is a best reply also in the entire unrestricted set of strategies Σ (and not only in Σ_3). It is immediate that naive players best reply (in Σ_1). For moderately sophisticated

---

[12] 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.

(henceforth, sophisticated) players, playing d_3 against non-naive opponents is strictly better than an earlier defection for small enough p, and playing d_2 against strangers and naive opponents is strictly better than earlier defections if $\mu(L_1)$ is large enough.

Next, we show that (μ^*, b^*) is balanced. In order to show this, we compare the fitness of naive and sophisticated agents against different opponents. Naive agents succeed more against an observing sophisticated opponent (who observed their type), because their observed naivety induces an additional round of mutual cooperation. Sophisticated 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.

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.

4 Uniqueness

4.1 Properness

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 fully rational agents (type L_M) who deviate 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. 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 neutral stability by requiring properness (Myerson (1978)). We begin by formally defining properness in this setup. A configuration is interior if every type has positive probability and every pure strategy (consistent with the type) is played with positive probability.

Definition 6. Configuration (μ, β) is *interior* if for each types $L_k, L_{k'} \in \mathcal{L}$ and for each *k-nice-retaliating* pure strategy $s_k \in S_k$: (1) $\mu(L_k) > 0$, (2) $\beta_k(\emptyset)(s_k) > 0$, and (3) $\beta_k(k')(s_k) > 0$.

Given some $\epsilon > 0$, an interior strategy configuration (μ, β) is *ϵ-proper* if for every type $L_k, L_{k'} \in \mathcal{L}$ and every signal-dependent k/k'-nice-retaliating signal-dependent pure strategy $b_k \in S_k$ and $b_{k'} \in S_{k'}$:

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

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

Definition 7. Configuration (μ, β) is proper, if for some sequence $(\epsilon_n)_{n\geq 1}$, there exist ϵ_n-proper interior configurations (μ^n, β^n) such that $(\mu^n, \beta^n) \to (\mu, \beta)$. That is, $\mu^n \to \mu$, and for each type L_k satisfying $\mu(L_k) > 0$ and for each pure strategy $s_k \in S_k$: (1) $\beta_k^n(\emptyset)(s_k) \to \beta_k(\emptyset)(s_k)$, and (2) for each $L_{k'} \in \mathcal{L}$:$\beta_k^n(k')(s_k) \to \beta_k(k')(s_k)$.

Remark 3. Note that:

1. It is immediate to see that in every proper configuration (μ, b) the profile b is a balanced Bayes–Nash equilibrium.
2. Definition 7 is closely related to Myerson's (1978) definition of proper equilibrium:
   (a) When the set of types is a singleton, then Definition 5 and Myerson's definition coincide.
   (b) Consider again the two-player "meta-game" in which each player chooses type L_k and signal-dependent *k-nice-retaliating* strategy. A symmetric (mixed) strategy profile in this meta-game is a proper equilibrium if and only if it is a proper configuration.
3. van Damme (1987) showed for normal-form games without types that evolutionary stability implies properness. As demonstrated by the configuration in the example above, this is not true in our setup.[13] Note that in any interior configuration, cooperating in the first $M-2$ stages gives a strictly better payoff when facing an observed type L_1. This implies that always defecting against type L_1 cannot be played with positive probability in a proper configuration.

---

[13] Recall that our setup differs in two main aspects: (1) having different types, and (2) a weaker notion of evolutionary stability that allows deviations off-equilibrium path to yield the same payoff as the incumbents.

A configuration is a proper naturally (evolutionarily) stable if it is both proper and naturally (evolutionarily) stable. In Section 6.2 we show that our results remain the same if one replaces properness with a weaker perfection notion that only requires players to use undominated strategies against mutant types.

4.2 Result (Uniqueness)

It is straightforward to show that (μ^*, b^*) is proper (Prop. 5, proved in Appendix A.3). Our next result shows that any proper neutrally stable configuration is equivalent to (μ^*, b^*).[14]

Theorem 2. *Let* $A > 4.57$ *and* $\frac{A}{(A-1)^2} < p < \frac{A-1}{A}$. *Then for a sufficiently small* $\delta > 0$, *if* (μ, β) *is a proper neutrally stable configuration, then it is equivalent to* (μ^*, b^*).

The sketch of the proof is as follows (see Appendix A.4 for the formal proof). 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 induce cooperation against them and, in doing so, outperform the incumbents; and (2) if the type is $L_k \neq L_M$, then mutants of type L_{k+1} 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 (as it is common knowledge that all players are rational at that stage), 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 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

---

[14] The assumption $A > 4.57$ is required to have uniqueness in the entire interval $\frac{A}{(A-1)^2} < p < \frac{A-1}{A}$. For lower A-s (as discussed in the appendix) the uniqueness may hold only in a sub-interval.

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.

5 Stability for 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 characterize the stable configurations in 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 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 cannot exist in a stable configuration (because the probability of being observed by the opponent is too low). For very low values of p-s, it implies that there exists a unique stable configuration in which all agents are fully rational and always defect. For low, but less extreme, values of p-s no stable configurations exist.

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 (if such a configuration exists) there must be some fully rational agents in the population, and these fully rational players defect at the first stage when they observe a fully rational opponent. Formally (see proof in Appendix A.5):

Theorem 3. *Let $\delta > 0$ be sufficiently small and assume that $A > 4.57$. Then:*

1. *Let $0 \leq p < \frac{1}{(M-2)\cdot(A-1)}$. Then there exists a proper evolutionarily stable configuration $\left(\tilde{\mu}, \tilde{b}\right)$ where all players have type L_M and they play d_M against strangers and type L_M, and d_{k+1} against observed "mutant" type $L_k < L_M$. Moreover, any other neutrally stable configuration is equivalent to $\left(\tilde{\mu}, \tilde{b}\right)$.*
2. *Let $\frac{1}{(M-2)\cdot(A-1)} < p < \frac{A}{(A-1)^2}$. Then no proper neutrally stable configurations exist.*
3. *Let $\frac{A}{(A-1)^2} < p < \frac{A-1}{A}$. Then (μ^*, b^*) is a proper evolutionarily stable configuration. Moreover, any other neutrally stable configuration is equivalent to (μ^*, b^*).*
4. *Let $\frac{A-1}{A} < p \leq 1$. Then in any proper neutrally 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.*

Tab. 2: Characterization of Proper Evolutionarily stable Configurations

Interval	Example $(A, M = 10)$	Characterization of Proper Evolutionarily stable Configurations
$0 \leq p < \frac{1}{(M-2)\cdot(A-1)}$	$0 \leq p < 1\%$	$\left(\tilde{\mu}, \tilde{b}\right)$ - Only fully rational agents who defect at all stages.
$\frac{1}{(M-2)\cdot(A-1)} < p < \frac{A}{(A-1)^2}$	$1\% < p < 12\%$	No stable configurations exist.
$\frac{A}{(A-1)^2} < p < \frac{A-1}{A}$	$12\% < p < 90\%$	(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 for stability: fully rational agents exist.

Parts 1 and 4 are proved in Appendix A.5. Parts 2 and 3 are immediately implied by Corollary 1 and Lemma 7 in the Appendix. Table 2 summarizes the properties of stable configurations in the different areas of the interval $p \in [0, 1]$ (for sufficiently small δ).

6 Robustness

In this section we demonstrate that our results are robust to plausible variants in: (1) the types and the signals (Section 6.1), (2) the perfection refinement (Section 6.2), and (3) the stability notion (Section 6.3).

6.1 Variants in the Types and Signals

6.1.1 Level 0

In the model we do not allow players to belong to "level-0" (L_0) who follow a *nice and retaliating* strategy in all rounds of the interaction. Such "level-0" players play a strictly dominated strategy (cooperating at the last stage). We chose to omit them from the model as such an extreme cognitive limitation seems implausible because playing strictly dominated strategies yields substantial losses when playing against Nature.

We note that our results are qualitatively robust to the addition of type L_0. Specifically, the main result (Theorems 1-2) would only shift a single step backwards: the naive players in the unique stable configuration (μ^*, b^*) would have type L_0 instead of L_1, and the moderately sophisticated players would have type L_2.

6.1.2 Asymmetric Type Observability

In the model we assume that any agent has the same probability to observe his opponent's type. In particular, lower types may identify the exact type of a more sophisticated opponent. One may argue (see, e.g., Mohlin (2012)) that it is more plausible that only higher types can identify the types of their opponents. We formalize this alternative assumption as follows. Before the interaction begins each agent independently obtains a signal about his opponent. With probability $1-p$ the signal is non-informative ($\emptyset$). With probability p the signal is informative:

1. If the opponent's type is strictly lower, then the agent exactly identifies it.
2. If the opponent's type is weakly higher, then the agent only observes that his opponent's type is weakly higher than his own type.

One can see that all of our results remain the same in this setup.

6.1.3 Small Perturbations to the Signal Structure

Our results remain qualitatively similar if the signal structure is slightly altered by any of the following perturbations:

1. A small positive correlation between the signal that each agent obtains about his opponent's type.
2. A small chance that the informative signal is incorrect.

That is, if the perturbation is small enough, then there exists a unique proper evolutionarily stable configuration that is closed to (μ^*, b^*)).

6.2 Weaker Perfection Refinement: Undominated Configurations

One may argue that the "properness" refinement is too strong to capture the intuition that incumbents in stable conflagrations cannot play "badly" against mutant types. In this section we present a weaker perfection refinement, which only requires incumbents to play undominated strategies against mutant types, and we show that our uniqueness result holds also with respect to this notion.

The payoff of an incumbent's strategy that is played against a mutant type depends on that mutant's strategy. One may expect that most of the time invading mutants will best reply to the incumbents because either (see, Swinkels (1992)): (1) "best-reply" mutants have higher fitness than other mutants, and thus they are expected to survive longer in the post-entry population; and/or (2) mutants choose their strategy by experimentation, and they are more likely to choose best-reply strategies. Formally:

Definition 8. Let (μ, b) be a configuration and let $L_{\tilde{k}} \in \mathcal{L} \backslash C(\mu)$ be a mutant type. A signal-dependent strategy $\tilde{b}_{\tilde{k}} : L \cup \emptyset \to \Sigma_{\tilde{k}}$ is a *best reply* if $u\left(\left(L_{\tilde{k}}, \tilde{b}_{\tilde{k}}\right), (\mu, b)\right) \geq u\left(\left(L_{\tilde{k}}, b'_{\tilde{k}}\right), (\mu, b)\right)$ for each alternative signal-dependent strategy $b'_{\tilde{k}} : L \cup \emptyset \to \Sigma_{\tilde{k}}$.

An incumbent strategy $b_k\left(\tilde{k}\right)$ is dominated by another strategy $b'_k\left(\tilde{k}\right)$ if it yields a strictly worse payoff against all best-reply strategies of mutants of type $L_{\tilde{k}}$. Strategy $b_k\left(\tilde{k}\right)$ is undominated if it is not dominated by any other strategy. Formally:

Definition 9. Let (μ, b) be a configuration, let $L_k \in C(\mu)$ be an incumbent type, let $L_{\tilde{k}} \in \mathcal{L} \backslash C(\mu)$ be a mutant type, let $b'_k\left(\tilde{k}\right) \in \Sigma_k$ be a strategy, and for each $k' \neq \tilde{k}$ let $b'_k(k') = b_k(k')$. Strategy $b_k\left(\tilde{k}\right) \in \Sigma_k$ is *dominated by* $b'_k\left(\tilde{k}\right)$ if for each best-reply signal-dependent strategy $\tilde{b}_{\tilde{k}} : \mathcal{L} \cup \emptyset \to \Sigma_{\tilde{k}}$: $u\left(\left(L_k, b_k\right), \left(L_{\tilde{k}}, b_{\tilde{k}}\right)\right) < u\left(\left(L_k, b'_k\right), \left(L_{\tilde{k}}, b_{\tilde{k}}\right)\right)$. Strategy $b_k\left(\tilde{k}\right) \in \Sigma_k$ is *undominated* if it is not dominated by any strategy $b'_k\left(\tilde{k}\right) \in \Sigma_k$.

A configuration is undominated if it includes only undominated strategies. Formally:

Definition 10. Configuration (μ, b) is *undominated*, if for each incumbent type $L_k \in C(\mu)$ and for each "mutant" type $L_{\tilde{k}} \in \mathcal{L} \backslash C(\mu)$, the strategy $b_k\left(\tilde{k}\right)$ is undominated.

Proposition 6 in Appendix A.3 shows that any proper configuration is undominated.

A configuration is *undominated neutrally (evolutionarily) stable* if it is both undominated and neutrally (evolutionarily) stable. With very minor adaptations the proof of Theorem 2 applies also to the weaker refinement of undominated configurations. Thus we get the following result (proof is omitted):

Proposition 1. *Let* $A > 4.57$ *and* $\frac{A}{(A-1)^2} < p < \frac{A-1}{A}$. *Then for a sufficiently small* $\delta > 0$, *if* (μ, β) *is an undominated neutrally stable configuration, then it is equivalent to* (μ^*, b^*).

6.3 Different Stability Notions

6.3.1 Focal Stability

One may argue that it is more plausible that the state of the population only specifies the behavior of players against existing types, and the behavior against mutants that introduce new types should be determined by a post-entry adaptation process (see, e.g., Dekel, Ely, and Yilankaya (2007)). In what follows we formalize this idea, and present an alternative notion of focal stability (which may be of independent interest in future research), and state that all our results remain the same with this stability notion.

A *compact configuration* is a pair consisting of a distribution of types and the strategy that each type uses against other types in the support of the distribution. Formally:

Definition 11. *Compact configuration* (μ, b) is a pair where $\mu \in \Delta(\mathcal{L})$ is the distribution of types in the population, and $b = (b_k)_{L_k \in C(\mu)}$ is the profile of signal-dependent strategies that is played by each type against strangers and incumbents. That is, for each type $L_k \in C(\mu)$, $b_k : C(\mu) \cup \emptyset \to \Sigma_k$ is a signal-dependent *k-nice-retaliating* strategy.

Internal mutant configurations (μ', b') do not introduce new types to the population; that is, they satisfy $C(\mu') \subseteq C(\mu)$. Internal mutants are interpreted as the combination of small perturbations to the frequency of incumbent types, and experimentation of new strategies by small groups of incumbents. A compact configuration is internally neutrally (evolutionarily) stable if any sufficiently small group of (non-equivalent) internal mutants obtain a weakly (strictly) lower payoff than the incumbents in the post-entry population. Formally:

Definition 12. Compact configuration (μ, b) is *internally neutrally (evolutionarily) stable* if for any internal mutant configuration (μ', b') ($(\mu', b') \not\approx (\mu, b)$) with $C(\mu') \subseteq C(\mu)$ there exists some $\epsilon_{\sigma'} \in (0,1)$ such that for every $0 < \epsilon < \epsilon_{\sigma'}$:

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

(resp., $u\left((\mu, b), \epsilon(\mu', b') + (1-\epsilon)(\mu, b)\right) > u\left((\mu', b'), \epsilon(\mu', b') + (1-\epsilon)(\mu, b)\right)$).

External mutants introduce a new type to the population. We assume that the incumbents and the new mutant interactively adapt their joint behavior, while taking the "focal" behavior of incumbents against other incumbents and strangers as fixed. We further assume that this adaptation process is fast enough relative to the evolution of types, such that the behavior in the post-entry population converges to a Bayes–Nash equilibrium (given the "focal" fixed play of the incumbents). A compact configuration is *(strictly) externally focally stable* if any new mutant type achieves a (strictly) worse payoff in the induced post-entry Bayes–Nash equilibrium. Formally:

Definition 13. Given compact configuration (μ, b), $\epsilon > 0$ and mutant type $L_{k'} \in \mathcal{L} \backslash \mathcal{C}(\mu)$, let $B(\mu, b, L_{k'}, \epsilon)$ be the set of *post-entry focal configurations* (μ', b') that satisfy:

1. The post-entry distribution is a mixture of ϵ mutants and $1-\epsilon$ incumbents: $\mu' = (1-\epsilon)\cdot\mu + \epsilon \cdot L_{k'}$.

2. Focal play: the incumbents continue to play the same as in the pre-entry configuration against strangers and other incumbents: $b'_k(\emptyset) = b_k(\emptyset)$ for each $L_k \in C(\mu)$, and $b'_k\left(\tilde{k}\right) = b_k\left(\tilde{k}\right)$ for each $L_k, L_{\tilde{k}} \in C(\mu)$.

3. Each incumbent type best replies when observing the mutant type $L_{k'}$.

4. The mutant type $L_{k'}$ best replies to all opponents.

Definition 14. Compact configuration (μ, b) is *(strictly) externally focally stable* if there exists $\bar{\epsilon} > 0$ such that for every mutant type $L_{k'} \in \mathcal{L}\backslash\mathcal{C}(\mu)$ and $0 < \epsilon < \bar{\epsilon}$ the mutants obtain a (strictly) lower payoff in any post-entry focal configuration $(\mu', b') \in B(\mu, b, k', \epsilon)$:

$$u\left((\mu, b'), (\mu', b')\right) \geq u\left((L_{k'}, b'), (\mu', b')\right)$$

(resp., $u\left((\mu, b'), (\mu', b')\right) > u\left((L_{k'}, b'), (\mu', b')\right)$).

Finally, a compact configuration is (strictly) *focally stable* if it is both neutrally (evolutionarily) stable and (strictly) externally focally stable. Simple adaptations to the proofs in the appendix yield the same result with focal stability. Formally (proof is omitted):

Proposition 2. *Let $\frac{A}{(A-1)^2} < p < \frac{A-1}{A}$ and let $\delta > 0$ be sufficiently small. Then the compact configuration (μ^*, b^*) is strictly focally stable. Moreover, if (μ, b) is a focally stable configuration then (μ, b) and (μ^*, b^*) are equivalent.*

6.3.2 DEY-Stability (Dekel, Ely, and Yilankaya (2007))

In our definition of focal stability the incumbents only approximately best reply in the post-entry population, because they do not adjust their play against incumbents and strangers to the presence of the new ϵ mutants. In some evolutionary setups, the adaptation process according to which agents choose their strategies might be much faster than the evolutionary process according to which the frequency of the types evolves. In these setups, it may be plausible to assume that the post-entry population adjusts their play to an exact Bayes–Nash equilibrium after any entry of mutants (both external and internal mutants).

Dekel, Ely and Yilankaya's notion of stability makes this assumption.[15] A compact configuration (μ, b) is *(strictly) DEY-stable* if:

---

[15] 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 less appropriate to the study of the stability of heterogeneous populations.

1. The strategy profile b 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 "mutant" 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 achieve a (strictly) lower payoff than the incumbents.

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." 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 levels of sophistication (L_5 and above) do not exist. Formally:

Proposition 3. *Let* $\max\left(\frac{1}{A-2}, \frac{2\cdot A-1}{A\cdot(A-1)}\right) < p < \frac{A-1}{A}$ *and let* $\delta > 0$ *be sufficiently small. Then the compact configuration* (μ^*, b^*) *is strictly DEY-stable. Moreover, any other DEY-stable configuration* (μ, b) *satisfies:*

$$\sum_{k\leq 4} \mu(L_i) = 1, \text{ and } 0 < \mu(L_1) < 1.$$

In this setup we only have the weaker "qualitative uniqueness" because Lemmas 6-7, which are required for a stronger uniqueness, do not apply to DEY-stability. The lemmas do not work in this setup because:

- Part (1) of Lemma 7 (and similarly, part (1) of Lemma 6) does not work because small perturbations in the frequencies that different actions are played cannot be represented as mutations in the setup of DEY-stability. Specifically, the lemma relies on a perturbation that slightly increases the frequency of agents with type $L_{\tilde{k}}$ who play d_3 (instead of d_2). This perturbation can be represented as a mutant configuration in the definitions of evolutionary stability and focal stability, but not in the definition of DEY-stability. Note that immediately after such a perturbation, the incumbents who play d_3 obtain a strictly higher payoff than those who play d_2, but DEY-stability implicitly assumes that the adjustment process to a new Bayes–Nash equilibrium works in the opposite direction: some incumbents who played d_3 change their play into d_2.

- Part (4) of Lemma 7 does not work because Dekel, Ely, and Yilankaya (2007)'s definition only considers entry of mutants with a single type, and it turns out that for some values of p, the three-layer configuration with a support of $C(\mu) = \{L_1, L_2, L_4\}$ can only be invaded by a heterogeneous mutant configuration. Note that such a heterogeneous invasion may simply represent a slight perturbation in the frequency of the different types in the population.

Finally, we note that if one adapts Dekel, Ely and Yilankaya's definition by 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 stronger uniqueness that is implied by Lemma 7, hold.

7 Concluding Remarks

1. **Other heuristics for long horizons**: In our model we assumed that all players use nice and retaliating heuristics 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, e.g., the set of "memory-1" strategies, which depend only on the actions of 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 "memory-1" strategy 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 (Theorem 1) to this setup, and to show that a simple adaptation of (μ^*, b^*), in which all players choose a nice and retaliating heuristic, is stable. We conjecture that there are only two sets of stable configurations in this extended setup: (1) efficient configurations: type distribution and strategies are equivalent to (μ^*, b^*), all players use nice heuristics, and a large enough proportion of each existing type chooses a retaliating heuristic; and (2) inefficient configurations in which all players defect at all stages (and use an "always-defect" heuristic).

2. **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 the interaction. Assume that the interaction lasts at least M rounds ($Pr(\mathbf{T} \geq M) = 1$), and that the continuation probability at each stage

$(Pr\,(\mathbf{T} > n|\mathbf{T} > n-1))$ is not too far from 1. Bounded forward-lookingness is modeled in this setup as the stage in which a player becomes aware of the timing of the final period: a 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 remark, a simple adaptation of (μ^*, b^*) is stable in this setup, and we have an analogous conjecture about the set of stable configurations.

3. **Analogy-based expectation equilibrium:** Our model of bounded forward-looking types could also be formulated using Jehiel's (2005) *Analogy-Based Expectation Equilibrium* (ABEE). In this formulation a player of type L_k bundles all nodes with a horizon larger than k into a single analogy class (while fully differentiating between nodes with horizons of at most k), and expects his opponent to play the same in all nodes of this class. The requirement that players play an evolutionary refinement of a balanced Bayes–Nash equilibrium in a configuration is replaced with the requirement that players play an analogous evolutionary refinement of a balanced ABEE in a configuration: at each stage every player best replies to his analogy-based expectations, and expectations correctly represent the average behavior in every class. As in the previous remarks, a simple adaptation of (μ^*, b^*) is stable, and we have an analogous conjecture about the set of stable configurations.

4. **Other games**: The formal analysis deals only with the repeated Prisoner's Dilemma. However, we conjecture 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 games 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 drove the biological evolution of human characteristics.

A Proofs

A.1 Neutral Stability Implies Balanced Bayes–Nash Equilibrium

Proposition 4. *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 $u((L_k,\beta_k),(\mu,\beta)) > u((\mu,\beta),(\mu,\beta))$. 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 $u((L_k,\beta_k'),(\mu,\beta)) > u((L_k,\beta_k),(\mu,\beta))$. By the first part of the proposition, $u((L_k,\beta_k),(\mu,\beta)) = u((\mu,\beta),(\mu,\beta))$. 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).

□

A.2 Stability of (μ^*, b^*)

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

Proof. In order to prove that is neutrally stable, 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 (induces the same payoff to all types in* $C(\mu^*)$*).*

Proof. Let $q = \mu(L_1)$ be the frequency of the naive players. A naive player gets $(L-1)\cdot A+1$ against a naive opponent, and $(L-2)\cdot A+1$ against a sophisticated opponent (type L_3). A sophisticated player gets $(L-2)\cdot A+(A+1)+1 = (L-1)\cdot A+2$ against a naive opponent, and against a sophisticated opponent he gets: $(L-3)A+3$ if both players identify each other, $(L-3)\cdot A+(A+1)+2 = (L-2)\cdot A+3$ if only he identifies his opponent, $(L-3)\cdot A+0+2$ if only his opponent identifies him, and $(L-2)\cdot A+2$ 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:

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

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

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

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

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

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

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

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

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

Lemma 2. *Strategy profile b^* is a Bayes–Nash equilibrium given the distribution μ^*. Moreover: (1) players of type L_3 best reply in the unrestricted set of strategies Σ, and (2) any deviation that induces a different play on-equilibrium path yields a strictly worse payoff.*

Proof. It is immediate that naive players (type L_1) best reply in Σ_1, as their only choice is between cooperating and defecting at the last stage, and the latter strictly dominates the former. We have to show that a sophisticated player (L_3) plays a best reply in the unrestricted set Σ, and that it is strictly better than any deviation on-equilibrium path. It is immediate that d_2 is a best reply against an observed naive opponent, and strictly better than 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 (A.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 immediate that d_2 is also strictly better (against strangers) than any deviation on-equilibrium path. We are left with showing that it is strictly 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 deviation 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):

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

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

□

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, b) $((\mu, b) \not\approx (\mu^*, b^*))$, 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 naive agents becomes larger (smaller) relative to its proportion in μ^*, then the naive agents achieve a lower (higher) payoff than the sophisticated agents. This is because naive agents obtain a strictly lower payoff than sophisticated agents when facing naive opponents (the sophisticated 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 3 stages before the end. Mutants of higher types (L_4 or more) obtain at most the incumbents' payoff when facing incumbents, 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. □

A.3 Proper and Undominated Configurations

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

Proposition 5. *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 non-existing type $L_k \notin \{L_1, L_3\}$ has $\mu^n(L_2) = O(\epsilon_n)$, while $\mu^n(L_1) = \mu(L_1) - O(\epsilon_n)$ and $\mu^n(L_3) = \mu(L_3) - O(\epsilon_n)$. Type L_1 (L_2) plays with probability of $1 - O(\epsilon_n)$ strategy d_1 (d_2), and plays with probabilities of $O(\epsilon_n)$ or smaller magnitudes all other pure strategies in Σ_1 (Σ_2) in a way that is consistent with ϵ_n-properness. Any other type L_k plays with probability of $1 - O(\epsilon_n)$ strategy b_3^*, and plays with probabilities of $O(\epsilon_n)$ or smaller magnitudes all other pure strategies in Σ_k 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^*). □

Next we show that any proper configuration is undominated (as defined in Section 6.2).

Proposition 6. *Let* (μ, β) *be a proper configuration. Then it is undominated.*

Proof. Assume to the contrary that there exists incumbent type $L_k \in C(\mu)$ and mutant type $L_{\tilde{k}} \in \mathcal{L} \backslash \mathcal{C}(\mu)$ such that the strategy $b_k\left(\tilde{k}\right) \in \Sigma_k$ is dominated. That is, there exists alternative strategy $b'_k\left(\tilde{k}\right) \in \Sigma_k$ that yields a strictly better payoff against any best-reply signal-dependent strategy $\tilde{b}_{\tilde{k}} : \mathcal{L} \cup \emptyset \to \Sigma_{\tilde{k}}$. If $(\mu_\epsilon, \beta_\epsilon)$ is an interior ϵ-proper configuration sufficiently close to (μ, β) then ϵ-properness implies that type $L_{\tilde{k}}$ plays a best-reply signal-dependent strategy with probability of $1 - O(\epsilon)$. This, in turn, implies that type L_k plays against $L_{\tilde{k}}$ strategy $b'_k\left(\tilde{k}\right)$ with a $\frac{1}{\epsilon}$-times higher probability than $b_k\left(\tilde{k}\right)$, but then no sequence of ϵ_n-proper configurations can converge to (μ, β). □

A.4 Uniqueness of (μ^*, b^*)

In this section we prove the uniqueness of (μ^*, b^*) in the interval $\frac{A}{(A-1)^2} < p < \frac{A-1}{A}$ (Theorem **2**). In addition, we show that there are no stable configurations in the interval $\frac{1}{(A-1)\cdot(M-2)} < p < \frac{A}{(A-1)^2}$ (Proposition 7, which is part (2) of Theorem 3).

Theorem 2. *Let* $A > 4.57$, *let* $\frac{A}{(A-1)^2} < p < \frac{A-1}{A}$, *and let* $\delta > 0$ *be sufficiently small. If* (μ, β) *is a proper neutrally stable configuration, then it is equivalent to* (μ^*, b^*).

For $A < 4.57$, the theorem holds if p satisfies either:

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

Proposition 7. *Let $\frac{1}{(M-2)\cdot(A-1)} < p < \frac{A}{(A-1)^2}$, and let $\delta > 0$ be sufficiently small. Then no proper neutrally stable configurations exist.*

Both Theorem **2** and Proposition 7 follow 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 $M-2$. Formally:

Lemma 3. *Let (μ,β) be a configuration such that β is a Bayes–Nash equilibrium given μ and let $\delta > 0$ be sufficiently small. 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.*
2. *Any type $L_k \neq L_{k_1}$ in the population defects (with probability 1) at horizon k_1+1.*
3. *If $k_1 < M$ and $\mu(L_{k_1}) = 1$ then the configuration* is *not neutrally stable.*
4. *If $k_1 = M$ and $\frac{1}{(M-2)\cdot(A-1)} < p$ then the configuration* is *not proper neutrally stable.*
5. *If $k_1 = M-1$ then the configuration* is *not neutrally stable.*
6. *If $\frac{1}{(M-2)\cdot(A-1)} < p$ then: $\mu(L_{k_1}) < 1$ and $k_1 \le M-2$.*

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 $k_1 < M$, then ϵ mutants of type L_{k_1+1} who play d_{k_1+1} and enter the population outperform the incumbents.
4. In any proper configuration incumbents must cooperate in the first $M-2$ stages when facing an observed mutant type L_1 (as defecting at an early stage against L_1 is strictly dominated). Consider ϵ mutants of type L_1 who invade the population. With

probability $1-p$ the mutant's type is unobservable and he obtains 1 less point 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 outperform the incumbents if:

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

If the lowest type in the population is L_{M-1}, then fully rational agents (type L_M) outperform agents with type L_{M-1} and the configuration cannot be neutrally stable.

5. It is immediately implied by the previous parts.

□

Given a configuration with more than one existing type, we call the lowest existing 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 $\delta>0$ be sufficiently small, 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 $\mu(L_{k_1})<1$, $k_1\leq M-2$, 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. *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 player in the population has a type strictly larger than L_{k_1+2}.*
4. *The population only includes types $\{L_{k_1},L_{k_1+2}\}$.*
5. $$\mu(L_{k_1})=\frac{p(A-1)-1+\delta\cdot(c(L_{k_1+2})-c(L_{k_1}))}{p(A-1)}$$

   for any $p>\frac{A}{(A-1)^2}$, and no neutrally stable configuration exists if $p<\frac{A}{(A-1)^2}$.
6. *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_{k_1+2} 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. This is immediate from the previous two parts.

5. In any balanced configuration the naive and the non-naive incumbents must have the same payoff. By repeating the calculation of Lemma 1, this can hold only if

$\mu(L_{k_1}) = \frac{p(A-1)-1+\delta\cdot\left(c\left(L_{k_1+2}\right)-c\left(L_{k_1}\right)\right)}{p(A-1)} > 0$. By repeating the calculations of Lemma 2, this configuration cannot be stable if $p < \frac{A}{(A-1)^2}$.

6. If $L_{k_1} = L_1$ then the previous parts imply that (μ, b) and (μ^*, 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 the incumbents of type L_{k_1} (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 the sophisticated opponent to postpone his defection). Thus the mutants achieve a strictly higher payoff if:

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

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

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 stranger. First, Lemma 5 shows that the frequency of the naive players is small, and that these naive players must be of 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 $\delta > 0$ be sufficiently small, 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 $\mu(L_{k_1}) < 1$, $k_1 \leq M-2$, and that there are agents who defect at horizons strictly 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:*
   1. *(a) No incumbent defects at horizons* > 3 *when facing cooperative strangers.*
   2. *(b) No incumbent defects at horizons* > 4 *when facing any cooperative incumbent.*
   3. *(c) No incumbent has a type strictly higher than* L_4*.*

Proof.

1. The fact that there are incumbents who defect with with positive probability at horizons strictly larger than k_1+1 against strangers implies that early defection (at horizon a 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:

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

$$\mu(L_{k_1}) \leq \frac{1}{A}. \tag{A.2}$$

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 (A.2) yields:

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

Assume to the contrary that $\mu(L_2)=0$. The balance property 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 (A.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.

2.

(a) Let $\mu(L_{3+})=1-\mu(L_1)-\mu(L_2)$. The balance property 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:

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

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

This implies, together with (A.2):

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

From the same argument as in part (1) of this lemma, if $\mu(L_2) > \frac{1}{A}$ 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.

□

All the lemmas presented so far are also valid for the alternative notion of DEY-stability (see Section 6.3). The following two lemmas do not hold for DEY-stability, as they rely on the assumption that the incumbents "stick" to their pre-entry play when facing incumbents and strangers. Using this assumption, we show in Lemma 6 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* $\delta > 0$ *be sufficiently small, 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* $\mu(L_{k_1}) < 1$ *and* $k_1 \leq M - 2$*. 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* $> k_1 + 2$ *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 $(L-1)\cdot A+1$ points when playing against itself. A random player with a type different than L_{k_1} who plays against L_1 gets at most $(L-1)\cdot A+1+1$ when he observes his opponent's type, and an expected payoff of at most $\eta\cdot((L-1)\,A+2)+(1-\eta)\cdot((L-2)\cdot A+3)$. 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 $(L-2)\cdot A+1$ from each payoff):

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

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

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

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

If (A.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:

$$\eta\cdot(1-p)\cdot(A-1)\leq1-\eta\cdot(1-p)\;\Leftrightarrow\;\eta\cdot(1-p)\cdot A\leq1\,(1-p)\;\Leftrightarrow\;\eta\leq\frac{1}{(1-p)\cdot A}.$$

Substituting (A.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*

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

let $\delta>0$ be sufficiently small, 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 $k_1=1$.*

Note that if $A>4.57$ 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 $\frac{1}{(A-1)\cdot(M-2)}<p<\frac{A-1}{A}$.

Finally, Lemma 7 shows that the configurations characterized by Corollary 1 cannot be proper evolutionarily stable. 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 is not valid for the notion of DEY-stability

(Section 6.3) due to: (1) the assumption that the incumbents fix the pre-entry play against incumbents and strangers, and (2) the assumption that "internal" mutant configurations (of existing types) can be heterogeneous.

Lemma 7. *Let $0 < p < \frac{A-1}{A}$, let $\delta > 0$ be sufficiently small, let (μ, β) be a configuration satisfying: (1) $L_1 \in C(\mu)$, (2) $\mu(L_1) < 1$, (3) $\mu(L_k) = 0 \; \forall k > 4$, and (4) 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'): (1) $\mu' = \mu$, (2) for each $k \neq \tilde{k}$, $b'_k = b_k$, (3) for each $L_k \in \mathcal{L}$, $b'_{\tilde{k}}(k) = b_{\tilde{k}}(k)$, and (4) $b'_{\tilde{k}}(\emptyset) = d_3$. That is, the mutants have the same distribution of types as the incumbents, and they play the same except that mutants of type $L_{\tilde{k}}$ always play d_3 when facing strangers. Observe that such mutants 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: type L_1 always plays d_1; type L_2 always plays d_2; type L_3 plays d_2 against observed L_1 and plays d_3 in all other cases; and type L_4 plays d_2 against observed L_1, d_3 against strangers and observed L_2, and plays 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:

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

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

Let $\mu_k = \mu\left(L_k\right)$. 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:

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

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 (A.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.

□

A.5 Stable Configurations Near 0 and 1

Theorem 3. (parts 1 and 4)

1. *Let* $0 \leq p < \frac{1}{(M-2)\cdot(A-1)}$. *Then there exists a proper neutrally stable configuration* $\left(\tilde{\mu}, \tilde{b}\right)$ *where all players have type* L_M *and they play* d_M *against strangers and type* L_M*, and* d_{k+1} *against observed "mutant" type* $L_k < L_M$*. Moreover, any other neutrally stable configuration is equivalent to* $\left(\tilde{\mu}, \tilde{b}\right)$.

2. *(part (4) in the main text) Let* $\frac{A-1}{A} < p \leq 1$. *Then in any proper neutrally 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 players best reply to each other. Consider ϵ mutants with type $k < L$ who invade the population. When facing incumbents, the mutants obtain 1 less fitness point when their type is unobserved, and $(A-1)\cdot(M-k-1)-1$ more fitness points when their type is observed. Thus for sufficiently small ϵ and δ, the incumbents achieve a strictly lower payoff if:
$$(1-p) > p\cdot((A-1)\cdot(M-k-1)-1) \Leftrightarrow 1 > p\cdot(A-1)\cdot(M-k-1)$$
$$p < \frac{1}{(A-1)\cdot(M-k-1)}.$$
This implies that for any $p < \frac{1}{(M-2)\cdot(A-1)}$, the configuration is proper neutrally stable. The uniqueness is immediately implied by Lemmas 3-7, which show that no configuration can be proper neutrally stable for $p < \frac{A}{(A-1)^2}$.

2. Let L_k be the highest type in the population. Let l be the largest horizon in which L_k begins 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 their L_k-s play d_l with probability 1 would achieve a strictly higher payoff in the post-entry population). 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 $l = k < M$, then ϵ mutants of type L_{k+1} 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.

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*, 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.

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(2), 187–218.

Costa-Gomes, M., and V. Crawford (2006): "Cognition and behavior in two-person guessing games: An experimental study," *American Economic Review*, 96(5), 1737–1768.

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.

Crawford, V. (2008): "Look-ups as the windows of the strategic soul: Studying cognition via information search in game experiments," *in Perspectives on the Future of Economics: Positive and Normative Foundations*, A. Caplin and A. Schotter, Oxford University Press.

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.

Fershtman, C., K. L. Judd, and E. Kalai (1991): "Observable contracts: Strategic delegation and cooperation," *International Economic Review*, 32(3), 551–559.

Frenkel, S., Y. Heller, and R. Teper (2012): "Endowment as a blessing," *mimeo.*

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

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.

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," *mimeo.*

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. (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.

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.

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. (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 chainstore 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.

SWINKELS, J. (1992): “Evolutionary stability with equilibrium entrants,” *Journal of Economic Theory*, 57(2), 306–332.

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.*