Learning Across Games*

Friederike Mengel†
University of Alicante

November 2006

Abstract

In this paper (reinforcement) learning of decision makers that face many different games is studied. As learning separately for all games can be too costly (require too much reasoning resources) agents are assumed to partition the set of all games into analogy classes. Partitions of higher cardinality are more costly. A process of simultaneous learning of actions and partitions is presented and equilibrium partitions and action choices characterized. The model is able to explain deviations from subgame perfection that are sometimes observed in experiments even for vanishingly small reasoning costs. Furthermore it is shown that learning across games can stabilize mixed equilibria in 2×2 Coordination and Anti-Coordination games and destabilize strict Nash equilibria under certain conditions.

JEL Classification: C70, C72, C73.

Keywords: Game Theory, Bounded Rationality, Reinforcement Learning, Analogies.

Very Preliminary and Incomplete !!!

*This paper has benefitted enormously from discussions with my supervisor Fernando Vega Redondo. I also wish to thank Larry Blume, David Easley, Ani Guerdjikova as well as seminar participants in Alicante, Cornell and Leuven for helpful comments.

† *Departamento de Fundamentos del Análisis Económico*, University of Alicante, Campus San Vicente del Raspeig, 03071 Alicante, Spain. *e-mail:* friederike@merlin.fae.ua.es

1 Introduction

Economic agents are involved in many games. Some of which can be quite distinct but many will share a basic structure (e.g. have the same set of actions) or be similar along other dimensions. A priori games can be similar with respect to the payoffs at stake, the frequency with which they occur, the context of the game (work, leisure, time of day/year...), the people one interacts with (friends, family, colleagues, strangers...), the nature of strategic interaction, or the social norms and conventions involved.[1] Distinguishing all games and learning separately for all of them requires a huge amount of alertness or reasoning costs. Consequently it is natural to assume that agents will partition the set of all games into analogy classes, i.e. sets of games they see as analogous.

In this paper we study (reinforcement) learning across games, i.e. decision makers that face many different games and *simultaneously* learn which actions to choose and how to partition the set of all games. Our approach does not presume an exogeneous measure of similarity nor do we make any assumption about what agents will perceive as analogous. Instead we focus on a much more instrumental view of decision-making and ask the question which games do agents *learn* to discriminate.

To fix ideas think about Rubinstein bargaining games that only differ in the discount factor, i.e. in the rate at which the pie shrinks in each of the games. If agents are involved in many such games it is natural to think that they will transfer their experience from some of the games to learn optimal actions in others. They might also learn though that experience in some games is a bad indicator for behavior in others and that transferring this experience will lead to bad decisions. In other words agents can learn to distinguish such games. In particular we consider the example of two players who interact repeatedly to play two bargaining games. One of the games has discount factor zero (ultimatum game) and the other one a strictly positive discount factor. We will show that even for vanishingly small reasoning costs there is an equilibrium in which players see the two games as analogous and where equilibrium actions always give a positive share of the pie to the responder. Learning across games in this example leads to predictions that fundamentally differ from those of learning in a single game. It constitutes a possible explanation for deviations from subgame perfection sometimes observed in experiments. This conclusion is our starting point to study the implications of partition learning for equilibrium selection in two-player games. Then, for vanishingly small reasoning costs, we establish the following results.

- Learning across games leads to approximate Nash equilibrium play in all games.[2]
- Nash equilibria in weakly dominated strategies that are unstable to learning in a single game can be stabilized by learning across games.

[1] Obviously social norms and conventions will typically arise endogenously.

[2] We say "approximately" Nash equilibrium because we consider a process of perturbed reinforcement learning.

- Strict Nash equilibria that are always stable to learning in a single game can be destabilized by learning across games.

- Mixed Nash equilibria in 2×2 coordination and anti-coordination games that are unstable to learning in a single game can be stabilized with learning across games.

Furthermore we try to characterize equilibrium partitions and find that if and only if the supports of the sets of Nash equilibria of any two games are disjoint, agents will distinguish these games in equilibrium. The cardinality of the action set thus provides an upper bound on the cardinality of equilibrium partitions.

The paper is organized as follows: In Section 2 the model is presented. In Section 3 we use stochastic approximation techniques to approximate the reinforcement learning process through a system of deterministic differential equations. In Section 4 we characterize equilibrium actions and show how learning across games leads to interesting new predictions. In Section 5 we characterize equilibrium partitions. In Section 6 we discuss related literature. Section 7 concludes. The proofs are relegated to an appendix.

2 The Model

Games and Partitions

There are 2 players indexed $i = 1, 2$ playing repeatedly a game randomly drawn from the set $\Gamma = \{\gamma_{1,}....\gamma_J\}$ according to probability measure $f_j > 0, \forall \gamma_j \in \Gamma$.[3,4] Denote by $\mathcal{P}(\Gamma)$ the power set (or set of subsets) of Γ and $\mathcal{P}^+(\Gamma)$ the set $\mathcal{P}(\Gamma) - \varnothing$. For both players $i = 1, 2$ all $\gamma \in \Gamma$ share the same action set A_i. Players partition the set of all games into subsets of games they learn not to discriminate or in other words to see as analogous. Denote G a *partition* of Γ with $\text{card}(G) = Z$. An element g of G is called an *analogy class.* For a given set of games Γ with cardinality J the number of possible analogy classes is thus $2^J - 1 = \text{card}(\mathcal{P}^+(\Gamma))$. The set of all possible partitions of Γ is given by $\mathcal{G}$ with $\text{card}(\mathcal{G}) = L$.

Reasoning Costs

There is a cost $\Xi(Z, \xi)$ of holding partitions reflecting the agents' limited reasoning resources. $\Xi(Z, \xi)$ is an increasing function, as partitions of higher cardinality are more costly. We make the following assumptions on the reasoning cost function.

(i) $Z_l \gtreqless Z_h \Leftrightarrow \Xi(Z_l, \xi) \gtreqless \Xi(Z_h, \xi)$ (strictly increasing costs).
(ii) $\forall\ Z, \xi > 0$: $0 < \Xi(Z, \xi) < \xi$.
(iii) $\xi < \left|\min \pi^i(a^t, \gamma^t)\right|$.

[3]The model is readily extended to n -player games at the cost of additional notational complexity. Many of our results will also hold for the n -player case.

[4]In the following I will - with some abuse of notation - denote both the random variable and its realization by γ.

Reasoning costs are strictly increasing and unimportant relative to the smallest possible payoff $\min \pi^i(a^t, \gamma^t)$ from any of the games. The case of small costs is the most interesting one. With high reasoning costs new predictions arise trivially.[5]

Notation

Before we proceed to describe the learning process let us point out the following notation.

1) Games are denoted $\gamma_j \in \Gamma = \{\gamma_1, ..., \gamma_J\}$.

2a) Actions for player 1 are denoted $a_m^1 \in A_1 = \{a_1^1, ..., a_{M1}^1\}$.

2b) Actions for player 2 are denoted $a_m^2 \in A_2 = \left\{a_1^2, ..., a_{M2}^2\right\}$.

3) Analogy classes are denoted $g_k \in \mathcal{P}^+(\Gamma) = \{g_1, ..., g_K\}$.

4) Partitions are denoted $G_l \in \mathcal{G} = \{G_1, ..., G_L\}$.

Throughout the paper the generic index h will be used whenever we want to distinguish between any game, action, analogy class or partition and a particular one.

Learning

Players learn simultaneously about partitions and actions. The model of learning employed is one of reinforcement (or sometimes called stimulus response) learning based on Roth and Erev (1995).[6] In these kind of models partitions and actions that have led to good outcomes in the past are more likely to be used in the future. More precisely players are endowed with *propensities* α_l^i to use partitions G_l and with *attractions* β_{mk}^i towards using each of their possible actions $a_m^i \in A_i$. Unlike in standard reinforcement learning where attractions are defined for a given game, in learning across games attractions depend on the analogy class $g_k \in \mathcal{P}^+(\Gamma)$. If there are M actions an agent thus holds $M(2^J - 1)$ attractions. Not all of them will be in use at all times. The number of attractions an agent uses is given by MZ_l where Z_l is the cardinality of the partition he holds. For L possible partitions an agent has L propensities. Players will choose partitions with probabilities q^i proportional to propensities and actions with probabilities p^i proportional to attractions according to the choice rules specified below.

Payoffs $\pi^i(a^t, \gamma^t)$ for player i at any time t depend on the game that is played γ^t and the actions chosen by both players a^t. Payoffs are normalized to be strictly positive and finite.[7] After playing a game players will update their propensities and attractions taking into account the payoff obtained.

State

At any point in time a player is thus completely characterized by his attractions and propensities: $(\alpha^{it}, \beta^{it})$ where $\alpha^{it} = (\alpha_l^{it})_{G_l \in \mathcal{G}}$ are the propensities for partitions and $\beta_i^t = ((\beta_{mk}^i)_{a_m \in A})_{g_k \in \mathcal{P}^+(\Gamma)}$ are the attractions for actions $a_m \in A$. The *state of player* i at time t is then $(\alpha^{it}, \beta^{it})$. The *state of the population* at the end of time t is given by the collection of the player's states $(\alpha^{1t}, \beta^{1t}, \alpha^{2t}, \beta^{2t})$.

[5] At the end of Section 5 this assumption will be discussed somewhat more.

[6] See also Erev and Roth (1998).

[7] This is a technical assumption commonly used in reinforcement models. (See among others Börgers and Sarin (1997)).

The Dynamic Process

The dynamic process unfolds as follows:

(i) First players choose a partition G_l with probability

$$q_l^t = \frac{\alpha_l^t}{\sum_{G_h \in \mathcal{G}} \alpha_h^t}. \tag{1}$$

Denote G^{it} the partition actually chosen by player i at time t.

(ii) A game γ_j^t is drawn from Γ according to $\{f_j\}_{j\in\Gamma}$ and classified into g_k^{it} according to G^{it}

(iii) Players choose action a_m with probability

$$p_{mk}^t = \frac{\beta_{mk}^t}{\sum_{a_h \in A} \beta_{hk}^t}. \tag{2}$$

Let a^{it} be the action actually chosen by player i at time t.

(iv) Players observe the record of play $w^{it} = \{G^{it}, g^{it}, a^{it}, \pi^i(a^t, \gamma^t)\}$.

(v) Players update attractions according to the following rule:

$$\beta_{mk}^{i(t+1)} = \begin{cases} \beta_{mk}^{it} + \pi^i(a^t, \gamma^t) + \varepsilon_0 & \text{if } g_k^i, a_m^i \in w^{it} \\ \beta_{mk}^{it} + \varepsilon_0 & \text{otherwise} \end{cases}. \tag{3}$$

The attraction corresponding to the action and analogy class just used is reinforced with the payoffs obtained $\pi^i(a^t, \gamma^t)$. In addition every attraction is reinforced by a small amount $\varepsilon_0 > 0$.[8] In the analogy class just used ε_0 is best interpreted as noise or experimentation. As ε_0 has a bigger effect on smaller β' s, it increases the probability that "suboptimal" actions are chosen. In analogy classes not used, it can be seen as reflecting forgetting. The intuition is simple. Think of an analogy class that is never used. Because of ε_0 all actions will eventually have the same attractions and will be used with the same probability in such an analogy class. ε_0 in this case leads to a reversal to the uniform distribution for action choices. That is why we say it reflects forgetting.[9]

(vi) Players update propensities as follows:

$$\alpha_l^{i(t+1)} = \begin{cases} \alpha_l^{it} + \left(\pi^i(a^t, \gamma^t) - \Xi(Z_l)\right) + \varepsilon_1 & \text{if } G_l \in w^{it} \\ \alpha_l^{it} + \varepsilon_1 & \text{if } G \notin w^{it} \end{cases} \tag{4}$$

where again $\varepsilon_1 > 0$ is noise. The payoffs relevant for partition updating are payoffs net of costs of holding partitions.[10]

[8] There are many alternative ways to model noise. One could see ε_0 as the exspected value of a random variable or allow noise to depend on choice frequencies without changing the results qualitatively. See Fudenberg and Levine (1998), Hopkins (2002) or Hofbauer and Hopkins (2005).

[9] Of course one could want to have ε differ in the two cases $g_k \in w_i^t$ and $g_k \notin w_i^t$. As the main interest of the paper is not to study perturbed learning we chose to stick to a simple formulation.

[10] Note that the algorithm is always well defined as $\left(\pi^i(a^t, \gamma^t) - \Xi(K_l)\right) > 0$ given our assumptions on the cost function. To allow for higher costs one could replace $\pi^i(a^t, \gamma^t) - \Xi(K_l)$ by $\max\{\left(\pi^i(a^t, \gamma^t) - \Xi(K_l)\right), 0\}$ in (4).

Note that agents need only very little information in this model. In particular they do not need to know the structure of the games that are played, nor do they need to make any distinction at all between the games that are seen as analogous, nor calculate best response or even know that games are played at all.

Flat Learning Curves and Step Size

Another characteristic property of this version of reinforcement learning is that learning curves get flatter over time. Note that the denominators of (1) and (2) ($\sum_{G_l \in \mathcal{G}} \alpha_l^{it} =: \alpha^{it}$ and $\sum_{a_m \in A_i} \beta_{mk}^{it} =: \beta_k^{it}$) are increasing with time. A payoff thus has a larger effect on action and partition choice probabilities in early periods. Unexperienced agents will learn faster than agents that have accumulated a lot of experience. Note also that the impact of noise or experimentation decreases over time. The step sizes of the process are given by $1/\beta_k^{it}$ and $1/\alpha^{it}$. The property of flat learning curves or decreasing step sizes is sometimes called "power law of practice" in psychology. It greatly simplifies the study of the asymptotic behavior of the process as we will see in the next section.

3 Asymptotic Behavior of the Process

Denote $x^{it} = (p^{it}, q^{it})$ the choice probabilities for actions and partitions of player i where $p^{it} = ((p_{mk}^{it})_{a_m \in A})_{g_k \in \mathcal{P}^+(\Gamma)}$ and $q^{it} = (q_l^{it})_{G_l \in \mathcal{G}}$. Let $x^t = (x^{1t}, x^{2t}) \in \mathbf{X}$ be the collection of these vectors. The main interest lies in the evolution of x^t. $\mathbf{X}$ is the space in which these choice probabilities evolve. It has dimension $(2^J - 1)(M_1 + M_2 - 2) + 2(L - 1)$, where $M_1 = \text{card}\, A_1$ and $M_2 = \text{card}\, A_2$.[11]

3.1 Mean Dynamics

Because of the high dimensionality of the system, the mean dynamics will be quite complicated expressions. We will first derive the mean motion in the general case and then illustrate the behavior of the algorithm in the special case of two 2×2 games.

Action Choice and Observed Play

First note that there is a difference between action choices actually made by the players and observed play in each game.

- *Action choice* is described by the probabilities $p_k^{it} = (p_{1k}^{it}, ..., p_{Mk}^{it})$ as defined in (2). These probabilities depend on the analogy class g_k of player i and are defined over the set of analogy classes $\mathcal{P}^+(\Gamma)$. They characterize a player's choice.

- *Observed play* in any game γ_j is described by the "phenotypic" probabilities $\sigma_j^{it} = (\sigma_{1j}^{it}, ..., \sigma_{Mj}^{it})$ defined over the set of games Γ. The σ_j^{it} do not

[11] There are J games with $2^J - 1$ non-empty subsets or possible analogy classes. Action choice probabilities are defined for each of the $M_1 + M_2$ actions of the two players depending on the analogy classes. There are L possible partitions of the set Γ for each of the players. Furthermore both action choice probabilities within each analogy class and partition choice probabilities have to sum to one.

characterize an agents choice but how an agent actually behaves in a given game.

In other words the observed play probability σ_{mj}^{it} captures the overall probability (across partitions) with which action m is chosen when the game is γ_j. It is generated from action and partition choice probabilities as follows: $\sigma_{mj}^{it} := \sum_{G_l \in \mathcal{G}} q_l^{it} \sum_{g_k \in G_l} p_{km}^{it} \mathrm{I}_{jk}$ where I_{jk} takes the value 1 if $\gamma_j \in g_k$ and zero otherwise.

Mean Motion

It is intuitively clear that the mean motion of action choice frequency p_{mk}^{it} will depend on how much action a_m is reinforced in analogy class g_k compared to other actions. Denote $\Pi_{mk}^{it}(x^t)$ the expected payoff of action m conditional on visiting analogy class g_k.[12] And let $S_{mk}^{it}(x^t)$ be the difference between the expected payoffs of action a_m and all actions on average at x^t conditional on visiting analogy class g_k.

$$S_{mk}^{it}(x^t) = \Pi_{mk}^{it}(x^t) - \sum_{a_h \in A_i} p_{hk}^{it} \Pi_{hk}^{it}(x^t). \tag{5}$$

The mean motion of action choice probabilities will of course also depend on how often the process visits analogy class g_k. Let $r_k^{it} := \sum_{G_l \in \mathcal{G}, \gamma_j \in \Gamma} q_l^{it} f_j \mathrm{I}_{kl} \mathrm{I}_{jk}$ - where $\mathrm{I}_{kl} = 1$ if $g_k \in G_l$ and zero otherwise - be the total frequency with which analogy class g_k is used. $\sum_{G_l \in \mathcal{G}} q_l^{it} \mathrm{I}_{kl}$ is the probability that a partition containing g_k is used and $\sum_{\gamma_j \in \Gamma} f_j \mathrm{I}_{jk}$ the (independent) probability that a game contained in g_k is played. We can state the following Lemma.

Lemma 1 *The mean change in action choice probabilities p_{mk}^{it} of player i is given by*

$$\left\langle p_{mk}^{i(t+1)} - p_{mk}^{it} \right\rangle = \frac{1}{\beta_k^{it}} [p_{mk}^{it} r_k^{it} S_{mk}^{it}(x^t) + \varepsilon_0 (1 - M p_{mk}^{it})] + O\left((\frac{1}{\beta_k^{it}})^2 \right). \tag{6}$$

Proof. Appendix A. ■

The mean change in action choice probabilities in analogy class g_k is determined by the payoff in g_k of the action in question (a_m) relative to the average payoff of all actions $(S_{mk}^{it}(x^t))$ scaled by current choice probabilities $p_{mk}^{it} r_k^{it}$. Similar laws of motion are characteristic of many reinforcement models. The second term in brackets is a noise term. Noise tends to drive action choice probabilities towards the interior of the phase space. The step sizes $\frac{1}{\beta_k^{it}}$ determine the speed of learning.

Partition choice probabilities are similarly determined by the relative payoff obtained when using the different partitions $S_l^{it}(x^t) = \Pi_l^{it}(x^t) - \sum_{G_h \in \mathcal{G}} q_h^{it} \Pi_h^{it}(x^t)$ where $\Pi_l^{it}(x^t)$ is the expected payoff net of reasoning costs obtained when using partition G_l. We can state Lemma 2:

[12] To write down $\Pi_{mk}^{it}(x^t)$ explicitly yields complicated expressions, which are stated in Appendix A.

Lemma 2 *The mean change in partition choice probabilities q_l^{it} of player i is given by*

$$\left\langle q_l^{i(t+1)} - q_l^{it} \right\rangle = \frac{1}{\alpha^{it}}[q_l^{it} S_l^{it}(x^t) + \varepsilon_1(1 - Lq_l^{it})] + O\left((\frac{1}{\alpha^{it}})^2 \right). \quad (7)$$

Proof. Appendix A. ■

Let us illustrate the expressions for mean motion (6)-(7) for the case of two 2×2 games.

Example 1 (Two 2×2 Games)

Let Γ consist of two games γ_1 and γ_2 drawn from the class of symmetric 2×2 games. Denote $A = \{H, L\}$ the action set and let the payoff matrices of the two games are given by

$$\gamma_1 = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix}, \gamma_2 = \begin{pmatrix} b_1 & b_2 \\ b_3 & b_4 \end{pmatrix} \quad (8)$$

where the first line gives the row player's payoffs associated with action H and the second line those associated with action L. There are two partitions $G_C = \{\gamma_1, \gamma_2\}$ and $G_F = \{\{\gamma_1\}, \{\gamma_2\}\}$ and three analogy classes $g_1 = \{\gamma_1\}, g_2 = \{\gamma_2\}$ and $g_3 = \{\gamma_1, \gamma_2\}$. The expected payoff of actions H and L conditional on visiting analogy class g_1 is given by $\Pi_{H1}^{it}(x^t) = \sigma_{H1}^{(-i)t} a_1 + (1 - \sigma_{H1}^{(-i)t}) a_2$ and $\Pi_{L1}^{it}(x^t) = \sigma_{L1}^{(-i)t} a_3 + (1 - \sigma_{L1}^{(-i)t}) a_4$. Note that $S_{H1}^{it}(x^t) = (1 - p_{H1}^{it}) \left[\Pi_{H1}^{it}(x^t) - \Pi_{L1}^{it}(x^t) \right]$. Substituting into (6) yields the equations for mean motion:

$$\begin{aligned} & \left\langle p_{H1}^{i(t+1)} - p_{H1}^{it} \right\rangle \quad (9) \\ = \; & \frac{1}{\beta_1^{it}} \left[r_1^{it} p_{H1}^{it} (1 - p_{H1}^{it}) \left[\Pi_{H1}^{it}(x^t) - \Pi_{L1}^{it}(x^t) \right] + \varepsilon_0 (1 - 2p_{H1}^{it}) \right] + O(\cdot) \end{aligned}$$

where $r_1^{it} = (1 - q_1^i) f_1$ is given by the product of the choice frequency of the fine partition and the frequency of the first game occurring. For analogy classes g_2 and g_3 this can be determined analogously. How does the analogous expression for partitions look like ? The expected net payoff from using the coarse partition G_C is given by $\Pi_C^{it}(x^t) = p_{H3}^{it} \Pi_{H3}^{it}(x^t) + (1 - p_{H3}^{it}) \Pi_{L3}^{it}(x^t) - \Xi(1)$. The expected net payoff of using the fine partition is given by $\Pi_F^{it}(x^t) = \sum_{k=1,2} p_{Hk}^{it} \Pi_{Hk}^{it}(x^t) + (1 - p_{Hk}^{it}) \Pi_{Lk}^{it}(x^t) - \Xi(2)$. Relative reinforcement for the coarse partition is given by $S_C^{it}(x^t) = (1 - q_1)[\Pi_C^{it}(x^t) - \Pi_F^{it}(x^t)]$ and the equation for mean motion can be obtained by substituting into (7).

3.2 Stochastic Approximation

Stochastic Approximation is a way of analyzing stochastic processes by exploring the behavior of associated deterministic systems. A stochastic algorithm like the one described in (1)-(4) can under certain conditions be approximated through a system of deterministic differential equations.[13] One of the conditions that make such an approach particularly suitable is the property of decreasing step sizes ($\sum_{t=1}^{\infty}\left(\frac{1}{\alpha^{it}}\right)^2 < \infty$ and $\sum_{t=1}^{\infty}\left(\frac{1}{\beta_k^{it}}\right)^2 < \infty, \forall i = 1, 2, \forall g_k \in \mathcal{P}^+(\Gamma)$). As this property is satisfied by reinforcement models obeying the "power law of practice", stochastic approximation is a convenient and often employed way of analyzing reinforcement models. There is one small complication though. While the vectors $x^{it} = (p^{it}, q^{it})$ are allowed to take values in $\mathbb{R}^d$ the step size is typically taken to be a scalar in standard models. Note though that here there are 2^{J+1} different step-sizes that are endogenously determined.[14] One possibility to deal with this problem is to introduce additional parameters that take account of the relative speed of learning. To reduce notational complexity though we focus on a simpler way of dealing with this problem that consists in normalizing the process. [15]

Normalization Assume that at each point in time $t-1$, $\forall i = 1, 2$ after attractions and propensities are updated according to (3) and (4), every attraction and propensity is multiplied by a factor such that $\alpha^{i(t)} = \mu + t\theta$ and $\beta_k^{it} = \mu + t\theta$ for some constant θ where $\mu = \alpha^0 = \beta_k^0$ (the sum of initial propensities and attractions) - but leaving $x^t = (p^t, q^t)$ unchanged.[16] Then there is a unique step size of order t^{-1}. Call the resulting process the *normalized* process.

We can state the following Proposition.

Proposition 1 *The normalized stochastic learning process can be characterized by the following system of ODE's:*

$$\dot{p}_{mk}^{i} = p_{mk}^{i} r_k^i S_{mk}^i(x) + \varepsilon_0(1 - M p_{mk}^i) \tag{10}$$

$$\dot{q}_l^{i} = q_l^i S_l^i(x) + \varepsilon_1(1 - L q_l^i). \tag{11}$$

$\forall a_m \in A_i, g_k \in \mathcal{P}^+(\Gamma), G_l \in \mathcal{G}, i = 1, 2.$

[13] See the textbooks of Kuschner and Lin (2003) or Benveniste, Metevier and Priouret (1990) on stochastic approximation theory. The relevant conditions are listed in Appendix A.

[14] For each of the two players there are $(2^J - 1)$ step sizes corresponding to attractions for actions in each of the analogy classes and 1 step size corresponding to propensities for partitions.

[15] See Hopkins (2002), Laslier, Topol and Walliser (2001) or an earlier version of this paper for approaches not based on normalization. Introducing additional parameters has the advantage that the relative speed of learning can be kept track of explicitly, but also complicates notation a lot. As none of our results hinges on the speeds of learning we decided for this simpler formulation. See Ianni (2002), Börgers and Sarin (2000) or Posch (1997) for approaches based on normalization.

[16] The factor needed is given by $(\mu + t\theta)/(\alpha^{i(t-1)} + \pi^{i(t-1)} + L\varepsilon_1)$ for all α_l^i and $(\mu + t\theta)/(\beta_k^{i(t-1)} + \pi^{i(t-1)} + M\varepsilon_0)$ for all β_{mk}^i. If one thinks of the process as an urn model, μ is the initial number of balls in each urn.

Proof. Appendix A. ■

The evolution of the choice probabilities $x^{it} = (p^{it}, q^{it})$ is closely related to the behavior of the deterministic system (10)-(11).[17] More precisely let us denote the vector field associated with the system (10)-(11) by $F(x(t))$ and the solution trajectory of $\dot{x} = F(x(t))$ by $x(t)$. Then with probability increasingly close to 1 as $t \to \infty$ the process $\{x^t\}_t$ follows a solution trajectory $x(t)$ of the system $F(x(t))$.[18] Furthermore if x^* is an unstable restpoint or not a restpoint of $F(x(t))$, then $\Pr\{\lim_{t\to\infty} x^t = x^*\} = 0$. If x^* is an asymptotically stable equilibrium point of $F(x(t))$, then $\Pr\{\lim_{t\to\infty} x^t = x^*\} > 0$.[19] In the following analysis we will thus focus on the asymptotically stable equilibria of (10)-(11).

4 Equilibrium Actions

Before starting the analysis we make the following assumption on noise:

(i) $\varepsilon_0 \to 0$ and (ii) $0 < \varepsilon_0/\varepsilon_1 < \infty$.

Noise is assumed to be vanishingly small and of the same order for both action and partition choices. (ii) ensures that there are no partitions whose choice probabilities converge faster to zero than noise ε_0. If this were the case in analogy classes that only occur with these partitions noise would dominate and a very wide range of outcomes would be trivially sustainable.

The first result we would now like to present establishes a close relation between the asymptotically stable equilibria $x^* = (p^*, q^*)$ of $F(x(t))$ and the set of Nash equilibria $E^{Nash}(\gamma)$ in any game γ. Denote $E(\varepsilon_0)$ the set of asymptotically stable equilibria of the system and the limit set $\lim_{\epsilon_0\to 0} E(\varepsilon_0) =: E^*$. The following proposition can be stated.

Proposition 2 *If $\xi \to 0$ any asymptotically stable equilibrium $x^* \in E^*$ must induce phenotypical behavior that is approximately Nash in every game $\gamma_j \in \Gamma$,* i.e. $\lim_{\varepsilon_0\to 0} (\sigma_j^1(\varepsilon_0), \sigma_j^2(\varepsilon_0)) \in E^{Nash}(\gamma_j), \forall \gamma_j \in \Gamma$.

Proof. Appendix B. ■

Whenever reasoning costs are small enough equilibrium action and partition choices will be such that approximately a Nash equilibrium is played in all games. Thus - unless reasoning costs are significant - learning across games does not lead to deviations from this basic prediction of game theory.[20]

[17] Equations (10)-(11) constitute some particular form of perturbed replicator dynamics. The relation between perturbed reinforcement learning and replicator dynamics has been analyzed by Hopkins (2002). Börgers and Sarin (1997), Ianni (2000) or Laslier, Topol and Walliser (2001) have examined the relation between unperturbed reinforcement learning and replicator dynamics.

[18] Kuschner and Lin (2003), Benveniste, Métivier and Priouret (1987).

[19] See Benaïm and Hirsch (1999), Benaïm and Weibull (2003), Benveniste, Métivier and Priouret (1987), Kushner and Lin (2003) or Pemantle (1990).

[20] Note though that if reasoning costs were high or partitions exogenous many deviations from Nash equilibrium can be observed. Endogenizing partition choice thus restricts the set of possible outcomes considerably.

Naturally now the question arises how learning across games selects between (possibly) many Nash equilibria ? We will see in the following subsections that learning across games can have more "bite" than one would expect and often leads to a very strong and clear-cut selection. Furthermore this selection can work in different directions than it does with learning in a single game. Learning across games thus leads to new and interesting predictions. In particular we will see that:

- Nash equilibria in weakly dominated strategies that are unstable to learning in a single game can be asymptotically stable to learning across games. This is particularly interesting in extensive form games with non-generic strategic form representations. Weakly dominated strategies in the strategic form representation of these games typically correspond to non subgame perfect behavior in the extensive form.
- Learning across games can stabilize mixed strategy equilibria in Coordination and Anti-Coordination Games. These equilibria are unstable to learning in a single game.
- Learning across games can sometimes destabilize strict Nash equilibria. These equilibria are always stable to learning in a single game.

We will begin each of the following subsections with an intuitive example that illustrates our main points and then proceed to state the general results.

4.1 Nash Equilibria in Weakly Dominated Strategies

The example we will use in this subsection are two bargaining games - one where all the pie is gone after the first offer (i.e. an ultimatum game) and one with a strictly positive discount factor. Afterwards we will generalize the insights from this example and identify a class of situations in which learning across games can stabilize equilibria in weakly dominated strategies.

4.1.1 Bargaining

The Rubinstein model describes a process of bargaining between two individuals, 1 and 2, who have to decide how to divide a pie of size 1. The bargaining process is modeled as a sequence of alternating offers, responses and counteroffers at discrete times. Assume without loss of generality that player 1 proposes first a certain division of the pie $(a, 1-a)$ where a denotes the share of the pie she wants to keep for herself. Player 2 can either accept or reject the offer and make a counter-offer. Then it is player 1's turn to accept or reject and make a counter-offer. The process continues until an agreement is reached. The size of the pie though shrinks over time at rate δ. It is thus preferable to reach an agreement as early as possible. At each decision node of the game κ a strategy of a player is characterized by two numbers (a^κ, b^κ) where a^κ is the proposal (the share of pie he wants to keep) and b^κ the acceptance threshold. Let $a^{i\kappa}$

and $b^{i\kappa}$ be from the finite grid $A = \{0, \frac{1}{M}, \frac{2}{M}, .. \frac{M-1}{M}, 1\}$.[21] In the following let us focus on stationary strategies, i.e. strategies that do not depend on decision node κ. The process is illustrated in Figure 1.[22]

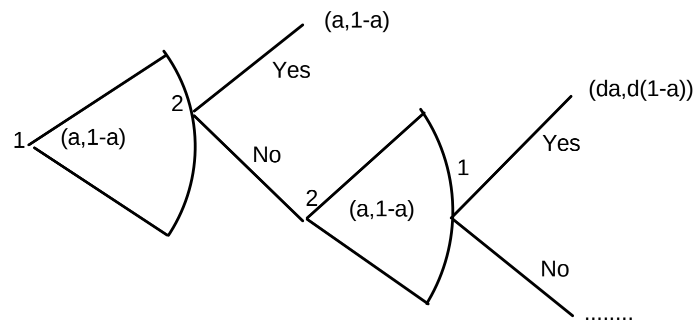

Figure 1: Rubinstein Bargaining

Consider now two Rubinstein games that differ only in the discount factor δ. At each point in time one of the games is randomly drawn and classified by the agents into an analogy class according to their partition choice probabilities. Then players choose an action according to their action choice probabilities and receive the (discounted) payoffs. Finally attractions and partitions are updated.

In particular let us consider the extreme case where one game γ_1 has discount factor $\delta_1 \to 1$, and the other game γ_2 has discount factor $\delta_2 = 0$. In γ_2 the whole pie is gone if the first offer is not accepted. γ_2 thus is essentially an Ultimatum Game. Even then both games have many Nash equilibria. There is a unique subgame perfect Nash equilibrium though in game γ_1 which involves $a^i = b^i = 1/2$, i.e. an equal split of the pie with an agreement reached in the first round. In γ_2, the ultimatum game, there are two SPNE involving either player 1 taking the whole pie and player 2 accepting all offers or player 1 proposing $\frac{M-1}{M}$ for himself and player 2 accepting all offers of at least $\frac{1}{M}$.[23] There are two possible partitions: A coarse partition in which players see the two games as analogous and a fine partition in which the games are distinguished. Denote the three analogy classes g_k with $k \in \{R, U, C\}$, corresponding to the Rubinstein game (with $\delta \to 1$), the Ultimatum game and the coarse partition. Whenever there is no reasoning cost ($\Xi(Z, \xi) = 0, \forall Z \in \mathbb{N}$) all asymptotically stable equilibria involve the fine partition and play of a subgame perfect Nash equilibrium in each of the games. For strictly positive reasoning costs (even if vanishingly small ($\xi \to 0$)) things change. Remembering that f_1 denotes the frequency with which game γ_1 occurs, we can state the following result.

[21] Assume that the grid A is fine enough s.th. it contains all equilibrium strategies described below. This assumption is a pure technicality facilitating the description of equilibria.

[22] In the graph the discount rate δ is denoted by d.

[23] This additional SPNE in the ultimatum game arises because of the discreteness of the action set. As $M \to \infty$, the two SPNE coincide.

Claim 1 $\forall \xi > 0$ *there exists an asymptotically stable equilibrium involving both players holding the coarsest partition* $G = \{\gamma_1, \gamma_2\}$, *player* 1 *demanding* $a^{1*} = \frac{1}{1+f_1}$ *and player* 2 *accepting all offers of at least* $b^{2*} = \frac{f_1}{1+f_1}$ *with asymptotic probability* 1.

Proof. Appendix B. ■

In this equilibrium players deviate from subgame perfection in both games. The equilibrium played is close to the SPNE in the Rubinstein game whenever this game is played with high probability and it is close to the SPNE of the Ultimatum Game whenever the latter is played very often. As the payoffs at stake are the same in both games agents will tend to play the more frequent game correctly (in the sense that equilibrium actions are closer to subgame perfection). Note though that the equilibrium from claim 1 is not unique. There is also an equilibrium in which the games are distinguished and the subgame perfect Nash equilibrium played in each game. The intuition for the result is as follows. Note that the equilibrium in which both games are seen as analogous induces approximate Nash play in both games and thus asymptotically there are no strict incentives to deviate from this equilibrium. A vanishingly small reasoning cost suffices to stabilize the equilibrium with the coarse partition provided that it is more important than noise. Whereas for (perturbed reinforcement) learning in a single game, asymptotic stability would select the SPNE in the ultimatum game, when there is learning across games deviations from subgame perfection can be observed. In fact there are many experiments that show that subjects often do not behave in accordance with subgame perfection.[24] If one thinks that the inclinations of experimental subjects to choose certain actions in the experiment have been shaped by a long process of reinforcement learning outside the laboratory, learning across games can provide an explanation for why deviations from subgame perfection are sometimes observed.

4.1.2 Equilibria in weakly dominated strategies

Note that the (non subgame perfect) equilibria from claim 1 correspond to Nash equilibria in weakly dominated strategies in the strategic form representation of the bargaining games. These equilibria are unstable to perturbed reinforcement learning in a single game. In fact whenever $\operatorname{card}\Gamma = 1$, i.e. whenever there is only one game, learning across games also predicts the instability of such equilibria. Whenever there is more than one game though learning across games can stabilize such equilibria. A case in which this is always true is whenever there are two games and the equilibrium in question is strict in the second game.

Proposition 3 *Let* $\widehat{\sigma}_1 = (\widehat{\sigma}_1^1, \widehat{\sigma}_1^2)$ *be a pure strategy Nash equilibrium in weakly dominated strategies in game* $\gamma_1 \in \Gamma$. *Then* $\forall \xi > 0$:

(i) If $\operatorname{card}\Gamma = 1$ *(learning in a single game), then* $\widehat{\sigma}_1$ *is not phenotypically induced at any asymptotically stable equilibrium* $x^* \in E^*$.

[24] See Binmore et al. (2002) and the references contained therein.

(ii) *If* $\operatorname{card}\Gamma > 1$ *this need not be true. Specifically if* $\operatorname{card}\Gamma = 2$ *and* $\widehat{\sigma}_2 = \widehat{\sigma}_1$ *is a strict Nash equilibrium in game* $\gamma_2 \neq \gamma_1$*, then there exists* $x^* \in E^*$ *which induces* $\widehat{\sigma}_1$ *in game* γ_1*.*

Proof. Appendix B. ■

Interesting implications of Proposition 3 concern extensive form games. The strategic form representations of these games are in general non generic and there is a relation between the concepts of weak dominance in the strategic form and subgame perfection in the extensive form. Generically in finite extensive form games with perfect information any equilibrium in weakly dominated strategies of the strategic form fails to be subgame perfect in the extensive form.[25] For example in the ultimatum game the (weakly dominated) equilibria in which the proposer offers a higher amount then $\frac{1}{M}$ to the responder do not satisfy the criterion of subgame perfection in the extensive form. As is shown in part (i) of Proposition 3, these equilibria are not selected for learning in a single game. We have already seen above though that such equilibria can be stable to learning across games. In fact the bargaining application also shows that the condition that $\widehat{\sigma}_1$ be a strict equilibrium in the second game is not necessary to stabilize such an equilibrium.

Next let us look at another class of situations where learning across games leads to new predictions.

4.2 Stabilization of Mixed Strategy Equilibria and Destabilization of Strict Nash Equilibria in Coordination Games

Interesting predictions of learning across games can also arise if Γ contains games with mixed strategy equilibria. In the following let it be a convention that whenever we say coordination games we refer to both Coordination games in the narrow sense and Anti-Coordination games. Again we start this subsection with an intuitive example before we move on to more general results.

4.2.1 2×2 games with mixed strategy equilibria

Consider the following payoff matrices:

$$M = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, M' = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

A set of games Γ_1 can be created by choosing either matrix M or M' for any player. Three strategic situations can arise: If both players have matrix M the resulting game γ_1 is one of pure Coordination.[26] If both players have matrix

[25] See chapter 6 in Osborne and Rubinstein (1994).

[26] A (pure) Coordination Game has two pure strategy Nash equilibria in which both agents choose the same action and a mixed strategy equilibrium.

M' the resulting game γ_2 is an Anti-Coordination Game.[27] And if player 1 has matrix M and player 2 has matrix M' the resulting game γ_3 is a game of Conflict, in which there is a unique equilibrium in mixed strategies (where both agents choose both actions with equal probability). These three games span the class of 2×2 games with a mixed strategy equilibrium. There are 5 possible partitions and $2^3 - 1 = 7$ possible analogy classes.

For learning in a single game the prediction in a model of perturbed reinforcement learning is that agents coordinate on one of the pure strategy equilibria in games γ_1 and γ_2 and play the mixed strategy equilibrium in game γ_3.[28] Simultaneous learning of actions and partitions leads to the same prediction whenever there are no reasoning costs ($\Xi(Z, \xi) = 0, \forall Z \in \mathbb{N}$). For vanishingly small costs ($\xi \to 0$) things already change. Denote the probability with which agent i chooses the first action in analogy class g_k by p_k^i and denote $g_c = \{\gamma_1, \gamma_2, \gamma_3\}$ the analogy class corresponding to the coarsest partition. The following result can be stated:

Claim 2 *Assume $f_j < 1/2, \forall j = 1, 2$. Then $\forall \xi > 0$ the unique asymptotically stable equilibrium for Γ_1 involves both players holding the coarsest partition $G_C = \{\gamma_1, \gamma_2, \gamma_3\}$ with asymptotic probability 1 and choosing $p_c^{i*} = 1/2, \forall i = 1, 2$.*

Proof. Appendix B. ■

Note that both pure strategies in the Coordination Games are a best response to the unique equilibrium in the Conflict Game. A small reasoning cost suffices to induce a tendency for the players to see all three games as analogous. The equilibrium with the coarse partition is stable whenever none of the Coordination Games is too important relative to the other two games. The reason is that if and only if $f_j < 1/2$ for $j = 1, 2$ the incentives of an agent who sees the three games as "one" correspond to those of a conflict game. Consequently, playing the mixed equilibrium with the coarse partition is asymptotically stable under this condition.

The example teaches us two things: on the one hand the presence of the conflict game destabilizes the otherwise asymptotically stable pure strategy equilibria in the Coordination Games (note that the equilibrium in claim 2 is unique). On the other hand the mixed equilibria in the Coordination games that are unstable to perturbed reinforcement learning in a single game, can be stabilized by learning across games.

4.2.2 Destabilization of strict Nash equilibria in Coordination Games

It is a well known result that strict Nash equilibria are asymptotically stable to many learning dynamics for a single game. Learning across games can sometimes destabilize strict Nash equilibria in Coordination and Anti-Coordination games

[27] An Anti-Coordination Game has two pure strategy Nash equilibria in which the agents choose different actions and a mixed strategy equilibrium.

[28] It is shown in Appendix B that the mixed strategy equilibrium in the conflict game is indeed asymptotically stable under perturbed reinforcement learning.

as we have seen in the previous example. This point is made more precise and general in the following proposition.

Proposition 4 *Let $\widehat{\sigma}_1 = (\widehat{\sigma}_1^1, \widehat{\sigma}_1^2)$ be a strict Nash equilibrium in $\gamma_1 \in \Gamma$. If γ_1 belongs to the class of 2×2 coordination games, then $\forall \xi > 0$:*

(i) If $\operatorname{card} \Gamma = 1$ *(learning in a single game), then there exists $x^* \in E^*$ that phenotypically induces $\widehat{\sigma}_1$.*

(ii) If $\operatorname{card} \Gamma > 1$ *this need not be true. Specifically let* $\operatorname{card} \Gamma = 2$ *and let γ_2 have a mixed equilibrium stable to learning in a single game. Then there exists $\widehat{f} > 0$ s.th. if $f_1/f_2 < \widehat{f}$ the strict Nash equilibrium $\widehat{\sigma}_1$ is not phenotypically induced at any asymptotically stable equilibrium $x^* \in E^*$.*

Proof. Appendix B. ■

The first part of this proposition shows that strict Nash equilibria are always stable to the perturbed reinforcement dynamics, if learning occurs in a single game. In fact it is a standard result for learning in a single game that strict Nash equilibria are always stable with respect to any deterministic payoff monotone dynamics.[29] If there are no reasoning costs ($\Xi(Z, \xi) = 0, \forall Z \in \mathbb{N}$) any strict Nash equilibrium can be induced in an asymptotically stable equilibrium even if there are many games. This is non surprising given that in this case the finest partition has the same reasoning cost, namely zero, as any other partition. These predictions change though once we have more than one game and allow for positive (even though arbitrarily small) reasoning costs. Specifically if the strict Nash equilibrium from some game is in the support of the unique stable equilibrium in a different game and furthermore the latter is a mixed strategy equilibrium, the strict equilibrium will be destabilized. The reason is that a) the mixed equilibrium will be observed in the second game in any asymptotically stable equilibrium as we know from Proposition 2 and b) the strict Nash equilibrium strategies are best responses to the mixed equilibrium. Even for a vanishingly small reasoning cost (provided that it is more important then noise) there will thus be tendency for agents to see the games as analogous and to save reasoning costs. Whenever the second game is sufficiently important learning across games stabilizes an equilibrium in which the strict Nash equilibrium is not played in game γ_1.

4.2.3 Stabilization of mixed Nash equilibria in Coordination Games

Similarly we have seen that mixed equilibria in 2×2 Coordination or Anti-Coordination games - that are unstable to learning in a single game - can be stabilized by learning across games.

[29] See for example proposition 5.11 in Weibull (1995).

Proposition 5 *Let $\widehat{\sigma}_1 = (\widehat{\sigma}_1^1, \widehat{\sigma}_1^2)$ be a mixed strategy Nash equilibrium in $\gamma_1 \in \Gamma$. If γ_1 belongs to the class of 2×2 coordination games, then $\forall \xi > 0$:*

(i) *If* $\operatorname{card} \Gamma = 1$ *(learning in a single game), then $\widehat{\sigma}_1$ is not phenotypically induced at any asymptotically stable equilibrium $x^* \in E^*$.*

(ii) If $\operatorname{card} \Gamma > 1$ *this need not be true. Specifically let* $\operatorname{card} \Gamma = 2$ *and let γ_2 has a mixed equilibrium $\widehat{\sigma}_2 = \widehat{\sigma}_1$ stable to learning in a single game. Then whenever $f_1/f_2 > \widehat{f}$, there exists $x^* \in E^*$ which induces $\widehat{\sigma}_1$ in game γ_1.*

Proof. Appendix B. ■

There has been a lot of research effort to investigate the stability properties of mixed equilibria. A very robust result from this literature is the instability of mixed equilibria in 2×2 pure Coordination and Anti-Coordination games in multipopulation models for very broad classes of dynamics.[30] Learning across games though can stabilize mixed equilibria in these games. Given the inherent instability of these equilibria for learning in a single game, it seems a reasonable conjecture that learning across games can stabilize mixed equilibria also in a far larger class of situations.

We have seen that learning across games often leads to interesting and new predictions for action choices. In the next section we will try to characterize the partition choices of agents.

5 Equilibrium Partitions

What partitions do agents choose in equilibrium ? The preceding sections suggest that the answer to this question will depend on a) reasoning costs and b) the degree of "overlap" between the Nash equilibria of the different games contained in Γ. Let us first continue with the assumption that reasoning costs are small and investigate the second conjecture and then shortly discuss reasoning costs.

Denote $S^{Nash}(\gamma_j)$ the support of the set of Nash equilibria $E^{Nash}(\gamma_j)$ of a game γ_j. Formally $S^{Nash}(\gamma_j) = \{a_m^i | \exists \sigma_j^i \in E^{Nash}(\gamma_j) \text{ with } \sigma_{mj}^i > 0\}$. The following proposition shows that if and only if the supports of the sets of Nash equilibria of the games in Γ are disjoint the finest partition will always emerge (unless reasoning costs are high).

Proposition 6 *If $\xi \to 0$ the finest partition G_F will be chosen with asymptotic probability $q_F^* = 1$ in all asymptotically stable equilibria if and only if $S^{Nash}(\gamma_j) \cap S^{Nash}(\gamma_h) = \varnothing$, $\forall \gamma_j \neq \gamma_h \in \Gamma$. Furthermore in this case the conclusions of part (i) of Propositions 3, 4 and 5 hold true in each of the games.*

[30] For learning in a single game results on the stability of mixed equilibria in multipopulation games are typically negative. Posch (1997) has analyzed stability properties of mixed equilibria in 2×2 games for unperturbed reinforcement learning. See also the textbooks by Weibull 1995, Vega-Redondo (2000) and Fudenberg and Levine (1998) or Hofbauer and Hopkins (2005) and Ellison and Fudenberg (2000) for recent research on this topic.

Proof. Appendix B. ■

The intuition is very simple: If the supports of the sets of Nash equilibria of two games are disjoint then seeing them as analogous necessarily involves choosing an action that is not a best response to the opponent's play for one of the players in one of the games. This player will gain from distinguishing these games. The following proposition establishes an upper bound on the cardinality of the partitions agents will use in equilibrium.

Proposition 7 $\forall \xi > 0, \forall i = 1, 2$ *any partition* $G_l \in \operatorname{supp} q^{i*}$ *has to satisfy* $\operatorname{card} G_l \leq \operatorname{card} A_i$.

Proof. Appendix B. ■

Equilibrium partitions have to be of smaller cardinality than a player's action set. The intuition behind the result is simple. If an equilibrium partition has higher cardinality then the action set, it will inevitably be the case that there exists an analogy class in which a best response is played to the opponent's play in all games of a different analogy class. Merging these analogy class will lead to asymptotically no payoff loss but save reasoning costs.

Reasoning Costs

Until now we have only considered the case of no or very small reasoning costs. Anything else would have been an arbitrary choice. We have seen that players will play approximately Nash equilibrium in all games. Obviously when reasoning costs are significant equilibrium outcomes can be quite different from Nash equilibria in some games. This raises the question of whether it is always optimal for an agent to have small reasoning costs (and thus to be able to hold partitions of high cardinality). If this were the case one could argue on evolutionary grounds that reasoning costs will most likely tend to be quite small. The following simple example shows that having smaller reasoning costs need not always lead to better outcomes for a player. Consider two games γ_1 and γ_2 with the following payoff matrices.

$$\gamma_1 = \begin{pmatrix} 1,1 & 4,3 & 3,1 \\ 1,3 & 5,1 & 1,2 \\ 2,4 & 2,1 & 1,1 \end{pmatrix}, \gamma_2 = \begin{pmatrix} 2,1 & 3,2 & 3,3 \\ 1,1 & 2,4 & 2,2 \\ 2,1 & 1,2 & 1,3 \end{pmatrix}$$

Assume both games occur with equal probability ($f_1 = f_2 = 1/2$). If reasoning costs are small both agents will use the fine partition in the unique asymptotically stable equilibrium and play the unique strict Nash equilibrium in each of the games. This leads to an outcome of $(2,4)$ in game γ_1 and $(3,3)$ in game γ_2. What would happen if player 1 had very high reasoning costs ? For high enough reasoning costs he would see both games as analogous.[31] It can be checked that the unique equilibrium in this case leads to an outcome of $(4,3)$ in game γ_1 and $(3,3)$ in γ_2. Player 1 is thus better off (both in terms of absolute and relative payoffs) if he has high reasoning costs. This example shows that

[31] Of course we have defined the process only for vanishingly small reasoning costs. Extending to general costs is no problem though. See footnote 9.

it is a priori not obvious in which direction evolutionary pressures will work on reasoning costs. To study this issue should be the object of further research.

6 Related Literature

The idea that similarities or analogies play an important role for economic decision making has long been present in the literature.[32] Most approaches have been axiomatic. Rubinstein (1988) gives an explanation of the Allais-paradox based on agents using similarity criteria in their decisions. Also Gilboa and Schmeidler (1995) argue that agents reason by drawing analogies to similar situations in the past. They derive representation theorems for an axiomatization of such a decision rule.[33] Jehiel (2005) proposes a concept of analogy-based reasoning. Seeing two games as analogous in his approach means having the same expectations about the opponent's behavior. Still agents act as expected utility maximizers in each game and can thus choose differently in games that are seen as analogous. All these approaches are static and partitions or similarity measures are exogenous.

LiCalzi (1995) studies a fictitious play like learning process in which agents decide on the basis of past experience in similar games. He is able to demonstrate almost sure convergence of such an algorithm in 2×2 games. Again similarity is exogenous in his model. Steiner and Stewart (2006) study similarity learning in global games using the similarity concept from Case-based decision theory.

Samuelson (2001) proposes an approach based on automaton theory in which agents group together games to reduce the number of (costly) states of automata. He finds that if agents - unlike in our paper - play in both player roles ultimatum games can be grouped together with bargaining games into a single state in order to save on complexity costs of automata with more states. The logic behind his result is quite different though from the logic behind claim 1 in this paper. While in his paper the existence of a tournament ensures high marginal costs for using additional states on the bargaining games, here the result holds also for vanishingly small marginal reasoning costs provided they are more important then noise.[34]

There is obviously also a relation to the literature on reinforcement learning. Conceptually related are especially Roth and Erev (1995) and Erev and Roth (1998) from which the basic reinforcement model is taken. Hopkins (2002) analyzes their basic model using stochastic approximation techniques. Also related are Ianni (2000), Börgers and Sarin (1997 and 2000) and Laslier, Topol and Walliser (2001) who rely on stochastic approximation techniques to analyze reinforcement models.[35]

[32] See Luce (1955) for early research on similarity in economics and Quine (1969) for a philosophical view on similarity.

[33] In Gilboa and Schmeidler (1996) they show that there is some conceptual relation between case-based optimization and the idea of satisficing on which reinforcement models are based.

[34] Other papers in the automaton tradition investigating equilibria in the presence of complexity costs are Abreu and Rubinstein (1988), Eliaz (2003) or Spiegler (2004).

[35] See also Karandikar et al. (1996), Posch (1997) or Hopkins (2005) for related learning

7 Conclusions

In this paper we have presented and analyzed a learning model in which decisionmakers learn simultaneously about actions and partitions of a set of games. We find that in equilibrium agents will partition the set of games according to strategic compatibility of the games. If the sets of Nash equilibria of any two games are disjoint agents will always distinguish these games in equilibrium. Whenever this is not the case though, interesting situations arise. In particular learning across games can destabilize strict Nash equilibria, stabilize mixed equilibria in 2×2 coordination and anti-coordination games and Nash equilibria in weakly dominated strategies. Furthermore learning across games can explain deviations from subgame perfection that are sometimes observed in experiments. Another recurrent observation in experiments is the existence of framing effects. A possible explanation for this phenomenon could be that different frames trigger different analogies. We strongly believe that analogy thinking and other instances of bounded rationality can constitute an explanation for many more experimental results. This line of research seems thus very worthwhile pursuing.

References

[1] Abreu, D. and A. Rubinstein (1988), The Structure of Nash Equilibrium in Repeated Games with Finite Automata, Econometrica 56(6), 1259-1281.

[2] Benaim, M. and M. Hirsch (1999), Mixed Equilibria and Dynamical Systems Arising from Fictitious Play in Perturbed Games, Games and Economic Behavior 29, 36-72.

[3] Benaim, M. and J. Weibull (2003), Deterministic Approximation of Stochastic Evolution in Games, Econometrica 71(3), 878-903.

[4] Benveniste A., M. Metevier and P. Priouret (1990), Adaptive Algorithms and Stochastic Approximation, Berlin: Springer Verlag.

[5] Binmore, K., J. McCarthy, G. Ponti, L. Samuelson and A. Shaked (2002), A Backward Induction Experiment, Journal of Economic Theory 104, 48-88.

[6] Börgers, T. and R. Sarin (1997), Learning through Reinforcement and Replicator Dynamics, Journal of Economic Theory 77, 1-14.

[7] Börgers, T. and R. Sarin (2000), Naive Reinforcement Learning With Endogenous Aspirations, International Economic Review 41(4), 921-950.

[8] Eliaz, K. (2003), Nash equilibrium when players account for the complexity costs of their forecasts, Games and Economic Behavior 44, 286-310.

models. Pemantle (1990), Benaim and Hirsch (1999) or Benaim and Weibull (2003) are technically related.

[9] Ellison, G. and D. Fudenberg (2000), Learning Purified Mixed Equilibria, Journal of Economic Theory 90, 84-115.

[10] Erev, I. and A.E. Roth (1998), Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria, American Economic Review 88(4), 848-881.

[11] Fudenberg, D. and D.K. Levine (1998), The Theory of Learning in Games, Cambridge: MIT-Press.

[12] Gilboa, I. and D. Schmeidler (1995), Case-Based Decision Theory, The Quarterly Journal of Economics, 110(3), 605-639.

[13] Gilboa, I. and D. Schmeidler (1996), Case-Based Optimization, Games and Economic Behavior 15, 1-26.

[14] Hofbauer, J., and E. Hopkins (2005), Learning in perturbed asymmetric games, Games and Economic Behavior 52, 133-152.

[15] Hopkins, E. (2002), Two Competing Models of How People Learn in Games, Econometrica 70(6), 2141-2166.

[16] Hopkins, E. (2005), Adaptive Learning Models of Consumer Behavior, working paper, University of Edinburgh.

[17] Ianni, A. (2000), Reinforcement Learning and the Power Law of Practice: Some Analytical Results, working paper, University of Southhampton.

[18] Jehiel, P. (2005), Analogy-based exspectation equilibrium, Journal of Economic Theory 123, 81-104.

[19] LiCalzi, M. (1995), Fictitious Play by Cases, Games and Economic Behavior 11, 64-89.

[20] Karandikar R., D. Mookherjee, D. Ray and F. Vega-Redondo (1998), Evolving Aspirations and Cooperation, Journal of Economic Theory 80, 292-331.

[21] Kuschner, H.J. and G.G. Lin (2003), Stochastic Approximation and Recursive Algorithms and Applications, New York: Springer.

[22] Laslier, J-F., R. Topol and B. Walliser (2001), A Behavioural Learning Process in Games, Games and Economic Behavior 37, 340-366.

[23] Luce, R.D. (1955), Semiorders and a Theory of Utility Discrimination, Econometrica 24(2), 178-191.

[24] Osborne, M.J. and A. Rubinstein (1994), A Course in Game Theory, Cambridge: MIT-Press.

[25] Pemantle, R. (1990), Nonconvergence To Unstable Points in Urn Models And Stochastic Approximations, The Annals of Probability 18(2), 698-712.

[26] Posch, M. (1997), Cycling in a stochastic learning algorithm for normal form games, Journal of Evolutionary Economics 7, 193-207.

[27] Quine, W.V. (1969), Natural Kinds, in: Essays in Honor of Carl G. Hempel, eds. Rescher, N., Reidel, D., Publishing Company Dordrecht, Boston.

[28] Roth, A.E. and I. Erev (1995), Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term, Games and Economic Behavior 8, 164-212.

[29] Rubinstein, A. (1988), Similarity and Decision-making under Risk (Is There a Utility Theory Resolution to the Alais Paradox?), Journal of Economic Theory 46, 145-153.

[30] Samuelson, L. (2001), Analogies, Anomalies and Adaptation, Journal of Economic Theory 97, 320-366.

[31] Spiegler, R. (2004), Simplicity of beliefs and delay tactics in a concession game, Games and Economic Behavior 47, 200-220.

[32] Steiner, J. and C. Stewart (2006), Learning by Similarity in Coordination Problems, mimeo CERGE-EI.

[33] Vega-Redondo, F. (2000), Economics and the Theory of Games, Cambridge University Press.

[34] Weibull, J. (1995), Evolutionary Game Theory, Cambridge: MIT-Press.

A Appendix: Proofs from Section 3

Proof of Lemma 1:
Proof. In the proof of Lemma 1 and 2 we will index player 2's actions by n instead of m to avoid confusion. Focus without loss of generality on player 1. It follows from (2) and (3) that the change in action choice frequency for action a_m in analogy class g_k is given by

$$p_{mk}^{1(t+1)} - p_{mk}^{1t} = \begin{cases} \frac{\beta_{mk}^{1t}+\pi^1(a^t,\gamma^t)+\varepsilon_0}{\sum_{a_h\in A_1}\beta_{hk}^{1t}+\pi^1(a^t,\gamma^t)+M\varepsilon_0} - \frac{\beta_{mk}^{1t}}{\sum_{a_h\in A_1}\beta_{hk}^{1t}} & \text{if } g_k, a_m \in w^{it} \\ \frac{\beta_{mk}^{1t}+\varepsilon_0}{\sum_{a_h\in A_1}\beta_{hk}^{1t}+\pi^1(a^t,\gamma^t)+M\varepsilon_0} - \frac{\beta_{mk}^{1t}}{\sum_{a_h\in A_1}\beta_{hk}^{1t}} & \text{if } g_k \in w^{it},\ a_m \notin w^{it} \\ \frac{\beta_{mk}^{1t}+\varepsilon_0}{\sum_{a_h\in A_1}\beta_{hk}^{1t}+M\varepsilon_0} - \frac{\beta_{mk}^{1t}}{\sum_{a_h\in A_1}\beta_{hk}^{1t}} & \text{if } g_k \notin w^{it} \end{cases} \tag{12}$$

or equivalently

$$p_{mk}^{1(t+1)} - p_{mk}^{1t} = \begin{cases} \frac{(1-p_{mk}^{1t})\pi^1(a^t,\gamma^t)+\varepsilon_0(1-Mp_{mk}^{1t})}{\sum_{a_h\in A_1}\beta_{hk}^{1t}+\pi^1(a^t,\gamma^t)+M\varepsilon_0} & \text{if } g_k, a_m \in w^{it} \\ \frac{-p_{mk}^{1t}\pi^1(a^t,\gamma^t)+\varepsilon_0(1-Mp_{mk}^{1t})}{\sum_{a_h\in A_1}\beta_{hk}^{1t}+\pi^1(a^t,\gamma^t)+M\varepsilon_0} & \text{if } g_k \in w^{it},\ a_m \notin w^{it} \\ \frac{\varepsilon_0(1-Mp_{mk}^{1t})}{\sum_{a_h\in A_1}\beta_{hk}^{1t}+M\varepsilon_0} & \text{if } g_k \notin w^{it} \end{cases} \tag{13}$$

The first event has the following probability

$\sum_{\gamma_j\in\Gamma} f_j\, \mathrm{I}_{jk} \sum\limits_{G_l\in\mathcal{G}} q_l^{1t}\mathrm{I}_{kl}p_{mk}^{1t} \sum_{a_n\in A_2}\left(\sum_{G_l\in\mathcal{G}} q_l^{2t}\sum_{g_k\in G_l} p_{nk}^{2t}\mathrm{I}_{jk}\right)$ where $\mathrm{I}_{jk}(\mathrm{I}_{kl}) = 1$ if $\gamma_j \in g_k$ $(g_k \in G_l)$ and zero otherwise.

Note that $\left(\sum_{G_l\in\mathcal{G}} q_l^{2t}\sum_{g_k\in G_l} p_{nk}^{2t}\mathrm{I}_{jk}\right) = \sigma_{nj}^{2t}$. The second event has probability

$\sum_{\gamma_j\in\Gamma} f_j\ \mathrm{I}_{jk} \sum\limits_{G_l\in\mathcal{G}} q_l^{1t}\mathrm{I}_{kl} \sum_{a_h\in A} p_{hk}^{1t}(1-\delta_{hm})\sum_{a_n\in A_2}\sigma_{nj}^{2t}$ where δ_{hm} is the Kronecker delta.[36]

The third event has probability $\sum_{\gamma_j\in\Gamma} f_j(1-\mathrm{I}_{jk}) + f_j\mathrm{I}_{jk}\sum\limits_{G_l\in\mathcal{G}} q_l^{1t}(1-\mathrm{I}_{kl})$.

Summing over all possible events (weighted with the probabilities) gives the mean change:

$$\begin{aligned}
&\left\langle p_{mk}^{1(t+1)} - p_{mk}^{1t}\right\rangle = \sum_{\gamma_j\in g_k} f_j \sum_{G_l\in\mathcal{G}} q_l^{1t}\mathrm{I}_{kl}\\
&\left[p_{mk}^{1t}\sum_{a_n\in A_2}\frac{(1-p_{mk}^{1t})\pi^1(a_m^1,a_n^2,\gamma_j)\sigma_{nj}^{2t}+\varepsilon_0(1-Mp_{mk}^{1t})}{\sum_{a_h\in A_1}\beta_{hk}^{1t}+\pi^1(a_m^1,a_n^2,\gamma_j)+M\varepsilon_0}\right.\\
&\left.+\sum_{a_\eta\neq a_m\in A_1} p_{\eta k}^{1t}\sum_{a_n\in A_2}\frac{-p_{mk}^{1t}\pi^1(a_\eta^1,a_n^2,\gamma_j)\sigma_{nj}^{2t}+\varepsilon_0(1-Mp_{mk}^{1t})}{\sum_{a_h\in A_1}\beta_{hk}^{1t}+\pi^1(a_h^1,a_n^2,\gamma_j)+M\varepsilon_0}\right]\\
&+\left(1-\sum_{\gamma_j\in g_k} f_j\sum_{G_l\in\mathcal{G}} q_l^{1t}\mathrm{I}_{kl}\right)\frac{\varepsilon_0(1-Mp_{mk}^{1t})}{\sum_{a_h\in A_1}\beta_{hk}^{1t}+M\varepsilon_0}
\end{aligned}\tag{14}$$

Denoting $\beta_k^{1t} = \sum_{a_h\in A_1}\beta_{hk}^{1t}$ and remembering that $\sum_{\gamma_j\in\Gamma} f_j\ \mathrm{I}_{jk}\sum\limits_{G_l\in\mathcal{G}} q_l^{1t}\mathrm{I}_{kl} = r_k^{1t}$ this can be rewritten concisely as follows:

$$\left\langle p_{mk}^{1(t+1)} - p_{mk}^{1t}\right\rangle = \frac{1}{\beta_k^{1t}}[p_{mk}^{1t}r_k^{1t}S_{mk}^{1t}(\cdot)+\varepsilon_0(1-Mp_{mk}^{1t})] + O\left((\frac{1}{\beta_k^{1t}})^2\right) \tag{15}$$

To see that the difference between the first term in (15) and expression (14) is

[36] $\delta_{hm} = 1$ if $h = m$ and $\delta_{hm} = 0$ otherwise.

indeed of order $(\frac{1}{\beta_k^{1t}})^2$ note that this difference can be written

$$
\begin{aligned}
& \frac{p_{mk}^{1t} r_k^{1t} S_{mk}^{1t}(\cdot) + \varepsilon_0 (1 - M p_{mk}^{1t})}{\beta_k^{1t}} - \left\langle p_{mk}^{1(t+1)} - p_{mk}^{1t} \right\rangle \\
= \quad & \frac{p_{mk}^{1t} r_k^{1t} S_{mk}^{1t} + \varepsilon_0 (1 - M p_{mk}^{1t}) - p_{mk}^{1t} r_k^{1t} S_{mk}^{1t} (1 + \frac{\pi^1(\cdot) + M\varepsilon_0}{\beta_k^{1t}})^{-1}}{\beta_k^{1t}} \\
& - \frac{\varepsilon_0 (1 - M p_{mk}^{1t})(1 + \frac{M\varepsilon_0}{\beta_k^{1t}})^{-1}}{\beta_k^{1t}} \\
= \quad & p_{mk}^{1t} r_k^{1t} S_{mk}^{1t} \frac{(\pi^1(\cdot) + M\varepsilon_0)}{\beta_k^{1t}(\beta_k^{1t} + \pi^1(\cdot) + M\varepsilon_0)} + \varepsilon_0 (1 - M p_{mk}^{1t}) \frac{M\varepsilon_0}{\beta_k^{1t}(\beta_k^{1t} + M\varepsilon_0)}.
\end{aligned}
$$

■

Proof of Lemma 2:

Proof. The changes in partition choice probabilities are given by

$$
q_l^{1(t+1)} - q_l^{1t} = \begin{cases} \frac{(1-q_l^{1t})\left(\pi^1(a^t,\gamma^t) - \Xi(Z_l)\right) + \varepsilon_1(1 - L q_l^{1t})}{\sum_{G_h \in \mathcal{G}} \alpha_h^t + \pi^1(a^t,\gamma^t) + L\varepsilon_1} & \text{if } G_l \in w_1^t \\ \frac{-q_l^{1t}\left(\pi^1(a^t,\gamma^t) - \Xi(Z_h)\right) + \varepsilon_1(1 - L q_l^{1t})}{\sum_{G_h \in \mathcal{G}} \alpha_h^t + \pi^1(a^t,\gamma^t) + L\varepsilon_1} & \text{if } G_l \notin w_1^t \end{cases} \tag{16}
$$

where $L = \operatorname{card} \mathcal{G}$. The first event occurs with probability $\sum_{\gamma_j \in \Gamma} f_j \sum_{A_1 \times A_2} q_l^t \left(\sum_{g_k \in G_l} p_{mk}^{1t} \mathrm{I}_{jk} \right) \sum_{a_n \in A_2} \sigma_{nj}^{2t}$. The second event occurs with probability $\sum_{\gamma_j \in \Gamma} f_j \sum_{A \times A} \sum_{G_h \neq G_l} q_h^t \left(\sum_{g_k \in G_h} p_{mk}^{1t} \mathrm{I}_{jk} \right) \sum_{a_n \in A_2} \sigma_{nj}^{2t}$. Multiplying delivers

$$
\left\langle q_l^{1(t+1)} - q_l^{1t} \right\rangle = \sum_{\gamma_j \in \Gamma} f_j
\left[\begin{array}{l} q_l^{1t} \sum_{a_n \in A_2} \frac{(1-q_l^{1t})\left(\sum_{g_k \in G_l} p_{mk}^{1t} \mathrm{I}_{jk}\right)\left(\pi^1(a_m^1, a_n^2, \gamma_j) - \Xi(Z_l)\right)\sigma_{nj}^{2t} + \varepsilon_1(1 - L q_l^{1t})}{\sum_{G_h \in \mathcal{G}} \alpha_h^t + \pi^1(a_m^1, a_n^2, \gamma_j) + L\varepsilon_1} \\ + \sum_{G_h \neq G_l} q_h^{1t} \frac{-q_l^{1t}\left(\sum_{g_k \in G_h} p_{mk}^{1t} \mathrm{I}_{jk}\right)\left(\pi^1(a_m^1, a_n^2, \gamma_j) - \Xi(Z_h)\right)\sigma_{nj}^{2t} + \varepsilon_1(1 - L q_l^{1t})}{\sum_{G_h \in \mathcal{G}} \alpha_h^t + \pi^1(a_m^1, a_n^2, \gamma_j) + L\varepsilon_1} \end{array} \right]
$$

Denoting $\sum_{G_l \in \mathcal{G}} \alpha_l^{1t} =: \alpha^{1t}$ the previous expression can be rewritten concisely as:

$$
\left\langle q_l^{1(t+1)} - q_l^{1t} \right\rangle = \frac{1}{\alpha^{1t}} [q_l^{it} S_l^{it}(x) + \varepsilon_1 (1 - L q_l^{it})] + O\left((\frac{1}{\alpha^t})^2 \right). \tag{17}
$$

■

Proof of Proposition 1:

Proof. Write the stochastic process $\{x^t\}_t$ in the form

$$
\begin{aligned}
p_{mk}^{i(t+1)} &= p_{mk}^{it} + \frac{1}{\beta_k^{it}} \widetilde{Y}_{mk}^{it} \\
q_l^{i(t+1)} &= q_l^{it} + \frac{1}{\alpha^{it}} Y_l^{it}
\end{aligned} \tag{18}
$$

$\forall i=1,2, \forall a_m \in A_i, \forall g_k \in \mathcal{P}^+(\Gamma), \forall G_l \in \mathcal{G}$. The Y^{it} and $\widetilde{Y}^{it}$ can be decomposed as follows:

$$\begin{aligned} \widetilde{Y}_{mk}^{it} &= \widetilde{y}_{mk}^{i}(x^t) + \widetilde{\omega}^{it}(c^t, d^t) + \widetilde{\upsilon}^{it} \\ Y_l^{it} &= y_l^i(x^t) + \omega^{it}(c^t, d^t) + \upsilon^{it} \end{aligned}$$

where the sequences $\{\upsilon^{it}\}_t$ and $\{\widetilde{\upsilon}^{it}\}_t$ are asymptotically negligible. The sequences $\{\omega^{it}\}_t$ and $\{\widetilde{\omega}^{it}\}_t$ are noise keeping track of the players randomizations at each period as well as of random sampling from Γ. In fact c^t is the indicator function for outcomes of players randomizations between actions and partitions and d^t the indicator function for outcomes of random sampling of games. And finally $\widetilde{y}_{mk}^i(x^t) = p_{mk}^{1t} r_k^{1t} S_{mk}^{1t}(\cdot) + \varepsilon_0(1 - M p_{mk}^{1t})$ and $y_l^i(x^t) = q_l^{it} S_l^{it}(\cdot) + \varepsilon_1(1 - L q_l^{it})$ are the mean motions derived before. Taking into account the normalization the stochastic process (18) can be rewritten as :

$$\begin{aligned} p_{mk}^{i(t+1)} &= p_{mk}^{it} + \frac{1}{\mu + t\theta} \widetilde{Y}^{it} \\ q_l^{i(t+1)} &= q_l^{it} + \frac{1}{\mu + t\theta} Y^{it} \end{aligned} \tag{19}$$

where $1/(\mu + t\theta)$ is the unique step size that is of order t^{-1}. It can be verified that the following conditions hold for the process (19): **(C1) :** $E[\omega^{it} | \omega^{in}, n < t] = 0$ and $E[\widetilde{\omega}^{it} | \widetilde{\omega}^{in}, n < t] = 0$. **(C2):** $\sup_t E\left|Y^{it}\right|^2 < \infty, \sup_t E\left|\widetilde{Y}^{it}\right|^2 < \infty$, **(C3)** $E\widetilde{y}^i(p^t, q^t)$ and $Ey^i(p^t, q^t)$ are C^2 (implying locally Lipschitz), **(C4)** $\sum_t \frac{1}{\mu + t\theta} \left|\upsilon^{it}\right| < \infty$ with probability 1 and **(C5)** $\sum_{t=0}^{\infty} \frac{1}{\mu + t\theta} = \infty$, $\frac{1}{\mu + t\theta} \geq 0, \forall t \geq 0$, and $\sum_{t=0}^{\infty} \left(\frac{1}{\mu + t\theta}\right)^2 < \infty$ (decreasing gains). Under these conditions the process (19) can be approximated by the deterministic system

$$\begin{aligned} \dot{p}_{mk}^{i} &= \widetilde{y}_{mk}^{i}(x) \\ \dot{q}_l^{i} &= y_l^1(x) \end{aligned}$$

$\forall i = 1, 2, \forall a_m \in A_i, \forall g_k \in \mathcal{P}^+(\Gamma), \forall G_l \in \mathcal{G}$ as standard results in stochastic approximation theory show.[37] ■

B Appendix: Proofs from Sections 4 and 5

Proof of Proposition 2:
Proof. By contradiction: Assume x^* is a restpoint that induces $(\sigma_j^1, \sigma_j^2)^* \notin E^{Nash}(\gamma_j)$. Let player i have a strictly better response in γ_j, denote $\widehat{a}_m$. If γ_j is an element of a singleton analogy class g_k the claim is straightforward, as the expected payoff of $\widehat{a}_m$ at x^* conditional on visiting g_k is strictly higher than that of all actions on average, i.e. $S_{mk}(x^*) > 0$. It follows then directly from (10)

[37] See the textbooks of Kuschner and Lin (2003) or Benveniste, Metevier and Priouret (1990).

that the growth rate function $\dot{p}_{mk}/p_{mk}$ of $\widehat{a}_m$ is strictly positive for all x in an open neighborhood of x^* (note that x^* is interior because of the perturbation).

Consider next the case where γ_j is an element of a non-singleton analogy class. Denote $\phi := \pi^i(\widehat{a}_m, \sigma_j^{-i}, \gamma_j) - \pi^i(\sigma_j^i, \sigma_j^{-i}, \gamma_j) > 0$ the payoff loss incurred by choosing σ_j^i instead of the better response $\widehat{a}_m$ in game γ_j. Consider a partition $G_h = \{g_h\}_{h=1}^{Z_h}$ in the support of q^{i*}. Assume that $\gamma_j \in \widetilde{g} \in G_h$. Partition $G_l = \{\{g_h - \widetilde{g}\}, \widetilde{g} - \gamma_j, \gamma_j\}$ coincides with partition G_h except for the fact that instead of analogy class $\widetilde{g}$ it contains two new analogy classes given by $\widetilde{g} - \gamma_j$ and the singleton analogy class γ_j. Consequently $\operatorname{card} G_l = (\operatorname{card} G_h) + 1$. We have seen above that in the singleton analogy class player i will play a best response to the the opponent's play. But then $\exists \widehat{\xi} < \phi$ such that $\forall \xi < \widehat{\xi}$: $\Pi_l^i(x^*) - \Pi_h^i(x^*) = \phi - (\Xi(Z_l) - \Xi(Z_h)) > 0$ and thus the growth rate function $\dot{q}_l/q_l$ is strictly positive for all x in an open neighborhood of x^*. Consequently x^* cannot be a stable restpoint. ■

Proof of Claim 1:

Proof. Consider the following strategic form game.

	$(a,0)$	$(0,\frac{1}{M})$	$(\frac{1}{M},\frac{1}{M})$	$..(\frac{1}{2},\frac{1}{2})..$	$(\frac{M-1}{M},1)$	$(1,1)$
$(0,b)$	$0,1$	$0,1$	$0,1$	$..0,1..$	$0,1$	$0,1$
$(\frac{1}{M},0)$	$\frac{1}{M},\frac{M-1}{M}$	$\frac{1}{M},\frac{M-1}{M}$	$\frac{1}{M},\frac{M-1}{M}$	$.\frac{1}{M},\frac{M-1}{M}.$	$\frac{\delta}{M},\delta\frac{M-1}{M}$	$0,0$
$(\frac{1}{M},\frac{1}{M})$	$\frac{1}{M},\frac{M-1}{M}$	$\frac{1}{M},\frac{M-1}{M}$	$\frac{1}{M},\frac{M-1}{M}$	$.\frac{1}{M},\frac{M-1}{M}.$	$\frac{\delta}{M},\delta\frac{M-1}{M}$	$0,0$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$.... \vdots ...$	$\vdots$	$\vdots$
$(\frac{1}{2},\frac{1}{2})$	$\frac{1}{2},\frac{1}{2}$	$\frac{1}{2},\frac{1}{2}$	$\frac{1}{2},\frac{1}{2}$	$..\frac{1}{2},\frac{1}{2}..$	$0,0$	$0,0$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$.... \vdots ....$	$\vdots$	$\vdots$
$(1,\frac{M-2}{M})$	$1,0$	$\delta,0$	$\delta\frac{M-1}{M},\frac{\delta}{M}$	$..0,0..$	$0,0$	$0,0$
$(1,\frac{M-1}{M})$	$1,0$	$\delta,0$	$\delta\frac{M-1}{M},\frac{\delta}{M}$	$..0,0..$	$0,0$	$0,0$
$(1,1)$	$1,0$	$\delta,0$	$0,0$	$..0,0..$	$0,0$	$0,0$

(20)

Game γ_1 is obtained by letting $\delta \to 1$ in matrix (20). Substituting $\delta = 0$ into (20) leads to the payoff matrix for the Ultimatum Game (γ_2).[38] Next consider the game with strategic form (20) and discount factor f_1. This is the expected discount factor when both games are seen as analogous (and game γ_1 occurs with frequency f_1). Note that player 1 proposing $\frac{1}{1+f_1}$ and player 2 accepting all offers above $\frac{f_1}{1+f_1}$ is the unique Nash equilibrium in weakly undominated strategies in this game. It can be shown that this equilibrium is asymptotically stable to perturbed reinforcement learning in a single game with discount factor f_1. Now we will show that the equilibrium x^* where both players hold the coarse partition and choose $a^1 = \frac{1}{1+f_1}, b^2 = \frac{f_1}{1+f_1}$ when visiting analogy class $g_C = \{\gamma_1, \gamma_2\}$ is asymptotically stable. When visiting the "off equilibrium" analogy classes g_U

[38]Note that in the Ultimatum game all strategies of player 2 are weakly dominated except the strategies $(a, 0)$ and $(a, \frac{1}{M})$. It is shown below (proof of proposition 3) that this implies that if the ultimatum game is played alone the unique equilibrium will have player 1 proposing $a^1 = \frac{M-1}{M}$ and player 2 randomizing between acceptance tresholds $b^2 = 0$ and $b^2 = \frac{1}{M}$. Note also that as there is no strict Nash equilibrium in game γ_1 (the bargaining game with $\delta \to 1$), part (ii) of Proposition 3 is not directly applicable.

and g_R, player 1 will play $(\frac{1}{1+f_1}, b)$, i.e. a best response to observed play of player 2. Player 2 will play $(\frac{1}{1+f_1}, \frac{f_1}{1+f_1})$ when visiting g_R but will randomize between strategies (a, b) with $b \leq \frac{f_1}{1+f_1}$ in g_U. Consider deviations in partition choice frequencies. Such deviations will increase the probability with which either player visits the fine partition. Given "off-equilibrium" play in analogy classes g_U and g_R increasing the probability of visiting the fine partition leads to a strict payoff loss for player 1. For player 2 there can be gains of $O(\varepsilon_0)$ for all strategies (a, b) with $b < \frac{f_1}{1+f_1}$ as $x^* \in \operatorname{int} \mathbf{X}$ and as $(\frac{1}{1+f_1}, \frac{f_1}{1+f_1})$ is weakly dominated for this player in the ultimatum game. Let $\mathcal{N}_{x^*}$ be an open neighborhood of x^* and denote $\Xi(x^*)$ the total reasoning cost at x^*, i.e. $\Xi(x^*) = \sum_{G_l \in \mathcal{G}} q_l^* \Xi(G_l)$. Then $\forall \xi > 0$:

$$\sum_{\Gamma} \left(\pi^i(x^*, x^{-i}, \gamma_j) - \pi^i(x, \gamma_j)\right) - (\Xi(x^*) - \Xi(x)) > 0, \forall x \in \mathcal{N}_{x^*} \cap \mathbf{X}, i = 1, 2.$$

Consider the (relative entropy) function associated with x^*, given by $D^i(x^*, x) = \sum_{A_1 \times A_2 \times \mathcal{G}} x^* \ln \frac{x_h^*}{x_h}$. Define the sum over the entropy functions for both players by $Q(x^*, x) = D^1(x^*, x) + D^2(x^*, x)$. It follows from the above equation that $\dot{Q}(x^*, x) < 0$. Thus $Q(x^*, x)$ is a strict Lyapunov function and x^* asymptotically stable. ■

Proof of Proposition 3:

Proof. (i) As card $\Gamma = 1$ there is trivially only one partition and one analogy class. Consequently action choice corresponds to observed play in the only game γ_1. Denote $\widehat{\sigma}_1^i$ the weakly dominated strategy and a^{*i} the strategy that weakly dominates $\widehat{\sigma}_1^i$ in game γ_1. It is clear that $\pi^i(a^{*i}, x^{-i}, \gamma_1) - \pi^i(\widehat{\sigma}_1^i, x^{-i}, \gamma_1) > 0, \forall x^{-i} \in \operatorname{int} \mathbf{X}_{-i}$. Consider a restpoint $\widehat{x}$ that induces $\widehat{\sigma}_1^i$. As $\widehat{x}$ is interior there exists a neighborhood $\mathcal{N}_{\widehat{x}}$ of $\widehat{x}$ such that

$$\pi^i(a^{*i}, x^{-i}, \gamma_1) - \pi^i(x, \gamma_1) + O(\varepsilon_0) > 0, \forall x \in \mathcal{N}_{\widehat{x}} \cap \operatorname{int} \mathbf{X}.$$

But then there exists $\vartheta > 0$ such that $\dot{p}_{a^{*i}}^i > \vartheta \; p_{a^{*i}}^i$ and integrating yields $p_{a^{*i}}^{it} \geq p_{a^{*i}}^{i0} \exp^{\vartheta t}, \forall x \in \mathcal{N}_{\widehat{x}} \cap \operatorname{int} \mathbf{X}, x \neq \widehat{x}$. As $p_{a^{*i}}^i$ increases exponentially at rate ϑ in the neighborhood of $\widehat{x}$ the latter cannot be a stable restpoint.[39]

(ii) We will show that the restpoint $\widehat{x}$ where the coarse partition $G_C = \{\gamma_1, \gamma_2\}$ is held with asymptotic probability $\widehat{q}_C = 1$ by all players that play a weakly dominated strategy in game γ_1 is asymptotically stable. Denote the three analogy classes by $g_1 = \{\gamma_1\}, g_2 = \{\gamma_2\}$ and $g_C = \{\gamma_1, \gamma_2\}$. Denote $\widehat{\sigma}^i$ the strategy that is weakly dominated for i in game γ_1 and a strict best response to x^{-i} in game γ_2.[40] For all x in an open neighborhood of $\widehat{x}$ we have that $\pi^i(\sigma^i, x^{-i}, \gamma_2) - \pi^i(x, \gamma_2) < \phi_2(\varepsilon_0), \forall \sigma^i \neq \widehat{\sigma}^i$ because $\widehat{\sigma}^i$ is a best response to x^{-i}. Furthermore $\pi^i(\sigma^i, x^{-i}, \gamma_1) - \pi^i(x, \gamma_1) < \phi_{10}(\varepsilon_0)$ and $\pi^i(\widetilde{a}^i, x^{-i}, \gamma_1) - \pi^i(x, \gamma_1) > \phi_{11}(\varepsilon_0)$ for some $\widetilde{a}^i \in A_i$ where the second inequality holds because $\widehat{\sigma}^i$ is weakly dominated in γ_1. Note that $\phi_2'(\varepsilon_0) < 0$ and $\phi_{11}'(\varepsilon_0) > 0 \forall \varepsilon_0 \in \mathbb{R}^+$.

[39]Part (i) of this proposition also follows from Proposition 5.8 in Weibull (1995).
[40]Of course $\widehat{\sigma}_1^i$ will be a pure strategy.

Furthermore $\phi_2(0) > 0$ and $\phi_{11}(0) = 0$ and thus $\lim_{\varepsilon_0 \to 0}(\phi_2(\varepsilon_0) - \phi_{11}(\varepsilon_0)) =: \phi > 0$. Furthermore note that in all "off-equilibrium" analogy classes best responses will be played and consequently deviations in partition choice frequencies will lead to at most gains of order ε_0. But then there exists a neighborhood $\mathcal{N}'_{\widehat{x}}$ of $\widehat{x}$ such that $\forall \xi > 0$,

$$\sum_{\Gamma} \left(\pi^i(\widehat{x}, x^{-i}, \gamma_j) - \pi^i(x, \gamma_j)\right) - (\Xi(\widehat{x}) - \Xi(x)) > 0, \forall x \in \mathcal{N}'_{\widehat{x}} \cap \mathbf{X}, i = 1, 2.$$

A strict Lyapunov function can be found as in the proof of claim 1. ■

Proof of Claim 2:

Proof. Let $G_1 = \{\{\gamma_1\}, \{\gamma_2\}, \{\gamma_3\}\}$, $G_2 = \{\gamma_1, \{\gamma_2, \gamma_3\}\}, G_3 = \{\gamma_2, \{\gamma_1, \gamma_3\}\}$, $G_4 = \{\{\gamma_1, \gamma_2\}, \gamma_3\}$ and $G_5 = \{\gamma_1, \gamma_2, \gamma_3\}$ be the five possible partitions of Γ. We will first argue that any restpoint where $q_l^i > 0$ for some $l = 1, 2, 3, 4$ and $i = 1, 2$ is unstable. Then we will show that the restpoint with $q_5^i = 1$ and $p_C^i = 1/2$, $\forall i = 1, 2$ is asymptotically stable.

(i) Note first that no stable restpoint $\widehat{x}$ can involve $q_l^i > 0$ for $l = 1, 4$. The reason is that in analogy class $g_3 = \{\gamma_3\} \in G_l, \forall l = 1, 4$ the unique Nash equilibrium strategy $p_3 = 1/2$ will always be played. But then also in analogy classes $g_k \neq g_3$ best responses to the opponent's observed play in g_3 are chosen. On the other hand there always exists a coarser partition where a best response to the opponent's play in the remaining analogy classes $g \in G_l, l = 1, 4$ is played. This destabilizes any such equilibrium. Consider for example the Jacobian matrix $\mathcal{M}$ associated with the linearization of the dynamics at restpoints $\widehat{x}$ that involve $q_1^i > 0$, for some $i = 1, 2$. Then if $f_1 > f_3$ a best response to the opponent's play in *both* games γ_1 and γ_3 will always be played in analogy class $\{\gamma_1, \gamma_3\}$. Consequently the diagonal element of $\mathcal{M}$ corresponding to q_3 will be strictly positive. More precisely $\left(\partial \dot{q}_3^i / \partial q_3^i\right) = \Xi(\widehat{x}) - \Xi(2) - 5\varepsilon_1 > 0$, as the coarse partition must have probability zero at $\widehat{x}$ (and thus $\Xi(\widehat{x}) > \Xi(2)$). If $f_2 > f_3$ it can be seen analogously that $\left(\partial \dot{q}_2^i / \partial q_2^i\right) > 0$ and if $f_3 > \max\{f_1, f_2\}$ we have that either $\left(\partial \dot{q}_3^i / \partial q_3^i\right) > 0$ or $\left(\partial \dot{q}_2^i / \partial q_2^i\right) > 0$ depending on the particular restpoint. What happens at restpoints that involve $q_4^i > 0$, for some $i = 1, 2$? If $f_3 > \max\{f_1, f_2\}$ we have that $\left(\partial \dot{q}_5^i / \partial q_5^i\right) > 0$ and if $f_3 > \max\{f_1, f_2\}$ either $\left(\partial \dot{q}_5^i / \partial q_5^i\right) > 0$ or a strictly better response will be played on average in either G_2 or G_3 (depending on the particular restpoint) that have the same reasoning costs. Thus G_1 or G_4 cannot be in the support of a stable restpoint.

Neither can a stable restpoint involve $q_l^{i*} > 0$ for $l = 2, 3$. Distinguish two cases: If $f_3 > \min\{f_1, f_2\}$, player 1 will play a fully mixed strategy $p_4^* = 1/2$ in $g_4 = \{\gamma_2, \gamma_3\}$ and player 2 will play the mixed strategy $p_5^* = 1/2$ in analogy class $g_5 = \{\gamma_1, \gamma_3\}$. It then follows immediately by arguments analogous to those above that $G_2 \notin \operatorname{supp} q^{1*}$ and $G_3 \notin \operatorname{supp} q^{2*}$. Furthermore note that any restpoint at which player 2 holds partition G_2 and player 1 partition G_3 cannot induce Nash play in all games and thus (by Proposition 2) cannot be asymptotically stable. If $f_3 < \min\{f_1, f_2\}$, either the same action probabilities are chosen in both analogy classes contained in G_3 (G_2) at any restpoint that induces Nash equilibrium play. But again this has as a consequence that $\left(\partial\dot{q}_5/\partial q_5\right) > 0$. Or there exist partitions where a strictly better response is played on average and thus G_l cannot be in the support of $q^{i*}$for $l = 2, 3$ at any stable restpoint.

(ii) The payoff matrix across all three games is

$$\begin{pmatrix} 2(f_1+f_3)+f_2 & 2f_2+f_1+f_3 \\ 2f_2+f_1+f_3 & 2(f_1+f_3)+f_2 \end{pmatrix} \text{ for player 1} \tag{21}$$

and

$$\begin{pmatrix} 2f_1+f_2+f_3 & 2(f_2+f_3)+f_1 \\ 2(f_2+f_3)+f_1 & 2f_1+f_2+f_3 \end{pmatrix} \text{ for player 2.} \tag{22}$$

Given the assumption that $f_j < 1/2$ for $j = 1, 2$ - (21) and (22) represent a conflict game with a unique Nash equilibrium in mixed strategies given by $(1/2, 1/2)$. Now we will show that (holding fixed $q_5^* = 1$) this equilibrium is asymptotically stable in the game (21) and (22).[41] Denote p^i the probability with which player i chooses the first action. The Jacobian matrix associated with the linearization of the perturbed dynamics at the equilibrium $\left(p^1, p^2\right) = (\frac{1}{2}, \frac{1}{2})$ is given by

$$\mathcal{M}_{(\frac{1}{2},\frac{1}{2})} = \begin{pmatrix} -2\varepsilon_0 & \frac{1}{2}(f_1+f_3-f_2) \\ \frac{1}{2}(f_1-f_2-f_3) & -2\varepsilon_0 \end{pmatrix}.$$

It can be verified easily that the spectrum of $\mathcal{M}_{(\frac{1}{2},\frac{1}{2})}$ is given by

$$\{\lambda_1, \lambda_2\} = \left\{ \frac{1}{2}(-4\varepsilon_0 \pm \sqrt{(f_1+f_3-f_2)(f_1-f_2-f_3)+16\varepsilon_0^2}) \right\}.$$

Given our assumptions on f_i the term under the square root is negative and thus $\mathrm{Re}\{\lambda_i(\varepsilon_0)\} < 0, \forall i = 1, 2$. Note also that as $(1/2, 1/2)$ is a Nash equilibrium in all games there is no analogy class in which a player i has a strictly better response to the opponent choosing $p^{-i} = 1/2$. But then as $q_5 = 1$ minimizes reasoning costs and sign $[O(\varepsilon_0)] \gtreqless 0 \Leftrightarrow p^i_{mk} \lesseqgtr \frac{1}{2}$ we know that x^* is asymptotically stable. ■

Proof of Proposition 4:

Proof. We will prove Proposition 4 for the case of a Coordination game. The case of Anti-coordination games is analogous. Unless otherwise indicated we will use the notation from Example 1.

(i) As $\operatorname{card}\Gamma = 1$ there is trivially only one partition and one analogy class $g = \gamma_1$. Consider again the symmetric game from example 1 in (8) with matrix

$$\gamma_1 : \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix}. \tag{23}$$

[41] Note that because of the perturbation equilibria are generically hyperbolic. This makes the proof in the perturbed case considerably easier then the proof of the unperturbed case. Posch (1997) has shown that unperturbed reinforcement learning leads to cycling in this class of games.

If $a_1 > a_3$ and $a_4 > a_2$ this is a Coordination game with two strict Nash equilibria where both players choose the same action. Consider the system of differential equations $(\dot{p}^1_H(p^1_H, p^2_H), \dot{p}^2_H(p^1_H, p^2_H))$ defined through (9) and (10). The Jacobian matrix associated with the linearization of the perturbed dynamics at the equilibrium $\left(p^1_H, p^2_H\right) = (1,1)$ is given by

$$\mathcal{M}^{\gamma_1}_{(1,1)} = \begin{pmatrix} -(a_1 - a_3) - 2\varepsilon_0 & 0 \\ 0 & -(a_1 - a_3) - 2\varepsilon_0 \end{pmatrix}.$$

Clearly this matrix is negative definite and the strict equilibrium $\left(p^1_H, p^2_H\right) = (1,1)$ asymptotically stable. Stability of $\left(p^1_H, p^2_H\right) = (0,0)$ is shown in exactly the same way.

(ii) Let the 2×2 game γ_2 have the payoff matrix

$$\gamma_2 : \begin{pmatrix} b_1, c_1 & b_2, c_3 \\ b_3, c_2 & b_4, c_4 \end{pmatrix}. \tag{24}$$

To have a mixed equilibrium that is stable to learning in a single game the game must be one of conflict. Assume wlg $b_1 > b_3, b_4 > b_2, c_1 < c_3$ and $c_4 < c_2$. The payoff matrix across games is given by

$$\begin{pmatrix} f_1a_1 + f_2b_1, f_1a_1 + f_2c_1 & f_1a_2 + f_2b_2, f_1a_3 + f_2c_3 \\ f_1a_3 + f_2b_3, f_1a_2 + f_2c_2 & f_1a_4 + f_2b_4, f_1a_4 + f_2c_4 \end{pmatrix}. \tag{25}$$

Whenever $f_1/f_2 < (c_3 - c_1)/(a_1 - a_3) =: \widehat{f}$ this matrix represents a game of conflict. Think of restpoints that induce the strict Nash equilibrium $\left(\sigma^1_{H1}, \sigma^2_{H1}\right) = (1,1)$ in game γ_1. If at such a restpoint the coarse partition G_C is used with asymptotic probability $q^*_C > 0$, then we need to have $p^i_{H3} = p^i_{H1} = 1$.[42] In order to induce Nash equilibrium also in game γ_2 one needs $p^i_{H2} = 0$ and $q^{i*}_C = \sigma^{i*}_{H2}$. But then $p^i_{HC} = 1$ is not a best response to the observed play of player $-i$ for at least one of the two players i. Consequently there cannot be such a restpoint. What about restpoints in which the fine partition is used with asymptotic probability 1 ? Then for at least one player action choice in the "off equilibrium" analogy class $g_C = \{\gamma_1, \gamma_2\}$ will be a best response to observed play in both games γ_1 and γ_2. As the coarse partition has smaller reasoning cost, the diagonal element $\left(\partial \dot{q}_C / \partial q_C\right) = \Xi(2) - \Xi(1) - 2\varepsilon_1$ of the Jacobian matrix at this equilibrium is strictly positive and the equilibrium thus unstable. ■

Proof of Proposition 5:

Proof. (i) Again as card $\Gamma = 1$ there is trivially only one partition and one analogy class $g = \gamma_1$, where γ_1 is given by (23). The Jacobian matrix $\mathcal{M}$ at the mixed equilibrium $\widehat{\sigma}^1_{H1} = \frac{a_4 - a_3}{a_1 + a_4 - (a_2 + a_3)} = \widehat{\sigma}^2_{H1}$ is given by

$$\mathcal{M} = \begin{pmatrix} -2\varepsilon_0 & \widehat{\sigma}^1_{H1}(1 - \widehat{\sigma}^1_{H1})(a_1 + a_4 - (a_2 + a_3)) \\ \widehat{\sigma}^2_{H1}(1 - \widehat{\sigma}^2_{H1})(a_1 + a_4 - (a_2 + a_3)) & -2\varepsilon_0 \end{pmatrix}.$$

[42] Remember from example 1 that $g_3 = \{\gamma_1, \gamma_2\}$ and $g_1 = \{\gamma_1\}$.

The characteristic polynomial is given by

$$(-2\varepsilon_0 - \lambda)(-2\varepsilon_0 - \lambda) - \left[\sigma^i_{H1}(1-\sigma^i_{H1})(a_1+a_4-(a_2+a_3))\right]^2 = 0.$$

Consequently the two eigenvalues are $\lambda_{1/2} = -2\varepsilon_0 \pm \sigma^i_{H1}(1-\sigma^i_{H1})(a_1+a_4-(a_2+a_3))$. Thus $\mathcal{M}$ has an eigenvalue with $\lim_{\varepsilon_0 \to 0} \mathrm{Re}\{\lambda_i(\varepsilon_0)\} > 0$ and $\widehat{\sigma}_1$ is unstable.

(ii) Let the 2×2 game γ_2 be again the game described in (24). We also assume that $\widehat{\sigma}^i_{H1} = \frac{a_4-a_3}{a_1+a_4-(a_2+a_3)} = \frac{b_4-b_3}{b_1+b_4-(b_2+b_3)} = \frac{c_4-c_3}{c_1+c_4-(c_2+c_3)}$, i.e. that both games have the same mixed strategy equilibrium. The mixed strategy equilibrium of the strategic form (25) is given by $\frac{f_1(a_4-a_3)+f_2(b_4-b_3)}{f_1(a_1+a_4-(a_2+a_3))+f_2(b_1+b_4-(b_2+b_3))} = \frac{a_4-a_3}{a_1+a_4-(a_2+a_3)}$. Consider the restpoint where both players hold the coarse partition and choose $p^i_{HC} = \frac{a_4-a_3}{a_1+a_4-(a_2+a_3)}$. This restpoint is asymptotically stable whenever $f_1/f_2 < \widehat{f}$ as can be shown in analogy to the proof of claim 2. ■

Proof of Proposition 6:

Proof. Consider any partition $G_l \neq G_F$. As G_l is not the finest partition there are two games, denote γ_1 and γ_2 that are seen as analogous and for which the same action choice is made. As $S^{Nash}(\gamma_1) \cap S^{Nash}(\gamma_2) = \varnothing$ by assumption no Nash equilibrium is played in at least one of the two games. It follows from Proposition 2 that $q^*_l = 0$. Consequently if $S^{Nash}(\gamma_j) \cap S^{Nash}(\gamma^{'}_j) = \varnothing$, $\forall \gamma_j, \gamma'_j \in \Gamma$ only restpoints that place probability one on the finest partition can be asymptotically stable. It is clear that if $\exists \gamma_1, \gamma_2 \in \Gamma$ s.t. $S^{Nash}(\gamma_1) \cap S^{Nash}(\gamma_2) \neq \varnothing$ the finest partition need not necessarily arise. Examples where this is the case have been analyzed above. ■

Proof of Proposition 7:

Proof. Consider a restpoint with partition $G_h = \{\{g_h\}^{Z_h-2}_{h=1}, g^{'}_k, g_k\}$ where $p^i_{mk} \in BR^i(\sigma^{-i}_j)$, $\forall \gamma_j \in \left(g^i_k\right)'$, i.e. where the action choice in analogy class g_k is a best response to the opponent's observed play in all games contained in g'_k. Consider the alternative partition $G_l = \{\{g_h\}^{Z_h-2}_{h=1}, g_k \cup g^{'}_k\}$, that differs from G_h only in the fact that the two analogy classes g_k and $g^{'}_k$ are merged, i.e. $\mathrm{card}\, G_l = (\mathrm{card}\, G_h) - 1$. But then $\forall \xi > 0 : \Pi_l(x^*) - \Pi_h(x^*) > 0$ and consequently G_h cannot be in the support of q^* at any stable restpoint $x^* = (p^*, q^*)$. Proposition 7 now follows from the observation that whenever $\mathrm{card}\, G_h > \mathrm{card}\, A_i$, there necessarily exist analogy classes $g^i_k \neq g^i_{k'} \in G_h$ with $p^i_{mk} \in BR^i(\sigma^{-i}_j), \forall \gamma_j \in g^i_{k'}$. ■